Cerebral Cortex, January 2017;27: 344–357

doi: 10.1093/cercor/bhw401
Advance Access Publication Date: 20 January 2017
Original Article

ORIGINAL ARTICLE

Neural Representations of Belief Concepts: A Representational Similarity Approach to Social Semantics

Anna Leshinskaya[1], Juan Manuel Contreras[2], Alfonso Caramazza[3,4] and Jason P. Mitchell[4,5]

[1]Center for Cognitive Neuroscience, University of Pennsylvania, [2]Capital One, McLean, VA, [3]Center for Mind/Brain Sciences, University of Trento, [4]Department of Psychology, Harvard University and [5]Center for Brain Science, Harvard University

Address correspondence to Anna Leshinskaya, Center for Cognitive Neuroscience, 3720 Walnut Street, Room B51, Philadelphia, PA 19104, USA. Email: alesh@sas.upenn.edu

Abstract

The present experiment identified neural regions that represent a class of concepts that are independent of perceptual or sensory attributes. During functional magnetic resonance imaging scanning, participants viewed names of social groups (e.g. Atheists, Evangelicals, and Economists) and performed a one-back similarity judgment according to 1 of 2 dimensions of belief attributes: political orientation (Liberal to Conservative) or spiritualism (Spiritualist to Materialist). By generalizing across a wide variety of social groups that possess these beliefs, these attribute concepts did not coincide with any specific sensory quality, allowing us to target conceptual, rather than perceptual, representations. Multi-voxel pattern searchlight analysis was used to identify regions in which activation patterns distinguished the 2 ends of both dimensions: Conservative from Liberal social groups when participants focused on the political orientation dimension, and spiritual from Materialist groups when participants focused on the spiritualism dimension. A cluster in right precuneus exhibited such a pattern, indicating that it carries information about belief-attribute concepts and forms part of semantic memory—perhaps a component particularly concerned with psychological traits. This region did not overlap with the theory of mind network, which engaged nearby, but distinct, parts of precuneus. These findings have implications for the neural organization of conceptual knowledge, especially the understanding of social groups.

Key words: conceptual knowledge, semantic memory, neuroimaging, social attributes, precuneus

Introduction

Much of what humans represent about the world refers to qualities never directly seen. We are able to represent objects in terms of their intended functions (Bloom 1996; Kelemen and Carey 2007; Lombrozo et al. 2007), and we can represent social groups—such as Atheists or Republicans—in terms of their mental attributes, such as "conservatism." Conceptual representations (such as the concept conservatism) must therefore allow us to distinguish non-physical attributes while generalizing across the varied physical qualities of the people or objects they describe. Generality and independence from sensory particulars are 2 hallmark properties of semantic representations that set them apart from lower level, sensory ones. Toward the aim of understanding how conceptual representations are neurally implemented, the present experiment identified neural areas that contain representations with such properties, focusing on a particular content domain: belief attributes.

Most prior work on semantic knowledge has not applied such criteria to identify higher level representations in cortex, though an interest in such representations has become of increasing interest (e.g. Skerry and Saxe 2015; Mason and Just 2016). One reason for the prior gap is a heavy focus on concrete object knowledge (see Martin 2007 for a review). Past findings have identified neural areas that respond more strongly to 1 category of object than another (Martin et al. 1996; Chao and Martin 2000; Anzellotti et al. 2011; Konkle and Caramazza 2013) or that contain pattern information about individual object kinds (Haxby et al. 2001; Kriegeskorte et al. 2008) and which do so across modalities of stimulus presentation (Simanova et al. 2013; Devereux et al. 2013; Fairhall and Caramazza 2013; Fairhall et al. 2013; Clarke and Tyler 2014). However, these approaches are not sufficient to identify semantic representations per se, because concepts referring to concrete categories point to both semantic and sensory knowledge: concepts like “banana” are inevitably associated with specific, sensory qualities, such as the color yellow. Accessing such a concept—whether through a picture or a word—thus triggers both higher level and lower level memory retrieval (see Mahon and Caramazza 2008 for discussion). Thus, in most past research, objects in similar categories possessed similar lower level properties, such as shape, even if those properties were retrieved from memory.

By studying concepts that refer to non-physical qualities, one can better target higher level, semantic knowledge specifically. The present research does just this, by identifying regions that represent mental attribute concepts, specifically of beliefs (henceforth, “belief concepts”). We selected belief concepts in particular for 3 reasons. First, as we describe below, belief concepts allowed us to meet both criteria: they are highly general, and are not confounded with any particular sensory qualities. Second, by choosing a specific subset of concepts, rather than many kinds, we avoided assuming that all non-physical attribute concepts are represented in a single, common neural region. Much of the prior research on “abstract” concepts makes such an assumption (see Skipper-Kallal et al. 2015 for a recent example, though cf. Wilson-Mendelhall et al. 2013). Yet this notion runs counter to repeated findings in cognitive neuropsychology that neural areas can be highly specialized for semantic knowledge of a particular domain or attribute (Warrington and Shallice 1984; Ochipa et al. 1989; Caramazza and Shelton 1998; Miceli et al. 2001; Vandenbulcke et al. 2006; Capitani et al. 2009). Assuming otherwise could limit one’s ability to identify critically important areas. We thus take the stance that a characterization of the set of neural regions supporting semantic knowledge will more likely succeed if content-generality is not assumed, and specific kinds of concepts are studied individually. We thus aim to identify areas that would count among the (likely many) brain regions involved in conceptual representation.

This stance raises the question of which specific kinds of concepts one should target. Thus, the third motivating factor for selecting belief concepts is that prior research in social neuroscience allows us to make a theoretically motivated prediction about their neural localization. This work has described a set of neural regions that are engaged when thinking about mental attributes of people, relative to thinking about their physical attributes (Mitchell et al. 2002, 2005; Saxe and Wexler 2005; Lombardo et al. 2010; Ross and Olson 2010). This network includes bilateral temporo-parietal junction (TPJ), precuneus, dorso-medial prefrontal cortex (MPFC), and anterior temporal lobe (ATL), resembling a set of regions involved in social reasoning and mental inference, often termed the mentalizing network (see Koster-Hale and Saxe 2013 for a review). We hypothesized that belief concepts may be represented in or near some of these regions, stemming from the broader theoretical position that conceptual representations are localized in spatial proximity to other cognitive processes where their content may be most computationally relevant (Leshinskaya and Caramazza 2016).

To test whether a region represents belief concepts, we required that its activation pattern contain information distinguishing different belief attributes—an important qualification for a region that allows us to understand their meanings. We thus tested which regions encode instances of the same belief concept as similar, and distinguish among belief concepts that are different, to explicitly establish that they contain information about specific belief concepts. This approach allowed us to identify regions that represent belief concepts without assuming that those regions represent only belief concepts, nor all non-physical attribute concepts.

Furthermore, and most critical for establishing that these representations are conceptual, we ensured that the belief concepts we targeted were not coincident with any particular sensory property. To do so, our stimuli were names of a wide range of social groups, each of which could be characterized by its position along 2 belief dimensions (Fig. 1): “political orientation” (“Liberal” to “Conservative”) and “spiritualism” (“Spiritualist” to “Materialist”). We tested each dimension separately by varying participants’ task (Fig. 2), and looking for neural regions that represented the distinction between Liberal versus Conservative social groups when participants attended to political orientation, and Spiritualist versus Materialist social groups when participants attended to spiritualism, while explicitly generalizing over the unattended dimension. For example, to show evidence of representing spiritualism and “materialism,” activation patterns in a neural region had to be similar (correlated) between “Spiritual Conservative” and “Spiritual Liberal” groups, such as Fortune Tellers and Rabbis, while being different (less correlated) for Spiritual Liberal and “Materialist Liberal” groups. Thus, evidence of representing these concepts required broad generalization over a range of distinct social groups, and sensitivity to properties common to all of these social groups, which is their common belief attribute. This implies that even if participants retrieved vivid mental images corresponding to particular social groups, those images could not drive the results.

Finally, the attention manipulation made it unlikely that findings are driven by something specific about the social groups themselves, and more likely that they were driven by the mental attributes retrieved about them. When participants attended to political orientation, for example, Rabbis and Fortune Tellers were predicted to be dissimilar, opposite to the prediction during the spiritualism task. In summary, we searched for neural regions that explicitly represented highly general belief concepts, which captured non-physical similarities among a wide range of social groups.

Materials and Methods

Participants

Three non-overlapping groups of participants were included in this study. Sixteen participants completed online behavioral

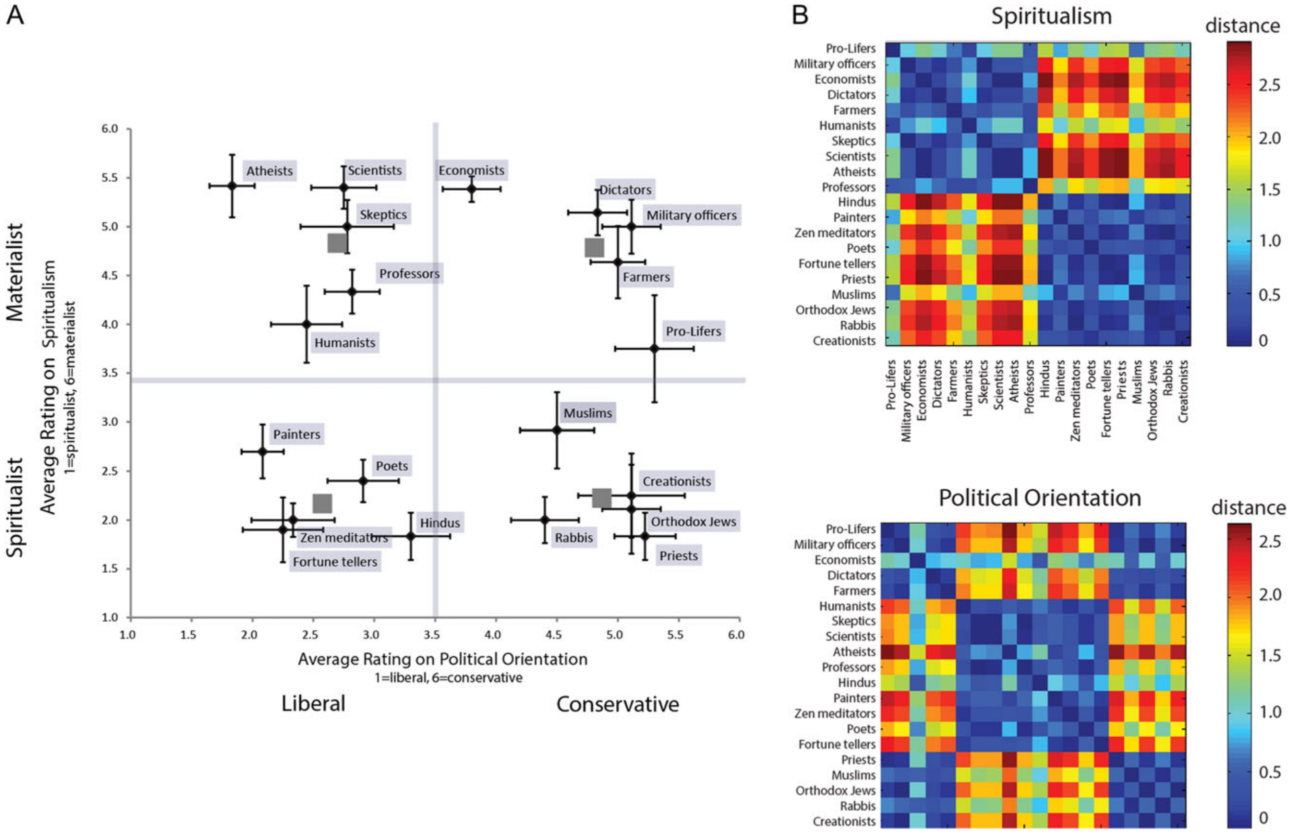


Figure 1. (A) Average value of each social group on each of 2 rating dimensions, Spiritualist to Materialist and Conservative to Liberal, on a rating scale from 1 to 6. Error bars indicate the standard error of the mean; n = 16. (B) Similarity norms: dissimilarity matrices for each rating scale, computed from the data in (A) by taking all absolute pairwise numerical differences between the social groups, on each rating scale separately.

tasks for stimulus norming; 32 participants performed a behavioral experiment in the laboratory; and 22 participants completed the functional magnetic resonance imaging (fMRI) experiment. All participants were college undergraduates recruited using an online subject pool and were native speakers of English; fMRI participants were further screened such that all were right-handed and had no history of neurological disorder. All procedures were approved by the Committee for the Use of Human Subjects of Harvard University. Two fMRI participants were excluded prior to data analysis due to scan interruptions (technical problems and/or participant discomfort, ending the sessions with insufficient data acquisition). There were 20 participants in the analyzed sample (10 male; mean age = 20 years; range 18–25).

Stimuli

The stimuli were names of 20 social groups, each of which fell along the spectrum from Spiritual to Materialist (spiritualism dimension) and from Liberal to Conservative (political orientation dimension). To select these 20 stimuli and obtain their exact position along these 2 dimensions, 16 participants were asked to rate 62 social groups on these scales. Spiritualists were described to participants as “people who believe in forces and beings that exist outside the physical world, including deities, spirits, and miracles.” Materialists were described as “people who believe primarily in those things that they can see.” Each social group name was presented one at a time with a 6-point rating scale below it, with the end points Liberal and Conservative (in one block) or Spiritualist and Materialist (in another block). An “unsure” response option was also available, in case participants were unfamiliar with the social group or did not know how to rate it. The task was administered online using Qualtrics survey software, and block order and scale order were counterbalanced across participants.

The final set of 20 social groups was selected by first excluding any item for which more than 30% of participants responded “unsure” on either rating scale. Responses to the remaining items were then averaged across participants, after removing outlier trials based on reaction time (shorter than 2 SD below and longer than 2 SD above the mean). On the basis of these ratings, items were then assigned to 1 of the 4 quadrants of the 2D belief space: Spiritualist Liberals, Spiritualist Conservatives, Materialist Liberals, and Materialist Conservatives. Five items per category were chosen to maximize several properties: items with large values on at least one dimension; low correlation between spirituality and political orientation; and comparable ranges in both dimensions—that is, the political distance spanned by the groups was equivalent to the spiritual distance spanned; and equivalent reaction times between Spiritualists and Materialists among the participants who judged spiritualism ($t[9] = -0.24$, $P = 0.82$) and between Liberals and Conservatives among the participants who judged political orientation ($t[9] = -0.70$, $P = 0.50$). These properties ensured that attribute dimensions were separable in this set of stimuli by virtue of being uncorrelated ($r = -0.08$, $P = 0.74$), were equally easy to retrieve, and had similar ranges in the stimulus set. Figure 1 shows the mean and standard deviation (SD) of each selected social group on each of the rated dimensions.

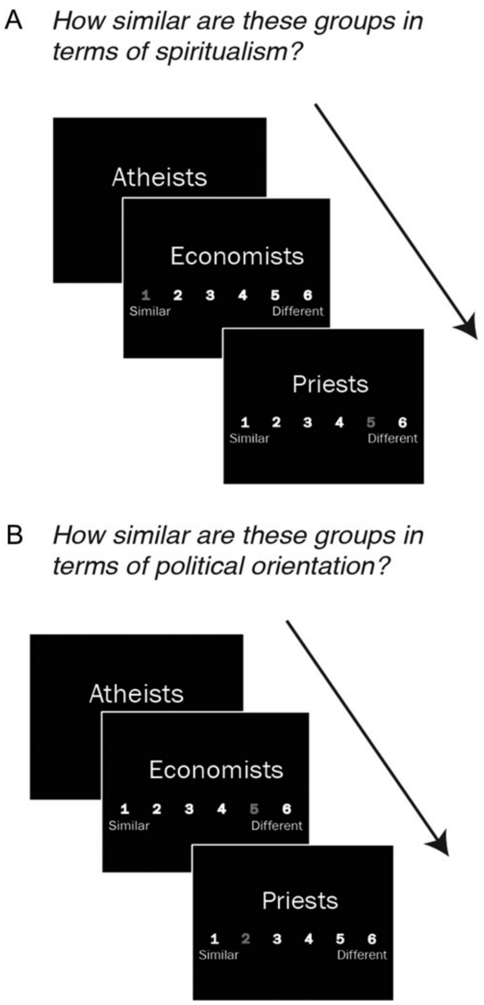


Figure 2. Illustration of the 2 one-back tasks in the neuroimaging experiment, with example responses in red. (*A*) The Spiritualism task; (*B*) The Political Orientation task. Results of stimulus norming study, shown for the social group names used in the neuroimaging experiment.

Experimental Task

The experimental task consisted of a one-back similarity judgment, based on one of the 2 dimensions (spiritualism or political orientation). Participants saw one social group name per trial (in white, 48-point Helvetica font on a black background), and compared it to the previous group by responding on a 1–6 scale from "Very Similar" to "Very Different," presented at the bottom of the screen (the order of the scale was counterbalanced across participants). Figure 2 illustrates a few example trials. If the attended dimension was spiritualism, participants were expected to compare the on-screen group to the previous group in terms of spiritualism. If both groups are Spiritual to an equal extent, they were asked to respond Very Similar; if they are both Materialist to an equal extent, they also should respond Very Similar.

In the presentation sequence, each social group was preceded by every other social group exactly twice. This ensured that each stimulus was the target of the same comparisons—to each other stimulus equally often. The neural response to each stimulus was then taken as the average across all of these different comparisons, allowing us to isolate the content of interest (the target trial). The optimized sequence thus contained 42 presentations of each stimulus. Trial duration was 3 s, during which the stimulus was on screen for 2.8 s and through all of which participants could make their response. The same stimuli were presented in both tasks, such that the only difference between the 2 tasks was the basis of the similarity judgment.

This experimental task was chosen for 3 reasons. First, it ensured that participants retrieved the relevant conceptual content about the groups (i.e. mental attributes). Second, the same stimuli could be presented while manipulating only which attributes about them are retrieved, thereby increasing confidence that effects are indeed about the belief attributes (political or spiritual) rather than any other properties of the groups. Lastly, it allowed dissociating the task response from the dimension of interest: social groups with similar belief attributes (e.g. high on conservatism) would not receive more similar task responses than groups with different belief attributes, because the responses on the task are based on the relation between the current and previous item, which varied across the trials in each condition.

To ensure that this task was not too difficult, and that one-back similarity judgments recapitulated the similarity relationships expected from direct ratings obtained in the stimulus norming study, 20 pilot participants were asked to perform the one-back similarity task in a laboratory setting to calibrate the timing and amount of practice needed to master the task, and another 12 to validate it. The validation sample demonstrated that participants' responses recapitulated the measures obtained from untimed, explicit ratings from the stimulus norming experiment.

fMRI Procedures

Pre-scan Behavioral Measures

Prior to scanning, fMRI participants were given a practice procedure with the one-back similarity task (as described in Experimental Task and Behavioral Piloting). They were first asked to rate either the Spiritualism or Political Orientation of each social group on a 1–6 scale, corresponding to the dimension they would be judging in the scanner, and were given practice on the one-back similarity task itself until they understood and felt comfortable with the task. They were given feedback on their response consistency and the number of missed responses, since there was no pure measure of accuracy.

Experimental In-Scan Task

During fMRI scanning, participants performed the one-back similarity task on 1 of the 2 dimensions. Ten participants performed the spiritualism task, and 10 performed the political orientation task; these were split between participants in order to keep each scan a reasonable length; note that the participant groups are never compared with each other. This task was split over 6 runs; the last item of each run was repeated as the first item of the subsequent run. This created runs with 148 trials (7.9 min in duration). Unique sequences were created by reassigning the items (social group names) to different trial codes in the sequence.

Theory of Mind Localizer

Between Runs 3 and 4 of the one-back similarity task, participants were given a theory of mind localizer (Dodell-Feder et al. 2011; www.saxelab.mit.edu). This task required participants to read 2 kinds of stories: those involving out-of-date beliefs ("false belief" condition) and those involving out-of-date physical representations such as photographs or maps ("false photograph" condition). All trials were composed of a story (10 s), followed by the true-or-false question (4 s), and a fixation block (12 s). Ten trials of each condition were presented across two 4.5-min runs. Theory of mind regions were localized by finding brain regions responding more to the false belief condition than the false photograph condition.

Post-Scan Behavioral Measures

After the scan, participants rated the groups on the dimension along which they did not rate the groups during the scan, as well as on the extent to which they personally liked each group (on a 1–6 scale, from Like to Dislike; scale order counterbalanced across participants) and how closely they identified with each group themselves (on a 1–6 scale, from Identify Closely to Do Not Identify; scale order counterbalanced across participants).

fMRI Acquisition Parameters

Neuroimaging data were collected with a 3-T Siemens Magnetom TrioTim scanner at the Harvard University Center for Brain Science, using a 32-channel head coil. Structural scans were acquired using an multi-echo magnetization prepared rapid gradient-echo sequence with 1-mm isotropic resolution and a 256 × 256 × 176 mm matrix size. Functional runs for the Theory of Mind Localizer were acquired with a gradient-echo, interleaved sequence (time repetition [TR] = 2.0 s, time echo [TE] = 28 s, flip angle = 85°, voxel size = 3 × 3 × 3 mm, gap = 2.5 mm, 32 slices, non-axial slices −35° to −40° from ACPC, matrix size = 108 × 108 mm, field of view = 216 mm). Functional runs for the experimental task were acquired with a similar sequence but optimized for higher spatial resolution (TR = 3.5 s, TE = 2.8 s , flip angle = 90°, voxel size = 2 × 2 × 2.5 mm, gap = 0 mm, 32 slices, non-axial slices −35° to −40° from ACPC, matrix size = 108 × 108 mm, field of view = 216 mm). The whole brain was covered except for up to 2 slices at the very superior tip of parieto-occipital cortex. The first 4 functional volumes of each sequence were discarded to ensure steady-state magnetization.

Preprocessing and Linear Modeling of fMRI Data

Functional data were preprocessed using AFNI software (Cox 1996). Slices in each volume were corrected for acquisition timing using Fourier interpolation (3dTshift). Each volume was then spatially aligned to the fourth volume of the first scan (3dVolReg). In each run, a Fourier high-pass temporal filter (0.008 Hz) was applied to remove low-frequency trends (3dDetrend), and image intensities were normalized. The data were spatially smoothed with a 4-mm full-width, half-maximum Gaussian kernel for the experimental runs, and a 6-mm kernel for the localizer runs. Two types of general linear models were fit to the experimental data: one that modeled activation individually for each of the 20 social groups, and one that modeled each of the 4 quadrants of the 2D belief space (Liberal Spiritualists, Liberal Materialists, Conservative Spiritualists, and Conservative Materialists). For the theory of mind localizer, predictors were created for the false photograph and false belief conditions, spanning both story and question presentation periods.

In all linear models, regressors were created by convolving their time-courses in the experiment with a gamma-modeled hemodynamic response. These convolved time-courses were used as predictors in a least-squares regression over the signal time-course in each voxel. The models also included regressors for motion, based on realignment parameter estimates in each of 4 directions and 2 rotations; as well as predictors for low-frequency linear trends across runs. The regression procedure produced a statistical map for each condition, representing a beta weight and *t*-statistic for each voxel, indicating the partial correlation between the signal in that voxel over the course of the experiment and the occurrence of that condition. The beta values are commonly interpreted as percent blood oxygen level-dependent (BOLD) signal change relative to the baseline condition, which here were the null trials.

Anatomical Surface Analysis

Anatomical data were processed using the Freesurfer software function recon-all (Fischl et al. 1999), which skull-stripped the volumes and used intensity gradients to segregate white and gray matter and generate inflated cortical surface maps for each individual. Inter-individual alignment was performed over the surfaces as follows: first, functional maps were aligned to each individual's native-space anatomical volume; the inflated surface based on this volume were then registered with other participants' surfaces using the AFNI function MapIcosohedron, and the alignment parameters from the volume to the resampled surface were used to align the functional data. These procedures were implemented using the Surfing Toolbox (available at http://surfing.sourceforge.net) and described in more detail by Oosterhof and colleagues (2011).

Multivariate Searchlight Analyses

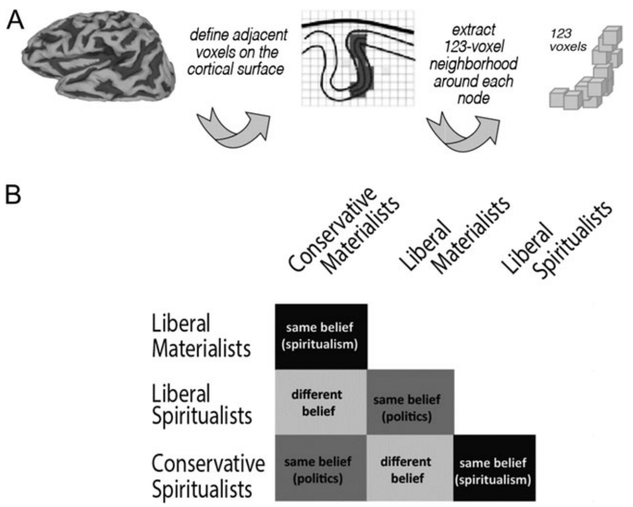
Neighborhoods surrounding each node (surface unit) were defined on the cortical surface by identifying 123 adjacent voxels, respecting the curvature of that participants' cortical surface (using the Surfing Toolbox [Oosterhof et al. 2011]). In contrast to neighborhoods defined volumetrically, this resulted in neighborhoods with a curved cylindrical shape that followed the contours of the sulci and gyri of each individual. By moving these surfaces into standardized space, these nodes are comparable across individuals.

Categorical Analysis

For each participant and in each neighborhood, the 123-voxel activation levels (*t*-values) associated with each quadrant condition were correlated pairwise, creating a condition × condition correlation matrix (as displayed in Fig. 3). These values were Fisher-corrected to normalize them and enable inferential testing. According to the belief dimension attended by that participant, each pair of conditions was then categorized as a "same-belief" or a "different-belief" pair. Thus, if a participant had attended political orientation, different conditions were categorized as same-belief pairs than if the participant had attended spiritualism, according to Figure 3B: same-belief pairs from the unattended dimension were used as different-belief pairs. For example, Liberal Materialists and Liberal Spiritualists were considered a same-belief pair for participants attending political orientation, but a different-belief pair for participants attending spiritualism. This ensured that our analysis uncovered regions which were sensitive to the attentional manipulation, as these models made largely opposite predictions.

The same-belief and different-belief correlation values were each averaged, and then subtracted from each other, creating a same versus different belief correlation difference for that neighborhood of voxels.

This created a correlation-difference map on the cortical surface of each individual, reflecting the discrimination of the 2 belief concepts attended by that participant. After this, the


A
define adjacent voxels on the cortical surface
extract 123-voxel neighborhood around each node
123 voxels
B

	Conservative Materialists	Liberal Materialists	Liberal Spiritualists
Liberal Materialists	same belief (spiritualism)		
Liberal Spiritualists	different belief	same belief (politics)	
Conservative Spiritualists	same belief (politics)	different belief	same belief (spiritualism)

Figure 3. (*A*) Schematic of the searchlight procedure: neighborhoods are defined by selecting sets of contiguous of voxels which follow the contours of the cortical surface. (*B*) Illustration of the categorical MVPA analysis. In each neighborhood, the voxel-wise BOLD signal is correlated between each pair of conditions, and same-belief and different-belief cells are, respectively, averaged and subtracted, yielding a measure of belief category discriminability. The same-belief quadrants were defined differently depending on the attended dimension (spiritualism vs. political orientation). The same-belief pairs of the unattended dimension were treated as different-belief pairs. Same-condition pairs are excluded from the diagram as these were not included in this analysis.

results from the task groups were combined, and group-level statistics were performed across the entire sample, collapsing across task. This was motivated by the assumption that a region that represents belief concepts should be sensitive to multiple kinds of beliefs, and thus, effects should hold across task groups, given the appropriate model for that task group. A one-tailed *t*-test against 0 was computed at each node, effectively testing that the same-belief correlation was greater than the different-belief correlation.

Full-model Analysis

An alternative multivoxel pattern analysis (MVPA) approach involved computing all the pairwise correlations among the 20 individual social groups in terms of the voxel-wise patterns in each neighborhood, and then correlating the lower half, off-diagonal of this matrix to the corresponding entries in the stimulus similarity norms for the appropriate dimension (Mur et al. 2009). Both matrices were symmetric. The norms were used because in-scanner responses likely included task-related error and other noise. This analysis produced a correlation value at each node, reflecting the fit between the full similarity model to the voxel-wise patterns in its neighborhood. These values were Fisher-corrected and submitted to a group *t*-test, collapsing across task, just as in the categorical MVPA analysis.

Multiple Comparison Correction

To correct for the multiple comparisons across the 10 000 nodes, a permutation analysis was used following Oosterhof et al. (2010), in which the sign of each participants' *r*-differences (categorical analysis) or *r*-value (full-model analysis) across the cortex was reversed with 50% probability, and the group *t*-test is repeated (this is equivalent to swapping the within and between correlation values before computing the *r*-differences). Ten thousand such *t*-maps were produced, and in each one, the maximal cluster size above an initial threshold of $P < 0.005$ uncorrected was extracted, providing a distribution of maximal cluster sizes expected under the null hypothesis (i.e. that same and different correlation values are equal). This distribution was used to assign corrected probability values to observed cluster sizes, by locating their position in this distribution. The multivariate procedures were performed using custom code in MATLAB (The MathWorks Inc 2009) and functions from the Surfing Toolbox (Oosterhof et al. 2011).

ROI and Mask Definition

The theory of mind localizer scan was used to find voxels more active during false belief than false photograph stories, thus localizing brain regions that were preferentially active when thinking about mental states ("theory of mind regions"). To define participant-specific regions of interest (ROIs), each theory of mind region was defined individually in each participant, using volumetric data to be consistent with prior approaches. First, a group-average mask—thresholded liberally at $P < 0.01$, uncorrected, to ensure inclusivity—was used to outline the location of each spatially distinct area most typically investigated as part of the theory of mind network: MPFC, precuneus/posterior cingulate, and left and right TPJ. Participant-specific regions were selected by finding that participant's top 200 contiguous voxels (400 mm^2) within the group boundary. This ensured that the number of voxels in each region and each participant was equal, and that the included voxels were the most selective ones to the false belief condition for that participant, without being excessively large (which could harm a correlation-based MVPA analysis).

Peak Region Analysis

To identify individual participants' peak coordinates in the searchlight analysis, the location of each participant's maximal correlation-difference value from the surface-based searchlight MVPA analysis that was closest to the group-average peak in the right precuneus was identified, projected back to the volume and transformed into Talairach space. Clusters of various sizes were then defined around the peaks, by taking the 25–200 voxels (in increments of 25) showing the largest *r*-value differences. A similar approach was used for theory of mind peaks. Proportion overlap was then computed at each cluster size by dividing the number of voxels shared between the 2 clusters by the total size of each. Because the cluster sizes were defined by relative rank, this analysis did not depend on trying to equate beta values and *r*-difference statistics.

Results

Behavioral Measures During the In-Scan Task

During fMRI scanning, participants performed a one-back similarity judgment over social group names, being asked to attend either to spiritualism or to political orientation (as depicted in Fig. 2). Participants responded on 96% of all trials. Moreover, similarity judgments of each pair of items (converted to distances) were highly correlated with the distances derived from the norming sample ratings (shown in Fig. 1*B*) on the appropriate dimension (mean $r = 0.58$, $t[19] = 16.40$, $P < 0.0001$, one-tailed) (all reported *r*-values are Fisher-corrected before being submitted to the *t*-test. Only the 190 unique, off-diagonal, pairs are used in correlation analyses between similarity matrices). These judgments also correlated with distances on the unattended dimension (mean $r = 0.10$, $t[19] = 6.50$, $P < 0.001$, 1-tailed) but to a significantly smaller degree ($t[19] = 12.40$, $P < 0.0001$, 2-tailed).

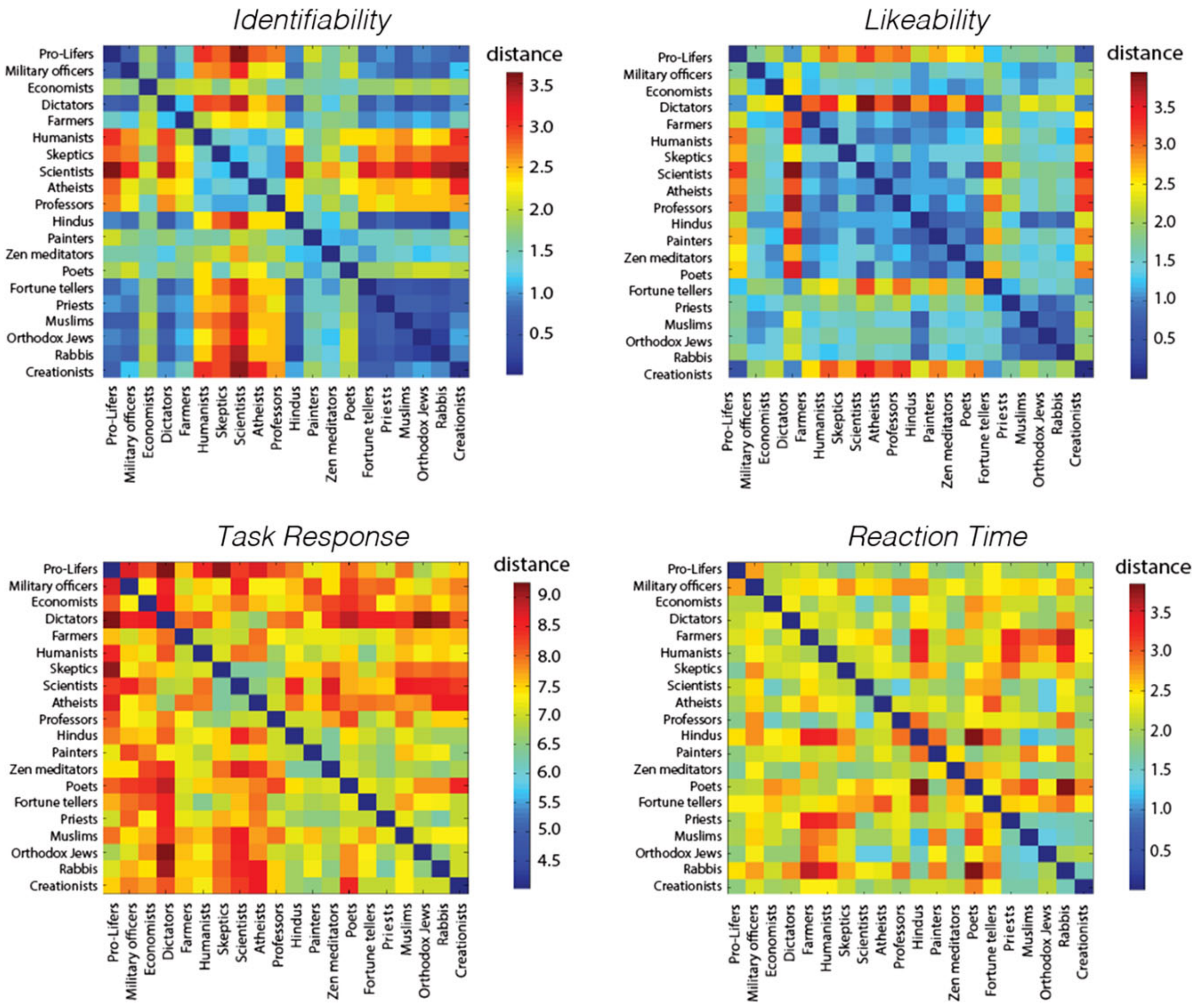


Figure 4. Similarity matrices showing absolute pairwise differences on 4 behavioral measures collected from the participants in the neuroimaging experiment, all considered models of no-interest: clockwise from the top, ratings on how closely the participants report identifying with each group; how much the participants like each group; the task response during scanning; reaction time of the task response during scanning.

This implies that the 2 belief concept dimensions were appropriately retrieved according to the task (perfectly selective attention would be indicated by $r = 0$ for the unattended dimension). Participants' judgments were also correlated to that of each other, within their task group (mean $r = 0.59$). Thus, participants successfully retrieved the relevant attributes during the in-scan task, and were consistent in their judgments, re-instantiating the similarity structure observed in the stimulus norms.

On the other hand, the task response and reaction time to each of the 20 social groups on "average" (across all comparisons) were not expected to correlate with belief-attribute distances, as each group was compared with every other group equally often. Indeed, pairs of social groups with similar belief attributes did not have more similar reaction times (t[19] = −1.34, P = 0.90, 1-tailed) or response values (t[19] = −0.49, P = 0.69, 1-tailed). The similarity matrices of these variables are shown in Figure 4.

To rule out potential confounds of reaction times and responses, we tested these variables in a fashion analogous to the categorical MVPA analyses of fMRI data. This involved testing whether responses for same-belief (within-category) groups were on average any more similar than different-belief (between-category) groups, according to the attended dimension (as shown in Fig. 3B). For each participant, the reaction times and responses to all social groups belonging to each of the 4 quadrant conditions were averaged (e.g. the average reaction time to all Liberal Materialists). Next, the absolute differences in responses and reaction times were computed between each pair of quadrant conditions. The within-category and between-category pairs were, respectively, averaged. The average within-category difference was subtracted from the average between-category difference, creating a within-versus-between response difference measure for each participant, which was tested against 0 at the group level. This difference was significant for responses (t[19] = 3.37, P = 0.001, 1-tailed) but not reaction times (t[19] = 0.80, P = 0.22, 1-tailed), indicating that responses were more different between different-belief groups. This effect in the categorical analyses was surprising, given the absence of such an effect in the full pairwise analysis of all 20 social groups (above), and given the design of the task. Given the lack of correlation between task responses and the dimensions of interest when using the full matrix of items (as reported above), this categorical reaction time effect was most likely driven by noise rather than any true relation with the dimensions of interest. Nonetheless, despite the fact that the differences in task responses were not explicitly related to the nature of the categories, the related properties of the response values and these categories make it essential that both a categorical and a full pairwise approach to the neural data are considered.

Behavioral Measures Acquired Post-Scan

Participants' likeability and identification ratings were treated in the same way as the task response variables, to ensure they were not confounded with belief-attribute similarity. As above, this was first tested item-wise: by assessing the correlation between distances in terms of these ratings and the belief-attribute distances. Second, it was also tested with a categorical analysis on each dimension to see whether groups in same-belief quadrants were given similar ratings relative to groups in different-belief quadrants, on average.

The distances from the belief similarity norms were weakly correlated with distances in likeability ratings (mean $r = 0.08$; $t[18] = 2.54$, $P = 0.01$) (Likeability and Identification ratings were missing for one participant due to time constraint. We report the correlations among the rest of the participants) and distances in identification ratings (mean $r = 0.12$; $t[18] = 3.63$, $P = 0.001$). However, groups in the same category as a whole (e.g. Spiritualists) were not more similar than those in different categories, in terms of either likeability ($t[18] = 0.29$, $P = 0.38$) or identity ($t[18] = -0.27$, $P = 0.61$). This implies that the categorical MVPA analyses of the fMRI data are specific to belief attributes, and are not confounded with either how likeable those groups are to these participants, or how closely the participants identified with them.

Categorical MVPA Searchlight

The aim of the categorical MVPA analysis was to identify neural regions that distinguished Liberal from Conservative groups when political orientation was attended, and Spiritualist from Materialist groups when spiritualism was attended. This was performed over the 4 quadrant conditions, Liberal Materialist, Liberal Spiritualist, Conservative Materialist, and Conservative Spiritualist, which were created by averaging across the 5 specific social groups belonging in them, as indicated in Figure 1*A*. Thus, in a whole-brain searchlight, we tested each neighborhood of voxels to determine whether its activation patterns were more similar for pairs of same-belief groups than for pairs of different-belief groups, according to the attended dimension. Figure 3*B* illustrates these predicted relations for each condition pair. For example, for participants performing the spiritualism task, this analysis identified neighborhoods where the voxel-wise pattern responses to Conservative Materialists correlated more with Liberal Materialists than with Conservative Spiritualists, indicating that such a neighborhood distinguished between Spiritualist and Materialist groups across large variation in political orientation. This prediction was reversed for the participants performing the political orientation task, where same-belief pairs from the spiritualism task became different-belief pairs. Thus, the similarity model applied was specific to the dimension attended by the participant, keeping constant the actual stimuli, therefore looking for regions distinguishing Spiritualists from Materialists, when this distinction was attended, and Conservatives from Liberals, when that distinction was attended.

By this procedure, at each node, a correlation-difference value was obtained for each participant, and a *t*-test was performed across the participants' values, collapsing across task. Task groups were combined because our interest was in a neural region that represents more than one kind of belief concept, and which should therefore show consistent effects across our participants. When correcting for multiple comparisons across the entire cortical surface, the right precuneus passed the significance threshold of $P < 0.05$ (Fig. 5). Below, the voxel-wise pattern distances (inverse correlations) between each of the 20 social groups are shown, separately for each task group; these were extracted from the peak node in each participant within the right precuneus cluster. As the graphs demonstrate, the patterns in this region differed according to attended dimension, additionally confirmed by a significant negative correlation between them ($r = -0.20$, $P = 0.005$).

Full-Model Analysis and Task Response Controls

To rule out a confound of task response, which correlated with the category distinction but not the full similarity space of all 20 groups, a full-model searchlight analysis was performed as a convergent test. This involved extracting, at each node, the voxel-wise patterns in response to each individual social group, and correlating these patterns with the similarity norms, according to the participant's attended dimension.

The full-model analysis must be treated only as complementary to the categorical analysis, as on its own, it has several limitations. First, it does not require explicit generalization across the opposite, unattended belief dimension, and thus could be significant in regions encoding only within-quadrant similarity (e.g. among the 5 Spiritual Materialist groups) rather than the general categories of interest here. The full similarity model is furthermore confounded with likeability and identity ratings (across subjects and thus across both dimensions; see Behavioral Measures Acquired Post-Scan); however, it is not confounded with task response, and thus serves as an essential convergent analysis.

It was indeed found that an overlapping portion of the right precuneus passed significance in the full-model searchlight (Fig. 6). Because this region was significant in both the categorical and the full-model analyses, its pattern of activation cannot be due to any of the nuisance variables considered. Regions identified only in the full-model analysis likely reflect nuisance variables that were not the target of the investigation, and thus cannot be interpreted.

To confirm that the very cluster observed in the categorical MVPA results was, on the whole, specifically driven by belief similarity and not task response, the full-model neural similarity matrix was extracted from the cluster found in the categorical searchlight analysis. Although testing the significance of the full-model fit in these ROIs is biased (because the categorical structure is also implicit within the full-model similarity matrix), it is equally biased for belief category and for task response, since both predict the same outcome in the categorical analysis. Instead, the aim of this analysis was to test whether the full similarity model captures unique variance attributable to the belief similarity model only, controlling for task response similarity using a partial correlation. This was indeed the case (mean partial $r = 0.06$, $t[19] = 3.72$, $P = 0.0007$), suggesting that the belief similarity model captured distinct variance from the task response model in the right precuneus.

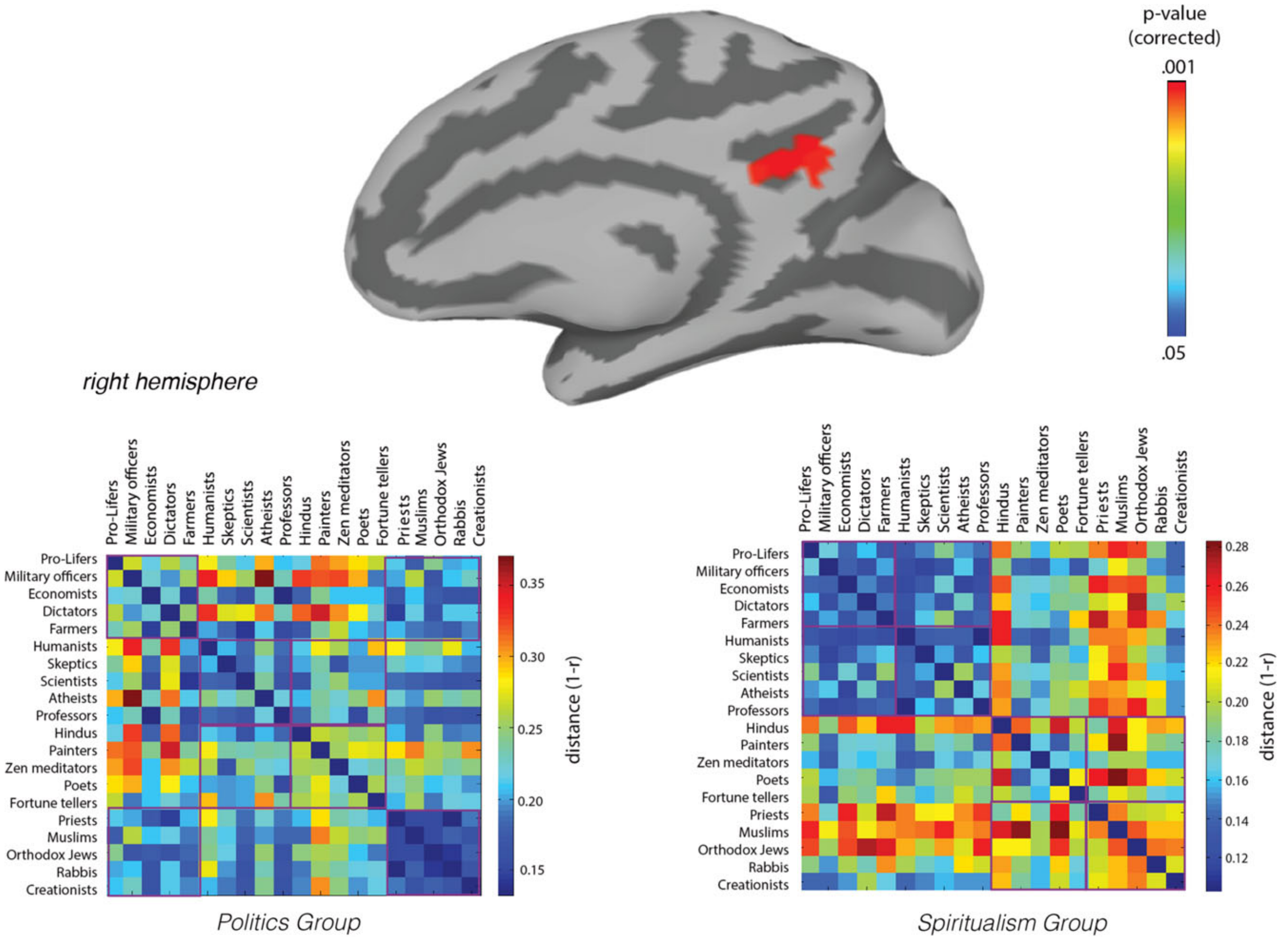


Figure 5. Categorical MVPA searchlight results, $n = 20$. Results from a whole-brain searchlight, showing the only significant cluster, in the right precuneus. Below, for illustration only, similarity matrices showing the pairwise distances between each pair of social groups, measured as the inverse correlation of the voxel-wise patterns between those items, as measured in each participant's peak node (searchlight neighborhood) within the precuneus cluster. Purple rectangles surround same-condition quadrants and within-category conditions: that is, pairs expected to be similar to each other.

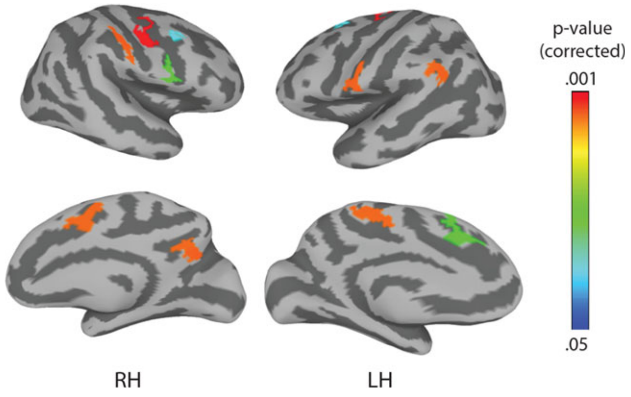


Figure 6. Full-model, whole-brain searchlight results; $n = 20$.

Interactions

To ensure that the searchlight categorical MVPA effects in the right precuneus were driven equally by both spiritualism and political orientation dimensions, mean correlation differences within the significant cluster, for each subject, were split by task group and compared. This comparison was not significant ($t[9] = 0.68$, $P = 0.51$). Because differences between the 2 belief dimensions could not be detected (main effects [e.g. looking at groups separately] are biased by the ROI selection and are thus not reported. At the whole-brain level, testing the effects separately in each group is insufficiently powerful with $n = 10$), they were unlikely to have been driven by one group more than the other. This supports the conclusion that the right precuneus represents both kinds of belief-attribute dimensions, though the optimal test would be to show independent effects for each, with a doubled sample size.

ROI Analysis in Theory of Mind Regions

To see whether belief category MVPA effects could be found within any theory of mind region, an ROI analysis was performed at the individual subject level. Group-level contrasts are anatomically imprecise and smooth over individual variability, increasing the probability of overlap (Saxe et al. 2006; Fedorenko et al. 2010). Thus, 4 canonical theory of mind regions (left and right TPJ, precuneus, and MPFC) were defined individually in each participant (see ROI and mask definition), and their voxel-wise patterns of response to each quadrant condition were extracted. A categorical MVPA analysis was then performed in the same way as in each searchlight neighborhood,

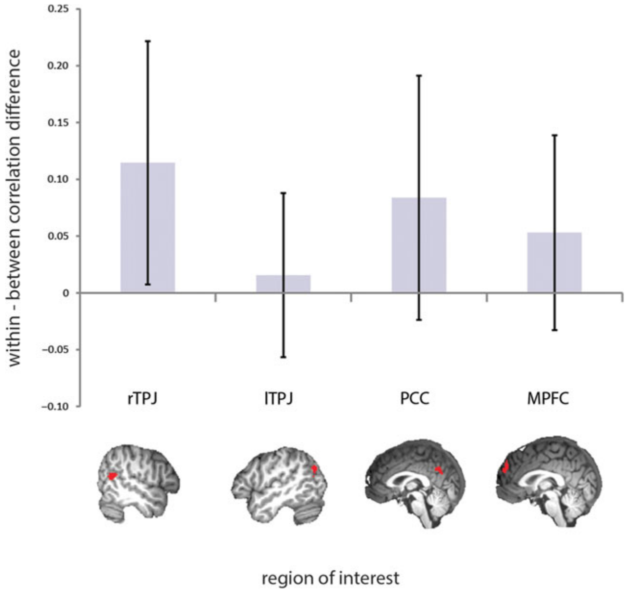


Figure 7. Results of a categorical MVPA analysis within individually defined theory of mind regions (200 voxels in size), as identified using an independent theory of mind localizer. Images of each region shown from one participant. Error bars show standard error of the mean; n = 20. All comparisons against 0 non-significant.

testing whether the difference between same-category and different-category condition pairs was greater than zero. These difference values, averaged across participants in each ROI, are shown in Figure 7. None was significant (all $P > 0.10$, 1-tailed), indicating no discernable belief-attribute representations in these theory-of-mind regions.

It was surprising to see no effect in the precuneus ROI, despite apparently nearby effects in the searchlight analysis. This is unlikely to be due to a difference in the number of voxels included, as searchlight neighborhoods and ROIs were of roughly similar sizes (123 vs. 200 voxels, respectively). However, one important difference is that medial regions on the cortical surface are by necessity split into hemispheres, differently from the way they are typically treated in volumetric analyses standardly used to define theory of mind ROIs; we followed prior approaches in defining only one, bilateral precuneus. However, splitting the regions is more accurate, as the 2 hemispheres are indeed physically unconnected even in their medial aspects. Thus, breaking with traditional theory of mind ROI definition, a specifically right precuneus ROI was defined; in most participants, a right precuneus peak was distinguishable from the larger, more medial cluster which appeared, in the volume, to span both hemispheres. The new ROI was furthermore defined to be 123 voxels in size to more perfectly match the searchlight size. However, belief category MVPA analysis in the right precuneus was, also, not significant (t[19] = 0.68, P = 0.25, 1-tailed). Thus, the null results in the theory of mind precuneus region are not simply due to artificial pooling of the bilateral precuneus.

Overlap and Divergence of Theory of Mind and Belief-Attribute Searchlight Peaks

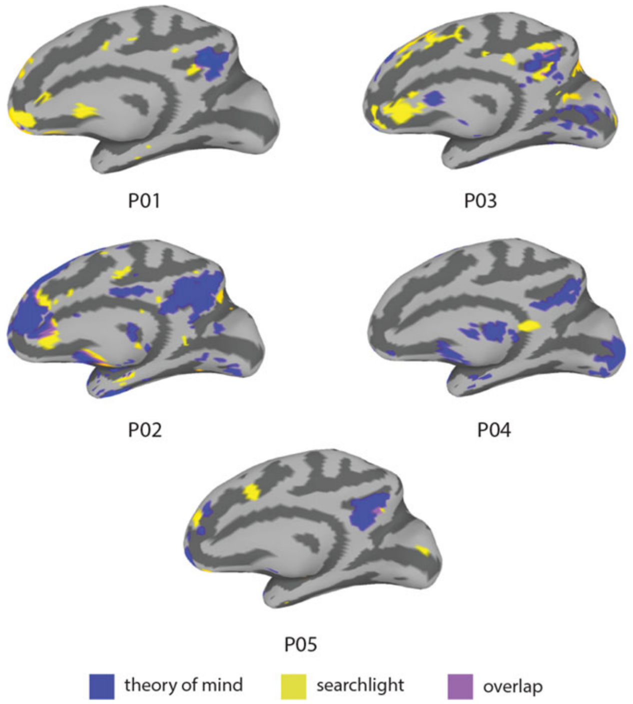


Figure 8. Overlays of theory of mind localizer and categorical MVPA searchlight effects, individual participant data (first 5 shown). Thresholded, for illustration purposes only, at $t > 2$ for theory of mind and r-difference > 0.5 for searchlight effects.

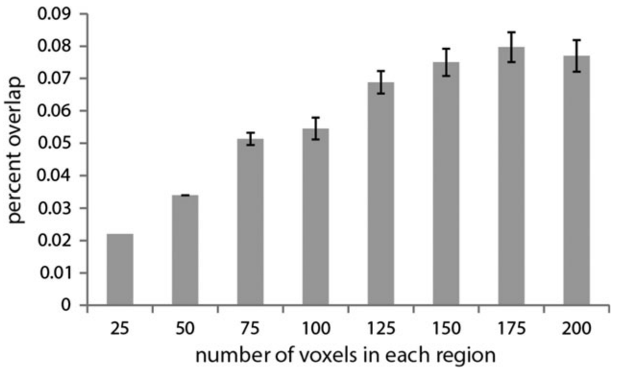


Figure 9. Average percent overlap between individual subject theory of mind and categorical MVPA searchlight regions (n = 20). Regions were defined by taking a certain number of the most active voxels surrounding each subject's peak. Percent overlap is plotted as a function of region size; error bars show standard error of the mean.

ROI results suggest that belief attributes are represented in regions close to, but distinct from, those maximally responsive during the False Belief task. The overlaps between theory of mind and searchlight results in each individual subject's right precuneus, as illustrated in Figure 8, are consistent with this account. As a more explicit test of the dissociation between these right precuneus loci, an overlap analysis was performed in each participant by defining a cluster of voxels around each of the 2 peaks, and evaluating their proportion of overlap. Clusters of different sizes were defined by taking a specified number of the most active voxels for that contrast (e.g. the top 25 voxels). As shown in Figure 9, the amount of overlap ranged from 2% to 8% across cluster sizes from 25 to 200 voxels, and decreased as the clusters became larger. This means that as the regions became more selective for their respective contrast, the proportion of overlap between them decreased, suggesting the overlap of the 2 contrasts is driven by weaker, rather than stronger, voxels.

Table 1 Locations of the categorical MVPA searchlight and theory of mind effects

Belief category searchlight MVPA			Theory of mind		
x (medial–lateral)	*y* (anterior–posterior)	*z* (inferior–superior)	*x* (medial–lateral)	*y* (anterior–posterior)	*z* (inferior–superior)
8.95 (3.14)	−53.60 (11.13)	27.05 (6.68)	8.95 (3.25)	−53.70 (4.65)	32.75 (6.03)

Notes: Units are mean Talairach coordinates of individual participants' peaks. SDs are shown in parentheses.

To test whether the 2 loci diverged in any systematic direction in space, the coordinate positions of their peaks were compared across participants along each spatial axis. These coordinates are reported in Table 1. While the *x* and y dimensions did not show significant differences, the *z* dimension (inferior to superior) showed a reliable difference ($t[19] = 2.25$, $P = 0.04$); though it should be noted that this effect does not survive correction for the number of dimensions, which would require $P < 0.016$. It is also notable that searchlight effects appeared to vary widely on the y (anterior to superior) axis, suggesting that the spatial non-overlap of theory of mind and MVPA effects was driven by the variability of searchlight effect localization along the y-axis, instead of (or as well as) a more systematic spatial discrepancy along the z-axis. Finally, a correlation analysis between the 2 sets of coordinates in each dimension revealed that they were negatively or undetectably correlated ($r = -0.22, -0.07, -0.58$, for *x*, *y*, and *z*, respectively). In summary, the location of theory of mind effects is not a good predictor of the localization of belief-attribute representations in the right precuneus in individuals, whether due to systematic spatial shift or simply independent variability in the searchlight effects. Although group-average effects appear broadly similar, this apparent similarly masks the underlying, individual anatomical dissociations between these effects.

Discussion

Using searchlight MVPA, we uncovered a cortical region in right precuneus that had distinct patterns of activation for Liberal versus Conservative social groups when political orientation was attended, and for Spiritualist or Materialist social groups, when spiritualism was attended. By using a wide variety of social groups, and keeping the social groups the same while making distinct predictions according to the dimension retrieved, we have been able to show that this region represents the belief-attribute concepts Liberal, Conservative, Spiritual, and Materialist.

These concepts are highly general: spiritualism characterized social groups as diverse as priests and poets, whereas Liberalism spanned poets to scientists. They are also non-sensory: no perceivable quality could define category membership. Finally, social groups in same and different belief categories were equal in how much the participants liked them and identified with them. Thus, the representations identified are semantic—general belief concepts—and not perceptual, affective, or situation-specific ones.

In addition, the portion of right precuneus in which we found these representations was adjacent to, but not overlapping with, the part of precuneus identified as part of the theory of mind network using a standard localizer (Dodell-Feder et al. 2011).

Semantic Role of Precuneus

The results of the MVPA identify precuneus as part of the brain's semantic system. This is an extension of past findings that have identified a network of regions, including the precuneus, which are engaged by semantic tasks (as compared with non-semantic tasks; see Binder et al. 2009 for a meta-analysis) and which allow cross-modal classification of various categories of objects and animals (Fairhall and Caramazza 2013). Although these prior approaches were able to identify candidate neural regions involved in semantics, they were not necessarily specific to semantics, because retrieving word meanings engages not just semantic representations—general aspects of meaning—but also mental imagery of specific sensory or episodic memories. These systems are hard to disentangle: the distinction between categories of objects (such as animals, fruits and tools) can be formulated in terms of their associated sensory or motor properties. Thus, regions that can distinguish between such categories, even across different modalities of presentation, are not necessarily semantic. Here, we circumvented this problem by using category distinctions that do not differ in terms of perceptual properties in their core definition, but denote kinds of beliefs (e.g. Liberal vs. Conservative). We thus provide further support for the role of the precuneus in conceptual representation.

Recently, Hassabis et al. (2014) also found that a portion of precuneus was able to distinguish between scenarios that participants constructed involving novel characters described as extraverted or introverted, suggesting it might represent personality trait concepts. However, these observations could have reflected properties of the constructed scenarios rather than the trait concepts themselves; in our case, the broad generalization across a variety of social groups, with no particular demand on constructing scenarios, makes this account unlikely. In contrast to previous research, the present work reports the presence of specifically semantic representations in the precuneus.

It is important to emphasize that the main thrust of our findings is to demonstrate a representational property of the precuneus: namely its capacity to represent information that is independent of sensory particulars and sensitive to non-physical attributes. It is this capacity that qualifies it to be part of the full set of regions that support conceptual knowledge. These findings were not intended to identify all conceptual regions, nor any particular region representing all semantic contents. Much more than a single finding will be needed to fully characterize the content and representational capacity of the right precuneus and its relation to other regions involved in semantic representations.

Distinction Between Semantics of Beliefs and Theory of Mind

Our findings also contribute to the resolution of a puzzle regarding the relation between semantics and theory of mind: that candidate semantic regions as described above (Binder et al. 2009; Fairhall and Caramazza 2013) resemble the theory of mind network, which is defined as those regions engaged more during mental state inference than equally semantic but non-social inference (e.g. the false photograph task as employed here; see Koster-Hale and Saxe 2013 for a review). Our findings rule out the idea that the same neural region underlies both semantics and theory of mind in right precuneus. Indeed, representations of belief concepts in precuneus were found outside individual participants' theory of mind precunei, they were absent within the latter regions, and their respective locations were not predictive of each other. Thus, at least some socially relevant concepts are represented outside the theory of mind network proper, making it unlikely that a singular, homogeneous network represents both theory of mind and social semantic information of all sorts (c.f. Spreng and Grady 2010; Hassabis et al. 2014).

General Principles of Semantic Organization: Adjacency

It should not be ignored, however, that the 2 right precunei—the one active for theory of mind and the one containing information about belief concepts—were adjacent, with nearly identical group-average coordinates in 2 dimensions (Table 1). Not all conceptual knowledge is represented in the precuneus—or indeed any one neural locus—and the principles by which this content is distributed remains a core issue in the neuroscience of semantic memory.

One possibility is that semantic content is localized according to the computation such content supports, carried out in regions connected to it (Mahon and Caramazza 2011). This possibility becomes more likely when the content of such knowledge is not determined by any sensory quality—as here—making it difficult to account for its localization based on an association with a sensory modality, as suggested by alternative accounts (Martin and Chao 2001; Thompson-Schill 2003). The computation principle also fits with our recent findings of object function representation in anterior inferior parietal lobule (Leshinskaya and Caramazza 2015), near coordinates previously found to respond to tools and their movements: both types of information would be relevant for using tools to achieve goals. Belief concepts, in turn, are relevant for understanding others' minds—the function clearly attributed to the theory of mind network. Subcomponents of this network, and adjacent conceptual regions, might each serve a complementary function toward this common goal. Characterizing these unique contributions can provide clues about how this goal is accomplished.

General Principles of Semantic Organization: Selectivity

It is well established that the semantic system contains subdivisions selective to certain kinds of contents; these are likely based on evolutionarily relevant domains such as animate and inanimate kinds (Caramazza and Mahon 2006). Is "social" such a domain? Recent neuropsychological evidence shows that semantic knowledge about social groups can be selectively impaired due to neurological damage (Rumiati et al. 2014) and neuroimaging work has found that retrieving such knowledge preferentially activates a number of regions, including the precuneus (Zahn et al. 2007; Contreras et al. 2012). However, the precise content of these neural areas remains to be characterized. If the portion of precuneus reported here is part of a socially selective semantic subsystem, then our findings would suggest that concepts of belief attributes are one of the things that a social semantic system represents.

The challenge lies in connecting past findings to each other. Prior work has found that part of the precuneus shows stronger responses for retrieving detailed semantic properties of famous people more than famous landmarks (Fairhall et al. 2013), and for mental/psychological adjectives than physical attributes (Mitchell et al. 2005; Lombardo et al. 2010; Skipper et al. 2011 though see Moran et al. 2011), including those of social groups (Contreras et al. 2013). Curiously, it also responds to the retrieval of psychological versus physical attributes of dogs (Mitchell et al. 2005) and was not found when comparing social adjectives to descriptors of animal behaviors such as "trainable" (Zahn et al. 2007; Ross and Olson 2010), overall suggesting that it is selective to attribute retrieval, rather than the humanness of the target. In summary, there is good reason to believe that parts of the precuneus show selective responses to psychological attributes.

Whether these past findings are in the same part of precuneus as the present ones is uncertain, but likely. One concern has been that selective responses to psychological adjectives may have been driven by either the operation of a socially selective semantic system, or the incidental engagement of social-inferential processes (Zahn et al. 2007)—predicting behavior, inferring desires, and beliefs—in the theory of mind network, which was not explicitly localized in those studies. However, person-selective precuneus scales with semantic demand (retrieving more vs. less specific content), and does so more during the retrieval of person-related rather than place-related knowledge (Fairhall et al. 2013). If the level of semantic demand does not also increase theory of mind demand, this suggests a content-selective, semantic effect (see Zahn et al. 2007 for a similar argument about the ATLs). It thus seems likely that representations of belief concepts are part of such a system.

Limitations

Our findings relied primarily on a categorical approach to data analysis to control for nuisance variables. This could have limited our ability to find representations that are highly sensitive to the differences within a category, while still encoding belief attributes. Furthermore, it could have missed areas that represented only one end of a spectrum consistently, because it required within-category similarity on both ends. Nonetheless, we sought a region that represented multiple kinds of belief concepts, for increased generality.

As argued in the Introduction, an approach that localizes specific concepts is more likely to succeed than one that assumes content-general conceptual regions and fails to find them. Nonetheless, even given this success, it remains for future research to characterize the representational scope of the identified region. Some regions, such as prefrontal cortex, are known to contain flexible, task-defined representations across a wide variety of contents (Freedman et al. 2001; Roy et al. 2010). Is the precuneus such a region, or is its domain limited? To answer this, our findings can be placed within the context of work in other domains that makes use of similar tasks or analytic approaches. The case of inanimate object properties in anterior temporal cortex (e.g. Peelen and Caramazza 2012) and emotion attributes in MPFC (Skerry and Saxe 2014) provides salient examples of content domains that are represented primarily in other regions. Nevertheless, identifying the precise scope of the precuneus awaits future research.

Conclusion

The signature of semantic representations is their capacity to represent properties not directly available to the senses—for example, belief concepts. We used this signature to identify a region in the right precuneus as part of the brain's semantic system, by showing its capacity to represent concepts like Liberal and Spiritual. We further observed that these concepts were represented in an adjacent—but non-overlapping—part of right precuneus to that engaged during theory of mind. Although the precise scope and selectivity of this part of precuneus has yet to be determined, it likely has a privileged role in representing mental attributes, and works together with other regions, such as the adjacent theory of mind precuneus, that encode distinct but complementary content for understanding our social environment. It is possible that the participation of belief concepts in social reasoning is part of the reason they are localized as they are. Finally, we anticipate that by characterizing the distinct components of this system, we can ultimately unravel the nature of the computations that allow us to think of people in terms of their internal qualities.

Funding

Natural Sciences and Engineering Research Council of Canada (to A.L.); Norman Anderson graduate student fund (to A.L.); National Science Foundation (to J.M.C.).

Notes

We would like to thank Tammy Moran and Ross Mair for their help with fMRI protocol development and optimization; Andrew Messing, Katharine Dobos, Grace Mahoney, and Ahn Ton for help with data collection; Nick Oosterhof for guidance in surface-based and permutation analyses; and Stefano Anzellotti and Mark Thornton for statistical advice. *Conflict of Interest*: None declared.

References

Anzellotti S, Mahon BZ, Schwarzbach J, Caramazza A. 2011. Differential activity for animals and manipulable objects in the anterior temporal lobes. J Cogn Neurosci. 23(8): 2059–2067.

Binder JR, Desai RH, Graves WW, Conant LL. 2009. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. Cereb Cortex. 19(12): 2767–2796.

Bloom P. 1996. Intention, history, and artifact concepts. Cognition. 60:1–29.

Capitani E, Laiacona M, Pagani R, Capasso R, Zampetti P, Miceli G. 2009. Posterior cerebral artery infarcts and semantic category dissociations: a study of 28 patients. Brain. 132(4): 965–981.

Caramazza A, Mahon BZ. 2006. The organization of conceptual knowledge in the brain: the future's past and some future directions. Cogn Neuropsychol. 23(1):13–38.

Caramazza A, Shelton JR. 1998. Domain-specific knowledge systems in the brain the animate-inanimate distinction. J Cogn Neurosci. 10(1):1–34.

Chao LL, Martin A. 2000. Representation of manipulable man-made objects in the dorsal stream. NeuroImage. 12(4): 478–484.

Clarke A, Tyler LK. 2014. Object-specific semantic coding in human perirhinal cortex. J Neurosci. 34(14):4766–4775.

Contreras JM, Banaji MR, Mitchell JP. 2012. Dissociable neural correlates of stereotypes and other forms of semantic knowledge. Soc Cogn Affect Neurosci. 7(7):764–770.

Contreras JM, Schirmer J, Banaji MR, Mitchell JP. 2013. Common brain regions with distinct patterns of neural responses during mentalizing about groups and individuals. J Cogn Neurosci. 25(9):1406–1417.

Cox RW. 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res. 29(3):162–173.

Devereux BJ, Clarke A, Marouchos A, Tyler LK. 2013. Representational similarity analysis reveals commonalities and differences in the semantic processing of words and objects. J Neurosci. 33(48):18906–18916.

Dodell-Feder D, Koster-Hale J, Bedny M, Saxe R. 2011. fMRI item analysis in a theory of mind task. NeuroImage. 55:705–712.

Fairhall S, Anzellotti S, Ubaldi S, Caramazza A. 2013. Person- and place-selective neural substrates for entity-specific semantic access. Cereb Cortex. 24(7):1687–1696.

Fairhall S, Caramazza A. 2013. Brain regions that represent a modal conceptual knowledge. J Neurosci. 33(25):10552–10558.

Fedorenko E, Hsieh P-J, Nieto Castanon A, Whitfield-Gabrieli S, Kanwisher N. 2010. A new method for fMRI investigations of language: defining ROIs functionally in individual subjects. J Neurophysiol. 104:1177–1194.

Fischl B, Sereno MI, Tootell RB, Dale AM. 1999. High-resolution intersubject averaging and a coordinate system for the cortical surface. Hum Brain Mapp. 8(4):272–284.

Freedman DJ, Riesenhuber M, Poggio T, Miller EK. 2001. Categorical representation of visual stimuli in the primate prefrontal cortex. Science. 291:312–316.

Hassabis D, Spreng RN, Rusu AA, Robbins CA, Mar RA, Schacter DL. 2014. Imagine all the people: how the brain creates and uses personality models to predict behavior. Cereb Cortex. 24:1979–1987.

Haxby JV, Gobbini MI, Furey ML. 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science. 293:2425–2430.

Kelemen D, Carey S. 2007. The essence of artifacts: developing the design stance. In: Laurence S, Margolis E, editors. Creations of the mind: artifacts and their representation. Oxford: Oxford University Press. p. 415–449.

Konkle T, Caramazza A. 2013. Tripartite organization of the ventral stream by animacy and object size. J Neurosci. 33(25):10235–10242.

Koster-Hale J, Saxe R. 2013. Functional neuroimaging of theory of mind. In: Baron-Cohen S, Lombardo M, Tager-Flusberg H, editors. Understanding other minds: perspectives from developmental social neuroscience. 3rd ed. Oxford: Oxford University Press. p. 132–163.

Kriegeskorte N, Mur M, Ruff DA, Kiani R, Bodurka J, Esteky H, Bandettini PA. 2008. Matching categorical object representations in inferior temporal cortex of man and monkey. Neuron. 60:1126–1141.

Leshinskaya A, Caramazza A. 2015. Abstract categories of functions in anterior parietal lobe. Neuropsychologia. 76: 27–40.

Leshinskaya A, Caramazza A. 2016. For a cognitive neuroscience of concepts: moving beyond the grounding issue. Psychon Bull Rev. 23(4):991–1001.

Lombardo MV, Chakrabarti B, Bullmore ET, Wheelwright SJ, Sadek SA, Suckling J, Baron-Cohen S. 2010. Shared neural circuits for mentalizing about the self and others. J Cogn Neurosci. 22(7):1623–1635.

Lombrozo T, Kelemen D, Zaitchik D. 2007. Inferring design: evidence of a preference for teleological explanations in patients with Alzheimer's disease. Psychol Sci. 18(11):999–1007.

Mahon BZ, Caramazza A. 2008. A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. J Physiol Paris. 102:59–70.

Mahon BZ, Caramazza A. 2011. What drives the organization of object knowledge in the brain? Trends Cogn Sci. 15(3):97–103.

Martin A. 2007. The representation of object concepts in the brain. Annu Rev Psychol. 58:25–45.

Martin A, Chao LL. 2001. Semantic memory and the brain: structure and process. Curr Opin Neurobiol. 11:194–201.

Martin A, Wiggs CL, Ungerleider LG, Haxby JV. 1996. Neural Correlates of Category Specific Knowledge. Nature. 379(15 February):649–652.

Mason RA, Just MA. 2016. Neural representations of physics concepts. Psychol Sci. 25:1–10.

Miceli G, Fouch E, Capasso R, Shelton JR, Tomaiuolo F, Caramazza A. 2001. The dissociation of color from form and function knowledge. Nat Neurosci. 4(6):662–667.

Mitchell JP, Banaji MR, Macrae CN. 2005. General and specific contributions of the medial prefrontal cortex to knowledge about mental states. NeuroImage. 28(4):757–762.

Mitchell JP, Heatherton TF, Macrae CN. 2002. Distinct neural systems subserve person and object knowledge. Proc Natl Acad Sci USA. 99(23):15238–15243.

Moran JM, Lee SM, Gabrieli JDE. 2011. Dissociable neural systems supporting knowledge about human character and appearance in ourselves and others. J Cogn Neurosci. 23(9): 2222–2230.

Mur M, Bandettini PA, Kriegeskorte N. 2009. Revealing representational content with pattern-information fMRI: an introductory guide. Soc Cogn Affect Neurosci. 4:101–109.

Ochipa C, Rothi LJG, Heilman KM. 1989. Ideational apraxia: a deficit in tool selection and use. Ann Neurol. 25:190–193.

Oosterhof NN, Wiestler T, Downing PE, Diedrichsen J. 2011. A comparison of volume-based and surface-based multivoxel pattern analysis. NeuroImage. 56(2):593–600.

Oosterhof NN, Wiggett AJ, Diedrichsen J, Tipper SP, Downing PE. 2010. Surface-based information mapping reveals crossmodal vision-action representations in human parietal and occipitotemporal cortex. J Neurophysiol. 104(2):1077–1089.

Peelen MV, Caramazza A. 2012. Conceptual object representations in human anterior temporal cortex. J Neurosci. 32(45): 15728–15736.

Ross LA, Olson IR. 2010. Social cognition and the anterior temporal lobes. NeuroImage. 49(4):3452–3462.

Roy JE, Riesenhuber M, Poggio T, Miller EK. 2010. Prefrontal cortex activity during flexible categorization. J Neurosci. 30(25): 8519–8528.

Rumiati RI, Carnaghi A, Improta E, Diez AL, Silveri MC. 2014. Social groups have a representation of their own: clues from neuropsychology. Cogn Neurosci. 5(2):85–96.

Saxe R, Brett M, Kanwisher N. 2006. Divide and conquer: a defense of functional localizers. NeuroImage. 30(4): 1088–1096.

Saxe R, Wexler A. 2005. Making sense of another mind: the role of the right temporo-parietal junction. Neuropsychologia. 43(10):1391–1399.

Simanova I, van Gerven MAJ, Oostenveld R, Hagoort P. 2013. Predicting the semantic category of internally generated words from neuromagnetic recordings. J Cogn Neurosci. 27(1):35–45.

Skerry AE, Saxe R. 2014. A common neural code for perceived and inferred emotion. J Neurosci. 34(48):15997–16008.

Skerry AE, Saxe R. 2015. Neural representations of emotion are organized around abstract event features. Curr Biol. 25(15): 1945–1954.

Skipper LM, Ross LA, Olson IR. 2011. Sensory and semantic category subdivisions within the anterior temporal lobes. Neuropsychologia. 49(12):3419–3429.

Skipper-Kallal LM, Mirman D, Olson IR. 2015. Converging evidence from fMRI and aphasia that the left temporoparietal cortex has an essential role in representing abstract semantic knowledge. Cortex. 69:104–120.

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(6):1112–1123.

The MathWorks Inc. 2009. MATLAB release 2009b. Natick, MA: The MathWorks, Inc.

Thompson-Schill SL. 2003. Neuroimaging studies of semantic memory: inferring "how" from "where". Neuropsychologia. 41:280–292.

Vandenbulcke M, Peeters R, Fannes K, Vandenberghe R. 2006. Knowledge of visual attributes in the right hemisphere. Nat Neurosci. 9(7):964–970.

Warrington EK, Shallice T. 1984. Category-specific semantic impairments. Brain. 107:829–854.

Wilson-Mendelhall C, Simmons K, Martin A, Barsalou LW. 2013. Contextual processing of abstract concepts reveals neural representations of nonlinguistic semantic content. J Cogn Neurosci. 25(6):920–935.

Zahn R, Moll J, Krueger F, Huey ED, Garrido G, Grafman J. 2007. Social concepts are represented in the superior anterior temporal cortex. Proc Natl Acad Sci USA. 104(15): 6430–6435.