Social identity shapes social valuation: evidence from prosocial behavior and vicarious reward

Leor M. Hackel,[1] Jamil Zaki,[1] and Jay J. Van Bavel[2]

[1]Department of Psychology, Stanford University, Stanford, CA, USA and [2]Department of Psychology and Center for Neural Sciences, New York University, New York, NY, USA

Correspondence should be addressed to Jay J. Van Bavel, Department of Psychology, New York University, 6 Washington Place, New York, NY 10003, USA. E-mail: jay.vanbavel@nyu.edu

Abstract

People frequently engage in more prosocial behavior toward members of their own groups, as compared to other groups. Such group-based prosociality may reflect either strategic considerations concerning one's own future outcomes or intrinsic value placed on the outcomes of in-group members. In a functional magnetic resonance imaging experiment, we examined vicarious reward responses to witnessing the monetary gains of in-group and out-group members, as well as prosocial behavior towards both types of individuals. We found that individuals' investment in their group—a motivational component of social identification—tracked the intensity of their responses in ventral striatum to in-group (*vs* out-group) members' rewards, as well as their tendency towards group-based prosociality. Individuals with strong motivational investment in their group preferred rewards for an in-group member, whereas individuals with low investment preferred rewards for an out-group member. These findings suggest that the motivational importance of social identity—beyond mere similarity to group members—influences vicarious reward and prosocial behavior. More broadly, these findings support a theoretical framework in which salient social identities can influence neural representations of subjective value, and suggest that social preferences can best be understood by examining the identity contexts in which they unfold.

Key words: social identity; prosociality; vicarious reward; intergroup relations; fMRI

Introduction

In 2014, alumni donations to U.S. universities reached $9.85 billion, surpassing contributions from non-alumni individuals and corporations by billions of dollars (Council for Aid to Education, 2015). This discrepancy provides one example of *group-based giving*—people's preference for prosocial behaviors that help members of their own communities (Bernhard *et al*., 2006). Classic work in social psychology indicates that group membership influences prosocial behavior. People often allocate more money to members of their in-group than members of an out-group (Tajfel *et al*., 1971), and are more likely to cooperate with others when shared group identity is salient (Brewer and Kramer, 1986). Although other primates also engage in group-

based prosociality, humans uniquely extend prosociality to anonymous in-group members (Burkart *et al*., 2009). Social groups, in other words, allow strangers to have prosocial interactions with one another. The current research was designed to better understand the psychological structure of group based prosociality.

Despite the ubiquity of group-based prosociality, it remains unclear why some people give more to in-group members. On one hand, people may feel obligated to group members or foresee in-group members reciprocating prosociality, leading to group-based giving that reflects strategic motives. Strategic models of intergroup bias have provided evidence that concerns about reputation lead to greater in-group altruism (Mifune *et al*., 2010), and expectations about reciprocity lead to greater in-

group cooperation (Yamagishi and Kiyonari, 2000). By highlighting conditions under which group-based prosociality is more likely to take place, these models have offered important insights into one potential source of intergroup behavior.

There is also reason to believe identification with social groups—termed social identification—may itself be motivationally meaningful, potentially imbuing the well-being of in-group members with intrinsic value. Extensive research suggests that social identification represents a key source of social motivation in humans, providing not only material resources and opportunities for social interaction but also psychological benefits such as self-esteem and a sense of belonging (Allport, 1954; Brewer, 1988; Baumeister and Leary, 1995; Correll and Park, 2005; Kurzban and Neuberg, 2005; Cikara and Van Bavel, 2014). Economic models suggest that people will pay more for outcomes that fulfill the norms of their social groups (Akerlof and Kranton, 2000), and social psychological models highlight identity as a motivating factor in decision-making and consumer preferences (Wicklund and Gollwitzer, 1982; Braun and Wicklund, 1989; Reed, 2004; Oyserman et al., 2007; Oyserman, 2009). These models suggest that people might feel better when fellow in-group members, as opposed to out-group members, benefit.

At the same time, research on identification suggests that not all members of a group will show such an in-group preference. People differ in the degree to which they identify with different social groups, and strength of identification determines the impact of identity on cognition and behavior (Ashmore et al., 2004). Indeed, people often do not want to be categorized as part of a group, either because they do not strongly identify with a particular group or because they want to be treated as a unique individual in a particular situation (Branscombe et al., 1999). When group identity is made salient, individuals low on social identification may experience negative emotions, distance themselves from the group, and even 'put down' other in-group members (Branscombe et al., 1999). These responses still demonstrate that identities are motivationally salient—since they represent affective responses to group categorizations—but also demonstrate that group-based responses can be positive or negative, depending on an individual's degree of identification. Therefore, an individual's degree of identification with a group should predict whether or not they value the outcomes of in-group members over out-group members.

Identification has been further decomposed into multiple components. For instance, 'group-level self-investment' indicates the extent to which people find group membership motivationally significant. Alternatively, 'group-level self-definition', indicates the extent to which people see themselves as similar to the group and group members as similar to one another (Leach et al., 2008). Consistent with the idea that motivational components of identity shape valuation, recent work found that group-level self-investment—but not self-definition—predicted evaluation of identity-relevant foods (Hackel et al., 2016). Individuals from the Southern United States with high group-level self-investment expected Southern foods (e.g. chicken fried steak) to be tastier than non-Southern foods (e.g. pizza). In contrast, Southerners with low group-level self-investment expected Southern foods to be *less* tasty than non-Southern foods, expressing a *negative* evaluation of the in-group. Identity motivations might similarly shape valuation of social outcomes: those who are strongly invested in a social group may prefer the well-being of in-group members, whereas those who dis-identify with a group may feel more negatively about the in-group's positive outcomes.

This work raises the possibility that group-based prosociality represents not only strategic concerns, but also the motivational effects of identification on social valuation. Motivational factors—including but not limited to group membership—broadly shape people's empathy towards others (Davis et al., 1999; Hein, et al., 2010; Cameron and Payne, 2011; Gutsell and Inzlicht, 2012; Cikara et al., 2014; Zaki, 2014). Thus, experiences of vicarious reward might depend on the motivational relevance of group members. This perspective leads to three predictions. First, individual differences in social identification should predict differential responses to in-group and out-group gains: those strongly identified with a group should prefer in-group gains in an intergroup context, whereas those weakly identified with a group should not. Second, motivational components of identification in particular should be associated with group-based biases, beyond effects of perceived similarity to others. Third, social identification should shape neural vicarious reward responses even when people merely witness the gains of other people—indicating that strategic explanations cannot fully account for group-based social preferences.

Recent neuroscience evidence suggests that people intrinsically value social outcomes (Zaki and Mitchell, 2011; Ruff and Fehr, 2014; Nook and Zaki, 2015). Brain regions commonly associated with value processing also respond during prosocial acts and when seeing other people receive rewards. When allocating money to the self or another person, prosocial decisions are associated with activation in ventral striatum and ventromedial prefrontal cortex (vmpFC)/medial orbitofrontal cortex (mOFC) (Rilling et al., 2002; Hare et al., 2010; Zaki and Mitchell, 2011; Zaki et al., 2014). For example, Zaki and Mitchell (2011) examined neural responses as people choose whether to allocate one amount of money to themselves or a different amount of money to another person. Sometimes, the participant stood to gain more than the other person (e.g. choosing between $3 for the self *vs* $1 for the other), and sometimes the other person stood to gain more than the participant (e.g. choosing between $1 for the self *vs* $3 for the other). Responses in mOFC were strongest when participants made prosocial choices, giving money to whichever person stood to gain the most—that is, when choices were 'efficient'. (Although Zaki and Mitchell referred to 'equitable' choices, these choices are better described as 'efficient', since they maximize total payouts to both players rather than equity between players.) In contrast, mOFC responses were weaker when participants kept money for themselves and the other person stood to gain more.

Both ventral striatum and vmPFC/mOFC have been linked to the computation of a common neural currency of value across multiple types of goods (Levy and Glimcher, 2012; Clithero and Rangel, 2014), including social goods (Ruff and Fehr, 2014). These regions have also been linked to vicarious reward processing when witnessing the outcomes of others (Morelli et al., 2015; Sul et al., 2015). In this manner, neuroimaging offers insight into subjective value that may not be fully accessible to behavior or self-report alone (Zaki et al., 2011). Moreover, since behavior and self-reports may be more susceptible to strategic and self-presentational motives, studying neural responses to the outcomes of others can more directly index the intrinsic value placed on those outcomes.

Based on this research, it is possible that people motivationally invested in a group preferentially value the outcomes of in-group members, and that this preferential value is reflected in ventral striatum activation. Recent research has indeed linked intergroup biases in evaluation to neural reward systems. For example, biased interpersonal evaluation of in-group members has been associated with activation in mOFC (Van Bavel et al., 2008). In other research, witnessing the success of a favored

team or the failure of a rival team was associated with activity in the ventral striatum (Cikara *et al.*, 2011). Moreover, individuals primed with an interdependent (*vs* independent) self-representation show greater vicarious reward responses in ventral striatum upon witnessing a friend win money (Varnum *et al.*, 2014). This body of evidence supports the idea that social identification may alter valuation of the outcomes of others.

Overview

Here, we used functional magnetic resonance imaging (fMRI) to test whether motivational components of identification shape social valuation. We hypothesized that people who are motivationally invested in a social group would more highly value the outcomes of in-group members, as reflected in prosocial behavior and neural markers of vicarious reward. To test this hypothesis, New York University (NYU) students witnessed monetary gains of two confederates who they believed were from NYU (in-group) and Columbia University (out-group). Participants also had opportunities to allocate money privately to themselves versus those targets. Afterwards, participants completed a measure of collective identification with NYU students, which included motivational components (group-level self-investment) and perceived similarity to the group (group-level self-definition). This design allowed us to test whether motivational components of identification predict social valuation for in-group versus out-group members, using converging behavioral and brain data. A purely strategic account would not predict differential giving to in-group and out-group members under private conditions, when reputational concerns are not applicable. At the same time, mixed motives can influence behavior—for instance, a motive to appear egalitarian to an experimenter might influence allocation decisions. Therefore, by examining neural responses in the absence of behavior, we were able to study responses even less likely to be affected by strategic concerns, providing a cleaner test of the intrinsic value hypothesis.

Method

Participants

Fifty New York University students (17 males; mean age = 20.90, s.d. = 3.23) participated in exchange for $30 plus the opportunity to earn a monetary bonus. The sample size was determined by collecting the largest sample possible given budgetary constraints (we sought to at least double the normal sample size of neuroimaging studies). All participants were right-handed and had normal or corrected-to-normal vision. Five participants were excluded from analyses because they did not believe the cover story. An additional participant was excluded from analysis because the participant chose to end the session early, leaving 44 participants for analysis. For three of these participants, one run was excluded from analyses of neuroimaging data (but not behavioral data) due to a scanner error. Participants gave informed consent in accordance with approval from the NYU University Committee on Activities Involving Human Subjects.

Procedure

During recruitment, participants were informed that the study was open to students from three New York area universities (New York University, Columbia University, and Hunter College). In reality, only NYU students were invited to complete the study. Upon arriving to the session, participants met two confederates posing as other participants. The participant and confederates each had their picture taken and filled out a personal profile listing their names, ages, and university affiliations. Through a rigged drawing, the true participant was assigned to complete a decision-making task in the scanner, while the confederates were dismissed to another room to complete a supposed face memory task.

Participants were informed that they would make decisions impacting monetary outcomes for themselves and the two confederates. As a performance incentive, participants were informed that five trials would be randomly selected and monetary bonuses would be rewarded to themselves or the confederates based on their responses. This ensured that they would treat trials seriously (Zaki and Mitchell, 2011).

This task involved two types of trials (Figure 1). In 'free choice' trials, participants made decisions to allocate money either to themselves or to one of the confederates (Zaki and Mitchell, 2011). In order to minimize reputational concerns, instructions emphasized that the confederates would have no knowledge of these decisions, but would simply be paid an extra bonus without explanation. On each free choice trial, participants saw their own face on one side of the screen and a confederate's face on the other side of the screen. Each face was associated with a different amount of money listed above it, and participants could press a button on a button box to make a monetary allocation to themselves or the confederate. To determine the amounts available, a random number (no greater than $4.00) was generated as the amount available for the participant. This number was then multiplied by a predetermined ratio to generate the amount available for the confederate. The ratios used to study costly choice (0.33, 0.5, 0.67, 1, 1.5, 2, 3) were symmetric around a 1:1 ratio, such that participants sometimes stood to gain more money and the confederate sometimes stood to gain more money. Participants also saw costless free choice trials in which they could choose between zero dollars for the self and a randomly determined amount for a confederate, or vice versa.

In 'forced choice' trials, participants saw only one face (self, in-group member, or out-group member) appear onscreen with an associated amount of money (randomly drawn with a maximum of $4.00). These trials indicated that the experimenter had assigned this amount to the target shown. Participants were asked to press a button to acknowledge this assignment, but could not make any decisions to influence these outcomes.

Altogether, participants completed four runs of 36 trials each in the scanner, divided into two in-group runs and two out-group runs; run order was randomized across subjects. Before each run began, a cue screen showed the photo of the confederate who would be viewed in that run as well as the profile they had ostensibly filled out. This cue indicated whether the target was an in-group member (NYU student) or out-group member (Columbia student). Which confederate was associated with which university was counterbalanced across participants, ensuring that effects of group membership could not be due to differences between the confederates. Confederates were blind to condition.

Across all runs, participants witnessed 12 'forced' trials in which the in-group member won money, 12 'forced' trials in which the out-group member won money, and 24 'forced' trials in which they themselves won money (due to an equal number of wins for self and other when playing with each target). Participants completed a total of 48 free choice trials involving each target, evenly divided across each ratio used (including zero). Forced choice trials and free choice trials were randomly intermixed in a different order for each subject. Following past work using a similar task (Zaki and Mitchell, 2011), each trial appeared onscreen for an initial "offer" phase lasting 2–8 s
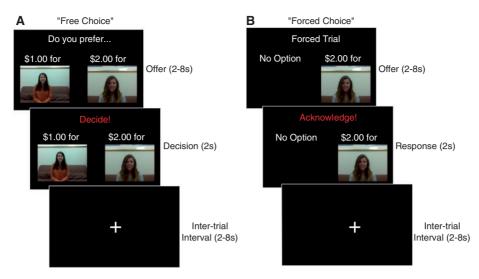
Fig. 1. Participants saw two types of trials involving monetary outcomes for themselves or another person. A cue before each block indicated whether the other person would be an in-group member (NYU student) or out-group member (Columbia student). (**A**) On 'free choice' trials, participants chose between one amount of money for themselves or a different amount of money for another person. (**B**) On 'forced choice' trials, participants passively witnessed a monetary gain issued to themselves, the in-group member, or the out-group member. In both trial types, an offer stage (2–8 s) was followed by a response stage (2 s). Images are not to scale.

(jittered to allow estimation of the hemodynamic response). Afterwards, participants had a 2-s response phase in which to enter a decision (free choice trials) or acknowledge an allocation (forced choice trials). Trials were followed by a jittered inter-trial interval (2–8 s). Stimuli were projected onto a screen at the rear of the magnet, which participants viewed through an angled mirror attached to the RF coil.

After scanning, participants completed a measure of identification with the in-group (Leach *et al.*, 2008). They indicated how strongly they agreed with 14 statements on a 5-point Likert scale anchored at 1 (*strongly disagree*) to 5 (*strongly agree*) ($\alpha = 0.86$, $M = 3.68$, s.d. $= 0.81$). This scale included two broad sub-components of identification. The first component, group-level self-investment ($\alpha = 0.87$, $M = 3.41$, s.d. $= 0.75$), measured the extent to which participants felt invested in the group, examining motivational components of identification (e.g. 'The fact that I am an NYU student is an important part of my identity'). The second component, group-level self-definition ($\alpha = 0.77$, $M = 2.14$, s.d. $= 0.74$), measured the extent to which participants defined themselves with reference to fellow group members, examining perceived similarity to group members and of group members to one another (e.g. 'I am similar to the average NYU student'). For the measure of group-level self-investment, 23% of responses fell below the midpoint, indicating disagreement with statements assessing investment in the group. For the measure of group-level self-definition, 86.4% of responses fell below the midpoint, indicating disagreement with statements assessing self-definition with the group. Finally, participants completed demographic information, were debriefed, and were paid their base pay plus a bonus computed by randomly instantiating the outcomes of five trials.

Image acquisition. Images were acquired with a 3T Siemens Allegra head-only scanner. Functional images (TR = 2000 ms; effective TE = 30 ms; flip angle = 82, 34.3 mm slices with a 0.45 mm gap for whole-brain coverage, matrix = 80 × 64; FOV = 240 × 192 mm; acquisition voxel size = 3 × 3 × 3.45 mm) were acquired using a customized multi-echo EPI sequence developed by the NYU Center for Brain Imaging to mitigate the effects of susceptibility artifacts. Five fixation scans were acquired at the start of each run and dropped from analysis to allow for magnet equilibration. Slices were collected parallel to the AC-PC line. Finally, T1-weighted high-resolution anatomical images (MPRAGE, 1 × 1 × 1 mm) were acquired for each subject for registration and group normalization purposes.

Image processing and analysis. Data were preprocessed and analyzed using SPM8 (Wellcome Department of Cognitive Neurology, London, United Kingdom). Data were corrected for slice-time acquisition and realigned to correct for subject motion, co-registered to structural images, transformed to conform to the default T1 Montreal Neurological Institute (MNI) brain interpolated to 3 × 3 × 3 mm, smoothed using a 6-mm full-width/half-maximum Gaussian kernel, corrected for artifacts (Mazaika *et al.*, 2007) and detrended (Macey *et al.*, 2004). The blood-oxygenation-level-dependent (BOLD) signal was modeled using a canonical hemodynamic response function, and filtered with a 128-s high-pass filter.

We first aimed to replicate prior findings regarding prosocial choices. Zaki and Mitchell (2011) observed increased activation in mOFC when people made efficient, as opposed to inefficient, choices—that is, when people allocated money to whomever stood to gain the most. We analyzed the decision phase (2 s) of costly choice trials, including onsets of these trials for in-group members and out-group members in separate conditions in a general linear model (GLM). These onsets were parametrically modulated by the (log) ratio of the amount available for the chosen versus unchosen recipient. (We took the logarithm of the ratios in order to linearize ratios around 1:1.) This parametric regressor served as a trial-by-trial measure of the efficiency of a choice, regardless of whether an allocation benefited the self or the other. For example, if a participant chose to keep $2.00 for the self instead of allocating $1.00 to another, this choice was more efficient (ratio = 2) than keeping $1.00 for the self instead of allocating $2.00 for another (ratio = 0.5).

This analysis strategy differed somewhat from prior work. Zaki and Mitchell (2011) divided choices into separate trial types based on whether choices were generous and efficient, self-serving and efficient, or self-serving and inefficient. We were unable to reproduce this exact analysis strategy because we had

fewer trials in our experiment, in order to shorten scan time and collect a larger sample to increase between-subject power. In addition, our trials were further sub-divided across in-group and out-group members. Had we followed this strategy for in-group and out-group trials, 18 participants would have zero trials in at least one cell, and a total of 29 participants would have fewer than three trials in at least one cell. However, the analysis we report is conceptually analogous: it examines the effect of choice efficiency while ignoring whether choices were self-serving or generous. Since it does not divide choices into different trial types, it does not suffer from data loss due to low trial numbers.

Next, to identify neural regions associated with vicarious reward to in-group versus out-group members, we included onsets of 'forced choice' trial offer stages in the GLM, modeled as impulses with zero duration following prior studies of reward receipt (Daw *et al.*, 2011; Doll *et al.*, 2015). Onsets were entered in separate conditions according to the identity of the target (self, in-group member, out-group member) and were parametrically modulated by the dollar amount won by the target indicated. To account for other task-related variance, we modeled onsets of response stages for forced choice trials of each condition, offer stages for costly choice trials of each condition, and response and offer stages for costless choice trials and any non-response trials. We concatenated data across runs, using the spm_fmri_concatenate function in SPM, and included a regressor of no interest indicating the start of new runs. Finally, we included the six motion covariates produced during realignment as nuisance regressors.

To examine neural correlates of prosocial choice, we created first-level contrasts of the parametric regressor of choice efficiency described above, and entered these contrasts into a whole-brain random-effects analysis. We corrected for multiple comparisons using cluster correction with a cluster-defining threshold of $P < 0.005$ and a cluster extent required to maintain a family wide error rate of $P < 0.05$, based on Gaussian random field theory as implemented in SPM (Friston *et al.*, 1994). Next, we separately interrogated responses to efficiency for in-group and out-group members in this region of interest (ROI) by extracting parameter estimates to in-group and out-group separately. To ensure independence between ROI definition and the subsequent test, we used a leave-one-subject-out procedure to extract parameter estimates (Esterman *et al.*, 2010): we recomputed the whole-brain random effects analysis once while leaving out each subject, and used this definition of the ROI to extract parameter estimates for the left-out subject.

For 'forced choice trials', our goal was to interrogate the relationship between social identification and vicarious reward responses. Therefore, we aimed to create a functional region of interest that showed increasing activity as participants won more money for themselves. We created first-level contrasts of the parametric modulator of amount won on 'forced choice' wins for the self, during the offer stage. Contrasts were entered into a whole-brain random effects analysis. Given strong a priori hypotheses about reward processing in ventral striatum (Clithero and Rangel, 2014; Garrison *et al.*, 2013), we examined results within an anatomical mask of ventral striatum. This mask was generated by obtaining a mask of the caudate and putamen from the Automated Anatomical Labeling atlas (Tzourio-Mazoyer *et al.*, 2002) and maintaining only ventral portions, defined as $z < 2$ for putamen and $z < 7$ for caudate (Postuma and Dagher, 2006). However, this contrast revealed only a cluster of 3 voxels in the right caudate and 5 voxels in the left putamen, which did not meet the requirements of a small

volume correction. Therefore, as an alternative method of constructing the ROI, we created a reverse inference map of the term 'reward' using the meta-analysis website Neurosynth (www.neurosynth.org; Yarkoni *et al.*, 2011). The resulting map was based on 617 studies. We intersected this map with our anatomical mask of ventral striatum.

To interrogate responses in this region to the monetary gains of others, we separately extracted parameter estimates in this region corresponding to the parametric modulators of 'amount won' for in-group and out-group trials, using the Marsbar toolbox (Brett *et al.*, 2002). That is, we examined how signal in this region changed in response to trial-by-trial changes in the monetary gains of in-group members and out-group members. We analyzed these parameter estimates as a function of target group and social identification (group-level self-investment and group-level self-definition). De-identified data have been made available at https://osf.io/6qbc4/.

Results

Prosocial behavior

We tested whether social identification shaped prosocial behavior toward each target. Specifically, we calculated the proportion of trials on which participants made altruistic decisions towards the in-group and out-group member and analyzed choices as a function of self-reported social identification with NYU. Since we had a continuous predictor (social identification) and a within-subjects factor (target group membership), we used Generalized Estimating Equations, which fits a marginal model that accounts for repeated measures in a regression framework (Liang and Zeger, 1986).

We hypothesized that strength of identification would predict differential giving to in-group and out-group members, even under private conditions that were likely to minimize reputational (and therefore strategic) concerns. In order to test whether group-based behavior was particularly driven by motivational importance of the group or by perceived similarity to others, we separated the identification scale into its components of group-level self-investment and group-level self-definition. We simultaneously entered both components (mean-centered) as predictors in one model in which they each interacted with target group membership. (Self-investment and self-definition were moderately correlated, $r = 0.35$.) We found a target group × self-investment interaction, $b = -0.02$, $SE = 0.01$, Wald $\chi^2 = 7.29$, $P = 0.007$ (Figure 2; Table 1). Those high on group level self-investment showed marginally significant favoritism toward the in-group, $b = -0.01$, $SE = 0.01$, Wald $\chi^2 = 3.30$, $P = 0.07$, whereas those low on group-level self-investment showed favoritism toward the out-group, $b = 0.02$, $SE = 0.01$, Wald $\chi^2 = 4.82$, $P = 0.03$. Group-level self-definition did not interact with target group ($P = 0.21$). These results indicate that motivational components of identification, rather than perceived similarity, shaped relative giving to in-group and out-group targets.

Neural correlates of prosocial choice

We next examined neural correlates of prosocial choices. We conducted a whole-brain analysis seeking regions showing parametrically increasing responses to the efficiency of choices. Again, efficiency was defined as the ratio of money available for the chosen versus unchosen beneficiary of an allocation, regardless of whether money was allocated to the self or another. Replicating findings of Zaki and Mitchell (2011),

increasing choice efficiency was associated with activation in mOFC (Figure 3A; Table 2). In other words, mOFC displayed greater activation when allocations favored a person who stood to gain more. By contrast, inefficient choice was associated with activation in anterior cingulate cortex (ACC), which has been found to track negatively with the value of a decision in past work (Doll *et al.*, 2015).
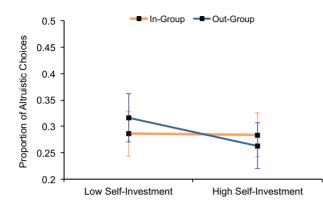

Fig. 2. Motivational components of group identification predicted prosocial behavior. Those high on group-level self-investment gave relatively more to in-group members, whereas those low on group-level self-investment gave relatively more to out-group members. Plot shows predicted proportions of giving from a regression analysis. Predicted proportions for high and low self-investment are plotted one standard deviation above and below the mean, respectively.

Table 1. Coefficients in regression of prosocial choice

Effect	b	SE	Wald χ^2	P
Intercept	0.287	0.03	91.81	<0.0001
Target Group	0.002	0.004	0.26	0.61
SI	−0.02	0.04	0.20	0.65
SD	−0.02	0.04	0.117	0.73
Group × SI	−0.02	0.007	7.29	0.007
Group × SD	0.008	0.006	1.55	0.21

Note: SI, group-level self-investment; SD, group-level self-definition.

We further extracted parameter estimates to efficiency from OFC for in-group and out-group trials separately, and entered these parameter estimates into a regression analysis using GEE. Specifically, we predicted parameter estimates as a function of target group membership, group-level self-investment, and group-level self-definition, as described in our behavioral analyses. We did not observe an interaction of group-level self-investment with target group membership, $b = -0.04$, $SE = 0.07$, Wald $\chi^2 = 0.27$, $P = 0.61$ (Table 3). However, follow-up analyses revealed that this null effect in mOFC was consistent with behavior. We expanded our GEE regression of behavior to analyze choices on each trial (0 = self-serving, 1 = prosocial) in a logistic model (Table 4). We included the log of each trial's ratio (other amount *vs* self-amount) as a predictor, as well as the interactions of ratio with target group membership and each identification scale. We observed two distinct effects: ratios predicted prosocial choices ($b = 1.89$, $SE = 0.24$, Wald $\chi^2 = 64.16$, $P < 0.0001$), and a group membership × self-investment interaction predicted prosocial choices ($b = -0.08$, $SE = 0.04$, Wald $\chi^2 = 4.58$, $P = 0.03$). In the analysis of mOFC responses, we did not observe a 3-way interaction between these predictors ($P = 0.35$). (We did observe an interaction of group and self-definition, such that those high on self-definition had stronger responses to efficiency for the out-group, $b = 0.15$, $SE = 0.06$, Wald $\chi^2 = 6.30$, $P = 0.01$. However, we refrain from interpreting this finding, as we did not observe similar effects of self-definition elsewhere.) Therefore, behavioral data suggest that group-level self-investment was associated with an overall tendency to give to the in-group or out-group, but not with an enhanced response to efficiency. This finding is consistent with the results in OFC, which tracked responses to efficiency, specifically.

Neural markers of vicarious reward

Finally, we examined putative neural markers of vicarious reward. A *strategic* model of intergroup relations would not predict differences in vicarious reward when a person merely witnesses in-group and out-group members win money. In contrast, an *intrinsic value* account would predict that a person's motivational investment in the in-group should shape relative responses to the gains of in-group and out-group others. To test these hypotheses, we created a region of interest for reward processing in ventral striatum using the meta-analytic website
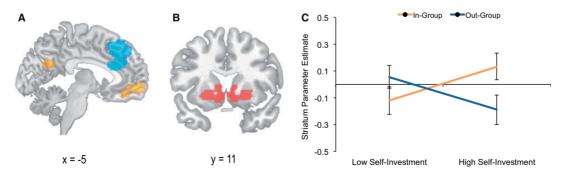

Fig. 3. Neural responses associated with prosocial behavior and witnessing in-group and out-group members win money. (**A**) During decisions to allocate money to the self or another person, medial orbitofrontal cortex (mOFC) showed increasing activation when participants allocated money to whomever stood to gain more (i.e. efficient choices), shown in yellow. Anterior cingulate cortex (ACC) showed increasing activation when participants allocated money inefficiently—for example, keeping money for the self when another person could gain more. (**B**) A portion of ventral striatum associated with the term 'reward' in an automated meta-analysis using Neurosynth. (**C**) Within this portion of ventral striatum, group-level self-investment predicted differential parametric responses to increasing amounts of money won by in-group members versus out-group members. The y-axis shows predicted parameter estimates (from a regression analysis) representing the strength of relationship between ventral striatum response and amount won by in-group or out-group members. Predictions for high and low self-investment are plotted one standard deviation above and below the mean.

Table 2. Brain regions that parametrically track choice efficiency

Contrast	Anatomical Region	Hemisphere	Volume (voxels)	MNI peak coordinates (mm) (x,y,z)	Maximum z score
Increasing efficiency	mOFC		182	9, 47, −8	4.05
	Supramarginal gyrus	R	195	60, −25, 34	4.18
	Precuneus		116	6, −55, 19	3.93
Decreasing efficiency	ACC/SMA		386	0, 29, 40	5.18
	Cerebellum	L	95	−45, −61, −32	3.68

Note: mOFC, medial orbitofrontal cortex; ACC, anterior cingulate cortex; SMA, supplementary motor area.

Table 3. Coefficients in regression of mOFC parameter estimates

Effect	b	SE	Wald χ^2	P
Intercept	0.20	0.04	22.25	<0.0001
Target Group	−0.06	0.07	0.68	0.41
SI	0.05	0.06	0.62	0.43
SD	0.05	0.06	0.73	0.39
Group × SI	−0.04	0.07	0.27	0.61
Group × SD	0.15	0.06	6.30	0.01

Note: SI, group-level self-investment; SD, group-level self-definition.

Table 5. Coefficients in regression of striatum parameter estimates

Effect	b	SE	Wald χ^2	P
Intercept	−0.03	0.04	0.62	0.43
Target Group	−0.04	0.06	0.43	0.51
SI	0.002	0.06	0.002	0.97
SD	−0.09	0.04	4.88	0.03
Group × SI	−0.17	0.08	4.29	0.04
Group × SD	0.15	0.13	1.43	0.23

Note: SI, group-level self-investment; SD, group-level self-definition.

Table 4. Coefficients in logistic regression of choice

Effect	b	SE	Wald χ^2	P
Intercept	−1.18	0.21	32.35	<0.0001
Ratio	1.89	0.24	64.16	<0.0001
Target Group	0.02	0.04	0.37	0.54
SI	0.47	0.34	1.97	0.16
SD	−0.001	0.35	<0.001	0.997
Ratio × Target	0.005	0.06	0.006	0.94
Ratio × SI	0.12	0.39	0.09	0.77
Ratio × SD	−0.29	0.34	0.71	0.40
Target × SI	−0.08	0.04	4.58	0.03
Target × SD	−0.02	0.03	0.30	0.58
Ratio × Target × SI	−0.08	0.09	0.88	0.35
Ratio × Target × SD	0.16	0.09	3.463	0.06

Note: SI, group-level self-investment; SD, group-level self-definition.

Neurosynth (Figure 3B). From this region, we extracted parameter estimates associated with the trial-by-trial amount won by in-group and out-group members. This analysis was restricted to trials on which participants merely witnessed outcomes without making any decisions, allowing us to examine responses when participants saw others receiving rewards.

As in our analysis of prosocial behavior, we found a target group × self-investment interaction, $b = −0.17$, SE = 0.08, Wald χ^2 4.29, P = 0.04 (Figure 3C; Table 5), suggesting that motivational components of identification were associated with differential responses in ventral striatum to in-group as opposed to out-group gains. For trials featuring in-group members, group-level self-investment had a marginally significant simple slope, indicating it was associated with more positive responses to in-group members, $b = 0.17$, SE = 0.10, Wald χ^2 = 2.99, P = 0.08. In contrast, self-investment did not significantly predict striatal responses when viewing out-group members, and the point estimate was negative, $b = −0.17$, SE = 0.10, Wald χ^2 = 2.75, P = 0.10. Among individuals with high self-investment, we observed a marginally significant simple effect of group, such that parameter estimates were stronger for the in-group than

the out-group, $b = −0.16$, SE = 0.09, Wald χ^2 = 3.59, P = 0.06. Among individuals with low self-investment, the coefficient for group membership ran in the opposite direction, although this difference was not statistically significant, $b = 0.09$, SE = 0.08, Wald χ^2 = 1.20, P = 0.27. These results suggest that participants' investment in the in-group predicted relative neural responses to the monetary gains of in-group and out-group members, even when they made no allocation decisions. (When we interrogated the small number of voxels identified in our reward localizer contrast, we observed the same interaction pattern in right caudate, $b = −0.36$, SE = 0.18, Wald χ^2=3.92, P = 0.05, but not left putamen, $b = −0.11$, SE = 0.28, Wald χ^2 = 0.31, P = 0.58.) Group-level self-definition was not associated with differential striatum responses to in-group versus out-group gains (P = 0.23). In sum, group-level self-investment (but not group-level self-definition) was associated with increasing responses in ventral striatum to witnessing increasing gains of an in-group as opposed to out-group member, linking a putative marker of vicarious reward to motivational components of social identification.

To ensure these effects were not due to objective differences in amounts won by each target despite the randomly generated monetary gains, we analyzed the amount won on forced win trials. We did not observe a clear effect of target group (P = 0.10), or a target group × self-investment interaction (P = 0.39). Therefore, neural responses were not due to objective differences in monetary gains, but rather to subjective responses to those gains.

Discussion

There is extensive evidence that social groups shape prosocial behavior (Brewer and Kramer, 1986; Tajfel *et al.*, 1971). However, group-based prosociality could stem from multiple psychological processes, including strategic concerns or the intrinsic value of helping motivationally important others. We tested whether people's motivational investment in a social group shapes the subjective value they place on social goods, beyond mere group categorization or perceived similarity to group members. By

studying a relatively large sample, we were able to parse individual differences in identification, which critically shape cognition and behavior in intergroup settings (Ashmore *et al.*, 2004). Behaviorally, we found that motivational components of group identification predicted differential giving towards in-group and out-group members. Using fMRI, we replicated prior findings demonstrating greater activation in medial orbitofrontal cortex when people make prosocial choices (Zaki and Mitchell, 2011).

Critically, fMRI further allowed us to examine responses when participants merely witnessed the gains of in-group and out-group members without making any allocation decisions. This approach helps preclude a role for strategic concerns. Motivational components of identification predicted differential ventral striatum responses when in-group and out-group members won increasing amounts of money. In past work, ventral striatum has been strongly associated with reward processing (Garrison *et al.*, 2013; Clithero and Rangel, 2014), including vicarious reward (Varnum *et al.*, 2014; Sul *et al.*, 2015). These findings suggest that social identification can shape the subjective value of another person's outcomes.

In theories of intergroup prosociality, some perspectives have viewed in-group favoritism as primarily driven by strategic concerns (Yamagishi and Kiyonari, 2000; Mifune *et al.*, 2010). However, we observed group-based differences in neural vicarious reward responses even when no choice was required. These effects cannot easily be explained by strategic concerns, and instead suggest that identity motivations can alter the intrinsic value placed on another person's outcomes. Of course, in daily life, multiple factors are likely to influence group-based prosociality, including strategic concerns. Future work could therefore compare conditions in which strategic motives are present versus absent—such as when others can or cannot reciprocate—in order to test their relative impact on behavior and vicarious reward. However, the present data suggest that these concerns cannot account for the full picture of intergroup social behavior.

More broadly, this finding advances understanding of both vicarious reward and intergroup behavior. Previous work has suggested that perceived similarity to another person moderates neural vicarious reward responses in ventral striatum (Mobbs *et al.*, 2009). In contrast, we found that the motivational significance of identification—but not perceived similarity to group members—was associated with striatal responses to in-group versus out-group gains. This finding introduces another variable—the motivational relevance of another person—that shapes vicarious reward. This work is consistent with a motivated account of empathic responding in intergroup relations (Cikara *et al.*, 2014; Zaki, 2014).

We did not observe a main effect of group membership on prosocial behavior; in fact, those low on group-level self-investment were more likely to give money to out-group members than in-group members. (The identification scale was anchored at 'strongly disagree', meaning that low ratings indicate disagreement with statements assessing identification with the group.) These findings are consistent with the idea that individuals who dislike a social group often react negatively when categorized as a group member (Branscombe *et al.*, 1999), and emphasize a motivational framework in which the value of identity-relevant outcomes depends on the significance of an identity to an individual (Hackel *et al.*, 2016). Although some contexts may evoke overall in-group favoritism, identity motivations may lead to more predictable patterns of group-based giving even in the absence of overall in-group bias.

In this vein, the pattern of responses in ventral striatum was partly driven by a *negative* response to in-group gains among those with *low* investment in the in-group. It is possible that a negative response indicates *displeasure* at seeing an in-group member win money, or at least displeasure that the in-group member won money rather than the self. Thus, social identity may not always lead to positive responses to in-group members, but instead may prompt responses consistent with one's own motivational orientation toward a group. For high-identifiers, it is possible that shared group membership mitigated the disappointment of not winning a reward oneself.

It is also not entirely clear whether this pattern was a result of low identification or dis-identification with the in-group. On one hand, previous work suggests that individuals low on group identification may resent being categorized as a group member (categorization threat), and may experience negative affect when group categorization is salient (Branscombe *et al.*, 1999), as was true in our experiment. On the other hand, even when individuals are unambiguously defined as group members, and do not experience threat from categorization, they may experience negative affect towards a group and a desire to leave the group—a phenomenon termed "dis-identification" (Becker and Tausch, 2014). Dis-identificaiton and categorization threat due to low identification have been considered distinct processes (Hamstra *et al.*, 2015). Therefore, future work should more directly test how each predicts prosocial behavior across groups. Altogether, however, the present findings are consistent with the hypothesis that identification can shape relative responses to in-group and out-group members.

Notably, people have multiple identities that can become salient in different contexts (Turner *et al.*, 1987; Van Bavel *et al.*, 2014), and identity motivations may most strongly impact valuation when the relevant identity is salient. Indeed, people show enhanced vicarious reward responses when they adopt an interdependent as opposed to independent self-representation (Varnum et al., 2014), indicating that salient self-representations can shape reward processing. Priming a group identity (as opposed to personal identity) may similarly enhance the influence of social identification on social valuation.

One limitation of the present data is that the contrast used to define a striatal region of interest did not survive correction for multiple comparisons, identifying only a small number of voxels. This may have been because we prioritized between-subjects power (i.e. many participants) over within-subjects power (i.e. many trials) in order to examine individual differences, leading to less robust contrasts. However, when interrogating a small number of voxels within right caudate that did respond to our localizer, we observed similar results as when interrogating an ROI generated from a Neurosynth map of the term 'reward'. Nonetheless, since these analyses departed from our initial analysis strategy, these results may be considered exploratory and would benefit from future replication with a more robust localizer.

Conclusion

The present research offers evidence that social identification can influence the subjective value of social goods. Using neuroimaging, the present study demonstrates that even neural responses measured in the absence of behavior—and thus in the absence of strategic or self-presentational motives—can reflect group identification, highlighting the value of neuroimaging for studying subjective value in social contexts. To the

extent that people feel invested in a social group, they may be particularly responsive to the outcomes of others in that group, and to the extent that people distance themselves from a group, they may devalue the in-group's well-being. More broadly, these findings support a framework in which motivationally salient identities can shape value-based decisions across multiple domains (Turner et al., 1987; Akerlof and Kranton, 2001).

References

Akerlof, G.A., Kranton, R.E. (2000). Economics and identity. *Quarterly Journal of Economics*, **115**(3), 715–53.

Allport, G.W. (1954). *The Nature of Prejudice*. Reading, MA: Addison-Wesley.

Ashmore, R.D., Deaux, K., McLaughlin-Volpe, T. (2004). An organizing framework for collective identity: articulation and significance of multidimensionality. *Psychological Bulletin*, **130**(1), 80.

Baumeister, R.F., Leary, M.R. (1995). The need to belong: desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, **117**(3), 497.

Becker, J.C., Tausch, N. (2014). When group memberships are negative: The concept, measurement, and behavioral implications of psychological disidentification. *Self and Identity*, **13**(3), 294–321.

Bernhard, H., Fischbacher, U., Fehr, E. (2006). Parochial altruism in humans. *Nature* **442**(7105),912–5.

Branscombe, N.R., Ellemers, N., Spears, R., Doosje, B. (1999). The context and content of social identity threat. In Ellemers N., Spears R., and Doosje B., editors. *Social Identity* (pp. 35–58). Oxford, England: Blackwell.

Braun, O.L., Wicklund, R.A. (1989). Psychological antecedents of conspicuous consumption. *Journal of Economic Psychology*, **10**(2), 161–87.

Brett, M., Anton, J.C., Valabregue, R., Poline, J.B. (2002) Region of interest analysis using an SPM toolbox. 8th International Conference on Functional Mapping of the Human Brain, June 2–6.

Brewer, M.B. (1988). A dual process model of impression formation. In Wyer R. S. and Srull, editors T. K. *Advances in Social Cognition* (pp. 1–36). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Brewer, M.B., Kramer, R.M. (1986). Choice behavior in social dilemmas: effects of social identity, group size, and decision framing. *Journal of Personality and Social Psychology*, **50**(3), 543.

Burkart, J.M., Hrdy, S.B., Van Schaik, C.P. (2009). Cooperative breeding and human cognitive evolution. *Evolutionary Anthropology: Issues, News, and Reviews*, **18**(5), 175–86.

Cameron, C.D., Payne, B.K. (2011). Escaping affect: how motivated emotion regulation creates insensitivity to mass suffering. *Journal of Personality and Social Psychology*, **100**(1), 1.

Cikara, M., Botvinick, M.M., Fiske, S.T. (2011). Us versus them: social identity shapes neural responses to intergroup competition and harm. *Psychological Science*, **22**, 306–13.

Cikara, M., Bruneau, E., Van Bavel, J.J., Saxe, R. (2014). Their pain gives us pleasure: how intergroup dynamics shape empathic failures and counter-empathic responses. *Journal of Experimental Social Psychology*, **55**, 110–25.

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

Clithero, J.A., Rangel, A. (2014). Informatic parcellation of the network involved in the computation of subjective value. *Social Cognitive and Affective Neuroscience*, **9**(9), 1289–302.

Correll, J., Park, B. (2005). A model of the ingroup as a social resource. *Personality and Social Psychology Review*, **9**(4), 341–59.

Council for Aid to Education. (2015). Voluntary support of education. Retrieved from http://cae.org/images/uploads/pdf/VSE_2014_Sample_Pages.pdf.

Davis, M.H., Mitchell, K.V., Hall, J.A., Lothert, J., Snapp, T., Meyer, M. (1999). Empathy, expectations, and situational preferences: personality influences on the decision to participate in volunteer helping behaviors. *Journal of Personality*, **67**, 469–503.

Daw, N.D., Gershman, S.J., Seymour, B., Dayan, P., Dolan, R.J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron*, **69**(6), 1204–15.

Doll, B.B., Duncan, K.D., Simon, D.A., Shohamy, D., Daw, N.D. (2015). Model-based choices involve prospective neural activity. *Nature Neuroscience*, **18**(5), 767–72.

Esterman, M., Tamber-Rosenau, B.J., Chiu, Y.C., Yantis, S. (2010). Avoiding non- independence in fMRI data analysis: leave one subject out. *Neuroimage*, **50**(2), 572–6.

Friston, K.J., Worsley, K.J., Frackowiak, R.S.J., Mazziotta, J.C., Evans, A.C. (1994). Assessing the significance of focal activations using their spatial extent. *Human Brain Mapping*, **1**(3), 210–20.

Garrison, J., Erdeniz, B., Done, J. (2013). Prediction error in reinforcement learning: a meta-analysis of neuroimaging studies. *Neuroscience & Biobehavioral Reviews*, **37**(7), 1297–310.

Gutsell, J.N., Inzlicht, M. (2012). Intergroup differences in the sharing of emotive states: neural evidence of an empathy gap. *Social Cognitive and Affective Neuroscience*, **7**(5), 596–603.

Hackel, L.M., Wohl, M.J.A., Coppin, G., Van Bavel, J.J. (2016). From groups to grits: Social identity shapes evaluations of food pleasantness.

Hamstra, M.R., Sassenberg, K., Van Yperen, N.W., Wisse, B., Rietzschel, E.F. (2015). Regulatory fit buffers against disidentification from groups. *Motivation Science*, **1**(3), 184.

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**(2), 583–90.

Hein, G., Silani, G., Preuschoff, K., Batson, C.D., Singer, T. (2010). Neural responses to ingroup and outgroup members' suffering predict individual differences in costly helping. *Neuron*, **68**(1), 149–60.

Kurzban, R., Neuberg, S. (2005). Managing Ingroup and Outgroup Relationships. In Buss D. M., editor, *The Handbook of Evolutionary Psychology* (pp. 653–675). Hoboken, NJ: John Wiley & Sons Inc.

Leach, C.W., van Zomeren, M., Zebel, S., *et al.* (2008). Group-level self-definition and self-investment: a hierarchical (multicomponent) model of in-group identification. *Journal of Personality and Social Psychology*, **95**(1), 144.

Levy, D.J., Glimcher, P.W. (2012). The root of all value: a neural common currency for choice. *Current Opinion in Neurobiology*, **22**(6), 1027–38.

Liang, K.Y., Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, **73**(1), 13–22.

Macey, P.M., Macey, K.E., Kumar, R., Harper, R.M. (2004). A method for removal of global effects from fMRI time series. *Neuroimage*, **22**(1), 360–6.

Mazaika, P., Whitfield-Gabrieli, S., Reiss, A., Glover, G. (2007). Artifact repair for fMRI data from high motion clinical subjects. Presentation at the meeting of the Organization of Human Brain Mapping, Chicago, IL.

Mifune, N., Hashimoto, H., Yamagishi, T. (2010). Altruism toward in-group members as a reputation mechanism. *Evolution and Human Behavior*, **31**(2), 109–17.

Mobbs, D., Yu, R., Meyer, M., et al. (2009). A key role for similarity in vicarious reward. *Science*, **324**(5929), 900.

Morelli, S.A., Sacchet, M.D., Zaki, J. (2015). Common and distinct neural correlates of personal and vicarious reward: a quantitative meta-analysis. *NeuroImage*, **112**, 244–53.

Nook, E.C., Zaki, J. (2015). Social norms shift behavioral and neural responses to foods. *Journal of Cognitive Neuroscience*, **27**(7), 1412–26.

Oyserman, D., Fryberg, S.A., Yoder, N. (2007). Identity-based motivation and health. *Journal of Personality and Social Psychology*, **93**(6), 1011.

Oyserman, D. (2009). Identity-based motivation: Implications for action-readiness, procedural-readiness, and consumer behavior. *Journal of Consumer Psychology*, **19**(3), 250–60.

Postuma, R.B., Dagher, A. (2006). Basal ganglia functional connectivity based on a meta-analysis of 126 positron emission tomography and functional magnetic resonance imaging publications. *Cerebral Cortex*, **16**(10), 1508–21.

Reed, A. (2004). Activating the self-importance of consumer selves: Exploring identity salience effects on judgments. *Journal of Consumer Research*, **31**(2), 286–95.

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**(2), 395–405.

Ruff, C.C., Fehr, E. (2014). The neurobiology of rewards and values in social decision making. *Nature Reviews Neuroscience*, **15**(8), 549–62.

Sul, S., Tobler, P.N., Hein, G., et al. (2015). Spatial gradient in value representation along the medial prefrontal cortex reflects individual differences in prosociality. *Proceedings of the National Academy of Sciences of the United States of America*, **112**(25), 7851–6.

Tajfel, H., Billig, M.G., Bundy, R.P., Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology*, **1**(2), 149–78.

Turner, J.C., Hogg, M.A., Oakes, P.J., Reicher, S.D., Wetherell, M.S. (1987). *Rediscovering the Social Group: A Self-Categorization Theory*. Cambridge, MA: Basil Blackwell.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*, **15**(1), 273–89.

Van Bavel, J.J., Xiao, Y.J., Hackel, L.M. (2014). Social identity shapes social perception and evaluation: Using neuroimaging to look inside the social brain. In Derks B., Scheepers D., Ellemers N., editors. *The Neuroscience of Prejudice*. New York: Psychology Press.

Van Bavel, J.J., Packer, D.J., Cunningham, W.A. (2008). The neural substrates of in-group bias: a functional magnetic resonance imaging investigation. *Psychological Science*, **19**(11), 1131–9.

Varnum, M.E., Shi, Z., Chen, A., Qiu, J., Han, S. (2014). When "Your" reward is the same as "My" reward: self-construal priming shifts neural responses to own vs. friends' rewards. *NeuroImage*, **87**, 164–9.

Wicklund, R.A., Gollwitzer, P.M. (1982). *Symbolic Self Completion*. Hillsdale, NJ: Routledge.

Yamagishi, T., Kiyonari, T. (2000). The group as the container of generalized reciprocity. *Social Psychology Quarterly*, 116–32.

Yarkoni, T., Poldrack, R.A., Nichols, T.E., Van Essen, D.C., Wager, T.D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods*, **8**(8), 665–70.

Zaki, J. (2014). Empathy: a motivated account. *Psychological Bulletin*, **140**(6), 1608–47.

Zaki, J., López, G., Mitchell, J.P. (2014). Activity in ventromedial prefrontal cortex co-varies with revealed social preferences: evidence for person-invariant value. *Social Cognitive and Affective Neuroscience*, **9**(4), 464.

Zaki, J., Mitchell, J.P. (2011). Equitable decision making is associated with neural markers of intrinsic value. *Proceedings of the National Academy of Sciences of the United States of America*, **108**(49), 19761–6.

Zaki, J., Schirmer, J., Mitchell, J.P. (2011). Social influence modulates the neural computation of value. *Psychological Science*, **22**(7), 894–900.