Behavioral/Cognitive

The Self-Concept Is Represented in the Medial Prefrontal Cortex in Terms of Self-Importance

Marie Levorsen,[1] **Ryuta Aoki,**[2] **Kenji Matsumoto,**[3] **Constantine Sedikides,**[1] **and Keise Izuma**[1,4,5]

[1]School of Psychology, University of Southampton, Southampton SO17 1BJ, United Kingdom, [2]Graduate School of Humanities, Tokyo Metropolitan University, Tokyo 192-0397, Japan, [3]Brain Science Institute, Tamagawa University, Machida, Tokyo 194-8610, Japan, [4]School of Economics & Management, Kochi University of Technology, Kochi 780-8515, Japan, and [5]Research Institute for Future Design, Kochi University of Technology, Kochi 780-8515, Japan

Knowledge about one's personality, the self-concept, shapes human experience. Social cognitive neuroscience has made strides addressing the question of where and how the self is represented in the brain. The answer, however, remains elusive. We conducted two functional magnetic resonance imaging experiments (the second preregistered) with human male and female participants employing a self-reference task with a broad range of attributes and carrying out a searchlight representational similarity analysis (RSA). The importance of attributes to self-identity was represented in the medial prefrontal cortex (mPFC), whereas mPFC activation was unrelated both to self-descriptiveness of attributes (experiments 1 and 2) and importance of attributes to a friend's self-identity (experiment 2). Our research provides a comprehensive answer to the abovementioned question: The self-concept is conceptualized in terms of self-importance and represented in the mPFC.

Key words: fMRI; MVPA; self-concept; self-identity; social neuroscience

Significance Statement

The self-concept comprises beliefs about who one is as an individual (e.g., personality traits, physical characteristics, desires, likes/dislikes, and social roles). Despite researchers' efforts in the last two decades to understand where and how the self-concept is stored in the brain, the question remains elusive. Using a neuroimaging technique, we found that a brain region called medial prefrontal cortex (mPFC) shows differential but systematic activation patterns depending on the importance of presented word stimuli to a participant's self-concept. Our findings suggest that one's sense of the self is supported by neural populations in the mPFC, each of which is differently sensitive to distinct levels of the personal importance of incoming information.

Introduction

The sense of self shapes human experience (Sedikides et al., 2021). The self (or self-concept) consists of knowledge that people possess about the kind of person they are, such as traits, physical attributes, preferences, beliefs, values, or ingroup (Sedikides and Gregg, 2003). The self has been of keen interest to psychologists since the birth of the discipline (James, 1890). From the late 90s (Craik et al., 1999) onward, cognitive neuroscientists have been investigating the neural basis of the self (Wagner et al., 2019). However, where and how the self is represented in the brain remains elusive.

Past neuroimaging studies on the self have identified a network of brain regions, including medial prefrontal cortex (mPFC) and posterior cingulate cortex (PCC), that are consistently active during self-reference judgment compared with other semantic judgements (Denny et al., 2012; Murray et al., 2012). However, the approach of contrasting neural responses during the self-reference versus control tasks to unveil the neural basis of the self has several limitations.

First, activation observed by simply comparing the strength of neural responses between the self-reference and control tasks may be because of cognitive processes unrelated to the self (Gillihan and Farah, 2005; Legrand and Ruby, 2009) such as autobiographical memory (Martinelli et al., 2013) and positive

Received Nov. 24, 2022; revised Mar. 27, 2023; accepted Mar. 30, 2023.

Author contributions: R.A. and K.I. designed research; K.M. and K.I. performed research; M.L. and K.I. analyzed data; M.L. wrote the first draft of the paper; R.A., K.M., C.S., and K.I. edited the paper; R.A., C.S., and K.I. wrote the paper.

This research was supported by the Economic and Social Research Council (ESRC), the South Coast Doctoral Training Partnership Grant Number ES/P000673/1 as a PhD studentship (to M.L.). This research was also supported by the Japan Society for Promotion of Science (JSPS) KAKENHI Grant Number JP19K24680 (to K.I.) and the Ministry of Education, Culture, Sports, Science and Technology (MEXT) as part of Joint Research Program implemented at Tamagawa University Brain Science Institute, in Japan. We thank Atsushi Miyazaki, Mika Oosawa, and Maoko Yamanaka for their assistance with fMRI data collection.

The authors declare no competing financial interests.

Correspondence should be addressed to Keise Izuma at izuma.keise@kochi-tech.ac.jp.

https://doi.org/10.1523/JNEUROSCI.2178-22.2023

affect (Bartra et al., 2013). This limitation is at least partially addressed by recent functional magnetic resonance imaging (fMRI) studies that used a multivariate pattern analysis (MVPA; Chavez et al., 2017; Yankouskaya et al., 2017; Feng et al., 2018; Courtney and Meyer, 2020; Koski et al., 2020; Parelman et al., 2022). Yet, these studies have come short of documenting what information about the self is specifically being processed.

Second, exceptions notwithstanding (Rameson et al., 2010; Jenkins and Mitchell, 2011), the bulk of the literature has used only trait adjectives as experimental stimuli. However, this practice likely limits researchers' ability to identify the neural representations of the self. As stated above, the self includes not only personality traits, but also physical characteristics, preferences, aspirations, abilities, and social groups (Linville, 1985); thus, personality traits comprise a narrow subset of the self (see del Prado et al., 2007).

Third, most neuroimaging research has operationalized the self-concept in terms of trait self-descriptiveness. There is an infinite number of characteristics that can describe an individual (e.g., "I sleep everyday"), but just because an item is self-descriptive does not necessarily mean it is a part of the self-concept. Instead, the self might be represented in the brain in terms of personal importance of each characteristic (hereafter, self-importance or centrality). The relevance of taking into account self-importance when assessing the self has also been recognized by psychologists (Markus, 1977). That is, whether information will influence one's behavior depends on its personal importance (Markus, 1983). For example, if being a mother is important to an individual, her behaviors as a mother are likely to be different (e.g., more attentive, responsible, or consistent) from her behaviors in other, less self-defining roles (Deutsch et al., 1988). How self-important (central) a personality trait is affects how a person seeks (Sedikides, 1993) and remembers (Sedikides et al., 2016) information about themselves. Thus, it is likely that there is a dedicated neural system in the brain, which encodes the self-importance of incoming information (Markus and Wurf, 1987; Sedikides, 1995). Indeed, a few studies have forayed into self-importance in the brain, suggesting that mPFC activation is correlated with the importance of possessing a personality trait (D'Argembeau et al., 2012) and with the personal significance of autobiographical memories (Lin et al., 2016).

We addressed the question of where and how the self-concept is represented in the brain in two fMRI experiments. We used the self-reference task with a broad range of stimuli combined with a representational similarity analysis (RSA; Kriegeskorte et al., 2008) of fMRI data to test how mPFC activation patterns are related to self-importance as well as self-descriptiveness while controlling for other factors (see Materials and Methods).

Materials and Methods

Experiment 1

Participants

We recruited 32 right-handed undergraduate students from Tamagawa University. The students had no history of psychiatric disorders. We excluded data from four participants because of excessive head movement (>3 mm; one participant), because their response consistency in the fMRI tasks was close to chance (one participant), because of no variance in the postscan memory rating (one participant), and because their self-reference rating reliability was low (one participant). In regard to the last case, each participant completed the self-reference task three times, and we computed correlation coefficient across the three ratings. The average correlation of this fourth participant was 0.21, which was >3 SDs below the group average of $r = 0.72$ (SD = 0.14), suggesting very poor compliance with the task instructions and/or having a highly unstable self-concept. The final sample consisted of 28 participants (16 women, 12 men) aged 19–22 years (M_{age} = 19.84, SD_{age} = 0.86). Participants clicked a box to indicate their consent before the online questionnaires. Additionally, we obtained written consent from all participants before the fMRI experiment. The study was approved by both the University of Southampton and Tamagawa University ethics committees. We remunerated each participant with 5000 Japanese yen.

Table 1. Examples of items used in the experiments

Physical	Social	Attributes	Other
Female	Art club	Talkative	Film
Bad eyesight	Flower arrangement club	Smart	Twitter
Tall	Department of Engineering	Hate prawns	Christmas
Hay fever	XX High School graduate	Open-minded	School trip
Born in Tokyo	Female	Compassionate	Rain
Sweaty	Softball team	Good singer	Piano
Brown hair	Japanese	Dog person	Earthquake
Sensitive skin	Basketball club	Play guitar	
Right-handed	Buddhist	Family-oriented	
20 years old	Kochi University of Technology student	Like to discuss ideas	

We used 959 unique items across the two experiments. Items in the "Other" category were mainly prepared by an experimenter. The four categories (Physical, Social, Attributes, and Other) are based on the coding scheme by Cousins, 1989).

Experimental procedure

The experiment comprised the three following parts, which took place on three separate days: (1) first online questionnaire, (2) second online questionnaire, (3) fMRI experiment. We administered the two online questionnaires in an effort to create the stimulus set for the fMRI experiment, a stimulus set that covers as widely as possible the content of each participant's self-concept (see below). The first and second online questionnaires were separated by an average of 6.25 d (SD = 3.26). The second online questionnaire and fMRI experiment were separated by an average of 8.29 d (SD = 2.19).

First online questionnaire. The first online questionnaire is similar to the Twenty Statement Test (TST; Kuhn and McPartland, 1954). During the online questionnaire, we instructed participants to provide at least 30 characteristics by responding to the prompt "I______." To facilitate this task, we gave participants examples such as physical characteristics (e.g., I am tall), personality (e.g., I am social), likes or dislikes (e.g., food, music, artists), and groups to which they belonged (university, department, clubs).

Second online questionnaire. The second online questionnaire included a total of 80 items some of which were provided by participants during the first online questionnaire, and others were added by an experimenter. Items prepared by the experimenter were intended to dissociate self-descriptiveness, self-importance, and other factors described below. For example, "right-handed" was added with an aim to dissociate self-descriptiveness and self-importance. "School trip" was added with an aim to dissociate self-descriptiveness/self-importance and autobiographical memory. "Convenience store" was added with an aim to dissociate self-descriptiveness/importance and familiarity. We instructed participants to rate each item in terms of (1) self-descriptiveness (1 = not descriptive at all, 7 = very descriptive), (2) importance to their self-identity (1 = not important at all, 7 = very important), (3) valence (1 = very negative, 7 = very positive), (4) familiarity (1= not familiar at all, 7 = very familiar), and (5) autobiographical memory or the extent to which "each word/phrase brought back memories of your past when you saw it" (1 = it did not evoke any memory at all, 7 = it evoked very vivid memory).

Stimulus set preparation. Based on the ratings, we selected a final stimulus set of 40 items under the stipulation that the ratings not be highly intercorrelated (i.e., effects on neural activities be dissociable). Specifically, we randomly picked 40 out of the 80 items used in the second questionnaire and computed correlations across the following six ratings/characteristics of the randomly selected 40 items: (1) self-descriptiveness, (2) valence, (3) familiarity, (4) autobiographical memory, (5) number of characters, (6) whether the

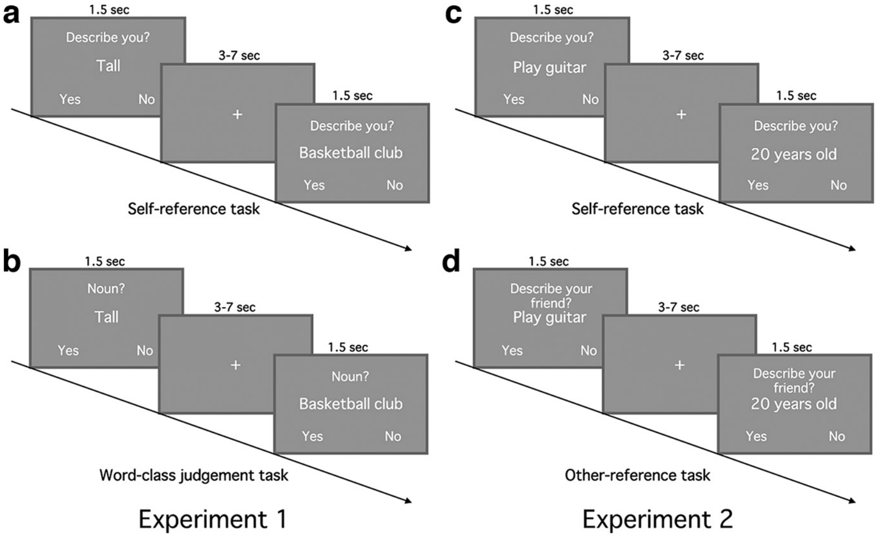


Figure 1. Experimental tasks. The self-reference task (***a***) and the word-class judgment task (***b***) in experiment 1. The self-reference task (***c***) and the other-reference task (***d***) in experiment 2.

item was provided by a participant during the first questionnaire (1) or not (0). We then recorded the highest correlation coefficient ($r_{highest}$). We repeated this process a large number of times (e.g., 1,000,000,000) and selected the final set of 40 items that had the lowest $r_{highest}$. For some participants, the self-descriptiveness and self-importance ratings were highly positively correlated. In such a case, we set a different criterion for that correlation. For instance, we computed $r_{highest}$ without considering $r_{self\text{-}descriptiveness/self\text{-}importance}$, and selected a set of 40 items whose $r_{highest}$ was the lowest given that $r_{self\text{-}descriptiveness/self\text{-}importance}$ was <0.6. For examples of items, see Table 1 (The coding scheme is based on Cousins, 1989).

fMRI experiment. Before the fMRI scan, participants received instructions regarding MRI safety and tasks they would perform inside the fMRI scanner. During the fMRI session, participants performed two tasks: self-reference and word-class judgment (Fig. 1*a*,*b*). We programmed both of them using Psychtoolbox (http://psychtoolbox.org/) with MATLAB software (version 2018a; http://www.mathworks.co.uk). Participants completed six runs of each task. Each run consisted of 40 trials, one item per trial. We presented the same set of 40 items in both tasks and in each run. We counterbalanced task (ABBAABBAABBA or BAABBAABBAAB), and randomized trial order within each run.

In the self-reference task, for each trial, participants viewed an item. On the screen, above the characteristic, they encountered the question "Describes you?" For each trial, they could answer "yes" or "no" to indicate whether the characteristic described them or not (Fig. 1*a*). For each trial in the word-class judgment task, participants viewed an item. On the screen, above the characteristic, they encountered the question "Noun?" They could answer "yes" or "no" to indicate whether the characteristic was a noun or not (Fig. 1*b*). For both tasks, each item was presented for 1.5 s, followed by intertrial interval (ITI; 3–7 s, mean = 5 s). Participants answered by pressing one of two buttons on a response box.

How vividly each item evoked a personal memory might differ between the second questionnaire and the fMRI task. Consequently, after the fMRI scan, we instructed participants to rate the same 40 items on autobiographical memory, namely, to what extent each item evoked an autobiographical memory when seeing it inside the fMRI scanner (1 = it did not evoke any memory at all, 7 = it evoked very vivid memory). Furthermore, to check for consistency of the self-descriptiveness judgment, we instructed participants to rate the 40 items again on self-descriptiveness using the same response scale. Next, participants completed a demographic questionnaire.

fMRI data acquisition. We acquired images using a 3-T Trio A Tim MRI (Siemens) scanner with a 32-channel head coil. For functional imaging, we used T2*-weighted gradient-echo echo-planar imaging (EPI) sequences with the following parameters: time repetition (TR) = 2500 ms, echo time (TE) = 25 ms, flip angle (FA) = 90°, field of view (FOV) = 192 mm^2, matrix = 64 × 64. We acquired, in an interleaved order, 42 contiguous slices with a thickness of 3 mm. In addition, we acquired a T1-weighted structural image from each participant.

Statistical analysis

Behavioral data analysis. Each participant rated each of the 40 items on self-descriptiveness, and they did so three times: (1) during the second online questionnaire, (2) during the fMRI scan, (3) after the fMRI scan. Although participants rated each item six times (across six fMRI runs) on a two-point scale (yes or no) during the fMRI scan, they rated each item once on a seven-point scale during the second questionnaire and post-fMRI rating task. Thus, for the self-descriptiveness rating data obtained during the fMRI scan, we computed a self-descriptiveness score for each item as a proportion of yes responses across six ratings of each item. We assumed that participants maintained a stable self-concept across a few weeks, and we tested this assumption by checking for consistency of their self-descriptiveness ratings obtained across the three times (or sessions).

fMRI data preprocessing. We conducted preprocessing and statistical analysis of the fMRI data using SPM12 (Welcome Department of Imaging Neuroscience), implemented in MATLAB (MathWorks). We discarded the first four volumes before preprocessing and data analyses to allow for T1 equilibration. We conducted preprocessing of the fMRI data with SPM 12's preproc_fmri.m script starting with realignment of all functional images to a common image. We spatially realigned all images within each run to the first volume of the run using seventh-degree B-spline interpolation, and we unwarped and corrected for motion artefacts. We segmented the T1-weighted structural image and normalized it into a common stereotactic space (MNI atlas). Subsequently, we applied the normalization parameters to the functional images and resampled them to 3 × 3 × 3 mm^3 isotropic voxels (i.e., original voxel size was retained) using seventh-degree B-spline interpolation. Following the normalization, we spatially smoothed the data [with a Gaussian kernel of 8-mm full-width at half-maximum (FWHM)] for the univariate analysis. To maintain fine grained activation patterns, we did not apply smoothing before the first-level data analysis for the RSA. We applied smoothing before the group analysis of the RSA outputs to account for individual variability in brain structure (with a Gaussian kernel of 4-mm FWHM).

fMRI data analysis: univariate analysis. We used three general linear models (GLMs) to analyze the fMRI data. In the first GLM, we compared the two conditions (self vs word), whereas we used the spmT images from the second GLM for the RSA. In the first GLM, we separately modeled 40 self-reference judgment trials and 40 word-class judgment trials using a box-car function convolved with the canonical hemodynamic response function.

In the second GLM, we investigate whether mPFC activities parametrically increase as a function of self-importance and/or self-descriptiveness ratings. As in the first GLM, we separately modeled 40 self-reference trials and 40 word-class judgment trials. In addition, we added to each of the self and word trial regressors the following seven parametric regressors: (1) self-descriptiveness, (2) self-importance, (3) valence, (4) familiarity, (5) autobiographical memory, (6) word-length, (7) whether each item was self-provided (1) or not (0). Given that SPM automatically performs orthogonalization for parametric regressors (see Mumford et al., 2015), we also tried another GLM where the order of self-descriptiveness and self-importance parametric regressors were switched (self-importance as the first parametric regressor, and self-descriptiveness as the second parametric regressor), but the results were

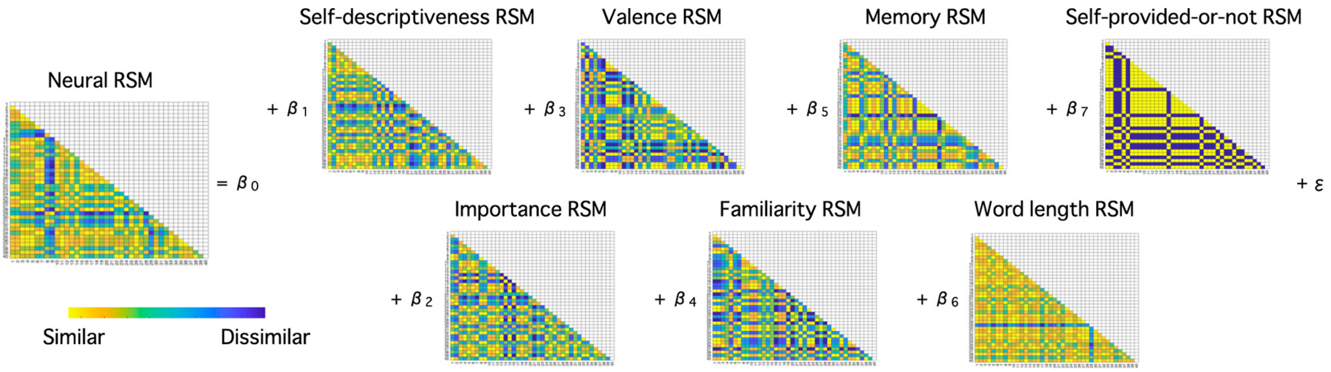


Figure 2. Neural and model RSMs in the multiple regression analyses in experiment 1. For each participant, we made seven model RSMs based on participant ratings and item characteristics. The value in each cell indicates a similarity between a pair of items on a given dimension. We made a neural RSM for each searchlight. We conducted a multiple regression analyses for each searchlight with the seven model RSMs.

virtually the same. For the first two GLMs, we submitted the contrast images to a second level analysis. We set statistical threshold at $p < 0.005$ with cluster-$p < 0.05$ [familywise error (FEW) corrected] within the mPFC mask. Outside of the mask, we set up the statistical threshold at $p < 0.001$ (uncorrected for multiple comparisons) with a cluster threshold of $p < 0.05$ (FWE corrected).

In the third GLM, we modeled separately each of the 40 items for each task. We used a total of 80 spmT images from the third GLM in subsequent RSA. In all the GLMs, we included six head motion parameters and session effects as nuisance regressors.

Representational similarity analysis (RSA): model representational similarity matrix (RSM)

To test the effect of self-descriptiveness and self-importance on neural responses in the mPFC, while controlling for other factors (i.e., valence, familiarity, autobiographical memory, word-length, whether items were self-provided or not), we used RSA with a searchlight approach. For each participant, we calculated a model RSM separately for the following seven dimensions: (1) self-descriptiveness, (2) self-importance, (3) valence, (4) familiarity, (5) autobiographical memory, (6) word-length, (7) whether each item was self-provided (1) or not (0). For self-descriptiveness, self-importance, valence, and familiarity, we used ratings from the second online questionnaire. For autobiographical memory, we used ratings from the postscan behavioral session. Each RSM was a 40 × 40 matrix (Fig. 2), where each cell represented the similarity of the ratings between two items. For ratings completed on a seven-point scale, we calculated similarity as seven minus the absolute difference between two ratings. For the word length, we calculated similarity as the maximum number of characters in the 40 items minus the absolute difference in the number of characters between two items. Lastly, for whether items were self-provided or not, we coded similarity as 0 if an item was provided by the participant but the other item was not (or vice versa), whereas we coded similarity as 1 otherwise (i.e., both items were provided by the participants or both items were provided by the experimenter). We standardized the values with the respective mean and SD for each rating before regression analyses.

RSA: neural RSM. We extracted local patterns of neural activity from searchlights with a three-voxel radius, so that each searchlight consisted of a maximum of 123 voxels (and less on the edges of the brain). For each searchlight, we calculated voxel-by-voxel correlations between each pair of the 40 items, which resulted in a 40 × 40 neural RSM (Fig. 2). Consequently, the correlation in neural activities between two items within a searchlight was represented by a cell in the respective neural RSM. We Fisher-Z transformed the correlation values before further analyses.

RSA: multiple regression analysis. In each searchlight, we conducted a multiple regression analysis, where the seven model RSMs were independent variables and the neural RSM was dependent variable (Fig. 2). We repeated this analysis for every searchlight across the mPFC region of interest (ROI; see below for more information about the mPFC mask applied) and the whole brain, resulting in a β-map for each of the seven independent variables for each of the two tasks (i.e., a total of 14 β-maps for each participant).

RSA: group analysis. We entered the β-maps into a group-level analysis that we computed with permutation testing (i.e., one-sample *t* test with 5000 permutations) using the Statistical NonParametric Mapping (SnPM) toolbox for SPM (Nichols and Holmes, 2002). Among several brain regions previously implicated in self-processing (Qin and Northoff, 2011; Denny et al., 2012; Murray et al., 2012), we focused especially on the mPFC, as a lesion study indicated that this region is necessary for a stable and accurate self-concept (but is not critical for knowledge about another person; Marquine et al., 2016). Accordingly, we applied an mPFC mask to the analysis to limit the group-analysis to voxels within the a priori ROI. We created the mask with the WFU PickAtlas toolbox for SPM (Maldjian et al., 2003). The mPFC ROI mask included Frontal_Sup_medial_L, Frontal_Sup_medial_R, Frontal_Mid_Orb_L, Frontal_Mid_Orb_R, Rectus_L, Rectus_R, Cingulum_Ant_L, and Cingulum_Ant_R (dilation factor = 2), which we took from the Anatomical Automatic Labeling (AAL) masks implemented in the WFU pickatlas toolbox. We applied the same statistical threshold as the univariate analyses above [within the mPFC mask, $p < 0.005$ with cluster-$p < 0.05$ (FWE corrected), and outside the mask, $p < 0.001$ with cluster-$p < 0.05$ (FWE corrected)].

Experiment 2

Preregistration

Before data analyses, we preregistered the hypotheses, sample size, data analytic plan, and exclusion criteria on the Open Science Framework (https://osf.io/agq3b). We followed the preregistration in all analyses reported below, unless otherwise noted.

Participants

As preregistered, final analyses included a sample of 35 undergraduate students (23 men, 12 women) at Kochi University of Technology ($M_{age} = 19.66$, $SD_{age} = 1.64$, range = 18–23 years). All of them were right-handed, and none had a history of psychiatric disorders. We scanned eight additional participants but excluded them from the final analyses, because they did not meet the preregistered inclusion criteria. In particular, we excluded one participant because of a brain anomaly, and seven participants because the reliability of either their self-reference or other-reference rating was low. In regard to these seven participants, we calculated correlations between their responses in the self-reference task during fMRI scanning and their self-descriptiveness rating in the second questionnaire, as well as the correlation between their other-reference task responses during fMRI scanning and their friend-descriptiveness rating in the second questionnaire. We considered the correlation value low, if it was <0.5. All participants ticked a box to indicate their consent before the online questionnaires, and they consented in writing before the fMRI experiment. The study was approved by the Kochi University of Technology ethics committee. They were remunerated with 2000 Japanese yen.

Power analysis

We conducted a power analysis using the Bootstrap procedure. First, we randomly sampled 35 participants from the 28 participants of experiment 1, with replacement. For each randomly-selected sample, we conducted a group analysis (one-sample *t* test). Given that in experiment 1 we found significant activations within the mPFC, we applied the same mPFC mask created via the WFU PickAtlas toolbox for SPM (Maldjian et al., 2003). We applied a voxel-wise threshold of $p < 0.005$ (uncorrected) and cluster-$p < 0.05$ (FWE corrected) to assess significance. We repeated these steps 2000 times, and counted the number of times we found significant activations within the mPFC mask. The result indicated that the sample size of $n = 35$ would achieve power of 91.85%.

Experimental procedure

The procedure was similar to that of experiment 1, consisting of three parts (two online questionnaires, fMRI experiment) on three separate days. The only alteration involved the control task. To examine whether the experiment 1 results were specific to the self, or whether the mPFC also encodes important information for a friend's identity, we used an other-reference task as control (see Fig. 1*c*,*d*). The first and second online questionnaires were separated by an average of 11.0 d (SD = 6.68). The second online questionnaire and fMRI experiment were separated by an average of 14.03 d (SD = 8.34).

Online questionnaires. As in experiment 1, in the first online questionnaire, participants provided at least 30 characteristics by responding to the prompt "I ____." Similarly, participants provided the name of a close friend and at least 30 characteristics they believed to be descriptive of or important for that friend. They did so by responding to the prompt "My friend _____."

In the second online questionnaire, participants rated 80 items, some of which were made available by participants during the first online questionnaire. In particular, they rated each item on the following seven dimensions: (1) self-descriptiveness (1 = not at all descriptive, 7 = very descriptive), (2) importance to self-identity (1 = not at all important, 7 = very important), (3) friend self-descriptiveness (1 = not at all descriptive, 7 = very descriptive), (4) importance to friend's identity (1 = not at all important, 7 = very important), (5) valence (1 = very negative, 7 = very positive), (6) familiarity (1 = not at all familiar, 7 = very familiar), and (7) autobiographical memory (1 = it did not evoke any memory at all, 7 = it evoked very vivid memory). We selected a stimulus set of 40 items as in experiment 1 (for item examples, see Table 1).

fMRI experiment. During the fMRI session, participants conducted the self-reference and other-reference tasks (Fig. 1*c*,*d*). We used the same set of 40 items for both tasks.

Just like in experiment 1, during the self-reference task, for each trial, the participants viewed one of the 40 items. On the screen, above the characteristic, they saw the question "Describes you?" For each trial, they answered "yes" or "no" to indicate whether the characteristic described them (Fig. 1*c*). For each trial in the other-reference task, participants similarly viewed an item on the screen. Above the item, they saw the question "Describes your friend?" and answered "yes" or "no" to indicate whether the characteristic described their friend, the same close friend they mentioned during the first online questionnaire (Fig. 1*d*). For both tasks, each item was presented for 1.5 s, followed by intertrial interval (ITI; 3–7 s, mean = 5 s). Participants indicated their answers by pressing one of two buttons on a response box.

Postscan behavioral session. After the scan, participants rated the previously presented words on autobiographical memory again (1 = it did not evoke any memory at all, 7 = it evoked very vivid memory). Next, they completed a demographic questionnaire.

fMRI data acquisition. We acquired images using a Siemens 3.0 T Verio MRI scanner with a 64-channel phased array head coil. For functional imaging, we used T2*-weighted gradient-echo echo-planar imaging (EPI) sequences with the following parameters: time repetition (TR) = 2500 ms, echo time (TE) = 25 ms, flip angle (FA) = 90°, field of view (FOV) = 192 mm^2, matrix = 64 × 64. We acquired 42 contiguous slices with a thickness of 3 mm, in an interleaved order. Moreover, we acquired from each participant a high resolution anatomic T1-weighted image (1-mm isotropic resolution).

Experimental procedure

fMRI data processing. We conducted preprocessing of fMRI data as in experiment 1. The preprocessing described in our preregistration stated that we would use an EPI-template when normalizing fMRI data to the standard MNI space. Although, based on visual inspection of normalized images, there was no issue with this method when analyzing the fMRI data from experiment 1, we noticed that fMRI images normalized with this method were consistently smaller in the anterior-to-posterior and left-to-right dimensions (possibly because of the difference in head-coil between the two experiments; 32 channels in experiment 1 vs 64 channels in experiment 2 (for a similar case, see Smith et al., 2018). Accordingly, we decided to use a T1-template when normalizing the fMRI data as implemented in the SPM 12's preproc_fmri.m script. For the sake of consistency, we re-analyzed fMRI data in experiment 1 with this new preprocessing steps, as reported above. In experiment 1, we report the re-analyzed data (note that the two preprocessing steps generated virtually identical results).

Univariate fMRI analysis. Similar to experiment 1, we used three GLMs. In the first GLM, we intended to compare the two conditions (self vs other). In the second GLM (not preregistered), we intended to test whether mPFC activities increase parametrically as a function of self-descriptiveness and self-importance ratings. Finally, we used the spmT images from the third GLM for the RSA.

In the first GLM, we separately modeled 40 self-reference judgment trials and 40 other-reference trials using a box-car function convolved with the canonical hemodynamic response function. In the second GLM, as in the first one, we separately modeled 40 self-reference trials and 40 other-reference judgment trials. In addition, we added the following nine parametric regressors to each of the self-reference and other-reference trial regressors: (1) self-descriptiveness, (2) self-importance, (3) friend-descriptiveness, (4) friend-importance, (5) valence, (6) familiarity, (7) autobiographical memory, (8) word-length, (9) whether the item was self-provided or not.

For the first two GLMs, we submitted the contrast images to a second level analysis. As preregistered, we employed the same mPFC mask as in experiment 1, and within the mPFC mask we set the statistical threshold at $p < 0.005$ (uncorrected for multiple comparisons) with a cluster threshold of $p < 0.05$ (FWE corrected). Outside of the mask, we set the statistical threshold at $p < 0.001$ (uncorrected for multiple comparisons) with a cluster threshold of $p < 0.05$ (FWE corrected).

In the third GLM, we modeled separately each of the 40 items for each task. In all three GLMs, we included six head motion parameters and session effects as nuisance regressors.

Model RSMs. We conducted searchlight RSA as in experiment 1. However, in addition to testing the effect of self-descriptiveness and self-importance, we tested the effect of friend-descriptiveness and friend-importance, on neural representations. So, for each participant, we calculated a model RSM separately for each of the following nine dimensions: (1) self-descriptiveness, (2) self-importance, (3) friend-descriptiveness, (4) friend-importance, (5) valence, (6) familiarity, (7) autobiographical memory, (8) word-length, and (9) whether the item was self-provided or not. For self-descriptiveness, self-importance, friend-descriptiveness, friend-importance, valence, and familiarity, we used the ratings from the second questionnaire. For autobiographical memory, we used the ratings from the postscan behavioral session.

Neural RSM. We created a neural RSM for each searchlight as in experiment 1.

RSA: multiple regression analysis. In each searchlight, we conducted a multiple regression analysis where the nine model RSMs were independent variables and the neural RSM was the dependent variable. We repeated the analysis for every searchlight across the brain, resulting in a β-map for each of the nine independent variables and each of the two tasks [a total of 18 (2 × 9) β-maps for each participant]. Although not preregistered, we attempted another RSA by adding a model RSM based on participants' average RT for each item (a total of 10 model RSMs), and this additional RSA produced results virtually identical to those reported below.

RSA: group analysis. We conducted the second-level group analysis as in experiment 1 (i.e., using SnPM). We applied the same statistical threshold as the univariate analyses above [within the mPFC mask, $p < 0.005$ with cluster-$p < 0.05$ (FWE corrected), and outside the mask, $p < 0.001$ with cluster-$p < 0.05$ (FWE corrected)].

Classifier-based MVPA (not preregistered). We also conducted a classifier-based MVPA analysis that directly compares the effects of self-importance and friend-importance on mPFC activation. We did so in search for evidence that a neural code for information importance is unique to the self. Specifically, we tested whether, during the other-reference task, the mPFC activation patterns evoked by items high (also middle or low) in self-importance task are distinct from activation patterns evoked by items high (also middle or low) in friend-importance.

Table 2. Average within-Person correlations (SD) across the three self-descriptiveness ratings in experiment 1

	Second questionnaire	fMRI task	Post-fMRI rating
Second questionnaire	—		
fMRI task	0.83 (0.09)***	—	
Post-fMRI rating	0.72 (0.14)***	0.67 (0.12)***	—

Each participant rated each of the 40 items on self-descriptiveness three times: (1) during the second online questionnaire, (2) during the fMRI scan, and (3) after the fMRI scan. ***$p < 0.001$ (corrected for multiple comparisons) based on one sample t test (one-tailed; correlation coefficients were Fisher-z transformed before the t tests).

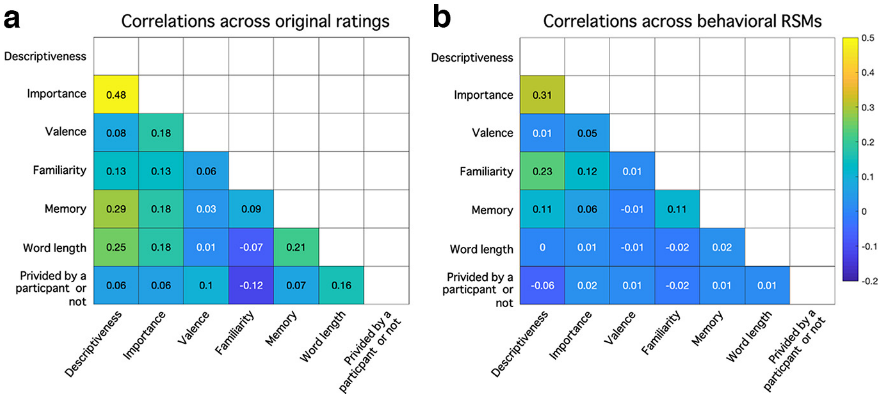


Figure 3. Average correlations across seven ratings (***a***) and seven model RSMs (***b***) in experiment 1.

First, we conducted another GLM analysis where each item was classified into one of the three categories depending on level of self-(and friend-)importance: high, middle, low. Given that the distribution of ratings was different across participants (e.g., with some frequently providing ratings of 6–7, and others frequently providing ratings of 1–2), we used different criteria for different participants when classifying each item into the three categories so that the three categories included roughly an equal number of items. Of note, within each participant, we used the same criterion for the self and other conditions. Thus, in this GLM, when modeling the fMRI data from the self-reference task, we classified 40 items into three categories based on self-importance ratings: (1) self-importance-high, (2) self-importance-middle, (3) self-importance-low. We modeled separately items in each of the three categories. Similarly, for the other-reference task fMRI data, we classified items into three categories in the same way based on the friend-importance ratings (high, middle, or low), and we modeled separately items in each of the three categories. We included six head motion parameters and session effects as nuisance regressors. We then computed an spmT map for each category per fMRI run resulting in 2 (tasks; self vs other) × 3 (level of importance) × 6 (runs) spmT images per participants, which we used in the subsequent MVPA.

To define independently a self-importance related mPFC ROI, we used a leave-one-participant-out cross-validation procedure (Esterman et al., 2010). We re-ran the second-level group analysis (the searchlight RSA group analysis described above) 34 times with a different single participant left out in each. We used each second-level analysis to determine an mPFC ROI for each left-out participant. For each participant, we extracted data from a three-voxel radius sphere surrounding the peak voxel within the mPFC most strongly associated with self-importance ratings. To ascertain that each participant's ROI was roughly from the same anatomic subregion within the mPFC, we searched a peak voxel for each participant within a 30-mm sphere surrounding the peak voxel identified by the group analysis with all 35 participants.

We used a linear support vector machine, which we conducted using Matlab in combination with LIBSVM (https://www.csie.ntu.edu.tw/~cjlin/libsvm/; Wake and Izuma, 2017) with a cost parameter of c = 1 (default). We paired each of the self run and friend run in the order of acquisition, and evaluated classification performances with a leave-one-pair-out cross-validation procedure. Thus, using the spmT images from the five runs of each task, we trained a classifier that discriminates activation patterns between self-importance-high versus friend-importance-high items. Then, using the spmT images from the left-out run of each task, we tested whether the classifier could discriminate between self-importance-high versus friend-importance-high items. We repeated the procedure six times so that each run-pair served as the testing set once. We averaged six classification accuracy values for each participant. We conducted the same analysis to test whether activation patterns are distinct between items low in self-importance versus items low in friend-importance (also, items middle in self-importance vs middle in friend-importance). We assessed statistical significance with permutation tests where we performed classifications with scuffled labels 1000 times to obtain a null distribution; p-values were set at 0.05 (one-tailed) and Bonferroni-corrected for three comparisons.

Data availability

Unthresholded group-level statistical maps and the mPFC mask image are available on NeuroVault (https://neurovault.org/collections/13069/).

Results

Experiment 1

Behavioral results

Participants rated each of the 40 items on self-descriptiveness three times: (1) during the second online questionnaire, (2) during the fMRI scan, (3) after the fMRI scan. Their responses were highly consistent across the three sessions (average within-individual correlation = 0.74; Table 2). This finding supports our assumption that participants' self-concept was stable over the weeks of testing.

We present average correlations between the behavioral ratings (self-descriptiveness, self-importance, valence, familiarity, autobiographical memory, word-length, whether items were self-provided) from the second questionnaire in Figure 3*a* (note that memory ratings were from the second memory rating task after the fMRI scan). Similarly, we present average correlations across the seven model RSMs in Figure 3*b*. The correlation between the self-descriptiveness and self-importance model RSMs was the highest [average $r = 0.31$ (SD = 0.26)]. We also calculated and checked the variance inflation factors (VIFs) for the seven independent variables within each participant. VIF provides an index of the degree to which the variance of a coefficient is increased because of collinearity, with values of above 10 often considered problematic. Across a total of 196 (7 variables × 28 participants) VIFs, 193 of them were below 2, and the maximum VIF was

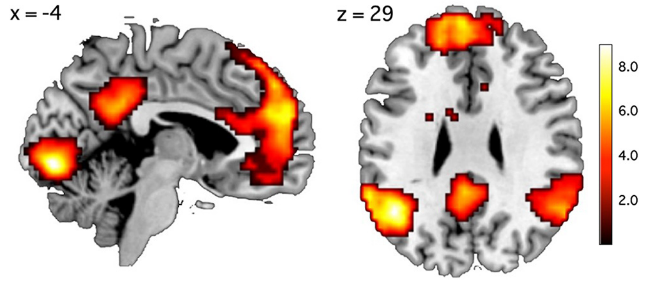


Figure 4. Group activation map for the self versus word contrast in experiment 1. For a display purpose, we set the voxel-wise threshold at $p < 0.005$ (uncorrected) and set the cluster size threshold at $p < 0.05$ (FWE corrected). For all activated areas, see Table 3.

Table 3. Brain regions showing significant activations during the self-reference task and the word class judgment task in experiment 1

		MNI coordinates				
Contrast	Location	*x*	*y*	*z*	Z	Cluster size (voxels)
Self > word	dmPFC	−9	41	53	5.62	1605
	amPFC	−3	50	23	5.42	
	dACC	6	20	20	4.19	
	MFG	−33	20	38	4.68	171
	Left STS	−60	−22	−10	5.55	736
	PCC	−3	−46	29	4.82	370
	Left TPJ	−45	−58	29	6.20	580
	Right TPJ	57	−58	29	5.83	339
	Lingual gyrus	−3	−85	−4	6.37	722
Word > self	Left IFG	−48	32	20	4.69	519

The statistical threshold was set at $p < 0.005$ (uncorrected for multiple comparisons) with a cluster threshold $p < 0.05$ (FWE corrected). dmPFC, dorsomedial prefrontal cortex; amPFC, anterior medial prefrontal cortex; dACC, dorsal anterior cingulate cortex; MFG, middle frontal gyrus; STS, superior temporal sulcus; PCC, posterior cingulate cortex; TPJ, temporoparietal junction; IFG, inferior frontal gyrus. Voxel size = 3 × 3 × 3 mm.

3.03, indicating reasonable ability to draw inferences on the unique variance explained by each variable in all participants.

fMRI results

Univariate analysis. Successfully replicating the previous studies (Qin and Northoff, 2011; Denny et al., 2012; Murray et al., 2012), we found that the self-reference versus word-class judgment contrast significantly activated the mPFC and PCC (Fig. 4). Other activated regions included left and right temporoparietal junction (TPJ), left superior temporal sulcus (STS), and lingual gyrus (Fig. 4; Table 3). The opposite contrast (word vs self) activated the left inferior frontal gyrus (IFG), which is known to play a major role in language processing (Ferstl et al., 2008; Table 3).

Parametric modulation analysis. Next, we tested whether mPFC activities parametrically increase as a function of self-descriptiveness and/or self-importance. Contrary to previous studies (Macrae et al., 2004; Moran et al., 2006; D'Argembeau et al., 2012; Koski et al., 2020; Elder et al., 2022), neither self-descriptiveness nor self-importance ratings were significantly associated with mPFC activations during the self-reference task. We note that other studies with a similar event-related design did not report results relevant to this association (Kelley et al., 2002; Ochsner et al., 2005; Heatherton et al., 2006; Yaoi et al., 2015), although they could have done so. Furthermore, the results of a recent study with direct neural recordings from human participants using electrocorticography (ECoG; Tan et al., 2022) suggest that the linear relationship between self-descriptiveness and mPFC activations is, if anything, small.

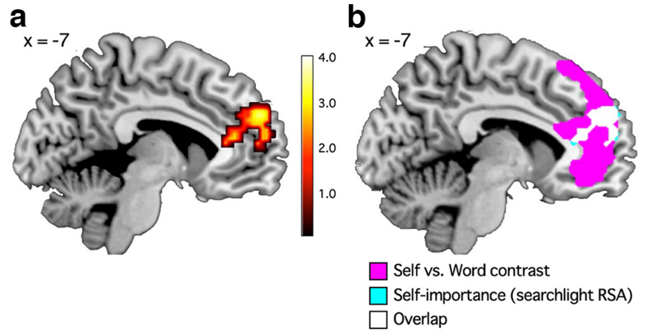


Figure 5. Searchlight RSA result in experiment 1. ***a***, Self-importance was significantly associated with activation patterns within the mPFC during the self-reference task (see also Table 4). $p < 0.005$ (uncorrected) and cluster-$p < 0.05$ (FWE corrected). ***b***, The mPFC areas significantly associated with self-importance (cyan) largely overlapped with the areas activated by the self-reference task compared with the word class judgment task (magenta; Fig. 4).

Table 4. mPFC regions from searchlight RSA showing significant association with self-importance ratings during the self-reference task in experiment 1

	MNI coordinates				
Location	*x*	*y*	*z*	Z	Cluster size (voxels)
dmPFC	−9	53	29	3.84	306
dACC	12	38	23	3.47	
amPFC	6	56	14	2.89	

The statistical threshold was set at $p < 0.005$ (uncorrected for multiple comparisons) with a cluster threshold $p < 0.05$ (FWE corrected). Voxel size = 3 × 3 × 3 mm.

Both self-descriptiveness and self-importance ratings were unrelated to mPFC activities during the word task.

Searchlight RSA result. We conducted searchlight RSA within the mPFC ROI to test whether self-descriptiveness or self-importance had an effect on the local patterns of activation within the mPFC. We found that different levels of self-importance were represented by different patterns of activation within the mPFC (medial superior frontal gyrus; $x = -9$, $y = 53$, $z = 29$, 306 voxels) during the self-reference task (Fig. 5; Table 4). However, self-descriptiveness was not significantly associated with activation patterns within the mPFC. Likewise, the remaining five variables were not significantly associated with mPFC activations. The mPFC region associated with self-importance (Fig. 5*a*) largely overlapped with the mPFC region activated by the self versus word contrast (Fig. 4). Out of the 306 voxels whose activities were significantly associated with self-importance, 194 voxels (63.3%) were included in the area significantly activated by the self-reference task compared with the word task (Fig. 5*b*).

We repeated the same searchlight RSA using the data from the word class judgment task, and found no significant results. The null effects indicate that the representation of self-importance within the mPFC is task dependent. Self-importance is represented within the mPFC only when performing a task that requires thinking about the self.

Outside of the mPFC ROI, different levels of word-length were represented by different patterns of activation within the visual cortex (lingual-gyrus) for both the word-class judgment task and the self-reference task, indicating that visually similar stimuli evoke similar activation patterns in the visual cortex regardless of task. No other significant results emerged.

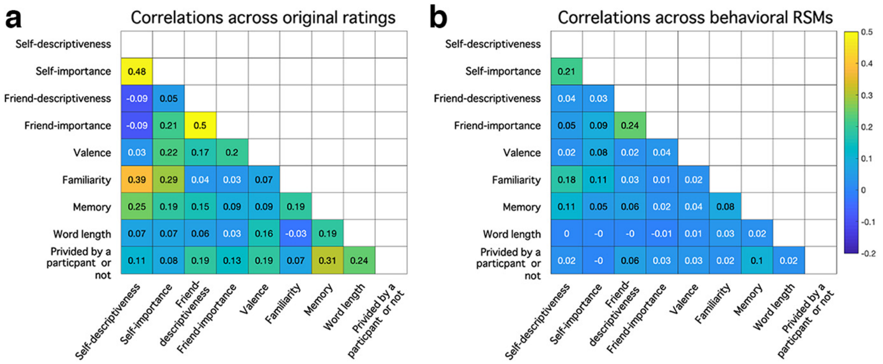
Taken together, we obtained initial evidence that the mPFC represents self-importance information. However, it is possible that the mPFC represents importance not specific to self-identity, but relevant to another person's identity as well; that is, the mPFC may not be specific to the self, but instead process person information in general. We addressed this possibility in experiment 2.

Experiment 2

Behavioral results

Participants rated each of the 40 items on self-descriptiveness and friend-descriptiveness twice: (1) during the second online questionnaire, (2) during the fMRI scan. Their responses were highly consistent across the two sessions. Average within-individual correlation for self-descriptiveness ratings was 0.78 (SD = 0.09), and average within-subject correlation for friend-descriptiveness was 0.74 (SD = 0.09).

We present average correlations between the behavioral ratings (self-descriptiveness, friend's self-descriptiveness, self-importance, friend's self-importance, valence, familiarity, autobiographical memory, word-length, whether items self-provided) from the second questionnaire in Figure 6. We checked the VIFs for the nine independent variables within each participant. Results showed that all 315 (9 variables × 35 participants) VIFs were below 2, indicating reasonable ability to make inferences on the unique variance explained by each variable in all participants.



Correlations across original ratings

	Self-descriptiveness	Self-importance	Friend-descriptiveness	Friend-importance	Valence	Familiarity	Memory	Word length	Privided by a particpant or not
Self-descriptiveness									
Self-importance	0.48								
Friend-descriptiveness	-0.09	0.05							
Friend-importance	-0.09	0.21	0.5						
Valence	0.03	0.22	0.17	0.2					
Familiarity	0.39	0.29	0.04	0.03	0.07				
Memory	0.25	0.19	0.15	0.09	0.09	0.19			
Word length	0.07	0.07	0.06	0.03	0.16	-0.03	0.19		
Privided by a particpant or not	0.11	0.08	0.19	0.13	0.19	0.07	0.31	0.24	

Correlations across behavioral RSMs

	Self-descriptiveness	Self-importance	Friend-descriptiveness	Friend-importance	Valence	Familiarity	Memory	Word length	Privided by a particpant or not
Self-descriptiveness									
Self-importance	0.21								
Friend-descriptiveness	0.04	0.03							
Friend-importance	0.05	0.09	0.24						
Valence	0.02	0.08	0.02	0.04					
Familiarity	0.18	0.11	0.03	0.01	0.02				
Memory	0.11	0.05	0.06	0.02	0.04	0.08			
Word length	0	-0	-0	-0.01	0.01	0.03	0.02		
Privided by a particpant or not	0.02	-0	0.06	0.03	0.03	0.02	0.1	0.02	

Figure 6. Average correlations across nine ratings (***a***) and nine model RSMs (***b***) in experiment 2.

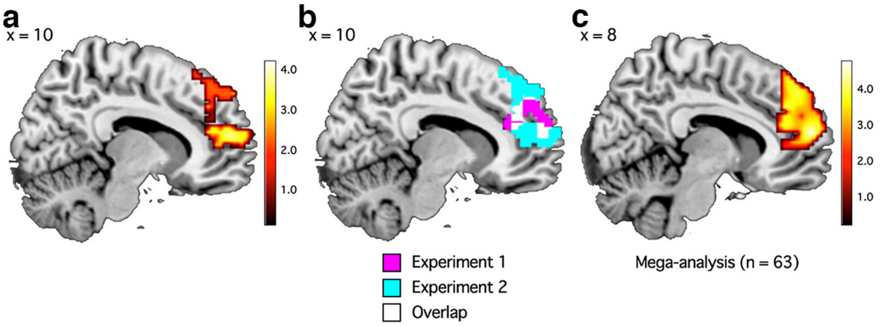


Figure 7. Searchlight RSA result in experiment 2 (***a***), activation overlaps between experiments 1 and 2 (***b***), and mega-analysis results (***c***). ***a***, Self-importance was significantly associated with activation patterns within the mPFC in experiment 2 (see also Table 5). $p < 0.005$ (uncorrected) and cluster-$p < 0.05$ (FWE corrected). ***b***, mPFC areas associated with self-importance in experiment 2 (magenta) overlap with areas associated with self-importance in experiment 1 (cyan). ***c***, Mega-analysis results ($n = 63$) showing an mPFC cluster that is significantly associated with self-importance (see also Table 7). $p < 0.005$ (uncorrected) and cluster-$p < 0.05$ (FWE corrected).

fMRI results

Univariate analysis. The self-reference versus other reference contrast did not reveal significant activation within the mPFC or across the whole brain. Although this result is in contrast to our preregistered hypothesis, previous studies have generated mixed findings regarding the difference between the self and other conditions, and our finding is consistent with studies that reported no difference (Schmitz et al., 2004; Ochsner et al., 2005; Vanderwal et al., 2008; Benoit et al., 2010; Tan et al., 2022). The opposite contrast also did not reveal significant activation in any region.

Parametric modulation analysis. We investigated whether mPFC activities parametrically increase as a function of self-descriptiveness and/or self-importance. However, as in experiment 1, neither self-descriptiveness nor self-importance ratings were significantly associated with mPFC activations during the self-reference task. Similarly, neither friend-descriptiveness nor friend-importance were significantly associated with mPFC activations during the other-reference task.

Searchlight RSA within the mPFC ROI. Based on our preregistered hypothesis that self-importance is encoded in the mPFC, we first limited the search area to within the mPFC by applying the anatomic mPFC mask. We conducted searchlight RSA to test whether self-importance information is represented in areas within the mPFC during self-reference task. As hypothesized, self-importance was reliably signaled in the mPFC during the self-reference task ($x = 3$, $y = 41$, $z = 50$; 280 voxels; Fig. 7*a*; Table 5). This mPFC cluster overlapped with the mPFC cluster related to self-importance in experiment 1, although the overlap was relatively small (a total of 25 voxel; Jaccard index = 0.046; Fig. 7*b*).

Table 5. mPFC regions from searchlight RSA showing significant association with self-importance during the self-reference task in experiment 2

	MNI coordinates				
Location	*x*	*y*	*z*	Z	Cluster size (voxels)
dmPFC	3	41	50	3.65	280
amPFC	9	53	14	3.43	
dACC	9	41	17	3.02	

The statistical threshold was set at $p < 0.005$ (uncorrected for multiple comparisons) with a cluster threshold $p < 0.05$ (FWE corrected). Voxel size = 3 × 3 × 3 mm.

In contrast, self-descriptiveness was not encoded in the mPFC during the self-reference task. Furthermore, neither self-importance nor self-descriptiveness were encoded in the mPFC during the other-reference task. These results replicate those of experiment 1. Processing of information about how important each stimulus is to the self in the mPFC is task dependent, and its neural representations emerge only when performing a task that requires thinking about the self.

In contrast, both friend-descriptiveness and friend-importance were not significantly associated with mPFC activation during the self-reference and other-reference tasks. Likewise, the remaining five variables were not significantly related to mPFC activation during either task.

Whole-brain searchlight RSA. We performed searchlight RSA throughout the whole-brain. Consistent with experiment 1, we found that different levels of word-length were represented by

Table 6. Classifier-based MVPA results

Comparison	Classification performance	P_{perm} (uncorrected)
High importance	50.48%	0.40
Middle importance	62.62%	$<$0.001**
Low importance	55.71%	0.003*

*$p < 0.05$ and **$p < 0.01$ (Bonferroni corrected). Significant classification performance means that activation patterns were distinct in the self versus other conditions for a given importance level.

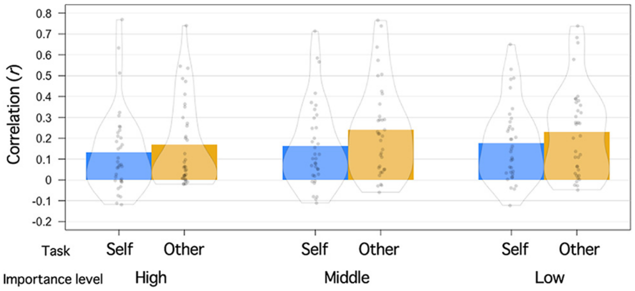


Figure 8. Average within-condition correlations for each of the six conditions [2 (task; self vs other) × 3 (importance-level; high, middle, or low)]. Each participants completed six runs of each task, and within-condition correlations were computed for all possible run-pairs (a total of 15) which was averaged within each participant. A 2 × 3 repeated-measures ANOVA revealed significant main effects of both task ($F_{(1,34)} = 6.71$, $p = 0.014$) and importance-level ($F_{(2,68)} = 3.67$, $p = 0.0307$), while a task × importance-level interaction was not significant ($F_{(2,68)} = 0.5262$ $p = 0.539$). Note that the correlation coefficients were Fisher-z transformed before conducting the ANOVA.

different patterns of activation within the visual cortex (lingual-gyrus) for both the self-reference and other-reference tasks. We also found that, during the self-reference task, familiarity ratings were related to activation patterns in left middle frontal gyrus (MFG; $x = -21$, $y = 4$, $z = 47$, 648 voxels) and in left inferior frontal gyrus (IFG; $x = -45$, $y = 11$, $z = 14$, 302 voxels). No other significant effects emerged.

Exploratory analysis directly comparing effects of self-importance and friend-importance

Although classification performance for items high in importance was not significant, the classifier-based MVPA successfully discriminated activation patterns between self-importance-middle versus friend-importance-middle and between self-importance-low versus friend-importance-low (Table 6), indicating that neural codes for information importance are largely unique to the self. Classification performance was not significant for items high in importance; however, our additional analysis found that the data were noisier for items high in importance (i.e., activation patterns evoked by the same item were less consistent across six runs of the same task; Fig. 8). It also found that the self-reference task data were noisier compared with those of the other-reference task. Thus, mPFC activation patterns evoked by items high in self-importance during the self-reference task were least consistent (i.e., noisiest) across six runs. This might be because performing the self-reference (vs other-reference) task and items high (vs low) in importance evokes other cognitive/affective processes (e.g., autobiographical memory, positive affect) that can influence mPFC activation (Bartra et al., 2013; Martinelli et al., 2013), and these unrelated processes might have impacted on mPFC activation patterns differently in each run. Regardless, the elevated level of noise observed in items high in self-importance explains, at least partially, the nonsignificance classification performance for items high in self- versus friend-importance.

Table 7. mPFC regions from searchlight RSA (mega-analysis; $n = 63$) showing significant association with self-importance during the self-reference task

	MNI coordinates				
Location	x	y	z	Z	Cluster size (voxels)
dACC	15	38	26	4.42	942
amPFC	9	56	14	4.34	
dmPFC	6	47	38	3.82	

We set up the statistical threshold at $p < 0.005$ (uncorrected for multiple comparisons) with a cluster threshold $p < 0.05$ (FWE corrected). Voxel size = 3 × 3 × 3 mm.

Mega-analysis (not preregistered)

Given that the self condition was common across experiments 1 and 2, we combined the self-task data from both experiments and ran a mega-analysis that included 63 participants. Self-importance information was reliably encoded in a large cluster in the mPFC ($x = 15$, $y = 38$, $z = 26$, 942 voxels; Fig. 7*c*; Table 7).

When we did not apply the anatomic mPFC mask, the above mPFC cluster extended laterally to right IFG ($x = 36$, $y = 38$, $z = 5$) and left IFG ($x = -33$, $y = 29$, $z = 29$) consisting of 2087 voxels. We observed no other significant cluster for self-importance. Familiarity information was represented in right superior frontal gyrus (SFG; $x = 27$, $y = 26$, $z = 53$; 303 voxels), and autobiographical memory information was represented in right lingual gyrus ($x = 27$, $y = -67$, $z = -1$; 440 voxels). Furthermore, unsurprisingly, word length was strongly associated with activation patterns in the visual cortex (right: $x = 12$, $y = -85$, $z = 2$, and left: $x = -9$, $y = -91$, $z = 2$; a total of 2413 voxels). We observed no significant result for self-descriptiveness and valence.

Discussion

Previous research has linked the self-reference task to neural activation in the mPFC (Wagner et al., 2012), but it is unclear which information about the self is represented in that brain region. Across two self-reference experiments, while controlling for potential confounds, we consistently demonstrated that the mPFC represents how important attributes are to one's self-identity. The results suggest that the self-concept is represented in the mPFC and conceptualized in terms of self-importance, not self-descriptiveness. Furthermore, in both experiments, the parametric modulation analysis found no significant activation in the mPFC. Thus, these results indicate that self-importance information systematically alters activation patterns, but does not affect overall activation magnitude in the mPFC. In experiment 2, we did not observe the relationship between mPFC neural responses and importance in the other (best friend) condition, and mPFC activation patterns associated with each of three levels of importance were generally distinct between self and best friend, suggesting that the mPFC represents information about the importance of information specifically to the self. Taken together, our research improves understanding about how and where the self is represented in the brain.

Although we found an association between self-importance and mPFC activation patterns across two experiments, the self-importance sensitive mPFC areas did not overlap widely (Fig. 7*b*). Given that locations of peak self-related activations reported in previous neuroimaging studies vary greatly along the z-coordinates (e.g., from $z = -10$ to $z = 70$; see also Denny et al., 2012), there might be considerable individual differences in functional dissociations within the mPFC. Thus, only focusing on a group average may prevent researchers from fully understanding the role of the mPFC in the self-concept (and the mPFC functions more generally). Follow-up research should consider and address this possibility.

Although neuroimaging research has documented the involvement of the default mode network in the self-reference task (Qin and Northoff, 2011; Wen et al., 2020), the current results indicate that only the mPFC is associated with self-importance. This finding is consistent with a previous lesion study, which illustrated the mPFC's crucial role in accurate and reliable trait knowledge of the self (Marquine et al., 2016). A patient (J.S., 74-year-old white male) had extensive damage to the medial prefrontal areas including orbitofrontal cortex and anterior cingulate gyrus. He and control participants completed a self-reference task on two occasions using the same trait adjectives. A male nurse who had known Patient J.S. for five years also rated patient J.S. on the same traits. Patient J. S.'s ratings were less consistent across two sessions and less consistent with ratings done by the nurse compared with the control group. On the other hand, when patient J.S. was asked to rate the nurse, his ratings were consistent across two sessions and consistent with ratings done by the nurse himself, indicating that trait knowledge of another person was preserved. In similar experiments with various patients (e.g., autism, ADHD, Alzheimer's disease), trait self-knowledge was remarkably resistant to neural and cognitive damage (Klein and Lax, 2010). Thus, to the best of our knowledge, damage to the mPFC is the only case where trait knowledge of self is impaired, which is in a sharp contrast to other nonself-related knowledge that is impaired after damage to parietal, temporal, or frontal areas (Gainotti, 2000; Neininger and Pulvermüller, 2003; Damasio et al., 2004).

We demonstrated across two experiments that self-importance is represented in the mPFC only during the self-reference task, whereas stimulus' perceptual properties (i.e., word length) are represented in the visual cortex regardless of the task involved. This task-specific neural representation is consistent with RSA studies on object representations (Bracci et al., 2017), which showed that information relevant to a given task is represented in prefrontal and parietal areas only while performing the task, whereas occipitotemporal areas mainly represent stimulus' perceptual properties (e.g., object shape) regardless of task involved. However, during the self-reference task, participants judged whether each personality trait describes them or not; this task does not explicitly require judging how important each trait is to one's identity. Hence, our results suggest an interesting possibility: individuals may actively use self-importance information of a stimulus when judging self-descriptiveness. Furthermore, given the lesion study described above (Marquine et al., 2016), self-importance information represented in the mPFC might be necessary for accurate and consistent self-knowledge.

Our findings have far-reaching implications. First, they can be contextualized in psychological models of the self-concept. One family of such models depicts the self-concept as an associative network structure where the self is a central entity (node) connected to a number of self-relevant features (e.g., "young," "university student") that are themselves connected to each other (Greenwald and Pratkanis, 1984; Kihlstrom and Cantor, 1984). Researchers further added associative strength to the network model so that some of features (nodes) are more or less strongly connected to the self (and each other; Greenwald et al., 2002). Although these researchers considered the strength of association as "the potential for one concept to activate another" (p 5), its psychological meaning was unspecified. Our findings suggest that, for links (edges) directly connected to the self, strength of association may be understood as degree of self-importance. Given that associative strength is considered responsible for reaction time facilitation or inhibition during the Implicit Association Test (IAT; Greenwald et al., 1998, 2002), our findings generate a hypothesis about reaction time facilitation (e.g., priming effects) based on information self-importance, but not self-descriptiveness (although factors other than self-importance are likely to affect reaction times, such as valence). For example, if being a writer is important to an individual, processing speed for the word "writer" will be facilitated after seeing a prime word "self" (or other highly self-important stimulus). Thus, scores on the self-esteem IAT (Greenwald and Farnham, 2000) might reflect the self-importance information processing function of the mPFC as well as the valence processing function of the reward related network. Largely consistent with this possibility, individual difference in implicit self-esteem as measured by the IAT are independently predicted by activation patterns in the mPFC and those in reward-related brain regions (Izuma et al., 2018).

Second, the findings have implications for psychological research on the link between the self-concept and mental health. For example, it is possible that people who have greater self-complexity (i.e., higher number of, and great differentiation between, self-aspects) are less likely to experience depression, physical illness, and stress in response to aversive events (Linville, 1987), especially when they perceive high control over their self-aspects (McConnell et al., 2005). Similarly, individuals who identify with multiple groups, compared with a single group, report lower stress levels (Binning et al., 2009). Other lines of research point to a link between mental conditions or disabilities and the self-concept. For instance, schizophrenia is associated with changes in self-identity (Conneely et al., 2021), and individuals with autism manifest atypical neural self-representation (Lombardo et al., 2010). How information self-importance is represented in these patients' brains might shed new light on the nature of mental health, including schizophrenia and autism.

Third, the findings have implications for the long-debated nature of the self among psychologists. One stream of research has emphasized the cognitive properties of the self (Kihlstrom and Klein, 1994; Kihlstrom et al., 2003), characterizing its cognitive structure as complex but ordinary (Greenwald and Banaji, 1989). Another stream of research has emphasized the motivational properties of the self (Kunda, 1990; Sedikides and Strube, 1997), emphasizing its uniqueness (Alicke and Sedikides, 2009; Sedikides, 2021). Our findings align with the second empirical stream. If the cognitive representation of the self is unique compared with the cognitive representation of other, this uniqueness lies in motivation (here, attribute self-importance) rather than cognition (here, attribute self-descriptiveness). Moreover, the findings have implications for the long-debated nature of the self among philosophers. Numerous philosophers have cast serious doubts on the mere existence of the self (Hofstadter, 2007; Metzinger, 2009; Baggini, 2011; Midgley, 2014). Here, we countered his viewpoint by providing evidence for the representation of the self in the brain, not only in terms self-descriptiveness, but also in terms of self-importance.

In conclusion, research on the self has a long history in psychology (James, 1890), and the question of "where is the self in the brain?" has attracted keen theoretical and empirical interest in the last two decades (Craik et al., 1999). While earlier neuroimaging studies found a robust link between mPFC activation and self-reference processing, what information about the self is processed in the mPFC during a self-reference task has eluded an answer. Our research pinned down the nature of the information about the self that is represented in the mPFC: across two experiments, we demonstrated that information about self-importance (how important a stimulus is to one's self-identity), but not self-descriptiveness, is represented in the mPFC. Put otherwise, the self-concept is represented in the mPFC in terms of self-importance. The mPFC is a neural locus of the self-concept, and this neural system may play a pivotal role in maintaining an accurate and consistent self-concept.

References

Alicke MD, Sedikides C (2009) Self-enhancement and self-protection: what they are and what they do. Eur Rev Soc Psychol 20:1–48.

Baggini J (2011) The ego trick: what does it mean to be you? London: Granta Publications.

Bartra O, McGuire JT, Kable JW (2013) The valuation system: a coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. Neuroimage 76:412–427.

Benoit RG, Gilbert SJ, Volle E, Burgess PW (2010) When I think about me and simulate you: medial rostral prefrontal cortex and self-referential processes. Neuroimage 50:1340–1349.

Binning KR, Unzueta MM, Huo YJ, Molina LE (2009) The interpretation of multiracial status and its relation to social engagement and psychological well-being. J Soc Issues 65:35–49.

Bracci S, Daniels N, Op de Beeck H (2017) Task context overrules object- and category-related representational content in the human parietal cortex. Cereb Cortex 27:310–321.

Chavez RS, Heatherton TF, Wagner DD (2017) Neural population decoding reveals the intrinsic positivity of the self. Cereb Cortex 27:5222–5229.

Conneely M, McNamee P, Gupta V, Richardson J, Priebe S, Jones JM, Giacco D (2021) Understanding identity changes in psychosis: a systematic review and narrative synthesis. Schizophr Bull 47:309–322.

Courtney AL, Meyer ML (2020) Self-other representation in the social brain reflects social connection. J Neurosci 40:5616–5627.

Cousins SD (1989) Culture and self-perception in Japan and the United States. J Pers Soc Psychol 56:124–131.

Craik FIM, Moroz TM, Moscovitch M, Stuss DT, Winocur G, Tulving E, Kapur S (1999) In search of the self: a positron emission tomography study. Psychol Sci 10:26–34.

Damasio H, Tranel D, Grabowski T, Adolphs R, Damasio A (2004) Neural systems behind word and concept retrieval. Cognition 92:179–229.

D'Argembeau A, Jedidi H, Balteau E, Bahri M, Phillips C, Salmon E (2012) Valuing one's self: medial prefrontal involvement in epistemic and emotive investments in self-views. Cereb Cortex 22:659–667.

del Prado AM, Church AT, Katigbak MS, Miramontes LG, Whitty M, Curtis GJ, de J, Vargas-Flores J, Ibáñez-Reyes J, Ortiz FA, Reyes JA (2007) Culture, method, and the content of self-concepts: testing trait, individual-self-primacy, and cultural psychology perspectives. J Res Pers 41:1119–1160.

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 24:1742–1752.

Deutsch FM, Ruble DN, Fleming A, Brooks-Gunn J, Stangor C (1988) Information-seeking and maternal self-definition during the transition to motherhood. J Pers Soc Psychol 55:420–431.

Elder J, Cheung B, Davis T, Hughes B (2022) Mapping the self: a network approach for understanding psychological and neural representations of self-concept structure. J Pers Soc Psychol 124:237–263.

Esterman M, Tamber-Rosenau BJ, Chiu YC, Yantis S (2010) Avoiding non-independence in fMRI data analysis: leave one subject out. Neuroimage 50:572–576.

Feng C, Yan X, Huang W, Han S, Ma Y (2018) Neural representations of the multidimensional self in the cortical midline structures. Neuroimage 183:291–299.

Ferstl EC, Neumann J, Bogler C, von Cramon DY (2008) The extended language network: a meta-analysis of neuroimaging studies on text comprehension. Hum Brain Mapp 29:581–593.

Gainotti G (2000) What the locus of brain lesion tells us about the nature of the cognitive defect underlying category-specific disorders: a review. Cortex 36:539–559.

Gillihan SJ, Farah MJ (2005) Is self special? A critical review of evidence from experimental psychology and cognitive neuroscience. Psychol Bull 131:76–97.

Greenwald AG, Pratkanis AR (1984) The self. In: Handbook of social cognition (Wyer RS, Srull TK, eds), pp 129–178. Hillsdale: Lawrence Erlbaum Associates Publishers.

Greenwald AG, Banaji M (1989) The self as a memory system: powerful, but ordinary. J Pers Soc Psychol 57:41–54.

Greenwald AG, Farnham SD (2000) Using the implicit association test to measure self-esteem and self-concept. J Pers Soc Psychol 79:1022–1038.

Greenwald AG, McGhee DE, Schwartz JLK (1998) Measuring individual differences in implicit cognition: the implicit association test. J Pers Soc Psychol 74:1464–1480.

Greenwald AG, Banaji MR, Rudman LA, Farnham SD, Nosek BA, Mellott DS (2002) A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. Psychol Rev 109:3–25.

Heatherton TF, Wyland CL, Macrae CN, Demos KE, Denny BT, Kelley WM (2006) Medial prefrontal activity differentiates self from close others. Soc Cogn Affect Neurosci 1:18–25.

Hofstadter D (2007) I am a strange loop. New York: Basic Books.

Izuma K, Kennedy K, Fitzjohn A, Sedikides C, Shibata K (2018) Neural activity in the reward-related brain regions predicts implicit self-esteem: a novel validity test of psychological measures using neuroimaging. J Pers Soc Psychol 114:343–357.

James W (1890) The principles of psychology. New York: Henry Holt and Company.

Jenkins AC, Mitchell JP (2011) Medial prefrontal cortex subserves diverse forms of self-reflection. Soc Neurosci 6:211–218.

Kelley WM, Macrae CN, Wyland CL, Caglar S, Inati S, Heatherton TF (2002) Finding the self? An event-related fMRI study. J Cogn Neurosci 14:785–794.

Kihlstrom JF, Cantor N (1984) Mental representations of the self. Adv Exp Soc Psychol 17:1–47.

Kihlstrom JF, Klein SB (1994) The self as a knowledge structure. In: Handbook of social cognition (Wyer RS, Srull TK, eds), pp 153–208. Hillsdale: Lawrence Erlbaum Associates Publishers.

Kihlstrom JF, Beer JS, Klein SB (2003) Self and identity as memory. In: Handbook of self and identity (Leary MR, Tangney JP, eds), pp 68–90. New York: Guilford Press.

Klein SB, Lax ML (2010) The unanticipated resilience of trait self-knowledge in the face of neural damage. Memory 18:918–948.

Koski JE, McHaney JR, Rigney AE, Beer JS (2020) Reconsidering longstanding assumptions about the role of medial prefrontal cortex (MPFC) in social evaluation. Neuroimage 214:116752.

Kriegeskorte N, Mur M, Bandettini P (2008) Representational similarity analysis - connecting the branches of systems neuroscience. Front Syst Neurosci 2:4.

Kuhn MH, McPartland TS (1954) An empirical investigation of self-attitudes. Am Sociol Rev 19:68–76.

Kunda Z (1990) The case for motivated reasoning. Psychol Bull 108:480–498.

Legrand D, Ruby P (2009) What is self-specific? Theoretical investigation and critical review of neuroimaging results. Psychol Rev 116:252–282.

Lin WJ, Horner AJ, Burgess N (2016) Ventromedial prefrontal cortex, adding value to autobiographical memories. Sci Rep 6:28630.

Linville PW (1985) Self-complexity and affective extremity: don't put all of your eggs in one cognitive basket. Soc Cogn 3:94–120.

Linville PW (1987) Self-complexity as a cognitive buffer against stress-related illness and depression. J Pers Soc Psychol 52:663–676.

Lombardo MV, Chakrabarti B, Bullmore ET, Sadek SA, Pasco G, Wheelwright SJ, Suckling J, Consortium MA; MRC AIMS Consortium; Baron-Cohen S (2010) Atypical neural self-representation in autism. Brain 133:611–624.

Macrae CN, Moran JM, Heatherton TF, Banfield JF, Kelley WM (2004) Medial prefrontal activity predicts memory for self. Cereb Cortex 14:647–654.

Maldjian JA, Laurienti PJ, Kraft RA, Burdette JH (2003) An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. Neuroimage 19:1233–1239.

Markus H (1977) Self-schemata and processing information about self. J Pers Soc Psychol 35:63–78.

Markus H (1983) Self-knowledge: an expanded view. J Pers 51:543–565.

Markus H, Wurf E (1987) The dynamic self-concept: a social psychological perspective. Annu Rev Psychol 38:299–337.

Marquine MJ, Grilli MD, Rapcsak SZ, Kaszniak AW, Ryan L, Walther K, Glisky EL (2016) Impaired personal trait knowledge, but spared other-person trait knowledge, in an individual with bilateral damage to the medial prefrontal cortex. Neuropsychologia 89:245–253.

Martinelli P, Sperduti M, Piolino P (2013) Neural substrates of the self-memory system: new insights from a meta-analysis. Hum Brain Mapp 34:1515–1529.

McConnell AR, Renaud JM, Dean KK, Green SP, Lamoreaux MJ, Hall CH, Rydell RJ (2005) Whose self is it anyway? Self-aspect control moderates the relation between self-complexity and well-being. J Exp Soc Psychol 41:1–18.

Metzinger T (2009) The ego tunnel: the science of the mind and the myth of the self. New York: Basic Books.

Midgley M (2014) Are you an illusion? London: Routledge.

Moran JM, Macrae CN, Heatherton TF, Wyland CL, Kelley WM (2006) Neuroanatomical evidence for distinct cognitive and affective components of self. J Cogn Neurosci 18:1586–1594.

Mumford JA, Poline JB, Poldrack RA (2015) Orthogonalization of regressors in FMRI models. PLoS One 10:e0126255.

Murray RJ, Schaer M, Debbané M (2012) Degrees of separation: a quantitative neuroimaging meta-analysis investigating self-specificity and shared neural activation between self- and other-reflection. Neurosci Biobehav Rev 36:1043–1059.

Neininger B, Pulvermüller F (2003) Word-category specific deficits after lesions in the right hemisphere. Neuropsychologia 41:53–70.

Nichols TE, Holmes AP (2002) Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum Brain Mapp 15:1–25.

Ochsner KN, Beer JS, Robertson ER, Cooper JC, Gabrieli JD, Kihsltrom JF, D'Esposito M (2005) The neural correlates of direct and reflected self-knowledge. Neuroimage 28:797–814.

Parelman JM, Doré BP, Cooper N, O'Donnell MB, Chan HY, Falk EB (2022) Overlapping functional representations of self- and other-related thought are separable through multivoxel pattern classification. Cereb Cortex 32:1131–1141.

Qin P, Northoff G (2011) How is our self related to midline regions and the default-mode network? Neuroimage 57:1221–1233.

Rameson LT, Satpute AB, Lieberman MD (2010) The neural correlates of implicit and explicit self-relevant processing. Neuroimage 50:701–708.

Schmitz TW, Kawahara-Baccus TN, Johnson SC (2004) Metacognitive evaluation, self-relevance, and the right prefrontal cortex. Neuroimage 22:941–947.

Sedikides C (1993) Assessment, enhancement, and verification determinants of the self-evaluation process. J Pers Soc Psychol 65:317–338.

Sedikides C (1995) Central and peripheral self-conceptions are differentially influenced by mood: tests of the differential sensitivity hypothesis. J Pers Soc Psychol 69:759–777.

Sedikides C (2021) The homeostatic model of identity protection: lingering issues. Psychol Inq 32:284–288.

Sedikides C, Strube MJ (1997) Self-evaluation: to thine own self be good, to thine own self be sure, to thine own self be true, and to thine own self be better. Adv Exp Soc Psychol 29:209–269.

Sedikides C, Gregg AP (2003) Portraits of the self. In: Sage handbook of social psychology (Hogg MA, Cooper J, eds), pp 110–138. London: Sage Publications.

Sedikides C, Alicke MD, Skowronski JJ (2021) On the utility of the self in social perception: an egocentric tactician model. Adv Exp Soc Psychol 63:247–298.

Sedikides C, Green JD, Saunders J, Skowronski JJ, Zengel B (2016) Mnemic neglect: selective amnesia of one's faults. Eur Rev Soc Psychol 27:1–62.

Smith JF, Hur J, Kaplan CM, Shackman AJ (2018) The impact of spatial normalization for functional magnetic resonance imaging data analyses revisited. bioRxiv 272302. https://doi.org/10.1101/272302.

Tan KM, Daitch AL, Pinheiro-Chagas P, Fox KCR, Parvizi J, Lieberman MD (2022) Electrocorticographic evidence of a common neurocognitive sequence for mentalizing about the self and others. Nat Commun 13:1919.

Vanderwal T, Hunyadi E, Grupe DW, Connors CM, Schultz RT (2008) Self, mother and abstract other: an fMRI study of reflective social processing. Neuroimage 41:1437–1446.

Wagner DD, Haxby JV, Heatherton TF (2012) The representation of self and person knowledge in the medial prefrontal cortex. Wiley Interdiscip Rev Cogn Sci 3:451–470.

Wagner DD, Chavez RS, Broom TW (2019) Decoding the neural representation of self and person knowledge with multivariate pattern analysis and data-driven approaches. Wiley Interdiscip Rev Cogn Sci 10:e1482.

Wake SJ, Izuma K (2017) A common neural code for social and monetary rewards in the human striatum. Soc Cogn Affect Neurosci 12:1558–1564.

Wen T, Mitchell DJ, Duncan J (2020) The functional convergence and heterogeneity of social, episodic, and self-referential thought in the default mode network. Cereb Cortex 30:5915–5929.

Yankouskaya A, Humphreys G, Stolte M, Stokes M, Moradi Z, Sui J (2017) An anterior-posterior axis within the ventromedial prefrontal cortex separates self and reward. Soc Cogn Affect Neurosci 12:1859–1868.

Yaoi K, Osaka M, Osaka N (2015) Neural correlates of the self-reference effect: evidence from evaluation and recognition processes. Front Hum Neurosci 9:383.