Supporting Appendix

Neural detection of socially valued community members

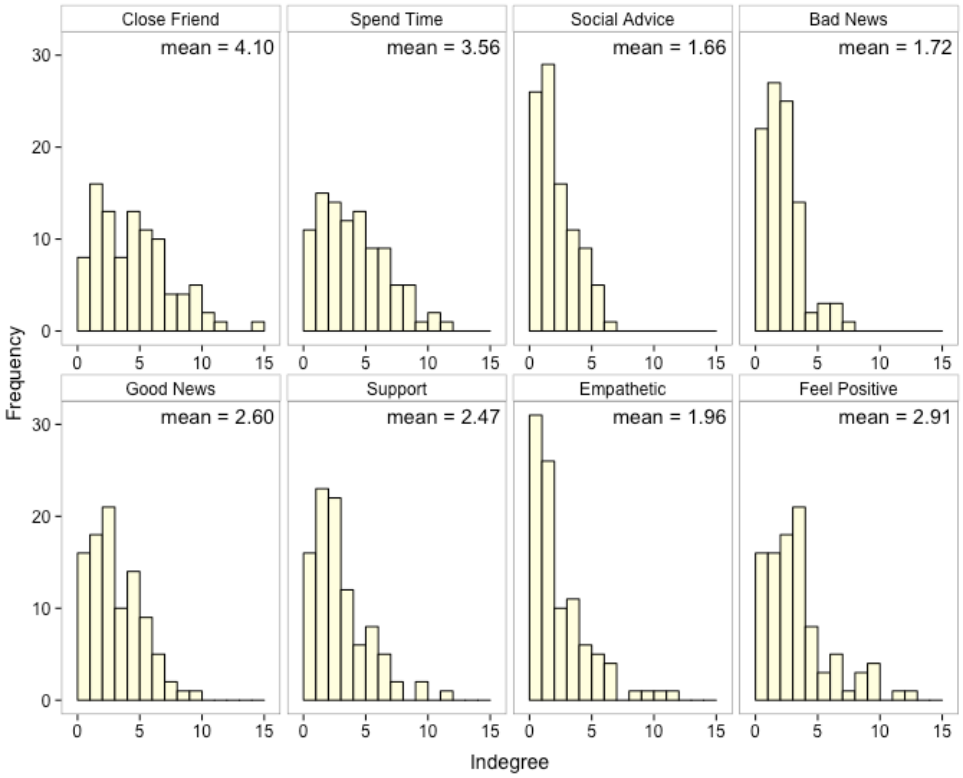
*Sylvia A. Morelli, *Yuan Chang Leong, Ryan W. Carlson, Monica Kullar, & Jamil Zaki

Supplemental Methods

Social Network Nominations

Factor analysis. For each social support question, we computed *indegree* by summing the number of ties directed to each person from the dorm [1] (Fig. S1).

Figure S1. Histograms of indegree for each of the 8 social network questions, collapsed across both dorms.



As a data reduction step, we then performed a factor analysis on indegree for each of the eight questions, using the full sample (i.e., 97 participants) to increase reliability. Using the

"psych" package in R, a parallel analysis (i.e., a factor retention method) recommended one factor in the exploratory analysis [2] (Fig. S2). As a result, we specified a one-factor model with unweighted least squares extraction (i.e., "minres") with no rotation. We evaluated model fit with the Tucker–Lewis Index (TLI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRSR). Generally TLI values above .90, as well as RMSEA and SRSR values of .08 or less indicate adequate fit [3]. The one-factor solution yielded acceptable fit across indices: TLI = .86, RMSEA = .21, SRSR = .05. In addition, factor loadings for this model indicated relatively high internal consistency (see Table S1), ranging from .72 to .91. Overall, these analyses revealed that indegree across nominations emerges as one composite factor.

Figure S2. Parallel analysis scree plots of indegree for each of the eight nomination questions

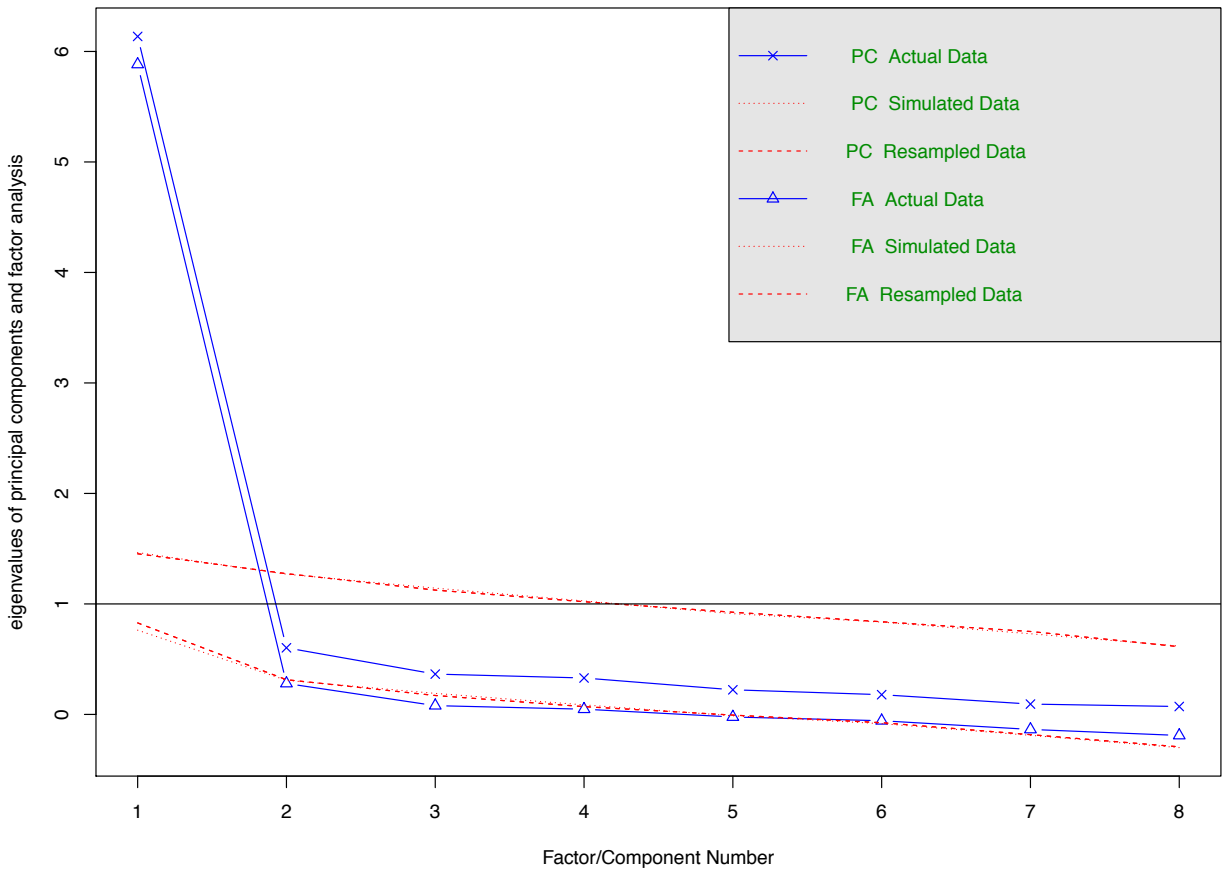


Table S1. Factor loadings for indegree of each of the eight social network nominations

Indegree	Full Sample (n = 97)
1. Who are your closest friends?	.91
2. Whom do you spend the most time with?	.91
3. Whom have you asked for advice about your social life?	.83
4. Who do you turn to when something bad happens?	.84
5. Whom do you share good news with?	.94
6. Who makes you feel supported and cared for?	.89
7. Who is the most empathetic?	.72
8. Who usually makes you feel positive (e.g., happy, enthusiastic)?	.81

Social Network Graphs

Our process for generating social network graphs in Fig. 1 was as follows: (1) we generated an adjacency matrix from all the nominations in each dorm for each of the eight questions, (2) we multiplied each matrix by its respective factor loading (Table S1), (3) we averaged these weighted matrices together, (4) we converted this new matrix into a list of edges that represented the strength and direction of a tie between individuals in the network. We then created a Gephi graph from this list of edges with a "force-based" algorithm (i.e., Force Atlas 2; https://github.com/gephi/gephi/wiki/Force-Atlas-2). This algorithm attracts linked nodes to each other and pushes non-linked nodes apart. Therefore, nodes that are closer together represent denser interconnections. Each node represents an individual in the dorm, and larger nodes indicate higher numbers of nominations received from the dorm. Arrows represent the strength and direction of connections between individuals. More specifically, the color and thickness of the arrow represents the weighted average of the connections between the source and the target.

Pre-Scan Ratings of Dorm Relationships

In an online survey, participants saw a photo of each dorm member and were asked if they had talked or interacted with him/her. Using a 7-point likert scale, they were then asked to rate each dorm member on various dimensions (in randomized order): (1) how close are you to this person? (2) how much do you like this person? (3) how empathetic is this person? (4) how happy does this person seem in general? (5) how much time do you usually spend with this person each week? (6) how attractive is this person? (7) how much do you think this person likes you?

Face Selection Algorithm

To select which 30 faces were displayed in the face-viewing task, we first removed dorm members with whom the participant had not interacted (see above). When the resulting list included less than 30 dorm members (only occurred for 3 participants), we randomly included dorm members that the perceiver had not interacted with. When the resulting list included over 30 dorm members, we created a distribution using the relationship index (i.e., a sum of questions 1-4 for pre-scan ratings, see Methods in main manuscript) and then randomly removed one participant from the modal bin until we reached a total of 30 dorm members. This approach ensured that we included a wide distribution of dorm members with varied levels of closeness with the perceiver.

Image Acquisition

Whole-brain fMRI data were acquired on a 3.0 Tesla GE magnetic resonance imaging scanner with a 32-channel head coil at the Stanford Center for Cognitive and Neurobiological Imaging. For all tasks, we used MATLAB with the Psychophysics Toolbox extensions [4] to present the tasks to participants and record their responses. Participants viewed the fMRI tasks via a mirror system and made their responses via an fMRI compatible button response box. For

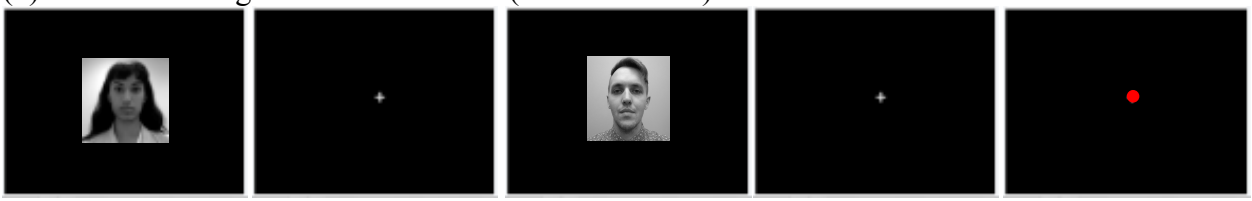
the face-viewing task, we collected 149 functional T2* weighted gradient echo pulse sequence image volumes in one functional run (slice thickness = 2.9 mm, no gap, 46 slices, TR = 2000 ms, TE = 25 ms, flip angle = 77°, interleaved acquisition). For the functional reward localizer, we collected 160 functional T2* weighted gradient echo pulse sequence image volumes in one functional run. High-resolution structural images were also acquired with a T1-weighted pulse sequence (slice thickness = 0.9 mm, flip angle = 12°, interleaved acquisition).

Image Preprocessing

For each participant's EPI time series, the first four volumes were discarded to account for T1-equilibration effects. Manual quality checking of all participants' scans was conducted through FSL view for potential dropout, distortion, radio frequency noise, spiking, motion, and other common artifacts. Through SPM8 (Wellcome Department of Imaging Neuroscience, London), we reoriented, realigned, co-registered to the structural scan, spatially normalized to a standard MNI (Montreal Neurological Institute) template using segmentation parameters, and smoothed (6 mm full width at half maximum Gaussian kernel) all functional images.

Face-Viewing Task

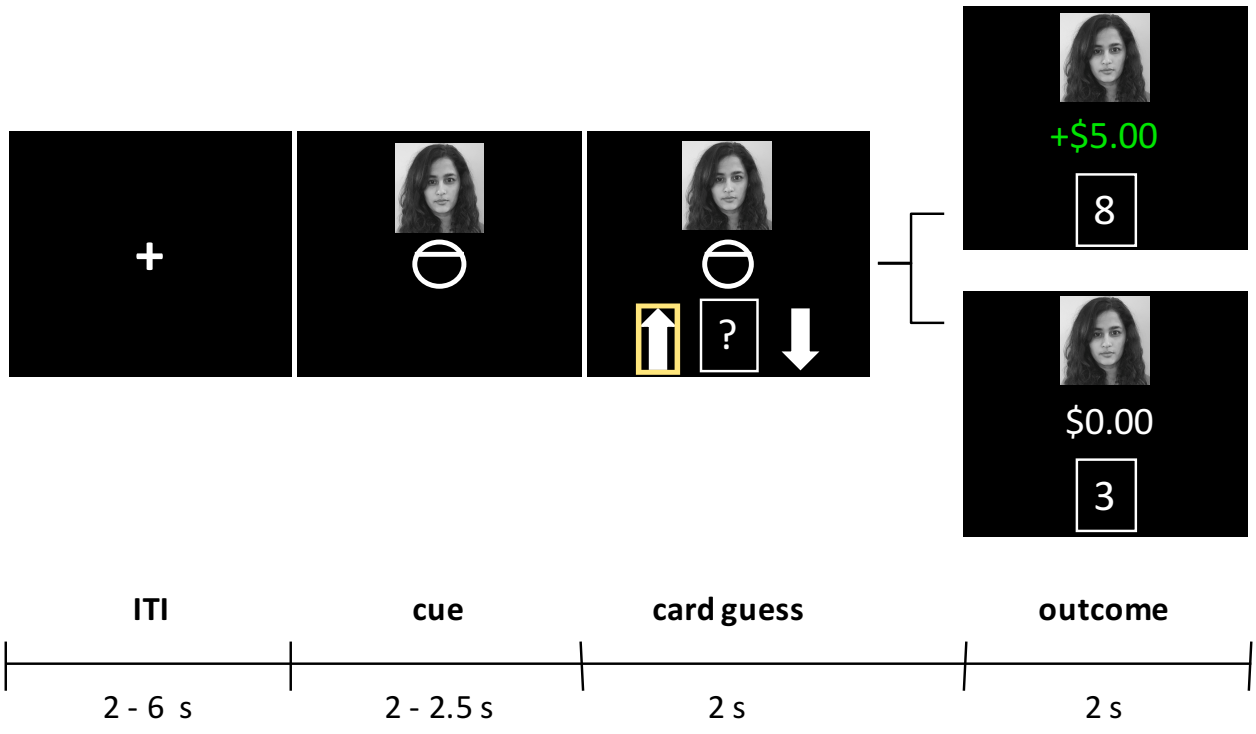
Figure S3. A schematic of the rapid event-related design for the face-viewing task. Participants viewed (1) photos of their dorm members, (2) a crosshair during the interstimulus interval, and (3) a red dot during the attention check (~16% of trials).



Functional Reward Localizer

Task description. The reward localizer (Fig. S4) was a novel combination of two popular reward tasks, blending elements of the Monetary Incentive Delay task [5] with a card-guessing task [6-8]. Before entering the scanner, the experimenter told participants that they would be playing a card-guessing game and could win real money for themselves (for two randomly selected games). Participants played games in which they could (i) potentially gain money (20 trials) or (ii) receive no money (10 trials). These 30 trials were fully randomized in one functional run. Because our main goal was to localize neural activity related to rewarding outcomes, we only describe potential gain trials below.

Figure S4. A schematic for potential gain trials in the card-guessing game. Participants viewed (1) a crosshair during the intertrial interval (ITI), (2) saw their own photo and a cue indicating that they could win $5, (3) guessed whether the card would be above or below 5, and (4) saw the value of the card and the amount of money they received.



For potential gain trials, participants viewed (1) a crosshair during the intertrial interval (ITI) for 2, 4, or 6 s, (2) saw their own photo along with a cue indicating that they could potentially win $5 (i.e., a circle with a line at the top) for 2, 2.25, or 2.5 s, (3) guessed whether the card would be above 5 (i.e., up arrow) or below 5 (i.e., down arrow) from a deck of cards from 1-9 (excluding 5) for 2 s, and (4) saw the actual value of the card (e.g., 8 or 3) and the corresponding amount of money they gained or did not gain (i.e., +5.00 or $0.00) for 2 s (Fig. S4). A correct guess yielded a monetary gain of $5 (in green), whereas an incorrect guess resulted in $0 (in white). If participants failed to make a card guess, participants saw the word "MISS" in red. Outcomes were preprogrammed to lead to 50% reward and 50% no-reward outcomes.

Analytical approach. We aimed to identify striatal regions related to the *receipt* of reward (not reward anticipation). We estimated first-level effects with the general linear model. In this model, the two covariates of interest included the 10 reward *outcomes* (i.e. +$5) and the 10 no-reward *outcomes* (i.e., $0) for the potential gain trials (see Fig. S4). If participants failed to make a guess, these "misses" were modeled as a covariate of no interest. In addition, all other parts of the potential gain trials (i.e., cues and card guesses) were modeled as separate nuisance regressors. Lastly, the six motion parameters and time points with excessive motion or in-brain global signal were also included as additional covariates of no interest. All trial components were modeled as a boxcar spanning their duration and convolved with a double-gamma HRF and high-pass filtered (cutoff of 128 s). Serial autocorrelations were modeled as an AR(1) process.

For each participant, we created contrast images comparing reward outcomes to no-reward outcomes. Then, we computed random effects analyses of the group using the contrast images generated for each participant. The threshold was raised until we could identify a distinct cluster for bilateral ventral and dorsal striatum, with a final threshold of $p < .00000001$

(uncorrected). We then used this cluster in the striatum as a region of interest (ROI) during leave-one-out cross-validation analyses. The unthresholded map is available on NeuroVault: http://neurovault.org/images/51937/.

We only selected the striatum because all additional regions that were activated in the reward localizer were redundant with other regions of interest. In addition to the striatum, the reward localizer activated the MPFC, DMPFC, VMPFC, and poster cingulate. However, these regions significantly overlapped with ROIs we generated on Neurosynth for the mentalizing network (see below). Therefore, we decided to use the ROIs generated from Neurosynth because those were derived from 124 studies (rather than a single study).

Regions of Interest

In addition to the ROI created for reward receipt (see above), we also created ROIs for regions related to mentalizing. First, we searched for the term "mentalizing" on Neurosynth (http://www.neurosynth.org/analyses/terms/mentalizing/) and found 124 studies associated with this key term. Next, we downloaded the reverse inference map (FDR-corrected at $p = .01$). From this map, we selected the following clusters as separate ROIs: (1) dorsomedial, medial, and ventromedial prefrontal cortex (labelled as "MPFC" in main manuscript), (2) precuneus/posterior cingulate (labelled as "PMC" in main manuscript), (2) R temporoparietal junction, (3) L temporoparietal junction, (4) R temporal poles, and (5) L temporal poles.

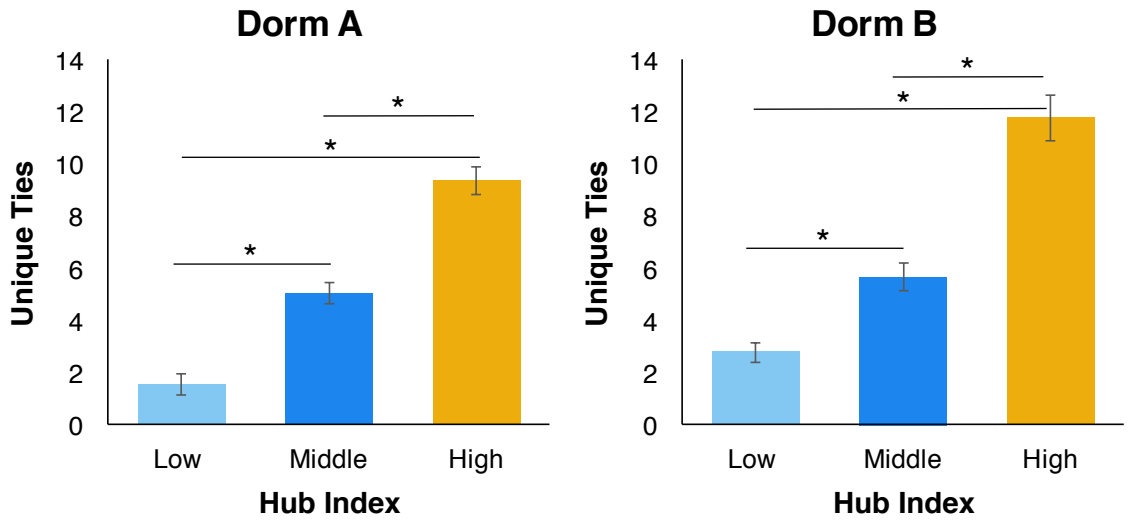
For comparison in leave-one-out cross-validation analyses, we also created an ROI of the primary visual cortex as a "control" region. We searched for the term "V1" on Neurosynth (http://www.neurosynth.org/analyses/terms/v1/) and found 89 studies associated with this key term. We then downloaded the reverse inference map (FDR-corrected at $p = .01$) and selected the largest and most medial cluster that included the highest number of voxels in Brodmann's Area 18. For all ROIs, individual clusters were: (i) selected in xjView and exported as masks, (ii)

imported into MarsBaR [9] and converted to ROIs, and (iii) and exported into the same image space as group-level analyses for the face-viewing task.

Supplemental Results

The Relationship Between Hub Index and Number/Strength of Ties

Figure S5. The average number of unique ties for individuals in the low, middle, and high hub categories. Independent-samples t-tests showed that all categories were significantly different from each other (* $p < .001$) in both dorms.



To decompose how the number and strength of ties relates to hub index, we conducted multiple regression analyses in each dorm. We entered two simultaneous predictors to predict hub index: (i) number of unique connections across all 8 questions and (ii) average number of questions nominated for (i.e., strength of ties). In both dorms, the number of ties positively related to hub index (Dorm A: $\beta = .92$, $p < .001$, Dorm B: $\beta = .93$, $p < .001$; Fig. S6). The strength of ties also positively related to hub index (Dorm A: $\beta = .27$, $p < .001$ Dorm B: $\beta = .29$, $p < .001$), but not as strongly. Stepwise regression analyses further revealed that the number of ties predicts 85% (Dorm A) - 87% (Dorm B) of the variance in hub index, after controlling for the strength of ties. In contrast, the strength of ties predicts 7% (Dorm A) - (Dorm B) 9% of the variance, after controlling for the number of ties. Thus, higher hub index scores were primarily driven by the number of unique connections and less so by the strength of these connections.

Figure S6. The relationship between the number of unique ties and hub index.

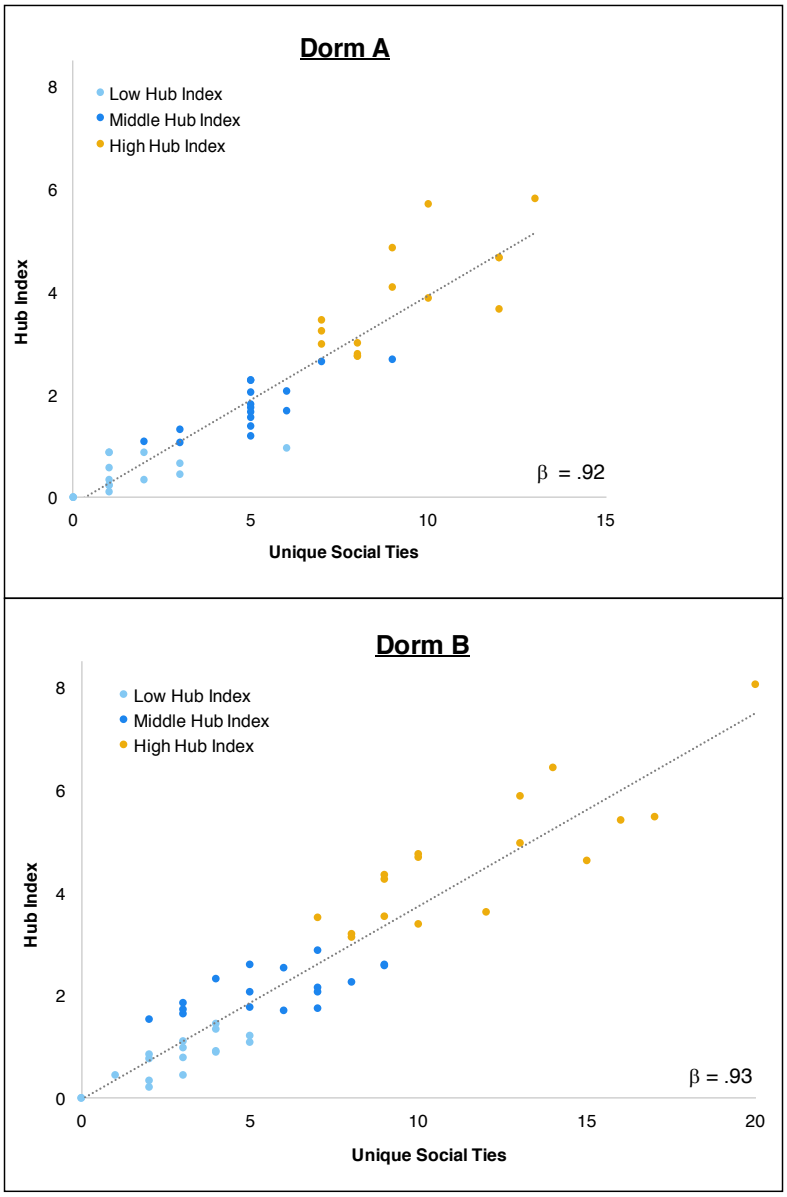


Table S2. Parametric effect of targets' idiosyncratic hub index, after controlling for the perceiver's nominations and their self-reported closeness to targets.

Region Name	Cluster size (voxels)	Peak t-score	Peak MNI coordinates (mm)		
			x	y	z
R caudate	96	5.350	24	17	19
		4.315	21	-16	28
L caudate	74	5.143	-15	5	22
		3.915	-21	23	13
R cerebellum	123	5.142	51	-73	-17
		3.900	18	-70	-29
		3.687	39	-91	-20
Precuneus/occipital lobe	75	4.683	3	-85	46
L temporal pole	79	4.599	-51	5	-29
Medial prefrontal cortex	111	4.536	9	47	4
Medial prefrontal cortex	47	4.224	3	68	22

Notes. Coordinates are all local maxima separated by at least 20 mm. L and R refer to left and right hemispheres. x, y, and z are Montreal Neurological Institute (MNI) coordinates in the left-right, anterior-posterior, and inferior-superior dimensions, respectively.

Table S3. Average within-subject correlation between predicted and actual hub category in different ROIs, separately for univariate and multivariate pattern prediction.

Region	R (univariate)	R (multivariate)
Mentalizing Network	.337 ± .100 **	.343 ± .090 **
MPFC	.318 ± .098 **	.405 ± .094 ***
PMC	.339 ± .095 **	.318 ± .099 **
L TP	.297 ± .098 **	.319 ± .091 **
R TP	.276 ± .105 **	.207 ± .101 *
L TPJ	.311 ± .095 **	.259 ± .095 *
R TPJ	.246 ± .098 *	.247 ± .095 *
Striatum	.136 ± .096	.298 ± .091 **
V1	.164 ± .103	.086 ± .100

Notes. Statistical significance was assessed using non-parametric permutation tests. Plus-minus values indicate standard error of the mean. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Figure S7. Univariate prediction of social value hubs**. A)** Neural activity was averaged in the depicted regions of interest (ROI) and used to predict hub category in leave-one-out cross-validation models. **B)** Forced-choice accuracy between different hub categories for each ROI. * $p < 0.05$ for a two-sided binomial test. MPFC = medial prefrontal cortex; PMC = posterior medial cortex; TP = temporal poles; TPJ = temporoparietal junction; R = right; L = left. Red: regions associated with mentalizing; blue: regions associated with value computation; grey: control region.

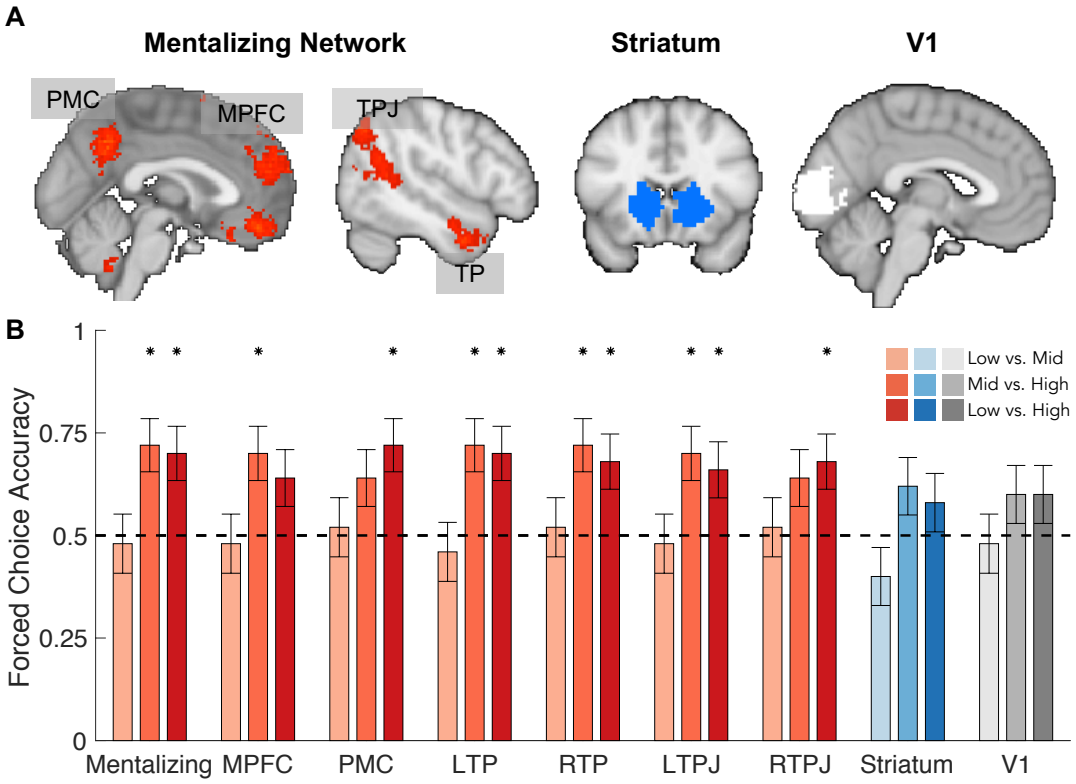
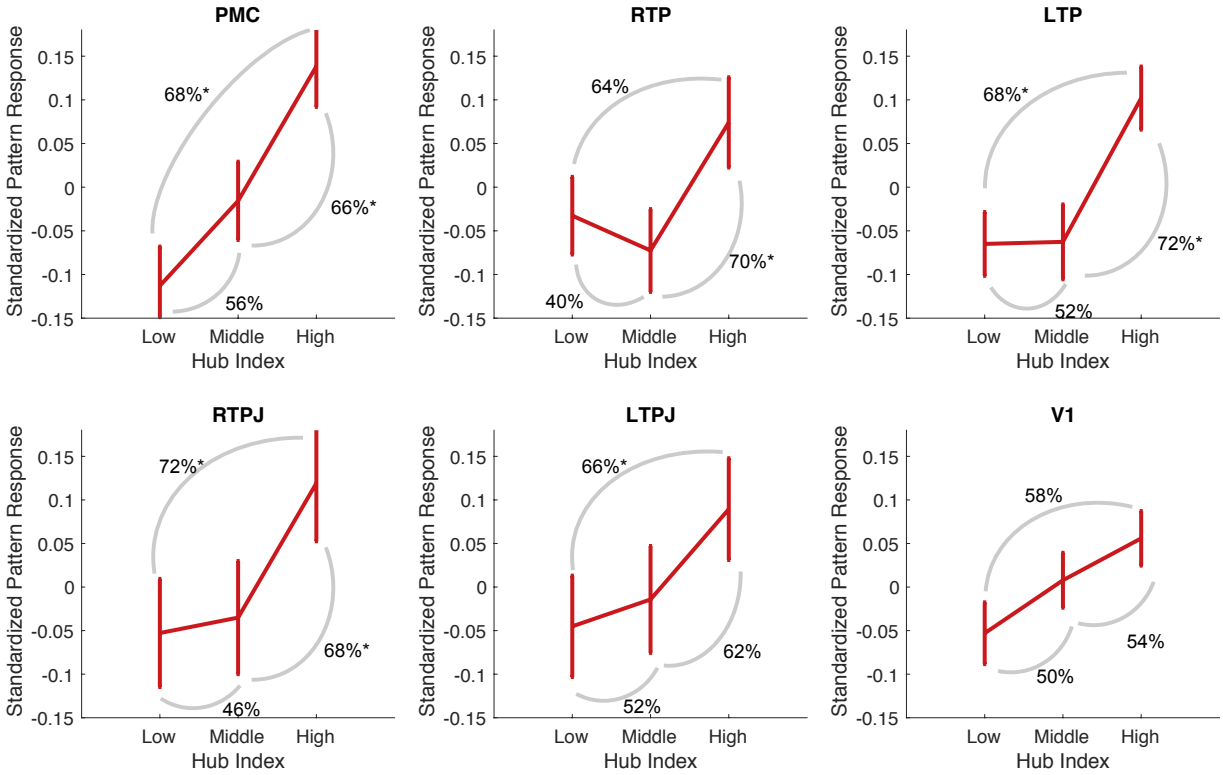


Figure S8. Average standardized pattern response to hub index. The standardized pattern response was computed as the spatial correlation between the activation map for each level of hub index and the multivariate pattern associated with increasing social hub index. Higher pattern response indicates that the activation map was more similar to the pattern associated with increasing hub index. Percentage indicates forced-choice classification between each pairwise comparison. Error bars indicate SEM across participants. * $p < 0.05$ for a two-sided binomial test. MPFC = medial prefrontal cortex; PMC = posterior medial cortex; TP = temporal poles; TPJ = temporoparietal junction; R = right; L = left.



Additional Evidence for Non-Linear Relationship Between Hub index and Neural Activity

If social value is tracked linearly, we would expect the classifiers to show increasing predicted values from low to medium hub category, and from medium to high hub category. Critically, the differences in predicted values between low vs. medium and medium vs. high hub category would be roughly equal to each other. In contrast, if social value is tracked in a non-linear manner, then we would expect that these same comparisons would not be equal. When we tested these differences in predicted value for the univariate (Table S4) and multivariate classifiers (Table S5), we showed that the increase in predicted response is greater from middle to high hub index, than from low to middle hub index. These results provide further evidence that the relationship between hub index and neural activity was in fact non-linear, with a much stronger neural response to targets high in social value.

Table S4. Increase in predicted response from univariate activity is higher from middle to high hub category than from low to middle hub category

ROI	Mid > Low	High > Mid	Difference
Mentalizing Network	-0.01 (0.06)	0.22 (0.06)	t(49) = 2.18, p = 0.017
MPFC	-0.03 (0.06)	0.21 (0.06)	t(49) = 2.46, p = 0.009
PMC	0.01 (0.07)	0.20 (0.06)	t(49) = 1.76, p = 0.042
R TP	-0.03 (0.05)	0.17 (0.05)	t(49) = 2.45, p = 0.009
L TP	-0.04 (0.06)	0.19 (0.05)	t(49) = 2.42, p = 0.010
R TPJ	-0.01 (0.05)	0.11 (0.04)	t(49) = 1.50, p = 0.070
L TPJ	0.03 (0.06)	0.15 (0.06)	t(49) = 1.16, p = 0.127
Striatum	-0.02 (0.03)	0.04 (0.03)	t(49) = 1.17, p = 0.124
V1	-0.01 (0.03)	0.04 (0.03)	t(49) = 1.00, p = 0.160

Notes. Parentheses denote the standard error of the mean. ROI = region of interest; MPFC = medial prefrontal cortex; PMC = posterior medial cortex; TP = temporal poles; TPJ = temporoparietal junction; R = right; L = left. p-value determined by a right-tailed paired t-test.

Table S5. Increase in predicted response from multivariate activity is higher from middle to high hub category than from low to middle hub category

ROI	Mid > Low	High > Mid	Difference
Mentalizing Network	-0.03 (0.09)	0.33 (0.08)	t(49) = 2.40, p = 0.010
MPFC	-0.04 (0.09)	0.41 (0.06)	t(49) = 3.74, p < 0.001
PMC	-0.03 (0.07)	0.22 (0.07)	t(49) = 1.69, p = 0.049
R TP	-0.04 (0.06)	0.21(0.06)	t(49) = 2.33, p = 0.012
L TP	-0.01 (0.08)	0.25 (0.06)	t(49) = 2.09, p = 0.021
R TPJ	-0.01 (0.06)	0.14 (0.05)	t(49) = 1.42, p = 0.081
L TPJ	0.01 (0.06)	0.16 (0.07)	t(49) = 1.39, p = 0.086
Striatum	0.01 (0.05)	0.16 (0.06)	t(49) = 1.35, p = 0.092
V1	-0.02 (0.05)	0.07 (0.06)	t(49) = 1.00, p = 0.160

Notes. Parentheses denote the standard error of the mean. ROI = region of interest; MPFC = medial prefrontal cortex; PMC = posterior medial cortex; TP = temporal poles; TPJ = temporoparietal junction; R = right; L = left. p-value determined by a right-tailed paired t-test.

Table S6. Comparison between prediction for mean activity versus multivariate patterns. The difference column reports the results from each paired-samples t-test, comparing root mean squared error (RMSE) from multivariate prediction to RMSE from univariate prediction.

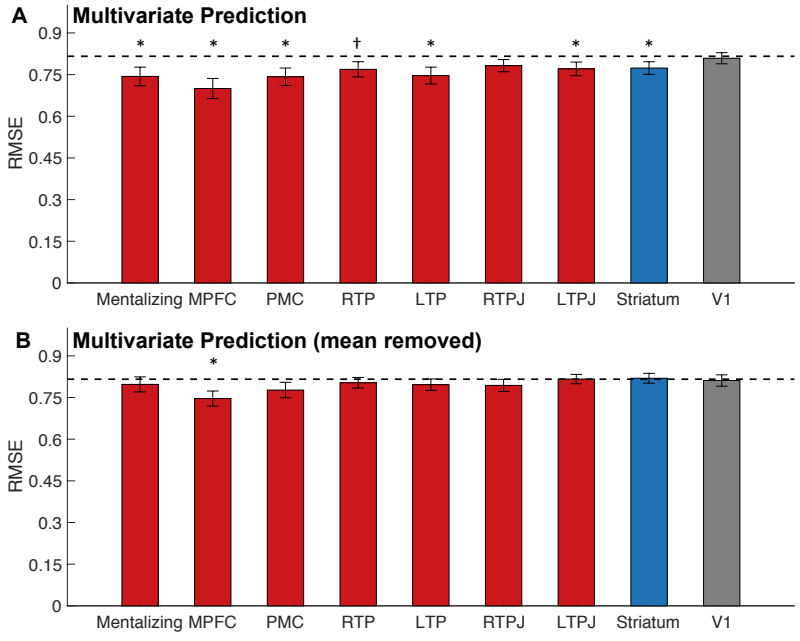
Region	Average Multivariate RMSE	Average Univariate RMSE	Difference
Mentalizing Network	.743	.752	t(49)= -0.367, p=0.716
MPFC	.700	.761	t(49)=-2.779 p=0.008**
PMC	.742	.753	t(49)=-0.436, p=0.665
R TP	.769	.776	t(49)=-0.552, p=0.583
L TP	.747	.769	t(49)=-1.571, p=0.123
R TPJ	.787	.782	t(49)=-0.377, p=0.708
L TPJ	.771	.764	t(49)=0.583, p=0.562
Striatum	.774	.813	t(49)= -2.055, p=0.045*
V1	.809	.808	t(49)=0.054, p= 0.957

Notes. * $p < 0.05$, ** $p < 0.01$. RMSE = root mean squared error; ROI = region of interest; MPFC = medial prefrontal cortex; PMC = posterior medial cortex; TP = temporal poles; TPJ = temporoparietal junction; R = right; L = left.

Multivariate Prediction With and Without Mean Signal Removed

We compared predictive accuracy of the multivariate model with and without the mean ROI signal removed. When the mean ROI signal was not removed, most ROIs in the mentalizing network and the striatum had above chance predictive accuracy (Fig. S8). We then removed the mean ROI signal by z-scoring the voxels in each ROI. Predictive accuracy was above chance only in the MPFC ($t(49) = -2.60$, $p = 0.012$), suggesting that fine-grained patterns contributed to the prediction in the MPFC, but not in the other ROIs.

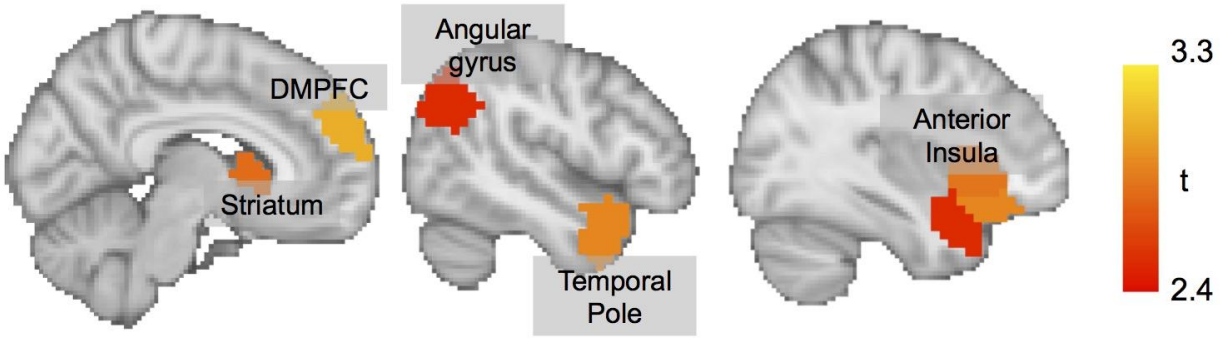
Figure S9. Multivariate patterns do not contribute substantially to predicting hub index outside of the MPFC. **A.** Leave-one-participant-out cross-validated root mean squared error (RMSE) when predicting hub index multivariate patterns, without removing mean ROI activity. RMSE was compared to "chance" RMSE of 0.816 (the RMSE if the algorithm always predicted the mean hub index; dotted line). There was better than chance accuracy in most ROIs in the mentalizing network and the striatum, but not V1. **B.** RMSE when predicting hub index from the multivariate patterns, removing mean ROI activity. RMSE was only better than chance in the MPFC. * $p < 0.05$, † < 0.10.



Whole-Brain Multivariate Prediction

We repeated the multivariate prediction analysis across 100 ROIs taken from a whole-brain functional parcellation [10]. For each ROI, we performed the same leave-one-participant-out cross-validated prediction analysis and computed the root mean squared error (RMSE) of the prediction for each participant. We then tested the RMSE against chance, defined as the RMSE if the algorithm always predicted the mean hub category, to obtain a t-statistic and p-value for each ROI. No ROI survived correction when controlling for false-discovery rate of $q < 0.05$. When we inspected the t-map thresholded at $p < 0.01$ (unthresholded t-map available at: https://neurovault.org/collections/2715/; Fig. S8), we found activity in regions related to mentalizing and value processing, including the dorsomedial prefrontal cortex, temporal poles, angular gyrus, and striatum significantly predicted hub status. Only one region outside of these networks – the anterior insula – explained significant variance in the prediction.

Figure S10. Whole-brain multivariate prediction. We repeated the multivariate prediction analysis across 100 ROIs taken from a whole-brain functional parcellation [10]. Resulting t-map was thresholded at $p < 0.01$. No ROI survived correction when controlling for false-discovery rate with $q < 0.05$.



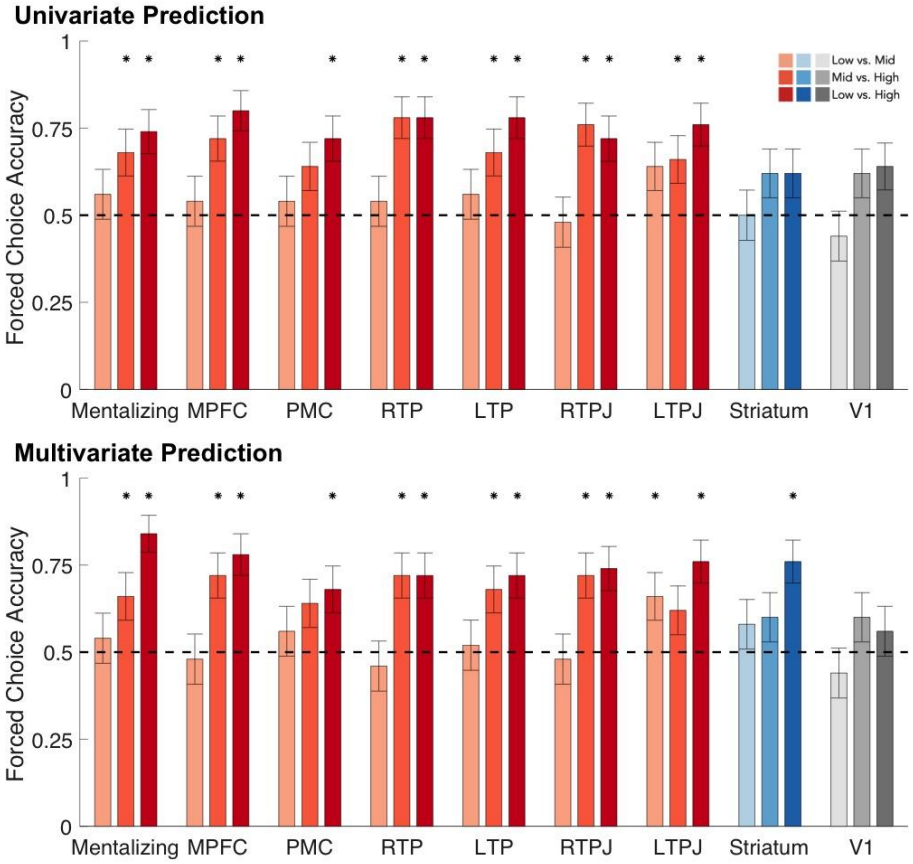
Unique Effects of Target Characteristics on Neural Activity

In order to examine the unique effects of other target characteristics, we conducted a parametric analysis with three parametric modulators in the general linear model: (i) hub index, (ii) pre-scan ratings of closeness, and (iii) pre-scan ratings of attractiveness. Activity in inferior frontal gyrus scaled with personal closeness. No significant clusters were parametrically modulated by attractiveness.

Neuroimaging Analyses Without Personal Relationship Covariates

To examine whether the results are robust and hold when removing covariates related to the personal relationship between the perceiver and the target, we repeated all neuroimaging analyses without any covariates. More specifically, we did *not* exclude personal nominations when calculating the hub index and did not control for subjective ratings of closeness. For the univariate parametric analyses (http://neurovault.org/collections/2715/), the results are similar to the results that control for personal relationships, with more robust clusters appearing in regions related to mentalizing (i.e., medial prefrontal cortex, posterior medial cortex, temporoparietal junction, and temporal poles) and value processing (i.e., striatum). We also repeated the univariate and multivariate prediction analyses without personal relationship covariates. Forced classification accuracy rates (Fig. S9) were highly similar when compared to results accounting for personal relationships.

Figure S11. Forced-choice classification results without personal relationship covariates. In both univariate and multivariate prediction analyses, accuracy was above chance when classifying between high vs. medium and high vs. low social support individuals.



Neuroimaging Analyses Controlling for Target Attractiveness

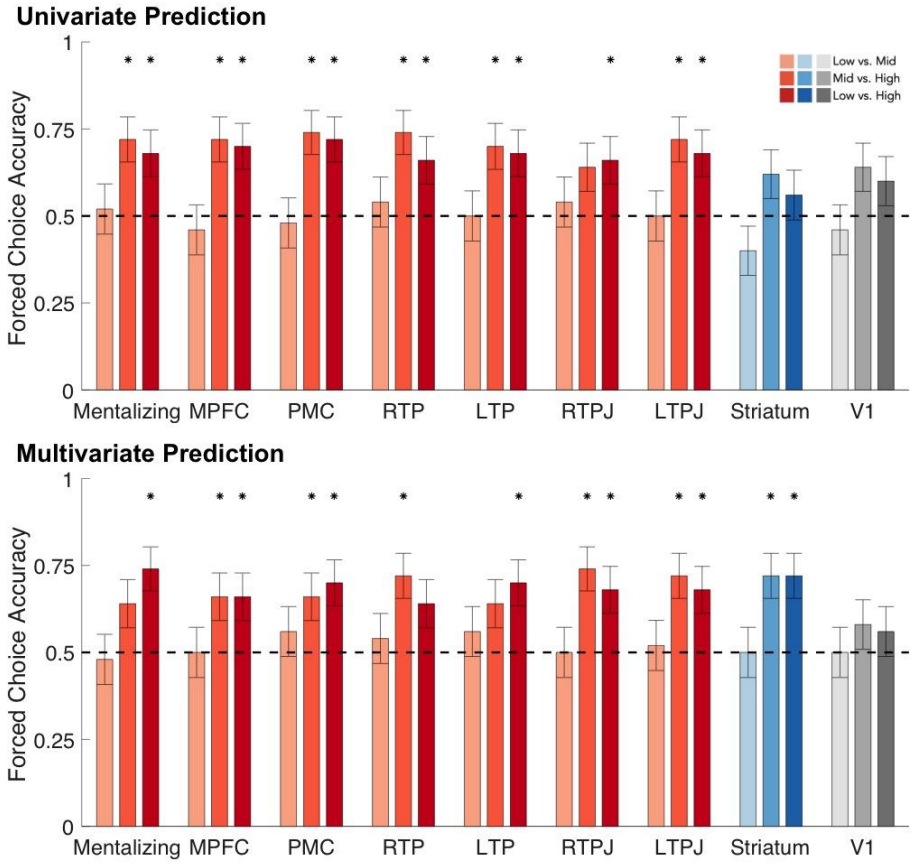
One alternative account for our results is that participants' brains might be picking up on the attractiveness of the faces, as attractive individuals might draw more social connections than less attractive ones. To rule out this possibility, we repeated all analyses while controlling for the subjective attractiveness of the faces (in addition to the original covariate of personal nominations and personal closeness). For the univariate parametric analyses (http://neurovault.org/collections/2715/), the results are very similar when controlling for attractiveness: significant clusters appear in regions related to mentalizing (i.e., medial prefrontal

cortex, temporal poles) and value processing (i.e., striatum). We also repeated the univariate and multivariate prediction analyses while adding the additional covariate of attractiveness. Forced classification accuracy rates (Fig. S10) also remained highly similar to our original results. Furthermore, the attractiveness ratings were only weakly correlated with idiosyncratic hub index (average within-subject correlation = .14), suggesting that attractiveness is not a confound.

Effect of Scan Date on Neural Activity

We ran additional analyses to test if scan date was associated with forced-choice classification accuracy for hub index. We did not find evidence that scan date impacted the ability to detect hubs. For each participant, we calculated the number of days between the social network nominations during Week 2 of the quarter and their fMRI scan. We then conducted binary logistic regression analyses with number of days as a predictor of forced-choice classification accuracy for (1) middle vs. low hub index, (2) high vs. middle hub index, and (3) high vs. low hub index. Results indicated that there were no significant associations for any of these models: $\chi^2(1) = .04$, $p = .84$ for Model 1; $\chi^2(1) = .00$, $p = .99$ for Model 2; and $\chi^2(1) = .9$, $p = .34$ for Model 3.

Figure S12. Forced-choice classification results controlling for face attractiveness, personal nominations, and subjective closeness of the participant. In both univariate and multivariate prediction analyses, accuracy was above chance when classifying between high vs. medium and high vs. low social support individuals



Individual Differences in Hub Detection

In two additional analyses, we investigated whether perceivers' characteristics –"hubness" and trait empathy – would impact their ability to detect hubs. However, neither of these characteristics was associated with enhanced hub detection in our data set. We ran binary logistic regression analyses with participants' hub index as a predictor of forced-choice classification accuracy for detecting targets with (1) middle vs. low hub index, (2) high vs. middle hub index, and (3) high vs. low hub index. Results indicated that there were no significant associations for

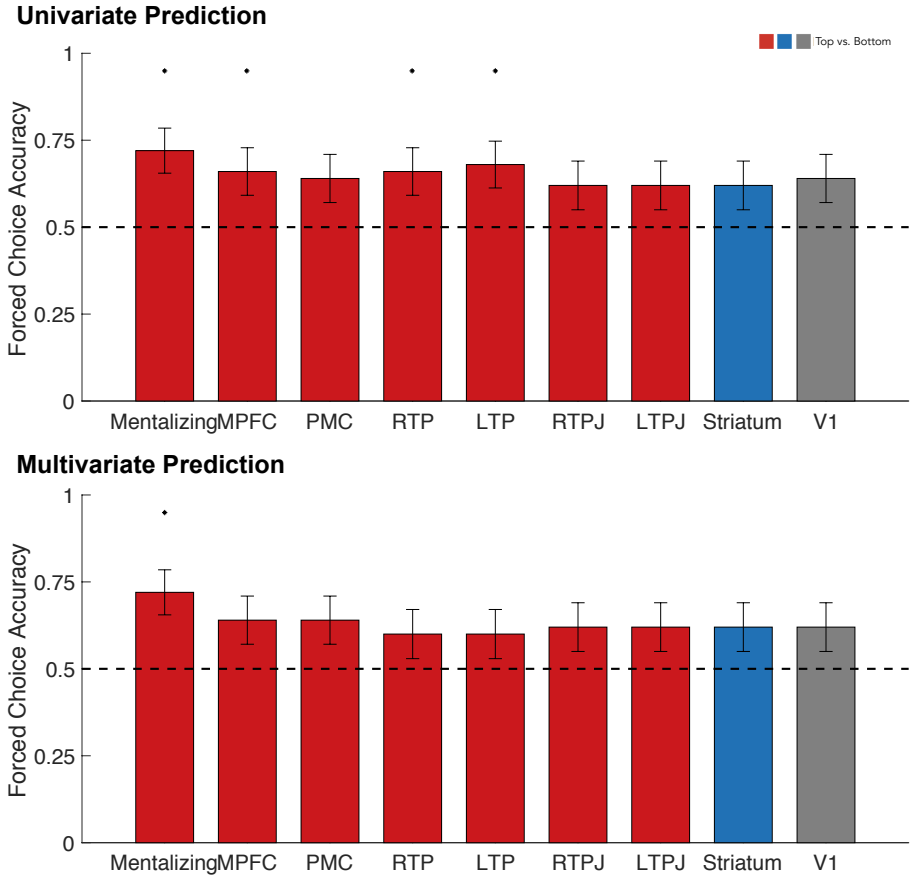
any of these models: $\chi^2(1) = 1.84$, $p = .18$ for Model 1; $\chi^2(1) = .2$, $p = .65$ for Model 2; and $\chi^2(1) = .81$, $p = .38$ for Model 3.

All participants in this sample completed a battery of 21 trait measures that broke down into four main trait clusters (i.e., empathy, life satisfaction, positive emotion, and negative emotion) [11]. The empathy factor included several social cognition measures (i.e., empathic concern, perspective-taking, positive empathy, prosociality, and agreeableness). Therefore, we conducted binary logistic regression analyses with the empathy factor as a predictor of forced-choice classification accuracy for (1) middle vs. low hub index, (2) high vs. middle hub index, and (3) high vs. low hub index. There were no significant associations for any of these models: $\chi^2(1) = 2.86$, $p = .09$ for Model 1; $\chi^2(1) = .4$, $p = .35$ for Model 2; and $\chi^2(1) = 1.23$, $p = .27$ for Model 3.

Forced-Choice Classification Results with Two and Four Hub Categories

When we binned the data into two bins, forced-choice classification accuracy based on mean ROI activity was above chance in the mentalizing network, MPFC, RTP, and LTP, but not different from chance in the other ROIs (Fig. S11). Classification based on multivariate activity was above chance in the mentalizing network but not in the other ROIs. Overall, these results suggest that we can distinguish between the top and bottom half of participants based on hub index, though the reliability across ROIs was diminished when predicting based on multivariate activity. One reason for the weaker results could be that by dividing the into two bins, the top bin would also include individuals with a medium social hub index, which we had shown in our original analysis to be indistinguishable from those with low social hub index.

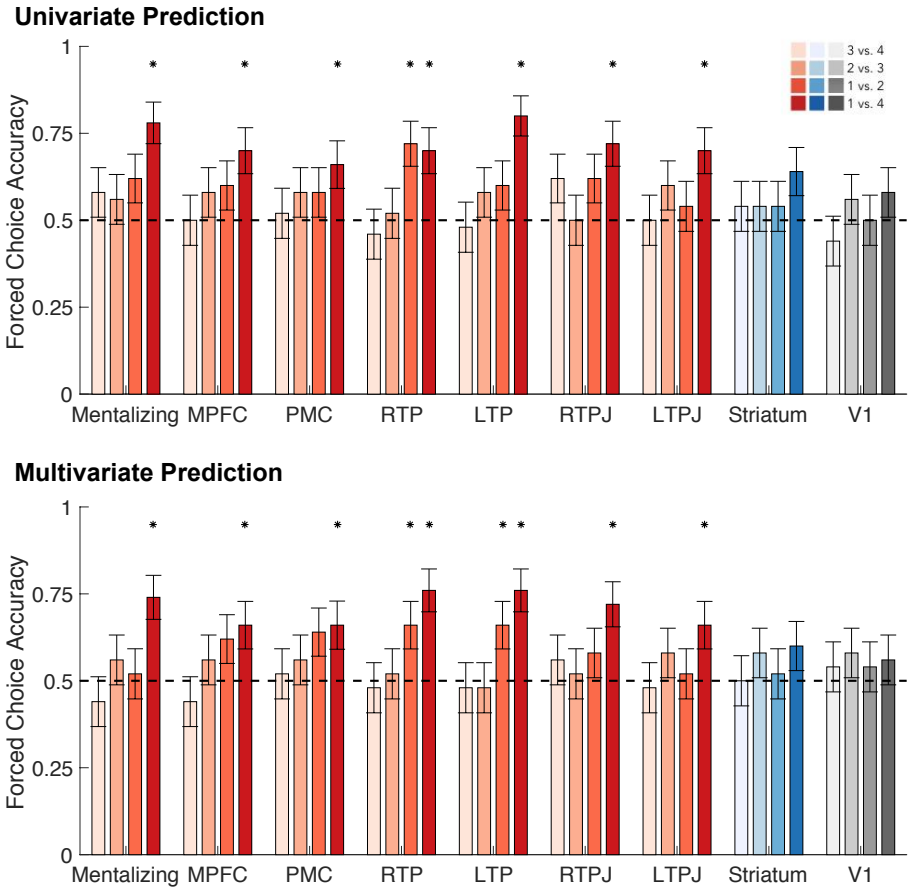
Figure S13. Forced-choice classification results when dividing targets into two hub categories. Classification based on mean ROI activity was above chance in the mentalizing network, MPFC, RTP and LTP, but not different from chance in the other ROIs. Classification based on multivariate activity was above chance in the mentalizing network.



When we binned the data into four bins, we could reliably distinguish between the top and bottom quartiles from both mean and multivariate activity in all ROIs in the mentalizing network (i.e., 1 vs 4; Fig. S12). In line with our previous findings, this suggests that hubs are closely tracked and can be distinguished from the least supportive individuals. In the bilateral temporal poles, we could accurately classify between the top quartile and the second quartile. In all other cases, forced-choice classification was at chance when comparing the top quartile to the middle quartiles (i.e., 1 vs. 2 and 1 vs. 3), as well as between consecutive quartiles (i.e. 1 vs. 2, 2

vs. 3, 3 vs. 4). These results suggest that the number of individuals in each bin might be inadequate for us to cleanly resolve between different quartiles.

Figure S14. Forced-choice classification results when dividing targets into four hub categories. In the mentalizing network, forced choice classification performance was better than chance when distinguishing between the top quartile (quartile 1) and the bottom quartile (quartile). In the bilateral temporal poles (RTP, LTP), forced choice classification performance was also better than chance when distinguishing between the top and second quartile. Forced-choice classification performance was at chance in all other cases and in all other ROIs.



Taken together, these additional analyses suggest that a median split and quartile splits divide the data into bins that are too coarse or too low in power, respectively. We believe this stems from the fact that brain responses to targets are non-linear—with sensitivity favoring high hub index targets—which is not well represented in the median split, and that a quartile split

decreases our power to detect relationships between brain activity and hub index. In contrast, dividing it into three bins allowed us to identify a top tercile containing individuals that evoked a distinct neural response than the other two terciles.

Within-Subject Prediction of Hub Category

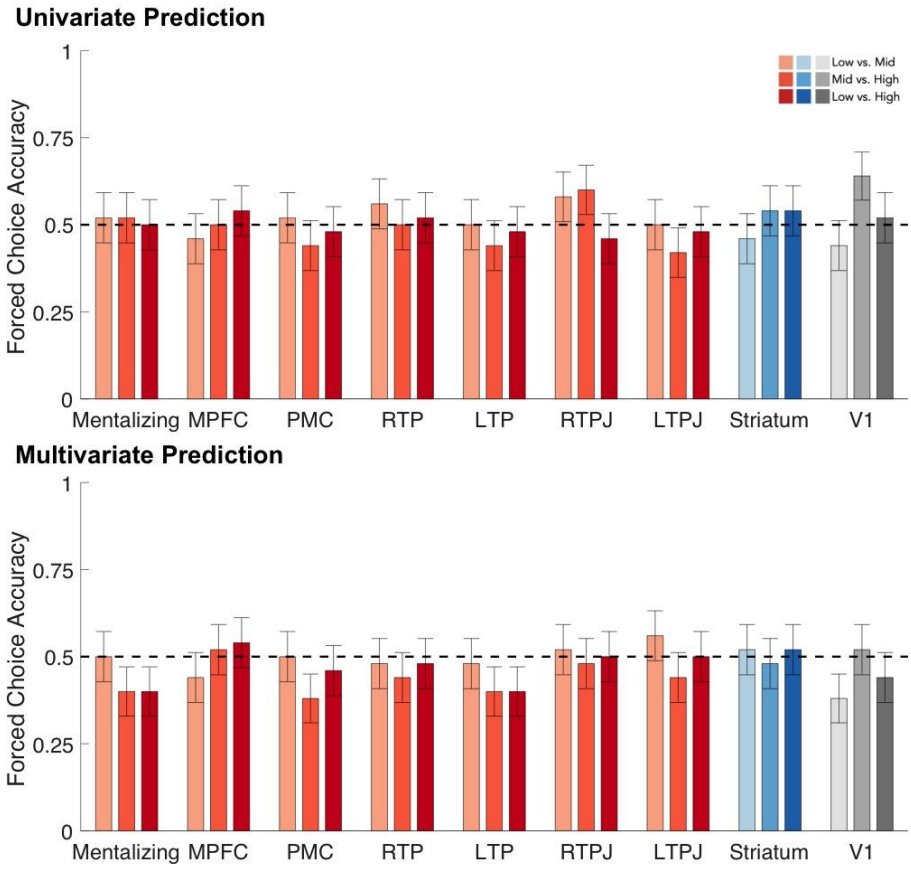
We randomly assigned the faces presented to each participant to one of two equally sized groups (N = 15 for each group). We trained the prediction algorithm on one group, and tested it on another group. Forced-choice classification accuracy was then averaged across subjects to assess prediction performance. Accuracy was at chance both when prediction was performed over average ROI activity and when prediction was performed over multivariate activity (Fig. S13). Given that each subject saw only 30 faces, it is likely that data from individual subjects are insufficient to train an accurate prediction algorithm. In our study, we were primarily interested in investigating whether there were between-subject similarities in how hubs are represented, and thus prioritized scanning *more subjects* rather than *more data per subject*.

Trait Characteristics of Social Value Hubs

To explore whether dorm members identify individuals as hubs on the basis of characteristics *specifically* related to social value as operationalized here (e.g., social support and prosociality), we used personality data collected from on each dorm member as part of a prior study [11]. All participants in this sample completed a battery of 21 trait measures. A factor analysis of these measures revealed four main trait clusters: (i) empathy, (ii) life satisfaction, (iii) positive emotion, and (iv) negative emotion (see Fig. 1 in Morelli et al., 2017). We used these four factors as simultaneous predictors of hub index in a multiple regression analysis. Empathy emerged as the only significant predictor of hub index ($\beta = .294$, $p = .006$) and predicted individuals' hub status above and beyond the effects of life satisfaction ($\beta = .177$, $p = .106$), positive emotion ($\beta = .004$, $p = .97$), and negative emotion ($\beta = .042$, $p = .697$). This suggests

that hub status reflects characteristics specific to empathy and social support, and not more general halo effects.

Figure S15. Within-subject prediction results. Forced-choice classification accuracy was at chance across all ROIs both when prediction was performed using average ROI activity (univariate prediction) and when prediction was performed using multivariate patterns (multivariate prediction).



Supplemental References

1. Jackson, M.O., *Social and economic networks*. 2008, Princeton, NJ: Princeton University Press.
2. Ledesma, R.D. and P. Valero-Mora, *Determining the number of factors to retain in EFA: An easy-to-use computer program for carrying out parallel analysis.* Practical Assessment, Research & Evaluation, 2007. **12**(2): p. 1-11.
3. Hu, L.t. and P.M. Bentler, *Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives.* Structural Equation Modeling: A Multidisciplinary Journal, 1999. **6**(1): p. 1-55.
4. Brainard, D.H., *The psychophysics toolbox.* Spatial Vision, 1997. **10**(4): p. 433-436.
5. Knutson, B., et al., *Anticipation of increasing monetary reward selectively recruits nucleus accumbens.* Journal of Neuroscience, 2001. **21**(16): p. RC159.
6. Fareri, D.S., et al., *Social network modulation of reward-related signals.* The Journal of Neuroscience, 2012. **32**(26): p. 9045-9052.
7. Mobbs, D., et al., *A key role for similarity in vicarious reward.* Science, 2009. **324**(5929): p. 900.
8. Preuschoff, K., P. Bossaerts, and S.R. Quartz, *Neural differentiation of expected reward and risk in human subcortical structures.* Neuron, 2006. **51**(3): p. 381-390.
9. Brett, M., et al., *Region of interest analysis using the MarsBar toolbox for SPM 99.* Neuroimage, 2002. **16**(2): p. S497.
10. Craddock, R.C., et al., *A whole brain fMRI atlas generated via spatially constrained spectral clustering.* Human Brain Mapping, 2012. **33**(8): p. 1914-1928.
11. Morelli, S.A., et al., *Empathy and well-being correlate with centrality in different social networks.* Proceedings of the National Academy of Sciences, 2017. **114**(37): p. 9843-9847.