Cerebral Cortex August 2014;24:1979–1987
doi:10.1093/cercor/bht042
Advance Access publication March 5, 2013

Imagine All the People: How the Brain Creates and Uses Personality Models to Predict Behavior

Demis Hassabis[1,*], R. Nathan Spreng[2,*], Andrei A. Rusu[3], Clifford A. Robbins[4], Raymond A. Mar[5] and Daniel L. Schacter[4]

[1]Gatsby Computational Neuroscience Unit, University College London, London WC1N 3AR, UK, [2]Laboratory of Brain and Cognition, Department of Human Development, Cornell University, Ithaca, NY 14853, USA, [3]Department of Computer Science, Vrije Universiteit, Amsterdam, The Netherlands, [4]Department of Psychology, Harvard University, Cambridge, MA 02138, USA and [5]Department of Psychology, York University, Toronto, ON, Canada M3J1P3

*Both the authors contributed equally to this work.

Address correspondence to: R.N. Spreng, Laboratory of Brain and Cognition, Department of Human Development, Cornell University, Ithaca, NY 14853, USA. Email: nathan.spreng@gmail.com

The behaviors of other people are often central to envisioning the future. The ability to accurately predict the thoughts and actions of others is essential for successful social interactions, with far-reaching consequences. Despite its importance, little is known about how the brain represents people in order to predict behavior. In this functional magnetic resonance imaging study, participants learned the unique personality of 4 protagonists and imagined how each would behave in different scenarios. The protagonists' personalities were composed of 2 traits: Agreeableness and Extraversion. Which protagonist was being imagined was accurately inferred based solely on activity patterns in the medial prefrontal cortex using multivariate pattern classification, providing novel evidence that brain activity can reveal whom someone is thinking about. Lateral temporal and posterior cingulate cortex discriminated between different degrees of agreeableness and extraversion, respectively. Functional connectivity analysis confirmed that regions associated with trait-processing and individual identities were functionally coupled. Activity during the imagination task, and revealed by functional connectivity, was consistent with the default network. Our results suggest that distinct regions code for personality traits, and that the brain combines these traits to represent individuals. The brain then uses this "personality model" to predict the behavior of others in novel situations.

Keywords: default mode network, fMRI, MVPA, personality traits, simulation, social neuroscience

Introduction

A key aspect to successfully navigating the social world is the ability to predict how people will behave in different situations. Individual differences in behavior can be predicted from an individual's personality traits, which encompass broad cognitive and behavioral tendencies (Costa and McCrae 1992; Roberts et al. 2007; Fleeson and Gallagher 2009). To make accurate predictions, a precise representation of an individual's personality traits must be created and this is known as a personality model (Park 1986; Park et al. 1994). Studies have shown that personality models can be quite accurate. In fact, close friends are as accurate at predicting each other's daily behavior as they are at predicting their own behavior (Vazire and Mehl 2008). Although there is a growing understanding of how an individual's personality is tied to the structure and function of his or her own brain (DeYoung 2010), little is known about how the personalities of other people are represented in order to predict behavior.

We propose that, when predicting or imagining the behavior of others based on their personality, the brain is likely to rely on the same network of regions that support other forms of mental simulation, such as remembering the past and planning for the future (Buckner and Carroll 2007; Hassabis and Maguire 2007; Schacter et al. 2007, 2008, 2012). Because autobiographical memories and future plans are both characterized by social events (Larocque and Oatley 2006), social information is a key component of event simulation (Spreng et al. 2009). Both planned and remembered events often involve the presence of other people, their thoughts, and their behaviors. Autobiographical recollection (remembering the past) involves reconstructing the spatial, temporal, and often social elements from a past event in one's life. This form of mental simulation is likely linked to personality models, as it is through previous experiences with others that one learns of their personality traits. These traits are then combined to build a model of that person's overall personality. However, predicting the future behavior of others differs from remembering in that it involves mentally simulated events that have yet to occur. This feature makes predicting behavior more akin to simulating a future event. Nevertheless, future event simulation engages many of the same core processes as remembering (Buckner and Carroll 2007; Hassabis and Maguire 2007; Schacter et al. 2007, 2008, 2012). This overlap includes the shared need to construct a scene, generating and maintaining a spatial context within which the future or past event unfolds (Hassabis and Maguire 2007; Hassabis, Kumaran, Maguire 2007; Hassabis, Kumaran, Vann, et al. 2007). We would therefore expect these same core regions to be activated when predicting someone's behavior, as scene construction is also required for mental simulating how someone might behave. These core areas include the posterior cingulate cortex (pCC), the medial temporal lobes (MTLs), the posterior inferior parietal lobule (IPL), and ventral medial prefrontal cortex (mPFC) (Hassabis, Kumaran, Maguire 2007; Hassabis, Kumaran, Vann, et al. 2007; Andrews-Hanna et al. 2010), all components of the default network.

Both the construction and application of personality models are a key component of social processing, because these models are essential for predicting and comprehending the behavior of others. Identifying trait tendencies in others relies on an ability to accurately read and interpret social cues, then linking these to broader cognitive and behavioral tendencies. The majority of research on the neuroscience of social processing has focused on the ability to infer the momentary mental

states of others, often referred to as mentalizing (Premack and Woodruff 1978; Carruthers and Smith 1996; Frith and Frith 2003). Currently, little is known about how the brain represents the broad cognitive and behavioral tendencies of other people, beyond momentary mental inferences. Based on past work, it appears that regions within the mPFC are involved in making trait inferences about others (Harris et al. 2005; Mitchell et al. 2006; Ma et al. 2011, 2012; Wagner et al. 2012). What is not known is how the brain models personality and how these personality models are employed to predict the behavior of others. Unlike the transient nature of mental states, personality traits are more enduring and generalize across time and situations. Moreover, our models of a person's personality are far more flexible and have broader utility. For example, personality models allow us to make predictions about people who are not currently present, or to predict a person's reaction to an entirely novel scenario. We can also construct personality models based on second-hand information, allowing us to make predictions about individuals we have only heard about. It remains unknown how the brain constructs and applies personality models, and what role the default network may play in these processes.

To address these questions, we employed functional magnetic resonance imaging (fMRI) to scan participants while they imagined short events, each involving 1 of 4 possible protagonists with distinct personality traits, in a variety of fictional situations. Prior to scanning, participants learned the distinct personalities of the protagonists using a "thin slice" approach (Borkenau et al. 2004). Participants were presented with profiles for 4 ostensibly real people, which included that person's name, a photograph of their face, and 12 statements about their personality; order was counterbalanced across participants. The 12 statements were created by modifying the items from 2 "Big Five" personality trait questionnaires (Costa and McCrae 1992; DeYoung et al. 2007). In this study, we focused on 2 of the 5 main personality traits: Agreeableness and Extraversion. Agreeableness reflects a tendency toward altruism, cooperation, and a valuing of harmony in interpersonal relationships as opposed to antisocial and exploitative behaviors (Costa and McCrae 1992). Extraversion, on the other hand, reflects a tendency toward the display of positive emotions and social affiliation as opposed to social withdrawal and reserve (Costa and McCrae 1992). We selected agreeableness and extraversion, because these traits are rotated variants of the major axes of the interpersonal circumplex. These traits therefore capture a broad spectrum of social behaviors and tendencies (e.g., McCrae and Costa 1989; DeYoung et al. 2012). The 4 protagonists were either high or low on these 2 dimensions, producing a 2×2 factorial design (Fig. 1*A*). After training, participants were capable of recognizing each protagonist and verbally describing that individual's personality. Participants were then asked to vividly imagine and to describe 12 novel locations for later recall in the scanner (e.g., a bar, restaurant, bank; Hassabis, Kumaran, Maguire 2007; Hassabis, Kumaran, Vann, et al. 2007).

In the subsequent scanning session, 12 text cues were presented, each describing a short event involving one of the protagonists in a preimagined location ("Protagonist" conditions). For example: "In a bar—someone spills their drink—Dave." Participants were instructed to mentally play out these "vignettes" over a 10-s period, concentrating on the actions, thoughts, and feelings of the protagonist. Participants then rated the imagined event for vividness and confidence in portraying the protagonist accurately. Two additional control conditions followed a similar format: (1) a "Self" condition where the participants imagined themselves as the main character, and (2) an "Empty Scene" condition, which involved simulating just the spatial scene, devoid of people and events. A final baseline condition involved counting the number of syllables in the text cue (the "Count" condition) yielding 84 trials in all (7 conditions $\times$ 12 trials).

To assess the engagement of the default network across the 3 imagination tasks, we compared these conditions to the baseline syllable Count condition. To examine how the brain generates and applies personality models, we compared the Self and Protagonist conditions with the Empty Scene control condition. This comparison allowed us to isolate the interpersonal components of imagining an event from scene-construction processes (Hassabis and Maguire 2007). Moreover, to determine whether the brain represents unique personality profiles with different brain areas, we employed multivariate pattern analysis (MVPA) (Haynes and Rees 2006; Norman et al. 2006). Doing so allowed us to determine whether spatially localized brain activity was reliable enough across individuals to infer who is being thought about, based solely on the pattern of brain activation. An additional functional connectivity analysis provided further clues as to how personality information is integrated in order to produce a model for behavioral predictions. Taken together, our results reveal the role of the default network in the representation and integration of social and personality information when predicting the future behavior of other people.

Materials and Methods

Participants

Participants were 19 healthy right-handed young adults (10 females; mean age = 21.4 ± 3.2 years) who gave written informed consent in accordance with the Harvard Institutional Review Board.

Prescan Training

Protagonist Training

Participants learned the 4 protagonists' personalities prior to scanning, through a combination of training and testing for both comprehension and recall (Landauer and Bjork 1978; Roediger and Karpicke 2006). Participants were introduced to each of the protagonists on paper. They were informed that the experimenters knew quite a lot about the 4 individuals, derived from interviews with them, their friends, and family. Participants were told that they would be given a little bit of information about each individual, and that the experimenters were interested in seeing how well the participant could predict how the individuals would behave based on that information. Protagonists were gender matched to the participant. Common popular names for the protagonists were selected from the US social security website (http://www.ssa.gov/cgi-bin/namesbystate.cgi) list of the top 10 most popular names given to children born in Massachusetts in 1990. Men's names were Mike, Chris, Dave, and Nick. Women's names were Ashley, Sarah, Nicole, and Jenny. No participant shared the same name as the protagonist. Each name was paired with 12 personality statements, 6 related to agreeableness and 6 related to extraversion, from items of the Big Five Aspect Scales (DeYoung 2010) and the Revised Neuroticism-Extraversion-Openness Personality Inventory (Costa and McCrae 1992). Personality statements for agreeableness included statements worded positively (e.g., "Likes to cooperate with others") and negatively (e.g. "Can be cold and aloof"). Likewise, statements for extraversion were both positive and negative (e.g., "Is outgoing, sociable"; "Is sometimes shy, inhibited"). While

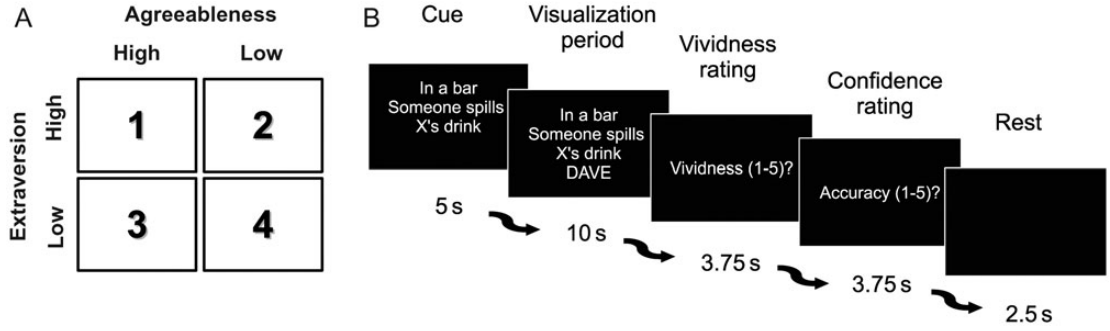

Figure 1. Study design. (*A*) The personalities of the 4 protagonists were varied high low on 2 trait dimensions, extraversion and agreeableness, to create a 2 × 2 factorial relationship between the 4 personalities. (*B*) Timeline showing the experimental design of a single trial. The text instruction cue is presented for 5 s containing the location and vignette information. The name of the protagonist is then displayed ("Dave" in this example trial), and the participant imagines that event happening to that person in that location for 10 s. Finally, participants give feedback on their imagined events via vividness and accuracy ratings followed by a rest period.

looking at the 12 statements, photograph, and name, participants rated 18 items about each protagonist, 1 protagonist at a time, as either true or false (72 items in total; e.g. "Likes the idea of being given a surprise party, TRUE or FALSE"). Participants were then given the opportunity to study the 12 statements prior to a free recall test where the 12 statements would be inaccessible. One at a time, participants were then instructed to "Describe your general thoughts and impressions of this individual in 3–4 sentences." Order of presentation, as well as personality, name, and photograph pairings, was randomized across participants.

Imagining Scenes Training

Participants were then asked to vividly imagine 12 common everyday locations such as a bar, restaurant, or bank, and to describe them in as much detail as possible (Hassabis, Kumaran, Maguire 2007; Hassabis, Kumaran, Vann, et al. 2007). They were explicitly told not to give an actual memory of a real place (or any part of one), but to instead construct something entirely novel. They were informed that instructions to reimagine these locations would appear multiple times during the scanning session, and that each time in response they were required to always think of the exact same location as closely as possible. These types of imagination tasks exemplify mental simulation (Hassabis, Kumaran, Maguire 2007; Hassabis, Kumaran, Vann, et al. 2007). Unlike autobiographical recollection, imagined events can be experimentally manipulated and systematized. All participants imagine the same scenarios and can do so multiple times, and the content can be tightly controlled (Hassabis, Kumaran, Maguire 2007; Hassabis, Kumaran, Vann, et al. 2007; Addis et al. 2009).

Immediately prior to scanning, participants were presented with a photograph for each protagonist and required to state the person's name and 3 words describing that person's personality. Following the scan, participants were shown the photos and again asked to name the person and provide 3 words to describe his or her personality; this process ensured the stability of the personality representation throughout the scanning interval. Participants also recounted the content of their imagined event for 2 randomly selected scenes for each of the 4 protagonists, as well as the self; this procedure allowed us to confirm task compliance.

Task and Procedure

During the scanning session, participants were presented with text cues describing a short event unfolding in each of the preimagined locations, with 1 of the 4 people named as the main protagonist. For example: "in the street—sees a homeless vet asking for change—Sarah." Participants were instructed to mentally play out 12 of these vignettes, each lasting the duration of a 10-s period, concentrating on the actions, thoughts, and feelings of the protagonist. Participants then rated the imagined event for vividness and confidence in accurately portraying the protagonist, both on a scale of 1–5. Each trial lasted 25 s, consisting of: A text cue containing information regarding the location, event, and protagonist (or "Self" or "Empty" cue) (5 s); a mental simulation period (10 s); vividness and confidence ratings (7.5 s); and a rest period (2.5 s). During the Count condition, the following information was displayed: A text cue (5 s); a counting period allowing the participant to determine the number of syllables in the verbal information (10 s); a question regarding whether the number of syllables was even or odd and a confidence rating for their judgment (7.5 s); and followed by a rest period (2.5 s). The trial order was counterbalanced across participants, with no 2 adjacent trials involving either the same location or the same condition. Trials with vividness and confidence scores over 3 (on the 5-point scale) were retained for subsequent fMRI analysis. The number of trials per condition were M(self) = 10.1 ± 2.7; M(high extra–high agree) = 9.6 ± 2.8; M(high extra–low agree) = 9.5 ± 2.8; M(low extra–high agree) = 9.4 ± 3.0; M(low extra–low agree) = 8.4 ± 3.1; and M(empty room) = 10.0 ± 2.8. Two additional conditions involving trait judgments were conducted, but are not germane to the current aims and are thus not reported.

Image Acquisition

Brain imaging data were acquired at the Harvard Center for Brain Science with a 3-T Siemens TimTrio MRI scanner with a 12-channel head coil. Anatomical scans were acquired using a T_1-weighted multiecho volumetric MRI (time repetition [TR] = 2530 ms; time echos [TE's] = 1.64, 3.5, 5.36, 7.22 ms; 7° flip angle; 1-mm voxel). Two 30 min 20 s blood oxygen level-dependent (BOLD) functional scans were acquired with a T_2*-weighted echo planar imaging (EPI) pulse sequence (TR = 2500; TE = 30 ms; 85° flip angle; 39 axial slices; 3 × 3 × 3 mm voxels).

fMRI Analysis

Partial Least Squares Preprocessing and Analysis

For the 5 pairwise contrasts and functional connectivity analysis, fMRI data were subjected to standard preprocessing steps and analyzed with partial least squares (PLS; McIntosh 1999; Krishnan et al. 2011). fMRI data were preprocessed using SPM2 (Wellcome Trust Center for Neuroimaging, London, UK). The first 4 volumes in each run were excluded from analyses to allow for T_1-equilibration effects. Data were corrected for slice-dependent time shifts and for head motion within and across runs using a rigid-body correction. Images were then spatially normalized to the standard space of the Montreal Neurological Institute (MNI) atlas. The volumetric time series was then spatially smoothed with a 6-mm full-width at half-maximum (FWHM) Gaussian kernel resulting in 4-mm cubic voxels. All coordinates are reported in MNI space.

PLS was utilized to examine BOLD signal changes at the group level. The method is sensitive to covariance in voxel response associated with tasks, making it well suited to study distributed patterns of activity. Additionally, PLS is impervious to scanner drift. This was essential as the longer runs needed for optimal MVPA analyses that exacerbated the potential for scanner drift. For the current analysis, we used the nonrotated version of task spatiotemporal PLS (ST-PLS),

enabling us to specify a priori nonorthogonal contrasts (as opposed to the data-driven version of rotated ST—PLS that identifies orthogonal latent variables). Five pairwise sets of contrasts were specified. Prior to analysis, activity at each time point, relative to trial onset, for each voxel is averaged across trials of a given condition and normalized to activity in the first TR of the trial. The data matrix is then expressed as voxel-by-voxel deviation from the grand mean across the entire experiment. This matrix is analyzed with singular value decomposition to derive the contrast effects in the data. Here, we applied PLS analysis to event-related fMRI data, and the results provide a set of brain regions wherein activity is reliably related to the task conditions at 8 poststimulus time points. Each voxel is given a singular value weight, known as a salience, which is proportional to the covariance of activity with the task contrast at each time point for the contrast. The significance of each contrast as a whole was determined by permutation testing, using 500 permutations. This analysis was accomplished by randomly reassigning the order of the conditions for each participant. PLS is recalculated for each permutation sample, and the frequency with which the permuted singular value exceeds the observed singular values is determined and expressed as a probability. All task contrasts were significant at $P = 0.002$. In a second, independent step, the reliability of the saliences for the brain voxels across subjects was determined by bootstrap resampling, using 100 iterations, to estimate the standard errors for each voxel. Clusters >100 mm^3 comprising voxels with a ratio of the salience to the bootstrap standard error values (i.e., the "bootstrap ratio"; BSR) >3 ($P < 0.005$) were reported. The local maximum for each cluster was defined as the voxel with a BSR higher than any other voxel in a 2-cm cube centered on that voxel. PLS identifies whole-brain patterns of activity in a single analytic step and, thus, no correction for multiple comparisons is required. Although most of the brain regions showed reliable activations across multiple time points, we report the BSR for the third TR (i.e., 7.5 s after the name appeared on the screen) as a representative index of brain activity in time.

The functional connectivity analysis was performed using "seed" PLS, a multivariate task-related functional connectivity analysis technique used to investigate the relationship between the activity of a seed region and the activity in the rest of the brain (McIntosh 1999; Krishnan et al. 2011). We used the BOLD signal in the mPFC (0, 50, 22) derived from the MVPA analysis, as seeds. We then assessed the task-related functional connectivity with the rest of the brain during the simulation of each of the 4 protagonists within the same analysis. Individually defined seed activity from the peak activation and 26 neighboring voxels in the third TR was correlated with activity in all brain voxels, across participants. This matrix was then analyzed as above. Mean BSR values from the peak personality classifier analyses in lateral temporal cortex (LTC) (−60, −34, −17), dorsal mPFC (−3, 41, 49), and pCC (−6, −49, 31), including the 26 neighboring voxels, were used to assess the significance of connectivity with the seeds.

MVPA Preprocessing and Analysis

The first 4 volumes were excluded from analysis to allow for T_1-equilibration effects. Using SPM8, the remaining volumes for individual subjects were realigned to the mean, resliced, and smoothed with an 8-mm FWHM kernel, and the first 5 s of each trial were selected for further analysis. Each trial was modeled with a separate regressor as a boxcar function and convolved with the canonical hemodynamic response function. A high-pass filter with a cutoff of 128 s was employed. Model estimation resulted in a "betamap" for each trial (Soon et al. 2008), which were then z-scored (Pereira et al. 2009) and inputted as the training data to the classifier.

Subsequent analyses were performed using multivariate pattern analysis in python (PyMVPA) (Hanke et al. 2009). We used a whole-brain "searchlight" MVPA approach (Kriegeskorte et al. 2006; Hassabis et al. 2009), which involved stepping voxel by voxel through each subject's brain, looking at localized patterns of activity in small spherical arrays of voxels (radius = 3 voxels) to yield an "accuracy" map. Standard linear Support Vector Machines ("libsvm" backend; Chang and Lin 2011) were trained (with the cost parameter C set to 1) using a leave-one-out cross-validation procedure to produce accuracy estimates across the brain (Pereira et al. 2009). The resulting accuracy maps for each participant were then normalized to the MNI space and smoothed with a 3-dimensional Gaussian kernel (cut at edges), to compensate for errors in normalization. Since optimal levels of smoothing can only be determined empirically (Hopfinger et al. 2000), several levels of smoothing may be used to analyze the data (Worsley et al. 1996). In this case, we used 11-mm FWHM for the protagonist and agreeableness conditions, and a 16-mm FWHM for extraversion. At a second-level analysis, voxel-wise t-values were computed across these individual subject accuracy maps, yielding a final group level "information heatmap", which illustrates the brain regions that carry sufficient information to discriminate between conditions.

We established the statistical significance of our findings using the standard approach of permutation testing (Nichols and Holmes 2002). For each classification problem, the whole procedure described above was repeated 20 times but with random permutations of class labels for the trials, and the maximum t-value in the whole brain computed. Then in accordance with standard practice (Nichols and Holmes 2002), the 0.95-quantile of this maximum t-statistic sample obtained from the permutation testing was selected as the significance threshold. The corresponding t-value map (computed with correct trial labels) was cut off below that value (t-value thresholds: 4.28, 2.82, 3.32 for agreeableness, extraversion, and protagonists classifications, respectively) yielding results significant at $P < 0.05$ corrected for multiple comparisons.

Table 1
Behavioral results

	Vividness		Confidence	
	Mean	SD	Mean	SD
High extraversion–high agreeableness	4.22	0.45	4.52	0.33
High extraversion–low agreeableness	4.07	0.39	4.28	0.46
Low extraversion–high agreeableness	4.17	0.53	4.36	0.46
Low extraversion–low agreeableness	4.00	0.50	4.19	0.60
All protagonists	4.11	0.37	4.34	0.39
Self	4.34	0.33	4.61	0.29
Empty scene	4.34	0.46	—	—

Results

Behavioral Data

Overall, participants' vividness and confidence ratings were high (average ratings >4 of 5; for details, see Table 1), indicating task compliance. The assessment of vividness ratings when imagining the 4 protagonists, the self, and the empty room revealed a significant main effect ($F_{5,14} = 3.87$, $P < 0.05$). However, no pairwise differences were observed when correcting for multiple comparisons (Bonferroni correction, $\alpha < 0.05$). The trend was toward a significant difference in vividness between the self condition and the 2 low agreeableness protagonists. There was a statistically significant difference in confidence ratings for how the participants and 4 protagonists would act ($F_{4,15} = 5.16$, $P < 0.01$). Participants' confidence in imagining their own behavior was higher than for the 2 low agreeableness protagonists (Bonferroni correction, $\alpha < 0.05$). Importantly, however, no significant differences were observed between the 4 protagonists. In determining whether an even or odd number of syllables were presented in the verbal material, participants performed at better-than-chance levels (mean percent accuracy = 60 ± 16, one sample $t_{(18)} = 2.55$, $P < 0.05$).

Neuroimaging Data

To confirm the engagement of the default network across the 3 imagination tasks, we compared these conditions to the

syllable counting task, in which participants only engaged the surface features of verbally presented information. In 3 pairwise contrasts, imagining the protagonists, the self, and an empty scene were compared with this baseline condition. The imagination tasks were all associated with an increased BOLD signal in default network regions, including the mPFC, pCC, MTL, LTC, temporal pole, IPL, and the superior and inferior frontal gyri (Fig. 2*A*–*C*, Supplementary Table 1). This pattern of activity is consistent with the activity observed when constructing a scene (Hassabis, Kumaran, Maguire 2007; Hassabis, Kumaran, Vann, et al. 2007) and imagining possible future events (Schacter et al. 2007, 2012).

Next, we parsed out the individual contributions of spatial and social processing to default network activity by conducting 2 contrasts: (A) imagining the protagonists > imagining an empty scene without people, and (B) imagining oneself > imagining an empty scene. Relative to imagining an empty scene, the contrasts for imagining the protagonists and the self revealed regions in the ventral, dorsal, and anterior mPFC, pCC, the temporal poles, and occipital cortex (Fig. 2*D*,*E*, Supplementary Table 1). These brain regions are more engaged by social processing during the simulation of a social interaction relative to those required to construct the scene. The contrast comparing imagining the self with imagining the protagonists revealed greater activity in hippocampus, mPFC, and other regions (see Supplementary Table 1 and Supplementary Fig. 1 for full results). No brain regions demonstrated a greater BOLD response for imagining the protagonists relative to the self.

Two key questions were then addressed: (A) Where in the brain is the personality information for protagonists represented? and (B) where is identity information for each of the 4 protagonists represented? To answer these questions, we utilized MVPA to determine whether brain regions carried information sufficient to discriminate between conditions. To examine where personality information was represented, a 2-way classification was performed with trials involving protagonists that shared a trait (e.g., high agreeableness) collapsed into one condition. We found clusters in the dorsal mPFC and left LTC (Fig. 3*A*) that distinguished between protagonists of high and low agreeableness. Protagonists with high and low extraversion were discriminated by differences in pCC response (Fig. 3*B*). To locate where identity information for the protagonists was represented, a 4-way classification was performed to discriminate between the 4 protagonists. Clusters in anterior and dorsal mPFC (superior to that observed for agreeableness) reliably discriminated between the 4 protagonists (Fig. 3*C*). Different personality models are therefore associated with unique and detectable patterns of brain activity in the mPFC. In other words, based on brain activation patterns alone, we were able to infer which of the 4 protagonists the participants were imagining.

Finally, we sought to evaluate the hypothesis that anterior mPFC assembles and updates the models of other people's personality traits processed by the LTC, dorsal mPFC, and pCC. To evaluate this hypothesis further, we conducted a functional connectivity analysis to determine whether the anterior mPFC was functionally connected with the LTC, the dorsal mPFC, and the pCC during mental simulation of the 4 protagonists. In this analysis, we assessed the task-dependent functional connectivity of the anterior mPFC in the 4 protagonist conditions. Across all 4 protagonists, activity in the anterior mPFC was significantly correlated with a distributed pattern of voxel response (Fig. 4). Functional connectivity with the anterior mPFC was significant for the LTC (mean BSR = 3.5, $P < 0.001$), dorsal mPFC (mean BSR = 4.1, $P < 0.001$), and pCC (mean BSR = 3.2, $P < 0.002$). During imagined social simulations, the brain regions that code for personality information are therefore functionally coupled with the region that codes for individual identities. Further, mPFC connectivity also extended to the entire default network, including the IPL, retrosplenial cortex, and hippocampus (Fig. 4 and Supplementary Table 1).

Gender Control Analysis

Given gender differences in social behavior, we reanalyzed our data to assess the effect of gender on our results. In the group analysis using PLS, brain activity for both gender groups covaried together, and there were no significant task by gender interactions. For the MVPA, we used gender as a

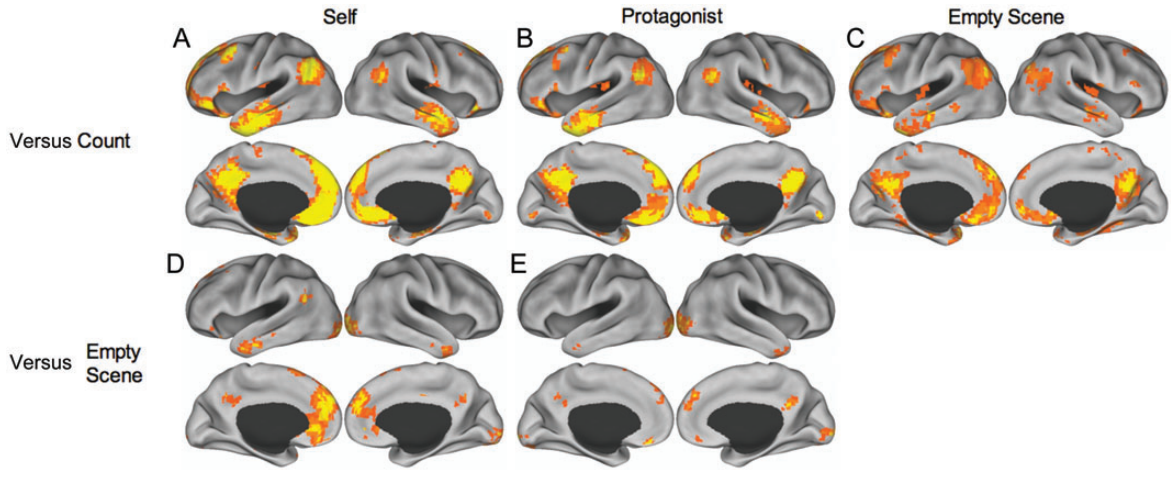

Figure 2. Pairwise contrasts of the experimental conditions. (*A*) Self > Count; (*B*) Protagonists > Count; and (*C*) Empty Scene > Count all demonstrate default network engagement across simulation conditions. (*D*) Self > Empty Scene and (*E*) Protagonists > Empty Scene show specific activity for social simulations. Images are displayed from TR 3 (7.5 s) and thresholded at $P < 0.005$ (PLS identifies whole-brain patterns of activity in a single analytic step and, thus, no correction for multiple comparisons is required). See Supplementary Table 1 for peak coordinates.

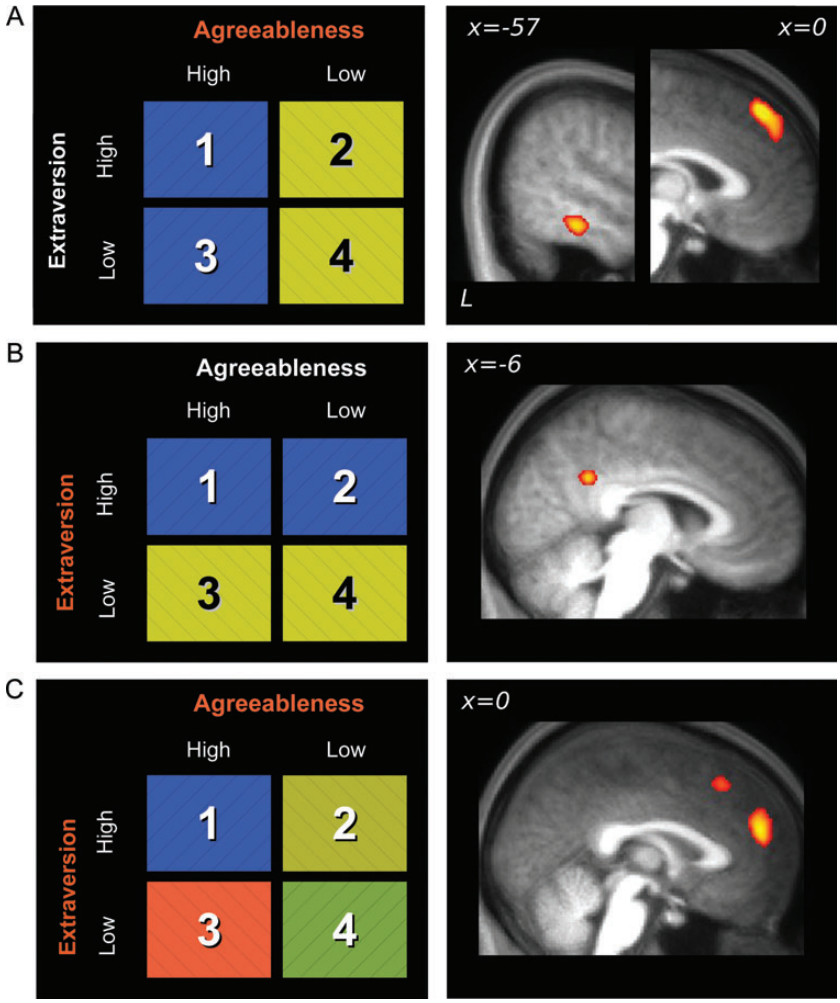

Figure 3. MVPA results showing significant voxels at the group level after the application of whole-brain searchlight classification techniques, overlaid on an averaged structural image. (*A*) Left LTC and dorsal mPFC discriminate between protagonists with high or low agreeableness. (*B*) pCC discriminates between protagonists with high or low extraversion. (*C*) mPFC discriminates between the 4 protagonists. All voxels significant at $P < 0.05$ corrected for multiple comparisons. See Supplementary Table 1 for peak coordinates.

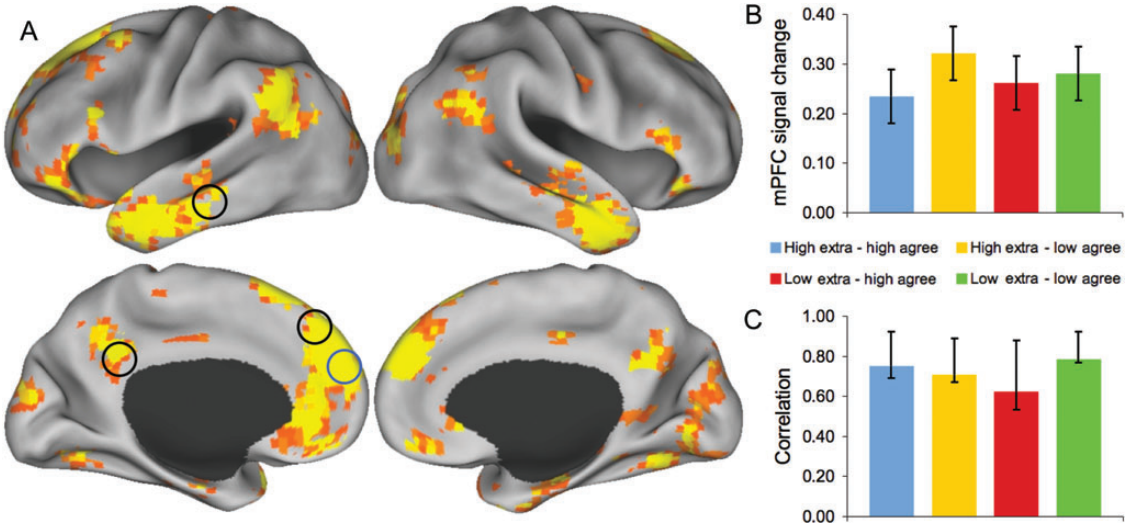

Figure 4. Functional connectivity analysis. (*A*) Whole-brain pattern of functional connectivity with the mPFC seed region identified by the MVPA analysis as discriminating between the 4 protagonists. Seed region is circled in blue. Black circles designate that regions identified by the personality classification are functionally connected with the mPFC region. Results displayed from TR 3 (7.5 s) and thresholded at $P < 0.005$ (PLS identifies whole-brain patterns of activity in a single analytic step and, thus, no correction for multiple comparisons is required). (*B*) BOLD signal change, relative to the trial onset, for the 4 protagonist conditions (error bars are the within-subject SEM). (*C*) Correlation between mPFC seed activity and composite measures of the distributed voxel response across the brain (confidence intervals were derived from bootstrap resampling). See Supplementary Table 1 for peak coordinates.

covariate of no interest in the second-level analysis and found the same pattern of results. Next, we assessed whether there where any reliable whole-brain differences in the regions that discriminate between levels of agreeableness and extraversion, as well as the 4 protagonists. No significant patterns were observed. Although we recognize that there are many reported gender differences in social cognition, we are unlikely to have the statistical power to observe them in our current sample.

Discussion

Personality models describe broad cognitive and behavioral tendencies, which are immensely useful for predicting how people will behave and react. Thus, these models are essential for successful navigation of the social world. In this study, we examined how the brain constructs and applies personality models. Default network activity was observed across 3 imagination tasks (consistent with this network's involvement in internally focused thoughts; Andrews-Hanna 2012), but subregions of this network were more involved in social and interpersonal information processing. Our results also identified specific brain regions where personality models are coded, as well as where individual trait information is represented in the brain. Connectivity analyses examined how protagonist identity and trait information interact, or how the brain associates specific personality traits with a given protagonist. These results suggest that personality information is integrated in the mPFC, producing a model for behavioral predictions.

A key question of interest was how personality models for different people are represented in the brain. Different patterns of activation in the anterior mPFC could reliably distinguish between the different people whose behavior was being imagined. We hypothesize that this region is responsible for assembling and updating personality models. We then examined how the brain represents 2 major traits: Agreeableness and extraversion. Agreeableness was associated with unique patterns of BOLD response in the dorsal mPFC and LTC, whereas extraversion was associated with the pCC. These results provide a novel neuroscientific demonstration that discriminable brain regions code for the specific personality traits of other people. Given the novelty of our observations, we encourage replication to increase confidence in these results. Notably, these brain regions are also reliably engaged when people are inferring what a person is momentarily thinking and feeling (Mar 2011) as well as when spontaneous and intentional trait inferences are being made (Ma et al. 2011). During these moments, a subprocess may be linking inferred mental states to broader personality traits. This idea is consistent with past behavioral work, showing that traits are rapidly and spontaneously inferred from behaviors (Uleman et al. 2008). As we get to know a person better, individual personality traits are combined to form a more holistic representation of the person's character—a personality model—which can be used to imagine and therefore predict the behavior and thoughts of individuals in hypothetical situations. As we hypothesized, the brain regions that code for personality traits were functionally coupled with the mPFC, which codes for individual identities during mental simulation.

Mentally simulated events are often composed of both social and spatial elements. To determine whether different components of the default network are associated with the social and spatial aspects of mental simulation, we controlled for scene construction to isolate the brain activity specific to interpersonal imagining. After doing so, imagining social elements was uniquely associated with the mPFC, pCC, and the temporal poles. These regions are reliably observed during studies of mental inference (Mar 2011) and are consistent with the observed overlap between autobiographical memory and mentalizing (Spreng et al. 2009; Rabin et al. 2010; Spreng and Grady 2010). They might also facilitate the integration of personal and interpersonal information for the strategic use of social conceptual knowledge (Spreng and Mar 2012). Integrating social knowledge could also support the generation of behavioral predictions based on personality models.

Event simulations consist of a rich spatial context within which complex social interactions can take place. These simulations reliably engage the default network, with recent work exposing 2 core regions of the default network: mPFC and pCC (Andrews-Hanna et al. 2010). Both of these regions are densely interconnected with 2 distinct subsystems: (1) the MTL subsystem, including the hippocampus, parahippocampus, restrosplenial cortex, the posterior IPL, and ventral mPFC, and (2) the dorsal mPFC subsystem, including dorsal mPFC, LTC, the temporal parietal junction, and the temporal pole (Andrews-Hanna et al. 2010). Generating the spatial context for an event simulation appears to be supported by the MTL subsystem of the default network (Andrews-Hanna et al. 2010; i.e., the "scene construction" network, Hassabis and Maguire 2007). Generating the social components of an event simulation, including the application of personality models, is supported by the core default structures and the dorsal mPFC subsystem (Wagner et al. 2012). Further, individual regions of the dorsal mPFC subsystem differentially code for personality traits and the unique identity of other people. During the process of simulation, these discrete regions are functionally coupled with each other and the MTL subsystem, with these associations suggesting an integration of social and personality features into a spatial context.

It has been suggested that mental simulations bestow an adaptive advantage on humans by allowing them to prepare for upcoming situations (Ingvar 1979; Taylor et al. 1998; Hassabis and Maguire 2007; Suddendorf and Corballis 2007; Schacter 2012). Planning for social situations by imagining the likely behavior of others may be especially critical for the success of a highly social species, such as humans. Interestingly, the anterior mPFC has been implicated in social cognition disorders, such as autism (von dem Hagen et al. forthcoming), and our results point to a possible inability to build accurate personality models of others for those with such disorders. Future work in this direction should examine how models of others are constructed, how personality models are updated based on new information (Rapp and Gerrig 2001), and how dysfunction might be treated.

Supplementary Material

Supplementary material can be found at: http://www.cercor.oxfordjournals.org/.

Funding

This work was supported by a Wellcome Trust grant to D.H., NIH grant MH060941 to D.L.S., and SSHRC grant to RAM.

Notes

We thank Jason Mitchell for sharing mental inference statements and Karen Spreng for comments on an earlier draft of this manuscript and Zoltan Szlavik for helpful analysis advice. *Conflict of Interest*: None declared.

References

Addis DR, Pan L, Vu MA, Laiser N, Schacter DL. 2009. Constructive episodic simulation of the future and the past: distinct subsystems of a core brain network mediate imagining and remembering. Neuropsychologia. 47:2222–2238.

Andrews-Hanna JR. 2012. The adaptive role of the default network in internal mentation. Neuroscientist. 18:251–270.

Andrews-Hanna JR, Reidler JS, Sepulcre J, Poulin R, Buckner RL. 2010. Functional-anatomic fractionation of the brain's default network. Neuron. 65:550–562.

Borkenau P, Mauer N, Riemann R, Spinath FM, Angleitner A. 2004. Thin slices of behavior as cues of personality and intelligence. J Pers Soc Psychol. 86:599–614.

Buckner RL, Carroll DC. 2007. Self-projection and the brain. Trends Cogn Sci. 11:49–57.

Carruthers P, Smith P, editors. 1996. Theories of theories of mind. Cambridge (MA): Cambridge University Press.

Chang CC, Lin CJ. 2011. LIBSVM: a library for support vector machines. ACM Trans Int Syst Technol. 2:1–27.

Costa PTJ, McCrae RR. 1992. NEO PI-R professional manual. Odessa (FL): Psychological Assessment Resources, Inc.

DeYoung CG. 2010. Personality neuroscience and the biology of traits. Social Personality Psychol Compass. 4:1165–1180.

DeYoung CG, Quilty LC, Peterson JB. 2007. Between facets and domains: 10 aspects of the Big Five. J Pers Soc Psychol. 93:880–896.

DeYoung CG, Weisberg YJ, Quilty LC, Peterson JB. forthcoming 2012. Unifying the aspects of the Big Five, the interpersonal circumplex, and trait affiliation. J Pers. doi:10.1111/jopy.12020.

Fleeson W, Gallagher P. 2009. The implications of Big Five standing for the distribution of trait manifestation in behavior: fifteen experience-sampling studies and a meta-analysis. J Pers Soc Psychol. 97:1097–1114.

Frith U, Frith CD. 2003. Development and neurophysiology of mentalizing. Philos Trans R Soc Lond B Biol Sci. 358:459–473.

Hanke M, Halchenko YO, Sederberg PB, Olivetti E, Fründ I, Reiger JW, Herrmann CS, Haxby JV, Hanson SJ, Pollman S. 2009. PyMVPA: a unifying approach to the analysis of neuroscientific data. Front Neuroinfor. 3:3.

Harris LT, Todorov A, Fiske ST. 2005. Attributions on the brain: neuroimaging dispositional inferences beyond theory of mind. Neuroimage. 28:763–769.

Hassabis D, Chu C, Rees G, Weiskopf N, Molyneux PD, Maguire EA. 2009. Decoding neuronal ensembles in the human hippocampus. Curr Biol. 19:546–554.

Hassabis D, Kumaran D, Maguire EA. 2007. Using imagination to understand the neural basis of episodic memory. J Neurosci. 27:14365–14374.

Hassabis D, Kumaran D, Vann SD, Maguire EA. 2007. Patients with hippocampal amnesia cannot imagine new experiences. Proc Natl Acad Sci USA. 104:1726–1731.

Hassabis D, Maguire EA. 2007. Deconstructing episodic memory with construction. Trends Cogn Sci. 11:299–306.

Haynes JD, Rees G. 2006. Decoding mental states from brain activity in humans. Nat Rev Neurosci. 7:523–534.

Hopfinger JB, Buchel C, Holmes AP, Friston KJ. 2000. A study of analysis parameters that influence the sensitivity of event-related fMRI analyses. Neuroimage. 11:326–333.

Ingvar DH. 1979. Hyperfrontal distribution of the cerebral grey matter flow in resting wakefulness: on the functional anatomy of the conscious state. Neurol Scand. 60:12–25.

Kriegeskorte N, Goebel R, Bandettini P. 2006. Information-based functional brain mapping. Proc Natl Acad Sci USA. 103:3863–3868.

Krishnan A, Williams LJ, McIntosh AR, Abdi H. 2011. Partial Least Squares (PLS) methods for neuroimaging: a tutorial and review. Neuroimage. 56:455–475.

Landauer TK, Bjork RA. 1978. Optimum rehearsal patterns and name learning. In: Gruneberg MM, Morris PE, Sykes RN, editors. Practical aspects of memory. New York: Academic Press. pp. 625–632.

Larocque L, Oatley K. 2006. Joint plans, emotions, and relationships: a diary study of errors. J Cultur Evol Psychol. 4:245–265.

Ma N, Vandekerckhove M, Van Hoeck N, Van Overwalle F. 2012. Distinct recruitment of temporo-parietal junction and medial prefrontal cortex in behavior understanding and trait identification. Soc Neurosci. 7:591–605.

Ma N, Vandekerckhove M, Van Overwalle F, Seurinck R, Fias W. 2011. Spontaneous and intentional trait inferences recruit a common mentalizing network to a different degree: spontaneous inferences activate only its core areas. Soc Neurosci. 6:123–138.

Mar RA. 2011. The neural bases of social cognition and story comprehension. Annu Rev Psychol. 62:103–134.

McCrae RR, Costa PT. 1989. The structure of interpersonal traits: Wiggins's circumplex and the five-factor model. J Pers Soc Psychol. 56:586–595.

McIntosh AR. 1999. Mapping cognition to the brain through neural interactions. Memory. 7:523–548.

Mitchell JP, Cloutier J, Banaji MR, Macrae CN. 2006. Medial prefrontal dissociations during processing of trait diagnostic and nondiagnostic person information. Soc Cogn Affect Neurosci. 1:49–55.

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

Norman KA, Polyn SM, Detre GJ, Haxby JV. 2006. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn Sci. 10:424–430.

Park B. 1986. A method for studying the development of impressions of real people. J Pers Soc Psychol. 51:907–917.

Park B, DeKay ML, Kraus S. 1994. Aggregating social behavior into person models: perceiver-induced consistency. J Pers Soc Psychol. 66:437–459.

Pereira F, Mitchell T, Botvinick M. 2009. Machine learning classifiers and fMRI: a tutorial overview. Neuroimage. 45:S199–209.

Premack D, Woodruff G. 1978. Does the chimpanzee have a theory of mind? Behav Brain Sci. 1:515–526.

Rabin JS, Gilboa A, Stuss DT, Mar RA, Rosenbaum RS. 2010. Common and unique neural correlates of autobiographical memory and theory of mind. J Cogn Neurosci. 22:1095–1111.

Rapp DN, Gerrig RJ. 2001. Readers' trait-based models of characters in narrative comprehension. J Mem Lang. 45:737–750.

Roberts BW, Kuneel NR, Shiner R, Caspi A, Goldberg LR. 2007. The power of personality: the comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. Perspect Psychol Sci. 2:313–345.

Roediger HL III, Karpicke JD. 2006. The power of testing memory: basic research and implications for educational practice. Perspect Psychol Sci. 181–210.

Schacter DL. 2012. Adaptive constructive processes and the future of memory. Am Psychol. 67:603–613.

Schacter DL, Addis DR, Buckner RL. 2008. Episodic simulation of future events: concepts, data, and applications. Ann N Y Acad Sci. 1124:39–60.

Schacter DL, Addis DR, Buckner RL. 2007. Remembering the past to imagine the future: the prospective brain. Nat Rev Neurosci. 8:657–661.

Schacter DL, Addis DR, Hassabis D, Martin VC, Spreng RN, Szpunar KK. 2012. The future of memory: remembering, imagining, and the brain. Neuron. 76:677–694.

Soon CS, Brass M, Heinze HJ, Haynes JD. 2008. Unconscious determinants of free decisions in the human brain. Nat Neurosci. 11:543–545.

Spreng RN, Grady CL. 2010. Patterns of brain activity supporting autobiographical memory, prospection, and theory-of-mind and their

relationship to the default mode network. J Cogn Neurosci. 22:1112–1123.
Spreng RN, Mar RA. 2012. I remember you: a role for memory in social cognition and the functional neuroanatomy of their interaction. Brain Res. 1428:43–50.
Spreng RN, Mar RA, Kim ASN. 2009. The common neural basis of autobiographical memory, prospection, navigation, theory of mind and the default mode: a quantitative meta-analysis. J Cogn Neurosci. 32:489–510.
Suddendorf T, Corballis MC. 2007. The evolution of foresight: what is mental time travel and is it unique to humans? Behav Brain Sci. 30:299–313.
Taylor SE, Pham LB, Rivkin ID, Armor DA. 1998. Harnessing the imagination: mental simulation, self-regulation, and coping. Am Psychol. 53:429–439.
Uleman JS, Adil Saribay S, Gonzalez CM. 2008. Spontaneous inferences, implicit impressions, and implicit theories. Annu Rev Psychol. 59:329–360.
Vazire S, Mehl MR. 2008. Knowing me, knowing you: the accuracy and unique predictive validity of self-ratings and other-ratings of daily behavior. J Pers Soc Psychol. 95:1202–1216.
von dem Hagen E, Stoyanova RS, Baron-Cohen S, Calder AJ. forthcoming. Reduced functional connectivity within and between "social" resting state networks in Autism Spectrum Conditions. Soc Cogn Affect Neurosci.
Wagner DD, Haxby JR, Heatherton TS. 2012. The representation of self and person knowledge in the medial prefrontal cortex. WIRES Cogn Sci. 3:451–470.
Worsley KJ, Marrett S, Neelin P, Evans AC. 1996. Searching scale space for activation in PET images. Hum Brain Mapp. 4:74–90.