Consistent Neural Activity Patterns Represent Personally Familiar People

Mark A. Thornton and Jason P. Mitchell

Abstract

■ How does the brain encode and organize our understanding of the people we know? In this study, participants imagined personally familiar others in a variety of contexts while undergoing fMRI. Using multivoxel pattern analysis, we demonstrated that thinking about familiar others elicits consistent fine-grained patterns of neural activity. Person-specific patterns were distributed across many regions previously associated with social cognition, including medial prefrontal, medial parietal, and lateral temporoparietal cortices, as well as other regions including the anterior and mid-cingulate, insula, and precentral gyrus. Analogous context-specific patterns were observed in medial parietal and superior occipital regions. These results suggest that medial parietal cortex may play a particularly central role in simulating familiar others, as this is the only region to simultaneously represent both person and context information. Moreover, within portions of medial parietal cortex, the degree to which person-specific patterns were typically instated on a given trial predicted subsequent judgments of accuracy and vividness in the mental simulation. This suggests that people may access neural representations in this region to form metacognitive judgments of confidence in their mental simulations. In addition to fine-grained patterns within brain regions, we also observed encoding of both familiar people and contexts in coarse-grained patterns spread across the independently defined social brain network. Finally, we found tentative evidence that several established theories of person perception might explain the relative similarity between person-specific patterns within the social brain network. ■

INTRODUCTION

The individual person is a fundamental unit of social life. To navigate our own social lives, we must have a detailed understanding of the individuals who populate it, including an ability to anticipate their idiosyncratic thoughts, feelings, and actions. As such, the mind and brain must represent knowledge of not just general human psychology—a lay theory of mind—but also of the particular psychologies of people who are important to us. Such person-specific knowledge could guide both direct interpersonal interactions as well as offline simulations of others' thoughts and actions. The present investigation aimed to characterize the neural encoding of our knowledge of personally familiar others. In addition to investigating which brain regions support person-specific representations, we examined how such representations contribute to the process of simulating other minds. In particular, we identified brain regions that are likely candidates for integrating person information with context information to form complete simulations of social events. We also tested the hypothesis that instating a "typical" person-specific pattern serves as a metacognitive clue to the predictive validity of such simulations. Finally, we examined several theories of person perception to determine which, if any, provide an appropriate taxonomy for our neural representations of personally familiar others.

Several previous studies have investigated the neural representation of personally familiar individuals. Regions including medial pFC, medial parietal cortex, the TPJ, and the STS have been consistently implicated in social thought, forming the so-called "social brain" network (for reviews, see Van Overwalle & Baetens, 2009; Mitchell, 2008). It is possible that many or all of these regions support person-specific representations, a position supported by studies examining the neural correlates of (person) familiarity (Cloutier, Kelley, & Heatherton, 2011; Gobbini, Leibenluft, Santiago, & Haxby, 2004). Evidence from an fMRI adaptation (i.e., repetition suppression) study also suggests that a broad array of brain regions may contain information specific to thinking about particular targets (Szpunar, Jacques, Robbins, Wig, & Schacter, 2014). However, both adaptation and activation-based fMRI studies have also suggested a more specific role for ventral medial pFC in representations of specific familiar others (Heleven & Van Overwalle, 2016; Welborn & Lieberman, 2015).

A reasonable question one might ask on the basis of such studies is to what extent the representation of other people is distributed across the social brain versus supported by a single, dedicated module? This debate over distributed versus modular functioning has long occupied cognitive neuroscientists generally (de Beeck, Haushofer,

Harvard University

Journal of Cognitive Neuroscience 29:9, pp. 1583–1594
doi:10.1162/jocn_a_01151

& Kanwisher, 2008). Understanding the functional brain architecture supporting representation of specific persons might help address questions of interest to psychologists. For example, if person-specific information appeared to be distributed throughout the social brain network, as previous studies have suggested, it would suggest that detailed knowledge about other people permeates our social cognitive machinery. In other words, it would hint that even social judgments that do not necessarily require any background knowledge may make use of person-specific information when it is available.

Identifying representations of individual people presents a challenge because traditional techniques for analyzing fMRI data are ill-suited to the task. The social brain network activates very differently in response to social and nonsocial stimuli, but imagining a set of familiar people would likely elicit comparatively similar levels of activity for each individual. Thus, traditional univariate approaches, which rely on differences in average activity levels to distinguish between stimuli, may prove unable to differentiate target people, even in brain regions that do contain person-specific representations. Fortunately, a nascent set of tools for exploring fMRI data may now make it possible to identify where these representations reside. These techniques, known collectively as multivoxel pattern analysis (MVPA), focus on fine-grained patterns of neural activity within brain regions, or coarse-grained activity patterns across multiple regions, rather than on gross average activity levels within individual regions (Haxby et al., 2001). As a result, MVPA can achieve greater sensitivity than traditional univariate statistical analyses and can detect multidimensional neural codes that may exist even in the absence of large-scale differences in average neural activity (Davis et al., 2014).

Here we apply the form of MVPA known as representational similarity analysis (RSA) to identify brain regions that demonstrate different patterns of neural activity when perceivers imagine different people (Kriegeskorte, Mur, & Bandettini, 2008). We entered into the investigation agnostic to the spatial scale at which person-specific information might be encoded, and thus, we applied RSA in search of both fine-grained intraregional activity patterns—via the use of searchlight mapping (Kriegeskorte, Goebel, & Bandettini, 2006)—and coarse-grained interregional patterns distributed across the social brain network as a whole.

An initial MVPA-based foray into the representations of individual people has already yielded promising results (Hassabis et al., 2014). In that study, patterns elicited by thinking about four fictional people in an episodic simulation task were reliably classified in two adjacent portions of dorsal medial pFC. The biographies of the fictional targets were written to create strong impressions of certain traits (agreeableness and extraversion). Additional brain regions were capable of these decoding individual personality traits, but only the medial prefrontal regions could distinguish between all four target people. It remains unclear whether the classification was being driven by the sort of rich, detailed knowledge that characterizes our representations of personally familiar others, or simply by the exaggerated social differences created by this experiment's central manipulation. To resolve this uncertainty, this study used a broad set of personally familiar targets.

This study was designed to cast light on more than just which brain regions support representations of specific people. When participants in our experiment imagined people they knew, they imagined them within a set of different contexts. This allowed us to assess which brain regions support context-specific patterns of neural activity and where representations of person and context identity overlap. The question of how knowledge of dispositions and situations are integrated when making attributions has been of long standing interest to social psychology (Gilbert & Malone, 1995; Jones & Harris, 1967). Recent neuroscientific investigations have suggested that increased activity in the social brain network predicts dispositional attributions, but have been inconsistent regarding the positive predictors of situational distributions (Moran, Jolly, & Mitchell, 2014; Kestemont, Vandekerckhove, Ma, Van Hoeck, & Van Overwalle, 2013). By focusing on fine-grained patterns rather than on raw activity levels, this study aimed to provide a clearer view of which region(s) are potential loci for the integration of situational and dispositional information. Although our design does not permit a detailed assay of the nature of such integration, localizing regions that support both types of information serves as a crucial first step in this direction.

By examining mental simulations of the same set of people over several iterations, we also aimed to shed light on the psychological process of imagining other people. In particular, we analyzed the patterns of activity during each simulation to reveal how perceivers arrive at metacognitive judgments of confidence in their imagination. Over the course of the study, participants were presumably able to integrate their knowledge of target people into some simulations better than others. For example, imagining an elderly relative having dinner with you is not particularly outlandish, whereas picturing that same person frolicking at the beach may prove rather more difficult. We hypothesized that the degree to which one can easily incorporate knowledge of a person into a simulation could signal how valid that simulation might be: that is, how likely to accurately reflect the real actions of a given person in a given situation. In the present paradigm, we measured the degree to which the pattern elicited on each trial resembles the patterns across all other trials featuring the same target person. This pattern typicality measure served as an indirect index of the incorporation of person knowledge into each simulation. We expected pattern typicality to be positively associated with trial-wise ratings of vividness and accuracy.

Finally, we planned to examine several theories which might explain the differences in person-specific neural representations. Determining how people naturally divide up the social world has produced many valuable theories in social psychology. Although these theories can each explain a variety of social behaviors, it is unclear which taxonomies characterize the information that perceivers spontaneously draw upon to simulate others' minds. To address this question, we examined a wide range of theories as candidate explanations for the similarities between person-specific activity patterns. These candidates included five major conceptions of person perception, self-reported and implicit measures of holistic similarity, and in-scanner ratings of accuracy and vividness. The extant social theories were the following: categorization of basic social groups (age, race, sex); the Five-Factor Model of Personality (Goldberg, 1990; McCrae & Costa, 1987), consisting of openness, conscientiousness, extraversion, agreeableness, and neuroticism; the dimensions of warmth and competence, which characterize the Stereotype Content Model (Fiske, Cuddy, Glick, & Xu, 2002); egocentric factors (similarity, familiarity, and liking); and dyadic relational model (sharing, trading, or ordering) between target and participant (Haslam, 2004; Haslam & Fiske, 1999).

Each of these different characterizations of the target people makes different predictions about which targets should elicit more similar or more different neural activity patterns. For example, two target people might belong to the same basic social groups but have very different personalities on the Big 5 factors. The social groups model would thus predict similar patterns of neural activity for these two targets, whereas the Five-Factor Model would predict relatively dissimilar patterns. Using RSA (Kriegeskorte, Mur, & Bandettini, 2008), we were able to compare these theory-based predictions about intertarget similarity with the actual similarity between patterns of neural activity. Thus, we could test hundreds of predictions simultaneously to determine which theory best accounts for the observed neural data. Directly comparing the theories to each other would require much more statistical power, and so in this study, we instead focused on examining whether there was evidence for each theory's influence on pattern similarity at all. We thus attempted to select as broad a range of theories as possible rather than a set of highly similar, competing theories.

METHODS

Raw behavioral data, processed imaging data, and custom experiment presentation and statistical analysis code for this study are freely available on the Open Science Framework (osf.io/gxhkr/). Raw imaging data have been deposited at the Harvard University Dataverse (dx.doi.org/10.7910/DVN/ZQQABJ). Shared data have been stripped of identifying features for the privacy of the participants.

Participants

A power analysis was conducted via Monte Carlo simulation to assess the number of participants and trials needed with the design specified below. This analysis targeted the final planned RSA, in which intertarget pattern similarity was to be modeled in terms of existing theories of person perception. We targeted this analysis rather than the primary trial-wise RSA to avoid the additional uncertainty of estimating context and run effect sizes at the trial level. Simulated patterns of neural activity were generated to embody an anticipated effect size of $r = .15$ (correlation between model and neural data), which we judged to be reasonable based on previous research (Kriegeskorte, Mur, & Bandettini, 2008; Kriegeskorte, Mur, Ruff, et al., 2008). This was achieved by creating a noise-free "model" pattern $\sim N(0, 1)$ and adding noise to create a simulated neural pattern matrix with the appropriate relationship to the model. A simulation-based search across noise parameters determined that a $\sim N(0, 2.4)$ distribution of noise yielded the appropriate effect size.

Additional, independently generated noise $\sim N(0, 3.4)$ was then added to these simulated neural patterns to produce patterns of activity for each simulated experimental trial. The trial-wise noise parameter was calculated by adjusting the same parameter used in a similar power analysis for a previous study (Tamir, Thornton, Contreras, & Mitchell, 2016) to account for the difference in trial duration (12.5 sec modeled in this study vs. 4.25 in the previous study). Each simulated pattern set consisted of 20 (target people) by 200 (voxel) elements. Because we could not anticipate the size of activations in advance, voxel count was set to near the average searchlight size. Patterns were then subjected to RSA as described below under imaging procedures, producing 20 × 20 similarity matrices for each simulated participant that could be compared with the correlation matrix of the original "model" pattern. To simplify computation, the general linear model (GLM) phase of this analysis was reduced to averaging the patterns within each participant across trials. This process was repeated 100 times with simulated participant numbers ranging from 2 to 25 and trial numbers ranging from 2 to 10 per target. These simulations indicated that 25 participants with six trials per person should be adequate to ensure 95% statistical power at a threshold of $\alpha = .001$.

Participants ($n = 25$) were recruited from the Harvard University Psychology Study Pool. Two were subsequently excluded: one due to a neurological anomaly detected during scanning and the other due to excessive head motion. All remaining participants (15 women; age range = 18–27 years, mean age = 21.1 years) were right-handed, neurologically normal, fluent in English, and had normal or corrected-to-normal vision. Participants provided informed consent in a manner approved by the committee on the use of human subjects in research at Harvard University.

Stimuli and Behavioral Procedure

Pretesting

We selected a set of 20 target stimuli to maximize observable differences in the imaging experiment. Participants first provided the names of 40 personally familiar adults. To choose a highly diverse set of targets, and to provide measures for RSA, we asked participants to rate these people on 16 social dimensions. These included Big 5 personality traits (Goldberg, 1990; McCrae & Costa, 1987), warmth and competence (Fiske et al., 2002), the degree to which their relationship with the person was predicated on communal sharing, equity matching, authority (Haslam, 2004; Haslam & Fiske, 1999), and similarity, familiarity, and liking. Participants also indicated the race, sex, and age of each target. Single-item ratings on 7-point Likert scales were used for all dimensions other than the demographics. Multi-item scales would have been preferable psychometrically but inordinately time-consuming and exhausting for participants. We provided definitions of the dimensions to ensure that participants used the scales consistently. The dimensions were presented in a unique random order for each participant, and the order of the target people was randomized for each dimension.

These ratings were used to select a subset of 20 target people who were diverse across all 16 dimensions. The selection process proceeded as follows: first the target most different (in Euclidean distance) from centroid of the group across the 16 measured dimensions was selected; on each subsequent iteration of the algorithm, the target person who would maximize the theory-weighted average standard deviation of the selected set would be added to it. The algorithm terminated once 20 targets had been selected. Although this procedure was not guaranteed to produce an optimal selection, it was very rapid to compute—a tradeoff we deemed worthwhile given that it had to be performed while participants waited. This approach succeeded in producing an increase in intertarget variance in 20 of 23 participants and yielded only minor decreases in variance in the other three.

Participants also provided holistic similarity measures in a separate behavioral task. On each of 380 trials, participants compared two target people to one reference target and indicated which of the two was more similar to the reference. The trial number allowed for two presentations of each unique pair of target people (in the nonreference positions). Participants' choices were used to derive a self-reported explicit holistic similarity matrix between the targets. This similarity matrix was generated by summing the times a particular pair of reference and selected comparison targets were judged to be the more similar of the two possible pairs and dividing this sum by the number of possible trials on which this judgment might have occurred. RTs were used to generate an equivalent implicit holistic similarity matrix, with longer RTs indicating greater similarity between the two choice targets.

Experimental Paradigm

Just before scanning, participants were asked to imagine each of six common situations: taking a long road trip, shopping for groceries, going to a fair, going to the beach, waiting for the doctor, and having dinner at a restaurant. They were instructed to imagine these situations as vividly as possible from their own perspective. During fMRI scanning, participants simulated each of the 20 target people in each of these contexts (Hassabis et al., 2014). Each trial began with a prompt (2.5 sec), indicating which context the participant should simulate (e.g., "eating at a restaurant"). Next, the name of 1 of the 20 targets appeared (2.5 sec), after which the screen went blank and participants had 10 sec to simulate that target in the specified context. Participants were instructed not to imagine new situations on each trial, but instead to place each target person in the same previously imagined context.

After the simulation period, participants rated the vividness of their simulation on that trial (2.5 sec) and how accurate they felt their simulation was with respect to what the target person would actually do in that context (2.5 sec). These ratings were averaged to form a "confidence" composite. The six contexts were fully crossed with the 20 target people for a total of 120 trials divided across six runs of 400 sec each. Note that each combination of target person and context occurred only once, making the task a unique trial design. Each target person was shown only once within each run, but the position of a given target person within a run was randomly determined (independently for each participant). The order of contexts was independently randomized for each target person within each participant. All experimental tasks were presented using Python 2.7 and the PsychoPy package (Peirce, 2007).

Imaging Procedure

Functional gradient-echo echo-planar images were obtained from the whole brain (43 interleaved slices of 2.5 mm thickness parallel to the AC–PC line, repetition time = 2500 msec, echo time = 30 msec, flip angle = 90°, in-plane resolution = 2.51 × 2.51 mm, field of view = 216 mm^2, matrix size = 86 × 86 voxels, 160 measurements per run) using a parallel imaging protocol and prospective acquisition correction. A high-resolution T1-weighted structural scan (multiecho MPRAGE, 1.195 mm isometric voxels, matrix size 192 × 192 voxels, 144 sagittal slices) was also acquired from each participant. Imaging data were acquired at the Harvard University Center for Brain Science with a 3-T Siemens Tim Trio scanner (Siemens, Erlangen, Germany) using a 32-channel head coil. Functional images were preprocessed and analyzed using SPM8 (Wellcome Department of Cognitive Neurology, London, UK) with the SPM8w extension (https://github.com/ddwagner/SPM8w) and in-house scripts in

MATLAB (The MathWorks, Natick, MA) and R (www.R-project.org/). Data were first spatially realigned via rigid body transformation to correct for head motion and then normalized to a standard anatomical space (2 mm isotropic voxels) based on the ICBM 152 brain template (Montreal Neurological Institute). The GLM was used to analyze each participant's data in preparation for MVPA. Two separate GLMs were run to allow for analysis at different levels: with respect to individual trials and at the level of target people.

In the first of these GLMs, each trial in the experiment was modeled separately (120 conditions of interest) as a boxcar regressor starting at the beginning of the target person presentation period and ending at the completion of the imagination period (12.5 sec). Note that modeling only this period of each trial allowed for relatively long (7.5 sec) intertrial intervals, minimizing multicollinearity in the design matrix. The boxcars regressors were convolved with a canonical hemodynamic response function and entered into the GLM. The model also included additional covariates of no interest: run trends and means, six motion realignment parameters, and outlier time points. The second GLM included only 20 conditions of interest, corresponding to the 20 personally familiar target people. Thus, each boxcar regressor modeled six trials, corresponding to imagining each target person in each of the six contexts presented. Otherwise, this GLM proceeded identically to the first, with the exception of adding temporal and dispersion derivatives of the conditions of interest to the model.

Searchlight Analyses

We extracted local patterns of neural activity from throughout the brain using an approximately spherical searchlight with four voxel radius (~9 mm). To ensure that the edges of the brain were included, searchlights with up to half of their voxels implicitly masked (i.e., outside the brain) were analyzed. Thus, searchlight size varied between 128 and 257 voxels. For each voxel in the brain, patterns of activity (*t* value maps from the first GLM) were extracted from the surrounding searchlight area for each condition of interest (trial in the experiment). We estimated the similarity between activity patterns by calculating the Pearson correlation between the values in each pair of patterns. These estimates were then rank transformed and regressed onto two binary variables encoding our hypotheses. Rank transformation is a recommended procedure for RSA, as it helps mitigate large-scale differences and violations of distributional assumptions that may otherwise compromise results (Kriegeskorte, Mur, & Bandettini, 2008). The person-representation variable predicted high similarity between trials with the same target person and low similarity between trials with different targets. Analogously, the context-representation variable predicted high similarity between trials with the same context and low similarity between trials with different contexts. We also included a similarly constructed run nuisance variables to control for pattern similarity between trials in the same run.

This regression analysis produced whole-brain unstandardized beta maps for each participant. These maps were subjected to smoothing with a Gaussian kernel (4 mm FWHM) to maximize intersubject alignment and were then entered into a random effect analysis across participants using one-sample *t* tests. This analysis was corrected for multiple comparisons using a MATLAB implementation (https://github.com/markallenthornton/MatlabTFCE) of maximal statistic permutation testing with threshold-free cluster enhancement (Smith & Nichols, 2009). The family-wise error rate was strictly controlled across three random effects *t* tests from the two searchlight analyses (the person- and context-representation mappings described above and the confidence analysis described below) by setting the permutation-corrected critical *p* value within each map to a Bonferroni-adjusted α = .0167. Results were rendered on the cortical surface using Connectome Workbench (Marcus et al., 2011).

Person-pattern typicality was estimated for each trial by averaging the similarity values between the trial in question and the other trials involving the same target. High typicality for a trial thus meant that the pattern for this trial was highly correlated with the patterns for the other trials on which the same target was presented. Each measure in the pattern typicality vector was the mean of five correlations (since each person was imagined in six contexts total). For each searchlight region, the pattern typicality vector was entered into a multiple regression as the dependent variable to be predicted by the confidence composite ratings. Additional nuisance regressors for target person were also included to control for the possibility that some targets might elicit more reliable patterns than others. As with the primary searchlight, this analysis produced whole-brain unstandardized beta maps for each participant. These maps were smoothed with a 4-mm FWHM Gaussian and then entered into random effects *t* tests, with multiple comparisons corrected as described above.

Feature-selected Representation Similarity Analysis

As reported in detail below, the person-representation searchlight analysis revealed an extended set of regions containing fine-grained person-specific activity patterns. The statistically significant regions in this analysis were used as a mask to select voxels for further analysis. Note that this selection process, despite drawing on dependent data, does not constitute statistically biased "double-dipping" for the purposes of the analyses we apply. This is because the independent variables that will be used to model the feature-selected patterns (i.e., participant ratings on the 16 social dimensions) were not themselves involved in voxel selection. Although on average this procedure will increase the observed size of real effects by disattenuating correlations, it will not systematically

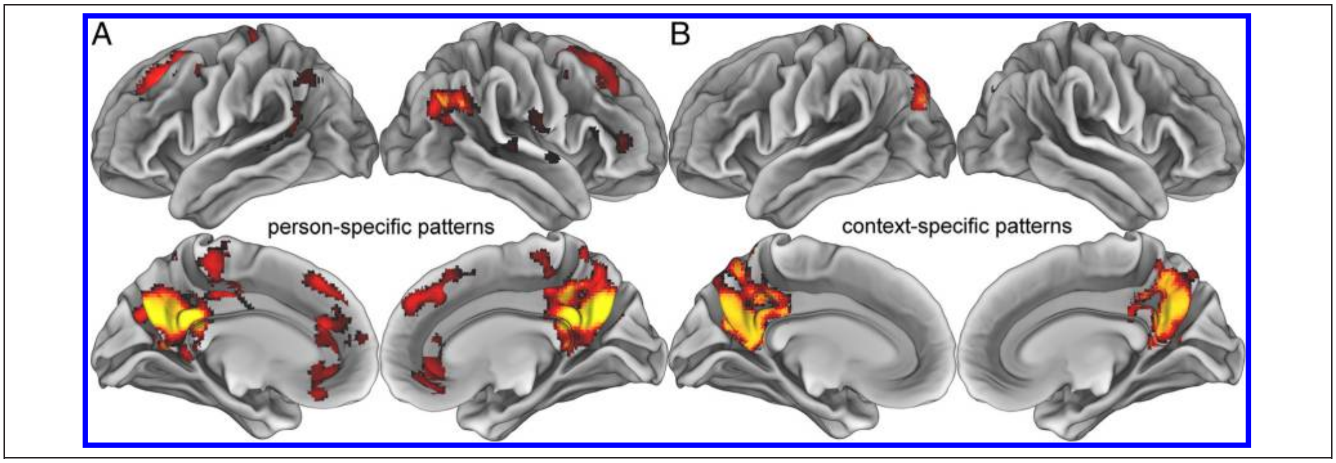

Figure 1. Person- and context-specific patterns. Consistent patterns of neural activity were elicited in the red/yellow regions during mental simulation for target people (A) or the context in which they were imagined (B). Results from searchlight mapping were combined across participants via *t* test and corrected for multiple comparison via maximal statistic permutation testing with threshold free cluster enhancement.

lead to the generation of spurious relationships between the models and the data. To avoid circularity, we do not test any models using the full deconvolution (trial-wise) GLM within this mask.

In addition to this feature selection approach—which should produce minimally reliability-attenuated estimates—we also repeated the same set of analyses using an independently defined mask from previous research (Tamir et al., 2016). The voxels in this mask were chosen by univariate differentiation of mental state concepts. Using this independent mask justifies stronger claims about the localization of representations within the regions sensitive to mentalizing per se and allows us to test for the presence of person-specific patterns at the trial level without circularity. This alternative feature selection approach also allows us to demonstrate the unbiased nature of our reliability-based feature selection.

For the 15,849 retained voxels in the reliability selection (Figure 1A and Table 1) and the 25,215 voxels in the independent mask (Tamir et al., 2016), patterns of

Table 1. Peak Voxel and Cluster Size for All Regions Obtained from the Searchlight Mappings (p < .05, Corrected)

Map/Anatomical Label	*x*	*y*	*z*	*Volume*	*Peak p*
Person-specific Patterns					
Medial parietal cortex/right TPJ	−2	−45	22	8651	<.0001
Medial pFC	−18	25	44	5256	.0045
Left TPJ	−44	−51	18	883	.0121
Right inferior frontal gyrus	40	39	4	651	.0115
Right STS	48	−33	−6	391	.0130
Right posterior insula	30	−25	12	17	.0163
Context-specific Patterns					
Medial parietal cortex/left superior occipital cortex	−6	−53	14	5968	.0010
Right superior occipital cortex	32	−75	32	207	.0121
Confidence Pattern Typicality Correlation					
Precuneus	14	−65	38	551	.0053
Posterior/mid-cingulate	2	−21	34	372	.0120
Precuneus	22	−51	44	15	.0158

Coordinates refer to Montreal Neurological Institute stereotaxic space. Volume refers to voxel number. Peak p indicates threshold-free cluster enhancement-corrected p value at peak.

neural activity (unstandardized beta maps) were extracted from the second GLM: for each participant, one pattern of activity across the feature-selected regions for each target person. Similarity estimates between these neural representations were again calculated by Pearson correlating the activity patterns. These estimates were then Spearman correlated with predictions of intertarget similarity made by models described above: the five theories of person perception, the two measures of holistic similarity, and the Euclidean space described by in-scanner ratings of vividness and accuracy, averaged by target person. Predictions of (dis)similarity between targets for each of the five theories of person perception were calculated by taking the Euclidean distance between ratings of target people within the *n*-dimensional space described the dimensions of the theory. In the case of the basic social groups theory, because the distance between categories is not a calculable quantity, the dissimilarity matrix was binary, with values indicating whether a pair of targets belonged to the same (sex, age, or race) group or not.

This RSA generated correlation coefficients for each of the seven models within each participant. The statistical significance of these results was assessed through complementary direct and indirect testing procedures. One-sample *t* tests across participants were conducted on the coefficients for each model. Fisher's *r*-to-*z* transformation was applied before the correlations before conducting these *t* tests. Additionally, nonparametric percentile bootstrapping was used to obtain robust 95% confidence intervals around the mean correlation coefficient for each model. Both testing procedures were conducted separately for the two feature selection methods.

The independent feature selection method allowed us to perform an additional test of our primary hypothesis: that patterns of activity in the social brain network encode personally familiar people. Repeating this analysis in the feature-selected regions had two primary advantages: It allowed for the unbiased estimation of a single summary measure of effect size, and it allowed us to examine whether person-specific patterns persist across spatial scales in the social brain network. This analysis was achieved by repeating the multiple regression RSA described above. However, in this case, the entire set of voxels in the independent mask was used to produce a single results neural similarity matrix for each participant. Results were combined across participants via random effects *t* tests on the regression coefficients from each participant. To determine whether this analysis depended on the same fine-grained patterns as the searchlight analysis or instead relied on coarse-grained patterns, it was repeated with unsmoothed and smoothed patterns of brain activity. The heavy degree of smoothing applied in the latter case (18 mm FHWM, equal to the diameter of the searchlight) served to ensure that any effects observed could be attributed to coarse-grained interregional patterns rather than fine-grained intraregional patterns.

RESULTS

Searchlight Results

Consistent patterns of neural activity for familiar others were observed across wide areas of cortex (Figure 1A). These areas primarily overlapped with regions previously implicated in social cognition, including medial pFC (dorsal and ventral), the TPJ and STS bilaterally, and most robustly, medial parietal cortex including portions of both precuneus and posterior cingulate (Table 1). Person-specific patterns were also observed in right ventral lateral pFC, bilateral dorsal pFC, medial precentral gyrus, and a portion of the posterior insula.

Context-specific patterns were observed in a more circumscribed set of regions (Figure 1B). These included medial parietal cortex and bilateral superior occipital cortex (Table 1). Medial parietal cortex—precuneus in particular—has long been implicated in mental imagery (Cavanna & Trimble, 2006; Fletcher et al., 1995), which may account for its role in representing context. The occipital areas may overlap with or be adjacent to the transverse occipital sulcus, which has been implicated in scene perception (Bettencourt & Xu, 2013). This would be consistent with their implication in context individuation in this study. Note that the only region of overlap between person- and context-specific patterns was in the medial parietal cortex.

A secondary searchlight analysis aimed to assess the relation between person-specific pattern typicality and by-trial ratings of confidence (accuracy and vividness). A positive relation between pattern typicality and confidence was detected in three adjacent regions within medial parietal cortex (Table 1). The most anterior of these regions was placed along the posterior cingulate gyrus, whereas the other two areas were located primarily within the precuneus (Figure 2). Within these regions, the more typical a pattern was for a particular target the more confidence participants would express in their simulation on a given trial. This result suggests both neural and psychological interpretations. Neurologically, it suggests that medial parietal lobe plays a particularly important role in actively "running" mental simulations or at least making them consciously available. Psychologically, this result suggests that people may use the degree to which they can integrate idiosyncratic person knowledge into a simulation as an indicator of simulation validity.

Feature-selected Representational Similarity Results

Person-specific activity patterns were extracted from the significant voxels in the corresponding searchlight analysis (Figure 1A) or using an independent mask for voxels sensitive to mental state representation (Tamir et al., 2016). The similarity between person-specific patterns was modeled in terms of feature sets via RSA. The results of this analysis provided consistent evidence that simulation

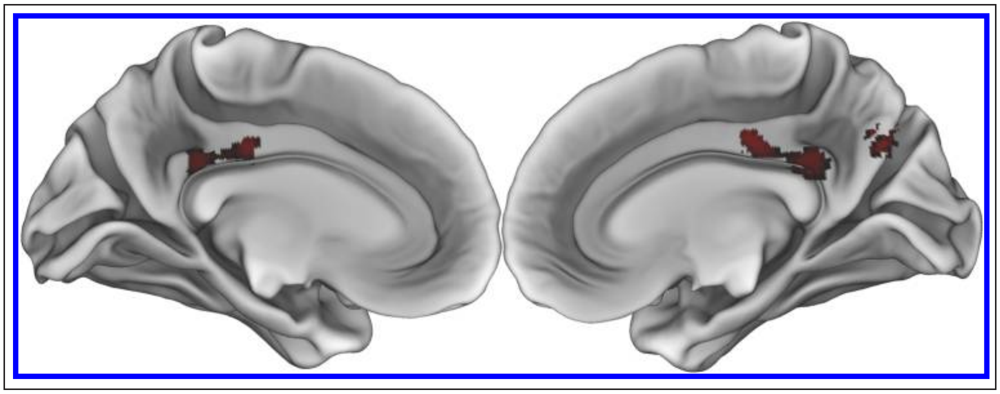

Figure 2. Pattern typicality predicts confidence. The regions in red showed a positive relation between the instatement of typical person-specific patterns and by-trial ratings of simulation vividness and perceived accuracy. Results from searchlight mapping were combined across participants via *t* test and corrected for multiple comparison via maximal statistic permutation testing with threshold free cluster enhancement.

confidence and holistic explicit similarity judgments predicted the neural representations of personally familiar others (Figure 3).

The average confidence (vividness and accuracy) of simulations involving each target was the most robust predictor of pattern similarity for both reliability-selected (mean r = .08, d = .66, p = .005) and independent (mean r = .09, d = .82, p = .0007) versions of the analysis. Although not contributing theoretical detail, self-reported behavioral ratings of holistic similarity also predicted pattern similarity in both the reliability selected regions (mean r = .03, d = .48, p = .03) and the independently selected regions (mean r = .03, d = .58, p = .01). Holistic implicit similarity was descriptively weakly positively associated with pattern similarity, but not at a significant level for either type of feature selection (ps > .1).

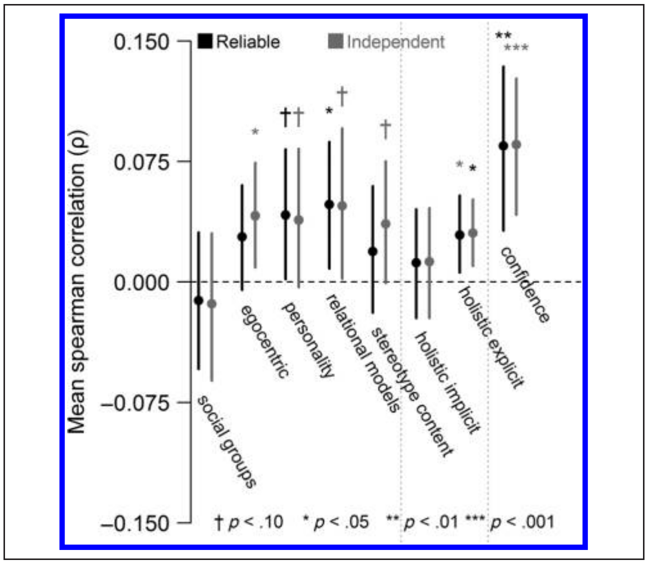

Figure 3. RSA. Person-specific patterns of neural activity across the feature-selected voxels were modeled in terms of a number of possible predictors. Two different forms of feature selection were used: Reliability-based selection was performed by using voxels that significantly encoded target people in the earlier searchlight results, and an independent mask of voxels sensitive to mental state representation was taken from earlier work (Tamir et al., 2016). Results were combined across participants via *t* tests on *r*-to-*z* transformed correlations. Error bars indicate 95% confidence intervals from percentile bootstraps of the mean (untransformed) correlation values. Dashed vertical lines separate models derived from different procedures: The five models on the left were based on self-report ratings, holistic similarity measures were calculated based on RT and choices in a triplet similarity judgment task, and the confidence model was derived from in-scanner ratings following each trial.

Results indicated weak evidence for the influence of several theories of person perception on neural pattern similarity (Figure 3). Egocentrism appeared to have a significant effect in the independent regions (mean r = .04, d = .50, p = .03) but was slightly less correlated with pattern similarity in the reliability selected areas and was not a statistical significant predictor therein (mean r = .03, d = .34, p = .12). Big 5 personality traits predicted pattern similarity at a marginally significant level under both the reliability-based (mean r = .04, d = .41, p = .06) and independent (mean r = .04, d = .36, p = .098) feature selection regimes. Relational models theory was a significant predictor of target pattern similarity in the reliability-based analysis (mean r = .05, d = .49, p = .03) but was only marginally significant for voxels within the independent mask (mean r = .05, d = .40, p = .07). Stereotype content was a marginally significant predictor, though only in the independent feature selection analysis (mean r = .04, d = .38, p = .08) and not the voxels chosen via reliability (mean r = .02, d = .19, p = .36). Similarity in terms of basic social groups (age, race, and sex) was slightly negatively correlated with pattern similarity, though the results were not significantly different from zero.

Despite some differences in categorical statistical significance due to near-threshold p values, the representationally similarity analysis results were generally highly similar across the two feature selection methods. Direct comparison between the theories was not undertaken, as the study was not designed with sufficient power for that purpose. Descriptive comparison of the models should also be approached cautiously, as differences in effect size may result from differences in reliability across models or—in the case of the accuracy and vividness judgments—greater psychological proximity to the simulations.

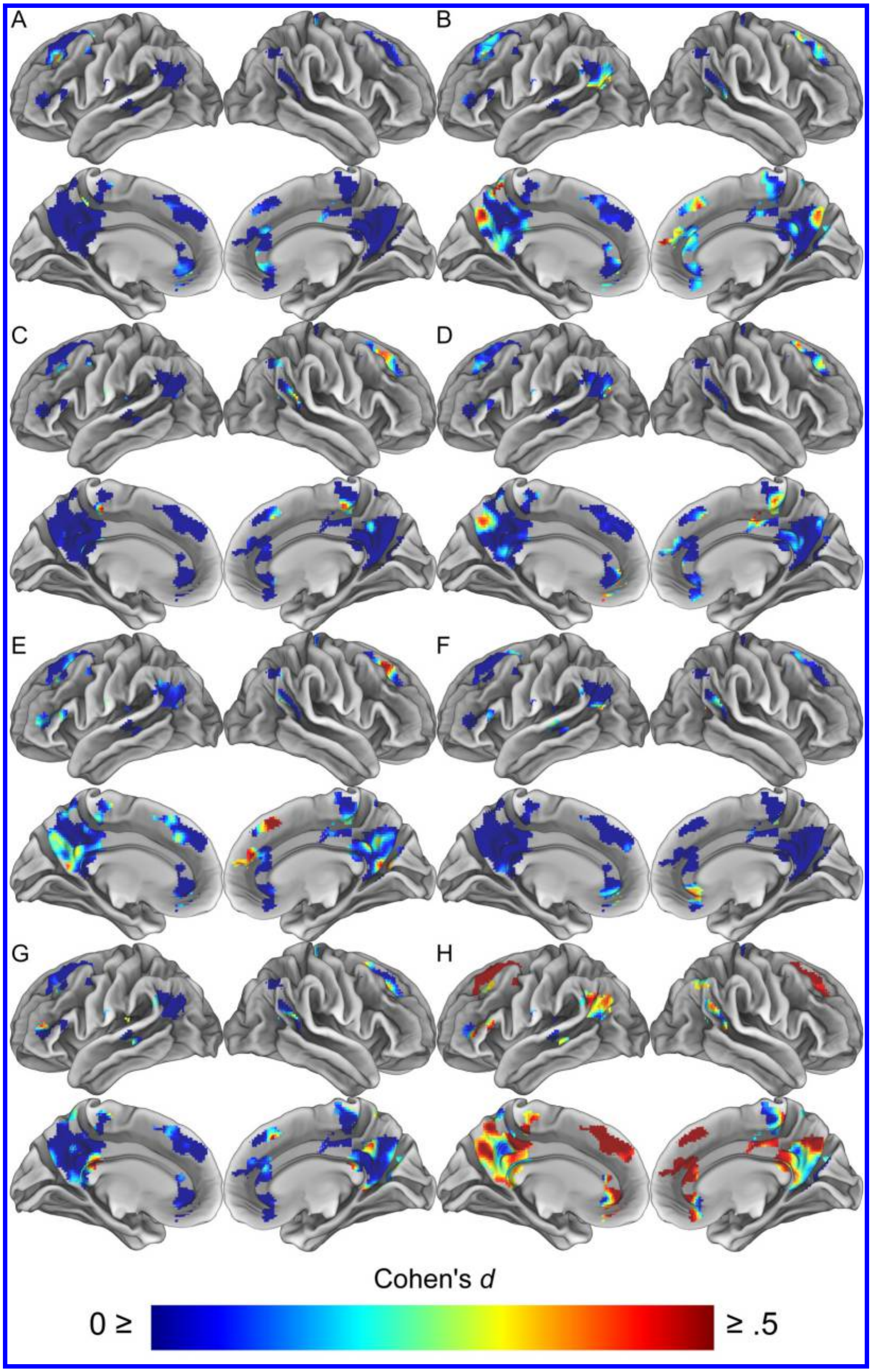

Figure 4. Descriptive mapping of theory effect sizes. Each map reflects the across-participant Cohen's *d* of one of the theories of person perception tested in the feature-selected RSA: (A) social groups, (B) egocentrism, (C) Big 5 personality traits, (D) relational models theory, (E) stereotype content model, (F) holistic implicit similarity, (G) holistic explicit similarity, and (H) confidence. Results reflect a whole-brain searchlight, masked by the voxels found to contain person-specific information in the primary searchlight analysis (i.e., the same voxels used in the reliability-selected RSA). The mapping is descriptive rather than inferential and thus is not corrected for multiple comparisons across voxels.

Pattern similarity on a trial-wise basis was analyzed within the voxels selected by the independently-defined mask. This multiple regression RSA mirrored the primary searchlight analysis described above, although over a larger spatial scale. Within these areas—previously identified as sensitive to mental state representation (Tamir et al., 2016)—we found that pattern similarity between experimental trials reflected the influence of both target person identity ($d = .89, p = .0003$) and context identity ($d = .78, p = .001$). Although not a hypothesis of substantive interest, we also note that the standardized effect of run on pattern similarity was very large ($d = 2.84$). The substantive effects were very similar in magnitude when heavily smoothed patterns were instead input into the RSA: person identity $d = .89, p = .0003$ and context identity $d = .68, p = .004$. These results indicate that the information contained in fine-grained patterns in the searchlight analysis is recapitulated at the coarse spatial scale of interregional differentials in the social brain network.

To further probe this result, we attempted to create a descriptive mapping of which voxels contributed most to the representation of person identity in the smoothed activity patterns. To this end, we repeatedly divided the independently feature-selected regions into 252 random parcels of approximately 100 voxels each. We repeated the person- and context-identity regression RSA using pattern similarity calculated separately across with each parcel of voxels. The regression coefficients could be tabulated on a voxelwise basis across iterations to determine the typical contribution of each voxel person-encoding coarse-grained patterns. Within each participant, the contribution of individual voxels could be calculated to arbitrary reliability across repeated simulation. The average split-half reliability with respect to voxels across 1000 parcellations was .83. However, the location of these voxels was not at all consistent across participants: The split-half reliability with respect to voxels across 23 participants was −.06. This result suggests that the coarse-grained patterns encoding person identity may rely on specific voxels within each brain, but if so, the location of these voxels is idiosyncratic. In other words, the spatial distribution of person-specific pattern codes appears to be uniformly distributed across the social brain network, at least at the level of the population.

We also computed a voxelwise descriptive mapping of standardized effect sizes for each of the substantive psychological theories tested in the feature-selected RSA. This analysis was conducted by repeating analysis testing these theories exactly as described above, but within the context of a searchlight analysis. The resulting correlation maps reflected the performance of each model in explaining pattern similarity within searchlights centered at each voxel in the brain. The maps were smoothed with a 4-mm FWHM kernel and then combined across participants using the formula for Cohen's d. The resulting group maps were masked by the significant results in the primary person-representation searchlight analysis (Figure 1), on the principle that meaningful theory fits could only occur in regions where person-specific patterns actually existed. The results (Figure 4) provide an indication of which regions may have particularly contributed to the performance of each theory in the feature-selected RSA.

Behavioral Results

During the imaging task, participants responded to the vividness probe on 92% of trials on average (*SD* = 8.1%) and responded to the accuracy probe on 95% of trials on average (*SD* = 6.0%). On average, 89% of trials had responses on both items and 98% of trials had responses on at least one item. The high response rates indicate that participants were consistently engaged with the task, especially considering the brief response windows. The mean vividness rating was 3.41 and the average of the standard deviations within each participant was .94. The mean accuracy rating was 3.30, and the average within-participant standard deviation was 1.07. The average correlation between vividness and accuracy was .67 (*SD* = .16), suggesting that these measures tapped related phenomenon, but were not completely redundant. Participants' ratings of targets yielded correlated results across several different conceptions of person perception (Figure 5), indicating that these social theories made comparable predictions.

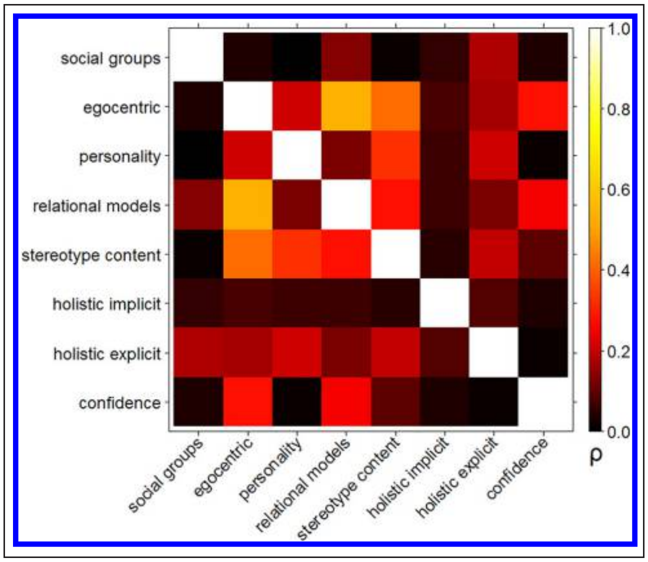

Figure 5. Correlations between potential predictors of neural similarity. Behavioral and self-report measures of intertarget similarity used in RSA were Spearman correlated with each other within each participant. Mean correlations across participants are shown in the heatmap, with larger values in warmer colors.

DISCUSSION

This study examined the neural representation of personally familiar people. Results of searchlight MVPA indicated that, during mental simulation, target person-specific patterns of neural activity were widely distributed across the social brain network. Person-specific activity patterns overlapped with context-specific patterns in only one area of the brain—medial parietal cortex—suggesting that this region may play a key role in combining person-specific knowledge with information about context in mental simulations. The central role of medial parietal cortex was further underscored by the fact that portions of this region showed a positive relation between pattern typicality and metacognitive perceptions of confidence in simulations of others. This relation suggests that the degree to which person knowledge is integrated into simulations by medial parietal cortex serves as a cue to how confident one should feel about those simulations—that is, how likely they are to predict the course of real world events. Finally, we observed that several measures of interpersonal similarity weakly predicted neural pattern similarity across brain regions that manifest person-specific patterns. The most consistent of these predictors were confidence in simulations (vividness and accuracy) and explicit (self-reported) perceptions of holistic similarity between target people.

The question of whether neural processes and representations are confined to discrete modular regions or are distributed across larger cortical networks has long been of interest to cognitive neuroscientists (de Beeck et al., 2008). The present research considered a particular social cognitive version of this issue: that is, whether person-specific information is widely distributed or confined to a single repository. Although many previous investigations of familiar people have concluded that such knowledge is broadly distributed (Szpunar et al., 2014; Cloutier et al., 2011; Gobbini et al., 2004), a few have emphasized the role of ventral medial pFC in particular (Heleven & Van Overwalle, 2016; Welborn & Lieberman, 2015). Our results fall firmly in line with former set of findings.

The current results indicate the presence of person-specific patterns in the social brain network, as defined by sensitivity to differences in others' mental states by a previous study. However, it is worth noting that not all of the regions implicated by our searchlight analysis are contained within the social brain as commonly defined. For instance, person-specific patterns were also detected within portions of the midcingulate, precentral gyrus, and insula—none of which are generally taken to be socially specific. Person-specific patterns in medial pFC also extended to quite posterior coordinates, overlapping considerably with the ACC. Given ACC's role in error and conflict monitoring (Kerns et al., 2004), this might suggest a role for ACC in keeping social simulations "on track"—that is, maximally plausible for individual target people.

The results of this study emphasize the central role that medial parietal cortex, including the precuneus and posterior cingulate, plays in mental simulations of other people. We have observed that (1) this region supports person-specific representations of familiar others, (2) this region also represents context-specific information, and (3) the presence of typical person-specific patterns in this region predicts vividness and accuracy judgments. The overlap between person and context representation hints that this region may contribute to integrating these forms of information, although this study does not provide direct evidence of this integration. The presence of context representations in medial parietal cortex—and particularly retrosplenial cortex—is highly consistent with previous research (Bar, 2004). However, medial parietal cortex has been associated with diverse mental functions over the history of cognitive neuroscience (for a review, see Cavanna & Trimble, 2006), so it seems unlikely that it is specifically devoted to social simulation. The pattern typicality results, however, do implicate it in supporting metacognitive access to such simulations when they occur.

RSA of person-specific patterns in the social brain network suggested several factors that may shape our mental simulations of other people. Holistic interpersonal similarity among the targets significantly predicted pattern similarity, suggesting that participants did have conscious access to some of the features contributing to their mental simulations. Confidence (vividness and accuracy) was the overall best predictor of neural pattern similarity between the target people. However, the explanatory success of confidence may be exaggerated by the fact that these ratings were interspersed in the simulation task itself.

Evidence for more specific influences on pattern similarity was weak and inconsistent. Relational models theory (Haslam, 2004; Haslam & Fiske, 1999), which provides a taxonomy of dyadic relationships based on their resource exchange logic, achieved the best performance descriptively but was still only marginally significant within independently selected voxels and statistically indistinguishable from the other theories tested. If relational models do shape patterns of activity during mental simulation, this would suggest that other people's functional role with respect to ourselves plays an important role in determining how we imagine them acting. We also observed weak evidence in support of the influence of several other theories including the Five-Factor Model of Personality (Goldberg, 1990; McCrae & Costa, 1987), egocentrism, and the stereotype content model (Fiske et al., 2002). However, the marginal nature of these findings suggests we should be highly tentative in drawing conclusions about which particular factors shape our mental representations of personally familiar people.

Altogether, the present study used advanced neuroimaging methods to further our understanding of the representation of familiar others. We found evidence for consistent fine-grained patterns of neural activity represent individuals in large sets of personally familiar targets, suggesting that the brain uses a distributed population-coding approach to support person knowledge. Moreover, we found that these person-specific patterns are widely distributed throughout the social brain network, rather than concentrated in a single region. Notably, we also detected coarse-grained person- and context-specific activity patterns, which spanned the social brain network. This suggests that social information encoded at one spatial scale may be recapitulated at other spatial scales. Finally, we found evidence that medial parietal cortex may contribute to the simulation of others' minds in two unique ways: by potentially integrating person and context knowledge and by providing metacognitive cues to simulation validity.

Acknowledgments

The authors would like to thank Rita Ludwig, Annie Rak, Stephanie Sasse, and Ava Zhang for their assistance. M. A. T. was supported by a National Science Foundation Graduate Research Fellowship (DGE1144152) and The Sackler Scholar Programme in Psychobiology.

Reprint requests should be sent to Mark Allen Thornton, Department of Psychology, Harvard University, William James Hall, 33 Kirkland Street, Cambridge, MA 02138, or via e-mail: mthornton@fas.harvard.edu.

REFERENCES

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience, 5,* 617–629.

Bettencourt, K. C., & Xu, Y. (2013). The role of transverse occipital sulcus in scene perception and its relationship to object individuation in inferior intraparietal sulcus. *Journal of Cognitive Neuroscience, 25,* 1711–1722.

Cavanna, A. E., & Trimble, M. R. (2006). The precuneus: A review of its functional anatomy and behavioural correlates. *Brain, 129,* 564–583.

Cloutier, J., Kelley, W. M., & Heatherton, T. F. (2011). The influence of perceptual and knowledge-based familiarity on the neural substrates of face perception. *Social Neuroscience, 6,* 63–75.

Davis, T., LaRocque, K. F., Mumford, J. A., Norman, K. A., Wagner, A. D., & Poldrack, R. A. (2014). What do differences between multi-voxel and univariate analysis mean? How subject-, voxel-, and trial-level variance impact fMRI analysis. *Neuroimage, 97,* 271–283.

de Beeck, H. P. O., Haushofer, J., & Kanwisher, N. G. (2008). Interpreting fMRI data: Maps, modules and dimensions. *Nature Reviews Neuroscience, 9,* 123–135.

Fiske, S., Cuddy, A., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology, 82,* 878–902.

Fletcher, P., Frith, C., Baker, S., Shallice, T., Frackowiak, R., & Dolan, R. (1995). The mind's eye—Precuneus activation in memory-related imagery. *Neuroimage, 2,* 195–200.

Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin, 117,* 21–38.

Gobbini, I. M., Leibenluft, E., Santiago, N., & Haxby, J. V. (2004). Social and emotional attachment in the neural representation of faces. *Neuroimage, 22,* 1628–1635.

Goldberg, L. R. (1990). An alternative "description of personality": The big-five factor structure. *Journal of Personality and Social Psychology, 59,* 1216–1229.

Haslam, N. (2004). *Relational models theory: A contemporary overview*. London, UK: Psychology Press.

Haslam, N., & Fiske, A. P. (1999). Relational models theory: A confirmatory factor analysis. *Personal Relationships, 6,* 241–250.

Hassabis, D., Spreng, R. N., Rusu, A. A., Robbins, C. A., Mar, R. A., & Schacter, D. L. (2014). Imagine all the people: How the brain creates and uses personality models to predict behavior. *Cerebral Cortex, 24,* 1979–1987.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science, 293,* 2425–2430.

Heleven, E., & Van Overwalle, F. (2016). The person within: Memory codes for persons and traits using fMRI repetition suppression. *Social Cognitive and Affective Neuroscience, 11,* 159–171.

Jones, E. E., & Harris, V. A. (1967). The attribution of attitudes. *Journal of Experimental Social Psychology, 3,* 1–24.

Kerns, J. G., Cohen, J. D., MacDonald, A. W., Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science, 303,* 1023–1026.

Kestemont, J., Vandekerckhove, M., Ma, N., Van Hoeck, N., & Van Overwalle, F. (2013). Situation and person attributions under spontaneous and intentional instructions: An fMRI study. *Social Cognitive and Affective Neuroscience, 8,* 481–493.

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences, U.S.A., 103,* 3863–3868.

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis—Connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience, 2,* 4.

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., et al. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron, 60,* 1126–1141.

Marcus, D. S., Harwell, J., Olsen, T., Hodge, M., Glasser, M. F., Prior, F., et al. (2011). Informatics and data mining tools and strategies for the human connectome project. *Frontiers in Neuroinformatics, 5,* 4.

McCrae, R. R., & Costa, P. T., Jr. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology, 52,* 81–90.

Mitchell, J. P. (2008). Contributions of functional neuroimaging to the study of social cognition. *Current Directions in Psychological Science, 17,* 142–146.

Moran, J. M., Jolly, E., & Mitchell, J. P. (2014). Spontaneous mentalizing predicts the fundamental attribution error. *Journal of Cognitive Neuroscience, 26,* 569–576.

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *The Journal of Neuroscience, 162,* 8–13.

Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage, 44,* 83–98.

Szpunar, K. K., Jacques, P. L. S., Robbins, C. A., Wig, G. S., & Schacter, D. L. (2014). Repetition-related reductions in neural activity reveal component processes of mental simulation. *Social Cognitive and Affective Neuroscience, 9,* 712–722.

Tamir, D. I., Thornton, M. A., Contreras, J. M., & Mitchell, J. P. (2016). Neural evidence that three dimensions organize mental state representation: Rationality, social impact, and valence. *Proceedings of the National Academy of Sciences, U.S.A., 113,* 194–199.

Van Overwalle, F., & Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: A meta-analysis. *Neuroimage, 48,* 564–584.

Welborn, B., & Lieberman, M. (2015). Person-specific theory of mind in medial pFC. *Journal of Cognitive Neuroscience, 27,* 1–12.