Socially connected brains:
Mechanisms that shape our social networks and positions within them

Noam Zerubavel

Submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2016

ABSTRACT

Socially connected brains: Mechanisms that shape our social networks and positions within them

Noam Zerubavel

The overarching goal of the present research is to gain a better understanding of mechanisms that shape our interpersonal ties and social networks by investigating their associated brain bases in naturally occurring groups. All three studies rely on a novel round-robin neuroimaging paradigm that incorporates group members as both participants in the fMRI scanner (perceivers) and stimuli (targets) presented during a naturalistic face-viewing task. Study 1 elucidates how group members' popularity is tracked by neural systems underlying valuation (i.e., processing reward value and evaluating others' motivational significance), which in turn engage social cognition systems that facilitate understanding others' mental states. Individual differences in the sensitivity of this neural mechanism are examined and found to correlate with perceivers' own popularity. Studies 2 and 3 extend the paradigm developed in Study 1 to incorporate social network data collected in a longitudinal context, and further test whether neural measures collected during the initial stages of group formation can prospectively predict group members' future liking ties (Study 2) and social network centrality (Study 3). In Study 2, neural activity in the aforementioned valuation systems predicts newly acquainted group members' future—but not current—idiosyncratic liking of one another. Further analyses suggest this effect reflects only one facet of a far more nuanced interpersonal phenomenon implicated in the eventual emergence of dyadic liking reciprocity: individuals' initial liking preferences are not *personally* tracked by their own brains' idiosyncratic valuation responses to particular group members, but rather *interpersonally* tracked by the neural valuation responses they uniquely

evoke in those particular group members; moreover, each dyad member's idiosyncratic valuation activity influences both their own and each other's future liking. Having established in Studies 1 and 2 a paradigm for measuring how social network members implicitly evaluate one another, Study 3 extends it to include oneself (i.e., the perceiver) as an evaluate target of social perception. Revisiting the Study 1 social network members' data, enhanced valuation activity in response to oneself (relative to others) correlates positively with questionnaire measures of dispositional narcissism (but not self-esteem) and negatively with sociometric popularity. Using the data from Study 2, the trait narcissism effect is replicated and extended to a context in which the "others" are newly acquainted group members. This longitudinal data also reveals that the neural measure of narcissistic self-valuation prospectively predicts future (un)popularity, even controlling for initial levels of popularity. Considered together, this research aims to integrate conceptual and methodological frameworks across social psychology (e.g., round-robin experimental designs), cognitive neuroscience (e.g., fMRI), and sociology (e.g., social network analysis).

Table of Contents

List of Figures and Tables…...ii

Introduction …....1

Study 1: Neural Mechanisms Tracking Popularity in Real-World Social Networks …..7

Introduction ….8

Methods ….12

Results….21

Discussion ….37

Study 2: The Neural Bases of Future Liking and Reciprocity…...42

Introduction …..43

Methods ….45

Results …..49

Discussion ….60

Study 3: Neural Measure of Enhanced Self-Valuation Tracks Narcissism and Predicts Unpopularity …..63

Introduction …...64

Methods ….66

Results …..67

Discussion ….74

References …...77

List of Figures and Tables

Figure 1. Social network structure of study participants..11

Figure 2. Popularity of targets predicted activity in valuation and social cognition regions of interest..22

Figure 3. Parametric whole-brain regression analysis isolating brain regions tracking target popularity during the face-viewing task..27

Figure 4. Primarily distinct patterns of neural activity associated with tracking popularity of targets and each perceiver's own liking of targets during the face-viewing task...........29

Figure 5. Activity in the valuation system mediated the observed relationship between target popularity and social cognition system activity..31

Figure 6. Interaction plot depicting popular perceivers' enhanced attunement to group members' status differences...34

Figure 7. Popular perceivers were more accurate in their predictions of how individual members of their organization assessed their personality across all trait items............................36

Figure 8. Valuation system regions of interest and social network structures of idiosyncratic valuation and liking relations..50

Figure 9. Valuation activity elicited by alters tracks their concurrent (T1) sociometric popularity, but does not predict their future (T2) sociometric popularity..54

Figure 10. Mutual reciprocation of idiosyncratic liking increased dramatically over the course of the summer program...57

Figure 11. Analyzing ego-to-alter idiosyncratic liking at T1 and T2 against ego-to-alter and alter-to-ego idiosyncratic neural valuation responses..59

Figure 12. Self-enhancement bias expressed in valuation system activity, a main effect of self-face which was amplified by individual differences in dispositional narcissism...........69

Figure 13. Self-enhancement bias expressed in valuation system activity, a main effect of self-face which was mitigated by sociometric popularity...69

Figure 14. Self-enhancement bias expressed in valuation system activity, a main effect of self-face which was not moderated by individual differences in self-esteem.......................70

Figure 15. Replicating self-enhancement bias expressed in valuation system activity, a main effect of self-face amplified by individual differences in dispositional narcissism.…...71

Figure 16. Self-enhancement bias in valuation activity predicts future (but not current) sociometric unpopularity...…73

Figure 17. Illustration of the negative relationship between the neural index of narcissistic self-valuation and sociometric popularity social networks separately together....................75

Table 1. Tabulation of group members included in each study phase……...................................14

Table 2. Results of linear mixed-effects models regressing activation parameter estimates against target popularity and covariates...25

Table 3. Regions parametrically tracking target popularity as identified whole-brain analysis...28

ACKNOWLEDGMENTS

This dissertation represents the collaborative effort of an incredible social network. First and foremost, I would like to thank my advisor, Kevin Ochsner, for making this dissertation research possible. I am lucky to have an advisor that provides his students so much intellectual freedom to pursue wacky ideas like this one. This research would have also been impossible without Peter Bearman, who has been its champion from the outset. Peter's contributions to this work are wide-ranging, but perhaps most invaluable to me has been his disregard for traditional disciplinary boundaries. I would also like to thank Niall Bolger for his immense statistical guidance and teaching me how to embrace the inherent interdependence of relational data. In additional, thank you to Adam Galinksy for showing me that social status is a fascinating and important topic for psychological research. Finally, Tory Higgins was not only the chair of my committee but also the reason I am—and will forever be—a social psychologist at heart.

I would also like to thank several other collaborators on this dissertation research who were not on my dissertation committee: Jochen Weber, Adam Reich, and Mark Hoffman. Jochen truly is the Neuroelf (also the name of the neuroimaging software he developed), and data analysis for this dissertation research would have been impossible without his programming magic. I am hugely grateful to Adam and Mark for all of the time and effort we have spent working together trying to make sense of Study 2.

I would also like to thank everyone in the SCAN Lab who I have come to consider my family over these years. In particular, my officemates for the majority of these six years—Bruce Dore, Seth Kallman, Rebecca Martin, Laura Martin Braunstein, and Jen Silvers—were consistently the best part of my graduate school experience.

There are so many other folks in the Psychology department to whom I am grateful for many enlightening conversations over the years, including James Curley, Dean Mobbs, Valerie Purdie-Vaughns, Daphna Shohamy, Janet Metcalfe, and Walter Mischel.

My family is small but wonderful. To Ima, Aba, Nogi, Dave, and my closest friends who might as well be family: I would not be me were it not for you, and hopefully I can do a better job of thanking you for everything in person rather than here in writing. Finally, thank you to the truly singular Kristin Nothwehr. You are the love of my life, and I am forever grateful to you and the internet and all the mysterious forces in the world that brought us together.

Thank you again to everyone listed above and to all those essential nodes in my network that I could not fit into these few paragraphs. We did this thing together.

Introduction

Humans are a fundamentally social species: we are embedded in uniquely complex social networks and possess unique social-cognitive capacities for navigating them. Our brains are thought to have evolved under selective pressures to automatically process and efficiently respond to relevant social information. Recent advances in neuroimaging technology have made it possible for cognitive neuroscientists as well as social psychologists to examine these neural processes at work. Typically, this research utilizes functional magnetic resonance imaging (fMRI) to investigate patterns of neural activation while viewing social stimuli that vary on the basis of experimentally manipulated dimensions like race, gender, or normative attractiveness – variables that can be engineered using computer-generated models or morphed faces.

However, many other social constructs of import to psychology and related fields—such as interpersonal affiliation, dyadic reciprocity, or social status—are not practically and/or ethically amenable to experimental manipulation. The irony is that while these interpersonal phenomena emerge organically and pervasively in real-world social contexts, they are difficult to convincingly translate into controlled lab settings. Given this methodological tension between ecological validity and experimental control, the majority of neuroimaging researchers have opted to utilize less naturalistic stimuli that they could artificially manipulate. For instance, fMRI studies of social status perception have operationally defined social status using performance on a dot array discrimination task (Zink et al., 2008), dominant versus submissive body posture (Freeman, Rule, Adams Jr, & Ambady, 2009; Marsh, Blair, Jones, Soliman, & Blair, 2009), and celebrities with known status relations (e.g., Prince Harry and Queen Elizabeth II) (Farrow et al., 2011). While neuroimaging studies benefit from the experimental control afforded by such stimuli, they leave open questions about the mental processes underlying status perception—and other interpersonal phenomena more generally—in real-world social contexts (Neisser, 1976;

Zaki & Ochsner, 2009; Zaki & Ochsner, 2012).

Many of these same social constructs have also been of longstanding interest to sociologists, even if not the associated psychological processes or neural mechanisms. The sociological approach—to researching these interpersonal phenomena and to social science theory more generally—emphasizes the analysis of social structure, that is, the patterned social arrangements that organize individuals and their interpersonal ties. Over the last century, sociologists have developed and enhanced the techniques of social network analysis (SNA) in order to precisely measure and quantify social-structural characteristics of individuals, dyads, and cliques embedded within social networks. Because sociologists are primarily interested in distilling the structural principles underlying the organization of interpersonal ties, these analytic techniques typically control for individual-level characteristics of interest to psychologists. In other words, the sociological perspective regards many psychological variables as idiosyncratic quirks—"noise" in the system that interferes with the social-structural "signal" of interest.

By contrast, social psychologists and neuroscientists are primarily interested in explaining phenomena in terms of intra-individual psychological processes and neural mechanisms, respectively. As such, they conventionally regard the social structures in which individual participants are embedded as background "noise" interfering with the psychological or neural "signal" they seek to capture within each individual. From the perspective of an experimental researcher, it should not be surprising that pre-existing interpersonal relationships between individuals participating in a study (or between a participant and another person incorporated as a stimulus in the study) constitute thorny methodological obstacles. Even social psychologists who are interested in interpersonal phenomena and study group processes often rely on minimal- or zero-acquaintance paradigms, in which the study's participant groups consist of individuals with no prior relationships. This experimental protocol provides a social-structural "clean slate" for psychological research by intentionally eliminating the very network structures studied in sociological research.

The dissertation research presented here attempts to integrate the conceptual and methodological frameworks of cognitive neuroscience, social psychology, and sociology. This interdisciplinary approach conceptualizes all three disciplines' levels of analysis—neural mechanism, psychological process, and social structure—as important "signals" of scientific research. Rather than ignoring any of these levels, we obtain measurements across all three; likewise, rather than statistically controlling for what might otherwise be considered "nuisance" variables by a particular research tradition, we analyze how interrelated variables mutually influence one other across different levels of analysis. How else can we hope to understand social phenomena that shape and are shaped by brains of individuals who interact in relationships that are embedded in social networks?

This work advances a radically social and naturalistic approach to studying the brain bases of social perception by emphasizing its fundamentally interpersonal and interdependent nature in the following ways: (1) recruiting real-world groups as study participants, (2) incorporating each group member as both a study participant and experimental stimulus, (3) integrating social network analysis to measure social-structural characteristics of individuals and relationships within these groups, and (4) directly challenging the assumption of independence that underlies empirical research in neuroscience and psychology.

To my knowledge, Study 1 represents the first fMRI study to recruit members of pre-existing, real-world groups (Zerubavel, Bearman, Weber, & Ochsner, 2015), an approach pursued throughout the entirety of the dissertation research. Studies 1-3 also rely on a novel round-robin neuroimaging paradigm that incorporates group members as both participants in the fMRI scanner (perceivers) and stimuli (targets of social perception) presented during a naturalistic face-viewing task. This round-robin design enables analyses at the level of the perceiver (Study 1), the target (Study 1 and Study 2), and perceiver X target interactions unique to each interpersonal relationship (Study 2 and—by conceptualizing self-perception as an interpersonal process in which perceivers are also the targets of their own social perception—Study 3; see Kwan, John, Kenny, Bond, & Robins, 2004). In addition, all three studies utilize sociometric instruments and social network analysis to precisely measure group members' relative positions and interpersonal ties within the network structure; then, participants' fMRI data can be analyzed as a function of these individual (i.e., perceiver-level or target-level) and relational social network variables. In Studies 2 and 3, the analytic strategy is flipped such that neural measures are leveraged to prospectively predict future liking (Study 2) and sociometric popularity (Study 3).

By recruiting group members with real-world relationships, incorporating them as both participants and stimuli, and utilizing social network analysis, the approach advanced by this dissertation fundamentally opposes a central tenant of psychology and neuroscience—the assumption of independence—which maintains that data from each individual unit (i.e., person) is unrelated to data from other units. The assumption of independence serves as the foundation for both fields' research paradigms (including participant recruitment and preparation of experimental stimuli), the statistical analysis of data, and—perhaps most critically—the conceptualization of persons as essentially independent entities. The present studies diametrically oppose this assumption of independence, instead capitalizing on the interdependence of group members methodologically (with the round-robin design), analytically (with social network analysis), and theoretically (by conceptualizing individuals as inextricably embedded within interpersonal contexts). Traditionalists in the fields of psychology and neuroscience would be right to criticize this approach as iconoclastic. So consider yourselves forewarned: this dissertation research on interpersonal phenomena and their associated intrapersonal processes violates the assumption that people are fundamentally independent.

Study 1:

Neural Mechanisms Tracking Popularity in Real-World Social Networks

Introduction

Humans are a fundamentally social species, and the social networks in which we are embedded significantly determine our physical and psychological well-being (Smith & Christakis, 2008). Effectively navigating interactions within these networks requires efficient mechanisms for processing social information about network members. This ability is so important that it may be among the foremost computational challenges that influenced primate evolution, particularly the dramatic development of our 'social brains' (Dunbar, 2012; Silk, 2007).

Differences in popularity reflect status inequalities that shape social interaction within virtually all human groups across an enormous array of contexts, from classrooms to military barracks to voluntary associations and beyond (Davis, 1970; Krantz & Burton, 1986; Lansu, Cillessen, & Karremans, 2013; Moreno, 1934; Vaughn & Waters, 1981). For decades, social scientists have used sociometric assessment and social network analysis (SNA) to measure the organization of groups and individuals' positions within them. Using these techniques, the extent to which each group member is collectively liked by group members – termed sociometric popularity – can be quantified (Moreno, 1934; Newcomb, 1963; Wasserman & Faust, 1994). Highly likeable, individuals attract group members and elicit their affiliation with warmth, altruism, and related traits like agreeableness (Henrich & Gil-White, 2001; Moreno, 1934; Newcomb, 1963; Wiggins & Trapnell, 1996). Sociometric popularity disparities arising from asymmetries in group members' liking ties are present in virtually all human groups and constitute a fundamental basis for status differentiation (Davis, 1970; Moreno, 1934).

The fact that differences in popularity have important behavioral consequences raises the question of how we recognize these differences in the first place. Consider, for example, that in our everyday social networks, we recognize that certain group members are collectively liked more than others, even when this consensus preference differs from our own. Adults and even children can perceive other group members' asymmetric liking ties, detect differences in their relative popularity, and accordingly orient attention and affiliative behavior toward popular individuals (Krantz & Burton, 1986; Lansu et al., 2013; Moreno, 1934; Vaughn & Waters, 1981). Achieving such acute sociometric awareness and attunement to popular group members might feel like second nature to us, yet little is known about the underlying neural mechanisms. Here, we combined functional magnetic resonance imaging (fMRI) and SNA to investigate how the human brain tracks the popularity of members of real-world social networks.

To provide new insights into the neural mechanisms that undergird navigation of our complex social worlds we addressed three inter-related questions: First, which brain systems track real-world popularity? Second, what is the functional organization of those systems? And third, does one's own status predict more or less neural attunement to others' status? Although no prior human research has investigated these questions, the extant literature suggests that two distinct types of brain systems may be involved in tracking popularity.

The first is comprised of the ventromedial prefrontal cortex (vmPFC), ventral striatum (VS), and amygdala. These densely interconnected regions (Haber & Knutson, 2009), henceforth referred to collectively as the *valuation system*, are consistently implicated in processing the affective value and motivational significance of various stimuli, including other people (Adolphs, 2003; Doré, Zerubavel, & Ochsner, 2014; Güroğlu et al., 2008; Haber & Knutson, 2009; Krienen, Tu, & Buckner, 2010; Zink et al., 2008). Although human neuroscience research has yet to investigate sociometric popularity, nonhuman primate researchers have found that neurons in these regions signal group members' dominance rank (Azzi, Sirigu, & Duhamel, 2012; Klein & Platt, 2013; Watson & Platt, 2012) and proposed that vmPFC, VS, and amygdala interact to encode, monitor, and signal other individuals' social value (Klein, Shepherd, & Platt, 2009). If tracking group members' popularity depends on the motivational significance and social value attributed to them, then valuation system activity should track targets' sociometric popularity.

The second network is comprised of the dorsomedial prefrontal cortex (dmPFC), temporoparietal junction (TPJ), and precuneus. These interconnected regions, henceforth referred to collectively as the *social cognition system*, are consistently activated in neuroimaging studies involving judgments about others' psychological characteristics, mental states, and intentions (Adolphs, 2003; Denny, Kober, Wager, & Ochsner, 2012; Doré et al., 2014) or the passive viewing of social stimuli—such as familiar faces—for which we might spontaneously make such attributions (Gobbini & Haxby, 2007). Although no neuroscience work has asked how these systems might track sociometric popularity, behavioral research shows that people are particularly concerned with understanding high-status individuals' mental states (especially how they are viewed by them) and predicting their intentions (Dépret & Fiske, 1993; Fiske, 1993; Snodgrass, 1985, 1992). If perceivers are preferentially motivated to understand popular (relative to unpopular) group members' mental states, then social cognition system activity should scale with targets' popularity.

Based on these findings both the valuation and social cognition systems are candidate neural networks for tracking group members' popularity. Our primary objective was to test these possibilities, recognizing that they are not mutually exclusive. Indeed, the two systems are functionally distinct but their interactions are often critical for diverse social behaviors (Doré et al., 2014).

To address these questions, two different groups of well-acquainted participants were recruited from two voluntary student organizations with equivalent size and affiliation network structures (see Fig. 1, Table 1, and Methods). Specifically, sociometric popularity was indexed by individuals' *degree prestige* within the directed liking network, standardized by group. This measure of popularity aggregates liking ratings received by each group member and thus intuitively reflects how much individuals are *collectively liked* by their fellow group members (Wasserman & Faust, 1994).

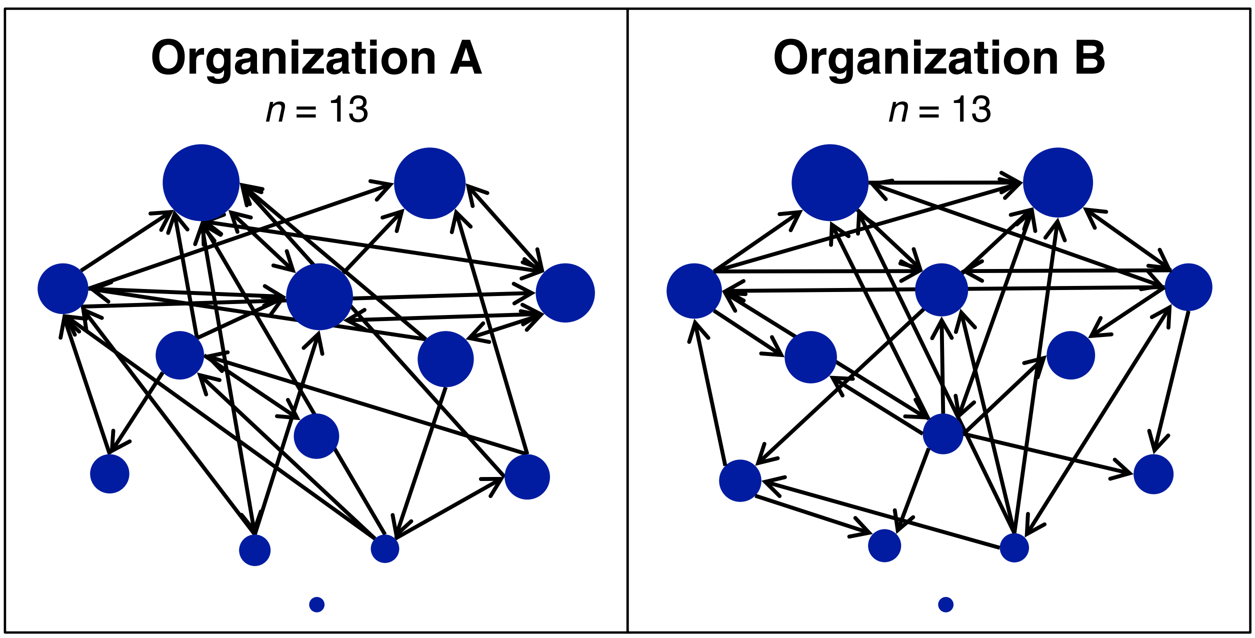

Fig. 1. Social network structure of study participants ($N = 26$) in two voluntary student organizations (clubs; participant information detailed in Methods and Table 1). Each network was comprised of 13 well-acquainted members. Each node represents one person. Directional arrows represent group members' directed liking relations (for visual clarity, only ties in the upper quartile are displayed). Node size reflects sociometric popularity – the extent to which the group collectively likes that person. Sociometric popularity was indexed by *degree prestige*, which we then standardized by group (Methods). Calculated by simply summing the weights of all liking ties received by an individual, this SNA metric represents an intuitive and straightforward index of popularity (Wasserman & Faust, 1994).

To model everyday social encounters within face-to-face social networks, we developed a round-robin neuroimaging paradigm in which group members were both the target stimuli presented during the scan and the perceivers that viewed them. A cover task guided perceivers to make simple judgments about briefly presented photographs of target faces. To provide a strong test of our hypotheses about the neural systems tracking targets' sociometric popularity, our primary analyses were based on independently identified valuation and social cognition networks that were localized using two additional tasks that were completed in the same scanning session (see Methods). We then used combinations of multi-level regression and mediation analyses to ask how activity within each network tracked targets' sociometric popularity during this face-viewing task, how activity in these systems interacted, and how a perceiver's own popularity impacted their sensitivity to differences in target popularity.

Methods

Participants. Participants were 26 healthy young adults (12m, 14f; mean age = 28.7, SD = 2.3) recruited from two different voluntary student club organizations with equivalent size and affiliation network structures (13 members from each; see Fig. 1 and Table 1) at a large university in the United States. Critically, this recruitment achieved several objectives. First, face-to-face organizations of this size were necessary to ensure that group members were sufficiently well-acquainted with one another (mean duration of relationship = 8.5 months, SD = 5.0). Second, for groups of this size if was feasible for each participant's face to be presented as a stimulus (target) with ten repetitions (multiple repetitions being necessary to estimate blood flow responses to each face) during the face-viewing task in the scan session. Third, selecting two such groups with equivalent size and affiliation network structures enabled us to aggregate their data and analyze between-subject (perceiver) popularity effects.

Initial recruitment yielded 100% member response rate in both organizations, however not all met the inclusion criteria to participate in each of the study phases. Out of 28 total individuals comprising both groups, 26 (93%) were eligible, willing, and able to participate in the study; among the 26 participants, all 26 (100%) completed the initial session in which the social network instruments were administered, 25 (96%) were photographed and incorporated as targets (face stimuli) in the subsequent fMRI face-viewing task, while 21 (81%) constituted perceivers who completed the fMRI scanning session (Table 1). Participants were English-speaking and had normal or corrected-to-normal vision. They were screened for a history of serious neuropsychiatric disorders, head injury, and other conditions that prevented scanning (e.g., a pacemaker, claustrophobia) prior to taking part in the fMRI scanning session.

Beyond these core participants, 40 additional participants were recruited via Mechanical Turk to provide normative ratings of stimuli used in the fMRI face-viewing task. 20 of these participants (7m, 13f; mean age = 35.9, SD = 14.6) rated faces based on attractiveness, and another 20 (10m, 10f; mean age = 34.8, SD = 12.9) rated faces based on trustworthiness. These normative ratings of stimulus faces could then be used as covariates (to rule out potential confounds associated with target facial attributes) in subsequent analyses.

All participants received monetary compensation and provided informed consent following the standards of the Columbia University Institutional Review Board.

Table 1. Tabulation of group members included in each study phase.

	Total	Organization A	Organization B
Non-Participants	*2 (2f)*	*2 (2f)*	--
Participants	**26 (12m, 14f)**	13 (5m, 8f)	13 (7m, 6f)
Targets: Incorporated in round-robin fMRI face-viewing task as stimuli	**25 (12m, 13f)**	13 (5m, 8f)	12 (7m, 5f)
Perceivers: Incorporated in round-robin fMRI face-viewing task as subjects	**21 (10m, 11f)**	9 (3m, 6f)	12 (7m, 5f)

Procedure and design. The study was comprised of two sessions. In a preliminary session, sociometric instruments and self-report questionnaires were administered, and photographs were taken of participants' faces (to be used subsequently in the fMRI face-viewing task). In a second session, participants underwent fMRI scanning while completing several tasks described below. For all computerized tasks in both sessions, stimulus presentation and behavioral data acquisition were controlled using E-Prime 2.0 (Psychology Software Tools, Inc.). For tasks completed in the fMRI scanning session, visual stimuli were displayed on a projection screen using a LCD projector and viewed via a rear-projecting mirror.

Sociometric assessment and social network analysis (SNA). Sociometric assessments of group members' affiliative relations and resulting network structure were collected from participants during the first session. These assessments were conducted via a computerized peer-rating paradigm in which participants rated how much they liked each group member (presented in randomized order) on a sliding visual analog scale anchored by the labels "not very" and "very" on opposite ends. This sociometric instrument provided a continuous measure of personal liking (i.e., affiliation tie strength) between group members that was used as a covariate in analyses and also to compute each group member's popularity. Specifically, sociometric popularity was indexed by individuals' degree prestige (alternatively referred to as indegree centrality) within the directed liking network (Wasserman & Faust, 1994), which we then standardized by group. In other words, liking ratings received by each group member were summed for that individual and then standardized to *z* scores within group. Using these sociometric assessments and network analyses thus generated a popularity index that reflects how much individuals are *collectively liked* by their fellow group members.

Round-robin fMRI face-viewing task. Stimuli for the fMRI face-viewing task were prepared from photographs of participants. During the preliminary session, participants' faces were photographed with affectively neutral facial expression and gaze directed straight at the camera. These photographs were cropped and converted to grayscale images with equal luminance. In addition, a "ghost face" stimulus image representing the superimposition of all group members' faces was prepared for each group following methods used in prior face perception research (Taylor et al., 2009). The face-viewing task implemented a rapid event-related design that included 10 repetitions of each stimulus face presented in pseudorandomized order. Faces were presented for 1000ms and interstimulus intervals (ISIs) consisting of white fixation cross on black background were jittered between 1500 ms and 11500 ms (mean duration of ISI=3500 ms). Perceivers viewed faces of targets while performing a simple cover task in order to maintain their alertness throughout. Specifically, participants were instructed to press a button with their pointer (second) finger each time a group member's face was presented and a different button with their ring (fourth) finger each time a "ghost face" was presented (~9% of total presentations).

Independent functional localizer task: Valuation system. Two functional localizer tasks were completed at the end of the scanning session. Participants completed the monetary incentive delay (MID) task (Knutson, Westdorp, Kaiser, & Hommer, 2000) to independently identify valuation regions active during the anticipation and receipt of monetary rewards (Tamir & Mitchell, 2012; Zaki, Schirmer, & Mitchell, 2011). The MID task included 30 trials in which it was possible to win a reward (reward-possible trials) intermixed with 15 trials in which winning a reward was not possible (neutral trials). Each trial of the MID task began with a 500 ms presentation of one of two cue symbols: a green square indicated that the current trial offered an opportunity to win $1 (reward-possible trial); a red square indicated that the current trial did not offer an opportunity to win money (neutral trial). Following the cue symbol, there was a delay interval (with randomly determined duration between 2000 and 2500 ms) and then a target stimulus (yellow star) was briefly presented. On reward-possible trials, participants would win $1 if they made a button press while the target stimulus was displayed and $0 if the button press was made prior to the target onset or after the target offset. On neutral trials, although no money could be earned, participants were instructed to still make a button press while the target stimulus was displayed. Following the target stimulus offset, feedback (i.e., the amount earned on that trial—either $0 or $1—and the total cumulative earnings) was presented for 500 ms. The duration of the target stimulus presentation was adjusted algorithmically (within the range of 150 to 550 ms) based on task performance up until that point; specifically, the algorithm was intended to generate a two-thirds success rate on reward-possible trials. The algorithm succeeded in adjusting task difficulty such that participants earned money on ~20 out of the 30 reward-possible trials (mean wins = 19.8). Five participants were unable to complete the MID task due to technical issues.

Trials in which participants won monetary rewards were contrasted to those in which they could not (win trials > neutral trials), encompassing both the anticipation and feedback phases of each trial. This analysis (thresholded at $P < 0.05$ corrected) revealed activation peaks consistent with previous studies using the MID task (Hommer et al., 2003; Knutson et al., 2000; Tamir & Mitchell, 2012; Zaki et al., 2011) in regions of *a priori* interest: vmPFC (-3, 48, -6), VS (0, 9, -3), and amygdala (-21, -6, -12 and 18, -3, -12). We then defined spherical ROIs with a radius of 8mm around these peaks (Tamir & Mitchell, 2012; Zaki et al., 2011) (Fig. 2A). The resulting spherical ROIs were 2109 voxels. For the subcortical structures (VS and amygdala), these spheres were then anatomically constrained using structural masks obtained from FSL: for VS, the Oxford-GSK-Imanova Structural-Anatomical Striatal Atlas constrained the ROI to 39 voxels; for amygdala, the Harvard-Oxford Atlas constrained the ROI to 1592 voxels).

Independent functional localizer task: Social cognition system. We adapted a well-validated person judgment task (Ochsner et al., 2005) as an independent functional localizer to identify social cognition regions supporting two kinds of judgments relevant in interactions with group members: evaluating target group members' mental states and traits (e.g., 'to what extent is [target] helpful?') and predicting how targets perceive them (e.g., 'to what extent does [target] see me as lonely?'). The task implemented a rapid event-related design comprising 240 judgment trials lasting 3500 ms each and inter-trial intervals (ITIs) consisting of a white fixation cross on black background were jittered between 1500 ms and 11000 ms (mean duration of ITI = 4000 ms). For each of 40 trait adjectives (20 positive, 20 negative), participants made six kinds of judgments using a five-point scale: you-about-you, you-about-other1, you-about-other2, other1-about-you, other2-about-you, and active baseline curved line judgments. On you-about-you trials, participants judged the extent to which the trait adjective described them (1 = not very; 5 =

very). On you-about-other1 and you-about-other2 trials, participants judged the extent to which the adjectives described one of two group members. On other1-about-you and other2-about-you trials, participants predicted the extent to which one of two group members would judge the adjective as describing them (i.e., the participant). On curved line trials, as an active baseline task that matched non-social aspects of the other judgment trial types, participants judged the extent to which the trait word contained curved lines as opposed to straight lines (1 = very few, <20% curved lines; 5 = very many, >80% curved lines). Trial types were presented in a pseudorandom counterbalanced order and distributed across the task's 4 runs such that each run included 10 trials (5 with positive traits, 5 with negative traits) of each judgment type. As detailed below, of interest for this study were activations common to both you-about-other and other-about-you judgment trials relative to active baseline curved line trials. Data relating to you-about-you trials were collected for other studies to be analyzed and reported separately.

We conducted a whole-brain conjunction analysis (thresholded at $P < 0.05$ corrected) to localize activation present in both you-about-other and other-about-you trials relative to active baseline curved line trials. This analysis revealed clusters with activation peaks in regions of *a priori* interest that were consistent with previous neuroimaging studies using similar social-cognitive tasks (Denny et al., 2012): dmPFC (0, 60, 21), precuneus (-3, -57, 21), and left (-60, -60, 24) and right TPJ (54, -60, 21). As with the valuation localizer, we defined spherical ROIs with a radius of 8mm around the observed activation peaks (Fig. 2B). For dmPFC and left TPJ, the 8mm-radius spheres were anatomically constrained so as not to extend beyond the boundaries of the brain (resulting in ROIs comprised of 2066 and 1701 voxels, respectively).

Imaging acquisition and analysis. Whole-brain fMRI data were acquired on a 1.5 Tesla GE system. Functional images were acquired with a T2*-sensitive EPI blood oxygenation level dependent (BOLD) sequence using the following parameters: TR = 2000 ms; TE = 34 ms; flip angle = 90°; field of view = 22.4cm × 22.4cm; matrix array size = 64 × 64; each volume consisted of 28 slices with slice thickness = 4mm and no inter-slice gap. High-resolution anatomical images with 1mm × 1mm × 1mm resolution were acquired with a T1-sensitive SPGR sequence at the end of the scan session.

In each of the two runs comprising the face-viewing task, 167 volumes (for participants in Organization A) and 157 volumes (for participants in Organization B) were acquired. (The difference in volumes acquired was due to the fact that the face task for Organization A included one more target than it did for Organization B.) For both groups, the person judgment task consisted of four runs of 230 volumes each, while the MID task consisted of one run of 115 volumes. The initial four "dummy" volumes of each run were discarded prior to analysis.

Functional images were preprocessed using SPM8 software (Wellcome Department of Cognitive Neurology, UCL), including slice-timing correction, motion-correction, realignment, coregistration between each participant's functional and anatomical data, normalization to a standard template (Montreal Neurological Institute; MNI) using segmentation parameters, 3mm isometric voxels, and spatial smoothing using a Gaussian kernel (full-width at half-maximum = 6mm).

Data for each of the tasks were subjected to a first level of regression, separately for each subject, using an ordinary-least-squares general linear model (GLM) implemented with Neuroelf v0.9c software (neuroelf.net). Task-based regressors are described below. Each of the GLMs included, in addition to the task-related regressors, the 6 motion parameters as estimated during realignment as well as a DCT-based basis set covering low-frequency up to 1/80Hz to account for signal variability introduced by head motion and temporal drifts. The GLM for the face-viewing task included one regressor for each target face (including the ghost face stimulus), representing the 10 repetitions of each respective face (12 repetitions for the ghost face). Each of these regressors was created by convolving the canonical hemodynamic response function (HRF) with a series of boxcars representing the 1000 ms intervals during which a particular face was presented. The GLM for the MID task included three task regressors corresponding to three trial types: wins, misses, and neutral. Each of these regressors was created by convolving the canonical HRF with a series of boxcars representing the 3000 ms interval encompassing anticipation (delay) and feedback phases of each trial. The GLM for the person judgment task included task regressors for each the 40 traits and 6 judgment types (i.e., you-about-you, you-about-other1, you-about-other2, other1-about-you, other2-about-you, and curved lines). Each of these task regressors was created by convolving the canonical HRF with a series of boxcars representing the 3500 ms duration of judgment trials.

The output of these first-level regressions was a series of parameter estimate (beta) maps used in the next group level of analyses. For the person judgment task, an additional intermediate step averaged a subset of these beta maps to obtain unbiased estimates of the BOLD response to a set of judgment trial types: you-about-other1 and you-about-other2 beta maps were combined into you-about-other beta maps; likewise, other1-about-you and other2-about-you were combined into other-about-you beta maps; additionally, the 40 individual trait beta maps were combined for each judgment type. For the face-viewing task, beta maps corresponding to trials on which participants viewed either the ghost face or themselves were discarded prior to the next level of analyses.

Results

Target popularity analyses: ROI approach. For our primary analysis, we first needed to independently localize regions of interest (ROIs) related to affective valuation and social cognition. Following the established analytic approach of previous neuroimaging studies, the monetary incentive delay (MID) task (Knutson et al., 2000) was used to independently localize regions active during anticipation and receipt of monetary rewards (Tamir & Mitchell, 2012; Zaki et al., 2011). The social cognition system localizer was a well-validated person judgment task (Ochsner et al., 2005) commonly used to identify regions involved in thinking about others' mental states and traits, here adapted such that perceivers made judgments about target group members and predicted targets' judgments of them.[1] For each functional localizer task we then defined 8mm radius spherical regions of interest (ROIs) surrounding activation peaks that fell within our *a priori* ROIs (see Methods). From the MID task we obtained anatomically constrained functional ROIs in vmPFC, ventral striatum, and amygdala (Fig. 2A). The person judgment task revealed clusters with peaks in dmPFC, precuneus, and bilateral TPJ (Fig. 2B). The activation peaks we found are consistent with previous neuroimaging studies using the MID (Hommer et al., 2003; Knutson et al., 2000; Tamir & Mitchell, 2012; Zaki et al., 2011) and person judgment tasks (Denny et al., 2012).

[1] As noted earlier, these are precisely the kinds of judgments which people are preferentially motivated to make about high-status (relative to low-status) targets.

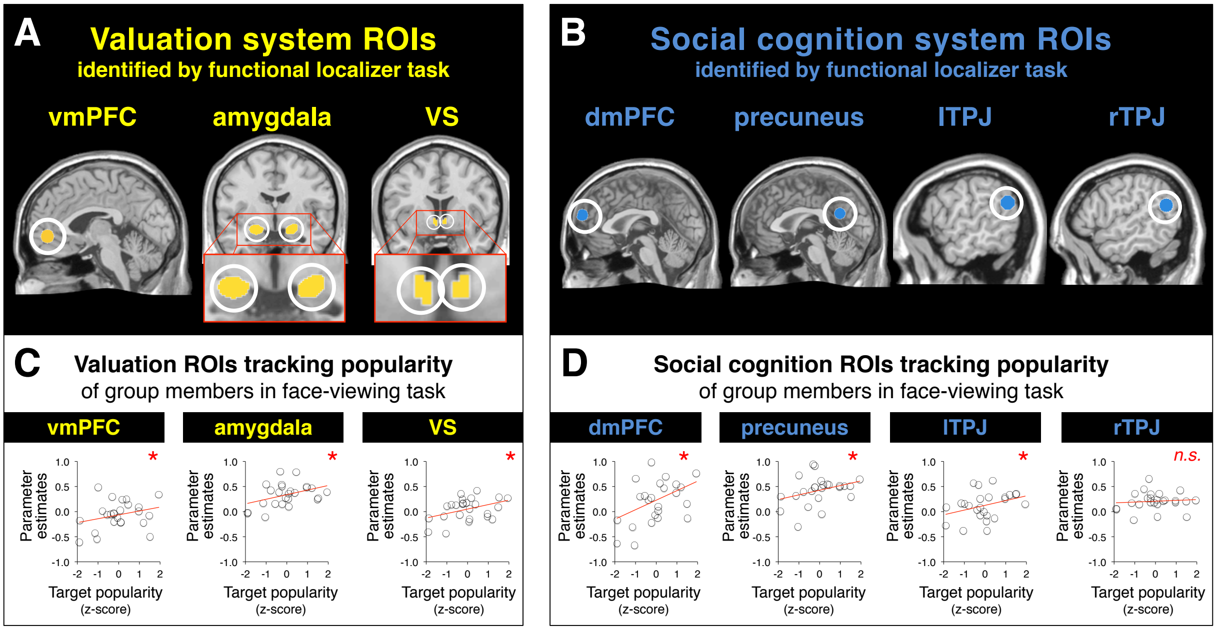

Fig. 2. Popularity of targets (group members presented as stimuli during the face-viewing task) predicted activity in each of the valuation and social cognition regions of interest (ROIs; all *P*s < 0.05) except rTPJ (*P* > 0.5), even when controlling for perceivers' own liking of target and other potential confounds (Table 2). Core brain regions underlying **(A)** valuation and **(B)** social cognition – and corresponding ROIs – were identified using two independent functional localizer tasks (Methods). Each task identified a set of commonly co-activated and strongly interconnected regions that are referred to collectively as the *valuation* and *social cognition* systems, respectively. Illustrations of the parametric relationship between target popularity and betas extracted from **(C)** valuation system ROIs and **(D)** social cognition system ROIs. Note that activity is averaged across perceivers for visual clarity. vmPFC, ventromedial prefrontal cortex; VS, ventral striatum; dmPFC, dorsomedial prefrontal cortex; lTPJ/rTPJ, left/right temporoparietal junction.

We then asked whether activation within these independently localized valuation and social cognition ROIs scaled with the popularity of targets presented in the face-viewing task. To answer this question we regressed activation parameter estimates (betas) extracted from each ROI against target popularity, controlling for each perceiver's liking of targets to ensure that analyses reflect neural sensitivity to how much target group members are *collectively* liked by the group and not merely *individually* liked by the perceiver.[2] Separate models were run for each ROI, with the dependent measure comprising beta values for all voxels within the ROI averaged together across the 10 repetitions of each target face presented to each perceiver. Linear mixed-effects models (lme4 and lmerTest packages for R) with random intercepts at the subject (perceiver) level were fitted using restricted maximum likelihood estimation (REML) with the appropriate degrees of freedom calculated using the Kenward-Roger method (Table 2)(Kenward & Roger, 1997). All hypotheses tests were two-sided with a statistical significance level of 0.05.

These analyses revealed that target popularity was positively associated with activity in each of the ROIs identified by the valuation localizer task (Fig. 2C; vmPFC parameter estimate ± SE: 0.122 ± 0.059, $P = 0.039$; amygdala: 0.089 ± 0.038, $P = 0.019$; VS: 0.083 ± 0.041, $P = 0.047$) and social cognition localizer task (Fig. 2D; dmPFC: 0.194 ± 0.079, $P = 0.015$; precuneus: 0.127 ± 0.055, $P = 0.022$; lTPJ: 0.124 ± 0.059, $P = 0.038$). The only ROI in which activity did not track target popularity was rTPJ (0.029 ± 0.045, $P > 0.5$), and was therefore was not included in the subsequent analyses.

To rule out alternative explanations—that activation in valuation and social cognition regions reflected variables potentially confounded with popularity rather than popularity *per se*—additional regression analyses were run that accounted for perceiver-target relational characteristics (i.e., length of relationship, frequency of contact, hours per week spent together, and subjective ratings of interpersonal closeness and similarity) and target attributes (i.e., age, sex, and normative ratings of facial attractiveness and trustworthiness) by including these

[2] It is important to note that correlation between perceivers' personal liking of targets and the rest of the group's liking of those targets was relatively low ($r = 0.22$), corresponding to less than 5% of variance explained. Moreover, whether or not perceiver's own liking of target was partialled out, target popularity predicted activity in both valuation and social cognition systems (Ps < 0.01; see below for details about system-level aggregation across valuation ROIs and social cognition ROIs, respectively).

covariates as fixed effects in the linear mixed-effect models described above. The positive association between target popularity and parameter estimates extracted from each ROI proved robust, remaining statistically significant after accounting for the effect of these potential confounds (all Ps < 0.05; Table 2).

To verify that these target popularity effects did not differ between groups, an additional model was run for each ROI that included all of the aforementioned parameters as well as two additional fixed effects for group membership (with -0.5 and 0.5 effect coding for each of the respective groups) and the interaction between group and target popularity. The main effect of target popularity on parameter estimates was robust to the inclusion of these additional parameters (all P's < 0.05, two-tailed; for VS, $P = 0.06$, two-tailed); importantly, there was no main effect of group or interaction between group and target popularity for any ROI (all Ps $>$ 0.2).

Table 2. Results of linear mixed-effects models (one model for each ROI) regressing activation parameter estimates against target popularity and covariates (i.e., to rule out alternative explanations and potential confounds). Bolded values connote $P < 0.05$, two-tailed. df reflects degrees of freedom calculated using the Kenward-Roger method (Kenward & Roger, 1997). ROI abbreviations as for Fig. 2.

	vmPFC				amygdala				VS			
target variables	**Est.**	**SE**	**df**	***P***	**Est.**	**SE**	**df**	***P***	**Est.**	**SE**	**df**	***P***
popularity	0.171	0.062	195.0	**0.006**	0.106	0.042	194.0	**0.013**	0.099	0.047	195.7	**0.038**
facial trustworthiness	-0.019	0.005	190.1	**0.001**	-0.009	0.004	190.1	**0.016**	-0.004	0.004	191.7	0.350
facial attractiveness	0.012	0.005	189.2	**0.013**	0.007	0.003	189.5	**0.049**	0.003	0.004	190.9	0.394
male	-0.342	0.107	184.9	**0.002**	-0.151	0.072	186.2	**0.039**	-0.120	0.081	187.4	0.142
age	0.019	0.024	185.3	0.436	0.017	0.017	186.6	0.311	-0.006	0.019	187.8	0.729
relational variables												
personal liking	-0.007	0.004	203.0	0.063	-0.003	0.003	202.0	0.224	0.000	0.003	202.9	0.983
perceived closeness	0.002	0.003	196.7	0.497	0.002	0.002	202.2	0.390	0.001	0.002	199.6	0.552
perceived similarity	0.004	0.003	199.7	0.183	0.002	0.002	197.6	0.266	0.002	0.002	199.5	0.311
length of relationship	-0.012	0.014	188.8	0.414	-0.002	0.010	199.6	0.838	-0.015	0.011	194.4	0.189
hrs/wk spent together	0.044	0.028	203.0	0.118	0.008	0.019	202.0	0.694	0.003	0.021	202.9	0.870
frequency of contact	0.005	0.014	196.2	0.739	-0.005	0.010	194.4	0.616	-0.008	0.011	196.5	0.432
(intercept)	0.106	0.715	192.1	0.882	0.106	0.486	194.8	0.828	0.300	0.546	194.2	0.583

	dmPFC				precuneus				lTPJ				rTPJ			
target variables	**Est.**	**SE**	**df**	***P***	**Est.**	**SE**	**df**	***P***	**Est.**	**SE**	**df**	***P***	**Est.**	**SE**	**df**	***P***
popularity	0.223	0.087	195.6	**0.011**	0.162	0.059	196.0	**0.007**	0.171	0.069	197.6	**0.014**	0.065	0.050	197.7	0.195
facial trustworthiness	-0.020	0.007	191.2	**0.007**	-0.012	0.005	191.6	**0.016**	-0.009	0.006	193.2	0.133	-0.006	0.004	193.6	0.154
facial attractiveness	0.015	0.007	190.3	**0.034**	0.012	0.005	190.7	**0.011**	0.007	0.005	192.2	0.177	0.005	0.004	192.7	0.200
male	-0.454	0.150	186.4	**0.003**	-0.325	0.102	186.8	**0.002**	-0.215	0.119	188.2	0.071	-0.221	0.086	189.0	**0.011**
age	0.016	0.034	186.9	0.640	-0.025	0.023	187.2	0.292	-0.042	0.027	188.4	0.123	-0.025	0.020	189.3	0.209
relational variables																
personal liking	-0.005	0.005	203.0	0.364	-0.008	0.004	202.9	**0.033**	-0.004	0.004	201.9	0.386	-0.003	0.003	202.3	0.357
perceived closeness	0.007	0.004	198.1	0.103	0.004	0.003	196.7	0.229	0.000	0.003	190.7	0.884	0.001	0.002	193.0	0.626
perceived similarity	0.000	0.004	199.8	0.955	0.002	0.003	200.2	0.478	0.000	0.003	201.7	0.934	0.000	0.002	201.5	0.924
length of relationship	-0.002	0.020	191.4	0.931	-0.002	0.014	189.1	0.901	0.020	0.016	179.7	0.214	0.002	0.011	183.5	0.859
hrs/wk spent together	0.019	0.039	203.0	0.627	0.031	0.027	202.9	0.249	-0.015	0.031	201.8	0.634	0.009	0.022	202.2	0.692
frequency of contact	-0.015	0.020	196.5	0.457	0.003	0.013	197.1	0.802	0.000	0.015	199.1	0.989	0.002	0.011	199.0	0.857
(intercept)	0.329	1.004	193.2	0.744	1.542	0.685	193.1	**0.025**	1.589	0.793	193.0	**0.046**	1.179	0.574	193.8	**0.041**

Target popularity analyses: Whole-brain approach. To validate these results and complement our hypothesis-driven ROI analyses with a data-driven analytic approach, we also conducted a random-effects, parametric whole-brain regression analysis at the group level. Perceivers' neural responses to group members (targets) during the face-viewing task were modeled as a function of targets' popularity (as a subject-level random slope parameter), again controlling for each perceiver's personal liking of individual targets (as a fixed effect) and including a random intercept for each perceiver. By fitting perceiver-specific random slopes and conducting a whole-brain search for regions in which these random slopes differed significantly from 0, this mixed-effects approach was best suited to answer the question: Are there any regions in which activity reliably (i.e., across perceivers) corresponds to target popularity?

This analysis replicated the ROI-based analysis: the same core valuation (vmPFC, amygdala, VS) and social cognition (dmPFC, precuneus, left TPJ) regions tracked significantly with target popularity (Fig. 3 and Table 3; whole-brain FWE-corrected $P < 0.05$ with uncorrected $P < 0.001$, k = 44 voxels; for amygdala and VS, FWE-corrected $P < 0.05$ using small volume correction). The analysis also revealed a cluster in middle frontal gyrus; as we had no specific predictions about the involvement of this region, the finding is reported without post hoc interpretation. It is worth noting that the whole-brain analysis utilized a two-tailed hypothesis to allow for testing of brain regions in which activity tracked *negatively* with target popularity; however, no such regions were found. See Fig. 4 for comparison of the distinct neural correlates of target popularity and liking.

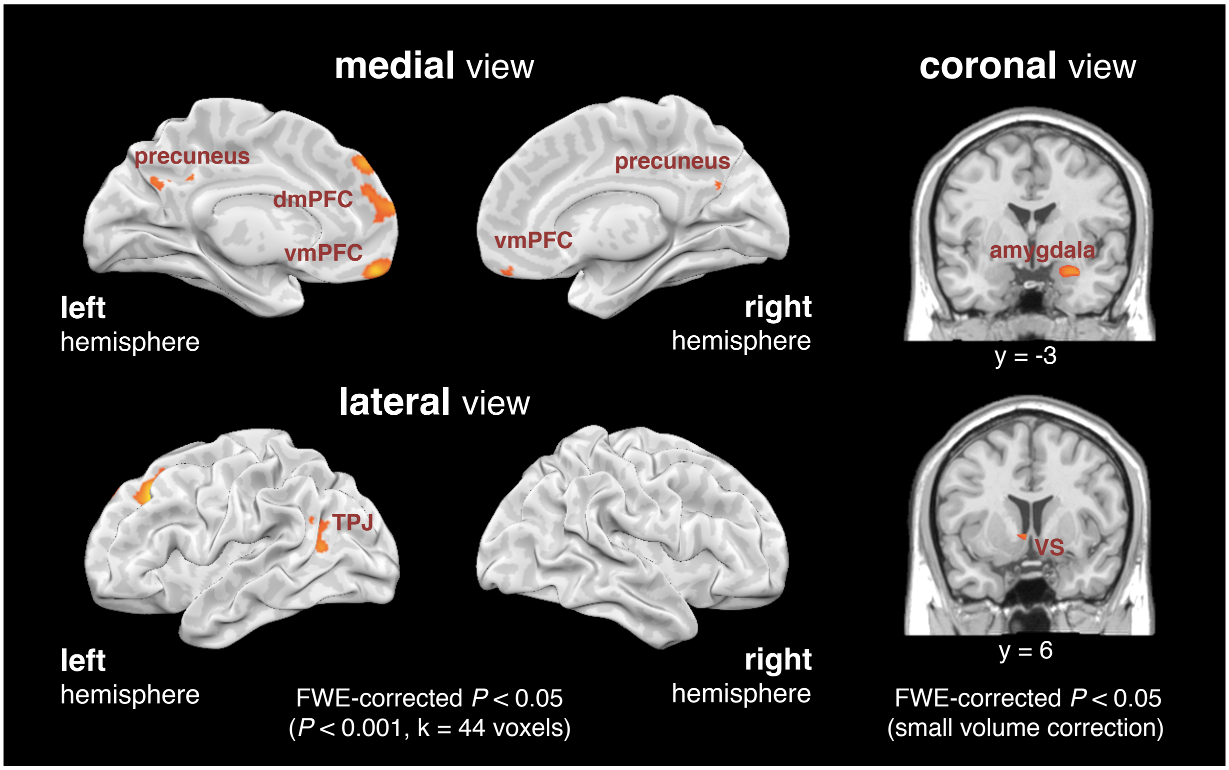

Fig. 3. Parametric whole-brain regression analysis isolating brain regions tracking target popularity during the face-viewing task (see Table 3). Clusters were thresholded at $P < 0.001$, k = 44 voxels (whole-brain FWE-corrected, $P < 0.05$, two-tailed) except as noted below. For ventral striatum and amygdala, subcortical structures of *a priori* interest, results were thresholded with small volume correction, FWE-corrected $P < 0.05$, two-tailed. Replicating the findings of the ROI-based approach, activity in the same core valuation (vmPFC, amygdala, ventral striatum) and social cognition (dmPFC, precuneus, left TPJ) regions tracked significantly with target popularity. The analysis also revealed a cluster in middle frontal gyrus, but was otherwise selective for our *a priori* hypothesized ROIs. Although the whole-brain analysis utilized a two-tailed hypothesis to allow for testing of brain regions in which activity tracked *negatively* with target popularity, no such regions were found. vmPFC, ventromedial prefrontal cortex; dmPFC, dorsomedial prefrontal cortex; TPJ, temporoparietal junction; VS, ventral striatum.

Table 3. Regions parametrically tracking target popularity during the face-viewing task, as identified by random-effects whole-brain analysis.

		MNI Coordinates				*t*-values	
Region		*x*	*y*	*z*	**k**	**Max**	**Mean**
vmPFC / Medial Frontal Gyrus (BA 11)	M	-6	51	-15	81	6.76	4.61
dmPFC / Superior Frontal Gyrus (BA 10)	M	-9	60	18	54	6.19	4.41
TPJ / Superior Temporal Gyrus (BA 39)	L	-39	-54	21	65	5.46	4.29
Precuneus (BA 7)	M	0	-48	39	60	4.71	4.21
Middle Frontal Gyrus (BA 8)	L	-24	30	39	189	7.43	4.84
Amygdala *	R	33	-3	-21	--	4.34	3.74
Ventral Striatum / Caudate *	L	-6	3	-3	--	4.36	3.34

Coordinates (in MNI space) refer to the peak activation in each cluster. All clusters were thresholded at $P < 0.001$, k = 44 voxels (whole-brain FWE-corrected $P < 0.05$, two-tailed), except as noted below. For subcortical structures of *a priori* interest, * reflects thresholding with small volume correction, FWE-corrected $P < 0.05$, two-tailed.

The clusters revealed by the whole-brain analysis were then subjected to the additional analyses conducted for the ROIs (described above) to rule out alternative explanations for the observed relationship and to verify that these target popularity effects did not differ between groups. We found that the positive association between target popularity and parameter estimates extracted from each cluster remained statistically significant even with the inclusion of fixed effects for all of the potential confounds listed above as well as group membership and the interaction between group and target popularity (all Ps < 0.01); moreover, there was no main effect of group or interaction between group and target popularity for any of the clusters (all Ps > 0.2).

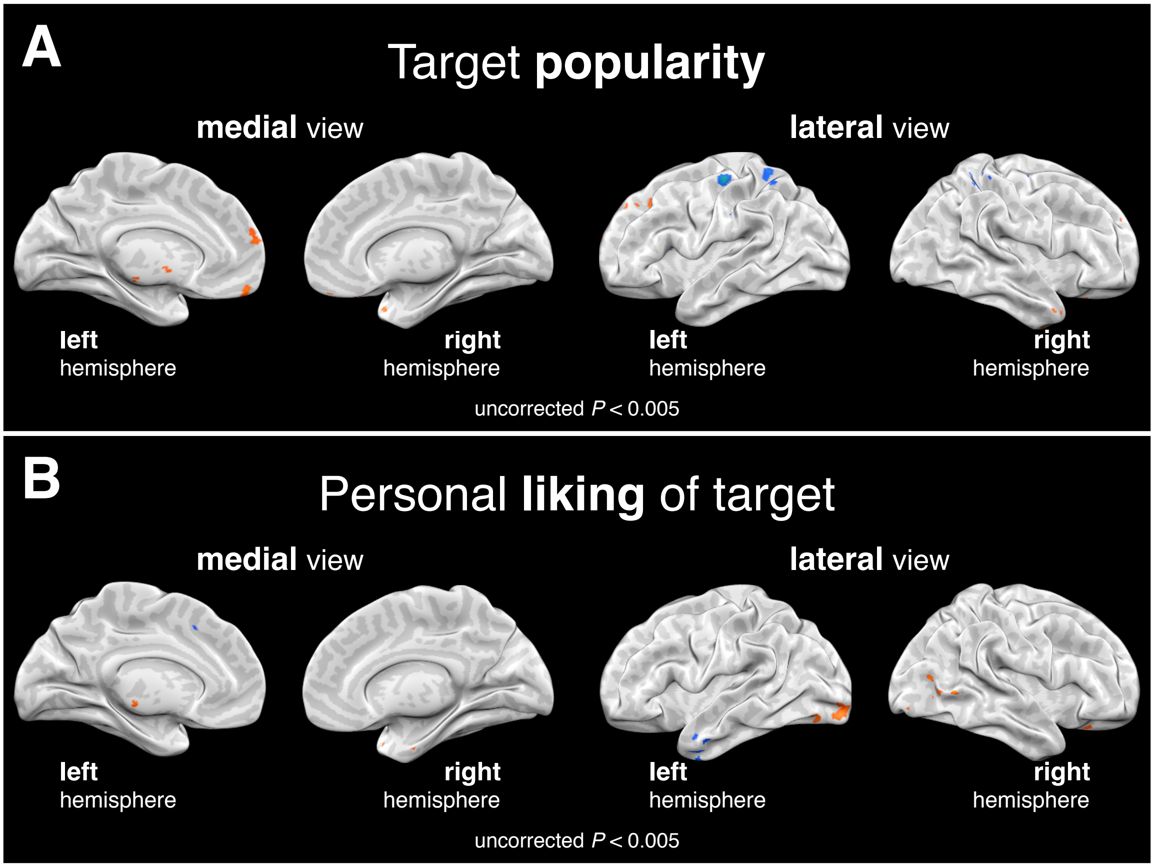

Fig. 4. Primarily distinct patterns of neural activity associated with tracking **(A)** popularity of targets and **(B)** each perceiver's own liking of targets during the face-viewing task. To enable more comprehensive comparison, parametric whole-brain regression analyses are presented with relaxed cluster threshold (uncorrected $P < 0.005$; orange and blue signify relative activation and deactivation, respectively). The results of this analysis provide empirical support for the conceptual distinction between sociometric popularity—the group's *collective* preference—and the *individual* preferences from which it is comprised. This is further corroborated by the observation that the perceiver's own liking of a target explains less than 5% of the variance in everyone else's liking for that target.

Mediation analyses. The observed correlations between target popularity and activity in valuation and social cognition regions confirmed our primary hypotheses, which led to our second question: do the two systems track popularity in parallel (independently) or serially, with one system assuming a primary role that mediates the popularity-activity relationship for the other? We predicted the valuation system would function as mediator based on the aforementioned literatures in social psychology [i.e., it is high-status group members' social importance that motivates others to predict their mental states (Dépret & Fiske, 1993; Fiske, 1993; Snodgrass, 1985, 1992)] and nonhuman primate neurophysiology [i.e., neurons in valuation regions encode social value and signal presence of high-status group members (Azzi et al., 2012; Klein & Platt, 2013; Klein et al., 2009; Watson & Platt, 2012)].

To test this prediction we performed multilevel mediation analyses, assessing whether valuation system activity explains the observed relationship between target popularity and social cognition system activity. First, parameter estimates extracted from vmPFC, amygdala, and VS—ROIs that had been independently localized by the MID task—were averaged to compute a composite measure of valuation system activity during the face-viewing task. Parameter estimates from dmPFC, precuneus, and lTPJ—ROIs that had been independently localized by the person judgment task—were likewise aggregated to compute a composite measure of social cognition activity while viewing group members' faces. Multilevel mediation analyses were then implemented via the gsem (generalized structural equation model) estimation command in Stata 13 (StataCorp, 2013), with social cognition activity as the dependent "Y" variable, valuation activity as the mediator "M" variable, and target popularity as the predictor "X" variable. As with the linear mixed-effects models in the target popularity main effect analyses above, the generalized structural equation models included random intercepts at the subject (perceiver) level and perceivers' personal liking ratings of targets as a covariate. Thus, the mediation analysis enabled us to quantify and statistically evaluate the extent to which increases in social cognition activity evoked by popular targets (independent of how much the individual perceiver liked them) were mediated by associated increases in valuation activity. The nlcom command in Stata 13, which computes 'delta method' standard errors, was used to conduct two-tailed significance tests of indirect paths.

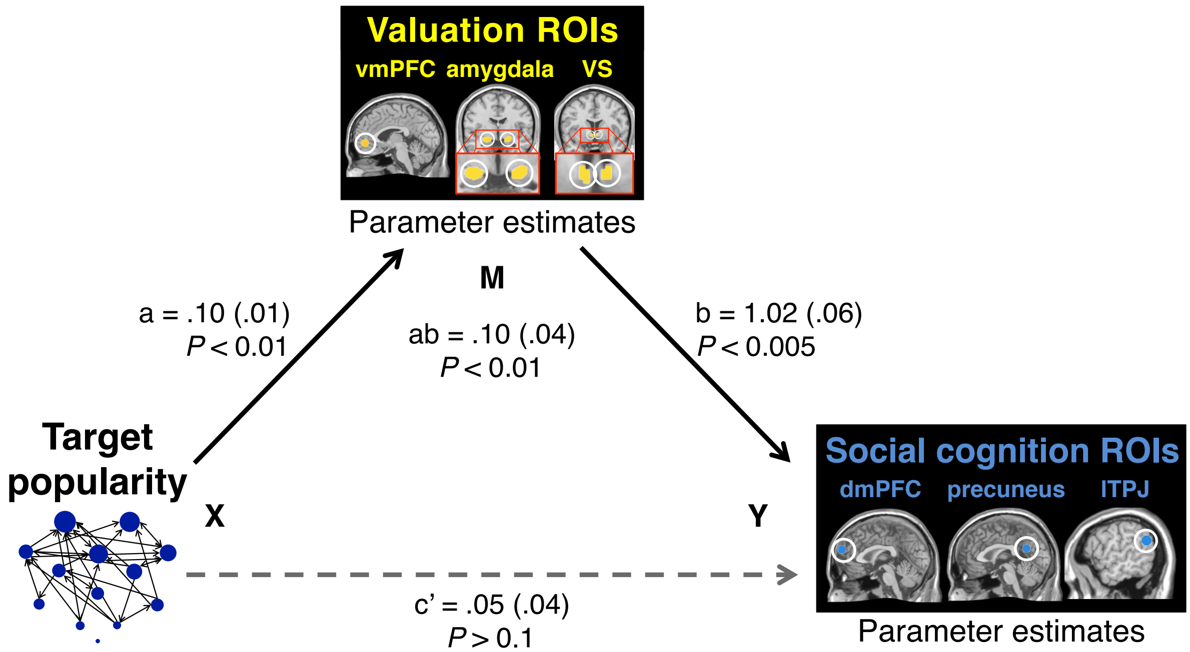

Fig. 5. Activity in the valuation system (vmPFC, amygdala, and ventral striatum ROIs independently localized by the MID task) mediated the observed relationship between target popularity and social cognition system activity (dmPFC, precuneus, and left TPJ ROIs independently localized by the person judgment task), with 64.6% of the total effect mediated ($P < 0.01$). See Fig. 2 and Methods for details on how these systems were defined and independently localized. Further analyses confirmed that the data supported this mediation model over both (1) the alternative serial organization in which social cognition system activity operated as the mediator, and (2) the parallel organization in which the two systems' activity independently tracked target popularity.

We found that valuation activity mediated the association between target popularity and social cognition activity (see Fig. 5; indirect effect parameter estimate ± SE: 0.100 ± 0.039, $P <$ 0.01), with 64.60% of the total effect mediated. Moreover, as assessed by Akaike's information criterion (AIC) and Bayesian information criterion (BIC), model fit was greater with valuation activity as the mediator (AIC = 734.43; BIC = 772.25) than with (1) the parallel organization in which the valuation activity and social cognition activity independently tracked target popularity (AIC = 902.10; BIC = 936.48), and (2) the alternative serial organization in which social cognition activity operated as the mediator (AIC = 736.96; BIC = 774.78). To evaluate the relative strength of these models and aid interpretation, AIC and BIC raw values were transformed into AIC and BIC weights, respectively (Wagenmakers & Farrell, 2004). According to either measure, the model with valuation activity as mediator had 3.54 times greater strength of evidence than did the other two models combined.

These results suggest that (1) a primary representation of sociometric popularity is value-based or motivational in nature, and (2) social cognitive systems may be engaged in the presence of popular group members to the extent that valuation systems signal their motivational significance. In such cases, social cognition systems may ready perceivers for effective interaction by supporting retrieval of knowledge about what target individuals are like and how they view us (precisely the two kinds of judgments elicited by the social cognition functional localizer task). This knowledge is useful for predicting high-status individuals' behavior and deciding how to act accordingly (Dépret & Fiske, 1993; Fiske, 1993; Snodgrass, 1985, 1992).

Perceiver popularity: Neural analysis of interpersonal sensitivity. The finding that valuation system activity directly tracked target popularity led to our third question: does the strength of this relationship (i.e., attunement to group members' popularity differences) relate to one's own popularity? In studies both of adults and children, popular individuals have more accurate perceptions of the affiliative social network structure that underlies differences in popularity (Bondonio, 1998; Casciaro, 1998; Krantz & Burton, 1986). In addition, human and nonhuman primate experiments have shown that although low-status individuals pay attention to group members of any status, high-status group members attend selectively to one another (Lansu et al., 2013; Shepherd, Deaner, & Platt, 2006). Therefore, we hypothesized that (1) perceiver popularity would amplify the effect of target popularity on valuation system activity, i.e., that valuation system activity of popular (relative to unpopular) perceivers would be more sensitive to status differences among group members, and (2) this effect would be driven by popular perceivers' attenuated responses to less popular targets.

Following the same analytic procedures as for the target popularity main effect analyses above, linear mixed-effects models (lme4 and lmerTest packages for R) with random intercepts at the subject (perceiver) level were used here to predict valuation activity parameter estimates. The models included fixed effects for target popularity, perceiver popularity, and their interaction term. As with the previous analyses, perceivers' personal liking ratings of targets were included as a covariate and linear mixed-effect models were estimated using REML with the appropriate degrees of freedom calculated using the Kenward-Roger method (Kenward & Roger, 1997).

We found that in addition to the main effect of target popularity (parameter estimate ± SE: 0.100 ± 0.037, $P < 0.01$), there was also an interaction such that the effect of target popularity on valuation activity was amplified for more popular perceivers (Fig. 6; 0.077 ± 0.037, $P < 0.05$). In other words, the valuation systems of popular perceivers were better calibrated to detecting the status differences among group members. This result is not an artifact of popular perceivers liking more popular targets. Consistent with our hypothesis and the aforementioned human and nonhuman primate findings (6, 37), the interaction effect was largely driven by an attenuation of responses to less popular targets in popular – but not unpopular – perceivers (Fig. 6). Moreover, the main effect of perceiver popularity showed a nonsignificant trend in the opposite (i.e., negative) direction (0.122 ± 0.078, $P = 0.13$). Considered in tandem, these results suggest that *popular* individuals demonstrate enhanced interpersonal *sensitivity* (i.e., attunement to group members' status differences), whereas *unpopular* individuals show more generalized interpersonal *responsiveness* (i.e., elevated valuation responses to all group members regardless of status).

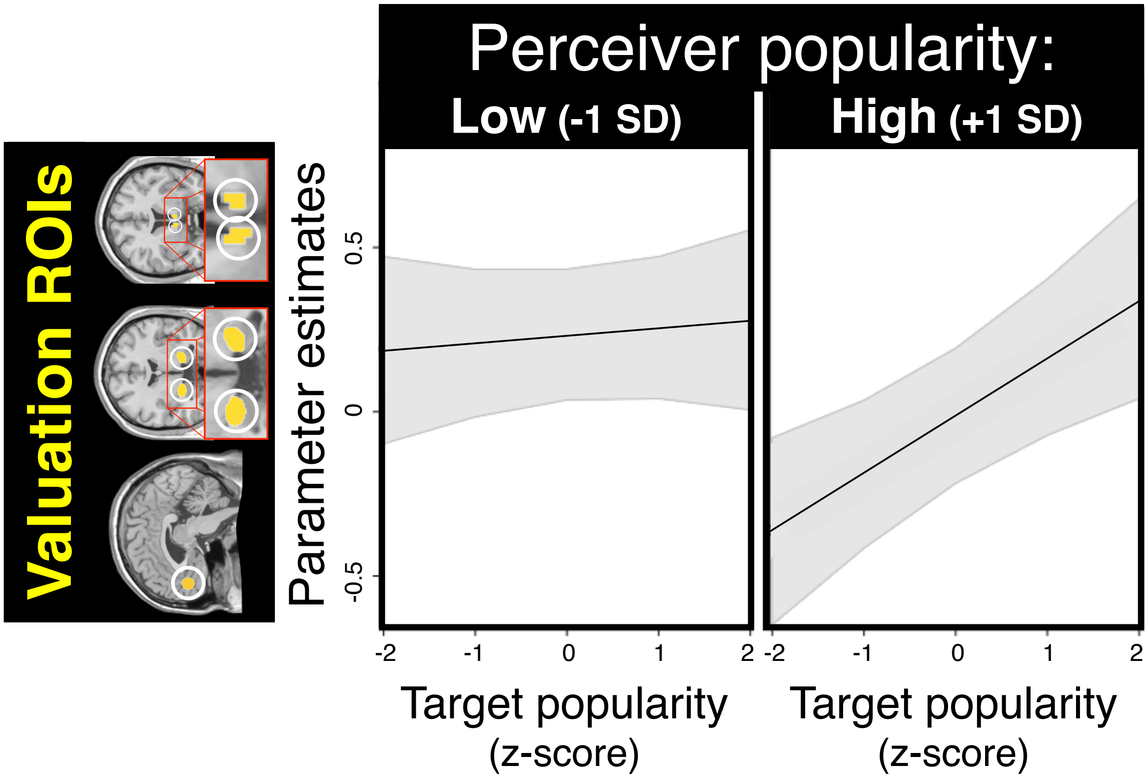

Fig. 6. Interaction plot depicting popular (+1 SD, relative to -1 SD unpopular) perceivers' enhanced attunement to group members' status differences (shaded area represents 95% CI). The main effect of target popularity on valuation activity ($P < 0.01$) was amplified for more popular perceivers ($P < 0.05$), suggesting their valuation systems were more sensitively calibrated to detecting status differences among group members. By contrast, there was a nonsignificant main effect trend of perceiver popularity in the opposite (i.e., negative) direction ($P = 0.13$), suggesting the valuation systems of unpopular individuals demonstrate greater generalized interpersonal responsiveness (i.e., elevated responses to all group members regardless of status).

Perceiver popularity: Behavioral analysis of interpersonal sensitivity. Having established a link between perceiver popularity and a neural measure of interpersonal sensitivity, we examined whether this finding would be corroborated using a social psychological (behavioral) measure of social acuity. During the behavioral (first) session, the same participants assessed each of their group members on a range of personality traits; in addition, they predicted how each of their group members' judged them on these same traits. In this computerized paradigm (E-Prime 2.0), participants used a sliding visual analog scale (anchored by the labels "not very" and "very" on opposite ends) to judge the extent to which trait adjectives described each group member and also predict how each group member would judge them on these traits. We computed the Pearson correlation between each perceiver's *predicted* personality profile (i.e., the perceiver's predictions about how a specific group member would judge the perceiver on various personality attributes) and the corresponding individual's *actual* personality profile of the perceiver (i.e., how that particular group member actually judged the perceiver on various personality attributes). In other words, a single correlation coefficient was computed for each dyadic pairing of matched *predicted*-*actual* personality profiles across all trait items. These dyadic *predicted*-*actual* profile relationships were transformed using Fisher's r-to-z transformation (i.e., from Pearson correlation coefficients to Fisher z scores) and aggregated to compute each perceiver's average profile relationship (across the perceiver's twelve *predicted-actual* profile relationships, one for each of the other twelve group members).

The resulting individual-difference measure of social acuity—termed overall meta-accuracy (Carlson, Furr, & Vazire, 2010)—was then correlated with perceiver popularity. Consistent with our prediction, perceivers' popularity was positively associated with their accuracy in predicting how individual group members assessed their personality across all items ($r = 0.44$; $P < 0.05$, two-tailed; see Fig. 7). This finding corroborates the neural perceiver popularity result and its interpretation as reflecting enhanced interpersonal sensitivity. Further, it dovetails with previous research demonstrating that popular adults and children more accurately perceive network members' interpersonal sentiments (7, 35, 36).

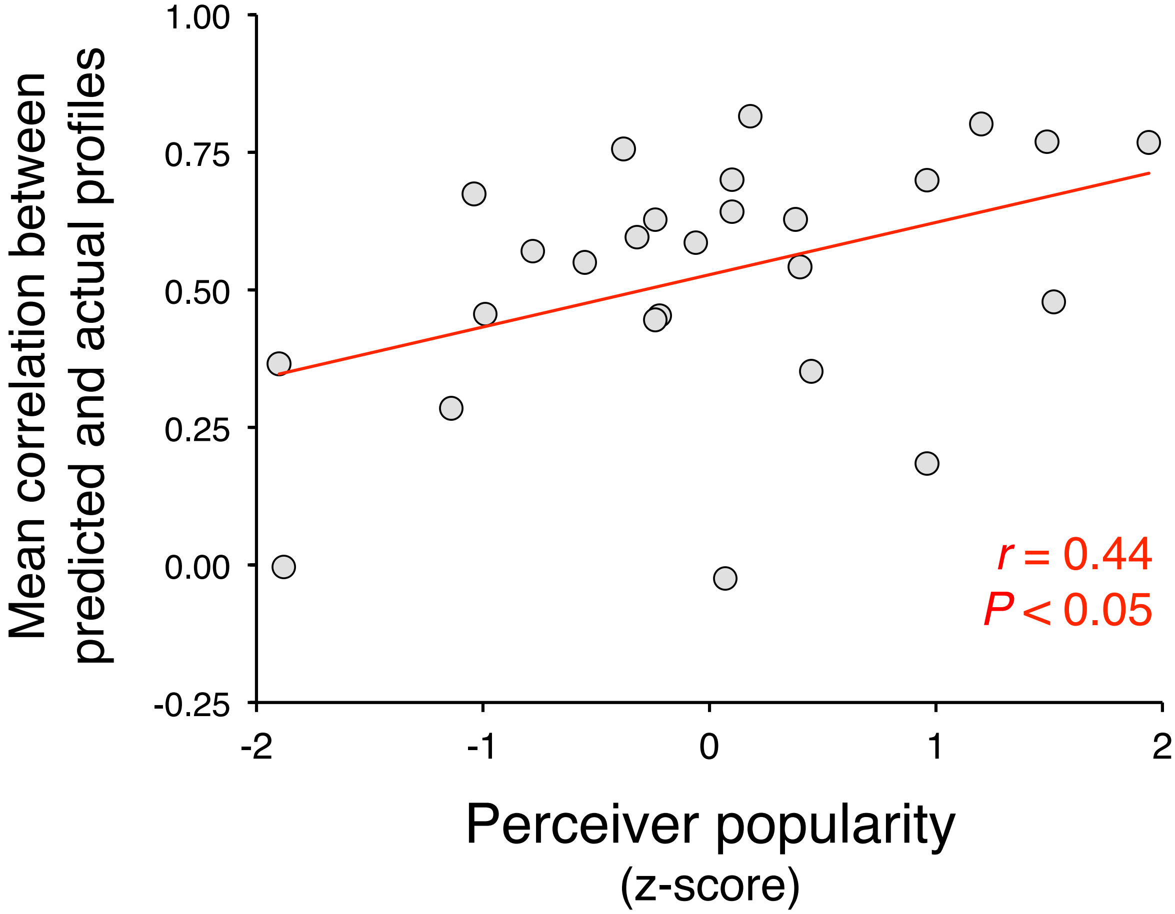

Fig. 7. Popular perceivers were more accurate in their predictions of how individual members of their organization assessed their personality across all trait items (r = 0.44; P < 0.05, two-tailed; n = 26). We computed the Pearson correlation between each perceiver's *predicted* personality profile (i.e., the perceiver's predictions about how a specific group member would judge the perceiver on various personality attributes) and the corresponding individual's *actual* personality profile of the perceiver (i.e., how that particular group member actually judged the perceiver on various personality attributes). In other words, a single correlation coefficient was computed for each dyadic *predicted-actual* personality profiles across all trait items. These dyadic pairings of matched *predicted-actual* profile relationships were transformed from Pearson correlation coefficients to Fisher z scores using Fisher's r-to-z transformation and aggregated to compute each perceiver's average profile relationship (across the perceiver's twelve *predicted-actual* profile relationships, one for each of the other twelve group members). The resulting individual-difference measure of social acuity (i.e., perceiver mean correlations plotted along the y-axis) reflects Fisher z scores that have been transformed back into Pearson correlations to aid interpretation.

Discussion

Taken together, the present results provide the first examination of neural mechanisms tracking popularity. Using a naturalistic face-viewing task, we identified two kinds of neural systems activated during encounters with members of real-world social networks. Affective valuation regions may assign motivational significance to group members based on their sociometric popularity and, in turn, may mediate engagement of social cognition regions that support understanding their mental states.

This neural mechanism presents adaptive features for navigating interactions within complex social networks. Tracking group members' status serves vital functions supported by valuation regions, e.g., assigning motivational importance to particular individuals, monitoring and detecting their presence, and signaling they deserve privileged status in attention and decision-making (Azzi et al., 2012; Klein & Platt, 2013; Klein et al., 2009; Krienen et al., 2010; Watson & Platt, 2012). In an experimental demonstration of this principle, rhesus macaques were willing to sacrifice fruit juice in order to view faces of high-status group members, while requiring overpayment of juice to view low-status monkeys' faces (Deaner, Khera, & Platt, 2005). Given the valuation system's critical role in reward processing and reinforcement learning (Haber & Knutson, 2009), this mechanism may also provide intrinsically rewarding reinforcement that motivates proximity and preferential attention to popular individuals as well as incentivizing interactions with them (Henrich & Gil-White, 2001; Klein et al., 2009; Lansu et al., 2013; Moreno, 1934; Vaughn & Waters, 1981). At the group-level, this neural mechanism may help stabilize social networks over time, thereby contributing to the self-reinforcing nature of social status (Magee & Galinsky, 2008; Solomon, 1942).

The mediation analysis suggests that the valuation system translates group members' popularity into motivational value signals that mediate activation of social cognition systems critical for explicit attributions about group members' psychological states and characteristics. Given our motivation to understand high-status individuals' mental states and predict their behavior (Dépret & Fiske, 1993; Fiske, 1993; Snodgrass, 1985, 1992), this neural mechanism may be both adaptive and socially advantageous: upon observing popular group members, it could proactively set in motion social-cognitive processes that facilitate social interaction.

The social advantage of this neural mechanism is further suggested by the results of our individual-differences analysis showing that perceivers' own popularity correlated with how strongly their valuation systems tracked network members' popularity. These intriguing findings are consistent with two views of how perceivers' own status relates to their perceptions of others.

One view comes from the social psychological literature on power, which suggests that having low power or subordinate status imbues other people with heightened relevance that motivates more careful attention to them and their perspectives (Fiske, 1993; Galinsky, Magee, Inesi, & Gruenfeld, 2006; Keltner, Gruenfeld, & Anderson, 2003). Our data suggest that differences in popularity may function in a similar way: as illustrated in Fig. 6, unpopular perceivers (left panel) demonstrated elevated valuation responses to all group members regardless of their status; by contrast, popular individuals (right panel) demonstrated valuation responses that scaled with targets' status. These results dovetail with evidence that although low-ranking monkeys and unpopular humans pay attention to group members of any status, their high-status counterparts attend selectively to one another (Lansu et al., 2013; Shepherd et al., 2006).

Another view consistent with our data is that popular individuals achieve their status because they are particularly skilled social perceivers. At the behavioral level, heightened interpersonal acuity has been linked to popularity in social networks of children (Krantz & Burton, 1986) and adults (Bondonio, 1998; Casciaro, 1998), and we likewise found that popular individuals more accurately predicted how individual group members viewed them. The findings in Fig. 6 could thus be interpreted as evidence at the neural level of popular individuals' enhanced social attunement, i.e., that their valuation systems were better calibrated to the social structure. On this view, perceivers' valuation responses to others might not reflect a *consequence* of perceivers' own status, but rather a *determinant* of how much status they ultimately achieve.

Consistent with this account which causally prioritizes valuation regions' functioning as influencing status, primate and rodent studies have shown that lesions to orbital prefrontal cortex and amygdala resulted in disrupted social behavior and loss of status, and manipulation of serotonergic neurotransmission and synaptic efficacy in mPFC influenced social skills, affiliative behavior, and changes in status (Wang, Kessels, & Hu, 2014). Although such experimental manipulations cannot be conducted in human research, the paradigm advanced here could be implemented longitudinally to investigate whether individual differences in the valuation system's social sensitivity are important determinants and/or consequences of one's ability and motivation to affiliate with group members and achieve status. Understanding the causal mechanisms underlying such individual differences in humans could have implications for clinical conditions such as depression and developmental disorders such as autism spectrum disorders, in which diminished interpersonal sensitivity, affiliative motivation, and social interaction have been linked to atypical valuation system structure and function (Chevallier, Kohls, Troiani, Brodkin, & Schultz, 2012; Healey, Morgan, Musselman, Olino, & Forbes, 2014).

More broadly, our findings are consistent with prior research showing that other aspects of network membership may also relate to the structure and function of valuation and social cognition systems. Recent studies [reviewed in (Dunbar, 2012)] have reported that individuals' social network size and/or complexity correlated with gray matter in vmPFC (Lewis, Rezaie, Brown, Roberts, & Dunbar, 2011; Powell, Lewis, Roberts, García-Fiñana, & Dunbar, 2012), amygdala (Bickart, Hollenbeck, Barrett, & Dickerson, 2012; Bickart, Wright, Dautoff, Dickerson, & Barrett, 2010; Kanai, Bahrami, Roylance, & Rees, 2012), and left TPJ (Lewis et al., 2011). Moreover, individual macaques' gray matter in mPFC and regions approximating human TPJ covary with both social network size (which was experimentally assigned) and social status (Noonan et al., 2014; Sallet et al., 2011). These findings support the proposition that affective valuation and social cognition systems are critical for navigating complex social networks and achieving high status within them.

Here it is important to note that prior neuroimaging studies examining processing of another dimension of social status – dominance – have not consistently implicated the valuation and social cognition systems observed here, but rather regions of lateral prefrontal cortex and inferior parietal lobe (Chiao, 2010). These differing findings could reflect the possibility that the relative dominance and sociometric popularity of group members are represented by different types of brain systems. But they could also reflect differences in methodology. Whereas our stimuli depicted members of participants' real-world groups in order to study naturally occurring variability in social status,[3] other human neuroimaging studies focusing on dominance have tended to experimentally manipulate social status with less naturalistic stimuli [see (Chiao, 2010) for review]. Our expectation is that for voluntary identity groups of comparable scale (8-79) where similar structural dynamics are observed, the findings reported here should be robust (Davis, 1970; Hallinan, 1974). As scale increases, mutual observation becomes impossible. Consequently the structuring dynamics of networks change (Bearman, 1997). Likewise, in groups with strong formal hierarchies, different dynamics may be observed. Future work could address these and other questions about the neural mechanisms that track popularity, specifically, and other kinds of social status in a wide range of social networks more generally.

In conclusion, this study advances an experimental paradigm that models group members' everyday encounters using a naturalistic task and personalized stimuli. In so doing, we provide an interdisciplinary framework that integrates theories and methods from social psychology, neuroscience (fMRI), and sociology (SNA) to enable research on the brain mechanisms underlying person perception and social cognition processes in real-world, status-laden social networks.

[3] Note that the nonhuman primate studies in which valuation regions were found to track group members' status (Azzi et al., 2012; Klein & Platt, 2013; Watson & Platt, 2012) also utilized similarly naturalistic stimuli (i.e., faces of group members).

Study 2:

The Neural Bases of Future Liking and Reciprocity

Introduction

Social scientists have long sought to understand the interpersonal forces that differentially attract group members to one another and shape how affective ties evolve within a network structure. For decades this line of research has been pursued primarily within a sociological framework that emphasizes social-structural phenomena, among the most important of which is reciprocity, long recognized as one of the key mechanisms underlying the evolution of social network ties in human groups (Bearman, 1997; Ekeh, 1968; Gouldner, 1960; Homans, 1958; E. E. Jones & Gerard, 1967; Leifer, 1981). Yet humans are a social species and it is likely that our brains have evolved under selective pressures to effectively navigate relationships within our dynamic social networks. Therefore, there are likely neural mechanisms associated with social-structural phenomena that shape interpersonal dynamics (e.g., reciprocity). At least since Freud, psychologists have postulated that an individual's intrapersonal processes and implicit motivations may explain the ways in which relationships unfold (Freud, 1910, 1912). With the emergence of behaviorism, interpersonal attraction was conceptualized as determined by reward value attributed to another person (Byrne, 1971; Newcomb, 1960) and reciprocity as emerging from the mutual reinforcement of this reward value between interacting dyad members (Newcomb, 1960). In this way a social-structural phenomenon like reciprocity—the key building block of social order—can be understood in regard to the intrapersonal processes through which it emerges. This article extends this line of inquiry by looking within the brain and identifies the neural underpinnings of reciprocity in human groups.

Advancements in functional neuroimaging (functional magnetic resonance imaging; fMRI) have made it possible for social scientists to investigate implicit psychological processes via their neural correlates. More recently, the *brain-as-predictor* approach has leveraged neural markers as *predictors* of participants' attitudes and behavior outside of the scanner (for review see (Berkman & Falk, 2013)). This approach has leveraged implicit neural measures of valuation to predict within-subject effects such as participants' unique preferences and choice behavior between various consumer products (Levy, Lazzaro, Rutledge, & Glimcher, 2011; Tusche, Bode, & Haynes, 2010), although to date such within-subject paradigm have not been extended to predict individuals' idiosyncratic social preferences. In neuroimaging research on social phenomena, the *brain-as-predictor* approach has typically been employed to explain differences between individuals; for instance, valuation system activity in response to erotic images, in-group happy facial expressions, and scenes depicting social interactions have been used to predict individual differences in participants' future sexual behavior, number of new in-group friendships formed, and daily time spent around other people, respectively (Demos, Heatherton, & Kelley, 2012; Powers, Chavez, & Heatherton, 2015). Human neuroimaging research has documented that the same brain regions which track individuals' subjective valuation of nonsocial objects also undergird social preferences: ventromedial prefrontal cortex (vmPFC), ventral striatum (VS), and amygdala, which together comprise the brain's core valuation system (Adolphs, 2003; Behrens, Hunt, Woolrich, & Rushworth, 2008; Chen, Welsh, Liberzon, & Taylor, 2010; Doré et al., 2014; Fareri & Delgado, 2014; Fehr & Camerer, 2007; Güroğlu et al., 2008; Izuma, Saito, & Sadato, 2008; R. M. Jones et al., 2011; Lin, Adolphs, & Rangel, 2012). Therefore, by extending the brain-as-predictor approach to within-subject social preferences, it should be possible to predict individual group members' personal allocation of affect and affiliation within the context of a complex social network.

We pursued this objective in the present study where we identify an implicit neural measure of interpersonal valuation that prospectively predicts how group members' liking ties evolve over time and reveal how reciprocity emerges in human groups. Specifically, we hypothesized that implicit measures of interpersonal valuation—operationalized by valuation-related neural activity in vmPFC, VS, and amygdala—could be leveraged to predict how new group members' liking ties would develop over the course of an intensive summer program. The study population consisted of 16 students involved in labor organizing who volunteered to spend nine weeks together. Over the course of the nine weeks, students spent time in smaller groups as well in the larger collective.

At the beginning of the program (T1), participants viewed faces of every other social network member while whole-brain fMRI data were collected. These neuroimaging data were analyzed against sociometric measures of liking collected at the beginning and end of the program (T2), as well as against individual node-level attributes (e.g., various demographic and personality variables). We find that greater neural activity in these valuation regions when an ego viewed an alter at T1 predicted ego's liking of the alter at T2, even when controlling for ego's initial liking as well as other potential confounds and predictors of eventual liking. More surprisingly, an ego's future liking of an alter was additionally predicted by *alter's* valuation activity at T1 when viewing ego. We propose several alternative interpretations of these intriguing results and suggest that they provide insight into the neural underpinnings of reciprocity. More broadly, this study advances a novel paradigm for researching the links between inter- and intra-personal mechanisms of social ties and their network structure.

Methods

Participants. Participants were 16 students who volunteered to spend nine weeks together to organize workers. All participants received monetary compensation and provided informed consent following the standards of the Columbia University Institutional Review Board.

Procedure and design. The T1 component of the study was comprised of two sessions. In a preliminary session, sociometric instruments and self-report questionnaires were administered, and photographs were taken of participants' faces (to be used subsequently in the fMRI face-viewing task). In a second session, participants underwent fMRI scanning while completing the face-viewing task described below tasks described below. For all computerized tasks in both T1 sessions, stimulus presentation and behavioral data acquisition were controlled using E-Prime 2.0 (Psychology Software Tools, Inc.). The T2 wave of data collection included sociometric assessments (described below) administered via Qualtrics online survey software after conclusion of the nine-week summer program. Additional data were collected for the purposes of other studies.

Sociometric assessment. Sociometric assessments of group members' affiliative relations and resulting network structure were conducted via a computerized peer-rating paradigm in which participants rated how much they liked each group member (presented in randomized order) on a sliding visual analog scale ranging from 0 to 100 anchored by the labels "not very" and "very" on opposite ends. This sociometric instrument provided a continuous measure of personal liking (i.e., affiliation tie strength) between group members.

Round-robin fMRI face-viewing task. Stimuli for the fMRI face-viewing task were prepared from photographs of participants. During the preliminary session, participants' faces were photographed with affectively neutral facial expression and gaze directed straight at the camera. These photographs were cropped and converted to grayscale images with equal luminance. In addition, a "ghost face" stimulus image representing the superimposition of all group members' faces was prepared for each group following methods used in prior face perception research (Taylor et al., 2009). The face-viewing task implemented a rapid event-related design that included 10 repetitions of each stimulus face presented in pseudorandomized order. Faces were presented for 1000ms and interstimulus intervals (ISIs) consisting of white fixation cross on black background were jittered between 1500 ms and 11500 ms (mean duration of ISI=3500 ms). Egos viewed faces of alters while performing a simple cover task (Taylor et al., 2009) in order to maintain their alertness throughout. Specifically, participants were instructed to press a button with their pointer (second) finger each time a group member's face was presented and a different button with their ring (fourth) finger each time a "ghost face" was presented (~9% of total presentations). Visual stimuli presented during the fMRI scanning session were displayed on a projection screen using a LCD projector and viewed via a rear-projecting mirror.

Valuation system regions-of-interest (ROIs). We first independently defined brain regions-of-interest (ROIs) related to valuation in a separate sample of participants (these data were published previously in (Zerubavel et al., 2015)). Following the established analytic approach of previous neuroimaging studies, the monetary incentive delay (MID) task (Knutson et al., 2000) was used to independently define regions active during anticipation and receipt of monetary rewards (Tamir & Mitchell, 2012; Zaki et al., 2011). We then defined 8-mm radius spherical ROIs surrounding activation peaks that fell within our a priori ROIs (and, for the subcortical structures, constrained them with anatomical masks). The activation peaks we found in the vmPFC, VS, and amygdala are consistent with previous neuroimaging studies using the MID (Fig. #) (Hommer et al., 2003; Knutson et al., 2000; Tamir & Mitchell, 2012; Zaki et al., 2011). Replicating our previous analytic protocol (Zerubavel et al., 2015), parameter estimates extracted from the vmPFC, VS, and amygdala and VS were averaged together to compute a composite measure of valuation system activity during the face-viewing task.

Imaging acquisition and analysis. Whole-brain fMRI data were acquired on a 1.5 Tesla GE system. High-resolution anatomical images with 1mm × 1mm × 1mm resolution were acquired with a T1-sensitive SPGR sequence at the end of the scan session. Functional images were acquired with a T2*-sensitive EPI blood oxygenation level dependent (BOLD) sequence.

Functional images were preprocessed using SPM8 software (Wellcome Department of Cognitive Neurology, UCL), including slice-timing correction, motion-correction, realignment, coregistration between each participant's functional and anatomical data, normalization to a standard template (Montreal Neurological Institute; MNI) using segmentation parameters, 3mm isometric voxels, and spatial smoothing using a Gaussian kernel (full-width at half-maximum = 6mm).

fMRI data were subjected to a first level of regression, separately for each subject, using an ordinary-least-squares general linear model (GLM) implemented with Neuroelf v0.9c software (neuroelf.net). Task-based regressors are described below. The GLM included, in addition to the task-related regressors, the 6 motion parameters as estimated during realignment as well as a DCT-based basis set covering low-frequency up to 1/80Hz to account for signal variability introduced by head motion and temporal drifts. The GLM also included one regressor for each target face (including the ghost face stimulus), representing the 10 repetitions of each respective face (12 repetitions for the ghost face). Each of these regressors was created by convolving the canonical hemodynamic response function (HRF) with a series of boxcars representing the 1000 ms intervals during which a particular face was presented. The output of these first-level regressions was a series of parameter estimate (beta) maps used in the next group level of analyses. Beta maps corresponding to trials on which participants viewed the ghost face were discarded prior to the next level of analyses.

Results

Independently defined valuation regions of interest (ROIs; see Methods) were interrogated for patterns of neural activity during the face-viewing task conducted at the beginning of the summer program (T1). Do group members' value-related neural responses to one another at T1 predict their ultimate liking after completing the summer program (T2)? To answer our primary question, we developed a model for regressing an ego's T2 liking against T1 valuation system activity (i.e., activation parameter estimates extracted from valuation ROIs; see Methods) as well as other potential predictors. This model and the subsequent analyses were approached using the nonparametric multiple regression quadratic assignment procedure (MRQAP) (Krackhardt, 1988), a standard method for analyzing social network data and performing the appropriate statistical significance tests with conservatively estimated standard errors (Dekker, Krackhardt, & Snijders, 2007). MRQAP handles row and column autocorrelation among observations in such data (e.g., interdependencies in liking ties sent and received by a given group member as ego or alter) by simultaneously permuting both the rows and the columns of matrix variables.

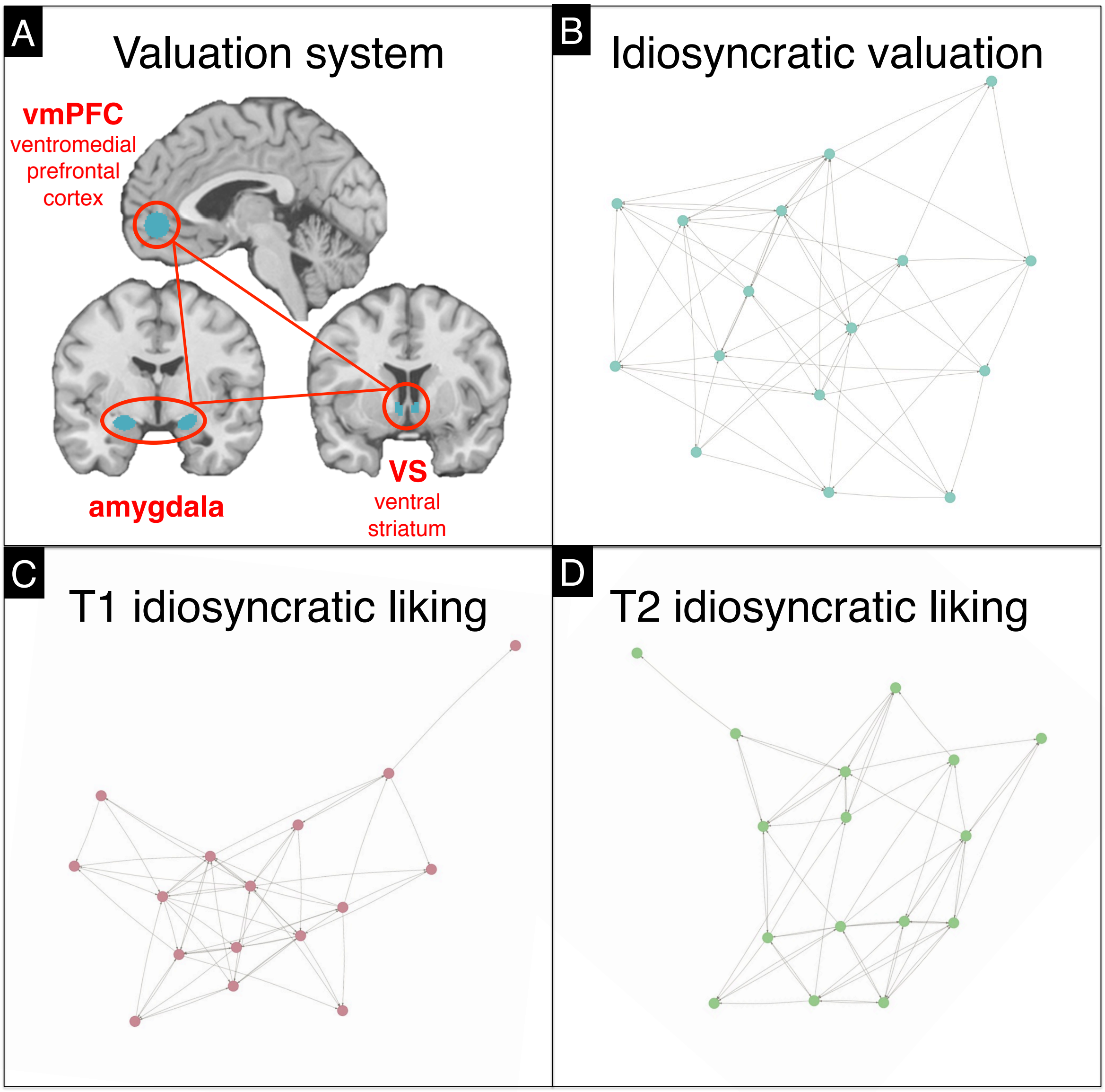

Fig. 8. (A) Valuation system regions of interest (ROIs). Core brain regions underlying valuation (vmPFC, VS, amygdala) were independently defined based on a functional localizer task conducted in a separate sample of participants (see Methods). Parameter estimates extracted from the vmPFC, VS, and amygdala and VS were averaged together to compute a composite measure of valuation system activity during the face-viewing task (15). **(B)** Social network structure of idiosyncratic valuation system activity during face-viewing task. Directional arrows represent idiosyncratic valuation activity of ego (participant in the fMRI scanner) while viewing alter (group member whose face was depicted as stimulus during fMRI scan). Social network structure of idiosyncratic liking relations directed from ego to alter **(C)** at T1 and **(D)** at T2. For visual clarity, only ties in the upper quartile are displayed in **(B)**, **(C)**, and **(D)**. vmPFC, ventromedial prefrontal cortex; VS, ventral striatum.

Predicting future liking. In our primary analysis regressing T2 liking based against T1 valuation system activity, we found that this relationship was indeed positive and significant (β = 0.175; P = 0.009). In other words, an ego's valuation activity in response to an alter presented during the face-viewing task at the beginning of the summer program (T1) predicted how much they ultimately liked that alter in the last week of the program (T2).

A different hypothesis that would be consistent with this initial result is that valuation activity was directly tracking concurrent (T1) liking, which in turn predicted T2 liking. To ensure that the association between activation of the valuation system and T2 liking was not merely due to both variables' association with T1 liking, we again used the MRQAP model to regress T2 liking against valuation activity, this time controlling for their initial (T1) liking. This analysis yielded several findings that clarified how these variables interrelated. Not surprisingly, liking at T1 predicted liking at T2 (β = 0.264; P < 0.001), although T1 liking only accounted for 6.95% of the variance in T2 liking (7.86% shared variance between T1 and T2 liking without including valuation activity as another predictor).

Critical to our primary hypothesis, the positive association between initial valuation activity and future liking remained significant after controlling for T1 liking (β = 0.144; P = 0.021), indicating this relationship is not merely driven by the effects of initial liking. On the contrary, these findings suggest that explicit (self-reported liking) and implicit (neural marker of valuation) measures at T1 operate as largely independent predictors of an ego's future liking of an alter. Considering together the results of this model and the primary analysis above, an ego's neural valuation of an alter predicted their future liking, whether or not concurrent (T1) liking was partialled out. As an additional test of the possibility that valuation activity directly tracked T1 liking (which in turn predicted T2 liking), another bivariate QAP regression model was conducted to assess the correlation between valuation activity and T1 liking. This association was not statistically significant ($\beta = 0.117$; $P = 0.114$) and descriptively weaker than that between valuation activity and T2 liking.

We then conducted several robustness checks to rule out alternative explanations. Extending the sociological models described above, these regression analyses incorporated additional covariates to control for other potential predictors of affiliation (including demographic and personality attributes of egos and alters, as well as homophily on these characteristics). The results of these analyses demonstrated that, even when controlling for each of these potential confounds, T1 valuation activity consistently remained a significant predictor of T2 liking (all Ps < 0.05).

This result, however, does not necessarily imply that group members' *idiosyncratic* neural valuations of one another at the beginning of the summer predict their *idiosyncratic* liking upon completing the program. Rather, our results might be driven by individual-level effects of egos, if the individuals who *generally* exhibited greater valuation activity (across all alters) were the same people who also *generally* liked everyone more at T2. The data did not support this alternative account, however, as the ego row averages (i.e., outdegree centrality effects) for valuation activity and T2 liking were not significantly correlated ($R = 0.230$; $P = 0.392$). Alternatively, individual-level alter effects might also account for our results, if certain alters who generally elicited greater valuation activity across all egos were also the same individuals who ultimately became at T2 the most collectively liked (i.e., sociometrically popular) group members. In this vein, in a previous neuroimaging study of longer-acquainted social network members, we found that the valuation system activity elicited by alters tracked their (i.e., alters') sociometric popularity even when this group consensus differed from the participant's own personal preferences (Zerubavel et al., 2015). In the current study, we observed that alter column averages (i.e., indegree centrality effects) for valuation activity and liking received by alters at T1 were highly correlated (R = 0.634; *P* = 0.008), meaning that valuation activity elicited by alters did track their concurrent (T1) sociometric popularity. But while the valuation activity elicited by alters did track their concurrent (T1) sociometric popularity, it did not predict their *future* (T2) sociometric popularity (R = 0.058; *P* = 0.831; see Fig. 9). Based on these results we conclude that the association between valuation activity and future liking is not simply due to an individual-level effect of either alter or ego characteristics (i.e., outdegree or indegree centrality effects, respectively).

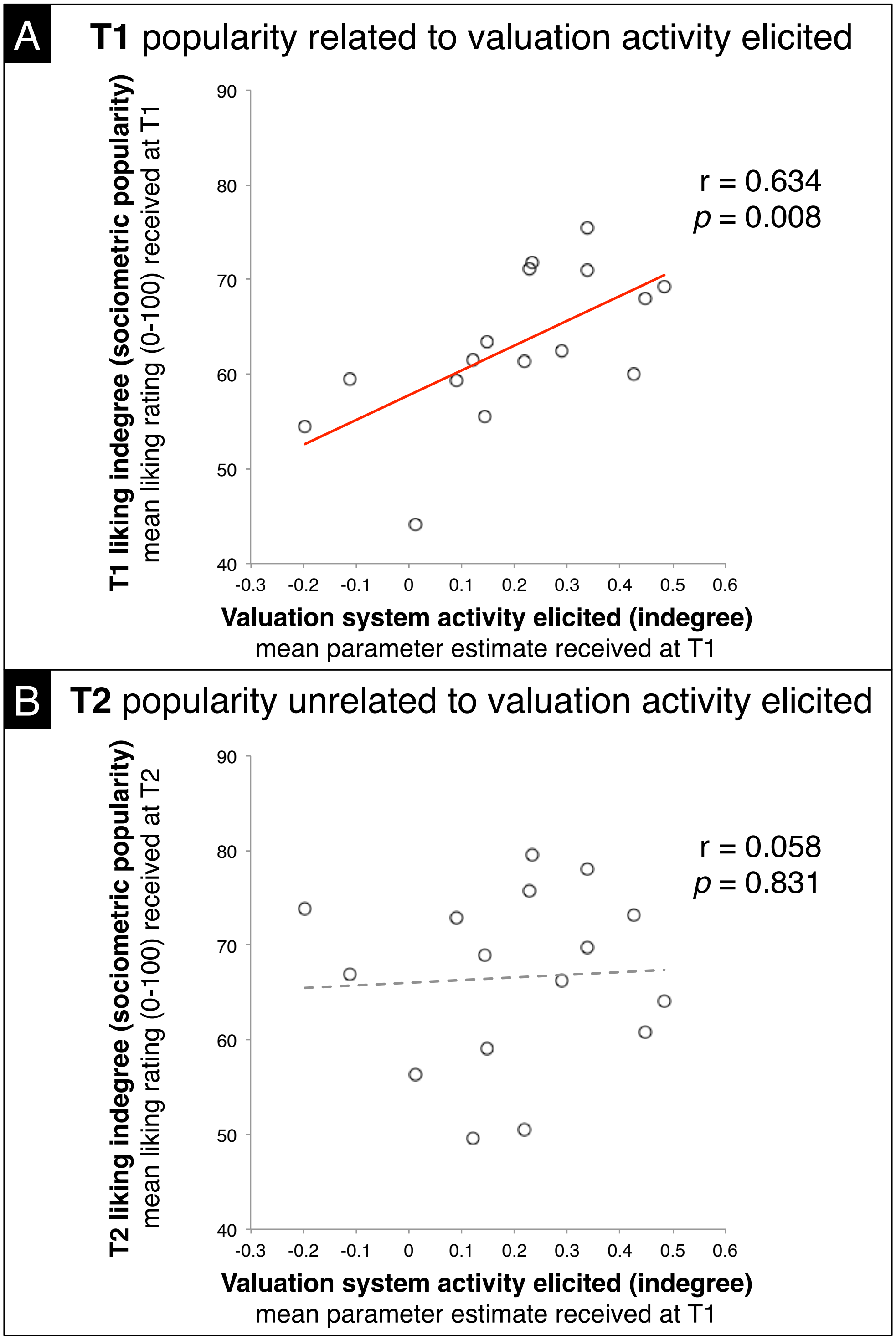

Fig. 9. Valuation activity elicited by alters tracks their concurrent (T1) sociometric popularity, but does not predict their *future* (T2) sociometric popularity. **(A)** Alter column averages (i.e., indegree centrality effects) for valuation activity and liking received by alters at T1 were highly correlated ($r = 0.634$; $P = 0.008$). **(B)** By contrast, this same neural index was not correlated with liking received by alters at T2 ($r = 0.058$; $P = 0.831$).

To directly test our interpretation that the results reflect *idiosyncratic* effects (i.e., unique to a particular ego about a particular alter) additional MRQAP models were implemented using variables that had been simultaneously mean-centered by ego and alter. Conceptualized through the lens of social network analysis, for each sociomatrix—valuation activity, T1 and T2 liking—this mean-centering procedure simply subtracted the row means and column means to remove any individual-level effects of outdegree and indegree centrality, respectively. Incorporating these variables mean-centered on ego and alter, we found that group members' unique neural valuation responses to one another did not track their unique patterns of idiosyncratic liking at T1 ($\beta = 0.089$; $P = 0.171$; see Fig. 11a). It is therefore striking that these idiosyncratic neural valuation responses *did* predict how much a particular ego would uniquely like a particular alter at T2 ($\beta = 0.179$; $P = 0.009$; see Fig. 11b), even when controlling for the ego's idiosyncratic liking of that alter at T1 ($\beta = 0.158$; $P = 0.014$). These analyses provide direct evidence in support of the hypothesis that neural processes link idiosyncratic valuation to future (T2) idiosyncratic liking.

Neural mechanisms underlying the emergence of reciprocity. What kind of mechanism might explain our findings that group members' unique valuation activity in response to one another predicts their unique liking at T2—but not in the present at T1? The previous analyses clearly indicated that a plausible mechanism will not operate on the individual-level (i.e., via ego and/or alter effects) and that we should instead focus our search for candidate mechanisms that link idiosyncratic valuation and liking variables at the dyadic or relational level.

Based on classical social theory and decades of empirical research spanning sociology, social psychology, and network analysis more generally, *reciprocity* is the relation-level phenomenon most implicated in the dynamics of affiliative ties and the evolution of their network structure (Bearman, 1997; Ekeh, 1968; Gouldner, 1960; Homans, 1958; E. E. Jones & Gerard, 1967; Leifer, 1981). In our data, at the behavioral level, the mutual reciprocation of idiosyncratic liking increased dramatically over the course of the summer program: at T1, unique liking from ego-to-alter and alter-to-ego shared only 2% variance (multiple $R^2 = 0.022$; adjusted $R^2 = 0.017$), compared to more than 20% by T2 (multiple $R^2 = 0.218$; adjusted $R^2 = 0.215$; see Fig. 10). Furthermore, an ego's idiosyncratic liking of an alter at T2 was predicted by the reciprocal alter's idiosyncratic liking of the ego at T1 ($\beta = 0.177$; $P = 0.007$), even when controlling for the ego's T1 idiosyncratic liking of the alter ($\beta = 0.143$; $P = 0.016$). These behavioral findings indicate that (a) the two dyad members' (idiosyncratic) interpersonal sentiments toward each other became much more closely aligned from T1 to T2, and (b) T2 idiosyncratic liking was influenced not just by one's own (i.e., ego-to-alter) T1 idiosyncratic liking, but also by the other dyad member's T1 idiosyncratic liking toward the ego.

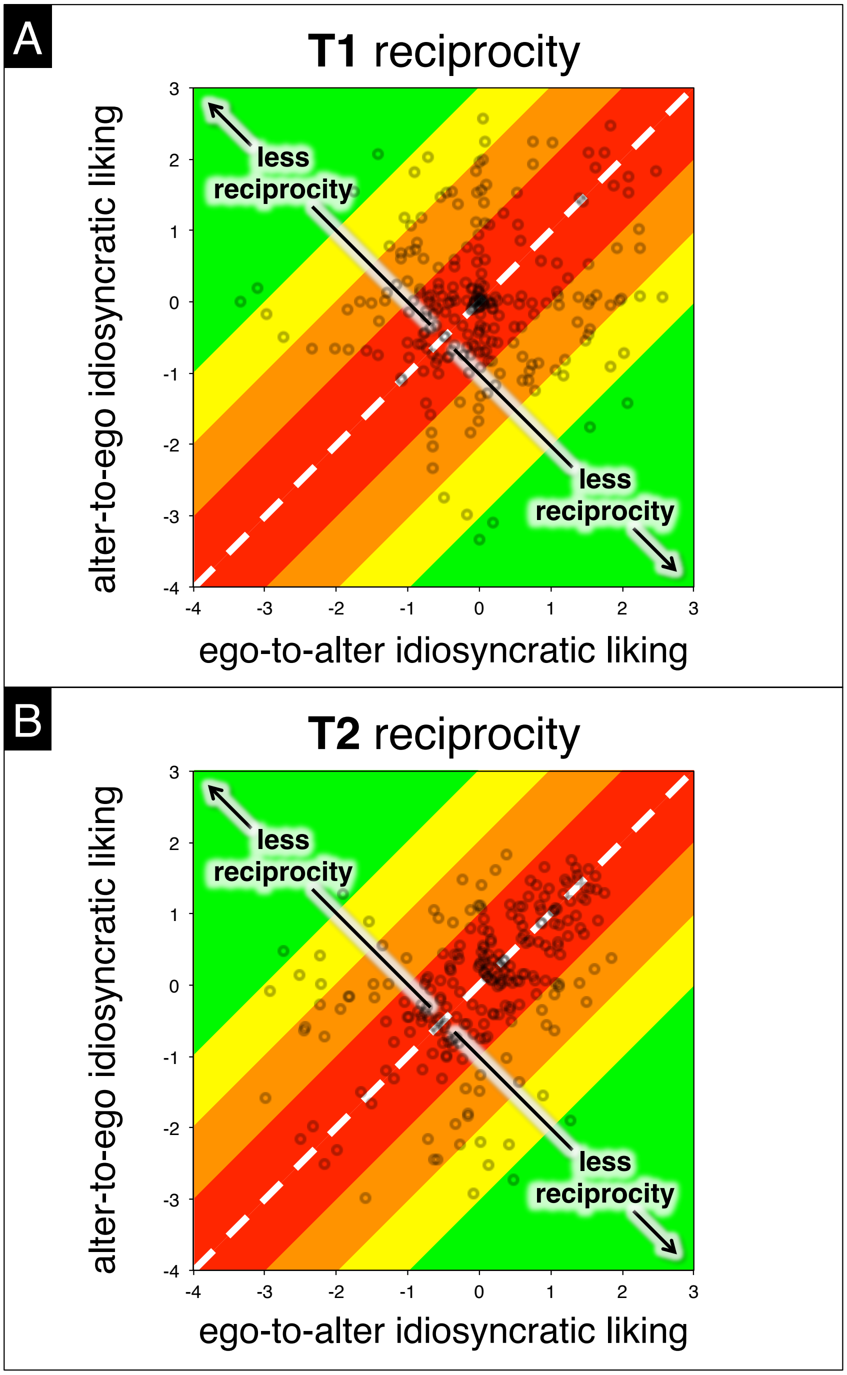

Fig. 10. Mutual reciprocation of idiosyncratic liking increased dramatically over the course of the summer program. **(A)** At T1, unique liking from ego-to-alter and alter-to-ego shared only 2% variance (multiple R^2 = 0.022; adjusted R^2 = 0.017). **(B)** By T2, this behavioral index of reciprocity had risen to more than 20% shared variance (multiple R^2 = 0.218; adjusted R^2 = 0.215). To illustrate the extent of reciprocity at each time point, idiosyncratic liking from ego-to-alter and alter-to-ego are plotted on the x-axis and y-axis, respectively. Perfect reciprocity between dyad members is indicated by the dashed line, y = x. Descending levels of reciprocity are indicated by the regions colored red (highest), orange, yellow, and green (lowest).

This led us to ask, *what are the neural mechanisms underlying how dyadic reciprocity emerges?* First, we considered the behavioral finding that an ego's T2 idiosyncratic liking of an alter was influenced by that alter's T1 idiosyncratic liking of ego and investigated the possibility of an analogous predictor at the neural level. Substantiating this premise, we found that unique ego-to-alter liking at T2 was predicted by unique *alter-to-ego* neural valuations at T1 ($\beta = 0.208$; $P = 0.0018$; see Fig. 11d). In relation to our primary hypothesis (i.e., a participant's neural valuation responses to others would predict his/her own unique liking of them in the future), this finding presents evidence in support of the reciprocal phenomenon: ego's unique neural valuation response to an alter predicted *that alter's idiosyncratic future liking of ego*. Finally, a comprehensive MRQAP model simultaneously regressed unique T2 ego-to-alter liking against four predictors: (1) an ego's T1 idiosyncratic liking of an alter; (2) the reciprocal alter's T1 idiosyncratic liking of ego; (3) an ego's idiosyncratic valuation response to an alter; and (4) an alter's idiosyncratic valuation responses to an ego. Corroborating the results of the previous analyses, we found that each of these four variables positively and significantly predicted an ego's idiosyncratic future liking of an alter (Ps < 0.05), with the exception of a trend-level effect in the same direction for an alter's unique liking of an ego at T1 ($P = 0.051$).

These results suggest that our future liking of group members can be jointly predicted by how our brains respond to them and how their brains respond to us. Moreover, these implicit neural measures of interpersonal valuation independently and reciprocally predict both dyad members' future liking above and beyond the effects of explicit measures (i.e., their self-reported liking of each other) collected at the same time.

Further elucidating the underlying mechanism, we found that an ego's unique neural valuations of an alter tracked how much that alter idiosyncratically liked the particular ego at T1 (β = 0.141; P = 0.029; see Fig. 11c), yet not their own (i.e., ego-to-alter) idiosyncratic liking at the same time point (β = 0.089; P = 0.171; see Fig. 11a). This pattern of results suggests that an ego's valuation system might be more sensitively tuned to an alter's unique interpersonal appraisals rather than to his/her own.

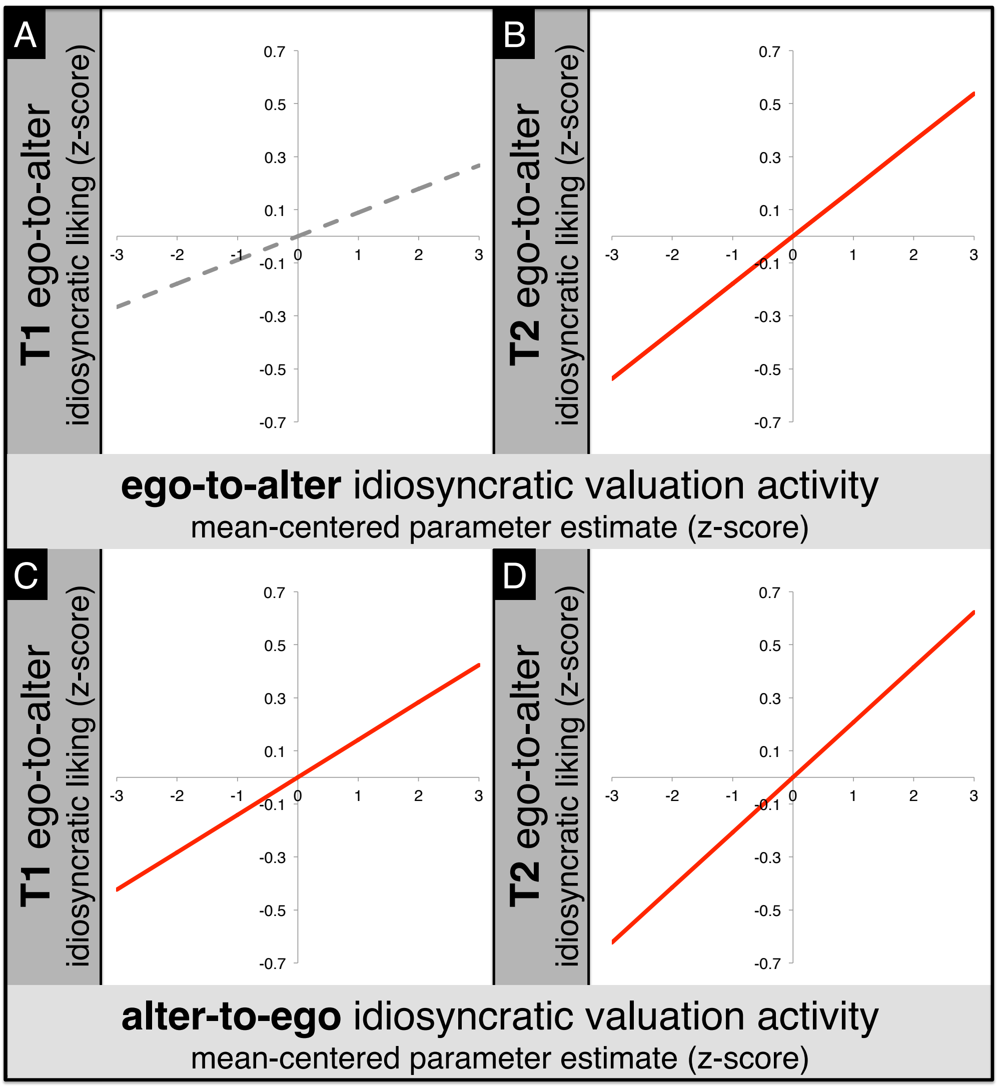

Fig. 11. (A) Egos' idiosyncratic neural valuation responses to alters did not significantly track their own (i.e., ego-to-alter) patterns of idiosyncratic liking at T1 (β = 0.089; P = 0.171). **(B)** These same (i.e., ego-to-alter) idiosyncratic neural valuation responses significantly predicted egos' future liking of alters at T2 (β = 0.179; P = 0.009). **(C)** The extent to which ego uniquely liked alter at T1 was tracked by alter's idiosyncratic valuation responses to ego (β = 0.141; P = 0.029). **(D)** These same (i.e., alter-to-ego) patterns of neural valuation activity also predicted ego's future idiosyncratic liking of alters at T2 (β = 0.208; P = 0.0018).

Could this finding indicate that an ego's patterns of valuation system activation to various alters reflect his or her metaperceptions of idiosyncratic alter-to-ego liking (i.e., predictions made by egos of how much they are uniquely liked by the other group members)? This hypothesis was tested—but not supported—using self-reported metaperception data we collected at T1. In other words, an ego's idiosyncratic neural valuations of an alter did not track how much that ego expected that alter uniquely liked them ($\beta = 0.063$; $P = 0.334$). Although metaperceptions are typically conceptualized as an ego's *explicit* judgments of how much they are liked by alters, perhaps valuation system activity reflects an ego's *implicit* metaperceptions that are not consciously accessible to them (i.e., based on their explicit self-report measures). In support of this interpretation, an ego's idiosyncratic valuation response to an alter tracked how much that alter idiosyncratically liked him/her at T1, even controlling for the ego's explicit metaperceptions of how much the alter idiosyncratically liked them. In fact, when both were entered as simultaneous predictors of an alter's idiosyncratic liking of an ego, ego's unique valuation responses remained significant ($\beta = 0.135$; $P = 0.039$) while his/her explicit metaperceptions did not ($\beta = 0.098$; $P = 0.13$).

Discussion

We set out to investigate neural predictors of future liking, hypothesizing that the brain's valuation system might function as an implicit measure of interpersonal attraction that could be unobtrusively probed in order to predict how liking ties evolve within a social network. In the technical sense, our hypothesis was indeed supported by our findings: the extent to which an ego's valuation system was engaged in response to an alter at the beginning of the summer program did predict their future liking of that same alter after completing the program, even controlling for initial (explicitly reported) liking. However, although the neural index we advanced did in fact predict future liking, the specific patterns of results we obtained suggest the underlying neural mechanism is much more interpersonally sophisticated than we had anticipated.

One interpretation follows closely from our hypothesis: the valuation system activity captured in our paradigm does in fact reflect an implicit measure of ego's affiliative motivation toward alter. In this regard, valuation activity could be conceptualized as a sort of "seed" representing latent interpersonal attraction yet to become manifest in interpersonal behavior, or even possibly the consciousness of ego. In support of this interpretation, we find that unique ego-to-alter valuation activity is not only associated with ego-to-alter idiosyncratic liking at T2, but also with alter's T1 metaperception of how much they are uniquely liked by ego. This would mean that one's own liking of another at T1 is not consciously accessible to oneself (or at least self-reported) at that time but *is* consciously accessible to the other. What remains unclear by this account is why ego-to-alter unique valuation activity is associated with *alter-to-ego*—but not ego-to-alter—idiosyncratic liking at T1 and, to a lesser extent, why it is (descriptively) more predictive of T2 idiosyncratic liking from *alter-to-ego* than from ego-to-alter.

Another interpretation is that unique ego-to-alter valuation activity does not primarily reflect ego's future liking of alter, but is instead attuned to how much alter uniquely likes ego. In support of this interpretation, we find that ego-to-alter idiosyncratic valuation tracks alter-to-ego idiosyncratic liking at T1. Yet if this account were true then it would also necessarily entail a discrepancy between implicit and explicit valuation measures; specifically, by this account, idiosyncratic valuation activity (a) reflects an implicit measure of alter-to-ego idiosyncratic T1 liking but (b) does not correlate with an ego's self-reported metaperception of how much an alter uniquely likes them.

The first account presents idiosyncratic valuation activity as a window into one's "true" idiosyncratic interpersonal attraction to others. We only realize it ourselves later, but others see it for what it is from the start and it tracks how much they like us already. If this is the case, we observe a truly *interpersonal* neural mechanism in so far as the implicit social valuation signal it generates is realized by others before us, and moreover guides, reflects, and tracks others' idiosyncratic liking before our own. By contrast, the second account presents the valuation system not as representing one's own (yet unrealized) social valuation of others, but rather others' unique valuation of us, which we do not yet appreciate explicitly but which we later reciprocate in our idiosyncratic liking of them. This interpretation also therefore presents the valuation system as fundamentally interpersonal in the sense that it is sensitively attuned to *others'* unique liking of us, which is later incorporated into our own liking of others. By both accounts, group members' valuation systems *interdependently* constitute the neural mechanisms underlying the emergence of reciprocity in affective ties and, more generally, the evolution of social network structure.

Study 3:

Neural Measure of Enhanced Self-Valuation Tracks Trait Narcissism and Predicts Unpopularity

Introduction

Throughout the history of psychology, theorists and researchers alike have been drawn to exploring phenomena at the interface of our intrapersonal and interpersonal realms. Psychologists have documented countless ways in which people differentially—and often preferentially—perceive, evaluate, and respond to information concerning themselves versus others. Considered in sum, the psychological literature indicates that such enhanced self-perception biases are motivated and pervasive (Alicke & Sedikides, 2009, 2011; S. C. Jones, 1973; Leary, 2007), possibly even universal (Allport, 1937). Yet they also evidence considerable variability across persons, indicated by individual-difference measures of dispositional narcissism and self-enhancement bias (e.g., discrepancy between self-perception and other-perception or an objective criterion) (Alicke, 1985; Campbell, Brunell, & Finkel, 2006; Campbell, Reeder, Sedikides, & Elliot, 2000; Grijalva & Zhang, 2016; Kwan, John, Kenny, Bond, & Robins, 2004; Morf, Horvath, & Torchetti, 2011).

This multifaceted individual-difference construct—*narcissistic self-enhancement*—incorporates complementary intrapersonal and interpersonal strategies that mutually enhance self-perception: inflated valuation of self (and relative devaluation of others); attentional focus on oneself (and insensitivity to others); impulsive gratification of egoistic motivations (at the expense of others' concerns and long-term welfare) (Campbell, Rudich, & Sedikides, 2002; Morf et al., 2011; Morf & Rhodewalt, 2001; Roberts & Robins, 2000; Vazire & Funder, 2006). Social psychologists have documented that narcissistic tendencies of enhanced self-perception pose fundamental long-term obstacles to maintaining positive interpersonal evaluations and successful relationships (Campbell & Campbell, 2009; Morf & Rhodewalt, 2001; Vazire & Funder, 2006). Although individual differences in dispositional narcissism and related self-enhancement tendencies may initially engender affiliation and favorable impressions, they ultimately predict sociometric unpopularity (i.e., being disliked by peers) in longitudinal studies of face-to-face groups (Leckelt, Küfner, Nestler, & Back, 2015; Paulhus, 1998). Establishing mechanistic links from intrapersonal processes underlying enhanced self-perception to such long-term interpersonal consequences has remained a principal aim for researchers in this field.

The present longitudinal study integrates functional neuroimaging (fMRI) and social network analysis (SNA) in order to investigate the intrapersonal mechanisms underlying narcissistic self-enhancement, the resultant long-term obstacles to maintaining sociometric popularity, and help elucidate the intrapersonal-interpersonal gap in between. We theorized that fMRI could be used to unobtrusively probe valuation processes that undergird inflated self-valuation and relative devaluation of others while group members engage in a naturalistic social perception task that elicits evaluative representations of each individual (including oneself).

Specifically, we hypothesized that brain regions involved in social valuation—ventromedial prefrontal cortex (vmPFC), ventral striatum (VS), and amygdala—would be implicated in this narcissistic self-enhancement pattern of increased activation elicited by oneself and relatively decreased activated elicited in response to others. These densely interconnected regions (Haber & Knutson, 2009), are consistently implicated in processing the affective value and motivational significance of other people (Adolphs, 2003; Doré et al., 2014; Güroğlu et al., 2008; Haber & Knutson, 2009; Krienen et al., 2010; Zerubavel et al., 2015; Zink et al., 2008). If the psychological processes underlying narcissistic self-enhancement depend on the relative value and motivational significance of oneself relative to others, then this pattern would likely be expressed in vmPFC, VS, and/or amygdala. Based on this logic, the present study sought to identify a neural measure of individual differences in inflated self-valuation that would relate to individual differences in trait narcissism (to establish convergent validity) but not self-esteem (to establish discriminant validity). Further, this neural measure should be inversely related to sociometric popularity in established social networks (i.e., groups in which members have already undergone extended acquaintanceship), but not in social networks whose group members are just minimally acquainted. Because the detrimental social outcomes associated with narcissism typically manifest only with intimacy and increasing levels of acquaintance (Leckelt et al., 2015; Paulhus, 1998), the longitudinal sample provided an opportunity to test whether the neural index of narcissistic self-valuation would (1) be unrelated (or perhaps even positively related) to initial sociometric popularity, and (2) prospectively predict perceivers' eventual unpopularity, even controlling for their initial (i.e., concurrent) unpopularity.

Methods

Study 3 incorporates both the Study 1 and Study 2 datasets. The cross-sectional data from Study 1 is used to identify a plausible neural measure of narcissistic self-enhancement; then, the longitudinal data from Study 2 is incorporated to replicate these correlational findings and extend them to prospectively predict sociometric unpopularity. As such, the methods for Study 3 recapitulate those delineated in Study 1 and Study 2. Several additional methodological details are noted below. For all participants, individual differences in dispositional narcissism were measured by the 16-item Narcissistic Personality Inventory (NPI-16)(Ames, Rose, & Anderson, 2006). For the participants in Study 1, individual differences in self-esteem were measured by the Single-Item Self-Esteem Scale (SISE)(Robins, Hendin, & Trzesniewski, 2001). Finally, Study 3 additionally utilized parameter estimate (beta) maps corresponding to trials on which participants viewed their own face (which had been discarded prior to second-level analyses in Study 2 and Study 3).

Results

Our primary analyses investigated reward-related neural activity during the face-viewing task as a function of whose face was presented (self versus other) and individual-difference measures (at the perceiver level) associated with narcissistic self-valuation. First, we tested whether individual differences in perceivers' dispositional narcissism predicted disproportionately high neural activity in response to seeing oneself versus other group members. We conducted separate mixed-effects models for each valuation ROI (with random intercepts included to address cross-nesting of perceiver and target levels), each regressing ROI neural activity against (1) an indicator variable to indicate whether the face being viewed at the time was one's own versus that of another participant (group member), (2) dispositional trait narcissism score as assessed by the NPI-16 questionnaire measure, and (3) the interaction of the previous two variables. Across all three ROIs, we observed a main effect of stimulus type such that viewing one's own face elicited more neural activity than did others' faces (all Ps < 0.05). Perceivers' trait narcissism significantly amplified the strength of this self-face enhancement effect in vmPFC ($P < 0.01$) and VS ($P < 0.05$), while only a trend-level self*NPI interaction was found in the amygdala ROI ($P = 0.08$).

This same modeling procedure was likewise implemented to test whether elevated engagement of valuation ROIs while viewing one's own face (relative to others') was related to participants' *un*popularity (i.e., a negative association with sociometric popularity). As with the analysis of trait narcissism above, we conducted separate mixed-effects models for each valuation ROI in which neural activity was regressed against (1) indicator variable for stimulus type (self-face versus other-face), (2) sociometric popularity, and (3) these two variables' interaction. Once again, neural activity in both vmPFC and VS exhibited a main effect of self-face enhancement (Ps < 0.05) that interacted with perceivers' individual differences, in this case a negative association with sociometric popularity (Ps < 0.05). In the amygdala ROI, however, perceiver popularity did not significantly dampen the self-face enhancement effect ($P > 0.3$).

Considered together, the analyses above converged upon the same pattern of results: in vmPFC and VS ROIs, the extent to which one's own face (versus others') elicited heightened neural activation was significantly associated with participants' trait narcissism (positively) and sociometric popularity (negatively); by contrast, in models of amygdala ROI activity, neither of these interaction effects surpassed statistical significance thresholds. The amygdala ROI was therefore not included in subsequent analyses, while parameter estimates extracted from vmPFC and VS ROIs were averaged together to compute a unitary measure of neural valuation.

These findings suggested to us that disproportionately high engagement of valuation ROIs (i.e., vmPFC and VS, collectively) during self-perception relative to other-perception could provide a neural index of narcissistic self-valuation. Convergent validity for this interpretation of the neural measure was evidenced by its relationship with individual differences in perceivers' trait narcissism ($P < 0.005$; see Fig. 12 and sociometric (un)popularity ($P < 0.005$; see Fig. 13).

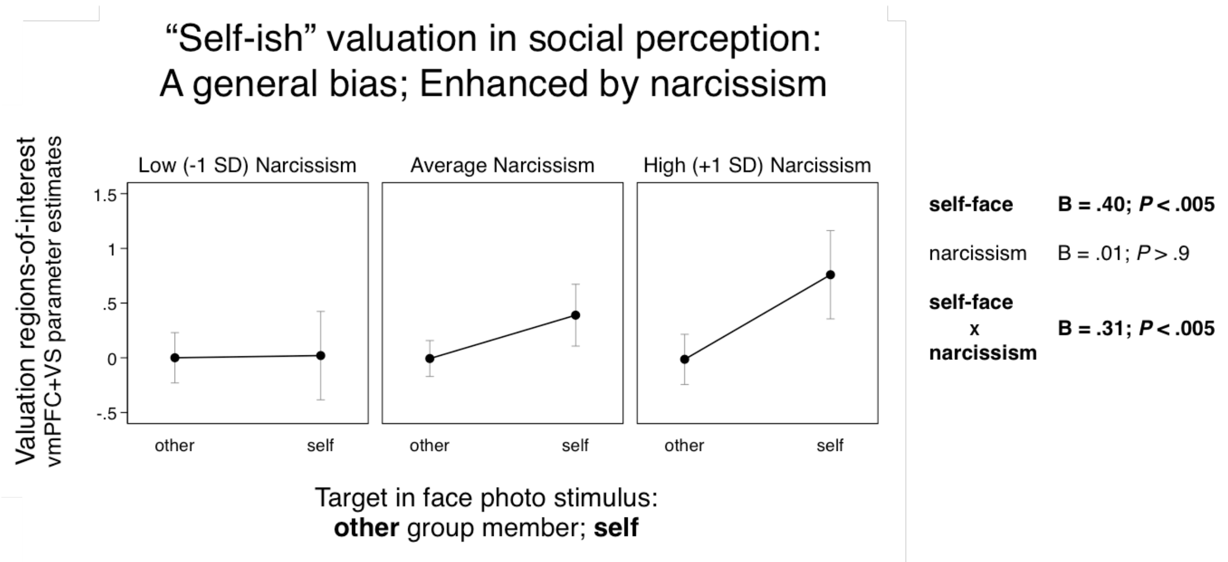

Fig. 12. Self-enhancement bias expressed in valuation system activity, a main effect of self-face which was amplified by individual differences in dispositional narcissism.

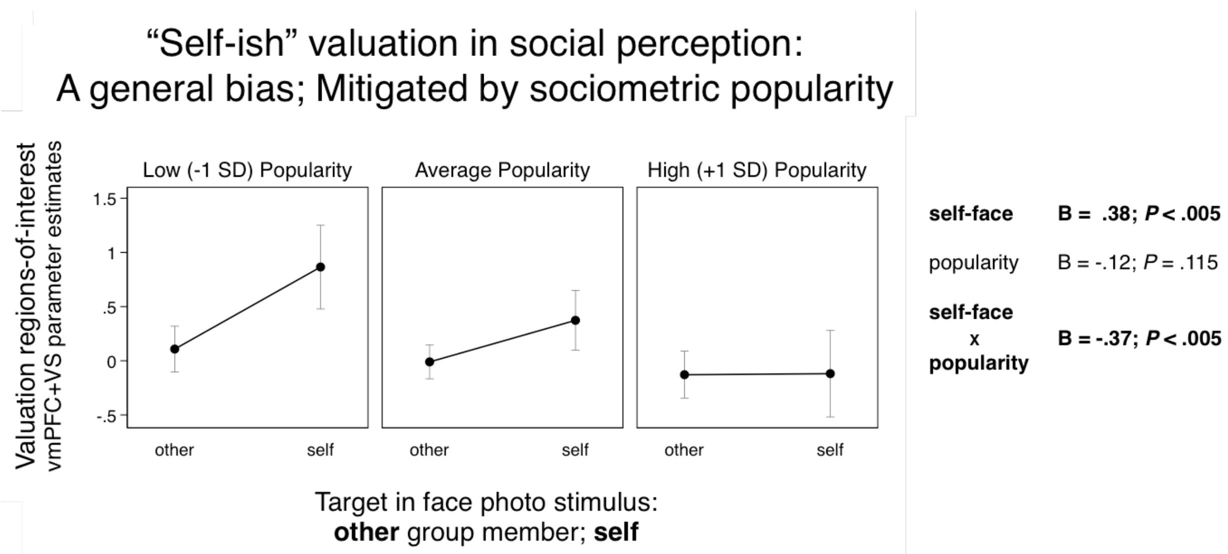

Fig. 13. Self-enhancement bias expressed in valuation system activity, a main effect of self-face which was mitigated by sociometric popularity.

In addition, we found of evidence of discriminant validity for this neural measure of narcissistic self-valuation in the result that it was unrelated to perceivers' self-report measure of self-esteem (*P* > 0.6; see Fig. 14).

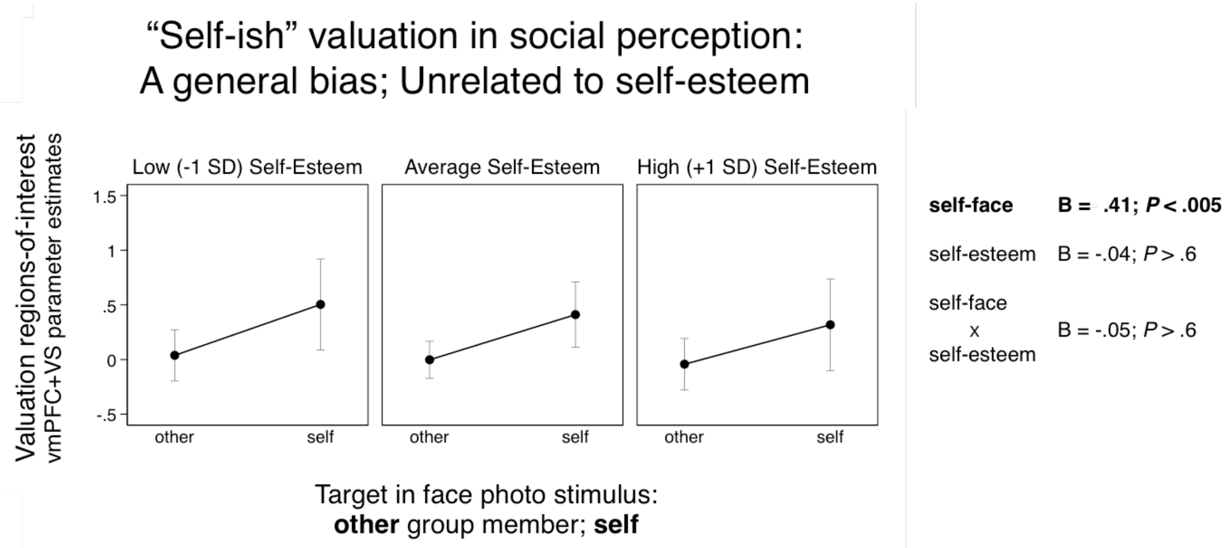

Fig. 14. Self-enhancement bias expressed in valuation system activity, a main effect of self-face which was not moderated by individual differences in self-esteem.

We then turned to a separate study sample in order to replicate our previous results and extend them in two important ways. This longitudinal study sample completed the fMRI face-viewing task and first round of sociometric data collection within less than a week of group members all meeting one another, which allowed us to ask:

(1) Do our previous findings of heightened valuation activity to one's own (relative to others') face generalize to social contexts in which the other viewed are those of newly acquainted group members?

(2) Does our neural index of narcissistic self-valuation prospectively predict future (un)popularity? If so, does this effect hold even controlling for initial popularity? Finally, does our implicit neural measure predict future (un)popularity better than does an explicit measure of narcissism?

Following the same mixed-effects modeling as before, we replicated the finding that valuation activity exhibited a main effect of self-face enhancement (*P*s < 0.05) that was amplified by perceivers' dispositional narcissism ($P < 0.05$; see Fig. 15).

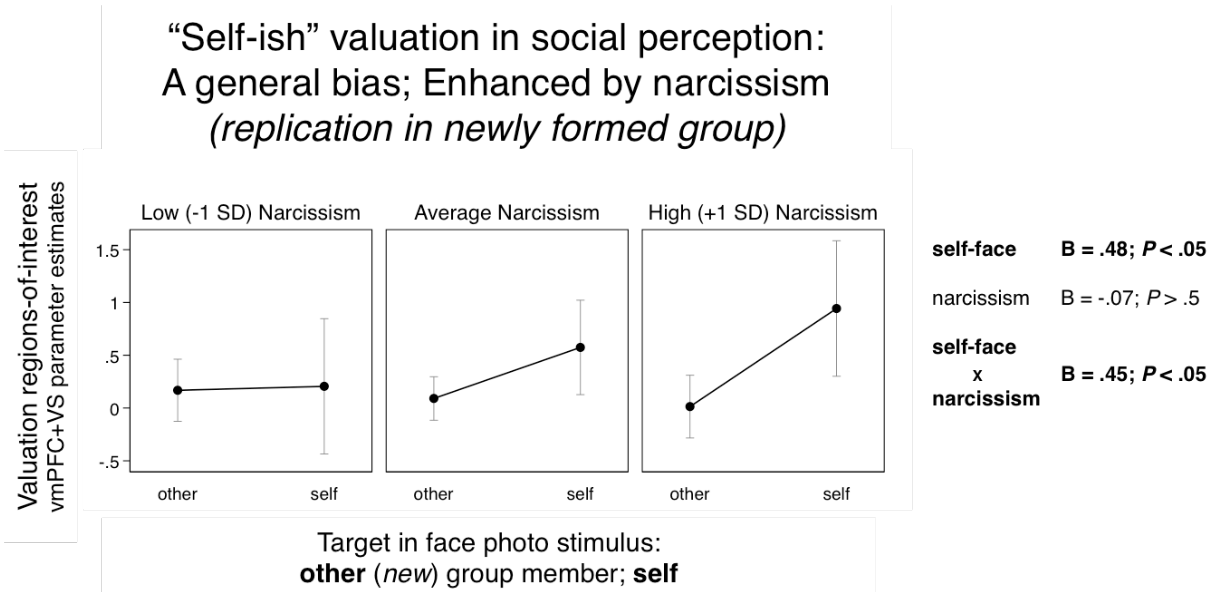

Fig. 15. Self-enhancement bias expressed in valuation system activity, a main effect of self-face which was amplified by individual differences in dispositional narcissism. This replicates the effects illustrated in Fig. 12 using a social context in which self-face was compared to faces of minimally acquainted—as opposed to well-acquainted—others.

In our earlier analysis of the relationship between initial sociometric (un)popularity and valuation activity, the latter was treated as the dependent variable. While no causal directionality of this relationship had been assumed, this modeling approach achieved greater statistical power by leveraging valuation activity as a dependent variable with multiple observations per participant.[4] In the present analysis, however, we sought to evaluate whether T2 (un)popularity could be *prospectively predicted* by a neural measure of narcissistic self-enhancement, and thus T1 brain data could only be appropriately characterized as a predictor variable. To this end, we computed an individual-difference variable reflecting narcissistic self-valuation using social relations analyses conducted in the *TripleR* package for R (Schönbrodt, Back, & Schmukle,

[4] This rationale likewise applies to the mixed-effect models in which dispositional narcissism was included as the individual-difference predictor variable. Whereas each subject-level variable (e.g., dispositional narcissism, sociometric popularity at T1 or T2) by definition provides only one data point per participant, the valuation activity data collected in this round-robin design constitute many observations per participant.

2012). Specifically, we utilized the self-enhancement index developed by Kwan and colleagues (Kwan, John, Kenny, Bond, & Robins, 2004) for round-robin designs, which simultaneously compares self-perceptions against perceptions *of* others and perceptions and perceptions *by* others. In other words, this social relations analysis conceptualizes self-enhancement bias as the extent to which one's self-about-self ratings exceed *both* self-about-other ratings and other-about-self ratings. Here we implemented this social relations modeling approach with interpersonal "ratings" consisting of neural activity in valuation ROIs; as with explicit trait ratings, these neural valuations were (implicitly) generated by each perceiver about each target (including, crucially, oneself). In the context of our data, Kwan and colleagues' (2004) self-enhancement index reflects the extent to which an individual perceiver disproportionately engaged valuation ROIs while viewing one's own face compared to *both* (a) self viewing others' faces as well as (b) others viewing self's face. This neural index of narcissistic self-valuation thus simultaneously accounts both for perceiver-level confounds (i.e., how much each perceiver *generally* engages valuation ROIs when viewing targets) and target-level confounds (i.e., how much valuation activity each target *generally* elicits from perceivers).

Using this neural measure, we found that self-enhancement bias in valuation activity at the beginning of the summer program (T1) was unrelated to perceivers' current sociometric popularity ($P > 0.9$; see Fig. 16a). However, it did predict their future (T2) unpopularity at the end of the summer program ($P < 0.05$; see Fig. 16b and Fig. 17c), even when controlling for initial popularity ($P < 0.05$). Moreover, T2 (un)popularity was better predicted by this implicit measure of narcissistic self-enhancement than by the explicit (self-report questionnaire) measure of narcissism. When both were entered as simultaneous predictors of T2 popularity, the neural index—but not the NPI-16—exhibited a statistically significant effect ($P < 0.05$ and $P > .7$, respectively). By itself, the neural measure of narcissistic self-valuation predicted almost 25% of the variance in T2 popularity ($R^2 = 0.241$).

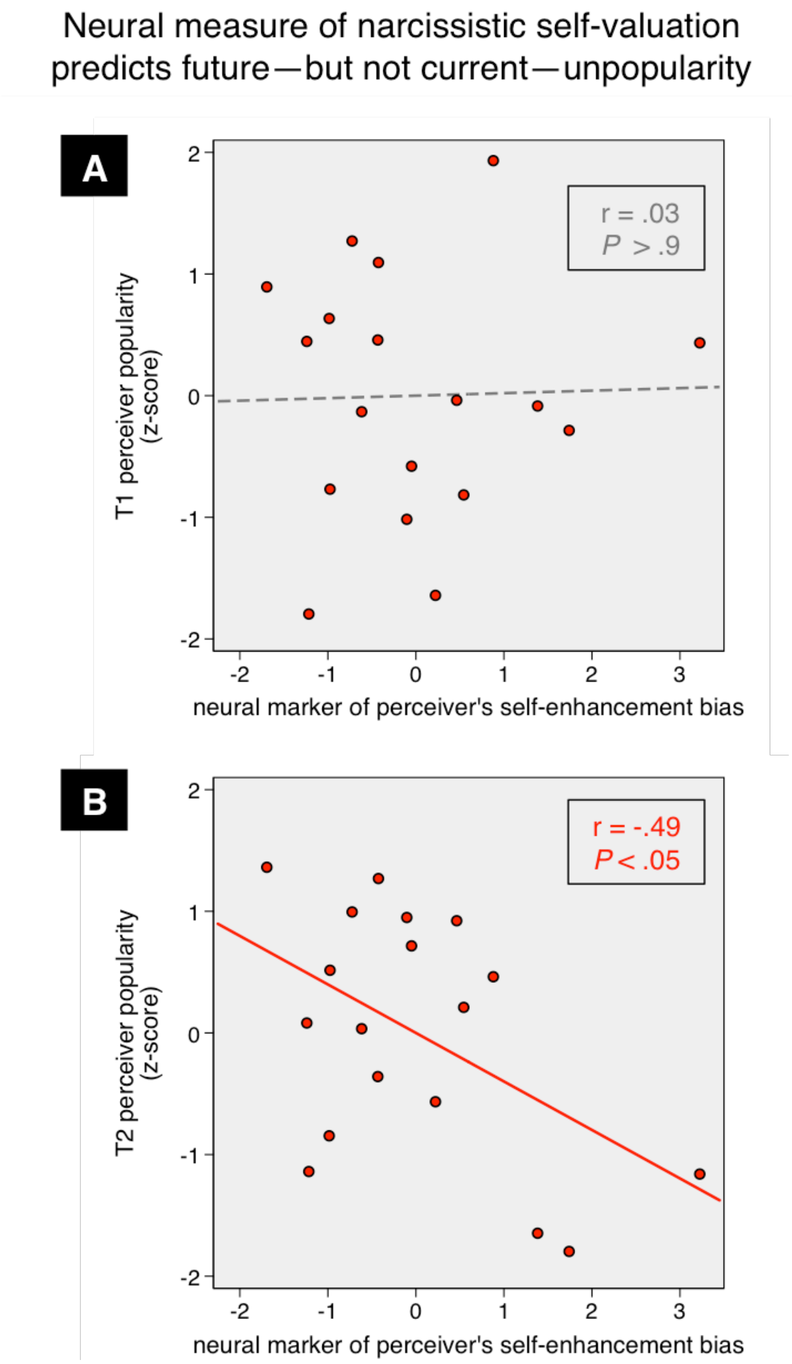

Fig. 16. Self-enhancement bias in valuation activity at the beginning of the summer program (T1) was **(A)** unrelated to perceivers' current sociometric popularity ($P > 0.9$), but **(B)** did predict their future (T2) unpopularity at the end of the summer program ($P < 0.05$; even when controlling for initial popularity).

Returning to the two social networks we had analyzed earlier, we now performed the social relations analysis of self-enhancement (Kwan, John, Kenny, Bond, & Robins, 2004) on their valuation activity data as well. This provided us an equivalent subject-level neural measure of narcissistic self-valuation for each participant across all three social networks. The relationship between this neural index and sociometric (un)popularity is illustrated in Fig. 17 separately for each network and also all three together. Aggregating the subject-level data across all three networks, valuation ROIs' self-enhancement bias was a consistent and robust predictor of perceiver (un)popularity, overall accounting for 24% of its variance ($P < 0.001$).

Discussion

The results of the present study demonstrate for the first time that fMRI can be used to pick up on neural markers of inflated self-valuation while participants merely view photos of their own and others' faces. This neural measure tracks individual differences in perceivers' dispositional narcissism (providing convergent validity) but not self-esteem (providing discriminant validity). In addition, it is negatively associated with concurrent sociometric popularity in social networks whose group members are well-acquainted. By contrast, in a social network whose group members are only minimally acquainted, this neural index of inflated self-valuation is unrelated to perceivers' concurrent (i.e., initial) popularity but prospectively predicts their ultimate unpopularity several months later. In fact, we found that T2 (un)popularity was better predicted by this implicit measure of narcissistic self-enhancement than by the explicit (NPI-16 questionnaire) measure of narcissism.

Social psychologists have long hypothesized and also found preliminary empirical support for the notion that narcissistic self-enhancement biases undermine healthy relationships by orienting disproportionately toward oneself rather than others. The results of the present study

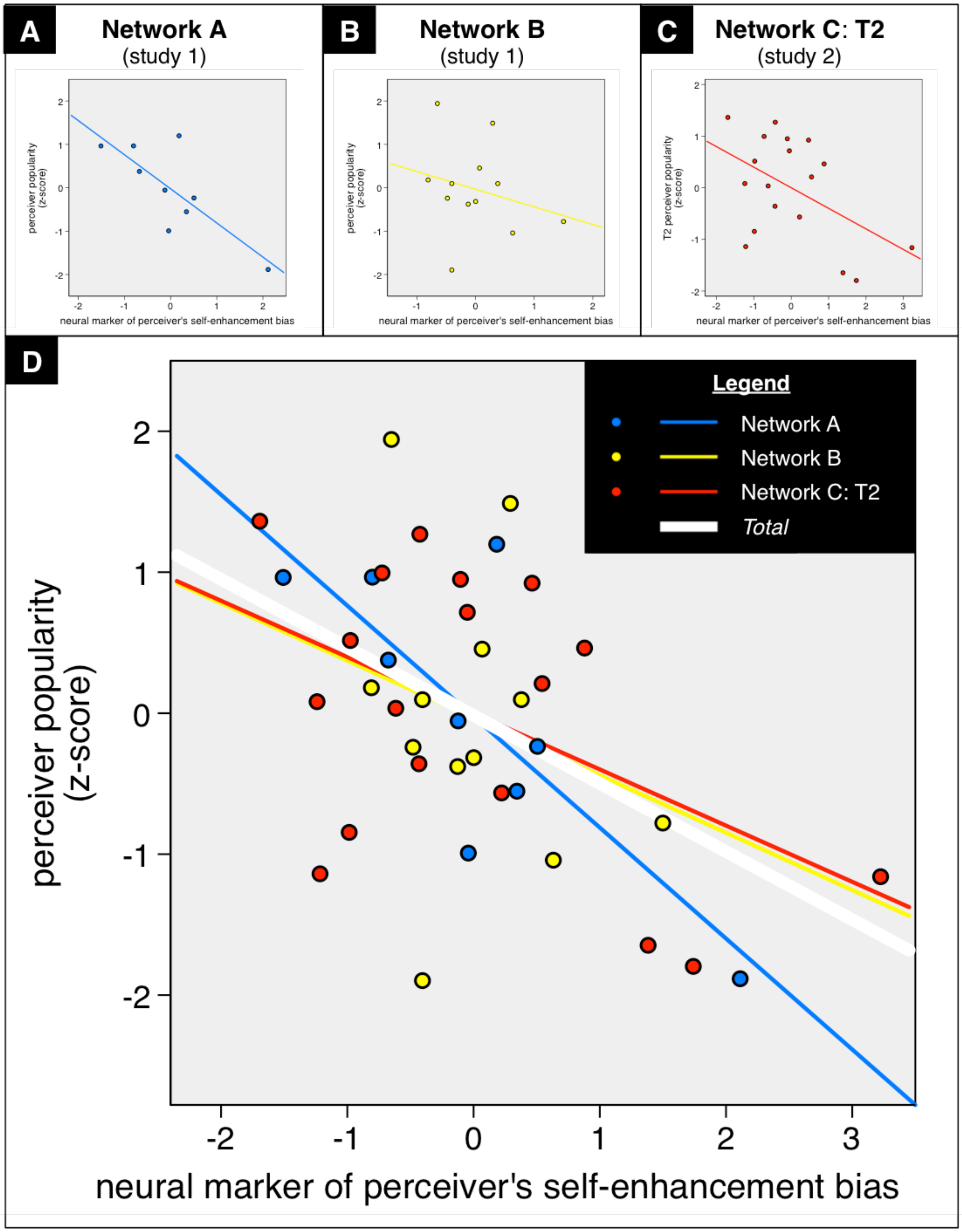

Fig. 17. Illustration of the negative relationship between the neural index of narcissistic self-valuation and sociometric popularity for each of the social networks separately **(A-C)** and also all three together **(D)**.

elucidate a plausible neural mechanism underlying this effect: heightened activation in brain regions underlying valuation—vmPFC and VS—engaged during self-perception (relative to other-perception). Given these brain regions' critical role in reward processing and reinforcement learning (Haber & Knutson, 2009), disproportionate engagement while orienting toward oneself (relative to others) may also provide intrinsically rewarding reinforcement that motivates such self-focused attention (rather than incentivizing proximity to others and social interactions with them). In other words, the neural mechanism of inflated self-valuation identified in this study may contribute to the self-reinforcing nature of narcissistic self-enhancement.

More generally, this study pioneers a novel approach to conducting neuroimaging research on social perception processes by integrating analytic techniques of Social Relations Modeling (SRM). In the present research, the SRM analytic approach was leveraged in order to generate implicit neural measures of narcissistic self-valuation, that is, enhanced activation of neural valuation regions when perceiving oneself relative to *both* perceiving others *and* being perceived by others. This approach could be extended to investigating the neural bases of other self-perception biases—not just enhanced self-valuation—during a naturalistic face-viewing task. The SRM conceptualization of self-perception recognizes that people function simultaneously as both perceivers and targets of their own social perception (Kwan, John, Kenny, Bond, & Robins, 2004). By integrating the SRM approach, neuroimaging research would benefit from a sociological insight first advanced by symbolic interactionists: self-perception is inextricably linked to interpersonal perception (Cooley, 1902; Goffman, 1978; Mead, 1934); by extension, neuroscientific and psychological understanding of the *intrapersonal* mechanisms underlying social phenomena would be greatly advanced by appreciating the *interpersonal* context in which they naturalistically occur.

References

Adolphs, R. (2003). Cognitive neuroscience of human social behaviour. *Nat Rev Neurosci, 4*(3), 165-178.

Alicke, M. D. (1985). Global self-evaluation as determined by the desirability and controllability of trait adjectives. *J Pers Soc Psychol, 49*(6), 1621.

Alicke, M. D., & Sedikides, C. (2009). Self-enhancement and self-protection: What they are and what they do. *European Review of Social Psychology, 20*(1), 1-48.

Alicke, M. D., & Sedikides, C. (2011). *Handbook of self-enhancement and self-protection*: Guilford Press.

Allport, G. W. (1937). *Personality*: Holt New York.

Ames, D. R., Rose, P., & Anderson, C. P. (2006). The NPI-16 as a short measure of narcissism. *Journal of Research in Personality, 40*(4), 440-450.

Azzi, J. C., Sirigu, A., & Duhamel, J.-R. (2012). Modulation of value representation by social context in the primate orbitofrontal cortex. *Proceedings of the National Academy of Sciences, 109*(6), 2126-2131.

Bearman, P. (1997). Generalized exchange. *American Journal of Sociology, 102*(5), 1383-1415.

Behrens, T. E., Hunt, L. T., Woolrich, M. W., & Rushworth, M. F. (2008). Associative learning of social value. *Nature, 456*(7219), 245-249.

Berkman, E. T., & Falk, E. B. (2013). Beyond brain mapping using neural measures to predict real-world outcomes. *Current Directions in Psychological Science, 22*(1), 45-50.

Bickart, K. C., Hollenbeck, M. C., Barrett, L. F., & Dickerson, B. C. (2012). Intrinsic amygdala–cortical functional connectivity predicts social network size in humans. *The Journal of Neuroscience, 32*(42), 14729-14741.

Bickart, K. C., Wright, C. I., Dautoff, R. J., Dickerson, B. C., & Barrett, L. F. (2010). Amygdala volume and social network size in humans. *Nat Neurosci, 14*(2), 163-164.

Bondonio, D. (1998). Predictors of accuracy in perceiving informal social networks. *Social Networks, 20*(4), 301-330.

Byrne, D. E. (1971). *The attraction paradigm* (Vol. 11): Academic Pr.

Campbell, W. K., Brunell, A. B., & Finkel, E. J. (2006). Narcissism, interpersonal self-regulation, and romantic relationships: An agency model approach. *Self and relationships: Connecting intrapersonal and interpersonal processes*, 57-83.

Campbell, W. K., & Campbell, S. M. (2009). On the self-regulatory dynamics created by the peculiar benefits and costs of narcissism: A contextual reinforcement model and examination of leadership. *Self and Identity, 8*(2-3), 214-232.

Campbell, W. K., Reeder, G. D., Sedikides, C., & Elliot, A. J. (2000). Narcissism and comparative self-enhancement strategies. *Journal of Research in Personality, 34*(3), 329-347.

Campbell, W. K., Rudich, E. A., & Sedikides, C. (2002). Narcissism, self-esteem, and the positivity of self-views: Two portraits of self-love. *Personality and Social Psychology Bulletin, 28*(3), 358-368.

Carlson, E. N., Furr, R. M., & Vazire, S. (2010). Do we know the first impressions we make? Evidence for idiographic meta-accuracy and calibration of first impressions. *Social Psychological and Personality Science, 1*(1), 94-98.

Casciaro, T. (1998). Seeing things clearly: Social structure, personality, and accuracy in social network perception. *Social Networks, 20*(4), 331-351.

Chen, A. C., Welsh, R. C., Liberzon, I., & Taylor, S. F. (2010). 'Do I like this person?' A network analysis of midline cortex during a social preference task. *Neuroimage, 51*(2), 930-939.

Chevallier, C., Kohls, G., Troiani, V., Brodkin, E. S., & Schultz, R. T. (2012). The social motivation theory of autism. *Trends Cogn Sci, 16*(4), 231-239

Chiao, J. Y. (2010). Neural basis of social status hierarchy across species. *Current opinion in neurobiology, 20*(6), 803-809.

Cooley, C. H. (1902). *Human nature and the social order*. New York: Scribner.

Davis, J. A. (1970). Clustering and hierarchy in interpersonal relations: Testing two graph theoretical models on 742 sociomatrices. *American Sociological Review, 35*, 843-851.

Deaner, R. O., Khera, A. V., & Platt, M. L. (2005). Monkeys pay per view: adaptive valuation of social images by rhesus macaques. *Current Biology, 15*(6), 543-548.

Dekker, D., Krackhardt, D., & Snijders, T. A. (2007). Sensitivity of MRQAP tests to collinearity and autocorrelation conditions. *Psychometrika, 72*(4), 563-581.

Demos, K. E., Heatherton, T. F., & Kelley, W. M. (2012). Individual differences in nucleus accumbens activity to food and sexual images predict weight gain and sexual behavior. *The Journal of Neuroscience, 32*(16), 5549-5552.

Denny, B. T., Kober, H., Wager, T. D., & Ochsner, K. N. (2012). A meta-analysis of functional neuroimaging studies of self-and other judgments reveals a spatial gradient for mentalizing in medial prefrontal cortex. *J Cogn Neurosci, 24*(8), 1742-1752.

Dépret, E., & Fiske, S. T. (1993). Social cognition and power: Some cognitive consequences of social structure as a source of control deprivation *Control motivation and social cognition* (pp. 176-202): Springer.

Doré, B. P., Zerubavel, N., & Ochsner, K. N. (2014). Social cognitive neuroscience: A review of core systems: APA Handbook of Personality and Social Psychology, Mikulincer, M., et al.,(Eds).

Dunbar, R. I. (2012). The social brain meets neuroimaging. *Trends Cogn Sci, 16*(2), 101-102.

Ekeh, P. P. (1968). Issues in exchange theory. *Berkeley Journal of Sociology*, 42-151.

Fareri, D. S., & Delgado, M. R. (2014). The Importance of Social Rewards and Social Networks in the Human Brain. *Neuroscientist*.

Farrow, T. F., Jones, S. C., Kaylor-Hughes, C. J., Wilkinson, I. D., Woodruff, P. W., Hunter, M. D., & Spence, S. A. (2011). Higher or lower? The functional anatomy of perceived allocentric social hierarchies. *Neuroimage, 57*(4), 1552-1560.

Fehr, E., & Camerer, C. F. (2007). Social neuroeconomics: the neural circuitry of social preferences. *Trends Cogn Sci, 11*(10), 419-427.

Fiske, S. T. (1993). Social cognition and social perception. *Annu Rev Psychol, 44*(1), 155-194.

Freeman, J. B., Rule, N. O., Adams Jr, R. B., & Ambady, N. (2009). Culture shapes a mesolimbic response to signals of dominance and subordination that associates with behavior. *Neuroimage, 47*(1), 353-359.

Freud, S. (1910). A special type of choice of object made by men (Contributions to the psychology of love I). *Standard Edition, 11*, 165-175.

Freud, S. (1912). The dynamics of transference. *Classics in Psychoanalytic Techniques*.

Galinsky, A. D., Magee, J. C., Inesi, M. E., & Gruenfeld, D. H. (2006). Power and perspectives not taken. *Psychol Sci, 17*(12), 1068-1074.

Gobbini, M. I., & Haxby, J. V. (2007). Neural systems for recognition of familiar faces. *Neuropsychologia, 45*(1), 32-41.

Goffman, E. (1978). *The presentation of self in everyday life*: Garden City, NJ: Doubleday.

Gouldner, A. W. (1960). The norm of reciprocity: A preliminary statement. *American Sociological Review*, 161-178.

Grijalva, E., & Zhang, L. (2016). Narcissism and Self-Insight A Review and Meta-Analysis of Narcissists' Self-Enhancement Tendencies. *Personality and Social Psychology Bulletin, 42*(1), 3-24.

Güroğlu, B., Haselager, G. J., van Lieshout, C. F., Takashima, A., Rijpkema, M., & Fernández, G. (2008). Why are friends special? Implementing a social interaction simulation task to probe the neural correlates of friendship. *Neuroimage, 39*(2), 903-910.

Haber, S. N., & Knutson, B. (2009). The reward circuit: linking primate anatomy and human imaging. *Neuropsychopharmacology, 35*(1), 4-26.

Hallinan, M. T. (1974). *The structure of positive sentiment* (Vol. 6): Elsevier Scientific Pub. Co.

Healey, K. L., Morgan, J., Musselman, S. C., Olino, T. M., & Forbes, E. E. (2014). Social anhedonia and medial prefrontal response to mutual liking in late adolescents. *Brain Cogn, 89*, 39-50.

Henrich, J., & Gil-White, F. J. (2001). The evolution of prestige: Freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. *Evolution and Human Behavior, 22*(3), 165-196.

Homans, G. C. (1958). Social behavior as exchange. *American Journal of Sociology*, 597-606.

Hommer, D. W., Knutson, B., Fong, G. W., Bennett, S., Adams, C. M., & Varner, J. L. (2003). Amygdalar recruitment during anticipation of monetary rewards. *Ann NY Acad Sci, 985*, 476-478.

Izuma, K., Saito, D. N., & Sadato, N. (2008). Processing of social and monetary rewards in the human striatum. *Neuron, 58*(2), 284-294.

Jones, E. E., & Gerard, H. (1967). Foundations of social psychology.

Jones, R. M., Somerville, L. H., Li, J., Ruberry, E. J., Libby, V., Glover, G., . . . Casey, B. (2011). Behavioral and neural properties of social reinforcement learning. *The Journal of Neuroscience, 31*(37), 13039-13045.

Jones, S. C. (1973). Self-and interpersonal evaluations: esteem theories versus consistency theories. *Psychol Bull, 79*(3), 185.

Kanai, R., Bahrami, B., Roylance, R., & Rees, G. (2012). Online social network size is reflected in human brain structure. *Proceedings of the Royal Society B: Biological Sciences, 279*(1732), 1327-1334.

Keltner, D., Gruenfeld, D. H., & Anderson, C. (2003). Power, approach, and inhibition. *Psychol Rev, 110*(2), 265.

Kenward, M. G., & Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics, 53*, 983-997.

Klein, J. T., & Platt, M. L. (2013). Social information signaling by neurons in primate striatum. *Curr Biol, 23*(8), 691-696.

Klein, J. T., Shepherd, S. V., & Platt, M. L. (2009). Social attention and the brain. *Curr Biol, 19*(20), R958-962.

Knutson, B., Westdorp, A., Kaiser, E., & Hommer, D. (2000). FMRI visualization of brain activity during a monetary incentive delay task. *Neuroimage, 12*(1), 20-27.

Krackhardt, D. (1988). Predicting with networks: Nonparametric multiple regression analysis of dyadic data. *Social Networks, 10*(4), 359-381.

Krantz, M., & Burton, C. (1986). The development of the social cognition of social status. *The Journal of genetic psychology, 147*(1), 89-95.

Krienen, F. M., Tu, P.-C., & Buckner, R. L. (2010). Clan mentality: evidence that the medial prefrontal cortex responds to close others. *The Journal of Neuroscience, 30*(41), 13906-13915.

Kwan, V. S., John, O. P., Kenny, D. A., Bond, M. H., & Robins, R. W. (2004). Reconceptualizing individual differences in self-enhancement bias: an interpersonal approach. *Psychol Rev, 111*(1), 94-110.

Lansu, T. A., Cillessen, A. H., & Karremans, J. C. (2013). Adolescents' Selective Visual Attention for High-Status Peers: The Role of Perceiver Status and Gender. *Child Dev, 85*(2), 421-428.

Leary, M. R. (2007). Motivational and emotional aspects of the self. *Annu. Rev. Psychol., 58*, 317-344.

Leckelt, M., Küfner, A. C., Nestler, S., & Back, M. D. (2015). Behavioral processes underlying the decline of narcissists' popularity over time.

Leifer, E. M. (1981). Competing models of political mobilization: The role of ethnic ties. *American Journal of Sociology*, 23-47.

Levy, I., Lazzaro, S. C., Rutledge, R. B., & Glimcher, P. W. (2011). Choice from non-choice: predicting consumer preferences from blood oxygenation level-dependent signals obtained during passive viewing. *The Journal of Neuroscience, 31*(1), 118-125.

Lewis, P. A., Rezaie, R., Brown, R., Roberts, N., & Dunbar, R. I. (2011). Ventromedial prefrontal volume predicts understanding of others and social network size. *Neuroimage, 57*(4), 1624-1629.

Lin, A., Adolphs, R., & Rangel, A. (2012). Social and monetary reward learning engage overlapping neural substrates. *Soc Cogn Affect Neurosci, 7*(3), 274-281.

Magee, J. C., & Galinsky, A. D. (2008). 8 Social Hierarchy: The Self-Reinforcing Nature of Power and Status. *The Academy of Management Annals, 2*(1), 351-398.

Marsh, A. A., Blair, K. S., Jones, M. M., Soliman, N., & Blair, R. J. (2009). Dominance and submission: the ventrolateral prefrontal cortex and responses to status cues. *J Cogn Neurosci, 21*(4), 713-724.

Mead, G. H. (1934). *Mind, self and society* (Vol. 111): Chicago University of Chicago Press.

Moreno, J. L. (1934). Who shall survive?: A new approach to the problem of human interrelations.

Morf, C. C., Horvath, S., & Torchetti, L. (2011). Narcissistic self-enhancement. *Handbook of self-enhancement and self-protection*, 399-424.

Morf, C. C., & Rhodewalt, F. (2001). Unraveling the paradoxes of narcissism: A dynamic self-regulatory processing model. *Psychological Inquiry, 12*(4), 177-196.

Neisser, U. (1976). *Cognition and reality: Principles and implications of cognitive psychology*: WH Freeman/Times Books/Henry Holt & Co.

Newcomb, T. M. (1960). Varieties of interpersonal attraction. *Group dynamics: Research and theory, 2*.

Newcomb, T. M. (1963). Stabilities underlying changes in interpersonal attraction. *The Journal of Abnormal and Social Psychology, 66*(4), 376.

Noonan, M. P., Sallet, J., Mars, R. B., Neubert, F. X., O'Reilly, J. X., Andersson, J. L., . . . Rushworth, M. F. (2014). A neural circuit covarying with social hierarchy in macaques. *PLoS biology, 12*(9), e1001940.

Ochsner, K. N., Beer, J. S., Robertson, E. R., Cooper, J. C., Gabrieli, J. D., Kihsltrom, J. F., & D'Esposito, M. (2005). The neural correlates of direct and reflected self-knowledge. *Neuroimage, 28*(4), 797-814.

Paulhus, D. L. (1998). Interpersonal and intrapsychic adaptiveness of trait self-enhancement: A mixed blessing? *J Pers Soc Psychol, 74*(5), 1197.

Powell, J., Lewis, P. A., Roberts, N., García-Fiñana, M., & Dunbar, R. (2012). Orbital prefrontal cortex volume predicts social network size: an imaging study of individual differences in humans. *Proceedings of the Royal Society B: Biological Sciences, 279*(1736), 2157-2162.

Powers, K. E., Chavez, R. S., & Heatherton, T. F. (2015). Individual differences in response of dorsomedial prefrontal cortex predict daily social behavior. *Soc Cogn Affect Neurosci*, nsv096.

Roberts, B. W., & Robins, R. W. (2000). Broad dispositions, broad aspirations: The intersection of personality traits and major life goals. *Personality and Social Psychology Bulletin, 26*(10), 1284-1296.

Robins, R. W., Hendin, H. M., & Trzesniewski, K. H. (2001). Measuring global self-esteem: Construct validation of a single-item measure and the Rosenberg Self-Esteem Scale. *Personality and Social Psychology Bulletin, 27*(2), 151-161.

Sallet, J., Mars, R. B., Noonan, M. P., Andersson, J. L., O'Reilly, J. X., Jbabdi, S., . . . Rushworth, M. F. (2011). Social network size affects neural circuits in macaques. *Science, 334*(6056), 697-700.

Schönbrodt, F. D., Back, M. D., & Schmukle, S. C. (2012). TripleR: An R package for social relations analyses based on round-robin designs. *Behavior Research Methods, 44*(2), 455-470.

Shepherd, S. V., Deaner, R. O., & Platt, M. L. (2006). Social status gates social attention in monkeys. *Curr Biol, 16*(4), R119-120.

Silk, J. B. (2007). Social components of fitness in primate groups. *Science, 317*(5843), 1347-1351.

Smith, K. P., & Christakis, N. A. (2008). Social Networks and Health. *Annual Review of Sociology, 34*(1), 405-429.

Snodgrass, S. E. (1985). Women's intuition: The effect of subordinate role on interpersonal sensitivity. *J Pers Soc Psychol, 49*(1), 146.

Snodgrass, S. E. (1992). Further effects of role versus gender on interpersonal sensitivity. *J Pers Soc Psychol, 62*(1), 154.

Solomon, R. (1942). The stability and some correlates of group-status in a summer-camp group of young boys. *The American Journal of Psychology*, 33-45.

Tamir, D. I., & Mitchell, J. P. (2012). Disclosing information about the self is intrinsically rewarding. *Proceedings of the National Academy of Sciences, 109*(21), 8038-8043.

Taylor, M. J., Arsalidou, M., Bayless, S. J., Morris, D., Evans, J. W., & Barbeau, E. J. (2009). Neural correlates of personally familiar faces: parents, partner and own faces. *Hum Brain Mapp, 30*(7), 2008-2020.

Tusche, A., Bode, S., & Haynes, J.-D. (2010). Neural responses to unattended products predict later consumer choices. *The Journal of Neuroscience, 30*(23), 8024-8031.

Vaughn, B. E., & Waters, E. (1981). Attention structure, sociometric status, and dominance: Interrelations, behavioral correlates, and relationships to social competence. *Developmental Psychology, 17*(3), 275.

Vazire, S., & Funder, D. C. (2006). Impulsivity and the self-defeating behavior of narcissists. *Personality and Social Psychology Review, 10*(2), 154-165.

Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic bulletin & review, 11*(1), 192-196.

Wang, F., Kessels, H. W., & Hu, H. (2014). The mouse that roared: neural mechanisms of social hierarchy. *Trends in neurosciences, 37*(11), 674-682.

Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications*: Cambridge University Press.

Watson, K. K., & Platt, M. L. (2012). Social signals in primate orbitofrontal cortex. *Curr Biol, 22*(23), 2268-2273.

Wiggins, J. S., & Trapnell, P. D. (1996). A Dyadic-lnteractional Perspective on the Five-Factor Model. *The five-factor model of personality: Theoretical perspectives*, 88.

Zaki, J., & Ochsner, K. (2009). The need for a cognitive neuroscience of naturalistic social cognition. *Ann N Y Acad Sci, 1167*(1), 16-30.

Zaki, J., & Ochsner, K. N. (2012). The neuroscience of empathy: progress, pitfalls and promise. *Nat Neurosci, 15*(5), 675-680.

Zaki, J., Schirmer, J., & Mitchell, J. P. (2011). Social influence modulates the neural computation of value. *Psychol Sci, 22*(7), 894-900.

Zerubavel, N., Bearman, P. S., Weber, J., & Ochsner, K. N. (2015). Neural mechanisms tracking popularity in real-world social networks. *Proceedings of the National Academy of Sciences, 112*(49), 15072-15077.

Zink, C. F., Tong, Y., Chen, Q., Bassett, D. S., Stein, J. L., & Meyer-Lindenberg, A. (2008). Know your place: neural processing of social hierarchy in humans. *Neuron, 58*(2), 273-283.