

<https://doi.org/10.1038/s44271-024-00148-8>

Self-esteem modulates the similarity of the representation of the self in the brains of others

Check for updates

Moriah S. Stendel , Taylor D. Guthrie, Victoria Guazzelli Williamson & Robert S. Chavez

Social neuroscientists have made marked progress in understanding the underlying neural mechanisms that contribute to self-esteem. However, these neural mechanisms have not been examined within the rich social contexts that theories in social psychology emphasize. Previous research has demonstrated that neural representations of the self are reflected in the brains of peers in a phenomenon called the ‘self-recapitulation effect’, but it remains unclear how these processes are influenced by self-esteem. In the current study, we used functional magnetic resonance imaging in a round-robin design within 19 independent groups of participants (total $N = 107$) to test how self-esteem modulates the representation of self-other similarity in multivariate brain response patterns during interpersonal perception. Our results replicate the self-recapitulation effect in a sample almost ten times the size of the original study and show that these effects are found within distributed brain systems underlying self-representation and social cognition. Furthermore, we extend these findings to demonstrate that individual differences in self-esteem modulate these responses within the medial prefrontal cortex, a region implicated in evaluative self-referential processing. These findings inform theoretical models of self-esteem in social psychology and suggest that greater self-esteem is associated with psychologically distanced self-evaluations from peer-evaluations in interpersonal appraisals.

Self-esteem—an individual’s subjective evaluation of their own worth—is believed to develop alongside social factors such as close relationships and social support. The connection between self-esteem and social relationships has been theorized as early as Cooley’s looking-glass self¹, the idea that our self-perceptions are informed by our perceptions of how others view us. Indeed, even competing models of self-esteem highlight interpersonal perception as a core process supporting this reciprocal development of self-esteem and social relationships. For instance, Swann’s self-verification theory² suggests that self-esteem is not dependent on being perceived positively by our peers, but rather by being perceived accurately. In contrast, self-enhancement theory³ states that self-esteem is maintained by an intrinsic motivation to view oneself in a positive light, even at risk of ignoring or minimizing negative peer-perceptions^{4–7}. The sociometer hypothesis proposes self-esteem as a system for monitoring others’ perceptions of us, providing an index of social valuation so that we can regulate our behaviors accordingly⁸. Although these competing theoretical models conflict with one another at points, they are unified by the notion that self-esteem can be shaped by either selectively incorporating or ignoring the appraisals of others.

A body of research also suggests that there might be individual differences in the degree to which interpersonal perception influences self-esteem or vice versa. Although self-esteem is responsive to value cues from others⁹, these value cues seem to exert more influence on the self-view of individuals with lower self-esteem¹⁰, and those with less positive self-image engage in more perspective taking during self-reflection¹¹ and are more likely to internalize peer rejection¹². Focusing on the appraisals of others also leads to immediate drops in state self-esteem¹⁰. These findings suggest that incorporating the views of others into one’s self-view can in fact be detrimental to self-esteem. In this view, having higher similarity between how we see ourselves and how our peers see us is associated with lower self-esteem.

Given these findings, it remains a possibility that distancing one’s view of oneself from the views of others might enhance self-esteem, and indeed the literature on self-enhancement demonstrates this. Here, individuals maintain or boost their self-esteem by preferentially ignoring the negative appraisals of others^{4,13–15}, over-estimating positive appraisals^{5,12,16}, or incorrectly interpreting ambiguous appraisals as self-flattering¹⁷. These findings suggest that possessing high self-esteem might be contingent on distancing one’s self-worth from the appraisals of others. Again, these findings support

Department of Psychology, University of Oregon, Eugene, OR, USA. e-mail: mstendel@uoregon.edu

the idea that having lower agreement between how we see ourselves and how our peers see us would support higher self-esteem.

In recent years, there has been an increasing amount of research identifying the brain systems supporting self-evaluation and self-esteem. Though many regions are involved in social cognition more broadly, such as the cortical midline structures, temporal-parietal junction (TPJ) and the posterior cingulate cortex (PCC), a consistent finding across paradigms has identified the medial prefrontal cortex (MPFC) as a brain structure central to self-evaluation^{6,16,18–20}. The MPFC is a region that has been broadly linked to self-referential processing and person-perception processes²¹. These findings have since been extended to show that MPFC activation following social feedback can predict individual differences in self-esteem^{16,22} and the MPFC seems to track evaluations from others²³ and prompt protective self-reflection in response to negative cues to buffer their debasing effect on self-esteem^{7,24}. When the MPFC is disrupted with transcranial magnetic stimulation, there is a corresponding disruption in self-positivity bias; such disruption leads individuals to become more likely to attribute positive traits to a peer than themselves²⁵ and predict others will socially accept them at chance level rather than above chance²⁶. Other research has linked individual differences in self-esteem to the connectivity of the MPFC with other brain systems linked to reward-related processes and positive affect such as the ventral striatum^{27,28}, suggesting that the MPFC is involved in integrating self-referential thought with evaluative processing. Taken together, this body of evidence suggests that the MPFC is a critical brain region for understanding how the brain reflects self-esteem in social contexts.

Self/other reflection during interpersonal perception has been observed at the neural level²⁹. Using advances in multivariate functional magnetic resonance imaging (fMRI), previous work has found that the patterns of neural activity when an individual engages in self-reflection correlate with the patterns of neural activity of their peers reflecting on them²⁹. Put differently, this provides evidence that neural representations of the self are reflected, or recapitulated, in the brains of other in-group members. This effect—which we henceforth refer to as the self-recapitulation effect—was found in the MPFC identified from an independent functional localizer task. However, this initial study was conducted within a single social group of participants and did not have a sufficient sample size to detect individual differences. Thus, it remains an open question whether these findings extend to multiple and more diverse social groups and how individual differences in self-esteem may influence these results.

In the present study, we sought to inform theories of social influences on self-esteem by examining the relationships between self-esteem and self/other interpersonal perception processes at the level of the brain. We had two primary goals: first, we wanted to replicate the self-recapitulation effect using the round-robin fMRI methods from Chavez and Wagner²⁹ but in a larger sample of participants drawn from multiple independent social groups and across larger brain networks implicated in social cognition. Second, we wanted to test whether individual differences in self-esteem were related to the strength of the self-recapitulation effect within the MPFC, given its established role in serving both social cognitive and self-evaluative processes. To this end, the current study used a multi-group round-robin design wherein previously acquainted individuals reflected on traits about themselves and their group members while undergoing fMRI. Critically, the round-robin design required each individual to serve as both a target and perceiver for all other group members of their group. From here, we can compare an individual's neural activity patterns during self-reflection with the neural activity pattern of their peers reflecting on them. This comparison is the basis for calculating the self-recapitulation effect, or the agreement between brain-to-brain activity patterns between people thinking of themselves and group members thinking of them, which we can then relate to self-esteem.

Methods

Participants

We recruited a total of 120 right-handed participants from 20 independent real-world social groups, 6 people per group, from the University of Oregon

and the local Eugene community. Individuals were eligible to participate in the study if they were over 18 years of age, had no MRI contraindications, and were part of a social group of at least five individuals who were all previously acquainted and all consented to participate in the study. We used posters and in-person recruitment to recruit individuals from a variety of social groups in Eugene, OR, including local places of employment, campus clubs, Reserve Officers' Training Corps, and sororities/fraternities. The sample size was determined by maximizing the number of groups and participants given the resources and feasibility of the design based on previous work²⁹ and is an order of magnitude greater than the previous study to increase power for detecting effects of self-esteem. All participants within each group were known to one another prior to entering the study. Across all groups, five participants failed to fulfill scheduled experimental session appointments, two subjects were excluded due to unusable imaging data, and three subjects were excluded for missing or incomplete behavioral data. After these exclusions, one group was eliminated for having too few subjects to calculate the recapitulation effect and the remaining three participants from that group were excluded. This left a final sample of 107 subjects across 19 groups in which all group members knew every other participant in the group. All participants were screened for MRI contraindications and had normal or corrected to normal vision. Sample demographics, as self-reported by participants, are as follows: age (range = 18–51, $M = 23.50$, $SD = 7.36$), gender (Women = 39, Men = 68), race (American Indian or Alaskan Native = 2, Asian = 3, Black or African American = 5, Native Hawaiian or Pacific Islander = 2, White or Caucasian = 83, multiple = 10, other = 2), and ethnicity (Hispanic = 17, Non-Hispanic = 90). The Internal Review Board (IRB) at the University of Oregon approved the study. Participants gave informed consent in accordance with the guidelines set by the IRB and all were paid for their participation.

Procedures

Behavioral measures. Participants completed two sessions: a behavioral-only session and an fMRI scanning session. Participants conducted both the behavioral and scanning sessions individually and not as a group. In the first session, trait self-esteem was measured using the Janis-Field Feelings of Inadequacy scale (JFS)³⁰, a thirty-six-item inventory of global self-esteem shown to have robust psychometric properties³¹. The JFS asks participants to use a 5-point Likert scale to respond to a variety of items that assess self-evaluative feelings (e.g. “How often do you have the feeling that there is nothing you can do well?”, “When you think that some of the people you meet might have an unfavorable opinion of you, how concerned or worried do you feel about it?”, or “How often do you dislike yourself?”). Four items were reverse-scored and the scale was summed, with higher scores indicating higher self-esteem. One of the primary motivations for collecting a much larger sample to increase our power to detect individual differences. Since individual differences in self-esteem were central to our research question, we opted to use the JFS, which is a longer, more comprehensive measure of self-esteem than other commonly used measures (e.g. 10-item Rosenberg Scale). We did not need to prioritize brevity for this measure as participants were only completing a self-report and not responding in a round robin fashion about other group members. All previously published work on self-esteem from members of our group that has used the Janis and Field measure of trait self-esteem when relating this construct to brain metrics (see: ^{27,28,32}). Data distribution was assumed to be normal but this was not formally tested. Participants completed additional personality and mental health measures (Big Five Inventory, PROMIS mental health measures, stereotype content model) as part of a larger study on the brain basis of interpersonal perception²⁷. All responses were recorded using the PsychoPy stimulus presentation software³³.

Neuroimaging procedures. In a second session, participants returned to complete the fMRI portion of the experiment, which, in replicating the protocol in Chavez & Wagner²⁹, consisted of two standard versions of a trait-judgment task widely used in the study of self-referential processing³⁴. The first task in the scanner was the self-referential

functional localizer. Participants were asked to make trait evaluations for themselves or an “Ordinary American” as a functional localizer to independently define the region of MPFC that responds to self-evaluation to be used in subsequent analyses. Participants were presented with the word “SELF” or “ORDINARY AMERICAN” above a valence-balanced trait adjective and asked to respond “yes” or “no” using a button box as to whether the adjective described the individual above. The words were presented for 2000 ms followed by 500 ms of fixation and intermittent passive fixation trials (2500–7500 ms). Jittered trials were optimized using Opseq³⁵. One run of this functional localizer task was completed and used to define a region of interest (ROI) within the MPFC that was maximally involved in self-referential processing and which was identified independently from the main task of interest. The functional localizer procedure is performed to avoid issues related to statistical non-independence or circular analysis which distort the interpretation of effects in neuroimaging studies^{36,37}.

The second scanner task was the main self/other round-robin task. In this task, we used a similar trait judgment task but modified to accommodate the round-robin nature of the interpersonal perception judgments we collected. Participants completed six runs of a round-robin version of a standard trait-judgment task, wherein participants were presented with a name from their small group alongside a trait-adjective and asked to respond via button-box as to whether this trait described that individual. For each trial, participants saw vertically-arranged white text on a black background with the top word either reading “SELF” or the name of one of the participant’s group members, and the bottom word displaying one of 360 valence-balanced trait-adjectives (e.g. “Happy”, “Clumsy”, “Smart”; Anderson, 1968) for 2000 ms followed by 2000 ms of fixation and jittered intermittent passive fixation trials (2000–12,000 ms). Participants were asked to make either a “yes” or “no” response using a button box in their right hand as to whether or not the trait adjective described either themselves or one of their group members. All targets were presented in every run, and there were a total of twelve trials per target per run. Individual traits were only presented once per target randomly across all runs in the experiment. No two participants were presented with the same target/trait-adjective order across the experiment, to account for potential order effects.

Crucially, we employed a full round-robin design in which each participant was both a target and perceiver when asked to reflect on themselves along with all other members within their group. This design allowed us to measure similarities in brain activity patterns to assess self-to-other similarity to test to what degree an individual’s neural activity pattern during self-reflection corresponds with the neural activity patterns of peers reflecting on the individual.

Image acquisition. Magnetic resonance imaging was conducted with a Siemens 3 T Skyra scanner using a 32-channel phased array coil. Structural images were acquired using a T1-weighted MP-RAGE protocol (175 sagittal slices; TR: 2500 ms; TE: 3.43 ms; flip angle: 7; 1 mm isotropic voxels). Functional images were acquired using a T2*-weighted echo planar sequence (TR: 2000 ms; 72 axial slices; TE: 25 ms; flip angle: 90; 2 mm isotropic voxels). For each participant, we collected one run of the functional localizer task and five runs of the main round-robin self-other task. In order to correct for distortion due to B0 inhomogeneity we also acquired a field map (TR: 6390 ms; TE: 47.8 ms; effective echo spacing: 0.345 ms). The total length of time for the entire scanning session was approximately 75 min.

Openness and transparency

We report how we determined our sample size, all data exclusions and manipulations, and we follow JARS³⁸. All data, analysis code, and research materials are available at <https://osf.io/n7pr2/>. This study’s design and its analysis were not pre-registered.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Analyses

Functional localizer preprocessing and analysis

All neuroimaging data including the functional localizer task were pre-processed and voxel responses were estimated using the FSL neuroimaging software package³⁹. For the functional localizer, data were first slice time corrected followed by a mean-based intensity normalization, high pass filtering (Gaussian-weighted least-squares straight line fitting, with $\sigma = 100$ s) and spatially smoothed with a 6-mm FWHM Gaussian smoothing kernel. First-level analyses investigated the self > ordinary American contrast. These results were then submitted to a group-level random effects analysis to isolate regions of the brain involved in self-referential processing, correcting for multiple comparisons using voxelwise family error rate corrected *p*-values generated from $n = 5000$ iteration nonparametric permutation test. Note, the familywise error correction here refers to whole-brain voxelwise values and is more conservative than cluster-based correction or false discovery rate procedures. Voxels with a corrected *p*-value < 0.05 from this analysis were then converted into a binary mask that were then used at the group level to extract self and target response patterns within the round-robin task.

Round-Robin task preprocessing and response pattern estimation

Preprocessing and registration of the main self-other task was identical to that of the functional localizer task described above except for the use of a smaller 4-mm FWHM Gaussian smoothing kernel (as is more typical of multivariate fMRI analyses where there is a need to preserve finer-scale neural responses). A multistep normalization procedure was used to register results to standard space. First, functional data were corrected for spatial distortions using a field map unwarping before aligning functional data to each participant’s anatomical scan using boundary-based registration⁴⁰ in conjunction with a linear registration with FSL’s FLIRT tool. These images were then warped to a 2-mm MNI template using nonlinear registration with FSL’s FNIRT tool and a 10-mm warp field. All first-level task-based analyses were performed in the participant’s native space before being warped into standard space for final analyses. Parameter estimates were separately estimated for the self-condition as well as for each of the other peers within each of the six runs. These responses were then combined in a second level within-subject fixed-effects analysis yielding a parameter estimate for the self as well as parameter estimates for each of the other group members participant’s conditions. Normalized (i.e., *z*-score) voxel responses for each condition within the ROI defined in the functional localizer task served as input to the subsequent round-robin analysis.

Multivariate Round-Robin task and self-esteem analyses

Here, we aimed to measure the degree to which a target’s activation pattern during self-trials correlates with aggregated activation patterns of group members reflecting on the target. For example, in a group with participants A, B, C, D, and E, a Spearman rank correlation of voxelwise parameter estimates within the regions identified by the functional localizer while A was thinking of A with the aggregate of voxelwise parameter estimates in the same ROI when B, C, D, and E were thinking of A. To this end, voxelwise parameter estimates within all areas identified by the functional localizer were flattened into one-dimensional response vectors to allow correlations across participants and conditions. Mirroring classic work in the interpersonal perception literature in which trait ratings are averaged across raters⁴¹, we used aggregated peer response patterns across voxelwise responses yielding a single group averaged response pattern for each target. As is standard in studies using fMRI to estimate multivariate similarity patterns (e.g.,³⁶), we used Spearman rank correlation distance metric (1 - Spearman-rho) to compute the dissimilarity between neural response patterns during the self-condition and the group-averaged peer responses vector from the other participants in each group. Thus, this distance metric has the range of 0–2 with smaller values indicating greater similarity and higher agreement between self/other brain-to-brain activation patterns.

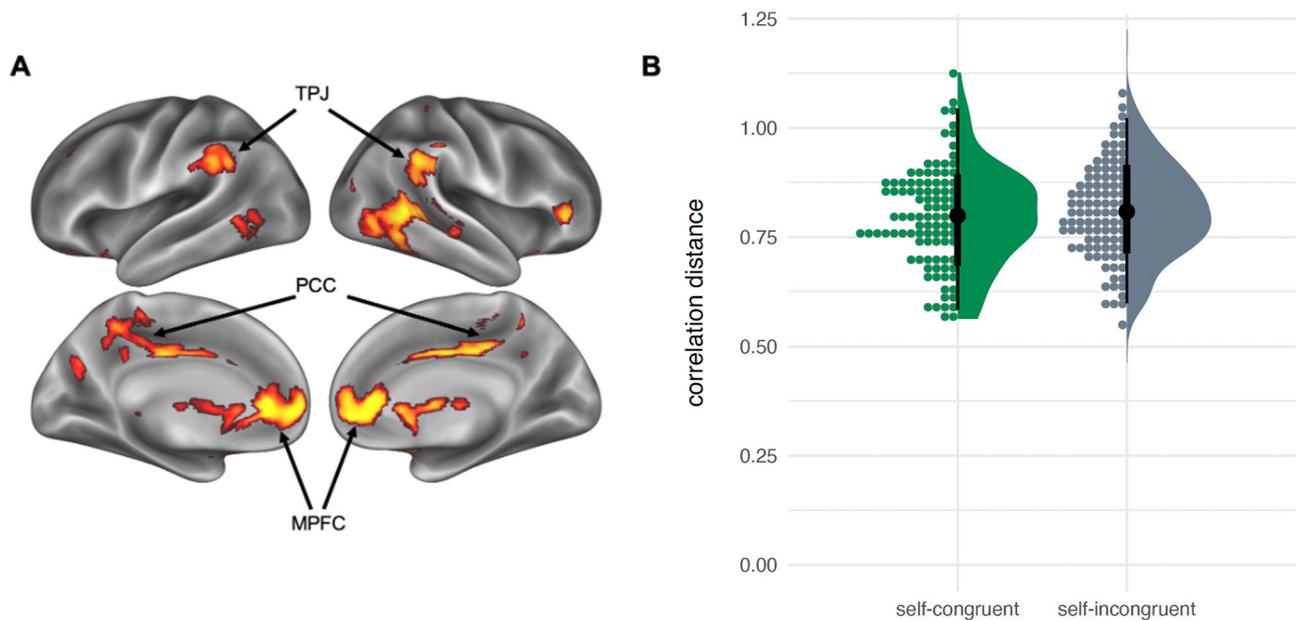


Fig. 1 | Results from the multi-group replication of the self-recapitulation effect from Chavez & Wagner²⁹. **A** Regions identified from the functional localizer task in which brain activity when making trait-judgments of the self was greater than making trait-judgements for others ($p < 0.05$, voxelwise-corrected). Here we identified areas previously implicated in self-referential processing and social cognition including the medial prefrontal cortex (MPFC), posterior cingulate cortex (PCC),

and the temporal-parietal junction (TPJ), and used these to independently define the brain areas examined in the round-robin task. $N = 107$. **B** We show that self-referential brain activity for each target participant shows greater accuracy (lower correlation distance) with the brain activity of group members when they are thinking about that target (self-congruent) than when they are thinking of different targets (self-incongruent). $N = 107$.

To provide evidence for the replication of the self-recapitulation effect, self-referential brain activity patterns within a target participant are correlated with those of others in their group when they are thinking of the target. However, rather than simply testing for a nonzero association between the target's self-responses and their peers' responses, we sought to test the more conservative hypothesis that these relationships are specific to the given target and, in general, exceed that for other persons. To accomplish this, we also computed the correlation distance between a target's self-responses and the group response vector for every other participant within each social group. This allowed us to then calculate the similarity between self-congruent target comparisons (i.e., when perceivers were thinking of the same target as self) and self-incongruent target comparisons (i.e. when perceivers were thinking of a different target than the self). Importantly, the incongruent comparisons did not include the self-target responses so as not to influence the dissimilarity with respect to their own congruent patterns. Next, we fit a nested multilevel model comparing correlation distances between the self/other responses when congruent versus incongruent with random intercepts for each participant within each group.

Finally, we tested the hypothesis that the strength of the self-recapitulation effect (i.e. the self-congruent effects from the replication analysis) from each participant would be related to individual differences in self-esteem within the MPFC. To select a region of the MPFC for this analysis, we thresholded the functional localizer results to only include the most significant regions identified from that analysis $t(106) = 8.5$, such that the only voxels that remained in the mask were those within a contiguous cluster within the MPFC. The self-esteem correlation analyses were then calculated within this MPFC region for the self-esteem correlation analysis, as well as in the mask from the localizer as a whole. This was done in order to test the specificity of the hypothesized results to the MPFC and not neural systems more broadly implicated in self-referential processing and social cognition.

Results

Replication of the neural self-recapitulation effect

Results from the functional localizer task identified a set of brain regions previously associated with self-referential processing and social cognition

(see Fig. 1A). These regions include the MPFC, precuneus/posterior cingulate cortex (PCC), temporal parietal junction (TPJ), lateral orbitofrontal cortex, and the ventral striatum; this analysis was performed on the social cognition network as a whole, including all of these regions as a single mask. Within this ROI, we then calculated the correlation distance between a target's activation pattern during self trials and the aggregated activation patterns of group members reflecting on the target. This was iterated for every target in the round-robin design across all groups in the study to test for the self-recapitulation effect using a multilevel model to test whether the correlation distance between brain response patterns for the self was lower (i.e. more similar) when group members were thinking of the self (i.e. self-congruent correlation distance) than when group members were thinking of a different person in the group (i.e. self-incongruent correlation distance). The results show that other's self-congruent responses were more similar to the self-referential patterns than they were compared against others (see Fig. 1B) after accounting for nested effects of groups and participants ($b = 0.02$, $SE = 0.01$, $t = 2.81$, $p = 0.0052$, $95\% \text{ CI} = [0.004, 0.025]$). These followed the same procedures as those from Chavez & Wagner²⁹ but in multiple groups of subjects and within a significantly larger sample. These results replicate the main finding of evidence for the self-recapitulation effect in which brain response patterns for the self are reflected in aggregated responses from group members. Moreover, we find that these results extend to show that these are present across multiple brain systems involved in both self-representation and social cognition.

Relationship of self-recapitulation strength and self-esteem

Next, we sought to determine whether the strength of the self-recapitulation effect from the multi-voxel fMRI patterns could be predicted by the target self-esteem. The most significant area identified in the functional localizer task identified a region of the MPFC that was used for the self-esteem correlation analysis (Fig. 2A). Within this region, we found a significant association between each individual's self-esteem and the similarity of patterns of self-representation to that of the aggregated group representations (i.e. the self-recapitulation strength), $R = 0.27$, $t(105) = 2.82$, $p = 0.006$, $95\% \text{ CI} = [0.08, 0.43]$ (Fig. 2B). These results show that the lower a person's

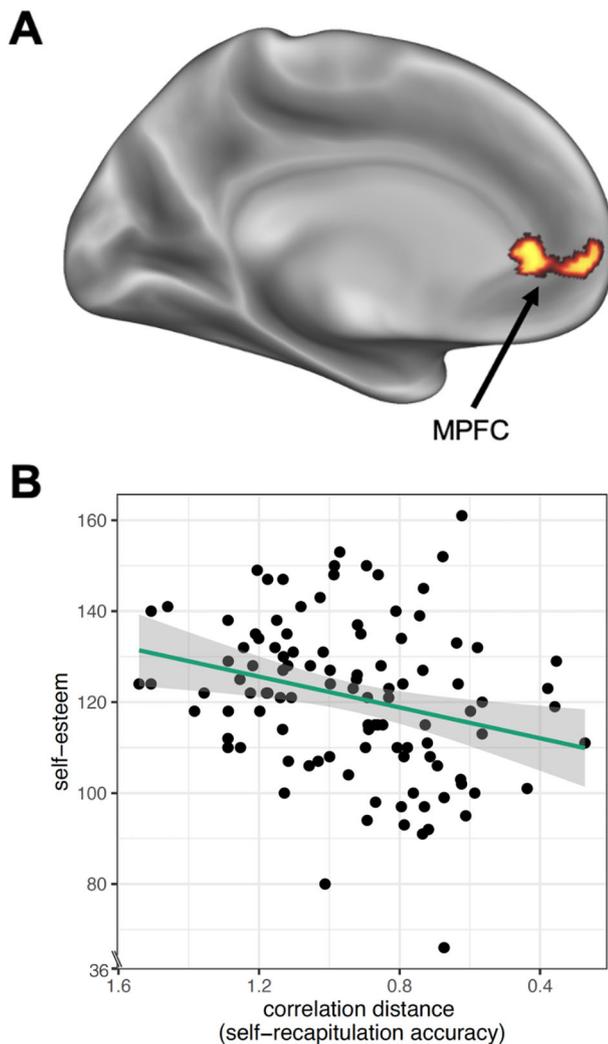


Fig. 2 | The results of the correlation of self-esteem with the strength of the recapitulation effect within the medial prefrontal cortex. **A** A portion of the medial prefrontal cortex (MPFC) that showed the most significant association with thinking of the self, relative to others from the functional localizer task and has been previously associated with self-evaluative processing. $N = 107$. **B** A scatter plot depicting the correlation between self-recapitulation strength and self-esteem within the MPFC. These results show that as self-esteem increases, self-recapitulation accuracy (i.e. correlation distance from the neural self-recapitulation analysis) decreases. Error bars reflect 95% confidence intervals. The scale of the values on the x-axis of the plot is inverted to increase the clarity of the interpretation of correlation distance as an inverse measure of self-recapitulation strength from these analyses. As the JFS has 36 questions scored on a five-point Likert scale, the lowest possible self-esteem score is 36, thus reflecting the lowest value on the y-axis. $N = 107$.

self-esteem, the more their neural representation within the MPFC self-reflection generally corresponds with those of their group members while reflecting on them. Additionally, self-esteem scores were not significantly correlated to the total number of in-scanner trait endorsements from group members for either positively valenced words ($R(105) = 0.11$, $p = 0.26$, 95% CI = $[-0.08, 0.29]$) or negatively valenced words ($R(105) = 0.11$, $p = 0.26$, 95% CI = $[-0.24, 0.14]$).

In order to test the specificity of this effect to the MPFC, we repeated the same correlation procedure but within the functional localizer mask as a whole from the replication analysis. From this analysis, we find that there was no significant correlation between the self-recapitulation effect strength and self-esteem in broader regions of the social brain identified with the localizer task $R = 0.09$, $t(105) = 0.93$, $p = 0.35$, 95% CI = $[-0.10, 0.28]$.

Additionally, we entered self-recapitulation values from both the MPFC and the remainder of the functional localizer in a multiple regression model as independent predictors of self-esteem. From this procedure, the full model remained significant, $F(2, 104) = 3.98$, $R^2 = 0.071$, $p = 0.022$. However, only MPFC self-recapitulation coefficient ($t = 2.65$, $p = 0.009$, 95% CI = $[4.2, 28.9]$) and not the rest of the localizer ($t = 0.27$, $p = 0.79$, 95% CI = $[-25.1, 32.7]$) remained a significant predictor in the model. To allow for a more comprehensive interpretation of the null effect from the localizer results, we also conducted an equivalence test based on the two one-sided tests procedure⁴². Here, we tested for the correlation of self-esteem to both the MPFC self-recapitulation values and those for the rest of the localizer, with equivalence boundaries of $R[-0.3, 0.3]$ and an $\alpha = 0.025$ using the TOSTER package in R⁴³. From this analysis, the MPFC region showed 95% CI $[0.08, 0.43]$ outside bands of the test boundaries indicating that this effect is not significantly equivalent to zero. The results from the remainder of the functional localizer showed a 95% CI $[-0.10, 0.28]$, which was within the bands of the test boundaries, indicating that this effect is significantly equivalent to zero. These results suggest that the effect of self-esteem modulates the self-recapitulation effect but only within the MPFC, which has previously been associated with self-evaluative processing, and provide increased specificity for the brain locations underlying this effect. These results underscore the specificity of the MPFC self-recapitulation and its relationship to self-esteem.

Discussion

We sought to test whether individual differences in self-esteem modulate the degree to which neural patterns overlap during self- and other-perception. To this end, we used a round-robin design wherein small groups of acquainted individuals reflected on traits about themselves and their group members while undergoing fMRI, and then determined the degree to which distributed patterns of brain activity in the MPFC during self- and other-reflection correlate. This design positions us to examine: (1) the degree to which an individual's neural activity pattern during self-reflection correlates with the neural activity patterns of peers reflecting on the individual, and (2) whether individual differences in self-esteem predict the magnitude of this effect. Replicating the main results from Chavez and Wagner²⁹, we find evidence for the self-recapitulation effect across multiple brain systems involved in both self-representation and social cognition. Furthermore, we also find that the self-recapitulation effect is moderated by self-esteem in the MPFC, such that lower self-esteem is associated with increased pattern overlap when that individual reflects on themselves and when their group reflects on them. These effects were specific to the MPFC, which has previously been identified as a region serving self-evaluative processes, and not found in other regions of the brain that serve other social cognitive processes. These findings have implications for both our understanding of the neural systems supporting interpersonal perception as well as social psychological theories of self-esteem.

A variety of competing theories have emphasized different primary sources of self-esteem. Intrapersonal perspectives on self-esteem propose self-esteem as a personal valuation of one's worth, in contrast to interpersonal perspectives that suggest that self-esteem is, at least in part, formed by the evaluation of others⁴⁴. To this end, we find that those who neurally represent themselves more similarly to how their peers represent them present with lower trait self-esteem. Although the present study was not designed to formally test theoretical models of self-esteem, our findings provide insight into these competing theories.

For example, some work has argued that higher self-esteem emerges from being perceived accurately by our peers (e.g. Self-Verification Theory;⁴⁵) with other research suggesting the opposite – that higher self-esteem relies on us representing ourselves distinctly from how others view us, particularly in an inaccurately positive fashion (e.g. Self-Enhancement Theory³). Our findings provide a neural account for self-enhancement theories of self-esteem, such that individuals with higher self-esteem represent themselves more differently from how their peers represent them. This finding connects to the extensive literature on self-enhancement biases,

which suggests that people tend to inflate their own self-worth even in the face of competing information^{4,7}. Critically, we find this effect in the MPFC, and recent neuroimaging work has identified the MPFC as a brain structure that is core to the process of self-enhancement^{6,18,20,46}, proposing the MPFC as a candidate region for the integration of social and self-evaluation that might compute self-protective responses in the face of negative social feedback⁹. The present study's findings suggest that the neural integration of value cues from peers might render one more vulnerable to lower self-esteem and/or that individuals with lower self-esteem might more accurately integrate value cues from peers.

Limitations

Although we found evidence that lower self-esteem is associated with increased neural overlap between a target's self-representation and how peers represent that target, these findings need to be interpreted in the context of their limitations. First, our interpretations rest on the assumption that the trait-adjective experimental paradigm is reliably evoking self and other- concepts. While this task is regularly used to elicit interpersonal representations in neuroimaging studies, the use of single trait adjectives might limit the richness of such representations. Even though this task represents one limited instantiation of self- and other- referential thought, the present findings suggest that similarities across even crude interpersonal representations are related to trait self-esteem. Moreover, while this task typically has individuals reflect on a distant other (e.g. a former president) or for varied targets across subjects (e.g. a parent), our round-robin paradigms allows all group members to reflect on the same, close others. We anticipate that conducting this task within real-world social groups increases the richness of person representations.

Another main limitation of this study is the inability to examine the effect of valence on the recapitulation effect as this study was not designed with enough trials of either positive or negative trait words to reliably test for differences between the two. Given that valence is central to many theories of self-esteem, future work would benefit from examining whether recapitulation is driven by positive traits, negative traits, or both, which would paint a clearer picture of how these findings relate to theories of self-esteem like self-enhancement.

One alternate explanation for our finding that those with higher self-esteem show lower recapitulation in the MPFC is that these individuals might have larger social groups, and thus their self-perceptions might be represented by a broader social network and thus less similar to the smaller group that participated in the study. While we cannot explicitly test this interpretation as we do not have data about the broader social networks that our participants are a part of, the degree of recapitulation was not significantly predicted by participant extraversion ($r(104) = 0.12$, $t = 1.22$, $p = 0.22$, 95% CI = $-0.07, 0.30$); we might also expect that more extraverted individuals have larger social groups, yet we do not find statistically significant evidence that this impacts recapitulation here. The recapitulation effect was also not predicted by relationship factors (knowing: $r(104) = -0.04$, $t = -0.44$, $p = 0.66$, 95% CI = $-0.23, 0.15$; similarity: $r(104) = -0.6$, $t = -0.61$, $p = 0.54$, 95% CI = $-0.25, 0.13$; liking: $r(104) = 0.02$, $t = -0.19$, $p = 0.85$, 95% CI = $-0.17, 0.21$; friendship: $r(104) = 0.09$, $t = 0.94$, $p = 0.35$, 95% CI = $-0.10, 0.28$; total relationship quality: $r(104) = -0.03$, $t = -0.26$, $p = 0.79$, 95% CI = $-0.22, 0.17$). Future research can collect data about the broader social networks participants are embedded in to more explicitly test whether these factors impact recapitulation.

Our interpretations are also limited by the homogeneity of our sample, which was entirely adult and from the United States, and largely white. Conceptions of selfhood and the importance of self-esteem both vary across cultures⁴⁷⁻⁵⁰ and development⁵¹, and our sample homogeneity limits our ability to generalize our findings to other populations. While cultural and developmental factors influence self-concept, across all humans, self-referential thought rests on psychological and biological processes at the level of the individual. We hope our findings can contribute to our understanding of such processes while also acknowledging how the homogeneity

of our sample limits external validity. It is necessary that future work emphasize both cultural and developmental considerations into the examination of self-concept and self-esteem, given the context-laden nature of such constructs.

Conclusions

There are numerous theories of neural and cognitive processes underlying interpersonal perception, self-esteem, and the possible interaction of these two via the integration of others' perspectives into one's self-concept. Although our study was not designed to directly test competing self-esteem models, our results are most consistent with theories that state that accurate self-perception is associated with lower self-esteem, such as sociometer theory or self-enhancement, rather than verification theories. Although future work is needed that explicitly tests out competing models of self-esteem using richer conceptions of self and other, the present findings suggest that similarities across even crude interpersonal representations are related to trait self-esteem. This reflects one possible neural mechanism underlying theoretical models that link self-esteem to individual differences in interpersonal perception.

Data availability

All data and research materials are available at <https://osf.io/n7pr2/>.

Code availability

All analysis code is available at <https://osf.io/n7pr2/>.

Received: 23 March 2024; Accepted: 15 October 2024;

Published online: 27 November 2024

References

- Cooley, C. H. *Human nature and the social order*. (Charles Scribner's Sons, 1902).
- Swann, W. B. *To be adored or to be known? The interplay of self-enhancement and self-verification*. *Foundations of Social Behavior* (Guilford Press, 1990).
- Krueger, J. Enhancement bias in description of self and others. *Pers. Soc. Psychol. Bull.* **24**, 505–516 (1998).
- Anderson, M. C. & Hanslmayr, S. Neural mechanisms of motivated forgetting. *Trends Cogn. Sci.* **18**, 279–292 (2014).
- Hepper, E. G., Hart, C. M., Gregg, A. P. & Sedikides, C. Motivated expectations of positive feedback in social interactions. *J. Soc. Psychol.* **151**, 455–477 (2011).
- Hughes, B. L. & Beer, J. S. Protecting the self: the effect of social-evaluative threat on neural representations of self. *J. Cogn. Neurosci.* **25**, 613–622 (2012).
- Korn, C. W., Prehn, K., Park, S. Q., Walter, H. & Heekeren, H. R. Positively biased processing of self-relevant social feedback. *J. Neurosci.* **32**, 16832–16844 (2012).
- Leary, M. M. R., Tambor, E. S. E., Terdal, S. K. & Downs, D. L. Self-esteem as an interpersonal monitor: the sociometer hypothesis. *J. Pers. Soc. Psychol.* **68**, 518–530 (1995).
- Weisbuch, M., Sinclair, S. A., Skorinko, J. L. & Eccleston, C. P. Self-esteem depends on the beholder: effects of a subtle social value cue. *J. Exp. Soc. Psychol.* **45**, 143–148 (2009).
- Tice, D. M. & Wallace, H. M. The reflected self: Creating yourself as (you think) others see you. in *Handbook of self and identity* 91–105 (The Guilford Press, 2003).
- Pfeifer, J. H. et al. Neural correlates of direct and reflected self-appraisals in adolescents and adults: when social perspective-taking informs self-perception. *Child Dev* **80**, 1016–1038 (2009).
- Rodman, A. M., Powers, K. E. & Somerville, L. H. Development of self-protective biases in response to social evaluative feedback. *Proc. Natl. Acad. Sci.* **114**, 13158–13163 (2017).
- Frey, D. L. & Gaertner, S. L. Helping and the avoidance of inappropriate interracial behavior: a strategy that perpetuates a

- nonprejudiced self-image. *J. Pers. Soc. Psychol.* **50**, 1083–1090 (1986).
14. Pyszczynski, T. & Greenberg, J. Depression and preference for self-focusing stimuli after success and failure. *J. Pers. Soc. Psychol.* **49**, 1066–1075 (1985).
 15. Schröder-Abé, M., Rudolph, A., Wiesner, A. & Schütz, A. Self-esteem discrepancies and defensive reactions to social feedback. *Int. J. Psychol.* **42**, 174–183 (2007).
 16. Somerville, L. H., Kelley, W. M. & Heatherton, T. F. Self-esteem modulates medial prefrontal cortical responses to evaluative social feedback. *Cereb. Cortex* **20**, 3005–3013 (2010).
 17. Flagan, T., Mumford, J. A. & Beer, J. S. How do you see me? The neural basis of motivated meta-perception. *J. Cogn. Neurosci.* **29**, 1908–1917 (2017).
 18. Van de Groep, I. H. et al. Overlapping and distinct neural correlates of self-evaluations and self-regulation from the perspective of self and others. *Neuropsychologia* **161**, 108000 (2021).
 19. Will, G.-J., Rutledge, R. B., Moutoussis, M. & Dolan, R. J. Neural and computational processes underlying dynamic changes in self-esteem. *eLife* **6**, e28098 (2017).
 20. Yankouskaya, A. & Sui, J. Self-positivity or self-negativity as a function of the medial prefrontal cortex. *Brain Sci.* **11**, 264 (2021).
 21. Wagner, D. D., Haxby, J. V. & Heatherton, T. F. The representation of self and person knowledge in the medial prefrontal cortex. *Wiley Interdiscip. Rev. Cogn. Sci.* **3**, 451–470 (2012).
 22. Eisenberger, N. I., Inagaki, T. K., Muscatell, K. A., Byrne Haltom, K. E. & Leary, M. R. The neural sociometer: brain mechanisms underlying state self-esteem. *J. Cogn. Neurosci.* **23**, 3448–3455 (2011).
 23. Ito, A. et al. The role of the ventromedial prefrontal cortex in automatic formation of impression and reflected impression. *Hum. Brain Mapp.* **41**, 3045–3058 (2020).
 24. Yoon, L., Somerville, L. H. & Kim, H. Development of MPFC function mediates shifts in self-protective behavior provoked by social feedback. *Nat. Commun.* **9**, 3086 (2018).
 25. Luber, B., Lou, H. C., Keenan, J. P. & Lisanby, S. H. Self-enhancement processing in the default network: a single-pulse TMS study. *Exp. Brain Res.* **223**, 177–187 (2012).
 26. Zhang, D. et al. Modulating social feedback processing by deep TMS targeting the medial prefrontal cortex: behavioral and electrophysiological manifestations. *NeuroImage* **250**, 118967 (2022).
 27. Chavez, R. S., Tovar, D. T., Stendel, M. S. & Guthrie, T. D. Generalizing effects of frontostriatal structural connectivity on self-esteem using predictive modeling. *Cortex* **146**, 66–73 (2022).
 28. Chavez, R. S. & Heatherton, T. F. Multimodal frontostriatal connectivity underlies individual differences in self-esteem. *Soc. Cogn. Affect. Neurosci.* **10**, 364–370 (2015).
 29. Chavez, R. S. & Wagner, D. D. The neural representation of self is recapitulated in the brains of friends: a round-robin fMRI study. *J. Pers. Soc. Psychol.* **118**, 407–416 (2020).
 30. Fleming, J. S. & Courtney, B. E. The dimensionality of self-esteem: II. Hierarchical facet model for revised measurement scales. *J. Pers. Soc. Psychol.* **46**, 404–421 (1984).
 31. Blascovich, J. & Tomaka, J. Measures of Self-Esteem. in *Measures of Personality and Social Psychological Attitudes* 115–160 (Elsevier, 1991). <https://doi.org/10.1016/B978-0-12-590241-0.50008-3>.
 32. Chavez, R. S. & Heatherton, T. F. Structural integrity of frontostriatal connections predicts longitudinal changes in self-esteem. *Soc. Neurosci.* <https://doi.org/10.1080/17470919.2016.1164753> (2016).
 33. Peirce, J. W. PsychoPy—psychophysics software in Python. *J. Neurosci. Methods* **162**, 8–13 (2007).
 34. Mitchell, J. P., Schirmer, J., Ames, D. L. & Gilbert, D. T. Medial prefrontal cortex predicts intertemporal choice. *J. Cogn. Neurosci.* **23**, 857–866 (2011).
 35. Dale, A. M. Optimal experimental design for event-related fMRI. *Hum. Brain Mapp.* **8**, 109–114 (1999).
 36. Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S. F. & Baker, C. I. Circular analysis in systems neuroscience: the dangers of double dipping. *Nat. Neurosci.* **12**, 535–540 (2009).
 37. Vul, E., Harris, C., Winkielman, P. & Pashler, H. Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspect. Psychol. Sci.* **4**, 274–290 (2009).
 38. Appelbaum, M. et al. Journal article reporting standards for quantitative research in psychology. *APA Publ. Commun. Board Task Force Rep. Am. Psychol.* **73**, 3 (2018).
 39. Smith, S. M. et al. Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage* **23**, 208–219 (2004).
 40. Greve, D. N. & Fischl, B. Accurate and robust brain image alignment using boundary-based registration. *NeuroImage* **48**, 63–72 (2009).
 41. McCrae, R. R. & Costa, P. T. Validation of the five-factor model of personality across instruments and observers. *J. Pers. Soc. Psychol.* **52**, 81–90 (1987).
 42. Lakens, D., Scheel, A. M. & Isager, P. M. Equivalence testing for psychological research: a tutorial. *Adv. Methods Pract. Psychol. Sci.* **1**, 259–269 (2018).
 43. Lakens, D. Equivalence tests: a practical primer for t tests, correlations, and meta-analyses. *Soc. Psychol. Personal. Sci.* **8**, 355–362 (2017).
 44. Yang, J., Xu, X., Chen, Y., Shi, Z. & Han, S. Trait self-esteem and neural activities related to self-evaluation and social feedback. *Sci. Rep.* **6**, 20274 (2016).
 45. Swann, W. B. Jr & Read, S. J. Self-verification processes: how we sustain our self-conceptions. *J. Exp. Soc. Psychol.* **17**, 351–372 (1981).
 46. Duran, K. A. et al. The medial prefrontal cortex: a potential link between self-deception and affect. *Int. J. Neurosci.* **131**, 701–707 (2021).
 47. Goto, S. G., Lewis, R. S. & Grayzel-Ward, S. Culture and Self-Conceptual in *The Oxford handbook of cultural neuroscience and global mental health*. 100. p. 288 (2021).
 48. Gurel-Atay, E., Xie, G. X., Chen, J. & Kahle, L. R. Changes in social values in the United States: 1976–2007: “Self-respect” is on the upswing as “a sense of belonging” becomes less important. *J. Advert. Res.* **50**, 57–67 (2010).
 49. Markus, H. R. & Kitayama, S. Cultural variation in the self-concept. in *The self: Interdisciplinary approaches* 18–48 (Springer, New York, NY, 1991).
 50. Triandis, H. C. The self and social behavior in differing cultural contexts. *Psychol. Rev.* **96**, 506 (1989).
 51. Robins, R. W. & Trzesniewski, K. H. Self-esteem development across the lifespan. *Curr. Dir. Psychol. Sci.* **14**, 158–162 (2005).

Acknowledgements

This study was funded by the University of Oregon (R.S.C.) and the Social Science and Humanities Research Council of Canada (M.S.S.). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

M.S.S.: conceptualization (lead), writing—original draft (lead), writing—review & editing (lead), analysis (lead), review & editing (lead), data collection (supporting), funding acquisition (supporting). T.D.G.: project administration (lead), data collection (lead), analysis (supporting), writing—review & editing (supporting). V.G.W.: conceptualization (supporting), writing—review & editing (supporting). R.S.C.: conceptualization (lead), writing—original draft (lead), writing—review & editing (lead), analysis (lead), funding acquisition (lead).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s44271-024-00148-8>.

Correspondence and requests for materials should be addressed to Moriah S. Stendel.

Peer review information *Communications psychology* thanks Elisa Baek and Yaara Yeshurun for their contribution to the peer review of this work. Primary Handling Editors: Yafeng Pan and Jennifer Bellingtier. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024