# Coupling social attention to the self forms a network for personal significance

Jie Sui[a,b,c,1], Pia Rotshtein[d], and Glyn W. Humphreys[a]

[a]Department of Experimental Psychology, Oxford University, Oxford OX1 3UD, United Kingdom; [b]Department of Psychology, and [c]Department of Biomedical Engineering, Center for Biomedical Imaging Research, School of Medicine, Tsinghua University, Beijing 100084, China; and [d]School of Psychology, University of Birmingham, Birmingham B15 2TT, United Kingdom

Prior social psychological studies show that newly assigned personal significance can modulate high-level cognitive processes, e.g., memory and social evaluation, with self- and other-related information processed in dissociated prefrontal structure: ventral vs. dorsal, respectively. Here, we demonstrate the impact of personal significance on perception and show the neural network that supports this effect. We used an associative learning procedure in which we "tag" a neutral shape with a self-relevant label. Participants were instructed to associate three neutral shapes with labels for themselves, their best friend, or an unfamiliar other. Functional magnetic resonance imaging data were acquired while participants judged whether the shape-label pairs were maintained or swapped. Behaviorally, participants rapidly tagged a neutral stimulus with self-relevance, showing a robust advantage for self-tagged stimuli. Self-tagging responses were associated with enhanced activity in brain regions linked to self-representation [the ventral medial prefrontal cortex (vmPFC)] and to sensory-driven regions associated with social attention [the left posterior superior temporal sulcus (LpSTS)]. In contrast, associations formed with other people recruited a dorsal frontoparietal control network, with the two networks being inversely correlated. Responses in the vmPFC and LpSTS predicted behavioral self-bias effects. Effective connectivity analyses showed that the vmPFC and the LpSTS were functionally coupled, with the strength of coupling associated with behavioral self-biases. The data show that assignment of personal social significance affects perceptual matching by coupling internal self-representations to brain regions modulating attentional responses to external stimuli.

Humans have the inherent ability to rapidly learn the social salience of a stimulus enhancing survival. There are a considerable number of studies on the effect of self-association in social psychology that have shown that there is enhanced importance assigned to self-associated objects (1), increased preference (2, 3), and stronger memory (4, 5). For example, by assigning participants to a specific team associated with specific symbols, participants typically rapidly orient their attention and prioritize the subsequent processing toward self-associated team members when asked to make social evaluations and allocate rewards (2, 3). These effects are not confined to high-level cognitive processes, however, Sui et al. recently demonstrated that self-associations with neutral geometrical shapes can rapidly alter perception (6), so that self-associated shapes are less affected by contrast reduction than shapes associated to other people. How this rapid perceptual effect of self-tagging emerges was investigated here.

In the past decade, there has been an increased interest in the neural mechanism that support self-related processing (7–14) and, in parallel, research that focuses on the way we process information about others (15–19). The work has revealed two complementary neural networks—a ventral midline network, including the ventral medial prefrontal cortex (vmPFC) engaged in self-processing, and a dorsal prefrontal (dPFC) network supporting processes related to others (20, 21). An additional neuroanatomical distinction is suggested between internal-focused self-processing (self-reflection and retrieval of self-knowledge) associated with the vmPFC, and externally focused social processing (attention and inferencing about others) associated with regions within the inferior parietal cortex, the superior temporal sulcus and the temporoparietal junction (Theory of Mind, social attention; 17–19, 22–24).

Self- and social processing, however, is not limited to the assessment of referential attributes, introspection, and inferences on mental states. An additional intriguing aspect of "the social self" is the reliance on ownership relations with the environment (25) and the flexibility by which the self "expands" within the environment through acquired associations (26, 27). Cunningham et al. demonstrated that randomly assigned shopping products as belonging to the self ("mine") versus to another (e.g., "Alex's") increased both the ability of participants to remember and the preferential value of products link with the self (26). It has been shown that this memory enhancement effect is mediated by the medial prefrontal cortex (mPFC) (28). Sui et al. have further shown that the self can also expand to simple geometric shapes (e.g., a circle) that have no real ecological value in the environment. This self-expansion affected the perceptual processing of the stimuli consistent with the self-tagged stimuli being perceptually more salient (6). The processing of salient information in the environment is suggested to involve the bilateral ventral frontoparietal attentional network (29, 30), with the left hemisphere (e.g., the temporoparietal junction and the frontal eye fields) specifically involved in processing salient information based on its potential behavioral relevance (31). Interestingly the posterior part of the ventral attentional network partly overlaps with regions that process social cues (refs. 24 and 32; also see refs. 17, 18, and 33). It has also been hypothesized that the attentional response in this latter region is tightly linked to the development of sensitivity to social cues and to joint attention in infancy (34). Therefore, it is possible that the association of a neutral shape to the self alters its saliency by modulating activity in the ventral network for social attention, generating a form of "social saliency." If so, we predicted that posterior regions within the ventral attentional network would be linked to the perceptual facilitation effect observed for self-tagged stimuli (6). In particular, we hypothesized that tagging a new neutral stimulus to the self would couple the posterior network supporting social attention to external stimuli to the ventral prefrontal network linked to internal representation of the self, to assign the stimuli personal significance.

Participants were instructed to form mental associations between three shapes (circle, triangle, and square) and three people, one with themselves, one with their best friend, and the third with a stranger. Immediately after this instruction, they carried out a matching task in a 3-T brain scanner, where they were required to indicate whether a shape-label pairing was correct, based on their previously learned

PSYCHOLOGICAL AND COGNITIVE SCIENCES

association (e.g., triangle-stranger). A unique aspect of our design is that it enabled us to examine the neural regions responsive to self-labels (triggering internal self-reflection), to self-tagged shapes (external self-related stimuli), and to their combination, and we used model-based analysis to test for functional connectivity changes based on self-association. To validate the observed results, we used individual differences and tested for correlations between brain responses (neural response strength and the coupling strength between brain regions) and the observed behavioral bias to the self. We hypothesized that the self-related label (You) may automatically activate internal self-reflection (22) associated with the engagement of cortical midline structures including the vmPFC (7, 8, 20). We also anticipated that responses within regions of the inferior parietal cortex/superior temporal sulcus would signify the presence of salient external stimuli associated with the self (the self label and the newly associated self shape). To change the personal social significance of the shape, we hypothesized that there would be modulation of functional connectivity between the vmPFC and ventral attentional regions.

We observed that the vmPFC and the left posterior superior temporal sulcus (LpSTS) responded to self-tagged shapes paired with self-related labels. Further, by examining the shape-label mismatching pairs, we were able to dissociate the functions of the two regions, showing that the vmPFC responded to the self-related label, whereas the LpSTS responded to the self-related shape along with the self-related label. Furthermore, using dynamic causal modeling (DCM), we showed that the projection from the vmPFC to LpSTS increased for self-associations but decreased for other associations. Responses of these regions predicted the extent of the self-advantage in behavior. The results suggest that perceptual effects of a new personal association rely on a neural network in which internally focused self-reflection is linked to external stimuli to enhance their social salience.

## Results

**Self Advantage in Behavior.** Repeated measures ANOVA for reaction times (RTs; Fig. 1*A*) showed a significant effect of matched associations, with faster responses associated with increasing proximity to the self: self < friend < stranger [$F_{(2, 30)} = 54.12$, $P < 0.001$, $\eta^2 = 0.78$]; pairwise comparisons showed reliable differences in all comparisons (self vs. friend, friend vs. stranger, self vs. stranger, $P < 0.001$). The analysis for accuracy also showed higher accuracy for self > friend > stranger, $F_{(2, 30)} = 22.87$, $P < 0.001$, $\eta^2 = 0.60$; all pairwise comparisons were reliably different, $P < 0.02$. In line with prior evidence (6), participants demonstrated a robust advantage for newly learnt self-associated stimuli in the matching task.

To assess the potentially distinct roles of the self-relevant label and shape, a repeated-measures ANOVA was conducted on nonmatching trials (across the conditions: self-label, self-shape, friend-stranger) (Fig. 1*B*). For the RT data, there was a significant effect of nonmatching pairs [$F_{(2, 30)} = 10.01$, $P < 0.001$, $\eta^2 = 0.40$]. Participants had faster responses when the self-label or self-shape stimuli appeared than when self-information was not present (i.e., friend-stranger, $P < 0.03$). There was also a marginal advantage for nonmatching pairs where the self-label was present compared with when the self-shape was present ($P = 0.05$). Accuracy did not vary across the nonmatching pairs ($P = 0.15$).

To examine the strength of self-advantage for matching vs. nonmatching trials, we compared responses to self-related stimuli when the shape and label matched and, on nonmatching pairs, when the self-label or -shape was present. There was a significant overall effect of condition for RTs [$F_{(2, 30)} = 29.09$, $P < 0.001$, $\eta^2 = 0.66$]; RTs were faster for shape-label matched pairs relative to nonmatching pairs ($P < 0.001$), and there was a borderline advantage for nonmatching self-label trials compared with non-matching trials with self-shape stimuli ($P = 0.05$). This effect was not observed for the accuracy data ($P = 0.62$).
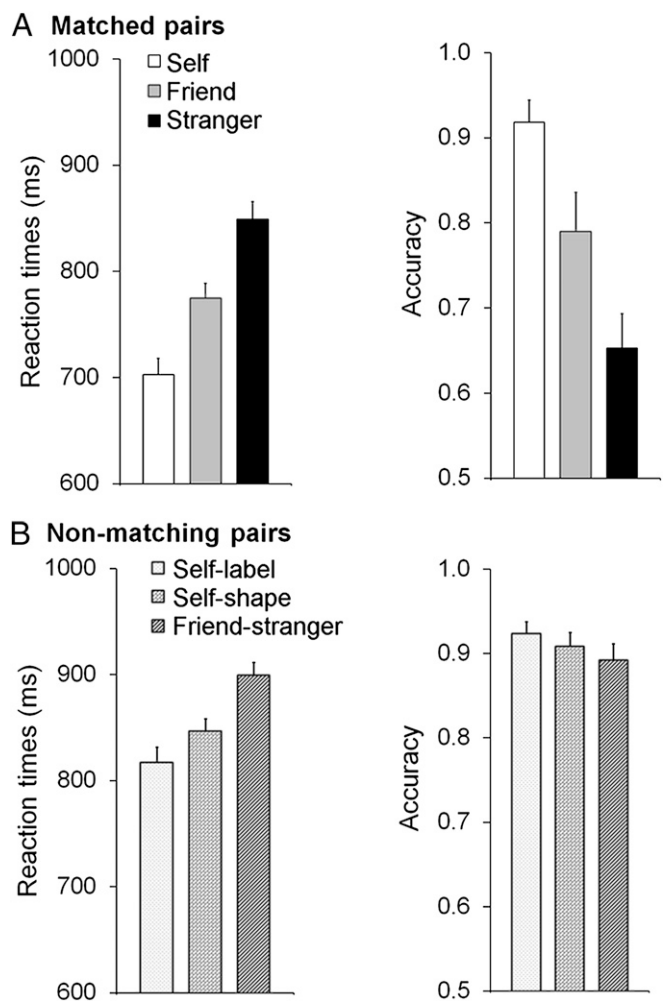


**Fig. 1.** Mean reaction times and accuracy as a function of a matched association (self, friend, or stranger) (*A*) and a nonmatching association (for self-label, self-shape, or friend-stranger pairs) (*B*). Error bars represent one SE.

**Involvement of the vmPFC and LpSTS in Creating New Personal Salience.** We next tested brain regions that showed either an increasing (self > friend > stranger) or a decreasing (self < friend < stranger) response for matching pairs depending on their proximity to the self. Relative to "stranger" stimuli, increased responses to self-associated stimuli were observed in the LpSTS and the vmPFC (Fig. 2*A* and Table S1); decreased responses for self-associations were seen in brain regions linked to executive functions, including the bilateral dorsal-lateral prefrontal cortex (DLPFC), the middle superior frontal gyrus, the bilateral inferior parietal cortex, and the right cuneus (Fig. 2*B* and Table S1). Modulation of responses in the LpSTS and vmPFC by the self-related stimuli correlated across participants ($r = 0.51$, $P < 0.05$), suggesting those two regions had synchronized responses when self-associated stimuli were matched relative to when stranger associations were matched (Fig. 2*C*). This effect was not observed in the DLPFC ($r = 0.32$ and $0.43$, $P = 0.23$ and $0.10$ for correlations with LpSTS and vmPFC, respectively). This result suggests that, in the context of the current task, the DLPFC was distinct from the LpSTS–vmPFC circuit.

To further validate our results, we assessed the relations between brain activity and behavior. Correlation analyses were conducted across participants by using self-stranger differences on matching trials, comparing the responses of each region and behavioral response efficiency [self-stranger, combining RTs, and
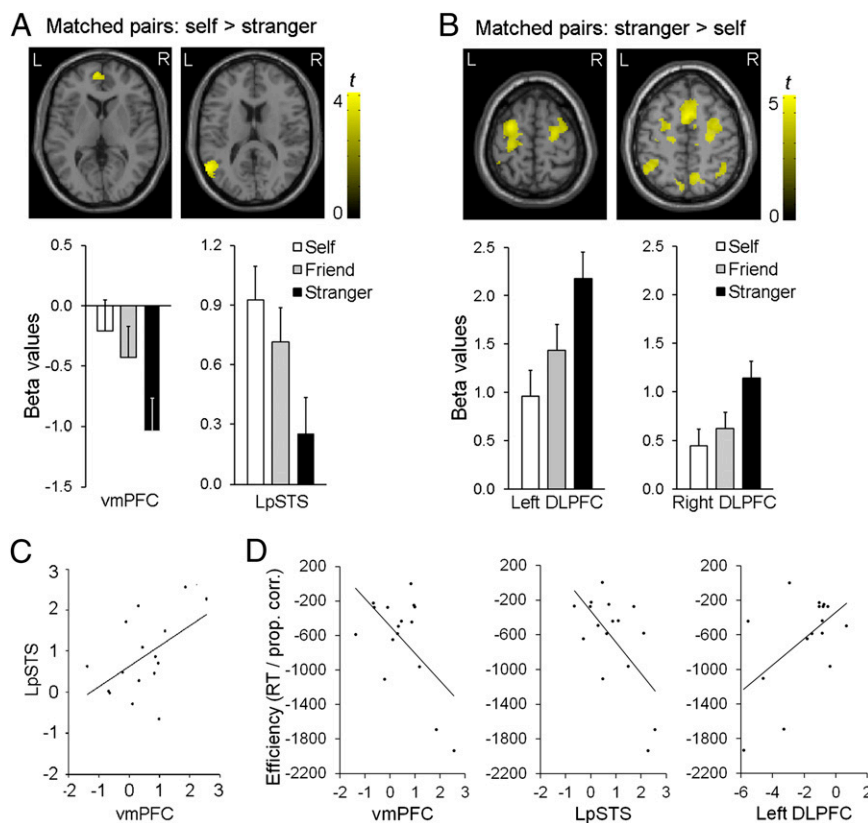
**Fig. 2.** Functional magnetic resonance imaging (fMRI) data for shape-label–matched association at $P < 0.005$ with an extent threshold of >70 voxels. (*A*) Self > stranger in the vmPFC and LpSTS, and beta values of brain responses in the vmPFC and LpSTS. Error bars represent one SE. (*B*) Stranger > self in the dorsal attentional network, and beta values of brain responses in the bilateral DLPFC. Error bars represent one SE. (*C*) Correlations between beta values of the vmPFC and LpSTS for self > stranger comparison. (*D*) Correlations between responses efficiency (RT/proportion correct) and beta values extracted from the vmPFC, LpSTS, and DLPFC using self-stranger differences from matching trials.

accuracy data to create a single measure: RT/Proportion (Prop.) correct; ref. 35]. The differential BOLD responses between the self- and stranger-related stimuli for each region predicted the behavioral advantage for self-related stimuli across participants (vmPFC: $r = -0.58$, $P < 0.02$, LpSTS: $r = -0.63$, $P < 0.01$, DLPFC: $r = 0.55$, $P < 0.03$). Gains in behavioral efficiency for the self (lower RT/Prop. correct scores) were associated with greater activity in the vmPFC and LpSTS and decreased activity in the left DLPFC (Fig. 2*D*).

To examine whether responses in the vmPFC and LpSTS were driven by the self-label (You) or the self-related shape, we assessed neural responses to the shape-label on mismatching trials. Increased activity in the vmPFC was exclusively associated with the presence of the self-label on nonmatching trials, whereas enhanced activity in the LpSTS was linked to the self-shape (Fig. S1 and Table S1) as well as the self-label. The results demonstrate a critical role for the vmPFC and LpSTS in self-association. To further assess the functions of these two regions, we conducted a repeated-measures ANOVA for the region (vmPFC vs. LpSTS) and the self-relevance condition (self-label vs. self-shape) for mismatched trials. No significant interaction was observed ($P = 0.33$). This result is consistent with effects of the self-labeling occurring in both regions, through effects of the shape were found only for the LpSTS. The functional connectivity critical to the operation of the regions was subsequently examined by using DCM analysis (see below). The correlation analyses additionally suggested that these two regions form a network that processes self-related information, distinct from the DLPFC. To evaluate this network in more detail, we used DCM analyses to determine the flow of information between the two regions within each participant.

**Coupling Strength Between the vmPFC and the LpSTS Predicts the Self-Advantage in Behavior.** A random effect analysis through Bayesian family inference was used to establish how stimulus input entered the network. Three family models were considered: input entering through the LpSTS, through the vmPFC, or through both two regions. The comparison revealed a winning family of models with the driving visual input coming into the vmPFC (Fig. 3*A*). The results are consistent with the evidence that the vmPFC responses to visual information are rapid (~130 ms) and precede responses in visual associative cortices (36).

Next we used Bayesian model averaging, including only models from the winning family (visual input to the vmPFC) to assess the strength of the intrinsic and effective modulator factors for each subject (37). There was positive intrinsic coupling between the vmPFC and LpSTS: vmPFC to LpSTS, $t_{(15)} = 2.98$, $P < 0.01$; LpSTS to vmPFC, $t_{(15)} = 2.70$, $P < 0.02$. Intrinsic connections within each region were inhibitory: LpSTS to LpSTS, $t_{(15)} = -881.41$, $P < 0.001$; vmPFC to vmPFC, $t_{(15)} = -400.28$, $P < 0.001$ (Fig. 3*B*). These connections varied between the conditions. The coupling from the vmPFC to LpSTS increased for matching self-pairs compared with nonmatching self-label pairs, $t_{(15)} = 2.53$, $P < 0.03$. There was also a decreased projection from vmPFC to LpSTS for nonmatching self-shape pairs (paired with the stranger label) and nonmatching self-label pairs (paired with the stranger shape), relative to when there were matching stranger shape-label pairs, $t_{(15)} = -2.25$ and $-3.01$, $P = 0.04$ and $0.009$. Projections from vmPFC to LpSTS increased when the self label was paired with the self shape; in contrast, pairing the self-shape or -label with a stranger label/shape decreased coupling, presumably indicating a mismatched self-association.
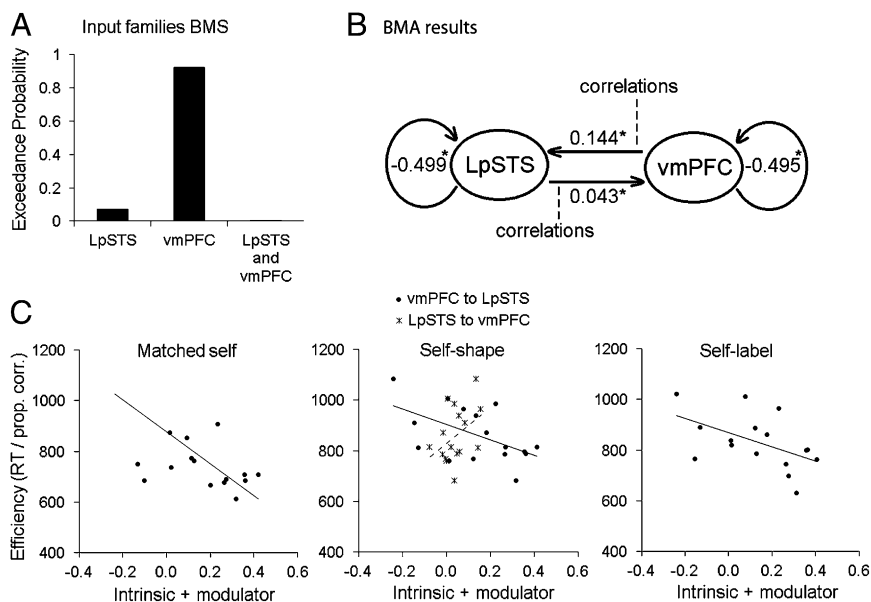
**Fig. 3.** (A) Exceedance probabilities from Bayesian model selection (BMS) procedure of 1,851 models dividing into three families based on the visual inputs to LpSTS, vmPFC, or both LpSTS and vmPFC. The family of models with input to the vmPFC had the greatest evidence. (B) Structure of the Bayesian model average (BMA) for the vmPFC family of models. Intrinsic parameter values are illustrated. Significant parameters are indicated by an asterisk. (C) Correlations between the values of intrinsic plus modulator and the efficiency of behavioral responses.

Finally, to validate the DCM results, we computed correlations across participants between connection strengths and behavioral performance (Fig. 3C). Connection strengths were estimated as the sum of the intrinsic and condition modulator effects. For matched self-pairs, there was a significant correlation between behavioral efficiency and connection strength from vmPFC to LpSTS, $r = -0.60, P < 0.02$. The stronger the effective connection from vmPFC to LpSTS, the more efficient (faster and more accurate) were behavioral responses for self-shape matching trials. Increased connection strength from vmPFC to LpSTS also correlated with more efficient behavior for the self-shape condition (when paired with the friend/stranger label), $r = -0.55, P < 0.03$, and the self-label condition (when paired with the friend/stranger shape), $r = -0.51, P < 0.05$. However, increased connectivity from LpSTS to vmPFC was associated with less efficient behavior, $r = 0.54, P < 0.04$ for the self-shape condition (when paired with a nonmatching label). This last result would follow if shape-driven activation of the self, from LpSTS to vmPFC, makes it more difficult to reject a nonmatching pair of stimuli.

These data suggest that the vmPFC received visual input before the LpSTS. Coupling between the vmPFC and LpSTS was then boosted when matched pairs were presented relative to nonmatching pairs. The strength of these projections predicted the self-prioritization effects observed in behavior.

## Discussion

Prior work indicates that the vmPFC is activated by self-related stimuli and by self-reflection (7, 8, 20, 38, 39). The LpSTS, however, is part of a ventral attentional network triggered by salient, contextual stimuli (31) that has been linked to the processing of social cues that help compute the mental states of others (17, 24, 40, 41), whereas the right posterior superior temporal sulcus/temporoparietal junction has been more strongly associated with theory of mind tasks (17, 18). Although it is not surprising that there is involvement of the vmPFC in the current study, given its prior links to self-based processing, we show that, of other brain regions activated in social cognition tasks, only the LpSTS is differentially recruited by self-association to neutral shapes. Our work indicates that the LpSTS was reliably linked to the simple perceptual matching of new

personal associations. Most notably, the effects of personal social significance on perception were based on the coupling of self-related responses in the vmPFC to the LpSTS. Consistent with this result, the analysis of the nonmatching trials indicated that the LpSTS was activated by the self-associated shape as well as the self label, whereas the vmPFC was activated only by the self-label. The presence of both the shape and the label linked to the self increased coupling between these brain regions, enhancing the behavioral response to the self. However, the presence of a self-associated stimulus in a nonmatching pair decreased the effective connectivity between the LpSTS and the vmPFC, presumably to enable a nonmatching response to be made. We suggest that the coupling between the vmPFC and LpSTS reflects a neural network that registers the social saliency of an external stimulus, reflecting its personal significance to the observer.

In contrast to the strong triggering of the social saliency network by self-associated stimuli, new associations between shapes and the label for a stranger activated a dorsal frontoparietal control network. The top-down attentional network is sensitive to new task demands (here, the formation of a novel association) (42). We assume that recruitment of the frontoparietal control network provides additional resources to enable the new association to be established and linked to an appropriate response. This result is consistent with Mitchell's study where they found that the vmPFC was associated with self-processing, while the dorsal part of mPFC was engaged when making judgments related to others who were dissimilar to self (20). The DLPFC activity found here, however, was distinct from both and more consistent with the recruitment of the dorsal attentional network (29), required to support performance in the more difficult task.

Interestingly, apart from being implicated in processing information related to the self, the vmPFC is linked to associative learning (43, 44). Behrens et al. demonstrated that two neighboring divisions of the anterior cingulate cortex were implicated in learning about social and reward-based information (43). However, when making a decision, the information learned by using these parallel streams was combined within vmPFC. Thus, we suggest that the vmPFC encodes information that is of high relevance to the individual, either through direct "self-tagging" of

a neutral stimulus to the self as implemented here, or by associating an unconditioned stimulus to a potential reward. Our connectivity analyses further demonstrate that information that becomes relevant because of its relation to the self, gains attentional saliency through the projections of the vmPFC upon the LpSTS. The results support the hypothesis that the projection from the vmPFC to LpSTS was key to forming a newly personal association by linking neutral external stimuli (processed via the LpSTS) with internal-focused self-reflection (triggered through the vmPFC).

Prior studies of self-prioritization have contrasted responses to participants' own names and faces, which both have an a priori association to the self and which also differ in familiarity relative to the stimuli they are typically contrasted with (45, 46). These studies contrast with our examination of newly learned associations with simple, neutral stimuli. There are other studies showing that self-association leads to privileged processing for neutral items in memory. For example, researchers have reported that subjects had better recall performance for the nouns associating with self-generated names than for nouns linked to names generated by others (47). Others have also shown that, after decision about self- or other-ownership during stimulus encoding, subjects showed better memory for self-related objects compared with those related to other people (26, 27). Our results extend the self-association benefit to the perceptual domain. A particular advantage of this study is that the neutral geometric shapes we used are easy to manipulate to investigate the nature of self-association at perceptual and attentional levels. The current experiment thus represents a methodological advance because it examines the rapid learning of a new self-association in perception, and it also uses neutral shape stimuli equated for familiarity and counterbalanced across participants. Through this methodological advance we have been able to establish how perceptual, personal significance is established at the level of neural circuits.

In conclusion, humans rapidly tag neutral stimuli with personal association by coupling self-representations to brain regions modulating sensory-driven attentional control. The current work demonstrates that the paradigm of self-tagging provides a powerful approach to enhance the efficiency of human behavior, and that connection strengths within the neural circuit underpinning stimulus–self association predict the magnitude of self-bias in perception.

## Materials and Methods

**Participants.** Sixteen healthy volunteers (three males, aged between 18–35 y, Mean = 22.25 ± 5.99) participated in the experiment. All participants were right handed and had normal or corrected-to-normal vision. Informed consent was obtained from all participants before the experiment according to procedures approved by the Institutional Review Board of Beijing Normal University, China.

**Stimuli.** Three geometrical shapes (triangle, square, and circle, each of 4.0° × 4.0°) and three words: "You," "Friend," and "Stranger" (3.1°/3.6° × 1.6°) were used as stimuli. The shape and label were presented above and below the fixation cross (0.8° × 0.8°), respectively. The distance between the center of the shape/label and fixation was 3.7°/2.7°. The stimuli were white and appeared on a gray background. E-Prime1.1 software (Psychology Software Tools) was used to present the stimuli and to record behavioral responses.

**Procedure.** Participants were instructed to imagine an association between a particular geometric shape (triangle, square, and circle) and, respectively, themselves, a personally familiar other (a named best friend), or an unfamiliar stranger, while the stimuli were not present (for details; ref. 6). The shapes were counterbalanced across participants. After this instruction, participants immediately performed a matching task in a brain scanner in which they had to judge whether paired shapes (triangle, square, or circle) and labels (You, Friend, or Stranger) either matched or mismatched. Each trial started with a central fixation cross for 400 ms, followed by a pairing of a shape and a label for 100 ms. There was then a blank interval for 1,100 ms, in which participants were expected to judge whether the shape was correctly assigned to the person by pressing one of the two response buttons as

quickly and accurately as possible. Feedback was given on the screen for 500 ms at the end of each trial. The pairing of the shape and the label was equally distributed, such that each shape was equally likely to appear with each word. Temporal jittering was introduced through the inclusion of 30% null events when the fixation stimulus was presented for 2.1 s. These events were randomly presented throughout the experiment but could occur on up to three consecutive presentations. The experiment was divided into four fMRI sessions. There were 169 volumes in each session comprising 13 trials of each pair (9 types of pairs: 3 shapes × 3 labels) and 52 null trials.

The behavioral data were analyzed by using repeated measures ANOVAs for matched associations (self, friend, or stranger) and then for nonmatching pairs (self-label, self-shape, or friend-stranger), respectively.

**fMRI Data Acquisition.** We used a Trio Siemens 3.0-T MRI scanner to acquire T2-weighted echo planar images (EPI) blood oxygenated level-dependent contrast. Thirty-nine slices were acquired with 2-mm thickness and 1-mm gap, with a plane resolution of 2.5 × 2.5 mm. We used 90° flip angle, 35-ms echo time and 2,300-ms slice repetition time. Images were acquired by using an eight-channel phase array coil with a sense factor of 2. The slices covered most of the brain including the entire temporal cortex, but excluding the most inferior parts of the cerebellum.

**Data Analysis.** The data were analyzed by using SPM8 (Wellcome Department of Imaging Neuroscience, London, United Kingdom).
*Preprocessing.* EPI volumes were spatially realigned and unwrapped to correct for interactions between movement artifacts and field inhomogeneities, transformed to the Montreal Neurological Institute (MNI) standard space (48), and resampled to a resolution of 2 × 2 × 2 mm. After this step, the data were smoothed by using a resolution of 8-mm Gaussian kernel to account for residual intersubject differences.
*Voxel-based statistical analysis.* Voxel-based analysis using a general linear model was performed in two steps. First, for each participant we computed the averaged estimated response across the four sessions for each experimental condition. We modeled the onset of each trial in each of the nine shape–label pairings conditions. These regressors were convolved with the canonical hemodynamic response function. In addition, to correct for signal changes due to head movement, the six realignment parameters were also included. An additional set of harmonic regressors were used to account for any low-pass frequency variance within the data across time, with a cutoff of 1/128 Hz and regressors modeling each session effect.

In the second step, we tested for consistent effects across participants (a random-effects second-level analysis) by using a factorial design including nine shape–label pairs (three shapes × three labels). In the model, we did not assume independency or equal variance across the conditions. We used a mixture of peak height and cluster extent threshold (49). Following our hypotheses, we first conducted t-contrast analyses to assess brain regions associated with enhancing or decreasing activity to self relative to stranger associations for matching trials. Next, to determine whether responses to the matched self condition were driven by the self-related label or by the newly learned self-related shape, we analyzed the nonmatching pairs. The analyses of the nonmatching conditions focused on the regions responding to the self-matched condition by using an inclusive mask, although we also explored the responses to these two types of information across the whole brain. There were three conditions for nonmatching pairs—self-label, self-shape, and friend-stranger. The self-label condition measured the effect of a self-label (paired with either a friend- or stranger-associated shape); the self-shape condition measured the effect of a self-associated shape (paired with either a friend or stranger label); the friend-stranger condition provided a baseline and reflected the case when friend and stranger labels and shapes were linked together. Paired t tests were performed for self-label or self-shape conditions vs. the friend-stranger baseline.

To ensure that the results were not driven by differences in the error and learning rates across the conditions, we computed a new first- (within participants) and second- (across participants) level analysis. In the first-level analysis, we modeled each of the nine experimental conditions but importantly only included the correct responses with the error trials modeled in a separate regressor. We further included a parametric modulation to capture learning effects across time. Individuals' estimated effect sizes from these models were entered into a second-level analysis. This new model produced similar results to the one reported above, suggesting that the observed pattern of results was driven by the shape–label content rather than by the number of errors or by differences in learning rate between conditions. To measure the effect of sex bias on the results, we also conducted the second-level analyses excluding three male participants. The data showed a similar pattern to that reported here.

We reported results based on the model not splitting trials based on accuracy, showing the amplitude of voxels surviving at $P < 0.005$ uncorrected across the whole brain and an extent threshold of 560 mm$^3$ (>70 voxels). It should be noted that we had a strong a priori hypothesis about the involvement of the vmPFC, given previous results strongly implicating the involvement of this region in self-evaluation.

To validate the observed self-effects on brain responses, we conducted correlation analyses between the behavioral response efficiency (RT/prop. corr.; ref. 35) and the estimated beta values extracted from the group peak response of three regions of interest (based on the above specified contrast). The correlation was based on the differential scores between the self and the stranger in the conditions where the shape and the label matched.

*DCM analysis.* A DCM analysis was undertaken to investigate how the interaction between the stimulus-driven attentional system and the self-network modulated self-biases across individuals. We focused on two regions—the LpSTS, which is part of the stimulus-driven attention network, and the vmPFC, which is part of the self network. To simplify the models, we included only four conditions: the matching self pairs, nonmatching trials where the self-shape was paired with another label, nonmatching trials with the self-label paired with another shape, and matching stranger trials. In all models, intrinsic connections between and within regions were always assumed. Across models, the selected four experimental conditions were systematically varied in a factorial way to account for context-dependent modulator effects on each of the four intrinsic connections (vmPFC → vmPFC, LpSTS → LpSTS, vmPFC → LpSTS, and LpSTS → vmPFC), for example, that each connection can be modulated by one, two, three, or four conditions differently combined and so forth. For each of the four connections,

we specified three families of model that differed in whether the driving visual input was assigned to: (*i*) the LpSTS, (*ii*) the vmPFC, or (*iii*) both regions. There was a total of 1,851 models for each subject. Because neither of the two regions (LpSTS and vmPFC) is a primary sensory area, we used a random effect analysis comparing across the three families of model to estimate the winning family, a function of the region in which the driving input enters the network. Bayesian model averaging was then used to assess the strength of the intrinsic and effective modulator connections across models within the wining family per subject (37). These parameter values were used in further statistical analyses in SPSS to assess the reliability of the effects across participants.

We first computed the liability of the intrinsic and modulator factors by using one-sample $t$ tests. Second, we examined whether the connection strength varied between conditions by using paired $t$ tests. Finally, to test whether changes in the connection strength between regions was directly related to the observed behavioral responses, we computed the correlation between (*i*) connection strengths between regions and (*ii*) the efficiency of behavioral responses for specific conditions. Again, because there was a consistent self-benefit on matching trials for both RTs and accuracy, a single efficiency measure was used as the behavioral index. Connection strength between the two regions was estimated as the intrinsic + condition modulator effect.

1. Kahneman D, Knetsch JL, Thaler RH (1990) Experimental tests of the endowment effect and the coase theorem. *J Polit Econ* 98(6):1325–1348.
2. Locksley A, Ortiz V, Hepburn C (1980) Social categorization and discriminatory behavior: Extinguishing the minimal intergroup discrimination effect. *J Pers Soc Psychol* 39(5):773–783.
3. Taijfel H (1970) Experiments in intergroup discrimination. *Sci Am* 223(5):96–102.
4. Rogers TB, Kuiper NA, Kirker WS (1977) Self-reference and the encoding of personal information. *J Pers Soc Psychol* 35(9):677–688.
5. Symons CS, Johnson BT (1997) The self-reference effect in memory: A meta-analysis. *Psychol Bull* 121(3):371–394.
6. Sui J, He X, Humphreys GW (2012) Perceptual effects of social salience: Evidence from self-prioritization effects on perceptual matching. *J Exp Psychol Hum Percept Perform* 38(5):1105–1117.
7. Northoff G, Bermpohl F (2004) Cortical midline structures and the self. *Trends Cogn Sci* 8(3):102–107.
8. Northoff G, et al. (2006) Self-referential processing in our brain—a meta-analysis of imaging studies on the self. *Neuroimage* 31(1):440–457.
9. Platek SM, Wathne K, Tierney NG, Thomson JW (2008) Neural correlates of self-face recognition: An effect-location meta-analysis. *Brain Res* 1232:173–184.
10. Philippi CL, Duff MC, Denburg NL, Tranel D, Rudrauf D (2012) Medial PFC damage abolishes the self-reference effect. *J Cogn Neurosci* 24(2):475–481.
11. Keenan JP, Nelson AM, O'Connor M, Pascual-Leone A (2001) Self-recognition and the right hemisphere. *Nature* 409(6818):305.
12. Turk DJ, et al. (2002) Mike or me? Self-recognition in a split-brain patient. *Nat Neurosci* 5(9):841–842.
13. Gallagher I (2000) Philosophical conceptions of the self: Implications for cognitive science. *Trends Cogn Sci* 4(1):14–21.
14. Schmitz TW, Johnson SC (2007) Relevance to self: A brief review and framework of neural systems underlying appraisal. *Neurosci Biobehav Rev* 31(4):585–596.
15. Happé F (2003) Theory of mind and the self. *Ann N Y Acad Sci* 1001:134–144.
16. Gallagher HL, Frith CD (2003) Functional imaging of "theory of mind." *Trends Cogn Sci* 7(2):77–83.
17. Saxe R, Kanwisher N (2003) People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind." *Neuroimage* 19(4):1835–1842.
18. Saxe R (2010) The right temporo-parietal junction: A specific brain region for thinking about thoughts. *Handbook of Theory of Mind*, eds Leslie A, German T (Taylor & Francis, Brighton, UK).
19. Decety J, Sommerville JA (2003) Shared representations between self and other: A social cognitive neuroscience view. *Trends Cogn Sci* 7(12):527–533.
20. Mitchell JP, Macrae CN, Banaji MR (2006) Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron* 50(4):655–663.
21. Uddin LQ, Iacoboni M, Lange C, Keenan JP (2007) The self and social cognition: The role of cortical midline structures and mirror neurons. *Trends Cogn Sci* 11(4):153–157.
22. Lieberman MD (2007) Social cognitive neuroscience: A review of core processes. *Annu Rev Psychol* 58:259–289.
23. Brother L (1990) The social brain: A project for integrating primate behavior and neurophysiology in a new domain. *Concepts Neurosci* 1:27–51.
24. Samson D, Apperly IA, Chiavarino C, Humphreys GW (2004) Left temporoparietal junction is necessary for representing someone else's belief. *Nat Neurosci* 7(5):499–500.
25. James W (1890) *The Principles of Psychology* (Holt, Rinehart, & Winston, New York), pp 291–372.
26. Cunningham SJ, Turk DJ, Macdonald LM, Neil Macrae C (2008) Yours or mine? Ownership and memory. *Conscious Cogn* 17(1):312–318.
27. Cunningham SJ, Brady-Van den Bos M, Turk DJ (2011) Exploring the effects of ownership and choice on self-memory biases. *Memory* 19(5):449–461.
28. Kim K, Johnson MK (2012) Extended self: Medial prefrontal activity during transient association of self and objects. *Soc Cogn Affect Neurosci* 7(2):199–207.
29. Corbetta M, Shulman GL (2002) Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci* 3(3):201–215.
30. Corbetta M, Patel G, Shulman GL (2008) The reorienting system of the human brain: From environment to theory of mind. *Neuron* 58(3):306–324.
31. DiQuattro NE, Geng JJ (2011) Contextual knowledge configures attentional control networks. *J Neurosci* 31(49):18026–18035.
32. Falk EB, Spunt RP, Lieberman MD (2012) Ascribing beliefs to ingroup and outgroup political candidates: Neural correlates of perspective-taking, issue importance and days until the election. *Philos Trans R Soc Lond B Biol Sci* 367(1589):731–743.
33. Cabeza R, Ciaramelli E, Moscovitch M (2012) Cognitive contributions of the ventral parietal cortex: An integrative theoretical account. *Trends Cogn Sci* 16(6):338–352.
34. Mundy P, Newell L (2007) Attention, joint attention and social cognition. *Curr Dir Psychol Sci* 16(5):269–274.
35. Townsend JT, Ashby FG (1983) *Stochastic Modelling of Elementary Psychological Processes* (Cambridge Univ Press, Cambridge, UK).
36. Bar M, et al. (2006) Top-down facilitation of visual recognition. *Proc Natl Acad Sci USA* 103(2):449–454.
37. Penny WD, et al. (2010) Comparing families of dynamic causal models. *PLOS Comput Biol* 6(3):e1000709.
38. Denny BT, Kober H, Wager TD, Ochsner KN (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.
39. Heatherton TF, et al. (2006) Medial prefrontal activity differentiates self from close others. *Soc Cogn Affect Neurosci* 1(1):18–25.
40. Allison T, Puce A, McCarthy G (2000) Social perception from visual cues: Role of the STS region. *Trends Cogn Sci* 4(7):267–278.
41. Gobbini MI, Koralek AC, Bryan RE, Montgomery KJ, Haxby JV (2007) Two takes on the social brain: A comparison of theory of mind tasks. *J Cogn Neurosci* 19(11):1803–1814.
42. Pammer K, Hansen P, Holliday I, Cornelissen P (2006) Attentional shifting and the role of the dorsal pathway in visual word recognition. *Neuropsychologia* 44(14):2926–2936.
43. Behrens TEJ, Hunt LT, Woolrich MW, Rushworth MFS (2008) Associative learning of social value. *Nature* 456(7219):245–249.
44. Izuma K, Saito DN, Sadato N (2008) Processing of social and monetary rewards in the human striatum. *Neuron* 58(2):284–294.
45. Gronau N, Cohen A, Ben-Shakhar G (2003) Dissociations of personally significant and task-relevant distractors inside and outside the focus of attention: A combined behavioral and psychophysiological study. *J Exp Psychol Gen* 132(4):512–529.
46. Tong F, Nakayama K (1999) Robust representations for faces: Evidence from visual search. *J Exp Psychol Hum Percept Perform* 25(4):1016–1035.
47. Greenwald AG, Banaji MR (1989) The self as a memory system: Powerful, but ordinary. *J Pers Soc Psychol* 57:41–54.
48. Ashburner J, Friston KJ (2005) Unified segmentation. *Neuroimage* 26(3):839–851.
49. Poline JB, Worsley KJ, Evans AC, Friston KJ (1997) Combining spatial extent and peak intensity to test for activations in functional imaging. *Neuroimage* 5(2):83–96.