

Network structure and dynamics of the mental workspace

Alexander Schlegel¹, Peter J. Kohler, Sergey V. Fogelson, Prescott Alexander, Dedeepya Konuthula, and Peter Ulric Tse

Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH 03755

Edited by Michael S. Gazzaniga, University of California, Santa Barbara, CA, and approved August 28, 2013 (received for review June 11, 2013)

The conscious manipulation of mental representations is central to many creative and uniquely human abilities. How does the human brain mediate such flexible mental operations? Here, multivariate pattern analysis of functional MRI data reveals a widespread neural network that performs specific mental manipulations on the contents of visual imagery. Evolving patterns of neural activity within this mental workspace track the sequence of informational transformations carried out by these manipulations. The network switches between distinct connectivity profiles as representations are maintained or manipulated.

occipital | frontoparietal | posterior parietal | precuneus | dorsolateral prefrontal

Albert Einstein described the elements of his scientific thought as “certain signs and more or less clear images which can be ‘voluntarily’ reproduced or combined” (1). Creative thought in science as well as in other domains such as the visual arts, mathematics, music, and dance requires the capacity to manipulate mental representations flexibly. Cognitive scientists refer to this capacity as a “mental workspace” and suggest that it is a key function of consciousness (2) involving the distribution of information among widespread, specialized subdomains (3).

How does the human brain mediate these flexible mental operations? Behavioral studies of the mental workspace, such as Shepard and Metzler’s work on mental rotation (4), have found that many mental operations closely resemble their corresponding physical operations. This finding supports the view that the mental workspace can simulate the physical world. Recent work in neuroscience has focused on mental representations instead of operations, showing that the contents of visual perception (5), visual imagery (6), and even dreams (7) can be decoded from activity in visual cortex. These results suggest that the same regions that mediate representations in sensory perception also are involved in mental imagery. However, how the mind can manipulate these representations remains unknown. Many studies have found increased activity in frontal and parietal regions associated with a range of high-level cognitive abilities (8, 9) including mental rotation (10), analogical reasoning (11), working memory (12), and fluid intelligence (13). Together, these findings suggest that a frontoparietal network may form the core of the mental workspace. We therefore hypothesized that operations on visual representations in the mental workspace are realized through the coordinated activity of a distributed network of regions that spans at least the frontal, parietal, and occipital cortices. A strong test of this hypothesis would be to ask whether patterns of neural activity in these regions contain information about specific mental operations and whether these patterns evolve over time as mental representations are manipulated.

In the present study, we tested this hypothesis by asking 15 participants to engage in either maintenance or manipulation of visual imagery while we collected functional MRI (fMRI) measurements of their neural activity. As stimuli, we developed 100 abstract parts that could be combined into 2×2 figures (Fig. 1*A* and *C*). In a series of trials, participants mentally maintained a set of parts or a whole figure, mentally constructed a set of four

parts into a figure, or mentally deconstructed a figure into its four parts (Fig. 1*B*). Stimuli were presented briefly at the beginning of each trial, followed by a task prompt and a 6-s delay during which the participant performed the indicated mental operation. At the end of the delay, the target output of the operation was presented along with three similar distractors, and the participant indicated the correct target (Fig. 1*D*). Adjusting the complexity of the stimuli allowed us to equate for task difficulty by maintaining an accuracy of two out of three correct responses for each participant in each of the four conditions (chance would be 1 out of 4 correct; Fig. 1*E*).

Results

As an initial procedure for selecting regions of interest (ROI) on the fMRI blood oxygenation level-dependent (BOLD) data, we carried out a whole-brain univariate general linear model (GLM) analysis to identify regions in which neural activity levels differed between mental manipulation (construct parts or deconstruct figure) and mental maintenance (maintain parts or maintain figure) conditions. This analysis revealed 11 bilateral cortical and subcortical ROIs (Fig. 2), suggesting that a widespread network mediates the manipulation tasks. All but two of the ROIs showed greater activation in manipulation than in maintenance conditions; the exceptions were the medial temporal lobe (MTL) and medial frontal cortex. In a separate control GLM analysis, we evaluated whether any regions showed differences in activity between the two manipulation conditions. No voxels were significant in this analysis, suggesting that overall activity levels were well matched between the manipulation tasks. We did not see a univariate effect in occipital cortex. This result is expected, because visual stimuli were equated across the

Significance

We do not know how the human brain mediates complex and creative behaviors such as artistic, scientific, and mathematical thought. Scholars theorize that these abilities require conscious experience as realized in a widespread neural network, or “mental workspace,” that represents and manipulates images, symbols, and other mental constructs across a variety of domains. Evidence for such a complex, interconnected network has been difficult to produce with current techniques that mainly study brain activity in isolation and are insensitive to distributed informational processes. The present work takes advantage of emerging techniques in network and information analysis to provide empirical support for such a widespread and interconnected information processing network in the brain that supports the manipulation of visual imagery.

Author contributions: A.S. and P.J.K. designed research; A.S., P.J.K., P.A., and D.K. performed research; A.S. and P.J.K. contributed new reagents/analytic tools; A.S., S.V.F., P.A., and D.K. analyzed data; and A.S., P.J.K., S.V.F., and P.U.T. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

¹To whom correspondence should be addressed. E-mail: schlegel@gmail.com.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1311149110/-DCSupplemental.

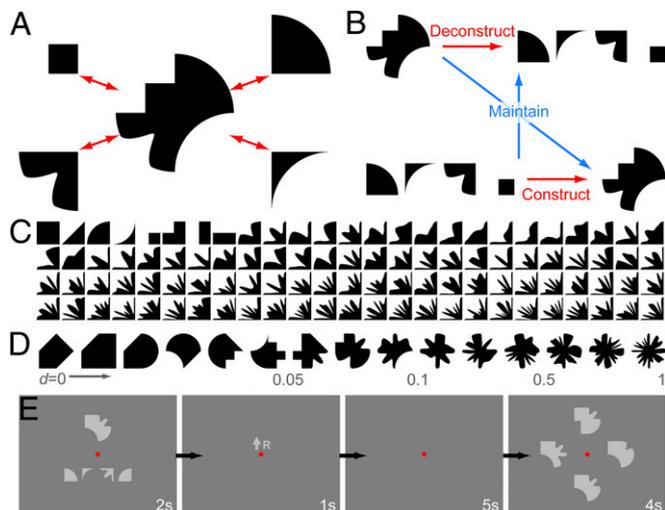


Fig. 1. Experimental design. (A) Parts could be constructed into 2×2 figures, and figures could be deconstructed into parts. (B) Participants performed four mental operations on stimuli: construct parts into figure, deconstruct figure into parts, maintain parts, or maintain figure. (C) The stimulus set of 100 abstract parts, ordered from simple to complex. (D) Example of figures. Parts and figures ranged from simple to complex according to an index, d . This index allowed the difficulty of the task to be equated across conditions. (E) Trial schematic. Trials begin with a figure and four unrelated parts presented for 2 s, followed by a task prompt for 1 s consisting of an arrow indicating the figure or the parts and a letter indicating the task. In this case, the participant is instructed to maintain the figure in memory. The task prompt is followed by a 5-s delay period during which no stimulus is shown and the participant performs the indicated operation. Finally, a test screen appears for 2.5 s. Depending on the task, four figures or four sets of parts are presented, and the participant indicates the correct output of the operation.

four conditions. However, because we hypothesized that visual cortex plays a role in mediating operations on visual imagery, we included an anatomically defined occipital mask in our set of ROIs. Thus we had 12 ROIs to investigate for informational content relevant to the mental operations.

We then attempted to decode the particular mental operations performed by participants based on spatiotemporal patterns of BOLD responses in each of these 12 ROIs. We carried out a multivariate pattern-classification analysis (5) within each ROI. In this analysis, a classifier algorithm first is trained by providing it with a set of BOLD response patterns from the ROI along with the mental operation associated with each pattern. Then a holdout pattern not involved in the training is used to test the classifier. If the classifier can predict above chance the mental operation associated with the holdout pattern, the ROI contains information specific to that particular mental operation and likely is involved in mediating that operation. We carried out two-way classifications in each ROI between construct-parts and deconstruct-figure conditions and between maintain-parts and maintain-figure conditions, with results shown in Fig. 3A. To evaluate the informational content of each ROI in a single analysis, we constructed the model confusion matrix that would be expected for regions that mediated the mental operations (Fig. 3B). A confusion matrix indicates the similarity between patterns from different conditions; if patterns are more similar, the classifier will be more likely to confuse them. In this case, we expected high similarity between patterns from the same condition, moderate similarity when both patterns were from either two manipulation or two maintenance conditions, and low similarity when one pattern was from a manipulation condition and the other was from a maintenance condition. We then carried

out correlation analyses between this model and the actual confusion matrix in each ROI derived from four-way classifications among the conditions (Fig. 3C). These analyses identified a subset of the ROIs, consisting of occipital cortex, posterior parietal cortex (PPC), precuneus, posterior inferior temporal cortex, dorsolateral prefrontal cortex (DLPFC), and frontal eye fields, in which we could decode the specific mental operations from patterns of neural activity. Additional control analyses confirmed that our results were not affected by ROI size or differences in response times between conditions (Fig. S1 and Table S1).

Each of the four operations followed a three-stage temporal sequence in which participants encoded an input into a mental representation, performed a mental operation (construct, deconstruct, or maintain) on that representation, and produced an output mental representation. Each of these stages entailed a unique relationship among the mental states associated with the four conditions (Fig. 4A). For example, the inputs to the construct-parts condition were similar to those of the maintain-parts condition, the operation performed during the construct-parts condition was similar to that of the deconstruct-figure condition, and the outputs from the construct-parts condition were similar to those of the maintain-figure condition. Thus, the relationship among the conditions evolved throughout the trial and provided a means of further exploring the informational content of the mental workspace. To do so, we carried out a four-way classification among the conditions at each time point and correlated the resulting confusion matrices with each of the three model similarity structures in Fig. 4A. High correlation between a confusion matrix and one of the model structures would indicate that a particular region was carrying out the corresponding stage of processing at that time. Fig. 4B shows the time course of correlations with each model in occipital cortex. In Fig. 4C, we report peak correlation times in each of the 12 ROIs. In the four regions with highest classification accuracies in Fig. 3A, correlation peaks progressed from input through operation to output, providing strong evidence that these four areas directly mediated the mental operations as they unfolded over time. It should be noted that the differences between test stimuli could have affected the output correlation time course (orange trace in Fig. 4B) because the output mental representations were similar to the stimuli presented during the test phase. Our experimental design did not allow us to evaluate the relative contributions of the output mental representations and of the test stimuli to the output correlation time course.

The above analyses show that a subset of ROIs supports the temporal evolution of information necessary to carry out particular mental operations. However, they do not provide

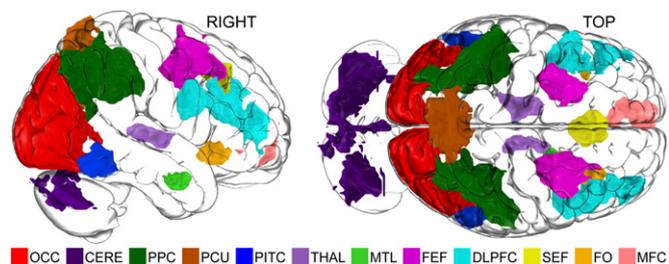


Fig. 2. Eleven ROIs showing differential activity levels in manipulation and maintenance conditions. An additional occipital cortex ROI was defined anatomically. CERE, cerebellum; DLPFC, dorsolateral prefrontal cortex; FEF, frontal eye fields; FO, frontal operculum; MFC, medial frontal cortex; MTL, medial temporal lobe; OCC, occipital cortex; PCU, precuneus; PITC, posterior inferior temporal cortex; PPC, posterior parietal cortex; SEF, supplementary eye field; THA, thalamus.

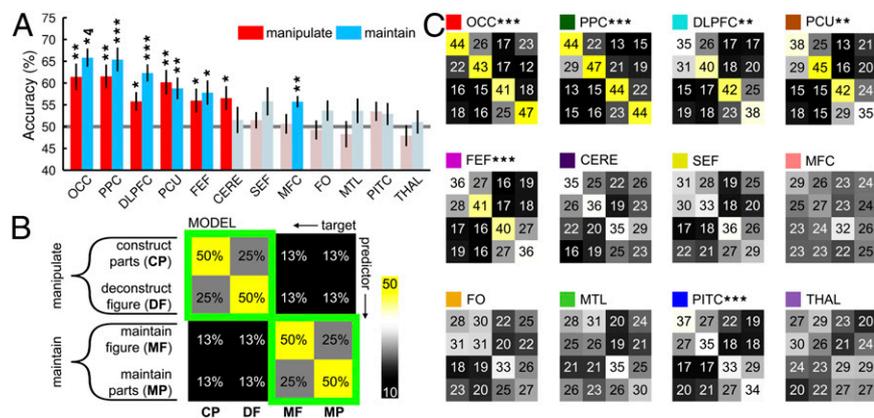


Fig. 3. Results of multivariate pattern classification. Asterisks indicate significance ($*P \geq 0.05$; $**P \geq 0.01$; $***P \geq 0.001$; $*4, P \geq 0.0001$). (A) Results for two-way classifications in each ROI between manipulation conditions and between maintenance conditions. The bar plot shows classification accuracies in descending order. Error bars indicate SEMs. Asterisks indicate accuracies significantly above chance ($P \leq 0.05$, FDR corrected across the 24 comparisons). Table S2 shows full statistical results. (B) Model similarity structure for regions that mediate the mental operations. Manipulation and maintenance conditions should be more similar within condition types than across condition types. (C) Confusion matrices from four-way classifications in each ROI. Values are percentages. Asterisks indicate regions in which confusion matrices correlated significantly with the model ($P \leq 0.05$, FDR corrected across the 12 comparisons). Because ROIs were selected based on differences in activity between manipulation and maintenance conditions, we considered only values within manipulation and maintenance conditions in the correlation (within the green squares in B). Table S3 shows full statistical results.

evidence about how these regions communicate within the mental workspace network. We investigated this communication by analyzing patterns of functional connectivity between the ROIs. For each condition, participant, and region, we constructed a time course by concatenating the mean BOLD signal within that region across the participant's correct-response trials for that condition. We calculated the functional connectivity, defined as the correlation between pairs of time courses, for each condition, participant, and pair of regions (14). This procedure yielded one network-wide pattern of functional connectivity for each condition and participant. A cross-subject classification analysis on these connectivity patterns successfully predicted whether participants mentally manipulated or maintained imagery with 61.7% accuracy [$t(14) = 2.4, P = 0.029$], thus indicating that patterns of connectivity between the network components changed depending on the operation that participants performed on the contents of their mental imagery. Investigating the weights that the classifier assigned to each pair of regions allowed us to determine which connections were most informative (Fig. 5A). Increases in connectivity between pairs with positive weights drove the classifier toward the manipulation conditions, whereas increases between pairs with negative weights drove it toward the maintenance conditions. Thus, stronger connectivities with the precuneus and with left posterior inferior temporal cortex indicated manipulation conditions, and stronger connectivities primarily with the MTL indicated maintenance conditions. In Fig. 5B, we plot the difference in functional connectivity between conditions. During manipulation conditions the precuneus and posterior inferior temporal cortex showed stronger connectivity with several frontal and parietal regions, whereas connectivity between the MTL and many regions became weaker. Thus, our data show not only that a distributed set of regions mediates mental operations but also that these regions communicate in an information-processing network. The network switches between two connectivity profiles depending on whether mental representations are maintained or manipulated.

Discussion

Our findings reveal a widespread cortical and subcortical network that operates on visual representations in the mental workspace. This network includes four core regions spanning the DLPFC, PPC, posterior precuneus, and occipital cortex that manipulate the contents of visual imagery. Within these regions

we decoded and tracked the evolution of mental operations over time. Several other areas showed a difference in BOLD responses between the manipulation and maintenance conditions but without the specificity found in the four core areas. Therefore it is likely that an extended network of regions is involved in the operations. Changes in patterns of connectivity between the mental workspace network's nodes reveal that the network supports at least two distinct modes of operation, depending on whether mental representations are maintained or manipulated. We discuss each of the identified components of the network below.

Frontoparietal Cortex. Our finding that the DLPFC and PPC directly mediate manipulation of visual imagery is supported by multiple studies suggesting that a network of frontal and parietal areas is involved in many high-level cognitive abilities in humans (10–13). Miller et al. (15) showed that the responses of neurons in DLPFC convey more information about the task relevance of stimuli than about their specific features and that this selectivity for task relevance is maintained over extended durations in the absence of stimulus input. Thus, the DLPFC appears to be part of a network that maintains representations in working memory via attention. Human neuroimaging studies have shown that both the DLPFC and the PPC are activated, regardless of the type of information that is held in working memory (16, 17). Selectivity for task rather than representation distinguishes this system from subsidiary systems that are capable only of maintaining particular classes of information (18). These findings support the view that the frontoparietal network is an executive system that recruits subsidiary systems, as proposed in Baddeley's (19) model of working memory. Modeling work by O'Reilly and colleagues (8, 9) has shown how prefrontal cortex may be able to self-organize abstract rules flexibly and later apply them to specific representations. This ability is common to many flexible cognitive processes in humans such as analogical reasoning, creativity (11), and fluid intelligence (13). Our data provide empirical support for this model by showing that the DLPFC and PPC mediate not only the maintenance of representations in working memory but also the manipulation of those representations. Thus, these areas may form the core of a system that mediates conscious operations on mental representations, in this case the contents of visual imagery represented at least partially in the occipital cortex.

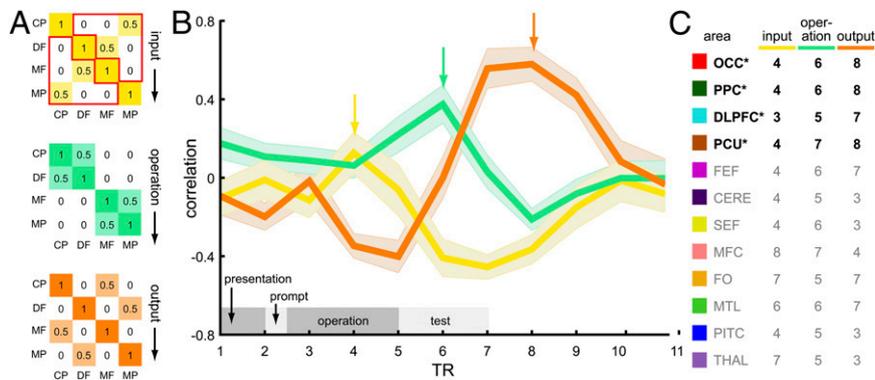


Fig. 4. Temporal progression of the neural informational structure during mental operations. (A) Model similarity structures between the four conditions based on the input mental representation, the mental operation performed, and the output mental representation. For example, constructing and maintaining parts have similar input representations, whereas constructing parts and maintaining figures have similar output representations. The red outline indicates values used in the following correlation time courses. (B) Time course of correlations in occipital cortex between model similarity structures and confusion matrices from individual time-point classifications. Shading indicates SEMs. The schematic along the x-axis shows the trial stages. (C) Peak correlation times for the 12 ROIs. In the four ROIs with highest classification accuracies in Fig. 3, the peaks in the correlation time courses followed a significant sequence from input mental representation, through operation, to output representation (asterisks indicate significant ROIs). Table S4 shows full statistical results.

Occipital Cortex. Several studies have found that the occipital cortex processes information relevant to internally generated visual experience. Harrison and Tong (6) used patterns of activity in early visual cortex to decode the orientation of gratings that participants maintained in working memory. Recently, Horikawa and colleagues (7) decoded the contents of participants' visual experience during dreaming from patterns in visual cortex. Thus, the visual cortex likely represents the contents of both internally and perceptually generated visual experience. Our results extend these findings to show that mental representations not only are formed but also are operated on in visual cortex. This result may generalize to other sensory domains, so that the brain mediates perceptual processes and operates on the corresponding mental representations in the same regions.

Precuneus. Margulies et al. (20) reported that the precuneus in humans is functionally connected to the lateral frontal, posterior parietal, and occipital cortices. The precuneus is one of the most connected regions of the cortex, suggesting that it may serve as a hub in several cortical networks. In their review, Cavanna and Trimble (21) cite a body of evidence that the precuneus is involved in visuospatial imagery, is relatively larger in humans than in nonhuman primates and other animals, and is one of the last regions to myelinate during development. Consistent with these findings, Vogt and Laureys (22) propose that the precuneus plays a central role in conscious information processing. Extending

this work, our data show that the posterior precuneus becomes more functionally connected to the DLPFC, PPC, and occipital cortex when participants manipulate mental visual representations and suggest that it acts as a hub in the mental workspace network.

Extended Network. Our findings reveal that the DLPFC, PPC, posterior precuneus, and occipital cortex are central to the mental workspace. However, several other regions activated during the experimental tasks. Current understanding of these areas' functions suggests possible roles they could play in mental operations. The cerebellum, long thought to be involved exclusively in motor coordination, now is known to connect strongly to prefrontal and posterior parietal cortices and to mediate attentional processes (23). Posterior regions of the inferotemporal cortex are involved in visual object processing (24). The thalamus is a hub for interaction between cortical areas and may play a critical role in consciousness (25). The MTL is a hub in memory formation and retrieval (26). This role is supported by our finding of stronger functional connectivity between the MTL and other ROIs during maintenance conditions. The frontal and supplementary eye fields play a role in controlling visual attention (27). Recently, Higo et al. (28) showed that the frontal operculum controls attention toward occipito-temporal representations of stimuli held in memory. The medial frontal cortex is a hub in the default mode network that plays a role in self-

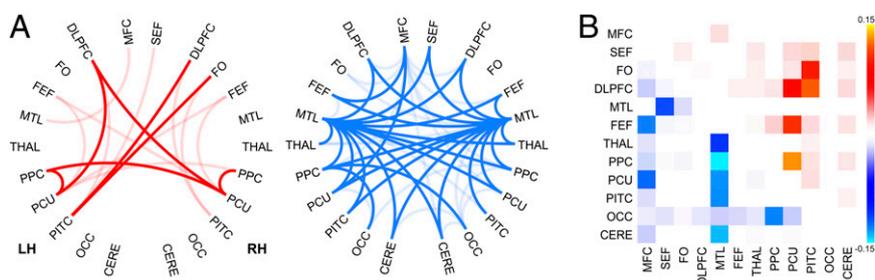


Fig. 5. Multivariate pattern analysis of functional connectivities. (A) Sensitivities for each pair of ROIs in a between-subject classification of functional connectivity in the manipulation and maintenance conditions. Red sensitivities are positive, driving the classifier toward choosing "manipulate." Blue sensitivities are negative, driving it toward "maintain." Only significant nonzero sensitivities are shown [$P \leq 0.05$, corrected for similarity between folds (36)]. Saturated colors indicate sensitivities that survived FDR correction across the 231 comparisons. (B) Difference in functional connectivity in the manipulation and maintenance conditions. Positive and negative differences are separated into the upper and lower diagonals, respectively. Only significant differences in connectivity are shown ($P \leq 0.05$), and differences surviving FDR correction are shown in saturated colors.

directed attentional processes (29). Thus, all these regions are likely involved in the mental operations performed by participants.

A significant finding of the present study is that connectivity in the mental workspace network switches between orthogonal modes of operation depending on whether the network maintains or manipulates representations. Although several network components represent information during both tasks, our data show that patterns of network connectivity associated with these tasks differ substantially. Maintenance of representations involves dense, bilateral interconnections across the entire network with the MTL acting as a hub, whereas manipulation of those representations recruits a sparse, slightly left-lateralized network with a hub in the posterior precuneus. Although the MTL hub does not contain specific information about either mental representations or manipulations, the posterior precuneus hub contains information specific to each operation. This finding suggests that these hubs serve distinct functions across the tasks. The MTL appears to bind network components together, whereas the posterior precuneus may exchange information within a sparse core of this network that itself supports manipulation of representations.

Previous studies have not been able to find evidence that the areas we identified play specific roles in manipulating representations. They have shown differences in BOLD or connectivity between maintenance and manipulation in certain areas (30–32) but have not shown that these areas are responsible for the manipulations themselves. An alternative explanation of these findings could be merely that attentional allocation is increased during manipulation as compared with maintenance tasks. In this study we investigated neural activity in two qualitatively distinct types of manipulations. We show that a subset of areas in the mental workspace network contains information specific to particular manipulations. We additionally show that the task-related informational structure of these areas evolves over time in accordance with the manipulations performed. These results provide specific evidence for the particular network components that directly mediate mental operations.

Human cognition is distinguished by the flexibility with which mental representations can be constructed and manipulated to generate novel ideas and actions. Dehaene (2) and others have proposed that this ability is a key role of a global neuronal workspace that in part realizes our conscious experience. Here we have shown that patterns of activity in just such a distributed neuronal network mediate the flexible recombination of mental images. Although the present study was limited to visual imagery, we anticipate that this network is part of a more general workspace in the human brain in which core conscious processes in frontal and parietal areas recruit specialized subdomains for specific mental operations. Understanding the neural basis of this workspace could reveal common processes central to the flexible cognitive abilities that characterize our species.

Materials and Methods

Participants. Sixteen participants (six females) age 19–30 y gave informed written consent according to the Institutional Review Board guidelines of Dartmouth College before participating. Data from one participant who could not achieve our task accuracy criterion were discarded before further analysis. Participation consisted of two sessions: an initial behavioral session during which participants practiced the tasks and an fMRI session.

Stimulus. The stimulus set consisted of 100 abstract parts (Fig. 1C). The first eight parts were defined manually, and each subsequent part was generated by randomly perturbing a quarter-circle with fixed endpoints (see *SI Materials and Methods* for details).

Task. Participants performed four mental operations with the stimuli: They mentally constructed four parts into a figure, deconstructed a figure into four parts, maintained four parts, or maintained a figure. At the start of each trial, both a figure and four unrelated parts were displayed to equate for low-level

image properties and attention across tasks. After 2 s, the stimulus disappeared and was replaced for 1 s by a prompt indicating the task to be performed. The participant then had 5 s to perform the operation, during which only a fixation dot appeared. Finally, a test screen appeared in which the target output of the operation was shown along with three distractors that were identical to the target except for a single part. The participant was instructed to indicate the target within 4 s of the test screen's appearance. The stimulus complexity was updated on each trial so that participants achieved an accuracy of two out of three correct responses in each trial type. See *SI Materials and Methods* for an extended description of the task.

MRI Acquisition and Preprocessing. Data were collected using a 3.0 T Philips Achieva Intera scanner with a 32-channel sense head coil at the Dartmouth Brain Imaging Center. Participants completed 10 functional runs consisting of 16 trials interleaved with 10-s blanks. fMRI data were preprocessed using FSL (33), and structural images were processed using the FreeSurfer image analysis suite (34). See *SI Materials and Methods* for a detailed description of acquisition parameters and preprocessing steps.

ROI Selection Procedure. A whole-brain GLM analysis was carried out on functional data using the FMRIB Software Library's FEAT tool. A first-level analysis for each participant used boxcar predictors for each of the four conditions, convolved with a double-gamma hemodynamic response function (HRF). Only trials in which participants made correct responses were considered (~27 per condition). The results of this analysis were passed to higher-level cross-subject analyses carried out in Montreal Neurological Institute space, in which *t* contrasts were defined for manipulate > maintain and for manipulate < maintain. Each *t*-contrast map was cluster thresholded at $z \geq 2.3$; clusters then were thresholded at $P \leq 0.05$ according to Gaussian Random Field theory (33). This analysis yielded 11 bilateral ROIs that then were transformed back into each participant's native space for further analysis. An additional occipital ROI was defined anatomically in each participant's native space using the following cortical masks from FreeSurfer: inferior occipital gyrus and sulcus; middle occipital gyrus and sulci; superior occipital gyrus; cuneus; occipital pole; superior occipital and transverse occipital sulci; and anterior occipital sulcus.

Multivariate Pattern Analysis: Classification. Multivariate pattern analysis (MVPA) was carried out using PyMVPA (35). Spatiotemporal patterns were constructed for each correct-response trial and ROI using the z-scored BOLD response from TRs 4–6 (the period during which the operation was performed, after shifting by a 4-s estimate of the HRF delay). Classification was carried out in each ROI between construct-parts and deconstruct-figure trials and between maintain-parts and maintain-figure trials using these patterns, a linear support vector machine (SVM) classifier, and leave-one-out cross-validation. Significance of accuracies was evaluated using one-tailed, one-sample *t* tests compared with chance (50%) and false-discovery rate (FDR) corrected across the 24 comparisons (one for each ROI and classification). A four-way classification also was carried out in each ROI to produce the confusion matrices in Fig. 3C. The correlation between each of these confusion matrices and the model similarity structure was calculated (Fig. 3B), and significance was determined at $P \leq 0.05$, FDR corrected across the 12 comparisons (one for each ROI).

MVPA: Correlation Time Courses. Four-way classification was carried out for each ROI and at each time point of the trial, here using only spatial patterns of the BOLD signal across all voxels within the ROI. This procedure produced a confusion matrix for each time point and ROI, and these confusion matrices were correlated with each of the model similarity structures in Fig. 4A. The first structure models similarities between the conditions based on whether the input representation is a set of parts or a figure. The second structure models similarities based on the two types of operations carried out, manipulation or maintenance. The third structure models similarities based on the outputs from each condition. For each ROI and model structure, we calculated the time point at which the mean correlation reached a maximum, yielding the table in Fig. 4C. These calculations were restricted to TRs 3–8, representing the pretest portion of the trial shifted by 4 s to account for hemodynamic lag. For each ROI we carried out one-way repeated-measures ANOVA on the peak correlation times to test whether the expected progression from input through operation to output occurred. We performed the analysis on trimmed, jackknifed data as recommended by Miller et al. (36) for latency analyses. In a jackknifed analysis with *N* subjects, *N* grand means of the data are calculated, each with one subject left out. The analysis then is performed on these grand means with corrections applied for the jackknife-induced decrease in variance. In the case of noisy

estimates, as occur when calculating latencies from single-subject time courses, this procedure provides cleaner results and does not bias estimates of significance. For each ANOVA we defined two orthogonal linear contrasts (input/output/output: $C1 = -1/-1/2$; $C2 = -1/1/0$) to evaluate the temporal order of the peaks. We determined that an ROI significantly followed the expected progression if and only if both of these contrasts were significant at $P \leq 0.05$ uncorrected.

Functional Connectivity. The functional connectivity (14), defined as the Fisher's z-transformed correlation between time courses, was calculated for each participant and condition across all pairs of the 24 unilateral ROIs and using data pooled across all correct trials. This procedure gave a single

connectivity pattern for each participant and condition. Unilateral ROIs were used to maximize the potential information in each pattern. We then carried out a cross-subject classification between manipulation and maintenance conditions, using these connectivity patterns and an SVM classifier. The sensitivities shown in Fig. 5A are significantly different from zero in a one-sample t test, corrected for the low variance because of the similarity between folds (36), and thresholded at $P \leq 0.05$.

ACKNOWLEDGMENTS. We thank Andrei Gorea for his input on the study design. This study was funded by a National Science Foundation Graduate Research Fellowship (to A.S.) and by Templeton Foundation Grant 14316 (to P.U.T.).

1. Hadamard J (1954) *The Psychology of Invention in the Mathematical Field* (Dover Publications, Inc., New York), p 142.
2. Dehaene S, Naccache L (2001) Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition* 79(1-2):1-37.
3. Baars BJ, Franklin S (2003) How conscious experience and working memory interact. *Trends Cogn Sci* 7(4):166-172.
4. Shepard RN, Metzler J (1971) Mental rotation of three-dimensional objects. *Science* 171(3972):701-703.
5. Haxby JV, et al. (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293(5539):2425-2430.
6. Harrison SA, Tong F (2009) Decoding reveals the contents of visual working memory in early visual areas. *Nature* 458(7238):632-635.
7. Horikawa T, Tamaki M, Miyawaki Y, Kamitani Y (2013) Neural decoding of visual imagery during sleep. *Science* 340(6132):639-642.
8. Rougier NP, Noelle DC, Braver TS, Cohen JD, O'Reilly RC (2005) Prefrontal cortex and flexible cognitive control: Rules without symbols. *Proc Natl Acad Sci USA* 102(20):7338-7343.
9. O'Reilly RC (2006) Biologically based computational models of high-level cognition. *Science* 314(5796):91-94.
10. Zacks JM (2008) Neuroimaging studies of mental rotation: A meta-analysis and review. *J Cogn Neurosci* 20(1):1-19.
11. Green AE, Kraemer DJM, Fugelsang JA, Gray JR, Dunbar KN (2012) Neural correlates of creativity in analogical reasoning. *J Exp Psychol Learn Mem Cogn* 38(2):264-272.
12. Salazar RF, Dotson NM, Bressler SL, Gray CM (2012) Content-specific fronto-parietal synchronization during visual working memory. *Science* 338(6110):1097-1100.
13. Jung RE, Haier RJ (2007) The Parieto-Frontal Integration Theory (P-FIT) of intelligence: Converging neuroimaging evidence. *Behav Brain Sci* 30(2):135-154, discussion 154-187.
14. Fox MD, et al. (2005) The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc Natl Acad Sci USA* 102(27):9673-9678.
15. Miller EK, Erickson CA, Desimone R (1996) Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *J Neurosci* 16(16):5154-5167.
16. Todd JJ, Marois R (2004) Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature* 428(6984):751-754.
17. Xu Y, Chun MM (2006) Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature* 440(7080):91-95.
18. Druzgal TJ, D'Esposito M (2003) Dissecting contributions of prefrontal cortex and fusiform face area to face working memory. *J Cogn Neurosci* 15(6):771-784.
19. Baddeley AD (1986) *Working Memory* (Oxford Univ Press, New York).
20. Margulies DS, et al. (2009) Precuneus shares intrinsic functional architecture in humans and monkeys. *Proc Natl Acad Sci USA* 106(47):20069-20074.
21. Cavanna AE, Trimble MR (2006) The precuneus: A review of its functional anatomy and behavioural correlates. *Brain* 129(Pt 3):564-583.
22. Vogt BA, Laureys S (2005) Posterior cingulate, precuneal and retrosplenial cortices: Cytology and components of the neural network correlates of consciousness. *Prog Brain Res* 150(05):205-217.
23. Bostan AC, Dum RP, Strick PL (2013) Cerebellar networks with the cerebral cortex and basal ganglia. *Trends Cogn Sci* 17(5):241-254.
24. Tanaka K (1996) Inferotemporal cortex and object vision. *Annu Rev Neurosci* 19:109-139.
25. Tononi G (2008) Consciousness as integrated information: A provisional manifesto. *Biol Bull* 215(3):216-242.
26. Ryan L, Lin C-Y, Ketcham K, Nadel L (2010) The role of medial temporal lobe in retrieving spatial and nonspatial relations from episodic and semantic memory. *Hippocampus* 20(1):11-18.
27. Schall JD (2004) On the role of frontal eye field in guiding attention and saccades. *Vision Res* 44(12):1453-1467.
28. Higo T, Mars RB, Boorman ED, Buch ER, Rushworth MFS (2011) Distributed and causal influence of frontal operculum in task control. *Proc Natl Acad Sci USA* 108(10):4230-4235.
29. Buckner RL, Andrews-Hanna JR, Schacter DL (2008) The brain's default network: Anatomy, function, and relevance to disease. *Ann N Y Acad Sci* 1124:1-38.
30. Ranganath C, Cohen MX, Dam C, D'Esposito M (2004) Inferior temporal, prefrontal, and hippocampal contributions to visual working memory maintenance and associative memory retrieval. *J Neurosci* 24(16):3917-3925.
31. Mourao-Miranda J, Ecker C, Sato JR, Brammer MJ (2009) Dynamic changes in the mental rotation network revealed by pattern recognition analysis of fMRI data. *J Cogn Neurosci* 21(5):890-904.
32. Formisano E, et al. (2002) Tracking the mind's image in the brain I: Time-resolved fMRI during visuospatial mental imagery. *Neuron* 35(1):185-194.
33. Smith SM, et al. (2004) Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* 23(Suppl 1):S208-S219.
34. Dale AM, Fischl B, Sereno MI (1999) Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* 9(2):179-194.
35. Hanke M, et al. (2009) PyMVPA: A python toolbox for multivariate pattern analysis of fMRI data. *Neuroinformatics* 7(1):37-53.
36. Miller J, Patterson T, Ulrich R (1998) Jackknife-based method for measuring LRP onset latency differences. *Psychophysiology* 35(1):99-115.

Supporting Information

Schlegel et al. 10.1073/pnas.1311149110

SI Materials and Methods

Participants. Sixteen participants (six females) age 19–30 y gave informed written consent according to the Institutional Review Board guidelines of Dartmouth College before participating. Data from one participant who could not achieve our task accuracy criterion were discarded before further analysis. Participation consisted of two sessions: an initial behavioral session during which participants practiced the tasks and a functional MRI (fMRI) session.

Stimulus. One hundred abstract parts served as the stimulus set (Fig. 1C). The first eight parts were defined manually. Each subsequent part was generated by randomly perturbing a quarter-circle while fixing the endpoints. For each part, 1,000 shapes were generated randomly, and the shape with lowest correlation to the previous shapes was chosen. The complexity of parts scaled with the number of control points used to generate them. Any four parts could be assembled into a 2×2 figure (Fig. 1A). A difficulty index d that scaled from 0–1 was used to specify the subset of parts to use, enabling us to control the difficulty of each task independently for each participant.

Task. Participants performed four mental operations with the stimuli: They mentally constructed four parts into a figure, deconstructed a figure into four parts, maintained four parts, or maintained a figure. Parts always were displayed in a horizontal row, rotated into the correct orientation so that, if constructed into a figure, they would be ordered clockwise starting with the upper right quadrant. During each 12-s trial, participants performed one operation. At the start of each trial, a figure and four unrelated parts were displayed, one above and the other below fixation (counterbalanced across trials). Both a figure and parts were displayed to equate for low-level image properties and attention across tasks. After 2 s, the stimulus disappeared and was replaced for 1 s by a task prompt consisting of an upward- or downward-facing arrow indicating whether the figure or parts would be used in the task and the letter “C” (for construct), “D” (for deconstruct), or “R” (for remember). The participant then had 5 s to perform the operation, during which time only a fixation dot appeared. Finally, a test screen appeared in which either four figures or four sets of parts (depending on the task) were shown for 2.5 s. One of these stimuli was the output of the instructed operation, and the other three were distractors that were identical to the target except for a single part. The participant was instructed to indicate the target within 4 s of the test screen’s appearance. During the behavioral session participants completed 50 trials of each operation type, with stimulus complexity set using a staircase procedure. From these data we estimated the d value for each operation at which each participant chose the correct target in two out of three trials.

MRI Acquisition. Data were collected using a 3.0 T Philips Achieva Intera scanner with a 32-channel sense head coil at the Dartmouth Brain Imaging Center. Whole-brain functional images were acquired using a T2*-weighted gradient-EPI scan [2,000 ms TR, 20 ms TE; 90° flip angle, 240×240 mm field of view (FOV); $3 \times 3 \times 3.5$ mm voxels; 0 mm slice gap; 35 slices]. Structural images were acquired using a T1-weighted magnetization-prepared rapid acquisition gradient echo sequence (8.176 ms TR; 3.72 ms TE; 8° flip angle; 240×220 mm FOV; 188 sagittal slices; $0.9375 \times 0.9375 \times 1$ mm voxels; 3.12 min duration). Participants completed 10 functional runs. Each run consisted of 16

trials interleaved with 10-s blanks, giving 40 trials for each condition. The d value was updated on each trial so that participants achieved an accuracy of two out of three correct responses for each trial type.

MRI Preprocessing. fMRI data were preprocessed using FMRIB Software Library (FSL) (1). Data were motion and slice-time corrected, high-pass filtered with a 100-s cutoff, and spatially smoothed with a 6-mm FWHM Gaussian kernel. Structural images were processed using the FreeSurfer image analysis suite (2).

Procedure for Selecting Regions of Interest. A whole-brain general linear model analysis was carried out on functional data using FSL’s FEAT tool. A first-level analysis for each participant used boxcar predictors for each of the four conditions, convolved with a double-gamma hemodynamic response function (HRF). Only trials for which participants made correct responses were included (~27 per condition). The results of this analysis were passed to higher-level cross-subject analyses, carried out in Montreal Neurological Institute space, in which t contrasts were defined for manipulate > maintain and for manipulate < maintain. Each t -contrast map was cluster thresholded at $z \geq 2.3$; clusters then were thresholded at $P \leq 0.05$ according to Gaussian Random Field theory (1). This analysis yielded 11 bilateral regions of interest (ROIs) that then were transformed back into each participant’s native space for further analysis. An additional occipital ROI was defined anatomically in each participant’s native space using the following cortical masks from FreeSurfer: inferior occipital gyrus and sulcus; middle occipital gyrus and sulci; superior occipital gyrus; cuneus; occipital pole; superior occipital and transverse occipital sulci; and anterior occipital sulcus.

Multivariate Pattern Analysis: Classification. Multivariate pattern analysis (MVPA) was carried out using PyMVPA (3). Spatio-temporal patterns were constructed for each correct-response trial and ROI using the z -scored BOLD response from TRs 4–6 of each trial (the period during which the operation was performed, after shifting by a 4-s estimate of the HRF delay). Classification was carried out in each ROI between construct-parts and deconstruct-figure trials and between maintain-parts and maintain-figure trials, using these patterns, a linear support vector machine (SVM) classifier, and leave-one-out cross-validation. Significance of accuracies was evaluated using one-tailed, one-sample t tests compared with chance (50%) and false discovery rate (FDR) corrected across the 24 comparisons (one for each ROI and classification). A four-way classification also was carried out in each ROI to produce the confusion matrices in Fig. 3C. Correlation analyses were carried out between each of these confusion matrices and the model similarity structure in Fig. 3B. Significance was determined at $P \leq 0.05$, FDR corrected across the 12 comparisons (one for each ROI).

MVPA: Correlation Time Courses. Four-way classifications were carried out at each time point of the trial, here using spatial patterns of BOLD signal across all voxels within each ROI. This process produced a confusion matrix for each time point, and these were correlated with each of the model similarity structures in Fig. 44. The first structure models similarities between the conditions based on whether the input representation is a set of parts or a figure. The second structure models similarities based on the two types of operations carried out (manipulation or maintenance). The third structure models similarities based on the outputs from each condition. For each ROI and model

structure, we calculated the time point at which the mean correlation reached a maximum, yielding the table in Fig. 4C. These calculations were restricted to TRs 3–8, representing the pretest portion of the trial, HRF shifted by 4 s. For each ROI we carried out a one-way repeated-measures ANOVA on the peak correlation times to test whether the expected progression from input through operation to output occurred. We performed the analysis on trimmed, jackknifed data, as recommended by Miller et al. for latency analyses (4). In a jackknifed analysis with N subjects, N grand means of the data are calculated, each with one subject left out. The analysis then is performed on these grand means with corrections applied for the jackknife-induced decrease in variance. In the case of noisy estimates, as occur when calculating latencies from single-subject time courses, this procedure provides cleaner results without biasing estimates of significance. For each ANOVA we defined two orthogonal linear contrasts (input/operation/output: $C1 = -1/-1/2$; $C2 = -1/1/0$) to evaluate the temporal order of the peaks. We determined that

an ROI significantly followed the expected progression if and only if both of these contrasts were significant at $P \leq 0.05$ uncorrected.

Functional Connectivity. The functional connectivity (5), defined as the Fisher's z -transformed correlation between time courses, was calculated for each participant and condition across all pairings of the 24 unilateral ROIs and using data pooled across all correct trials. This calculation yielded a single connectivity pattern for each participant and condition. Unilateral ROIs were used to maximize the potential information in each pattern. We then carried out a cross-subject classification between manipulation and maintenance conditions, using these connectivity patterns and an SVM classifier. The sensitivities shown in Fig. 5A are significantly different from zero in a one-sample t test, corrected for the artificially low variance because of the similarity between folds (4) and thresholded at $P \leq 0.05$.

1. Dale AM, Fischl B, Sereno MI (1999) Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* 9(2):179–194.
2. Hanke M, et al. (2009) PyMVA: A python toolbox for multivariate pattern analysis of fMRI data. *Neuroinformatics* 7(1):37–53.
3. Miller J, Patterson T, Ulrich R (1998) Jackknife-based method for measuring LRP onset latency differences. *Psychophysiology* 35(1):99–115.
4. Miller J, Patterson T, Ulrich R (1998) Jackknife-based method for measuring LRP onset latency differences. *Psychophysiology* 35(1):99–115.
5. Fox MD, et al. (2005) The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc Natl Acad Sci USA* 102(27):9673–9678.

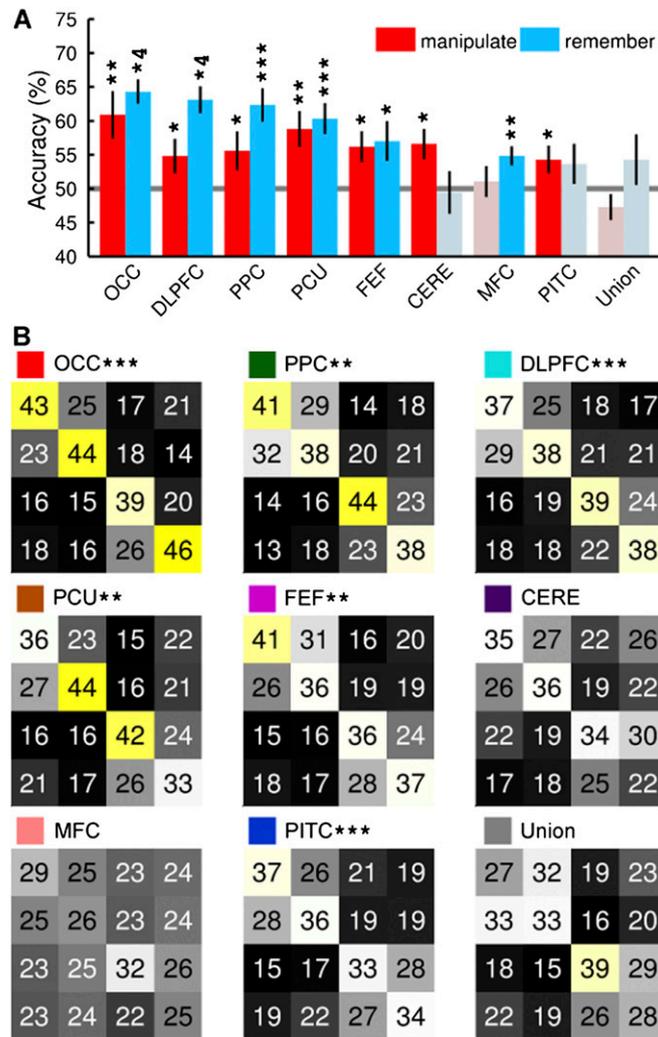


Fig. S1. ROI classification to control for ROI size. Asterisks indicate significance ($*P \geq 0.05$; $**P \geq 0.01$; $***P \geq 0.001$; $*4, P \geq 0.0001$). (A) Classification results using the procedure described in Fig. 3A, with two modifications. First, to verify that the inability to classify between conditions in SEF, FO, MTL, and THAL was not caused by ROI size, the union of these ROIs was constructed and the classification was performed within this "Union" ROI. The average size of this ROI across participants was 360 voxels. Second, to verify that the results of classification in the remaining eight ROIs did not depend on ROI size, we constructed new ROIs by eroding each original ROI until it consisted of the same number of voxels as the smallest of the eight ROIs (127 voxels on average across subjects). CERE, cerebellum; DLPFC, dorsolateral prefrontal cortex; FEF, frontal eye fields; FO, frontal operculum; MFC, medial frontal cortex; MTL, medial temporal lobe; OCC, occipital cortex; PCU, precuneus; PITC, posterior inferior temporal cortex; PPC, posterior parietal cortex; SEF, supplementary eye field; THA, thalamus. (B) Confusion matrices and results of correlation analysis as in Fig. 3C but using the ROIs described above.

