

# Brain System for Social Categorization by Narrative Roles

Yorai Ron,<sup>1\*</sup> Amnon Dafni-Merom,<sup>1\*</sup> Noam Saadon-Grosman,<sup>1</sup> Moshe Roseman,<sup>1</sup> Uri Elias,<sup>1</sup> and Shahar Arzy<sup>1,2,3</sup>

<sup>1</sup>Neuropsychiatry Lab, Department of Medical Neurosciences, Faculty of Medicine, Hebrew University of Jerusalem, Jerusalem 9190501, Israel,

<sup>2</sup>Department of Neurology, Hadassah Hebrew University Medical School, Jerusalem 9112001, Israel, and <sup>3</sup>Department of Brain and Cognitive Sciences, Hebrew University of Jerusalem, Jerusalem 9190501, Israel

The cognitive system applies categorical thinking to facilitate perception of the rich environment around us. In person cognition, research has focused on the roles of gender, race, age, or appearance in social categorical thinking. Here we investigated how narrative roles, as portrayed by different cinematic characters, are categorized in the neurocognitive system. Under functional MRI, 17 human participants (7 females) were asked to make different judgments regarding personality traits of 16 renowned cinematic characters representing four roles: hero, sidekick, mentor, and villain. Classification analysis showed a brain network, comprising the dorsal medial prefrontal cortex, the precuneus and the temporoparietal junction bilaterally, and the left occipital face area (OFA), to discriminate among the four roles. No such classification was found between other individual attributes including age or the associated film. Moreover, regions overlapping the default mode network (DMN) were found to better discriminate between roles, rather than the individual characters, while the OFA was found to better discriminate between individuals. These results demonstrate the inherent role of roles in person cognition, and suggest an intimate relation between roles categorization and self-referential activity.

**Key words:** categorical thinking; default mode; film; fMRI; person memory; social cognition; social neuroscience

## Significance Statement

Social categorization, the assignment of different people in our social network to subgroups, is a powerful strategy in social cognition. How is this managed by the brain? We provide evidence that different characters from different stories, representing similar roles in their corresponding narrative, elicit similar brain activation patterns, as revealed by functional MRI. Unlike previous studies of social categorization, these brain activations were similar to those elicited by social cognition rather than face processing, and included regions at the prefrontal cortex, the precuneus, and the temporoparietal junction. The identified brain network significantly overlapped the default mode network. We suggest that social categorization by roles is fundamental to the cognitive system, relying on brain regions related to social cognition.

## Introduction

*“Every man acts the part assigned to him—by himself or others—in this life.”* (Six Characters in Search of an Author; Luigi Pirandello, 1921)

As humans, we familiarize ourselves with thousands of people throughout our lifetimes (de Sola Pool and Kochen, 1978; Degenne and Forse, 1999; Jenkins et al., 2018). How does the

cognitive system manage to keep unique representations of so many individuals? One method is through social categorization; that is, the sorting of people one encounters into different groups (Allport, 1954; Bruner, 1957; Fiske and Neuberg, 1990; Bodenhausen and Macrae, 1998; Goldenberg et al., 2020; Peer et al., 2021). Social categories may be governed by various individual attributes such as gender, age, or association with a particular group (Lippmann, 1922; Allport, 1954; Tajfel, 1969). This function appears to be fundamental to human cognition, as evidenced by its emergence in the early stages of development (Lieberman et al., 2017). Furthermore, social categorization has been shown to be applied automatically and rapidly when encountering novel faces (Ito and Urland, 2003; Van Bavel and Cunningham, 2010; Amodio et al., 2014), to positively affect person memory (Srull and Wyer, 1989), and to guide decision-making (Macrae and Bodenhausen, 2000).

Previous neuroimaging studies of social categorization have mainly focused on visual perception of faces (Freeman et al., 2010; Hehman et al., 2014; Volpert-Esmont and Bartholow, 2021),

Received July 11, 2021; revised Apr. 28, 2022; accepted May 11, 2022.

Author contributions: Y.R. and S.A. designed research; Y.R. performed research; Y.R., A.D.-M., N.S.-G., M.R., U.E., and S.A. analyzed data; Y.R., A.D.-M., and S.A. wrote the paper.

The study was supported by the Israel Science Foundation (Grant 1306/18) and the Center for Interdisciplinary Data Science Research of the Hebrew University of Jerusalem. Y.R. is supported by the Samuel & Lottie Rudin Scholarship Foundation. We thank Dr. Gregory Peters-Founshtein for comments on the manuscript and help in data analysis.

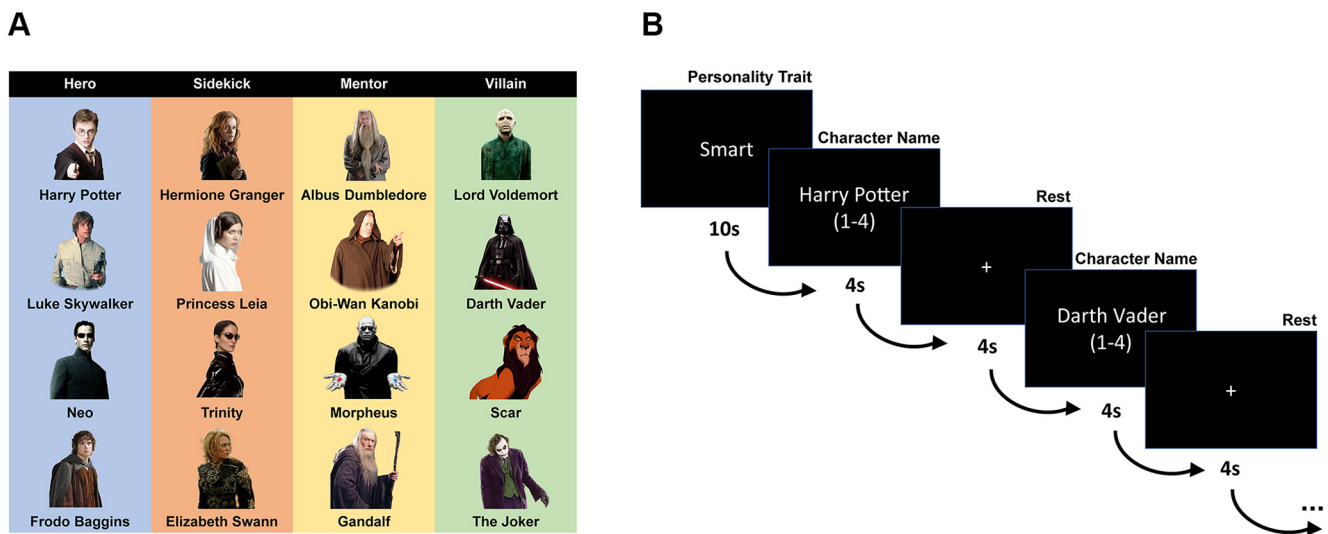
\*Y.R. and A.D.-M. contributed equally to this work.

The authors declare no competing financial interests.

Correspondence should be addressed to Shahar Arzy at shahar.arzy@ekmd.huji.ac.il.

<https://doi.org/10.1523/JNEUROSCI.1436-21.2022>

Copyright © 2022 the authors



**Figure 1.** Experimental paradigm. **A**, Stimuli. The 16 characters used in the experiment, categorized by their narrative role. Participants were not exposed to this categorization at any point during the experiment. **B**, Experimental design. Participants were presented with a quantitative personality question (e.g., “To what extent is the character smart/impulsive/cynical/etc.?”) followed by the names of each of the 16 characters. Response was then given on a scale of 1–4.

though it was shown that social categorization does not require a visual depiction of the person categorized (Cikara et al., 2017). Furthermore, most studies investigated the neural substrates of in-group versus out-group representations, usually by explicit differences, such as race and gender (Contreras et al., 2013; Ratner et al., 2013; Stoler and Freeman, 2017; Brooks et al., 2018). Interestingly, even ad hoc assignments to arbitrary groups elucidated similar differences (Cikara et al., 2017). In another line of work, Hassabis et al. (2014) trained participants to learn the personalities of four protagonists, varying in two personality traits (agreeableness and extraversion), thereby constructing a 2-by-2 factorial design. Importantly, participants were unaware of this design and were told that the experimenters were interested in seeing how well they could predict each protagonist’s behavior. Multivariate pattern analysis (MVPA) identified brain regions associated with personality-trait processing, while activity patterns in the medial prefrontal cortex (mPFC) successfully discriminated between the protagonists’ individual identities. These findings suggest that the brain represents individuals by combining these traits, enabling the prediction of others behavior. Yet, complex human behavior is not solely attributed to personality traits, as the context and the relations to the other individuals in the situation is essential as well.

Here we investigated social categorization by roles using narrative information. Narratives can be viewed as simulations of real-life social processes (Mar and Oatley, 2008), offering a refined, “stripped-down” model for social interactions and roles (Simony et al., 2016; Baldassano et al., 2018; Nguyen et al., 2019). Fictional characters have been used in neuroimaging studies to evoke cognitive elements of real-life social experience in controlled environments (Kumaran et al., 2012; Hassabis et al., 2014; Tavares et al., 2015). Thus, to investigate how the human brain processes roles, we turned to fictional characters representing canonical examples of these roles that recur in several fictional stories (namely: hero, villain, sidekick, mentor). We hypothesized that different characters from different stories, representing similar roles in their corresponding narrative will elicit similar brain activation patterns.

## Materials and Methods

**Participants.** Seventeen right-handed healthy young adults (7 females; mean  $\pm$  SD age,  $25.8 \pm 2.6$  years) with normal or corrected-to-normal vision and with no history of neurologic or psychiatric disorders participated in the study. All participants have previously watched each of the films used in the experiment and were specifically questioned to ensure high familiarity with each of the films and characters. The study was approved by the ethical committee of the Hadassah Medical Center, and all participants provided written informed consent.

**Stimuli and procedure.** Following consultations with three professional film critics, 16 well known characters, appearing in blockbuster films (Fig. 1), were chosen to represent four narrative roles, recurring in several films: (1) hero; (2) sidekick; (3) mentor; and (4) villain (Fig. 1A). During the experiment, participants were presented with a question regarding a personality trait (e.g., “To what extent is this character smart?”) followed by the name of one of the characters (Fig. 1B). For each of the 16 characters, participants were instructed to rate how much this trait characterized the specific character on a scale of 1–4 (Fig. 1B) (Peer et al., 2021). Thus, participants had to retain a mental image of the specific character, yet in a manner not directly related to the research question as examined here. Participants were instructed to respond accurately, but as fast as possible. To ensure that characters did not differ in these traits, repeated-measures ANOVA was run across traits within each group (hero, sidekick, mentor, and villain). None of the comparisons were found significant (all  $p$ -values  $> 0.2$ ), suggesting that the characters chosen for each role match on different trait dimensions. In addition, repeated-measures ANOVA was run across participants on the different rating of traits also in between roles. No differences were found here as well (all  $p$ -values  $> 0.1$ ). Overall, questions related to 18 traits were presented, 3 in each of the six experimental runs, giving rise to 288 character presentations during the experiment. Each trial lasted 124 s, starting with a trait-related question (10 s), followed by 16 consecutive stimuli (character names) presented in randomized order (4 s each), with an interstimulus fixation cross (4 s). Preceding the experiment, participants underwent a short learning session, which consisted of two full trials with trait-related questions unused in the actual experiment. The question sequence was counterbalanced across participants. Stimuli were presented using the Presentation software (version 18.3; Neurobehavioral Systems; [www.neurobs.com](http://www.neurobs.com)). Following the experiment, participants were asked to rate their familiarity with each film and character, as well as to specify when they last saw each film. These ratings were further examined with respect to participant performance in judging the characters’ traits using Pearson correlation. No significant correlations were found (all corrected  $p$  values  $> 0.26$ ).

**Data acquisition.** All participants were scanned at the same site using a Siemens Skyra 3 T system (32-channel head coil) with the same imaging sequence. Blood oxygenation level-dependent fMRI was acquired using a whole-brain, gradient-echo echoplanar imaging [repetition time (TR) = 2000 ms; echo time (TE) = 30 ms; flip angle, 75°; field of view, 192 mm; matrix size, 64 × 64; functional voxel size, 3 × 3 × 3 mm; 37 axial slices; descending acquisition order; 10% gap]. In addition, T1-weighted high-resolution (1 × 1 × 1 mm, 160 slices) anatomic images were acquired for each subject along the same orientation as the functional images using the MPRAGE protocol (TR = 2300 ms; TE = 2.98 ms; flip angle, 9°; field of view, 256 mm).

**Preprocessing.** Preprocessing was performed using the BrainVoyager QX 20.6.2.3266 software package (R. Goebel, Brain Innovation), NeuroElf ([www.neuroelf.net](http://www.neuroelf.net)), and in-house, MATLAB-based software (MathWorks). It included slice scan time correction (cubic spline interpolation), high-pass filtering (cutoff frequency, two cycles per scan), and 3D motion correction by realignment to the first run image (trilinear interpolation for detection and sinc for correction), exclusion of runs with maximal motion above a single voxel size (3 mm) in any direction. No spatial smoothing was applied (Op de Beeck, 2010; Weaverdyck et al., 2020). Functional data were coregistered to the anatomic T1 images. Anatomical brain images were corrected for signal inhomogeneity and skull stripped. All images were subsequently normalized to Montreal Neurologic Institute (MNI) coordinates (ICBM-152).

**General linear model analysis.** To assess the selective activations elicited by different experimental conditions, we applied a general linear model (GLM) analysis. The model predictors were defined by a convolution of a condition box-car time course with a standard hemodynamic response function (two-gamma). For each subject, we performed a first-level analysis for each of the six runs independently, which resulted in 16  $\beta$ -weights, one for every character used. Six motion parameters were added to the GLM as covariates of no interest to remove motion-related noise.

**Multivoxel pattern analysis (MVPA).** MVPA was performed using the CoSMoMVPA toolbox (Oosterhof et al., 2016) in MATLAB. In CoSMoMVPA, we performed a whole-brain searchlight analysis (Kriegeskorte et al., 2006) separately for each subject on the voxel-wise  $\beta$ -weights (GLM). First,  $\beta$ -weights were demeaned to verify that classification could not be governed by a difference in the amount of activity by condition across all voxels (simple univariate difference). Second, a linear discriminant analysis (LDA) classifier was trained on labeled data from four of the six functional runs. Following, the trained classifier was presented with naive data (unlabeled) from the two left-out runs (Lee et al., 2017; Pauli et al., 2019). This procedure was repeated 15 times, testing different runs each time (leave-two-out cross-validation). These cross-validated analyses were performed within overlapping spherical ROIs of 123 voxels tiling the entire brain. This yielded a whole-brain map for each subject in which the center voxel of each ROI is assigned with a classification accuracy. To determine whether the classifier performed above chance at the group level, we used random-effects Monte Carlo cluster statistics corrected for multiple comparisons (as implemented by the CosmoMVPA toolbox; Maris and Oostenveld, 2007; Oosterhof et al., 2016). Threshold-free cluster enhancement (TFCE; Smith and Nichols, 2009) was used as a cluster-forming statistic. To correct for multiple comparisons, the Monte Carlo technique used by CosmoMVPA generates 100 null searchlight maps for each participant by performing a sign-permutation test, swapping the signs of the classification accuracy results at random at each data point (voxel). The 1,000,000 null TFCE maps are then constructed by randomly sampling from these null datasets to estimate a null TFCE distribution (Stelzer et al., 2013), obtaining a group level  $z$  score map of the classifier results. The  $z$ -score threshold was set to 3.45 (Herrmann et al., 2012, see for use of  $z$ -score threshold), yielding five distinct clusters for roles classification (cluster threshold set to 300 voxels). Since the characters used in this experiment share commonalities in age and films that are not related to the roles investigated here, we ran the LDA classifier within each ROI for each participant while assigning balanced

**Table 1. Classification parameters**

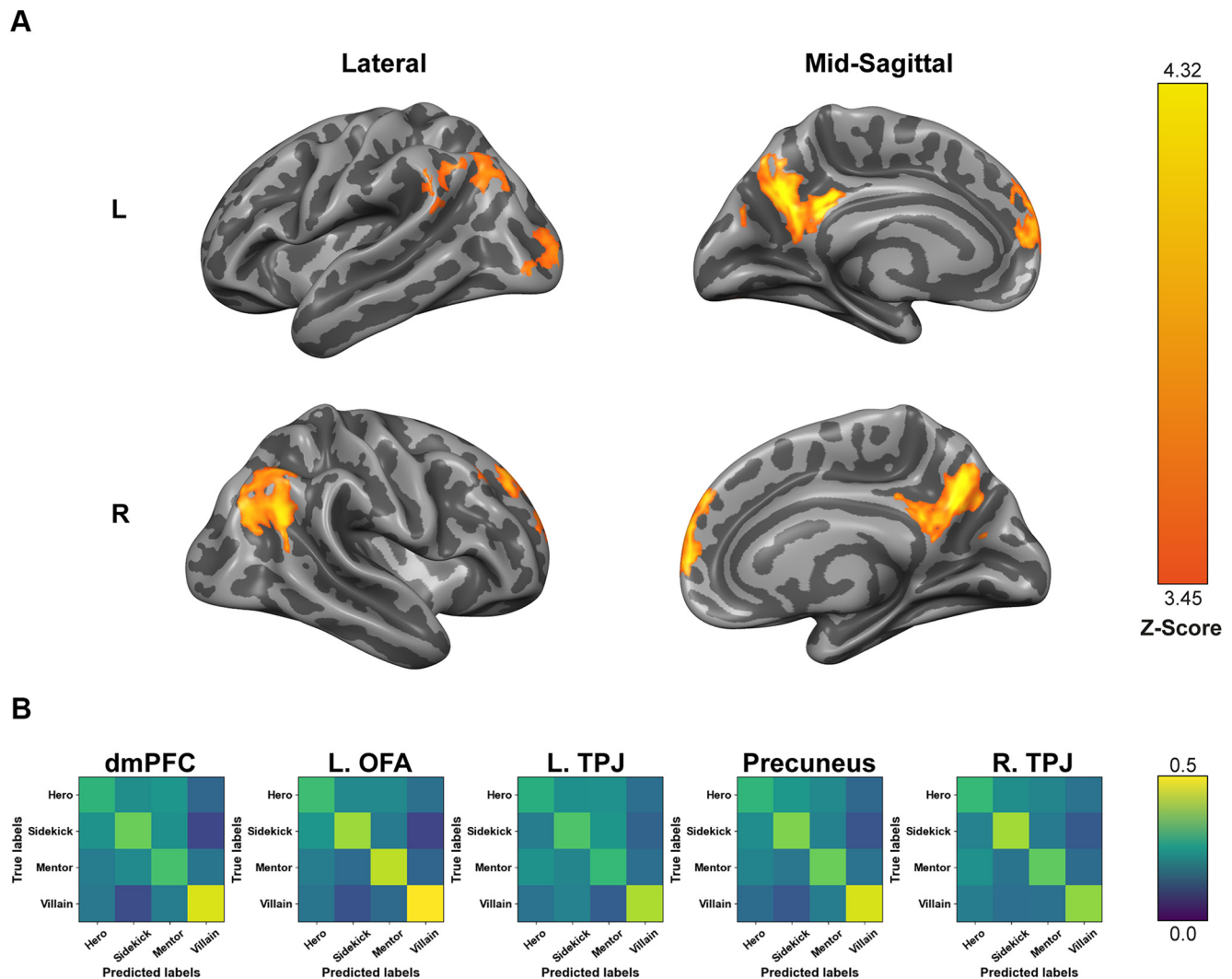
Condition	Samples distribution (%)	Training samples	Test samples
Role	25, 25, 25, 25	64	32
Age	12.5, 37.5, 25, 25	64	32
Film	15, 31, 31, 23	52	26
Gender	25, 75	64	32
Characters	25, 25, 25, 25	16	8

Number and distribution of samples in each condition and class in the training and testing sets (Roles: hero, sidekick, mentor, villain; Age: adolescents, young adults, middle-aged adults, older adults; Film: *Harry Potter*, *The Lord of the Rings*, *Star Wars*, and *The Matrix*; Gender: Females and males; Characters: 4 characters of 16 with 1820 permutations for all possible choices). To avoid potential bias (e.g., film–character pairing), the film condition consisted of a minimum of two characters per class, resulting in using four films with 13 characters in total. For the characters themselves, to enable equal comparison with roles, four characters were selected for each comparison (see Materials and Methods). Of the six runs, four were used for training the classifier, and two for testing classification.

training labels that reflect similarity in age and film (all targets were used at least once). Classification accuracies were averaged across all participants after removal of chance level (25%), and a Wilcoxon signed-rank test was applied for statistical comparisons ( $p$  values < 0.01 were considered significant). The classifier trained on four classes for each of the conditions (Roles: hero, sidekick, mentor, villain; Age: adolescence, young adulthood, middle-age, late adulthood; Films: *Harry Potter*, *The Lord of the Rings*, *Star Wars*, and *The Matrix*; Table 1). While the classification labels of the Roles condition followed a strict pattern of four balanced classes (four characters per class), the Age and the Film conditions required an additional step since the classes were not naturally balanced. We applied the minimum-occurrence method of CosmoMVPA (threshold set to 1), in which each fold consists of a sample set that includes all samples (training and test combined), where some samples are duplicated to ensure an equal number of labels for each class. All Age and Film classes met a requirement of a minimum of two characters per class. A separate classifier was run for gender with two classes: male and female. Since the dataset consists of 12 males and four females, to avoid bias we applied the same balancing scheme as described above. This balancing method is equivalent to applying a threefold higher penalty for male false classification errors than females. In a separate classifier for individual characters, each of the 16 characters was assigned its own class. Importantly, to account for the difference in number of classes between roles (4) and individual characters (16), we evaluated classification accuracy for individual characters by running MVPA separately on subsets of four individuals at a time. A total of 1820 subsets were tested, accounting for all possible permutations [16 combinations of 4 ( ${}_{16}C_4$ )]. Results of all permutations were averaged to yield an overall classification accuracy, and  $p$ -values for each voxel were calculated accordingly. Finally, as an additional validation measure for the Roles classification, we generated whole-brain maps through a small-scale version of the pipeline depicted here for each alternative subgroup (requiring only 100,000 permutations) and superimposed each of the maps (peak values) on the original group map.

**Comparison of scale-selective activations to large-scale resting-state networks.** A previously published whole-brain parcellation into seven large-scale brain networks was used as a template for resting-state networks location (Yeo et al., 2011). Next, we characterized the involvement of the overlapping resting-state networks in role versus individual classification to assess whether role and individual classifications are processed by similar brain networks by computing the Jaccard index (intersection over union) between the group-level map and a cortical parcellation to seven large-scale brain networks. We then computed accuracy levels for roles and individual characters in the default mode network (DMN) and the visual network. To keep all comparisons on the same level, we have applied a permutation analysis, which includes four characters in each cycle across all comparisons and conditions (1820 potential combinations). The  $p$ -value of each classifier on each voxel was calculated accordingly and was followed by false discovery rate (FDR) multiple-comparisons correction. In a second step, overlap was calculated with the 17 network parcellations available at the above resource (Yeo et al., 2011) to compute the overlap with the three subdivisions of the DMN.





**Figure 2.** MVPA results. **A**, Color-coded z-score maps ( $z = 3.45$ ,  $p < 0.0001$ ) illustrating brain areas that discriminate between roles. Clusters were identified in the dmPFC, the precuneus and the TPJ bilaterally, and the left OFA. **B**, Confusion matrices yielded by the classifiers at the five aforementioned clusters.

**Table 2.** Clusters identified in multivariate pattern analysis that discriminate between roles

Region	<i>x</i>	<i>y</i>	<i>z</i>	<i>z</i> score	<i>p</i> value
Left TPJ	−57	−58	17	3.81	0.000139
Right TPJ	45	−55	28	4.13	0.000036
Left OFA	−30	−97	7	3.88	0.000104
dmPFC	3	56	20	4.13	0.000036
Precuneus	−2	−62	33	4.24	0.000022

Shown are peak coordinates of clusters in MNI space (*x*, *y*, *z*).

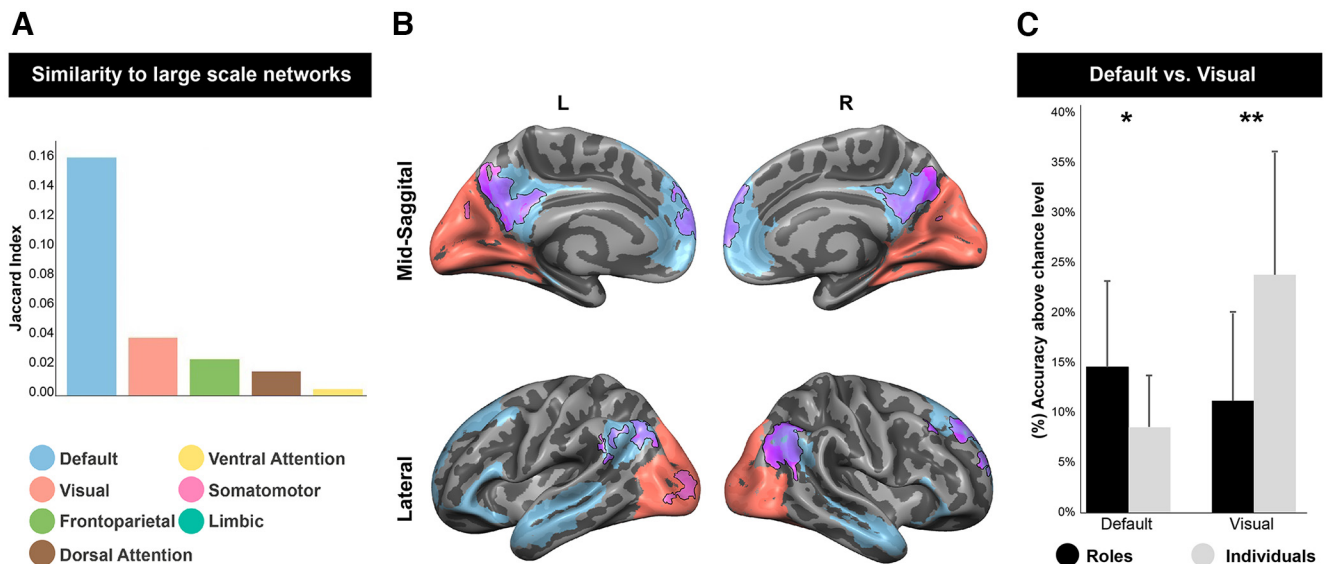
## Results

Applying a classification analysis on our fMRI data identified five clusters that significantly distinguished among the four roles ( $p < 0.0001$ , FDR corrected) in the dorsal mPFC (dmPFC), the precuneus, the superior temporal and inferior parietal cortex [the temporoparietal junction (TPJ)] bilaterally, and the left occipital face area (OFA; Fig. 2A, Table 2).

Classification based on narrative roles performed significantly better than chance in each of the five identified clusters ( $p < 0.05$ , corrected for multiple comparisons), while classifications for age and film were not above chance level when inspecting each condition in each cluster separately (all  $p$  values  $> 0.09$ ).

Classification based on gender only discriminated the OFA bilaterally. Confusion matrices for the classifiers were calculated for each of the five identified clusters, exhibiting similar patterns (Fig. 2B).

Next, we compared the network of brain regions as identified here, with a parcellation of the brain into seven cortical resting-state fMRI networks, based on a dataset from 1000 subjects (Yeo et al., 2011). Voxelwise quantification showed that the majority of the voxels significantly decoding categorization according to roles (71%) overlapped with the DMN. Jaccard index, calculated between the group-level map and a cortical parcellation to seven large-scale brain networks (Yeo et al., 2011), showed that most dominant in its similarity to the current results was the DMN (0.16), followed by the visual (0.04), frontoparietal (0.03), and dorsal attention (0.02) networks (Fig. 3A,B). Within the DMN, we observed higher accuracy levels for the detection of roles than that of the individual characters (paired  $t$  test:  $t_{(16)} = 3.64$ ,  $p = 0.002$ ; Fig. 3C, left), while a reverse pattern was found for the visual system (paired  $t$  test:  $t_{(16)} = 5.5$ ,  $p < 0.001$ ; Fig. 3C, right). Finally, within the DMN, the DMN-A subnetwork was found to mostly overlap with the regions significantly decoding roles (47.7%), more than DMN-B (18.4%) and DMN-C (9.6%).



**Figure 3.** Role-based classification and the DMN. **A**, Jaccard index (intersection over union) was calculated between the group-level map and a cortical parcellation to seven large-scale brain networks. Results show that most dominant in its similarity to the group-level map was the DMN (0.16), followed by the visual (0.04), frontoparietal (0.03), and dorsal attention (0.02) networks. **B**, Brain areas depicting role categorization (purple) and the two main networks overlapping it, the DMN in blue and the visual network in pink, are shown on an inflated brain. **C**, Accuracy levels for roles (black) and individual characters (gray) in both the DMN and the visual network (\* $p < 0.05$ , \*\* $p < 0.001$ ). Note that while the DMN is mainly explained by roles, the visual network is explained more by the characters themselves (“individuals”).

## Discussion

This work provides evidence for the neurocognitive implementation of narrative roles in social categorization. This function is processed by a brain network that includes the dmPFC, the precuneus, and the OFA bilaterally, as well as the left TPJ. Such a classification was not found for other factors, including age or the film involved. The identified brain network significantly overlapped the DMN. Moreover, while regions that overlapped visual areas better discriminated between specific characters, the DMN overlap showed better discrimination for narrative roles. These results are discussed below with respect to social categorization, person memory, the brain system involved, and the implication of the DMN.

Our mental capacity to readily deduce abundant social information from newly acquainted individuals has been extensively studied. One of the most common and efficient processes that facilitates social knowledge and perception is social categorization (Allport, 1954; Fiske and Neuberg, 1990; Macrae and Bodenhausen, 2000). Social categorization research has focused mainly on age, race, and gender (Ito and Urland, 2005; Amodio, 2008; Kubota et al., 2012; Cloutier et al., 2014). More conceptual categories were examined in studies of person memory; that is, the ways in which representations of people are organized in and extracted from one’s memory (Hastie and Kumar, 1979; Hamilton et al., 1980; Srull, 1981). Conceptual categories such as occupation or social classes have been shown to affect the remembering of people, their attributes, and their actions (Sherman et al., 1998; Macrae et al., 1999; Klauer et al., 2004). Conceptual information has also been shown to enhance face recognition, even more than perceptual information (Schwartz and Yovel, 2016). Self-reference may play a role in social categorization as well (Lau et al., 2020; Hayman and Arzy, 2021). Our work highlights a different, yet not less important, type of social categorization—that of roles played. Social categorization in general was found to have important downstream consequences for how we construe others; for example, stereotyping and prejudice (Taylor et al., 1978; Fiske and Neuberg, 1990; Wheeler and

Fiske, 2005). Our results suggest that this also applies for categorization by roles.

Social categorization according to race, age, and gender was found to rely on a network of brain regions, which in most studies involved areas dedicated to face identification, namely, the fusiform face area (FFA) and OFA (Cloutier et al., 2008; Feng et al., 2011; Wiese et al., 2012; Contreras et al., 2013; Ratner et al., 2013; Wei et al., 2014; Stolier and Freeman, 2017). While the OFA has been shown to process lower-level features such as detecting faces and categorization by gender or age (Sergent et al., 1992; Haxby et al., 1999; Pitcher et al., 2011), the FFA has been associated with higher-level processing such as the recognition of individual faces (Haxby et al., 1994). Notably, all these studies have used paradigms that included the presentation of faces. Our results point to the OFA, but not the FFA, as being active in social categorization by roles. More specifically, the left OFA was found here to better identify individuals compared with identifying different roles, coinciding with the literature, which emphasizes its role in individual recognition (Solomon-Harris et al., 2013; Ambrus et al., 2017). As for the FFA, it may be suggested that its activation in social categorization tasks is bound to the stimuli used.

The other regions identified in the present work, namely, the dmPFC, precuneus, and TPJ, were less frequently found in previous studies of social categorization. The dmPFC was involved in categorization according to group membership (comparison of “in-group” and “out-group” words; Morrison et al., 2012). Categorization by race using face images also revealed dmPFC activation (Feng et al., 2011). Moreover, comparing in-group to out-group categorization found the precuneus as a main hub of activity (Bruneau and Saxe, 2010). Precuneus activation was also elicited by categorization by race in addition to FFA activity (Wei et al., 2014). Interestingly, these regions were identified in our previous study (Peer et al., 2015) as being more active during social orientation judgments compared with spatial and temporal orientation judgments. To the best of our knowledge, TPJ activity was not mentioned previously in social categorization studies,

though it is a key region in Theory of Mind (ToM; Saxe and Kanwisher, 2003), which may be a basic step in categorization. It may be the case that faces facilitate categorization according to gender or race to the extent that ToM is less needed, compared with more complicated abstract categories such as narrative roles. These suggest that high-order functions are crucial for social categorization by roles.

The ensemble activity of dmPFC, precuneus, and TPJ may be better conceptualized in view of their involvement in the DMN, which is associated with internal mentation and self-referential mental activity (Gusnard et al., 2001; Buckner et al., 2008). More specifically, the DMN was found to be related with social cognition in the forms of ToM, cognitive distance, and self-projection in the social domain (Krienen et al., 2010; DiNicola et al., 2020; Hayman and Arzy, 2021). Voxelwise quantification showed that the majority of the voxels active for role categorization (71%) overlapped with the DMN (as demonstrated by the Jaccard index; Fig. 3A). The involvement of the DMN in social categorization may be related to the mutual relations between self-reference and social processing, as was previously hypothesized (Feng et al., 2011). Our data extend the implication of the DMN (specifically, the core subnetwork or DMN-A) in social cognition to the concept of social categorization by roles. Furthermore, the overlap with regions involved in self-projection and cognitive distance in the social domain may suggest a shared representational framework, in which unique roles serve as landmarks within a social feature space (Tavares et al., 2015; Kaplan and Friston, 2019; Arzy and Kaplan, 2022).

According to the social learning theory of Bandura (1986), characters within fictional narratives serve as models of social behavior, possibly shaping the way we understand and relate with others in actual life (Black et al., 2021). Advances in functional neuroimaging have invited researchers to explore the relationship between films and cognitive neuroscience, capitalizing on the rich ecological and narrative information they carry (Hasson et al., 2008b; Baldassano et al., 2018; Lee et al., 2020). One prominent line of research has focused on intersubject correlation of brain activity in response to film viewing; that is, how the same film or scene elicits similar responses in the brains of different individuals (Hasson et al., 2008a, 2010). Other recent fMRI studies showed that regions within the DMN accurately classify schematic knowledge of television characters (Vodrahalli et al., 2018; Raykov et al., 2020), a finding that corresponds with our results. The narratives depicted in films offer cognitive neuroscientists a more realistic and naturalistic setting, which may provide important insights into the neurocognitive mechanisms underlying human memory, social cognition, and person perception.

Our study was not free of limitations. Since most Hollywood genre films are vastly inspired by archaic patriarchal myths, and in view of the homogeneity required by the experimental settings, the ratio of male and female characters within each category is uneven. This bias has also prevented us from comparing social categorization by roles to categorization by gender. Nonetheless, a separate classifier for the gender condition showed discrimination in the OFA bilaterally, suggesting that this classifier relates to gender, a classification based on visual face-related features (Chiu et al., 2011; Wiese et al., 2012). Furthermore, while cinematic characters mimic to some extent the social roles seen in daily life, the use of real-life people from the participants' own life would have been more ideal. However, such a design would not enable cross-subject comparison, as analyzed here. Additionally, the

roles found here may result from a combination of personality traits. However, no such coherent pattern of traits was found. A prevailing theory of human social interactions emphasizes the performative aspects of social behavior (Jones et al., 1961). This theoretical framework reconceptualizes personal identity (and subsequently, person perception) as performed in a specific place or moment, and for a specific audience, thereby outlining another perspective, complementary to a relatively stable constellation of traits. In this respect, investigating the neural mechanisms that underlie the processing of roles could offer valuable insights, complementary to investigations of trait-based person perception. Finally, participants' affect toward the different characters may also play a role in our results. However, this is not very likely since classical affect-related regions (e.g., anterior insula, anterior cingulate cortex, amygdala, ventral striatum; Knutson et al., 2014; Lindquist et al., 2016) were not elucidated in our results.

In conclusion, this study has shown that the neurocognitive system intuitively classifies characters according to narrative roles. The brain network identified here to underlie social categorization by roles was found to be closely related to the DMN, highlighting the involvement of high-order self-referential processes in this activity. Together, these findings may help to further understand the way our cognitive system classifies other people according to the roles they play and the narrative they tell, paving the way to better understand the relations between internally generated and externally perceived information (Arzy and Schacter, 2019) in social cognition.

## References

- Allport GW (1954) The nature of prejudice. Boston: Addison-Wesley.
- Ambrus GG, Windel F, Burton AM, Kovács G (2017) Causal evidence of the involvement of the right occipital face area in face-identity acquisition. *Neuroimage* 148:212–218.
- Amodio DM (2008) The social neuroscience of intergroup relations. *Eur Rev Soc Psychol* 19:1–54.
- Amodio DM, Bartholow BD, Ito TA (2014) Tracking the dynamics of the social brain: ERP approaches for social cognitive and affective neuroscience. *Soc Cogn Affect Neurosci* 9:385–393.
- Arzy S, Schacter DL (2019) Self-agency and self-ownership in cognitive mapping. *Trends Cogn Sci* 23:476–487.
- Arzy S, Kaplan R (2022) Transforming social perspectives with cognitive maps. Social cognitive and affective neuroscience. Available at <https://doi.org/10.1093/scan/nsac017>.
- Baldassano C, Hasson U, Norman KA (2018) Representation of real-world event schemas during narrative perception. *J Neurosci* 38:9689–9699.
- Bandura A (1986) Social foundation of thought and action. Hoboken, NJ: Prentice Hall.
- Black JE, Barnes JL, Oatley K, Tamir DI, Dodell-Feder D, Richter T, Mar RA (2021) Stories and their role in social cognition. In: *Handbook of empirical literary studies* (Kuiken D, Jacobs AM, eds), pp 229–250. Boston: De Gruyter.
- Bodenhausen GV, Macrae CN (1998) Stereotype activation and inhibition. *Adv Soc cogn* 11:1–52.
- Brooks JA, Stoller RM, Freeman JB (2018) Stereotypes bias visual prototypes for sex and emotion categories. *Soc Cogn* 36:481–493.
- Bruneau EG, Saxe R (2010) Attitudes towards the outgroup are predicted by activity in the precuneus in Arabs and Israelis. *Neuroimage* 52:1704–1711.
- Bruner JS (1957) On perceptual readiness. *Psychol Rev* 64:123–152.
- Buckner RL, Andrews-Hanna JR, Schacter DL (2008) The brain's default network: anatomy, function, and relevance to disease. *Ann N Y Acad Sci* 1124:1–38.
- Chiu YC, Esterman M, Han Y, Rosen H, Yantis S (2011) Decoding task-based attentional modulation during face categorization. *J Cogn Neurosci* 23:1198–1204.



- Cikara M, Van Bavel JJ, Ingbreten ZA, Lau T (2017) Decoding “us” and “them”: neural representations of generalized group concepts. *J Exp Psychol Gen* 146:621–631.
- Cloutier J, Turk DJ, Macrae CN (2008) Extracting variant and invariant information from faces: the neural substrates of gaze detection and sex categorization. *Soc Neurosci* 3:69–78.
- Cloutier J, Freeman JB, Ambady N (2014) Investigating the early stages of person perception: the asymmetry of social categorization by sex vs. age. *PLoS One* 9:e84677.
- Contreras JM, Banaji MR, Mitchell JP (2013) Multivoxel patterns in fusiform face area differentiate faces by sex and race. *PLoS One* 8:e69684.
- de Sola Pool I, Kochen M (1978) Contacts and influence. *Soc Networks* 1:5–51.
- Degenne A, Forse M (1999) *Introducing social networks*. London: Sage.
- DiNicola LM, Braga RM, Buckner RL (2020) Parallel distributed networks dissociate episodic and social functions within the individual. *J Neurophysiol* 123:1144–1179.
- Feng L, Liu J, Wang Z, Li J, Li L, Ge L, Tian J, Lee K (2011) The other face of the other-race effect: an fMRI investigation of the other-race face categorization advantage. *Neuropsychologia* 49:3739–3749.
- Fiske ST, Neuberg SL (1990) A continuum of impression formation, from category-based to individuating processes: influences of information and motivation on attention and interpretation. *Adv Exp Soc Psychol* 23:1–74.
- Freeman JB, Rule NO, Adams RB, Ambady N (2010) The neural basis of categorical face perception: graded representations of face gender in fusiform and orbitofrontal cortices. *Cereb Cortex* 20:1314–1322.
- Goldenberg A, Sweeny T, Spiegel E, Gross JJ (2020) Is this my group or not? The role of ensemble coding of emotional expressions in group categorization. *J Exp Psychol Gen* 149:445–460.
- Gusnard DA, Akbudak E, Shulman GL, Raichle ME (2001) Medial prefrontal cortex and self-referential mental activity: relation to a default mode of brain function. *Proc Natl Acad Sci U S A* 98:4259–4264.
- Hamilton DL, Katz LB, Leirer VO (1980) Cognitive representation of personality impressions: organizational processes in first impression formation. *J Pers Soc Psychol* 39:1050–1063.
- Hassabis D, Spreng RN, Rusu AA, Robbins CA, Mar RA, Schacter DL (2014) Imagine all the people: how the brain creates and uses personality models to predict behavior. *Cereb Cortex* 24:1979–1987.
- Hasson U, Furman O, Clark D, Dudai Y, Davachi L (2008a) Enhanced inter-subject correlations during movie viewing correlate with successful episodic encoding. *Neuron* 57:452–462.
- Hasson U, Landesman O, Knappmeyer B, Vallines I, Rubin N, Heeger DJ (2008b) Neurocinematics: the neuroscience of film. *Projections* 2:1–26.
- Hasson U, Malach R, Heeger DJ (2010) Reliability of cortical activity during natural stimulation. *Trends Cogn Sci* 14:40–48.
- Hastie R, Kumar PA (1979) Person memory: personality traits as organizing principles in memory for behaviors. *J Pers Soc Psychol* 37:25–38.
- Haxby JV, Horowitz B, Ungerleider LG, Maisog JM, Pietrini P, Grady CL (1994) The functional organization of human extrastriate cortex: a PET-rCBF study of selective attention to faces and locations. *J Neurosci* 14:6336–6353.
- Haxby JV, Ungerleider LG, Clark VP, Schouten JL, Hoffman EA, Martin A (1999) The effect of face inversion on activity in human neural systems for face and object perception. *Neuron* 22:189–199.
- Hayman M, Arzy S (2021) Mental travel in the person domain. *J Neurophysiol* 126:464–476.
- Hehman E, Ingbreten ZA, Freeman JB (2014) The neural basis of stereotypic impact on multiple social categorization. *Neuroimage* 101:704–711.
- Herrmann B, Obleser J, Kalberlah C, Haynes JD, Friederici AD (2012) Dissociable neural imprints of perception and grammar in auditory functional imaging. *Hum Brain Mapp* 33:584–595.
- Ito TA, Urland GR (2003) Race and gender on the brain: electrocortical measures of attention to the race and gender of multiply categorizable individuals. *J Pers Soc Psychol* 85:616–626.
- Ito TA, Urland GR (2005) The influence of processing objectives on the perception of faces: an ERP study of race and gender perception. *Cogn Affect Behav Neurosci* 5:21–36.
- Jenkins R, Dowsett AJ, Burton AM (2018) How many faces do people know? *Proc R Soc B Biol Sci* 285:20181319.
- Jones EE, Davis KE, Gergen KJ (1961) Role playing variations and their informational value for person perception. *J Abnorm Soc Psychol* 63:302–310.
- Kaplan R, Friston KJ (2019) Entorhinal transformations in abstract frames of reference. *PLoS Biol* 17:e3000230.
- Klauer KC, Ehrenberg K, Wegener I (2004) Components of homogeneity: a multiple-process model of social categorization. In: *The psychology of group perception: perceived variability, entitativity, and essentialism* (Yzerbyt VYA, Judd CM, eds), pp 147–160. New York: Psychology Press.
- Knutson B, Katovich K, Suri G (2014) Inferring affect from fMRI data. *Trends Cogn Sci* 18:422–428.
- Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. *Proc Natl Acad Sci U S A* 103:3863–3868.
- Krienen FM, Tu PC, Buckner RL (2010) Clan mentality: evidence that the medial prefrontal cortex responds to close others. *J Neurosci* 30:13906–13915.
- Kubota JT, Banaji MR, Phelps EA (2012) The neuroscience of race. *Nat Neurosci* 15:940–948.
- Kumaran D, Melo HL, Duzel E (2012) The emergence and representation of knowledge about social and nonsocial hierarchies. *Neuron* 76:653–666.
- Lau T, Gershman SJ, Cikara M (2020) Social structure learning in human anterior insula. *Elife* 9:1–17.
- Lee H, Bellana B, Chen J (2020) What can narratives tell us about the neural bases of human memory? *Curr Opin Behav Sci* 32:111–119.
- Lee YS, Zreik JT, Hamilton RH (2017) Patterns of neural activity predict picture-naming performance of a patient with chronic aphasia. *Neuropsychologia* 94:52–60.
- Lieberman Z, Woodward AL, Kinzler KD (2017) The origins of social categorization. *Trends Cogn Sci* 21:556–568.
- Lindquist KA, Satpute AB, Wager TD, Weber J, Barrett LF (2016) The brain basis of positive and negative affect: evidence from a meta-analysis of the human neuroimaging literature. *Cereb Cortex* 26:1910–1922.
- Lippmann W (1922) Stereotypes. In: *Public opinion*, pp 79–94. New York: MacMillan.
- Macrae CN, Bodenhausen GV (2000) Social cognition: thinking categorically about others. *Annu Rev Psychol* 51:93–120.
- Macrae CN, Bodenhausen GV, Schloerscheidt AM, Milne AB (1999) Tales of the unexpected: Executive function and person perception. *J Pers Soc Psychol* 76:200–213.
- Mar RA, Oatley K (2008) The function of fiction is the abstraction and simulation of social experience. *Perspect Psychol Sci* 3:173–192.
- Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. *J Neurosci Methods* 164:177–190.
- Morrison S, Decety J, Molenberghs P (2012) The neuroscience of group membership. *Neuropsychologia* 50:2114–2120.
- Nguyen M, Vanderwal T, Hasson U (2019) Shared understanding of narratives is correlated with shared neural responses. *NeuroImage* 184:161–170.
- Oosterhof NN, Connolly AC, Haxby JV (2016) CoSMoMVPA: multi-modal multivariate pattern analysis of neuroimaging data in matlab/GNU octave. *Front Neuroinform* 10:27.
- Op de Beeck HP (2010) Against hyperacuity in brain reading: spatial smoothing does not hurt multivariate fMRI analyses? *Neuroimage* 49:1943–1948.
- Pauli WM, Gentile G, Collette S, Tyszka JM, O’Doherty JP (2019) Evidence for model-based encoding of Pavlovian contingencies in the human brain. *Nat Commun* 10:1–11.
- Peer M, Hayman M, Tamir B, Arzy S (2021) Brain coding of social network structure. *J Neurosci* 41:4897–4909.
- Peer M, Salomon R, Goldberg I, Blanke O, Arzy S (2015) Brain system for mental orientation in space, time, and person. *Proc Natl Acad Sci U S A* 112:11072–11077.
- Pitcher D, Walsh V, Duchaine B (2011) The role of the occipital face area in the cortical face perception network. *Exp Brain Res* 209:481–493.
- Ratner KG, Kaul C, Van Bavel JJ (2013) Is race erased? Decoding race from patterns of neural activity when skin color is not diagnostic of group boundaries. *Soc Cogn Affect Neurosci* 8:750–755.
- Raykov PP, Keidel JL, Oakhill J, Bird CM (2020) The brain regions supporting schema-related processing of people’s identities. *Cogn Neuropsychol* 37:8–24.
- Saxe R, Kanwisher NG (2003) People thinking about thinking people: the role of the temporo-parietal junction in “theory of mind”. *NeuroImage* 19:1835–1842.
- Schwartz L, Yovel G (2016) The roles of perceptual and conceptual information in face recognition. *J Exp Psychol Gen* 145:1493–1511.

- Sergent J, Ohta S, Macdonald B (1992) Functional neuroanatomy of face and object processing: a positron emission tomography study. *Brain* 115:15–36.
- Sherman JW, Lee AY, Bessenoff GR, Frost LA (1998) Stereotype efficiency reconsidered: encoding flexibility under cognitive load. *J Pers Soc Psychol* 75:589–606.
- Simony E, Honey CJ, Chen J, Lositsky O, Yeshurun Y, Wiesel A, Hasson U (2016) Dynamical reconfiguration of the default mode network during narrative comprehension. *Nat Commun* 7:12141.
- Smith SM, Nichols TE (2009) Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage* 44:83–98.
- Solomon-Harris LM, Mullin CR, Steeves JKE (2013) TMS to the “occipital face area” affects recognition but not categorization of faces. *Brain Cogn* 83:245–251.
- Srull TK (1981) Person memory: some tests of associative storage and retrieval models. *J Exp Psychol Hum Learn* 7:440–463.
- Srull TK, Wyer RS (1989) Person memory and judgment. *Psychol Rev* 96:58–83.
- Stelzer J, Chen Y, Turner R (2013) Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): Random permutations and cluster size control. *Neuroimage* 65:69–82.
- Stolier RM, Freeman JB (2017) A neural mechanism of social categorization. *J Neurosci* 37:5711–5721.
- Tajfel H (1969) Cognitive aspects of prejudice. *J Biosoc Sci* 1:173–191.
- Tavares RM, Mendelsohn A, Grossman Y, Williams CH, Shapiro M, Trope Y, Schiller D (2015) A map for social navigation in the human brain. *Neuron* 87:231–243.
- Taylor SE, Fiske ST, Etcoff NL, Ruderman AJ (1978) Categorical and contextual bases of person memory and stereotyping. *J Pers Soc Psychol* 36:778–793.
- Van Bavel JJ, Cunningham WA (2010) A social neuroscience approach to self and social categorisation: a new look at an old issue. *Eur Rev Soc Psychol* 21:237–284.
- Vodrahalli K, Chen PH, Liang Y, Baldassano C, Chen J, Yong E, Honey C, Hasson U, Ramadge P, Norman KA, Arora S (2018) Mapping between fMRI responses to movies and their natural language annotations. *Neuroimage* 180:223–231.
- Volpert-Esmond HI, Bartholow BD (2021) A functional coupling of brain and behavior during social categorization of faces. *Pers Soc Psychol Bull* 47:1580–1595.
- Weaverdyck ME, Lieberman MD, Parkinson C (2020) Tools of the trade: multivoxel pattern analysis in fMRI: a practical introduction for social and affective neuroscientists. *Soc Cogn Affect Neurosci* 15:487–509.
- Wei W, Liu J, Dai R, Feng L, Li L, Tian J (2014) Different brain activations between own- and other-race face categorization: an fMRI study using group independent component analysis. In: *Medical imaging 2014: biomedical applications in molecular, structural, and functional imaging*, Vol 9038 (Molthen RC, Weaver JB, eds), pp 903807–903851. Bellingham, WA: SPIE.
- Wheeler ME, Fiske ST (2005) Controlling racial prejudice: social-cognitive goals affect amygdala and stereotype activation. *Psychol Sci* 16:56–63.
- Wiese H, Kloth N, Güllmar D, Reichenbach JR, Schweinberger SR (2012) Perceiving age and gender in unfamiliar faces: an fMRI study on face categorization. *Brain Cogn* 78:163–168.
- Yeo BT, Krienen FM, Sepulcre J, Sabuncu MR, Lashkari D, Hollinshead M, Roffman JL, Smoller JW, Zöllei L, Polimeni JR, Fisch B, Liu H, Buckner RL (2011) The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J Neurophysiol* 106:1125–1165.