NeuroImage 183 (2018) 99–111

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Putting the pieces together: Generating a novel representational space through deductive reasoning

Katherine L. Alfred [a,*], Andrew C. Connolly [b], David J.M. Kraemer [a]

[a] Department of Psychological and Brain Sciences, Dartmouth College, 6207 Moore Hall, Hanover, NH, 03755, USA
[b] Geisel School of Medicine at Dartmouth, 582K01 Borwell, DHMC, NH, 03756, Lebanon, USA

ARTICLE INFO

Keywords:
Transitive reasoning
Representational similarity analysis
Intraparietal sulcus

ABSTRACT

How does the brain represent a newly-learned mental model? Representational similarity analysis (RSA) has revealed the neural basis of common representational spaces learned early in development, such as categories of natural kinds. This study uses RSA to examine the neural implementation of a newly-learned mental model–i.e., a representational space created through deductive reasoning–and study the structure of previously found parietal activity in reasoning tasks. Specifically, all the information in this mental model could only be obtained through abstract transitive reasoning, as there were no predictive differences between observable features in the stimuli, and stimuli were counterbalanced across participants. Participants were shown unfamiliar face portraits paired with names and asked to learn about the height of each person pictured in the portraits through comparison to other individuals in the set. Participants learned the relative heights only of adjacent pairs in the set and then used transitive reasoning to generate a linear ranking of heights (e.g., "Matthew is taller than Thomas; Thomas is taller than Andrew; therefore Matthew is taller than Andrew"). During fMRI, participants recalled the approximate height of each individual based on these inferences. Using a predictive model based on the relative heights of the set of individuals, RSA revealed three brain regions in the right hemisphere that encoded this newly-learned representational space, located within the intraparietal sulcus, precuneus, and inferior frontal gyrus. These findings demonstrate the value of RSA for analyzing structure within patterns of activity and support theories asserting that logical reasoning recruits spatial processing mechanisms.

Suppose you are meeting a coworker's family at a dinner party. You learn that her daughter Samantha is older than her son Max. Later, she tells you that her son Rob, who was not able to make it to dinner, is older than Samantha. You would be able to conclude that because Rob is older than Samantha, he is also older than Max. A cognitive system that supports this kind of inferential reasoning must accommodate incoming information, integrate that information with pre-existing knowledge, and reject conclusions that are incorrect or unhelpful.

Given the highly varied contexts that require inferential reasoning, how does the human mind support this kind of cognition? Mental model theory (Johnson-Laird, 1983, 2010; Knauff et al., 2002) proposes a flexible and robust system to support reasoning processes. At its core, a mental model is an organized representation of the elements of a problem. While reasoning, one first constructs a mental model that represents a true (or at least plausible) state of affairs based on the given premises, and then one inspects this model to evaluate a given conclusion.

How does the brain implement the representation of these models? Both the parietal cortex and the anterior prefrontal cortex are implicated in the neural instantiation of these mental models (for comprehensive review of this literature, please refer to Wendelken, 2015; Krawczyk, 2012; and Vendetti and Bunge, 2014). Increased recruitment of the right dorsolateral prefrontal cortex as well as the superior parietal cortex was found to be the key factor in improvements in adolescents' abilities to work with object representations and mental models (Crone et al., 2006). Most related to this study, Wendelken and Bunge (2010) used a transitive reasoning task where participants viewed pictures that indicated the relative weight of balls, and used that to determine whether a picture at the bottom was accurate or inaccurate. While encoding of the relationships of relative weights was associated with increased hippocampal activity as well as bilateral parietal cortex (both superior and inferior parietal lobule), relational integration itself was associated with increased activity in the rostrolateral prefrontal cortex.

These studies point to the involvement of parietal and prefrontal cortex in the process of abstract relational reasoning, but they do not

* Corresponding author.
E-mail address: katherinealfred.dartmouth@gmail.com (K.L. Alfred).

https://doi.org/10.1016/j.neuroimage.2018.07.062
Received 21 November 2017; Received in revised form 25 July 2018; Accepted 27 July 2018
Available online 4 August 2018

separate the reasoning process itself from the creation and representation of a mental model. In this study, we examine the querying of information from a novel mental model that had been created prior to the collection of neural data. Further, the previous work on abstract reasoning has focused on mean activity differences in specific neural regions. This leaves an open question: does this activity reflect the process of reasoning, the representation of the mental model, or both? Lastly, the previous work has used reasoning problems for which the relevant information is directly observable (e.g. reasoning about images depicting comparative weights; Wendelken and Bunge, 2010) or which draw on information from long term knowledge (e.g. analogies; Green et al., 2006). Therefore, previous work cannot address the question of whether current imaging tools can reveal the representational structure of a mental model built based on newly-learned information. Critically, we test this question using information that cannot be perceived directly, but for which relevant relationships must be inferred. Therefore, this task is specifically designed to investigate the neural implementation of a specific mental model by separating the transitive reasoning process from the unique representation of a newly-created mental model constructed through abstract relational integration.

To achieve this goal, we use neural decoding techniques relying on multivariate pattern analysis (MVPA; Haxby et al., 2001) to compare the patterns of theoretical constructs, such as mental models, with observed patterns of neural activity. The MVPA approach referred to as representational similarity analysis (RSA; Kriegeskorte et al., 2008) is designed to model the neural response to a set of stimuli as a function of a chosen dimension of similarity among the items in the set. Importantly, RSA can assess a representational space based on a dimension of similarity that is not directly observable, such as relative predacity of animals across taxonomical boundaries (Connolly et al., 2016).

Here, participants are shown unfamiliar face portraits paired with names and asked to learn about the height of each person pictured in the portraits through comparison to other individuals in the set. Participants learned the relative heights only of adjacent pairs in the set and then used transitive reasoning to generate a linear ranking of heights (e.g., "Matthew is taller than Thomas; Thomas is taller than Andrew; therefore Matthew is taller than Andrew"). During fMRI, participants recalled the approximate height of each individual based on these inferences. Participants are not given any visual information that could be indicative of exemplar height, and different participants learned different rank orderings of individual portraits, eliminating the possibility that certain faces could look like they belonged to a taller person. Critically, the representational model used in RSA (a dissimilarity matrix based on pairwise item comparisons) is isomorphic to the mental model itself. Both structures are used to represent relationships between abstract concepts that vary along a specified dimension. In this study, the neural implementation of the mental model of relative heights is directly queried using the dissimilarity matrix in RSA to represent the exact relationship between each pair of items in the stimulus set. In this way, the neural activity that is substantially isomorphic to these representations is shown to correlate significantly with this model. Localizing the representation (or representations) of this information informs theories of mental models and sheds light on how disparate networks of brain regions support abstract reasoning.

In particular, if it is the case that mental models are constructed to support transitive reasoning (e.g., when one is presented with linear syllogisms; Johnson-Laird, 2010), and these models are represented using spatial processing regions of the brain (Knauff et al., 2003), then we predict that RSA will reveal activity patterns in superior parietal regions (e.g. precuneus and intraparietal sulcus; IPS) that correlate significantly with these mental models. Whereas previous studies have seen precuneus and IPS activation during deductive reasoning tasks (Knauff et al., 2003; Goel and Dolan, 2001; Ruff et al., 2003), Neurosynth, which collects term-based meta-analyses (www.neurosynth.org; Yarkoni et al., 2011), reveals that studies that have been associated with the precuneus also cover varied topics such as viewing social interactions, schizophrenia, solving analogies, and self-attribution. Thus, examining mean activity levels in the IPS – broadly defined – does not specifically implicate this activity in the representation of a given mental model. In contrast, we posit that if participants are creating a mental model during the pre-scan session, then it is possible that RSA can reconstruct the precise structure of this model by decoding neural activity.

RSA is uniquely well suited to answer the question about the structure and format of representational spaces. At its core, the dissimilarity space model used in the analysis is essentially analogous to the structure of a mental model. By using searchlight representational similarity analysis, the pattern of neural activity in each region of the brain is compared to a specific pre-defined model – in this case, the model of the heights relative to each other. This allows us to determine not only which brain regions are broadly associated with this type of task, but also what regions represent this exact pattern of information. Ultimately, this will help us understand both the structure and format used by the brain to represent mental models created through transitive reasoning.

1. Method

1.1. Participants

Twenty-seven participants (14 female, mean age = 20) undergraduate and graduate students, who were right-handed native English speakers with normal or corrected to normal vision took part in this study. None of the participants had any history of neurological or psychiatric disorders. All participants were compensated with a choice of cash or course credit for their participation, in accordance with the Dartmouth Committee for the Protection of Human Subjects.

1.2. Materials and methods

1.2.1. Transitive reasoning task

Participants were trained on the relative heights of twelve fictitious "people." Each person exemplar consisted of a picture–each a male face with closed mouth, and a neutral expression (NimStim; Tottenham et al., 2009), associated with a name taken from a normed list of the most popular two-syllable names from the 1990s (https://www.ssa.gov). Participants were presented with statements that took the form, "[Person A] is [taller/shorter] than [Person B]." The statements were counterbalanced such that participants were asked to either identify exemplars associated with *taller than* or *shorter than* spatial relationships. For example, participants saw that "Thomas is taller than William" as well as "William is shorter than Thomas". Participants were told to use these statements to try to determine the relative height of each person in the group. The mappings of height were consistent between trials and sessions, so that each participant could reason about a stable hierarchy of heights, however the mappings of height to person exemplar were counterbalanced across participants. That means that a participant in Group A had a separate person who was learned as the tallest than a participant in Group B. Though the groups had separate randomized mappings of pictures to height rankings, the groups were analyzed as one group based on the rank order. Because this was a transitive reasoning task, participants were only presented with adjacent pairs of people, and had to reason about the entire hierarchy through their knowledge of other relationships. In order to know that A is taller than C, the participant must reason that A is taller than B, and B is taller than C, so A is taller than C. Participants were shown each pairwise relationship in each direction (11 possible direct connections between 12 exemplars, total of 22 possible statements) a total of 4 times per reasoning task training session.

1.2.2. Height probe

Participants were shown a black and white "police-style" lineup including silhouettes of three individuals of different heights for five seconds (see Fig. 1A). The participants were instructed to think about where in the lineup the following person would be; i.e., whether that

person was a tall, average, or short exemplar. The lineup then disappeared and the name and picture of one of the exemplars appeared for five seconds while the participant thought about the relative height of the currently viewed exemplar. The name and picture then disappeared, and the silhouette lineup reappeared with the numbers 1–3 randomly assigned to each silhouette superimposed in red font. Participants were instructed to press the number that corresponded with the silhouette that best matched the height of the exemplar they had just seen. Accuracy to this question was not assessed, as it was designed only to encourage participants to retrieve knowledge of the height of the given exemplar and to think about each individually presented exemplar's position in the ordered line of heights, and not to directly compare the individual

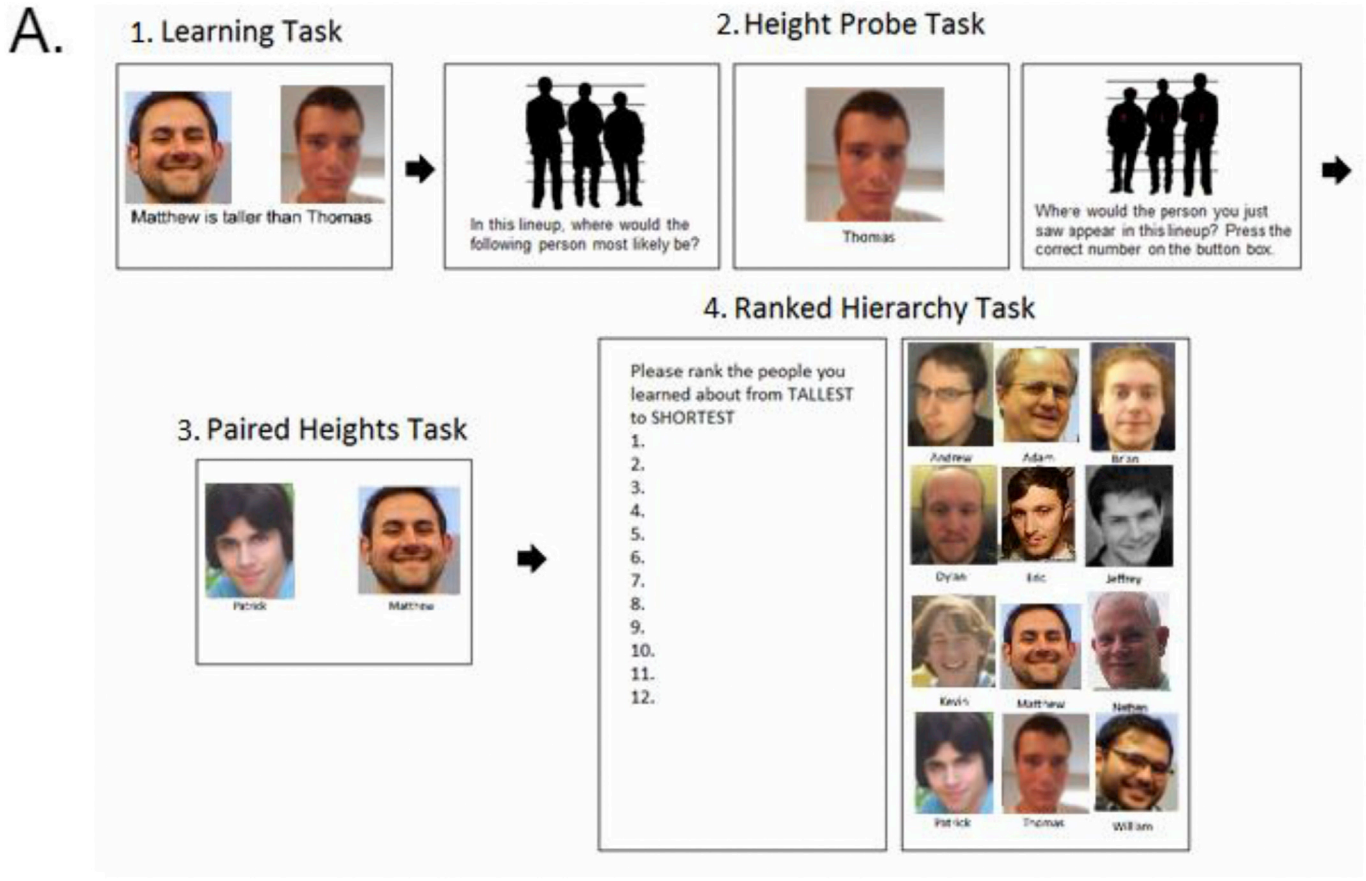

Fig. 1. Overview of the experimental design. *A*. The experimental design of the training session (completed twice on two separate days). Participants were first trained on the relative heights of the items, probed about the height of each person relative to the group, then asked to compare each pair of items in the study (including pairs never learned together). Finally, participants were asked to rank the people from tallest to shortest, with an alphabetized sheet of the items learned about (to probe the hierarchy, not simply memory for the items). NimStim faces have been replaced with lab members and friends. *B*. The fMRI procedure for the critical task. The window of time used in the later representational similarity analysis is during the 5s window when the participant was shown the face of the person for whom they were attempting to recall the height. During this time, participants were unable to prepare a motor response. Participants also completed the Paired Heights task in the scanner identically to how it was performed in training. The Ranked Hierarchy was completed outside of the scanner afterwards.

stimulus to any other. Participants were not given any feedback based on their responses, and due to the nature of the question, accuracy on this task cannot be analyzed.

1.2.3. Pairwise comparisons

Participants were instructed that they would see two of the exemplars they learned about presented together. Unlike during the transitive reasoning task described above, every pairwise comparison of exemplars appeared during this task. Participants pressed one button to indicate that the exemplar on the left was taller, and another button to indicate that exemplar on the right was taller. The presentation of particular exemplars on left or right side of screen was counterbalanced.

1.2.4. Hierarchy reconstruction

Participants were given a blank 1–12 numbered list and instructed to write the names of each of the people in order from tallest to shortest (referred to as the "Ranked Hierarchy" task). To aid participants that were largely relying on faces and not names, an alphabetized (and not height ordered) guide of face/name mappings was provided for participants to reference while making their decisions.

1.2.5. Experiment overview

Each of the training sessions was a half hour in length. First, participants completed the transitive reasoning task, and then completed the height probe, next they completed the pairwise comparisons, and lastly filled out the hierarchy on paper. The training sessions were required to be at least 24 h apart and no more than 48 h apart. The second training session had to be more than 24 h and less than 48 h before the scanning session.

1.3. Experimental design and statistical analysis

1.3.1. Training sessions

Each participant underwent two separate behavioral training sessions to be familiarized with the task and become familiar with the hierarchy that they were reasoning about. Participants were first trained on adjacent exemplars (or "people") in the hierarchy. They were shown that one person is taller than or shorter than another person, and were reminded that "taller than" means "immediately taller than" and not just generally taller than that other person to indicate that there is a specific height order that can be deduced. Participants were instructed to pay close attention to how tall each person is relative to the rest of the group. Participants were also told that they should do their best with the information they have, and that they can change their placement of people in the hierarchy as they learn more about the people in the lineup. Each exemplar was presented with both the exemplar taller than and shorter than it, and this was repeated four times over the session. After the initial learning session, participants were given a test with a blank police-style line up of silhouettes with one that is tall, one that is average, and one that is short, and were asked, "Where in this lineup would the following person most likely be?" The lineup was then removed, and one of the exemplars was shown alone on the screen as the participant thought about the height of that item. The exemplar was then removed from the screen, and the lineup was presented again, each silhouette numbered for a key response from the participant. The numbered silhouettes did not appear until after the presentation of the exemplar being considered in order to separate thoughts about the height of that item from preparing a motor response. Participants were then given the Paired Heights task, where they were presented with two exemplars, and asked to press F if the person on the left is taller or J if the person on the right is taller. This task compared every possible combination of exemplars, including exemplars that were never learned together, and whose relative heights could only have been deduced through transitive reasoning. Lastly, participants were given a blank list numbered 1 through 12 and were asked to write the names of the exemplars in order from tallest to shortest ("Ranked Hierarchy" task). Participants were provided with a reference sheet that contained an alphabetized list of the exemplars (with no height order information preserved). A diagram of the training sessions can be seen in Fig. 1A.

1.3.2. MRI scan session overview

The MRI scanning session lasted one hour and had the same basic structure as training sessions. Participants reviewed the transitive reasoning task during anatomical scans. The height probe and pairwise comparison tasks were completed during functional runs (Fig. 1B). During the scan session, the height probe task was completed twice in order to maximize data points of participants thinking about the height of each exemplar. The full hierarchy was completed again on paper after the completion of the scanner study.

1.3.3. MRI scanning parameters

All MRI scans took place at the Dartmouth Brain Imaging Center. The scanner used to obtain the imaging data was a Phillips 3 T Achieva Intera with a 32 channel SENSE head coil. For the functional runs, there were two runs of 176 vol per run for a total of 352 functional (T2*) volumes with a TR of 2.5s. The functional scans were a gradient-echo EPI with 42 transverse slices at 3 mm per slice. TE was 35, flip angle was 90°. The scan acquisition order was Philips interleaved.

1.3.4. Height probe univariate functional imaging analysis

This univariate analysis for the neural data from the Height Probe task was conducted to obtain item level beta values for the representational similarity analysis (discussed below). Neural data was preprocessed with FSL tools for motion correction and registration (Jenkinson et al., 2002). Each participant's neural data set was modeled using a canonical HRF was used (6 s to peak), and were smoothed using a 5 mm FWHM Gaussian kernel. Beta values were obtained through the contrast between each exemplar item and unmodeled baseline. Neural responses for each of the exemplars were modeled, as well as the button response periods after the stimulus presentation window. Neural activity related to the preparation and execution of button presses was separated from the participant's thoughts about the height of the pictured person, as described above. The portion of the trial in which motor preparation and execution was separated to avoid activity confounded by movements or movement preparation. Regressor covariance estimates generated by FSL confirmed that these portions of the trial were statistically separable due to the jittered fixation periods inserted in between sections of each trial. Anatomical data for the searchlight portion of the analysis were prepared using FreeSurfer (Fischl, 2012).

1.3.5. Height probe time course univariate analysis

Following the registration and preprocessing steps described above, a second univariate analysis was conducted to examine the timecourse of activity over the successive sections of the trial during the main task. Each trial was split into four explanatory variables: one made of a variable amount (5 s–12.5 s) of the intertrial fixation period as well as the entire 5 s in which the initial lineup was presented, one made of the 5 s period of the face and name presentation, one made of the variable duration (5–20 s) fixation cross between person presentation and response, and one made of the 3 s period where participants were presented with the numbered lineup and made their response. Unmodeled portions of the inter-trial were used as baseline. A canonical HRF was used (6 s to peak), and were smoothed using a 5 mm FWHM Gaussian kernel. Beta values for each of the variables were obtained through the contrast with the unmodeled baseline.

1.3.6. Representational similarity searchlight analysis

We used a surface-based searchlight mapping technique (Oosterhof et al., 2011) to produce a whole-brain map for each subject that reflected the Pearson correlation between local neural representational structure and a target similarity structure. The target similarity structure was based on the height relationships between our stimuli. Specifically, a

dissimilarity matrix (see Fig. 4) for the stimuli was created using the rank order of the heights of the people in the study, with the tallest person having 0 dissimilarity to himself, 1 dissimilarity to the next tallest person, and so on. At each searchlight location, the local neural dissimilarity matrix was computed using correlation distance between activity patterns for all pairs of stimuli (132 pairwise distances). Activity patterns were defined by the voxel-wise estimated hemodynamic responses from GLM analysis of the functional data collected during the two height probe sessions. We used a searchlight radius of 8 mm. These analyses were performed using Python and PyMVPA (http://www.pymvpa.org; Hanke et al., 2009), SciPy (http://scipiy.org), and NumPy (http://numpy.scipy.org).

To determine the likelihood that the observed correlations occurred due to chance, we conducted a permutation test to compare our observed results to a distribution of possible results based on a distribution of 10,000 random permutations of the target labels. The probabilities associated with our results were thus calculated as the number of times the average correlation at a given searchlight across subjects for permuted observations exceeded the actual observed average correlation, divided by 10,000.

2. Results

2.1. Behavioral performance

Ranked Hierarchy: A Spearman correlation was calculated to determine the correspondence accuracy between each participant's generated ranked list and the correct ranked list of heights. In the first session, participants were averaging $r_s(25) = 0.39$ with the actual order. At the end of the second training session, participants had median $r_s(25) = 0.98$, with a mean $r_s(25) = 0.77$, and a standard deviation of $r_s(25) = 0.36$. By the scanner session, participants had a median $r_s(25) = 1$, with a mean of $r_s(25) = 0.82$, and a standard deviation of $r_s(25) = 0.28$ (Figs. 2A–3). This indicates that by the time that neural data was collected, participants had learned the hierarchy to ceiling.

Paired Heights: Similar to the behavioral results from the ranked hierarchy task, participants performed well on the paired heights task. When judging which person was taller from the set of all possible combinations, participants performed close to chance during training sessions, and improved to just below 90% on average at the scan session (see Figs. 2B–3). In some cases, participants failed to respond in the 3s window allowed for responses, and those trials were removed (mean = 4.16; median = 1; stdev = 7.37; min = 0; max = 30). With the pairwise comparisons, participants need to compare not only people who were never presented together, but people who may be one or two ranks away. Distances of that magnitude may be difficult, because those people were never presented in conjunction with each other, they have similar heights, and it may take a significant amount of time to reason about who is taller (this is the likely source of the large number of errors of omission). Regardless, participants who had time to formulate their response largely responded very accurately.

In addition, analysis of the Paired Heights behavioral data from the scanner session found that judgments of pairs who were closer together in distance were more difficult than pairs that were further apart. Distances between exemplars were calculated such that there was one distance between an exemplar and its adjacent partner, 2 for the exemplar on the other side of its partner, and so forth. Because of this, there were a large number of pairs that were 1 distance apart (11) and only one pairing that was 11 distance apart (the tallest and shortest exemplar). In order to create relatively balanced groups, we binned pairs that were 1–2 distance apart and called them "near" (21 pairs). Distances that were 3–5 were "mid" (24 pairs), and 6–11 were called "far" (21 pairs). There was a significant effect of distance on accuracy, [$F(2,1609) = 4.51$, $p < 0.01$, $\eta^2 = 0.006$]. Post hoc testing revealed that this was driven by significantly higher accuracy for pairs that were far apart compared to mid distance pairs ($p < 0.01$), and there was no significant difference between near and mid, nor for near and far. There was also a significant effect of distance on reaction time as well, [$F(2,1602) = 43.49$, $p < 0.001$, $\eta^2 = 0.05$]. Post hoc testing revealed that not only were participants significantly faster to respond to pairs that were far compared to near ($p < 0.001$), participants were significantly faster to respond to pairs with a mid distance than near ($p < 0.01$) and participants were significantly faster to respond to pairs that were far apart than participants with a mid distance ($p < 0.001$). This demonstrates the symbolic distance effect previously found in studies of spatial mental comparisons (Moyer and Bayer, 1976; McNamara, 1986).

2.2. Neural representational similarity analysis

The searchlight RSA resulted in a correlation value for different

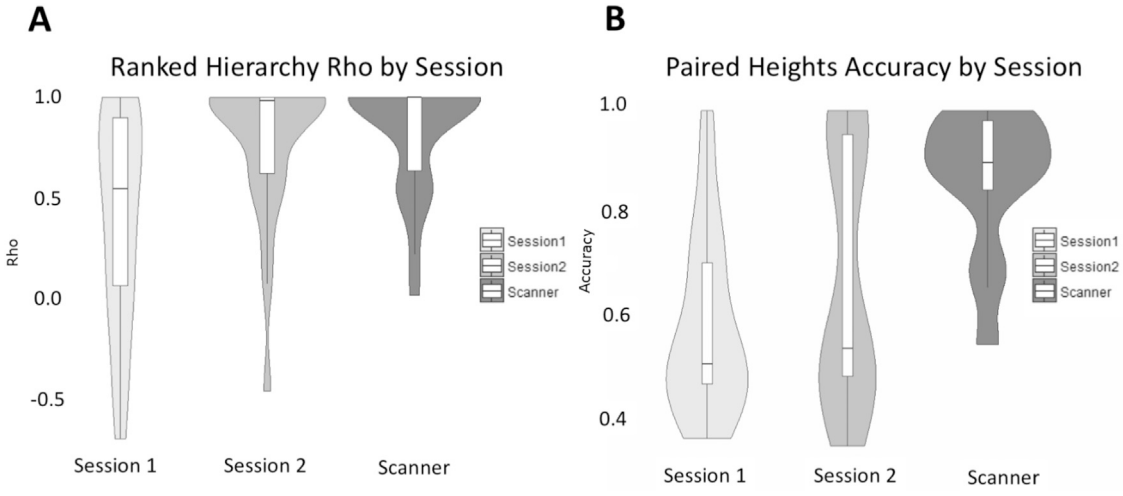

Fig. 2. Behavioral results by session. *A.* Violin plot with overlaid box plot to show learning by session as measured by Spearman's rho correlation between a participant's generated hierarchy on the Ranked Hierarchy task with the actual ranked order. The width along the x-axis of each shaded element in the plot is a histogram indicating the number of participants at the level of accuracy indicated by the y-axis. Whiskers on the box plot indicate the largest and smallest values within 1.5 * IQR. *B.* Violin plot with overlaid box plot to show learning by session as measured by accuracy on the Paired Heights task, where participants determined the taller person of each possible pair. The width along the x-axis of each shaded element in the plot is a histogram indicating the number of participants at the level of accuracy indicated by the y-axis. Whiskers on the box plot indicate the largest and smallest values within 1.5 * IQR.

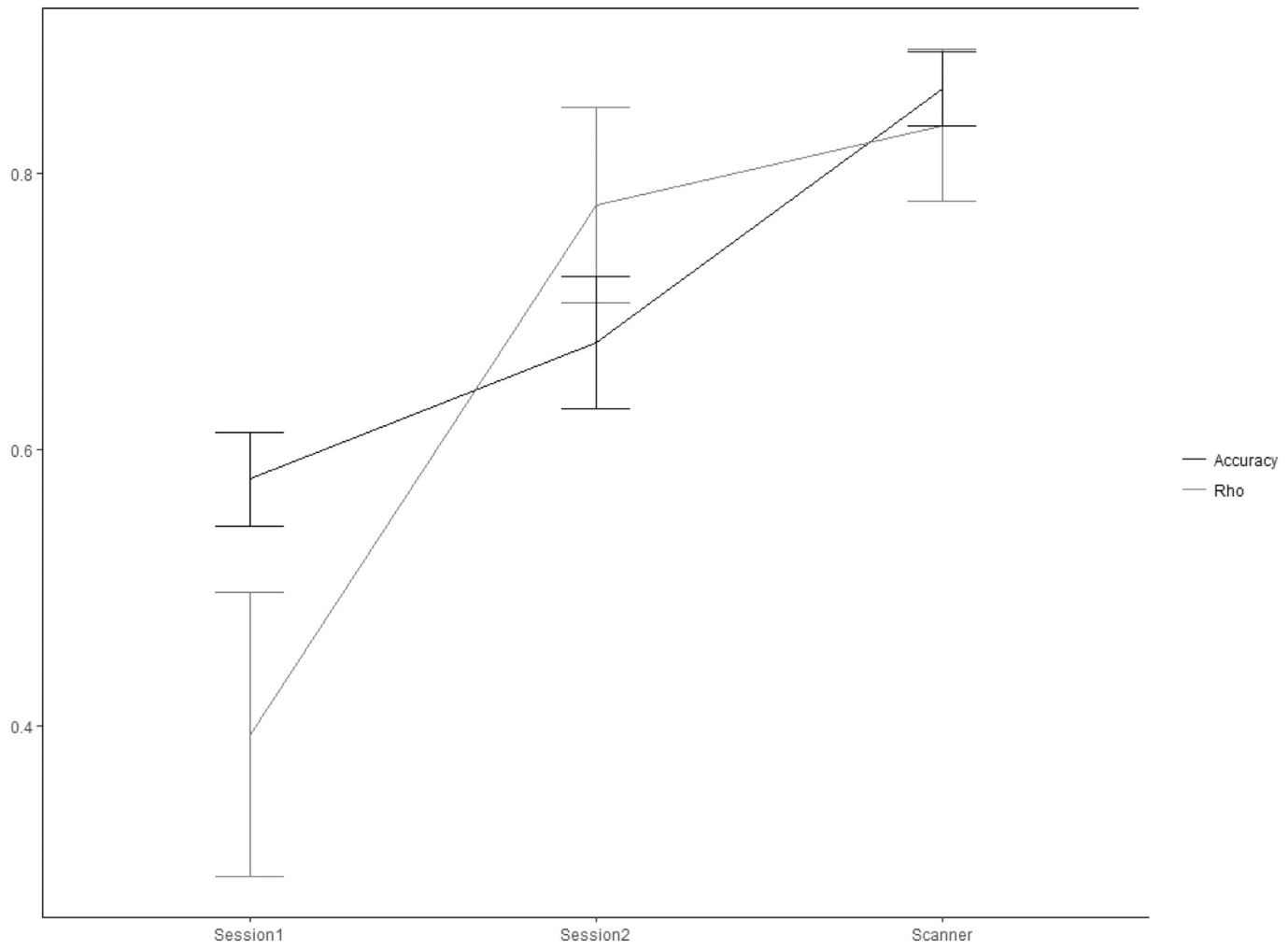

Fig. 3. Line graph of behavioral learning by session. The black line indicates mean accuracy on the Paired Heights task, and they grey line indicates mean Spearman's rho correlation between a participant's generated hierarchy on the Ranked Hierarchy task with the actual ranked order. Error bars indicate standard error.

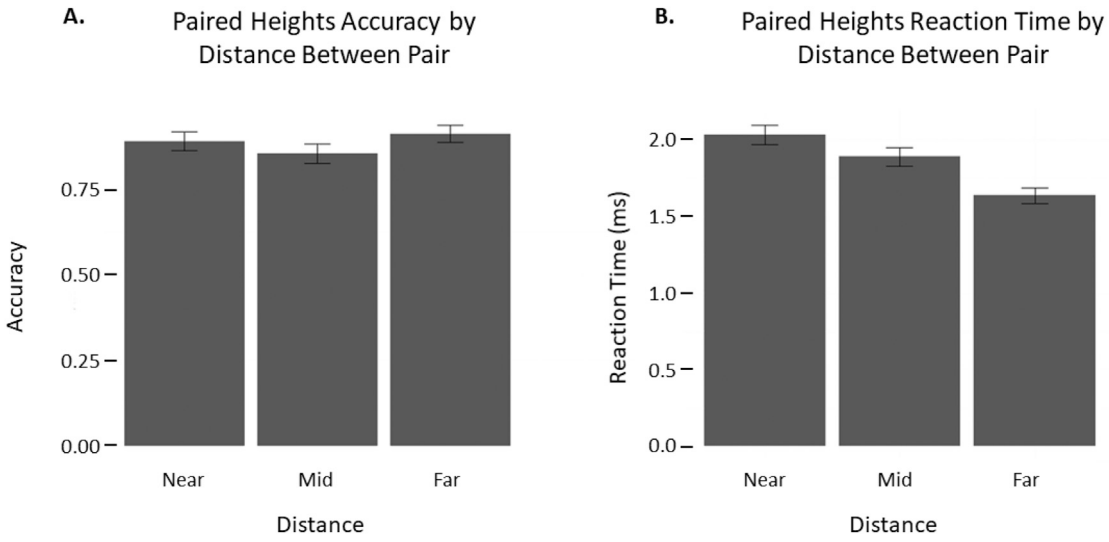

Fig. 4. Paired Heights behavioral results by distance. *A*. Accuracy on the Paired Heights task by binned distance group. "Near" contains pairs that are 1–2 distance apart, "Mid" contains 3–5, and "Far" contains 6–11. Pairs that are in the "Mid" distance bin have significantly lower accuracy than "Far" ($p < 0.01$). *B*. Reaction time on the Paired Heights task (in milliseconds) by binned distance group. Participants were significantly faster with "Far" pairs compared to "Mid" ($p < 0.01$), and were also significantly faster for "Mid" compared to "Near" ($p < 0.001$).

participants based on how closely the pattern of neural activity in that region mirrored the pattern of the mental model given in the dissimilarity matrix for the stimuli (Fig. 5). Following the permutation test we found $p < 0.01$ permutation-corrected probability for these correlations in two predicted regions: IPS and precuneus (both implicated in Knauff et al., 2003). We additionally found activity associated with another region that we had not predicted: the right IFG. These correlations are reported by region in Table 1 and depicted in Fig. 6. All of the RSA analyses were conducted at the individual subject level, and the permutation test and the resulting permuted t-map in Fig. 6 are at the group level.

As was predicted based on the results of Knauff et al. (2003), the model correlated strongly with activity in the IPS and precuneus. Precuneus activity has also been found in previous work to play a role in spatial mental rotation problems (Schlegel et al., 2013). The activity was bilateral, but had a higher correlation with regions in the right hemisphere (median $r_{RH} = 0.29$, median $r_{LH} = 0.24$). A heat map (histogram) was generated by creating spheres with a 1 mm radius around the voxel within the right IPS that had the highest correlation with the model within that subject (Supplementary Fig. 3). These spheres were then overlaid on top of each other to show the area within the IPS that subjects tended to have the strongest correlation with the model. Peak correlations were distributed to a degree, though there were nearby clusters within each region that showed the highest degree of overlap for peak correlations. While there were a high degree of individual differences and little overlap in terms of the location of each participant's whole brain peak RSA correlation (Supplementary Table 1), a whole brain heat map (histogram) was created using the nodes that contain each participant's top 5% (1025 nodes) or permuted t values with the height RSA (Supplementary Fig. 3). Unlike in Fig. 6, these nodes were not restricted to any *a priori* region of interest. The region containing the maximal number of participants' top 5% of nodes was the right parietal lobule, with 21 participants in the right postcentral sulcus 19, subjects in the right supramarginal gyrus, and 18 participants the right IPS (Table 2; full table: Supplementary Table 3).

The IPS and precuneus activity resulting from the searchlight RSA is consistent with the prediction that the mental model is spatial in nature. It is important to note that the correlations between the model and the pattern of neural activity need not be reflective of how active these regions are during the task, but instead reflective of the amount that the model of the height order is similar to the pattern of activity in that region. This means that these parietal regions are not just involved in the process of retrieving information from a mental model, as might be demonstrated by increased mean levels of activity, but more specifically contain the predicted structure of information in the predicted spatial format.

To further support our assertion that the involvement of the right IPS represents spatial information, we used the reverse inference map from Neurosynth with the "spatial" keyword. This map would indicate areas that are selectively active for spatial information (created through meta-analysis of 1157 studies, thresholded at FDR corrected 0.01), and is an objective, external method to generate networks based on keywords. The spatial reverse inference map was binarized and was overlaid on top of the significant permuted t-values ($p < 0.05$), which represent the *t*-test between our data and a random distribution created by scrambling the labels on our data 10,000 times, to correct for multiple comparisons. The significant permuted t-values and the spatial reverse inference map overlapped in the right IPS (Fig. 7A). The cluster of significant *t*-test

Table 1
RSA correlation and coordinates by region.

Region	Centroid Coordinates (X,Y,Z), MNI	Median Correlation	Maximum Correlation
Right IPS	38, −42, 40	0.297	0.581
Right IFG	43, 13, 22	0.304	0.482
Precuneus	11, −51, 36	0.189	0.604
Right RLPFC	42, 40, 4	0.116	0.558

Note: Coordinates for each region were determined by locating the centroid of the cluster of voxels in which each participant had their maximum correlation with the RSA model.

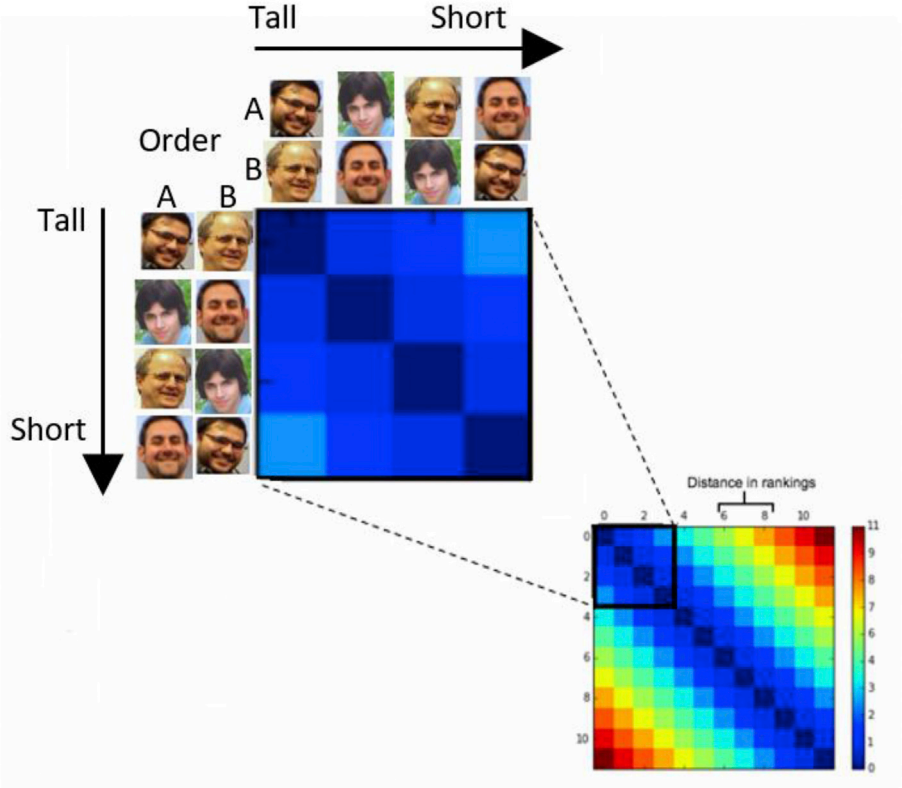

Fig. 5. The representational dissimilarity matrix for the hierarchy of heights. The dissimilarity matrix shows each row and column representing a person from the hierarchy that participants learned. The values represent the number of ranks that each item was away from each other item. For example, the first item had 0 dissimilarity to itself, and 1 from the next ranked item. NimStim faces have been replaced with lab members. The two randomized mappings between height rank order and image were used to control for the possibility of facial similarity confounding the rank ordering.

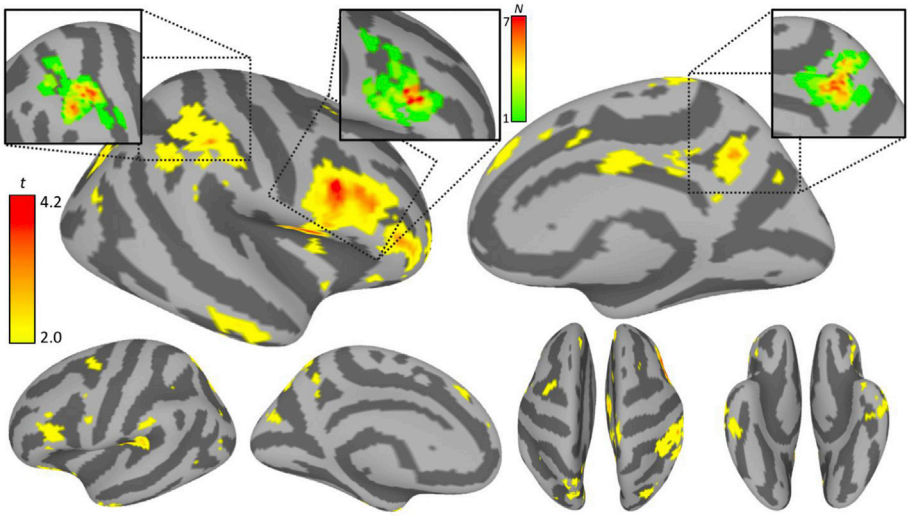

Fig. 6. RSA results for the model of height similarity. Intensity values correspond to permuted t-values, as indicated by the color bar on left. Brain maps in the lower two rows are displayed at a permuted threshold of $\alpha < 0.05$. The insets along the top row present a histogram in which intensity values correspond to the number of subjects for whom a peak correlation between the RSA model and brain activity occur in each voxel, as indicated by the color bar at the top. These maps are displayed (left to right) for the IPS, IFG, and precuneus, respectively.

Table 2
Anatomical locations of overlap from RSA.

Anatomical Region	Number of Participants
Right Postcentral sulcus	21
Right Superior frontal gyrus (F1)	21
Left Lateral occipito-temporal gyrus(fusiform gyrus, O4-T4)	19
Left Paracentral lobule and sulcus	19
Right Supramarginal gyrus	19
Right Intraparietal sulcus (interparietal sulcus) and transverse parietal sulci	18
Right Middle frontal gyrus (F2)	18
Right Superior temporal sulcus (parallel sulcus)	18
Right Unknown	18
Left Parieto-occipital sulcus (or fissure)	17
Left Pericallosal sulcus (S of corpus callosum)	17
Left Sulcus intermedius primus (of Jensen)	17
Left Superior frontal sulcus	17
Left Superior segment of the circular sulcus of the insula	17
Right Inferior frontal sulcus	17
Left Inferior occipital gyrus (O3) and sulcus	16
Left Superior occipital sulcus and transverse occipital sulcus	16
Right Inferior temporal gyrus (T3)	16
Right Opercular part of the inferior frontal gyrus	16
Right Postcentral gyrus	16
Right Precuneus (medial part of P1)	16
Right Superior parietal lobule (lateral part of P1)	16
Left Orbital part of the inferior frontal gyrus	15
Left Superior parietal lobule (lateral part of P1)	15
Right Angular gyrus	15
Right Orbital gyri	15
Right Pericallosal sulcus (S of corpus callosum)	15
Right Planum temporale or temporal plane of the superior temporal gyrus	15

Note: Anatomical location localized using the Destrieux et al. (2010). The top 5% of values were calculated individually for each participant, then the maps were binarized and summed to count the number of participants who had a node in the top 5% of their results in that location.

values in the right IPS was contained within the spatial reverse inference mask, further reinforcing that this area is selectively spatial, and that the correlation with the model in this region indicated that the predicted structure is spatial. This means that not only did the model correlate with patterns of neural activity in the IPS, but also that the region within the IPS where the correlations occur has frequently been previously associated with spatial processing, suggesting that the mental model of that information may be in a spatial format.

In order to confirm that the results of this RSA are due to the learned structure of the mental model and not due to visual information linked to the face/name pairs, we ran a second RSA using an HMAX model of visual similarity (T. Miconi, https://scholar.harvard.edu/tmiconi/pages/code). This HMAX model processes each picture and outputs values from the C2B cells of increasing receptive field size as well as C3 cells (which approximately map onto human inferotemporal cortex; Serre et al., 2007). The dissimilarity matrix was created by calculating the dissimilarity between the output vectors for each pairwise combination of people. The RSA using the HMAX model of visual similarity (Supplementary Fig. 4) primarily correlated with regions in the lateral fusiform gyrus (primarily left lateralized, but bilateral to an extent), ventromedial prefrontal cortex (left), superior temporal sulcus (bilateral), middle occipital gyrus (right), transverse parietal sulcus (right), superior occipital sulcus (right) and middle temporal gyrus (left). Importantly, the right IPS, precuneus, and IFG, which were most highly correlated with the model of height similarity, did not correlate with the HMAX model. This indicates that, as intended, activity in those regions covary with relative heights of exemplars in the height lineup and not with any visual information present in the stimuli.

In addition, the searchlight RSA revealed activity in the right rostrolateral prefrontal cortex (RLPFC; Fig. 6). The RLPFC has been widely implicated in the transitive reasoning process, specifically with relational integration (Christoff et al., 2001, 2003; Crone et al., 2006; Vendetti and Bunge, 2014; Wendelken et al., 2012; Prado et al., 2011; Whitaker, 2012; Bunge et al., 2009). Most relevant to the current study, this region has been previously found to be involved in relational integration (greater than relational encoding) in transitive reasoning tasks (Wendelken and Bunge, 2010; Green et al., 2006, 2010, 2012; Ramnani and Owen, 2004). Similarly, the task in the present study requires relational integration to properly place each of the exemplars in the lineup. The same Wendelken study (2010) noted that hippocampal involvement in relational transitive reasoning tasks (such as seen in Frank et al., 2003 in rats; Heckers et al., 2004; Van Opstal et al., 2008; Van Opstal et al., 2009; Zalesak and Heckers, 2009) is due to demands on relational memory when an inference needs to be made across learned associated pairs. In the present study, the step that would have involved this process occurred outside of the scanner, and while this particular analysis was conducted on the surface, we would not necessarily have predictions for hippocampal activity during this task. While the hippocampus is very important in relational reasoning, this scanner task was designed to test the specific prediction that mental models for transitive reasoning are created and then later queried (Johnson-Laird, 2010). Therefore, the creation of the mental model had to occur separately from querying the model to provide a response in the scanner.

The searchlight RSA also revealed activity in the right inferior frontal

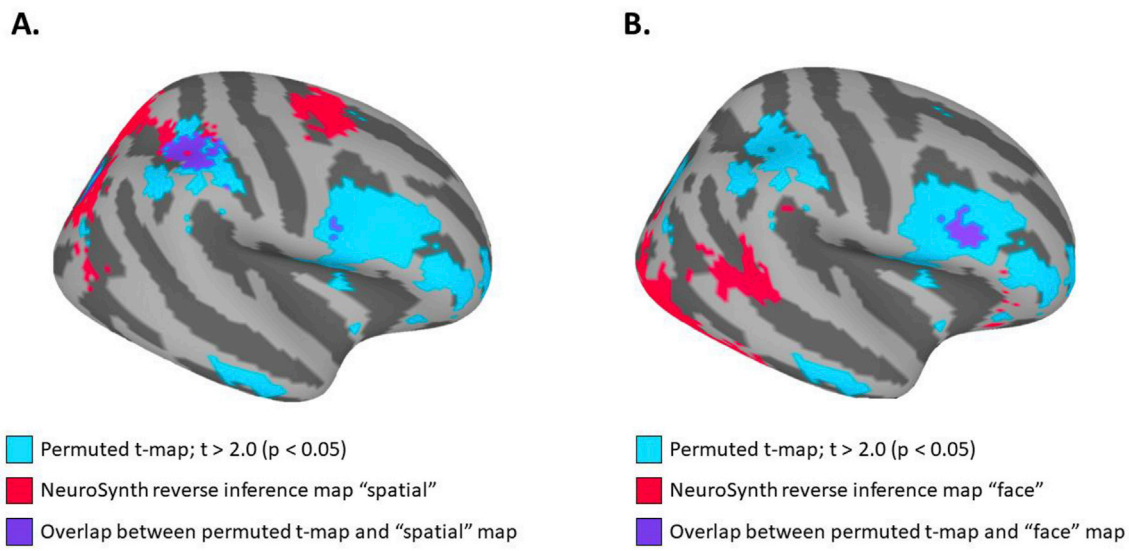

Fig. 7. Representational similarity analysis (RSA) results from the model of height rank compared with the NeuroSynth "spatial" and "face" reverse inference maps. *A.* RSA results for the model of height similarity overlaid by the spatial reverse inference mask from Neurosynth (relative selectivity for spatial information; mask thresholded at FDR corrected 0.01). Permuted t-values were thresholded to show significant permuted t-values above 2.0 ($\alpha < 0.05$), and then binarized to show all nodes where $t > 2.0$. The cluster in the right IPS is both significant and contained within the spatial reverse inference mask, indicating that region performs selectively spatial processing. *B.* RSA results for the model of height similarity masked by the face reverse inference mask from Neurosynth (relative selectivity for faces; mask thresholded at FDR corrected 0.01). Permuted t-values were thresholded to show significant permuted t-values above 2.0 ($\alpha < 0.05$), and then binarized to show all nodes where $t > 2.0$. The outlines indicate boundaries of clusters on the full original map. This indicates that the activity in the right IFG is encoding information about face selection and retrieval during the task.

gyrus (IFG). Although not predicted *a priori*, activity in the right IFG may be due to the retrieval of visual information related to faces or other objects (Kelley et al., 1998; Casasanto, 2003; Pihlajamäki et al., 2003; Wig et al., 2004), in this case, the retrieval of information about the face exemplars encountered in the task. Correlations and centroid coordinates for the voxels of peak participant RSA correlation for each region can be found in Table 1. Similar to our earlier analysis showing the overlap between the NeuroSynth "spatial" map and our permuted t-values in the right IPS, we ran an analysis comparing the NeuroSynth reverse inference map for "face" and our permuted t-map. The "face" map and the permuted t-values in the right IFG overlapped, indicating that the involvement of the right IFG is likely due to face retrieval (Fig. 7B). Though this region of IFG has been associated with studies of face processing, it is also worth noting that it is commonly co-activated with the IPS across numerous other tasks as well (www.neurosynth.org; Yarkoni et al., 2011). While there were additional clusters that reached significance in the RSA, none of the others were as prominent as the right IFG. While they will not be discussed further to avoid excessive *post hoc* explanations of results, the full set of results can be seen in Fig. 6.

2.3. Neural timecourse general linear model

In order to query the timecourse of activity in a set of *a priori* anatomical regions of interest, we examined the univariate response in the same reverse inference maps used in the multivariate analysis. The general linear model for the relative timecourses of activation for regions in the Neurosynth "spatial" reverse inference map (e.g. right IPS, precuneus) and the Neurosynth "face" reverse inference map (e.g. right IFG, ventral occipitotemporal cortex) revealed that the regions in the two masks (Fig. 8A) show different patterns for the level of activity based on the timing within the trial (Fig. 8B). As seen in Fig. 8, during early portions of the trial the face-responsive regions are more active than the spatial regions, and during the later portions of the trial the spatial regions become more active than the face regions. This timecourse of activity is consistent with the hypothesized processing that occurs during the task. In particular, we interpret these results to indicate that the face and name presentation early in the trial cues a recollection of where in the spatial hierarchy of heights that individual stimulus is located, which is the basis for the participant's response in the latter portion of the trial. Further it is worth noting that the spatial regions are active throughout the entire trial, whereas the face regions are less active than baseline during the variable time period between the face presentation and response periods. In this time, the participant is likely thinking about where in the hierarchy the particular stimulus is located, however the participant cannot yet prepare a motor response as the button cues do not appear until the response period. Therefore the fact that the dorsal stream spatial network remains active, as well as the results of the RSA which also implicate this network, indicates that processing of spatial information related to the stimulus occurs during this period even in the absence of the face cue onscreen.

3. Discussion

One purpose of this study was to show that RSA can be used to reliably detect novel representational spaces constructed through transitive reasoning. While RSA is commonly used to examine the structure of representational spaces that are learned over the course of the lifetime (Connolly et al., 2012; Kriegeskorte and Kievit, 2013), this study adds to the growing body of work that shows that RSA has applications in studying representational spaces transiently created to solve a given task (Collin et al., 2015; Garvert et al., 2017; Mack et al., 2016). Even though this study only trained participants for two half-hour sessions, RSA is a sensitive enough technique to examine the structure and uncover the neural instantiation of this representational space.

Moreover, by localizing the neural representations correlating with the putative mental model of ranked heights, these results inform our understanding of the networks of brain regions that coordinate to support using a mental model in the service of abstract reasoning. If the representation was primarily visual in nature (e.g., referring to images of the exemplars), then RSA would have revealed activity in relatively early regions of visual cortex, such as the V2 activity observed by Knauff et al. (2003), or possibly in face-preferential regions of fusiform cortex (Kanwisher et al., 1997; Guntupalli et al., 2016). Instead, the high correlations observed with neural activity in the IPS and IFG indicates a

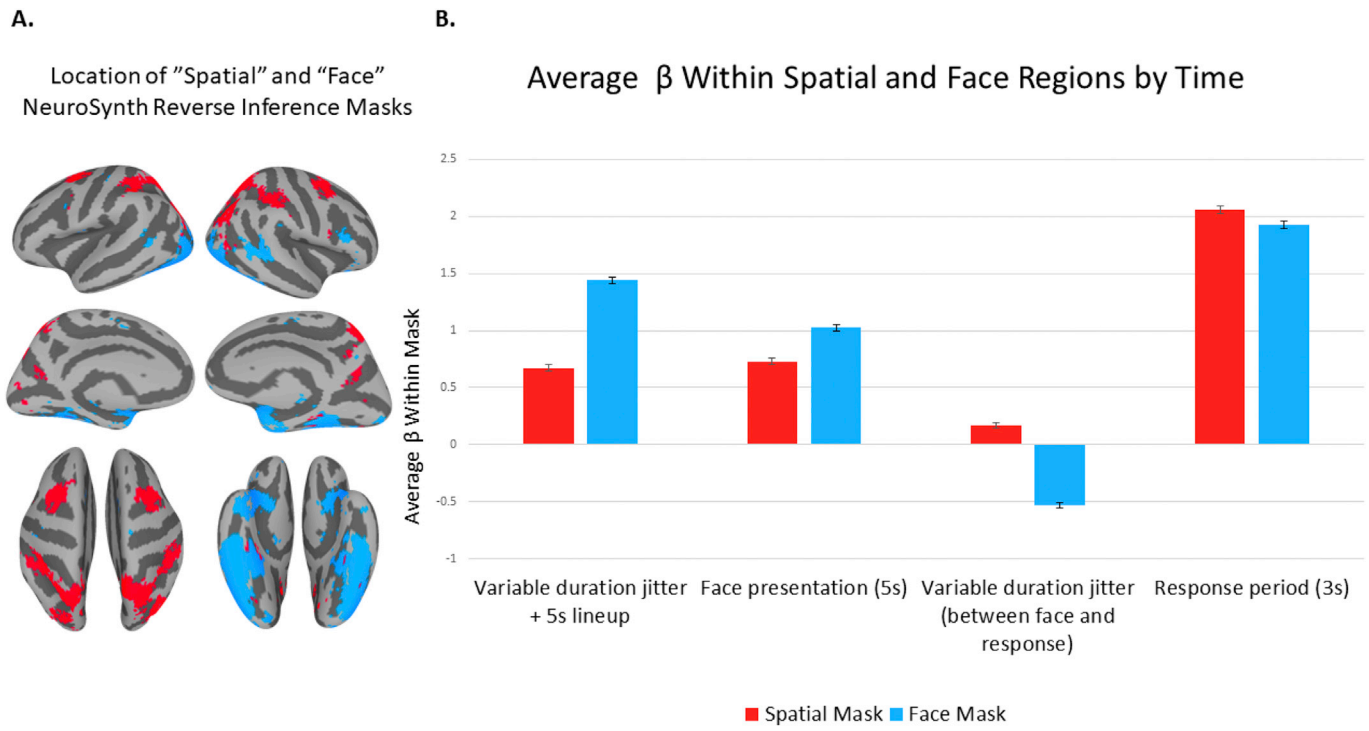

Fig. 8. Results from the height probe timecourse analysis. *A*. Locations of the regions contained within the NeuroSynth "spatial" reverse inference map (red) and the NeuroSynth "face" reverse inference map (blue). *B*. Bar graph of the average β value within each of the two NeuroSynth maps by chronological portion of the height probe trials. Error bars indicate a 95% CI for the average β values.

coordination between brain regions that support spatial transformation (IPS) and those that support the retrieval of face-specific information (right IFG). Reverse inference maps drawn from a large sample of studies (www.neurosynth.org; Yarkoni et al., 2011) confirmed this interpretation. Since there was absolutely no perceptual information with which to make this distinction, this means that the observed structure of the pattern of neural responses is due to the participants' mental model, drawn from inferential reasoning supported by these brain regions.

Another purpose of the study was to use the present results to build on prior theories of mental models. Early literature examining the reasoning process used systematically differing problems to determine what made one reasoning problem more difficult than another. Byrne and Johnson-Laird (1989) used problems that would require one or multiple mental models, or internal representations of the problem space, to be solved and problems that had more or fewer inferential steps. Ultimately, it was the number of mental models that needed to be constructed to solve the problem, not the number of inferential steps that affected the number of correct responses. Later research into the neural instantiation of mental models (Goel, 2007; Goel & Dolan, 2001, 2004; Knauff et al., 2003; Ruff et al., 2003) has shown that both reasoning in general and deductive reasoning in particular, correspond to parietal activation, possibly indicating that mental models are spatial in format. In a study run by Knauff et al. (2003) four separate types of deductive reasoning problems were presented auditorially to participants. Some of the deductive reasoning problems were easy to represent visually and spatially (e.g., above-below), whereas some were easy to represent visually but not spatially (e.g., cleaner-dirtier), and still others were difficult to represent both visually and spatially (e.g., better-worse). Regardless of problem type, they found activation in the precuneus and the right superior parietal gyrus. Problems that were difficult to visualize spatially but were easy to visualize with images additionally showed activity in visual cortex (V2) as well as in the precuneus and right superior parietal gyrus. Problems that elicited V2 activity ultimately had a slower response time, indicating that while images may have been generated to solve the problem, they were not helpful to the process.

The examples above do not necessarily support the strong conclusion that the nature of mental models is spatial. For example, parietal cortex activity does not exclusively correspond to spatial processes, but is also involved in non-spatial cognitive operations, such as the initiation and sustained maintenance of task sets for various types of stimuli and tasks (Dosenbach et al., 2006). Furthermore, even if the parietal activation observed by Knauff et al. (2003) does represent spatial processes, it is unclear whether it represents a complete mental model or evaluation of sets of premises. Is the reasoner constructing an entire scenario that represents the true premises prior to evaluating the conclusion, or does the reader evaluate on the fly each individual premise in relation to the other premise and to the conclusion? Either approach may involve spatial processes. However, the claim of mental model theories (e.g., Johnson-Laird, 2010), is that in the course of reasoning, one first constructs a mental model that represents a true (or at least plausible) state of affairs based on the given premises, and then one inspects this model to evaluate a given conclusion. In order to test this specific claim, the construction and querying of the mental model would need to happen separately. In this case, the representational space was created through transitive reasoning outside of the scanner. However, this does not mean that reasoning only took place outside of the scanner and that participants were simply recalling the exemplars they had learned in the scanner task. The behavioral results from the Paired Heights task (Fig. 4) showed that participants took longer to respond to pairings of exemplars that were near to each other in the space compared to participants who were further away from each other. If participants were simply able to retrieve the correct answer, there is no reason to believe why a comparison between any one pairing of items would be significantly different from any other pairing. However, if participants needed to use reasoning to determine whether the exemplar was taller than another exemplar, we might see an effect like this. If the exemplars are further apart (e.g. separated by 7 other exemplars), participants have more ways that they can validate their conclusion. If the exemplars are near (e.g. separated by only 1 exemplar between them) participants need to make a specific inference which is more difficult to validate (and therefore, takes longer

to respond). Similarly, in the height probe task, participants were performing a more generalized version of this task, where instead of comparing every possible pair of exemplars one by one, participants compared an exemplar to each of the other exemplars in the group in one stage. Therefore, while the initial transitive reasoning inference happened outside of the scanner, the reasoning process was ongoing in the scanner.

The previous studies were only able to draw conclusions about areas with higher mean levels of activity during reasoning processes (Knauff et al., 2003), networks of regions that are implicated in different subprocesses (Goel, 2007), or restricted their analysis to the hippocampus (Collin et al., 2015). In contrast, the present study demonstrates that the theoretically predicted structure of a mental model created to solve a transitive deductive reasoning problem is reflected in patterns of neural activity in the parietal lobe, specifically the precuneus and intraparietal sulcus. This shows that the precuneus and IPS are not simply involved in reasoning, but the pattern of neural activity reflects the mental model generated to solve the problem. The IPS has been historically implicated in magnitude based judgments (Holloway and Ansari, 2010), amodal representation of quantity (Dehaene, 2007), and is activated when quantities are represented in non-symbolic sets of visual/auditory objects or events (e.g. dot arrays or series of auditory tones) as well as in symbolic annotation (numerals or spoken number words; Pinel et al., 2001; Lyons et al., 2015). The association between the activity in the IPS and precuneus and the model of the novel representational space provides evidence in support of the assertion that mental models are spatial. In addition, this result is consistent with the theory that syllogistic reasoning relies on constructing a mental model, which is then consulted when making subsequent inferences. These findings are consistent with mental models theory (Johnson-Laird, 1983, 2010; Knauff et al., 2002; Jahn et al., 2007; Knauff, 2013), in which mental models are first created and then information is queried and retrieved from the model.

One of the benefits of using RSA to approach this problem is that it allows for the examination of the nature and structure of the representational space in a powerful and precise manner. RSA has primarily been used in domains that are learned over the course of the lifetime, such as the biological taxonomy example above. In contrast, mental models (and resulting representational spaces) used in transitive reasoning are constructed *ad hoc* to serve the reasoning process in solving a specific problem, which may only be encountered once. This study additionally showed that RSA has applications for studies that aim to examine the neural representations of newly-learned stimuli, even for stimulus sets with relatively subtle differences between items.

One limitation of the present study is that the relationship used in the transitive reasoning problems (height) is inherently spatial. Therefore, it is possible that despite some evidence that implicates parietal activation regardless of the type of relationship used in the reasoning problem (Knauff et al., 2003), the IPS activity seen here may reflect judgments made specifically about spatial information. One meta-analysis (Prado et al., 2011) found that activation in bilateral posterior parietal cortex (near the IPS) was significantly associated with transitive reasoning, even when the studies that relied on non-linguistic stimuli were excluded. Additionally, Acuna et al. (2002) found that ordering of shapes with arbitrary heights, learned through transitive reasoning, activated bilateral posterior parietal cortex near the IPS more than simply ordering the shapes by height (which was activated significantly less and restricted to the left hemisphere). This would seem to indicate that the involvement of the IPS and PPC more broadly are not inherently dependent on spatial problem presentation or content. However, in order to fully divorce the spatial content of the problems from the spatial nature of the mental model, future studies are necessary to disentangle the reasoning process from the content domain. The results from those studies will allow us to see more clearly whether parietal involvement is due to the structure of the mental model or if the precise parietal location is influenced by the content of the problem being solved.

An additional limitation of this study is that the use of an individual relational reasoning task is unable to fully account for the rich and diverse findings in the field of abstract transitive reasoning. In order to focus on testing a prediction made by mental model theory (i.e., that models are first created and then later queried; Johnson-Laird, 2010), our task was unable to capture the full spectrum of processes that go into supporting transitive inference. Notably, a large body of work has focused on the role of the hippocampus in transitive inference. Frank et al. (2003) argued that the hippocampus, unlike in accounts by Dusek and Eichenbaum (1997) was not flexibly comparing memories, but was more likely involved in shaping associative weights during training. Work from Heckers et al. (2004) showed that the hippocampus was involved specifically in making the transitive inferences, not simply recognizing stimuli presented together. Additionally, a follow up study by Frank et al. (2005) in humans similarly argued that the hippocampus likely plays a limited role in relational reasoning outside of the learning process. Van Opstal and colleagues further found that the hippocampus activity did not correlate with performance on transitive inference tasks, implying a more general role for the hippocampus than playing an active role in transitive reasoning specifically.

Beyond the limitations of our task, transitive reasoning is only one of a vast and diverse set of deductive reasoning tasks, and these results, while informative on the creation of mental models in transitive reasoning, cannot necessarily generalize to tasks with analogical reasoning or propositional reasoning. Even some studies of different types of transitive reasoning (e.g., set inclusion versus linear transitive reasoning) have shown different patterns of results (Prado et al., 2012). While linear transitive reasoning showed the typical inverse effect in the right PPC of decreasing activity with increasing distance, set inclusion showed increasing activity in the left IFG and left PPC for further items. Therefore, based on this study design, we cannot determine if we would see these same results if the participants were forming the mental model in another way.

Analogical reasoning work has found both overlapping and unique areas for those types of problems. Compared to transitive reasoning, analogical reasoning studies show much higher involvement of left PFC, particularly the left anterior PFC and DLPFC (Bunge et al., 2005; Green et al., 2006, 2010; Wharton et al., 2000). However, analogical reasoning studies have also highlighted and further reinforced the role of the RLPFC in relational integration (Bunge et al., 2005; Krawczyk et al., 2010), which has been seen in other types of reasoning tasks, and is not limited to analogical reasoning. Work on propositional reasoning has found some similar regions to those reported in transitive inference tasks. Using MVPA to analyze discriminability between different logical connectors revealed that a region in the anterior inferior frontal gyrus (overlapping with the RLPFC) seems to encode the full meaning of logical compounds. By contrast, the left IPS seems to be more involved in constructing the multiple mental models required to solve "IF" problems, and further involved in the pruning and selection of the specific model used (Baggio et al., 2016). Additionally, a comparison between propositional logical connectors (e.g. AND, OR, IF) and syntactic transformation with ditransitive verbs (e.g. GIVE, SAY, TAKE) showed the RLPFC and the left IPS for specifically logical inference, and interpreted their results to further reinforce the RLPFC's role in relational integration and the IPS's role in representing the structure of logical arguments (Monti et al., 2009). The varied yet often overlapping set of results found in different studies of deductive reasoning more broadly seems to support conclusions by Goel (2007) and Prado et al. (2011) that reasoning is a fractionate system which is dynamically configured to suit the task at hand.

Nevertheless, we believe these results support the argument that transitive inference is a type of reasoning rather than an associative gradient (Frank et al., 2003, 2005; Vasconcelos, 2008). Those findings are based on the apparent ability for rats and humans not consciously aware of the task are able to perform transitive inference. They argue that the structure seen in transitive inference tasks are the result of an associative gradient, since the first item in the series ("A") is always rewarded and the last item ("F") is never rewarded. They find that items with

similar reinforcement value that are similarly distant from the endpoints are not consistently solved correctly. However, this pattern was not seen in the human participants who became aware of the task (Frank et al., 2005), and they did not show significant differences with the comparisons predicted the be significantly different due to associative gradient. Similarly, Titone et al. (2004) found similar behavioral results to the rats and unconscious humans with schizophrenic patients. Interestingly, 75% of the schizophrenic patients who performed like the participants in Frank and colleagues' (2005) study reported being completely unaware or only partially aware of the transitive structure (compared to 73.3% of normal controls who reported being fully aware). It seems possible that there are different processes underlying unconscious transitive inference (best explained by associative gradients) and conscious transitive reasoning. Additionally, our study did not reward participants, nor was any feedback provided based on responses given (not simply in the height probe task, but in every task in this study). The endpoints in this task were equally unrewarded, and participants were not given any instructions about how to organize the items in order learn the overall lineup.

Ultimately, the present results for this transitive reasoning task are consistent with mental models theory (Johnson-Laird, 2010) and are not adequately explained by any existing alternative theory. Moreover, the present work clearly indicates that mental representations created to solve transitive reasoning problems can be neurally localized using RSA. These results represent a proof of concept that RSA can be used to assess learning on a wide range of tasks, and that it can be a particularly useful tool in assessing the neural basis of reasoning.

Acknowledgements

The authors would like to gratefully acknowledge funding to DJMK from Dartmouth College and the Nelson A Rockefeller Center at Dartmouth College.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.neuroimage.2018.07.062.

References

Acuna, B.D., Eliassen, J.C., Donoghue, J.P., Sanes, J.N., 2002. Frontal and parietal lobe activation during transitive inference in humans. Cerebr. Cortex 12 (12), 1312–1321.
Baggio, G., Cherubini, P., Pischedda, D., Blumenthal, A., Haynes, J.D., Reverberi, C., 2016. Multiple neural representations of elementary logical connectives. Neuroimage 135, 300–310.
Bunge, S.A., Helskog, E.H., Wendelken, C., 2009. Left, but not right, rostrolateral prefrontal cortex meets a stringent test of the relational integration hypothesis. Neuroimage 46 (1), 338–342.
Bunge, S.A., Wendelken, C., Badre, D., Wagner, A.D., 2005. Analogical reasoning and prefrontal cortex: evidence for separable retrieval and integration mechanisms. Cerebr. Cortex 15, 239–249.
Byrne, R.M., Johnson-Laird, P.N., 1989. Spatial reasoning. J. Mem. Lang. 28 (5), 564–575.
Casasanto, D., 2003. Hemispheric specialization in prefrontal cortex: effects of verbalizability, imageability and meaning. J. Neurolinguistics 16 (4), 361–382.
Christoff, K., Prabhakaran, V., Dorfman, J., Zhao, Z., Kroger, J.K., Holyoak, K.J., Gabrieli, J.D., 2001. Rostrolateral prefrontal cortex involvement in relational integration during reasoning. Neuroimage 14 (5), 1136–1149.
Christoff, K., Ream, J.M., Geddes, L., Gabrieli, J.D., 2003. Evaluating self-generated information: anterior prefrontal contributions to human cognition. Behav. Neurosci. 117 (6), 1161.
Collin, S.H., Milivojevic, B., Doeller, C.F., 2015. Memory hierarchies map onto the hippocampal long axis in humans. Nat. Neurosci. 18 (11), 1562–1564.
Connolly, A.C., Sha, L., Guntupalli, J.S., Oosterhof, N., Halchenko, Y.O., Nastase, S.A., Haxby, J.V., 2016. How the human brain represents perceived dangerousness or "predacity" of animals. J. Neurosci. 36 (19), 5373–5384.
Connolly, A.C., Guntupalli, J.S., Gors, J., Hanke, M., Halchenko, Y.O., Wu, Y.C., Haxby, J.V., 2012. The representation of biological classes in the human brain. J. Neurosci. 32 (8), 2608–2618.
Crone, E.A., Wendelken, C., Donohue, S., van Leijenhorst, L., Bunge, S.A., 2006. Neurocognitive development of the ability to manipulate information in working memory. Proc. Natl. Acad. Sci. Unit. States Am. 103 (24), 9315–9320.
Dehaene, S., 2007. Symbols and quantities in parietal cortex: elements of a mathematical theory of number representation and manipulation. In: Haggard, P., Rossetti, Y. (Eds.), Attention & Performance XXII. Sensori-motor Foundations of Higher Cognition, pp. 527–574.
Destrieux, C., Fischl, B., Dale, A., Halgren, E., 2010. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. Neuroimage 53 (1), 1–15.
Dosenbach, N.U., Visscher, K.M., Palmer, E.D., Miezin, F.M., Wenger, K.K., Kang, H.C., Petersen, S.E., 2006. A core system for the implementation of task sets. Neuron 50 (5), 799–812.
Dusek, J.A., Eichenbaum, H., 1997. The hippocampus and memory for orderly stimulus relations. Proc. Natl. Acad. Sci. Unit. States Am. 94 (13), 7109–7114.
Fischl, B., 2012. FreeSurfer. Neuroimage 62 (2), 774–781.
Frank, M.J., Rudy, J.W., Levy, W.B., O'Reilly, R.C., 2005. When logic fails: implicit transitive inference in humans. Mem. Cognit. 33 (4), 742–750.
Frank, M.J., Rudy, J.W., O'Reilly, R.C., 2003. Transitivity, flexibility, conjunctive representations, and the hippocampus. II. A computational analysis. Hippocampus 13 (3), 341–354.
Garvert, M.M., Dolan, R.J., Behrens, T.E., 2017. A map of abstract relational knowledge in the human hippocampal–entorhinal cortex. eLife 6, e17086.
Goel, V., 2007. Anatomy of deductive reasoning. Trends Cognit. Sci. 11 (10), 435–441.
Goel, V., Dolan, R.J., 2001. Functional neuroanatomy of three-term relational reasoning. Neuropsychologia 39 (9), 901–909.
Goel, V., Dolan, R.J., 2004. Differential involvement of left prefrontal cortexin inductive and deductive reasoning. Cognition 93 (3), B109–B121.
Green, A.E., Fugelsang, J.A., Kraemer, D.J., Shamosh, N.A., Dunbar, K.N., 2006. Frontopolar cortex mediates abstract integration in analogy. Brain Res. 1096 (1), 125–137.
Green, A.E., Kraemer, D.J.M., Fugelsang, J.A., Gray, J.R., Dunbar, K.N., 2010. Connecting long distance: semantic distance in analogical reasoning modulates frontopolar cortex activity. Cerebr. Cortex 20 (1), 70–76.
Green, A.E., Kraemer, D.J., Fugelsang, J.A., Gray, J.R., Dunbar, K.N., 2012. Neural correlates of creativity in analogical reasoning. J. Exp. Psychol. Learn. Mem. Cognit. 38 (2), 264.
Guntupalli, J.S., Wheeler, K.G., Gobbini, M.I., 2016. Disentangling the representation of identity from head view along the human face processing pathway. Cerebr. Cortex.
Hanke, M., Halchenko, Y.O., Sederberg, P.B., Hanson, S.J., Haxby, J.V., Pollmann, S., 2009. PyMVPA: a python toolbox for multivariate pattern analysis of fMRI data. Neuroinformatics 7 (1), 37–53.
Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., Pietrini, P., 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293 (5539), 2425–2430.
Heckers, S., Zalesak, M., Weiss, A.P., Ditman, T., Titone, D., 2004. Hippocampal activation during transitive inference in humans. Hippocampus 14 (2), 153–162.
Jahn, G., Knauff, M., Johnson-Laird, P.N., 2007. Preferred mental models in reasoning about spatial relations. Mem. Cognit. 35 (8), 2075–2087.
Jenkinson, M., Bannister, P., Brady, M., Smith, S., 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 17 (2), 825–841.
Johnson-Laird, P.N., 1983. Mental Models: towards a Cognitive Science of Language, Inference, and Consciousness (No. 6). Harvard University Press.
Johnson-Laird, P.N., 2010. Mental models and human reasoning. Proc. Natl. Acad. Sci. Unit. States Am. 107 (43), 18243–18250.
Kanwisher, N., McDermott, J., Chun, M.M., 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. J. Neurosci. 17 (11), 4302–4311.
Kelley, W.M., Miezin, F.M., McDermott, K.B., Buckner, R.L., Raichle, M.E., Cohen, N.J., Petersen, S.E., 1998. Hemispheric specialization in human dorsal frontal cortex and medial temporal lobe for verbal and nonverbal memory encoding. Neuron 20 (5), 927–936.
Knauff, M., 2013. Space to Reason: a Spatial Theory of Human Thought. MIT Press.
Knauff, M., Fangmeier, T., Ruff, C.C., Johnson-Laird, P.N., 2003. Reasoning, models, and images: behavioral measures and cortical activity. J. Cognit. Neurosci. 15 (4), 559–573.
Knauff, M., Mulack, T., Kassubek, J., Salih, H.R., Greenlee, M.W., 2002. Spatial imagery in deductive reasoning: a functional MRI study. Cognit. Brain Res. 13 (2), 203–212.
Krawczyk, D.C., 2012. The cognition and neuroscience of relational reasoning. Brain Res. 1428, 13–23.
Krawczyk, D.C., McClelland, M., Donovan, C.M., 2010. A hierarchy for relational reasoning in the prefrontal Cortex. Cortex.
Kriegeskorte, N., Mur, M., Bandettini, P., 2008. Representational similarity analysis–connecting the branches of systems neuroscience. Front. Syst. Neurosci. 2 (4).
Kriegeskorte, N., Kievit, R.A., 2013. Representational geometry: integrating cognition, computation, and the brain. Trends Cognit. Sci. 17 (8), 401–412.
Lyons, I.M., Ansari, D., Beilock, S.L., 2015. Qualitatively different coding of symbolic and nonsymbolic numbers in the human brain. Hum. Brain Mapp. 36 (2), 475–488.
Mack, M.L., Love, B.C., Preston, A.R., 2016. Dynamic updating of hippocampal object representations reflects new conceptual knowledge. Proc. Natl. Acad. Sci. Unit. States Am., 201614048
McNamara, T.P., 1986. Mental representations of spatial relations. Cognit. Psychol. 18 (1), 87–121.
Monti, M.M., Parsons, L.M., Osherson, D.N., 2009. The boundaries of language and thought in deductive inference. Proc. Natl. Acad. Sci. Unit. States Am. 106 (30), 12554–12559.
Moyer, R.S., Bayer, R.H., 1976. Mental comparison and the symbolic distance effect. Cognit. Psychol. 8 (2), 228–246.

Oosterhof, N.N., Wiestler, T., Downing, P.E., Diedrichsen, J., 2011. A comparison of volume-based and surface-based multi-voxel pattern analysis. Neuroimage 56 (2), 593–600.
Pihlajamäki, M., Tanila, H., Hänninen, T., Könönen, M., Mikkonen, M., Jalkanen, V., Soininen, H., 2003. Encoding of novel picture pairs activates the perirhinal cortex: an fMRI study. Hippocampus 13 (1), 67–80.
Prado, J., Chadha, A., Booth, J.R., 2011. The brain network for deductive reasoning: a quantitative meta-analysis of 28 neuroimaging studies. J. Cognit. Neurosci. 23 (11), 3483–3497.
Prado, J., Mutreja, R., Booth, J.R., 2012. Fractionating the neural substrates of transitive reasoning: task-dependent contributions of spatial and verbal representations. Cerebr. Cortex 23 (3), 499–507.
Ramnani, N., Owen, A.M., 2004. Anterior prefrontal cortex: insights into function from anatomy and neuroimaging. Nat. Rev. Neurosci. 5 (3), 184–194.
Ruff, C.C., Knauff, M., Fangmeier, T., Spreer, J., 2003. Reasoning and working memory: common and distinct neuronal processes. Neuropsychologia 41 (9), 1241–1253.
Schlegel, A., Kohler, P.J., Fogelson, S.V., Alexander, P., Konuthula, D., Tse, P.U., 2013. Network structure and dynamics of the mental workspace. Proc. Natl. Acad. Sci. Unit. States Am. 110 (40), 16277–16282.
Serre, T., Oliva, A., Poggio, T., 2007. A feedforward architecture accounts for rapid categorization. Proc. Natl. Acad. Sci. Unit. States Am. 104 (15), 6424–6429.
Titone, D., Ditman, T., Holzman, P.S., Eichenbaum, H., Levy, D.L., 2004. Transitive inference in schizophrenia: impairments in relational memory organization. Schizophr. Res. 68 (2), 235–247.
Tottenham, N., Tanaka, J.W., Leon, A.C., McCarry, T., Nurse, M., Hare, T.A., Nelson, C., 2009. The NimStim set of facial expressions: judgments from untrained research participants. Psychiatr. Res. 168 (3), 242–249.
Van Opstal, F., Fias, W., Peigneux, P., Verguts, T., 2009. The neural representation of extensively trained ordered sequences. Neuroimage 47 (1), 367–375.
Van Opstal, F., Verguts, T., Orban, G.A., Fias, W., 2008. A hippocampal–parietal network for learning an ordered sequence. Neuroimage 40 (1), 333–341.
Vasconcelos, M., 2008. Transitive inference in non-human animals: an empirical and theoretical analysis. Behav. Process. 78 (3), 313–334.
Vendetti, M.S., Bunge, S.A., 2014. Evolutionary and developmental changes in the lateral frontoparietal network: a little goes a long way for higher-level cognition. Neuron 84 (5), 906–917.
Wendelken, C., 2015. Meta-analysis: how does posterior parietal cortex contribute to reasoning? Front. Hum. Neurosci. 8, 1042.
Wendelken, C., Bunge, S.A., 2010. Transitive inference: distinct contributions of rostrolateral prefrontal cortex and the hippocampus. J. Cognit. Neurosci. 22 (5), 837–847.
Wendelken, C., Chung, D., Bunge, S.A., 2012. Rostrolateral prefrontal cortex: domain-general or domain-sensitive? Hum. Brain Mapp. 33 (8), 1952–1963.
Wharton, C.M., Grafman, J., Flitman, S.S., Hansen, E.K., Brauner, J., Marks, A., Honda, M., 2000. Toward neuroanatomical models of analogy: a positron emission tomography study of analogical mapping. Cognit. Psychol. 40 (3), 173–197.
Whitaker, K., 2012. Doctoral dissertation retrieved from. http://escholarship.org/uc/item/3vs463b7?query=Whitaker#.
Wig, G.S., Miller, M.B., Kingstone, A., Kelley, W.M., 2004. Separable routes to human memory formation: dissociating task and material contributions in the prefrontal cortex. J. Cognit. Neurosci. 16 (1), 139–148.
Yarkoni, T., Poldrack, R.A., Nichols, T.E., Van Essen, D.C., Wager, T.D., 2011. Large-scale automated synthesis of human functional neuroimaging data. Nat. Methods 8 (8), 665–670.
Zalesak, M., Heckers, S., 2009. The role of the hippocampus in transitive inference. Psychiatr. Res. Neuroimaging 172 (1), 24–30.