Neural precursors of future liking and affective reciprocity

Noam Zerubavel[a,b,1], Mark Anthony Hoffman[b,c], Adam Reich[b,c], Kevin N. Ochsner[d], and Peter Bearman[b,c,1]

[a]Center for Science and Society, Columbia University, New York, NY 10027; [b]Interdisciplinary Center for Innovative Theory and Empirics, Columbia University, New York, NY 10027; [c]Department of Sociology, Columbia University, New York, NY 10027; and [d]Department of Psychology, Columbia University, New York, NY 10027

Contributed by Peter Bearman, March 15, 2018 (sent for review February 7, 2018; reviewed by Lisa Feldman Barrett and James W. Moody)

Why do certain group members end up liking each other more than others? How does affective reciprocity arise in human groups? The prediction of interpersonal sentiment has been a long-standing pursuit in the social sciences. We combined fMRI and longitudinal social network data to test whether newly acquainted group members' reward-related neural responses to images of one another's faces predict their future interpersonal sentiment, even many months later. Specifically, we analyze associations between relationship-specific valuation activity and relationship-specific future liking. We found that one's own future (T2) liking of a particular group member is predicted jointly by actor's initial (T1) neural valuation of partner and by that partner's initial (T1) neural valuation of actor. These actor and partner effects exhibited equivalent predictive strength and were robust when statistically controlling for each other, both individuals' initial liking, and other potential drivers of liking. Behavioral findings indicated that liking was initially unreciprocated at T1 yet became strongly reciprocated by T2. The emergence of affective reciprocity was partly explained by the reciprocal pathways linking dyad members' T1 neural data both to their own and to each other's T2 liking outcomes. These findings elucidate interpersonal brain mechanisms that define how we ultimately end up liking particular interaction partners, how group members' initially idiosyncratic sentiments become reciprocated, and more broadly, how dyads evolve. This study advances a flexible framework for researching the neural foundations of interpersonal sentiments and social relations that—conceptually, methodologically, and statistically—emphasizes group members' neural interdependence.

liking | affective reciprocity | fMRI | dyadic | reward

In all known human groups, members of the group end up liking each other to varying degrees. This is partly due to group members' individual differences: some group members tend to like most everyone, and likewise some individuals are generally more liked relative to their peers. However, the vast majority of variation in liking is due to relationship effects, that is, group members having unique attractions to one another (1). Individuals' eventual liking ratings of particular group members are only modestly associated with their initial preferences, which evolve substantially over weeks of sustained interaction (2). Group members' unique liking sentiments develop interactively in the natural course of socializing, bonding, and forming relationships. In fact, they ultimately exhibit dyadic reciprocity—a fundamental feature of liking among interacting group members—which occurs when individuals we like also like us, or vice versa (1–6). This study leverages neuroimaging techniques to predict changes in group members' liking and elucidate how their initial idiosyncrasies develop into dyadic bonds of shared sentiment.

Social scientists have long sought to understand the interpersonal forces that attract group members to one another, generate dyadic ties of mutual affection, and shape how their social networks evolve over time. For decades, this line of research has been pursued primarily within a framework that emphasizes social-structural phenomena. One critical element of this research program has rested on the assumption that social relations tend toward affective reciprocity (1–6). However, this assumption, like the axiom that reciprocity is normative (4), prevents us from asking and answering deeply powerful questions, including the following: Why do we end up liking certain group members more than others, even if we initially did not? To what extent are such changes in specific attractions predictable in advance? How do mutually reciprocated affective ties arise (i.e., by what mechanism do dyads—the fundamental units of social relations—emerge from individuals)? Predicting group members' unique liking sentiments—and by extension, their emergent reciprocation—is our key focus.

Over the same period, psychologists have emphasized intrapersonal processes undergirding our liking preferences, affective ties, and social behaviors. Freud and others posited that individuals' attractions may be foreshadowed—or critically shaped—by intrapersonal processes of which they are not necessarily consciously aware (7). Building on reward-reinforcement research, social psychologists have theorized that interpersonal attraction is determined by the reward value individuals attribute to and elicit from one another (3, 8) and have proposed that affective reciprocity emerges from the mutual reinforcement of this reward value between interacting dyad members (2, 3). In this way, a social-structural phenomenon like affective reciprocity can be understood in regard to the intrapersonal processes through which it emerges.

This article extends and integrates both the intrapersonal and interpersonal lines of inquiry by testing a theoretically driven neural predictor of group members' future attractions. Specifically, we test a hypothesis involving interpersonal engagement of the brain's reward valuation system to identify the neural precursors

Significance

When joining a group, we may initially like some individuals more than others. Likewise, certain group members may be particularly drawn to us. Over months of interaction, these attractions inevitably change and typically become reciprocated. This study uses fMRI to predict such changes in liking. Specifically, we measure newly acquainted group members' reward system responses to images of one another's faces. We find that T1 neural responses predict whom one will like in the future. More strikingly, we find that others' T1 neural responses to us predict whom we will like months later, at T2. This brain-based mechanism helps explain how group members' initially unreciprocated liking sentiments become mutually reciprocated. This study reveals how our brains interdependently shape interpersonal relationships.

Author contributions: N.Z. and P.B. designed research; N.Z. performed research; N.Z., M.A.H., A.R., K.N.O., and P.B. analyzed data; and N.Z., M.A.H., A.R., and P.B. wrote the paper.

Reviewers: L.F.B., Northeastern University; and J.W.M., Duke University.

The authors declare no conflict of interest.

[1]To whom correspondence may be addressed. Email: nz2104@columbia.edu or psb17@columbia.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1802176115/-/DCSupplemental.

Published online April 9, 2018.

of relational liking and its reciprocation in human groups. To anticipate the main findings of this research, we show that (*i*) one's own (actor's) future liking of another group member (partner) at T2 can be intrapersonally predicted from our neural valuation of them measured months earlier at T1; (*ii*) our T2 liking of a partner can be interpersonally predicted by that partner's T1 reward system response to us; (*iii*) these actor and partner forecasting effects are robust when statistically controlling for each other, both individuals' initial attractions, and other potential predictors of affiliation; (*iv*) these reciprocal predictive effects also help explain how actor's and partner's liking of one another—which are initially unrelated—become reciprocated. These results offer insight into the neural precursors of interpersonal attraction and its emergent reciprocation, that is, the fundamental ingredients of human sociality from pair-bonding to group cohesion. More broadly, this study advances a paradigm for researching the links between the interpersonal and intrapersonal mechanisms undergirding the formation of social preferences, social ties, and their consequent social network structure.

Reward Value as a Precursor of Interpersonal Attraction

Psychologists theorize that interpersonal reward is an antecedent of liking and reciprocity (2, 3, 8). Their rationale is based on principles of positive reinforcement extended to social relations. Consider two hypothetical group members A (Anita) and B (Buddy): if Anita experiences reward during social encounters with Buddy, it will motivate Anita to affiliate with Buddy. Because Anita's interactions with Buddy are inextricably linked to Buddy's interactions with Anita, a positive-feedback loop plays out at the dyadic level, unfolding interdependently between Anita and Buddy. As Newcomb suggests, "[I]nsofar as both partners are rewarded, another evening of duets or another set of tennis is likely to ensue, together with still further opportunities for reciprocal reward" (3). Based on this reward-reinforcement theoretical framework, we hypothesized that group members' reward-related responses to one another could effectively forecast how their interpersonal attractions develop.

Functional magnetic resonance imaging (fMRI) can be used to unobtrusively probe and quantify neural activity thought to be associated with particular psychological processes, such as reward. Hundreds of fMRI studies have consistently implicated ventromedial prefrontal cortex (vmPFC) and ventral striatum (VS) in individuals' anticipation and processing of rewards as well as their subjective valuation of both social and nonsocial stimuli (9–15). Critically, these densely interconnected regions of the brain's core reward system (13) encode socially valuable rewards such as being liked and anticipating positive social feedback (14–20). Recent brain-as-predictor studies have utilized fMRI measures of both brain regions' activity as neural markers of reward value to predict participants' attitudes and behavior outside of the scanner (see ref. 21 for review). For example, participants' idiosyncratic preference judgments and choice behaviors across various consumer products were predicted by brain activity in vmPFC and VS, which had been measured earlier using fMRI while participants simply viewed the relevant consumer products (22, 23).

The psychology literature posits reward as an antecedent of interpersonal attraction, and the neuroscience literature suggests this interpersonal reward value can be implicitly measured and operationalized by neural activity in targeted regions of interest (ROIs) underlying reward valuation, namely, vmPFC and VS. We integrate these psychological theories and neuroimaging methods to test whether interpersonal engagement of neural reward systems (i.e., ROI activations elicited by briefly viewing each peer's face) can prospectively predict how group members' interpersonal attractions develop over time and become reciprocated. Because this study focuses on relational phenomena in which participants are inherently interdependent—liking, being liked, and forming mutually reciprocated bonds—we embed the brain-as-predictor approach within a dyadic framework such that one's outcomes can be predicted by one's own and others' neural data.

Analytical Approach

As our paradigm emphasizes group members' relational interdependence, it differs from previous brain-as-predictor studies in several respects. First, our study population consists of an interacting group whose members formed organic relationships. Second, the fMRI task models social encounters between group members by presenting a given participant's face to every other participant being scanned. Thus, each presentation corresponds to a particular relationship and the resulting neural data are inherently dyadic, inextricably linked to the person being scanned and the person whose face is presented. Third, the primary outcome we seek to predict—group members' unique attractions after months of interaction—consists of interdependent observations. If Anita's liking of Buddy and Buddy's liking of Anita are related at T2, this dyadic linkage of individuals' outcomes must be addressed for both conceptual and statistical reasons (24). In this vein, our sociological phenomenon of interest—mutually reciprocated attraction—is not a characteristic of individuals, but rather of dyadic relations.

The reciprocation of T2 liking suggests that participants' attractions are interdependent and at least partly shaped by interpersonal processes. In practical terms, this means that when we seek to predict an actor's outcome (i.e., unique attraction to partner at T2), we consider as potential predictors both actor inputs and partner inputs. We thus sociologically extend the brain-as-predictor approach—which conventionally uses our own neural responses to predict our own behavior—to consider how our own behavior could be reciprocally predicted by others' neural responses to us. We simultaneously assessed these predictive pathways using structural equation modeling (SEM) to implement the actor–partner interdependence model (APIM) (24, 25). The APIM conceptualizes relational dyads—each consisting of two interdependent individuals—as the fundamental units of analysis. It therefore allows for the possibility of correlated outcomes between these two individuals (for every possible pairing of participants). For each predictor variable, APIM simultaneously estimates both its intrapersonal and interpersonal predictive effects on the outcome variable. This discussion is captured in Fig. 1. Fig. 1*A* depicts the intrapersonal brain-as-predictor approach, in which actor's reward system activity in response to viewing partner's face at T1 predicts actor's future liking of partner at T2. Fig. 1*B* shifts our focus from intrapersonal processes to interpersonal dynamics. Specifically, we consider how others' neural responses to us predict our own future liking of them, and because the system is symmetrical, how our neural responses to them predict their future liking of us.

The analysis of liking is complicated by the fact that individuals differ with respect to how much they generally like others in their group, and similarly, how much other group members generally like them. To analyze the evolution of liking and emergence of reciprocity as dyadic relationship processes, however, we need to capture the uniquely relational component of liking that is specific to each relationship—rather than to each individual—in the group (1, 24). In other words, we want to isolate how much Anita uniquely likes Buddy (i.e., taking into account Anita's overall tendency to explicitly like others and Buddy's overall tendency to be the recipient of others' explicit liking). Likewise, our relationship-specific fMRI parameter needs to capture Anita's unique neural valuation of Buddy (i.e., taking into account how Anita's reward system generally responds to others and Buddy's overall tendency to elicit reward responses). Due to the round-robin design of our fMRI task and liking assessments, the respective measures of neural reward and explicit liking can be partitioned into relationship-specific and person-level components using TripleR (1, 26). We then incorporate these variables into models that expressly analyze associations between relationship-specific neural activity and relationship-specific future liking (*SI Text*).

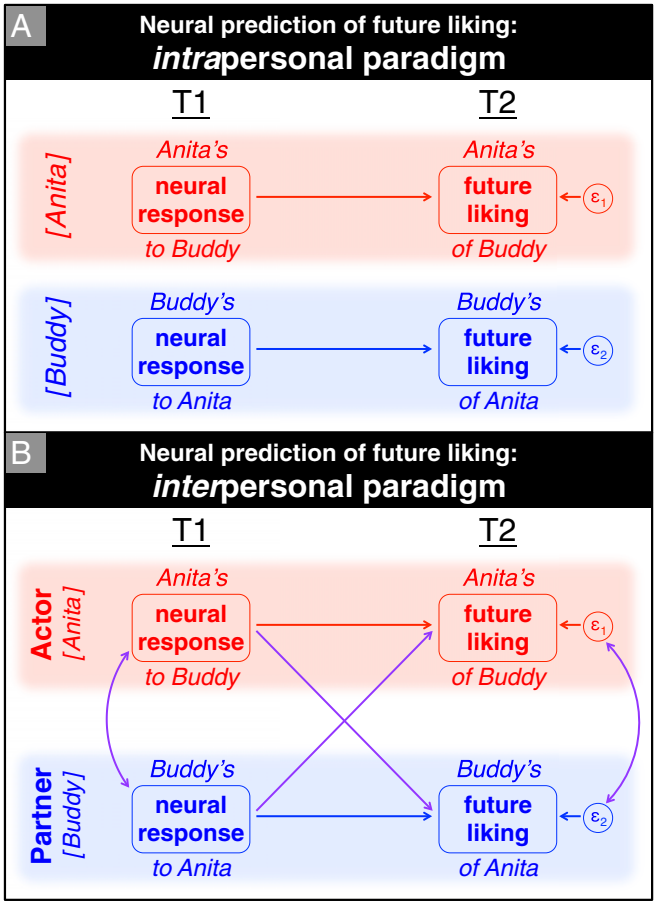

Fig. 1. Comparison of two conceptual paradigms for predicting future (T2) explicit liking based on initial (T1) explicit liking and implicit neural measure of reward value. (*A*) The intrapersonal model conceptualizes actor's outcome (i.e., T2 liking of a given partner) as predicted solely by actor's inputs at T1 (i.e., actor's implicit neural valuation of partner). More generally, individuals' T2 liking sentiments are assumed to be independent of one another. (*B*) By contrast, the interpersonal model allows for and quantifies such interdependence of outcomes between the two group members—actor and partner—comprising each dyad. This paradigm treats each predictor variable as potentially capable of exhibiting both intrapersonal and interpersonal effects (i.e., predicting actor's and partner's T2 liking, respectively). By the same token, this dyadic model necessarily implies that T2 liking can have both intrapersonal and interpersonal predictors. One-sided arrows depict directional paths from T1 predictor to T2 outcome: Horizontal paths colored blue/red are intrapersonal actor effects, and diagonal paths colored purple are interpersonal partner effects. Curved paths with two arrowheads depict symmetric correlations without specified directionality (including dyad members' correlated error terms, ε_1 and ε_2).

The study aimed to prospectively predict T2 relational liking, that is, specific actors' unique attractions to particular partners. The study population consisted of an interacting group of 16 college-age students involved in an intense summer of labor organizing (*Methods* and *SI Text*). Over the course of the 9 wk, participants spent time in smaller teams as well as in the larger collective. At the beginning of the program (T1), they viewed faces of every other social network member while fMRI data were collected (of specific interest, activation of reward system ROIs in vmPFC and VS, which were independently defined; *Methods*). Group members were thus implicated as both the sources and targets of interpersonal neural valuations. We analyzed reward system activity (average of vmPFC and VS ROIs) to predict participants' unique liking of each other at T2, controlling for T1 liking and social-structural factors that have been shown to be crucial drivers of tie formation (e.g., homophily).

Results

Actor Effects: Intrapersonal Predictors of T2 Liking. Focusing on the intrapersonal precursors of future liking, we asked, Do group members' initial reward responses to one another predict their future attractions? In other words, does Anita's reward system activity while viewing Buddy's picture at T1 predict how much Anita will ultimately like Buddy at T2? In support of our primary hypothesis, the APIM analysis demonstrated that actors' unique neural valuations of partners at T1 predicted their unique liking sentiments at T2 (Fig. 2*B* and *SI Text*; $\beta = 0.16$; $P < 0.01$).

To ensure that the association between neural activity and T2 liking was not merely due to both variables' association with T1 liking, our model includes a baseline control measure of initial liking at T1 (Fig. 2*A*). The path from T1 liking to T2 liking ($\beta = 0.23$; $P < 0.005$) measures the extent of affective stability over the course of the program. Critically, actor's T1 neural valuation of partner remained a significant predictor of future liking even after controlling for their T1 liking. The actor effect we observe is not merely an artifact driven by initial liking. On the contrary, explicit (self-reported liking) and implicit (neural marker of valuation) T1 measures were found to be distinct predictors of future liking. Considered together, these actor effects indicate that Anita's neural valuation of Buddy at T1 predicted how much she would ultimately like Buddy at T2, even taking into account her initial liking of him. This neural predictor thus explains change in interpersonal attraction above and beyond its temporal stability.

Partner Effects: Interpersonal Predictors of T2 Liking. In addition to actor effects of neural valuation and initial liking at T1, this model simultaneously assessed the respective partner effects of both variables. These partner effects correspond to predicting Anita's outcome (liking Buddy at T2) using Buddy's—rather than her own—T1 liking and fMRI data. As shown in Fig. 2*A*, the partner effect of T1 liking was positive and significant ($\beta = 0.19$; $P < 0.005$), meaning that Anita's ultimate liking of Buddy was predicted by Buddy's initial liking of Anita. This is consistent with the intuition that we often come to like people who like us. Actor's T2 liking of partner was likewise predicted by partner's T1 neural reward response to that actor ($\beta = 0.21$; $P < 0.005$), even controlling for both of their T1 liking ratings ($\beta = 0.17$; $P < 0.005$). In relation to our primary hypothesis (i.e., Anita's neural valuation of Buddy at T1 predicts how much she will ultimately like him at T2), this finding presents evidence in support of the reciprocal phenomenon (i.e., Buddy's neural valuation of Anita predicts how much she will later like him). This partner effect can be conceptualized in terms of our initial neural responses to each group member foreshadowing their unique sentiments toward us at T2. Equivalently, how much we ultimately like certain individuals at T2 can be predicted by their unique neural valuations of us at T1 (i.e., the neural reward responses we specifically activate in each of them).

Additional Analyses and Robustness Checks. Actor and partner effects of neural valuation are robust to each other, as well as to actor and partner effects of T1 liking. However, it is possible that other mechanisms are at play, specifically, sociological predictors of tie formation. We tested these alternative explanations. Each APIM iteration included a sociological predictor (e.g., homophily on demographic or personality attributes) in addition to those described above (actor's and partner's T1 liking ratings and neural responses). None of these sociological variables significantly predicted T2 liking outcomes (all values of $P > 0.2$; see *SI Text* for additional details). We also conducted APIM analyses using raw liking ratings and neural parameters instead of their uniquely relational components. Controlling for both dyad members' baseline liking ratings at T1, neural valuation evidenced both actor ($\beta = 0.13$; $P < 0.01$) and partner effects ($\beta = 0.14$; $P < 0.05$).

Correlations Among Predictor Variables. Although our analyses were primarily intended for modeling T2 liking against various T1 predictors, we also assessed interrelations among these predictor

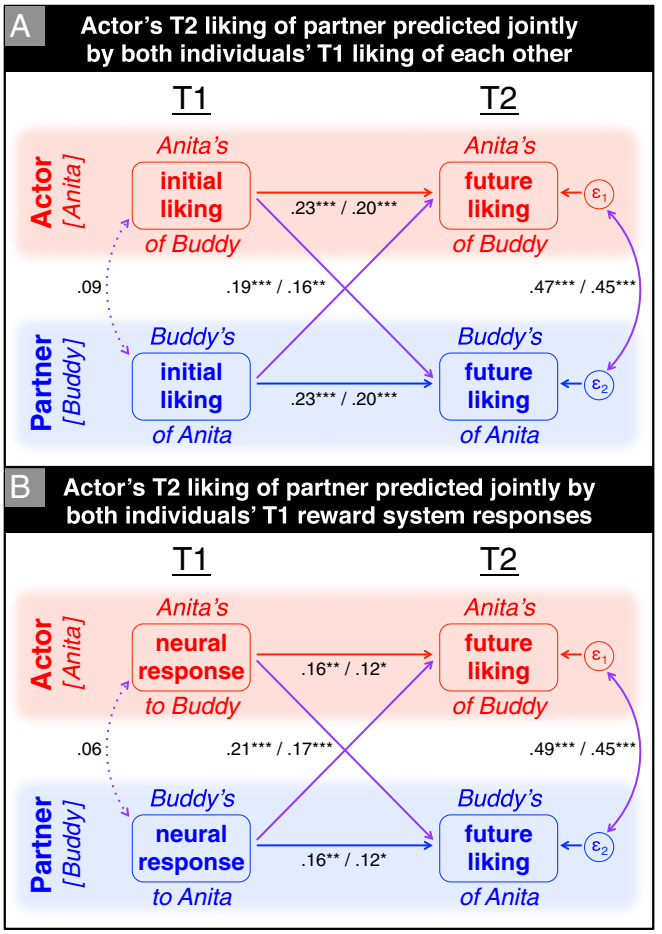

Fig. 2. Results of APIM analysis reveal how two group members' T1 predictors—initial liking and neural valuations of each other—independently and interdependently predict both individuals' future liking of each other. (*A*) Actor's T2 liking of partner was jointly predicted by its own T1 baseline measure (i.e., initial actor-to-partner liking; $\beta = 0.23$; $P < 0.005$) as well as partner's T1 liking of actor ($\beta = 0.19$; $P < 0.005$). These two predictor variables were not significantly correlated ($\beta = 0.09$; $P > 0.4$), indicating that relationship-specific liking was initially unreciprocated. (*B*) Actor's T2 liking of partner was mutually predicted by actor's neural reward response to partner ($\beta = 0.16$; $P < 0.01$) and partner's neural response to actor ($\beta = 0.21$; $P < 0.005$), even controlling for both members' initial liking ratings (actor effect: $\beta = 0.12$; $P < 0.05$; and partner effect: $\beta = 0.17$; $P < 0.005$). Note that, for visual clarity, the two neural predictors in *B* are illustrated in a separate panel from the two initial liking covariates in *A*; however, our analyses did in fact simultaneously model all four predictors' effects. Path coefficients preceding the slash correspond to models with (*A*) only initial liking predictors or (*B*) only neural predictors, and those following the slash to the combined model with both sets of predictors. The shape and color of paths follow Fig. 1. Paths are illustrated as solid lines if the effect is significant at $P < 0.05$; otherwise, they are illustrated as dotted lines. *$P < 0.05$; **$P < 0.01$; ***$P < 0.005$.

variables. Two of these correlational pathways are depicted as curved, double-sided arrows in Fig. 2: actor's and partner's unique liking of each other at T1 were not significantly correlated ($\beta = 0.09$; $P > 0.4$), nor were their T1 neural valuations of each other ($\beta = 0.06$; $P > 0.5$). In addition, actor's neural reward response to partner did not track with actor's own T1 liking of partner ($\beta = 0.08$; $P > 0.3$); however, it did correlate with that partner's T1 liking of actor ($\beta = 0.17$; $P < 0.05$). Critically, we ensured that our brain-as-predictor findings were robust to any such correlation by incorporating all four T1 predictors (i.e., actor's and partner's T1 liking ratings and neural responses) in the APIM analysis described above (Fig. 2 and *SI Text*).

The Emergence of Reciprocity. Between T1 and T2, participants' relational liking sentiments nearly doubled in variance and became predominantly dyadic as opposed to idiosyncratic (Fig. 3*A*). This pattern of results indicates that actors' and partners' unique attractions to each other became statistically coupled and spread out from the distribution's mean toward its tails. In other words, liking variance increased as dyads became differentiated from one another on the basis of dyad members' mutual liking. As depicted in Fig. 3 *B* and *C*, these data also reveal how the mutual reciprocation of relational liking dramatically increased—in fact, came into existence—over the course of the summer program: at T1, dyad members' unique attractions were uncorrelated with virtually zero (0.09) shared variance ($P > 0.4$), compared with 0.52 shared variance by T2 ($P < 0.005$). Such robust dyadic interdependence of T2 liking also empirically validates our rationale for modeling this outcome measure within the dyadic APIM framework (24, 25). Moreover, the APIM analysis estimates that 27% of the T2 liking reciprocity correlation is explained by the interpersonal brain-as-predictor model depicted in Fig. 2 (i.e., with T1 liking and neural predictors); crucially, even in its rudimentary form (i.e., based solely on dyad members' neural valuations of each other but not their initial liking ratings), a brain-only predictive model explains 14% of T2 liking reciprocation.

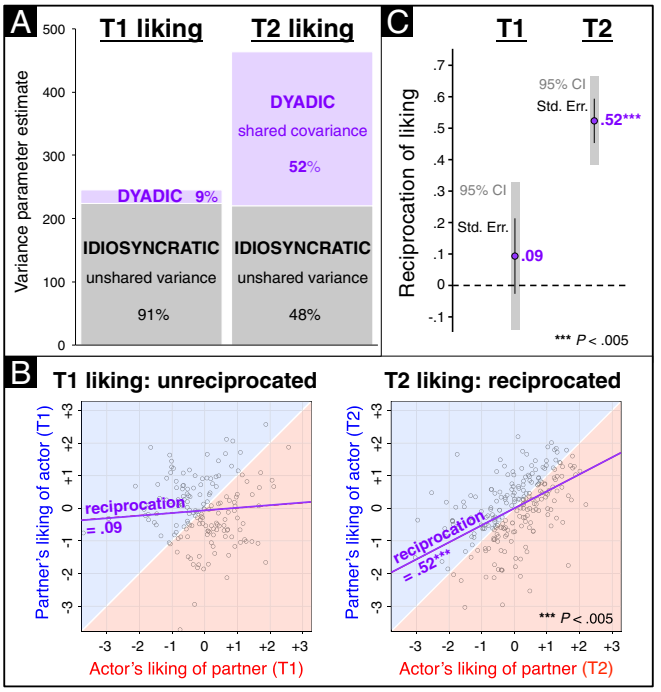

Fig. 3. Liking became a dyadic phenomenon over the course of the summer program, as individuals' idiosyncratic liking of particular group members at T1 developed into mutually reciprocated liking at T2. (*A*) Participants' relational sentiments nearly doubled in variance from T1 to T2, an effect driven by the rise of dyadic (shaded purple)—but not idiosyncratic (shaded gray)—liking variance components. This indicates that dyads became differentiated on the basis of mutual liking, that is, liking ratings underwent coupling and spreading (in pairs) away from the mean. In proportional terms, initial liking was almost solely idiosyncratic, whereas dyad members shared approximately one-half of T2 liking variance. (*B*) At T1, there was no association between actor's liking of partner and partner's liking of actor ($\beta = 0.09$; $P > 0.4$). By T2, this correlation had become strongly positive ($\beta = 0.52$; $P < 0.005$). (*C*) Mutually reciprocated attraction, which was not evident at T1, emerged prominently by T2. This reciprocity can be equivalently formulated as either (*A*) the proportion of liking variance due to dyads or (*B*) the correlation between actor-to-partner liking and partner-to-actor liking.

Discussion

We set out to test a hypothesized neural precursor of liking and reciprocity in human groups. When this group was just forming (T1), there was virtually no association between Anita particularly liking Buddy and Buddy particularly liking Anita. However, after months of interaction (T2), dyadic reciprocity had mushroomed from absence to predominance, explaining approximately one-half of the variance in group members' relational sentiments. This empirical reality precluded analyzing individuals' outcomes as if they were independent of each other (24, 25), the norm in fMRI and psychological research; in other words, we could not disregard the fact that Anita's unique liking of Buddy had become linked to Buddy's unique liking of Anita. Rather, by embedding these interdependent outcomes within their dyadic context, our analyses predicting individuals' future liking sentiments also modeled the emergence of reciprocity.

Future Liking Predicted by One's Own and Partner's Neural Reward Responses. A consequence of adopting APIM's dyadic framework is that it enabled us to model an actor's future liking of a particular partner as potentially predicted by both the actor's and the partner's neural responses to each other. Thus, while our analyses allowed us to test the original hypothesis—that relational engagement of the brain's reward system could serve as a neural predictor of interpersonal attraction—embedding this predictive pathway within a dyadic context greatly expanded the range of processes that could be considered as precursors of liking ties and their reciprocation. In a narrow sense, the straightforward brain-as-predictor hypothesis was indeed supported by the study findings: An actor's unique neural valuation of their partners at the beginning of the summer program predicted how much that actor would uniquely like each of them after completing the program, even controlling for initial liking.

In a broader sense, however, this individualistic actor-oriented paradigm could not anticipate the possibility of partner effects and therefore precluded their predictive potential. By incorporating dyadic APIM analyses, we could model neural reward responses as both intrapersonal and interpersonal predictors of future liking. This led us to discover that an actor's outcome (i.e., unique liking of partner at T2) could be prospectively predicted from their partner's neural reward activity when initially viewing an image of the actor's face; moreover, the predictive power of this partner effect was distinct from that of the actor effect and was of equivalent magnitude. This method for assessing neural partner effects represents an approach with untapped potential for researchers' growing usage of neural reward measures to predict individuals' unique preferences among various objects (21–23). By extending the brain-as-predictor approach to interpersonal preferences (i.e., sentiments about fellow study participants as opposed to, for instance, consumer products), we demonstrate how individuals' preferences can be predicted by their neural valuation of particular targets (as in previous neuroeconomic studies cited above) as well as the reciprocal reward response they elicit from each target. Such an extension is necessary for research into the determinants of group social structure and organizational culture, since members of such groups are, by definition, interdependent.

That said, both the actor and partner neural effects can be deemed predictive but not necessarily causal since brain function was measured rather than manipulated (21); hence, we ground these results and their interpretations in the broader neuroscience literature. Many fMRI studies have shown that vmPFC and VS respond to depictions of individuals that we like or who like us in the present (10, 14–19), and our study further demonstrates that these regions' activity prospectively predicts both kinds of outcomes months later (even controlling for their initial levels). These results are consistent with psychological theories of interpersonal attraction based on the self-reinforcing reward value that group members attribute to one another (3, 8), particularly given extensive neuroscientific evidence implicating our ROIs in processing intrinsic value, anticipating reward, and reinforcing behaviors associated with those rewards (11–17). Of course, vmPFC and VS are involved in many processes, not simply reward (11). One possibility is that these brain regions are consistently associated with interpersonal attraction and affect due to their allostatic functions (27).

We also found that the confluence of actor and partner neural effects elucidates how individuals' interpersonal attractions—which were initially unreciprocated at T1—became dyadically coupled by T2. This finding dovetails with a recent fMRI speed-dating study in which vmPFC and VS tracked one's own desires, being desired, and—above all—reciprocation of romantic interest (17). More broadly, neuroscience studies of wide-ranging relations (e.g., affiliation, romantic love, sexual partnership, long-term pair-bonding, and mother–infant attachment) have consistently implicated these brain regions in humans (28, 29) and other animals (30, 31). Considered in sum, the mechanisms we identify are consistent with neuroscience literature on vmPFC and VS, particularly their roles in encoding value, anticipating reward, and perpetuating—that is, forming, maintaining, and reinforcing—the dyadic bonds most valuable to our species.

Future Liking Predicted by One's Own and Partner's Initial Liking. Our finding that T1 liking predicted T2 liking is consistent with social psychological research on the temporal stability of interpersonal attraction (2). It also underscores the importance of testing whether T2 liking was predicted by (implicit) neural measures above and beyond (explicit) self-report measures of T1 liking (21). By controlling for T1 liking as a covariate—specifically, one that indexed baseline measurements of our outcome variable—we could model the evolution of interpersonal attraction, estimating how well each predictor variable forecasts future changes in liking. These changes in liking were profound: although individuals' unique attractions at T2 evidenced statistically significant traces of their initial affinities, the latter explained less than 7% of T2 liking variance (i.e., based on R^2 of model simply regressing T2 liking against T1 liking).

In addition to an actor effect of T1 liking, we also found a corresponding partner effect of comparable magnitude, suggesting that our unique preferences for particular individuals depend on both initially liking and being liked by them. These findings are consistent with the intuition articulated by Newcomb that "attraction breeds attraction" (3), both in the sense that liking is self-reinforcing and—as evoked by the metaphor of breeding—that both members of the dyad contribute to its reproduction.

Conclusion

Our neural findings suggest that our future liking of group members can be jointly predicted by how our neural reward system uniquely responds to them and how theirs uniquely responds to us. Moreover, these implicit measures of neural activity in reward-related ROIs mutually predict both dyad members' future liking above and beyond the predictive effects of explicit measures collected at the same time (i.e., both actor's and partner's initial self-reported liking of each other) as well as sociological antecedents of liking. It is worth noting that we observe these neural measures' long-term prognostic effects despite an experimental context rife with unpredictability; specifically, as young adults both working and living together in small groups for a summer, their relationships were multifaceted and complicated by intense interactions across domains. That the neural responses observed at the start of the program predicted their sentiment months later suggests that they serve as powerful drivers structuring our social relations. Given our study's specific context and small N, however, future research will be needed to assess how well these results replicate and generalize.

Considered together, our findings suggest that the neural reward systems engaged when two group members encounter each other may interdependently shape their future liking and facilitate its mutual reciprocation. As such, this study offers insight into the mechanisms underlying how we ultimately end up liking and attracting certain people, how affective reciprocity emerges, and more broadly, how individuals bond to form dyads. Finally, we advance a framework for researching the neural foundations

of interpersonal sentiments and social relations that emphasizes human relational interdependence.

Methods

Participants. Participants were 16 students who volunteered to spend 9 wk together to organize workers and collect oral histories (*SI Text*). They received monetary compensation and provided informed consent following the approval of all experimental procedures by the Columbia University Institutional Review Board.

Procedure and Design. The T1 component of the study consisted of two sessions. In a preliminary session, sociometric instruments of liking (described below) and self-report questionnaires were administered, and photographs were taken of participants' faces. In a second session, participants underwent fMRI scanning while completing the face-viewing task 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). The T2 wave of data collection included sociometric assessments and was administered via Qualtrics online survey software after conclusion of the 9-wk summer program. For both T1 and T2, sociometric assessments were conducted via a computerized peer-rating paradigm in which participants rated how much they liked each group member on a sliding visual analog scale anchored by the labels "not very" and "very" on opposite ends (later converted to 0–100 values). Additional data were collected for the purposes of other studies.

Round-Robin fMRI Face-Viewing Task. Methods relating to various aspects of this fMRI face-viewing task (e.g., round-robin experimental design, stimulus preparation, and participant procedures) were developed and described in our previous work (9). To prepare task stimuli, 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. The face-viewing task implemented a rapid event-related design that included 10 repetitions of each stimulus face presented for 1,000 ms in pseudorandomized order. Participants performed a simple cover task to maintain their alertness, pressing one button each time a group member's face was presented and a different button each time a "ghost face" (superimposition of all face stimuli) was presented (~5% of total presentations). Interstimulus intervals consisting of white fixation cross on black background were jittered between 1,500 and 8,500 ms (mean, 3,500 ms). Stimuli were displayed on a projection screen using a LCD projector and viewed via a rear-projecting mirror.

Image Acquisition. Data were acquired on a 3-T GE system with a 32-channel RF head coil. A T1-weighted sagittal 3D BRAVO sequence yielded high-resolution anatomical images with 1-mm^3 isometric voxels. Functional images were acquired with a T2*-sensitive echo-planar–imaging blood oxygenation level-dependent sequence using the following parameters: repetition time, 2,000 ms; echo time, 25 ms; flip angle, 77°; field of view, 19.2 cm × 19.2 cm; and each volume consisted of 45 slices with 3-mm thickness and no interslice gap. Functional volumes were collected in three runs, each consisting of 167 volumes (and four initial "dummy" volumes discarded before analysis). See *SI Text* for preprocessing parameters and further details.

Reward/Valuation System ROIs. We independently defined a priori ROIs underlying valuation and reward processes in a separate participant sample (these data were published previously in ref. 9). Following the protocol of prior studies (32), we used an established functional localizer task (33) to identify ROIs engaged in anticipating and receiving monetary reward; specifically, we defined spherical ROIs with 8-mm radius surrounding activation peaks in vmPFC (−3, 48, −6) and VS (0, 9, −3), constrained using an anatomical mask for VS. Parameter estimates extracted from vmPFC and VS ROIs were averaged together for a composite neural measure of reward value (9). Results for ROI analyses using parameter estimates extracted from only vmPFC or only VS are reported in Table S1.

ACKNOWLEDGMENTS. We thank the participants in this study, J. Weber for help with fMRI data analysis, and N. Bolger for advice with statistical analysis of dyadic data. This work was supported by a Columbia University Interdisciplinary Center for Innovative Theory and Empirics seed grant.

1. Kenny DA (1994) *Interpersonal Perception: A Social Relations Analysis* (Guilford Press, New York).
2. Newcomb TM (1963) Stabilities underlying changes in interpersonal attraction. *J Abnorm Soc Psychol* 66:376–386.
3. Newcomb TM (1956) The prediction of interpersonal attraction. *Am Psychol* 11: 575–586.
4. Gouldner AW (1960) The norm of reciprocity: A preliminary statement. *Am Sociol Rev* 25:161–178.
5. Homans GC (1958) Social behavior as exchange. *Am J Sociol* 63:597–606.
6. Bearman P (1997) Generalized exchange. *Am J Sociol* 102:1383–1415.
7. Freud S (1912) The dynamics of transference. *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, trans ed Strachey J (Hogarth Press, London), Vol 12, pp 97–108.
8. Lott AJ, Lott BE (1974) The role of reward in the formation of positive interpersonal attitudes. *Foundations of Interpersonal Attraction*, ed Huston TL (Academic, New York), pp 171–192.
9. Zerubavel N, Bearman PS, Weber J, Ochsner KN (2015) Neural mechanisms tracking popularity in real-world social networks. *Proc Natl Acad Sci USA* 112:15072–15077.
10. Chen AC, Welsh RC, Liberzon I, Taylor SF (2010) 'Do I like this person?' A network analysis of midline cortex during a social preference task. *Neuroimage* 51:930–939.
11. Doré BP, Zerubavel N, Ochsner KN (2014) Social cognitive neuroscience: A review of core systems. *APA Handbook of Personality and Social Psychology*, eds Mikulincer M, et al. (American Psychological Association, Washington, DC), pp 693–720.
12. Bartra O, McGuire JT, Kable JW (2013) The valuation system: a coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *Neuroimage* 76:412–427.
13. Haber SN, Knutson B (2010) The reward circuit: linking primate anatomy and human imaging. *Neuropsychopharmacology* 35:4–26.
14. Fareri DS, Delgado MR (2014) Social rewards and social networks in the human brain. *Neuroscientist* 20:387–402.
15. Ruff CC, Fehr E (2014) The neurobiology of rewards and values in social decision making. *Nat Rev Neurosci* 15:549–562.
16. Jones RM, et al. (2011) Behavioral and neural properties of social reinforcement learning. *J Neurosci* 31:13039–13045.
17. Cooper JC, Dunne S, Furey T, O'Doherty JP (2014) The role of the posterior temporal and medial prefrontal cortices in mediating learning from romantic interest and rejection. *Cereb Cortex* 24:2502–2511.
18. Davey CG, Allen NB, Harrison BJ, Dwyer DB, Yücel M (2010) Being liked activates primary reward and midline self-related brain regions. *Hum Brain Mapp* 31:660–668.
19. Gunther Moor B, van Leijenhorst L, Rombouts SA, Crone EA, Van der Molen MW (2010) Do you like me? Neural correlates of social evaluation and developmental trajectories. *Soc Neurosci* 5:461–482.
20. Izuma K, Saito DN, Sadato N (2008) Processing of social and monetary rewards in the human striatum. *Neuron* 58:284–294.
21. Berkman ET, Falk EB (2013) Beyond brain mapping: Using neural measures to predict real-world outcomes. *Curr Dir Psychol Sci* 22:45–50.
22. Tusche A, Bode S, Haynes J-D (2010) Neural responses to unattended products predict later consumer choices. *J Neurosci* 30:8024–8031.
23. Levy I, Lazzaro SC, Rutledge RB, Glimcher PW (2011) Choice from non-choice: predicting consumer preferences from blood oxygenation level-dependent signals obtained during passive viewing. *J Neurosci* 31:118–125.
24. Kenny DA, Kashy DA, Cook WL, Simpson J (2006) *Dyadic Data Analysis*. Methodology in the Social Sciences (Guilford, New York).
25. Olsen JA, Kenny DA (2006) Structural equation modeling with interchangeable dyads. *Psychol Methods* 11:127–141.
26. Schönbrodt FD, Back MD, Schmukle SC (2012) TripleR: an R package for social relations analyses based on round-robin designs. *Behav Res Methods* 44:455–470.
27. Kleckner IR, et al. (2017) Evidence for a large-scale brain system supporting allostasis and interoception in humans. *Nat Hum Behav* 1:0069.
28. Acevedo BP (2015) Neural correlates of human attachment: Evidence from fMRI studies of adult pair-bonding. *Bases of Adult Attachment* (Springer, New York), pp 185–194.
29. Bickart KC, Hollenbeck MC, Barrett LF, Dickerson BC (2012) Intrinsic amygdala-cortical functional connectivity predicts social network size in humans. *J Neurosci* 32: 14729–14741.
30. Young LJ, Lim MM, Gingrich B, Insel TR (2001) Cellular mechanisms of social attachment. *Horm Behav* 40:133–138.
31. Curtis JT, Liu Y, Aragona BJ, Wang Z (2006) Dopamine and monogamy. *Brain Res* 1126:76–90.
32. Zaki J, Schirmer J, Mitchell JP (2011) Social influence modulates the neural computation of value. *Psychol Sci* 22:894–900.
33. Knutson B, Westdorp A, Kaiser E, Hommer D (2000) FMRI visualization of brain activity during a monetary incentive delay task. *Neuroimage* 12:20–27.