THE MENTAL WORKSPACE AS A DISTRIBUTED NEURAL NETWORK

A Thesis

Submitted to the Faculty

in partial fulfillment of the requirements for the

degree of

Doctor of Philosophy

in

Cognitive Neuroscience

by

Alexander Schlegel

DARTMOUTH COLLEGE

Hanover, New Hampshire

2015 July

Examining Committee:

____

(chair) ***Peter Tse, Ph.D.***

____

Patrick Cavanagh, Ph.D.

____

David Kraemer, Ph.D.

____

Frank Tong, Ph.D.

____

F. Jon Kull, Ph.D.
Dean of Graduate Studies

Abstract. The brain is a vastly interconnected information processing network. In humans, this network supports the rich mental space at the root of the imagination and enables many flexible cognitive abilities such as scientific and artistic creativity. How the brain implements these creative processes remains one of the greatest mysteries in science, and solving this mystery carries with it a possibility for deep understanding of human nature, human potential, and machine intelligence. Logie has proposed that a key substrate for human cognition is a "mental workspace" that enables mental representations such as visual imagery to be formed and manipulated flexibly (1). However, the neural basis of this workspace remains poorly understood, partially because existing experimental methods have limited ability to study complex, higher-order mental functions. Here we develop new methods to probe the structure and dynamics of the large scale networks underlying complex cognition. We use these methods to show that the mental manipulation of visual imagery is mediated by a fundamentally distributed network that spans structures throughout the human brain. Our findings conflict with dominant models that posit an anatomically modular basis for working memory and related processes. Instead, the component processes underlying the mental workspace appear to transcend anatomical modules, occurring at a level of organization that is fundamentally distributed across the brain. Rather than having a fixed anatomical basis, the mental workspace appears to be mediated by a core network that can dynamically and flexibly recruit existing cortical and subcortical subnetworks for specific tasks. These findings call for a shift in cognitive neuroscience research away from functional localization and localized neural circuits and toward the study of organizational principles that govern the large scale integration of information processing in the brain.

Preface

The studies presented here and related work would not have been possible without the support and collaboration of many people:

Study 1 (Ch. 2). I thank Andrei Gorea for his input on the study design. **Contributions.** The author (AS) and Peter J. Kohler (PJK) designed the study. AS, PJK, Prescott Alexander (PA), and Dedeepya Konuthula (DK) collected the data. AS and PJK developed analytic tools. AS, Sergey V. Fogelson (SVF), PA, and DK analyzed data. AS, PJK, SVF, and Peter U. Tse (PUT) wrote the paper. **Funding.** National Science Foundation (NSF) Graduate Research Fellowship 2012095475 to AS and Templeton Foundation Grant 14316 to PUT.

Study 2 (Ch. 3). **Contributions.** AS designed the study and developed analytical tools. AS and PA collected and analyzed data. AS, PA, and PUT wrote the paper. **Funding.** NSF Graduate Research Fellowship 2012095475 to AS and Templeton Foundation Grant 14316 to PUT.

Study 3 (Ch. 4). **Contributions.** AS designed the study and developed analytical tools. AS, DK, and PA collected the data. AS, DK, PA, and Ethan Blackwood (EB) analyzed the data. AS, PA, and PUT wrote the paper. **Funding.** NSF Graduate Research Fellowship 2012095475 to AS and Templeton Foundation Grant 14316 to PUT.

Related work. I thank Justin Rudelson and PUT for their collaboration on a study of white matter changes with second language learning (2); PA, Walter Sinnott-Armstrong, Adina Roskies, PUT, and Thalia Wheatley for their collaboration on a series of studies of free will (3–5); PA, SVF, Xueting Li, Zhengang Lu, PJK, Enrico Riley, PUT, Ming Meng, Brenda Garand, and Douglas Van Citters for their collaboration on a study of neural changes with visual art and other creative learning (6); and Tetsuro Matsuzawa and Ikuma Adachi for hosting me at the Kyoto University Primate Research Institute in Inuyama Japan while I studied mental rotation in chimpanzees.

Finally, I would like to thank my dissertation committee members for helping to shape this thesis, and especially Peter Tse for being a mentor and friend for the past 13 years.

Table of Contents

List of Tables

List of Figures

Ch. 1: Introduction

A hallmark of human cognition is the ability to create flexible mental representations through conscious effort. Such abilities have been studied via several psychological constructs including working memory (7), mental imagery (8), visuospatial ability (9), mental models (10), analogical reasoning (11), and the mental workspace (1). The ability to work flexibly with mental representations underlies much of human life from mundane tasks such as planning seating arrangements at family get-togethers to our species' greatest artistic and scientific achievements. For instance, Albert Einstein wrote that his scientific thought process consisted primarily of “certain signs and more or less clear images which can be 'voluntarily' reproduced and combined” (12). In contrast, chimpanzees, our closest living evolutionary relative, appear to lack fundamental aspects of our flexible cognitive machinery such as symbolic thought (13) and imagination (14, 15). How has the human brain enabled these extraordinary abilities?

In a seminal experiment on the mental manipulation of visual imagery, Shepard and Metzler (1971) had participants mentally rotate three-dimensional objects to determine whether they were the same as other similar three-dimensional objects (see Figure S4.1 for example stimuli). Participants' reaction times correlated strongly with the angle of rotation necessary to align the two objects, suggesting that they had mentally rotated an internal model in much the same way as one would rotate a physical object. Subsequent behavioral research explored other operations such as mental paper folding

(17), the generation and analysis of mental analog clocks (18), and mental simulations of mechanical systems (10), a primary result being that mental operations resemble the corresponding physical ones. Other work has documented similar processes in domains such as mental time travel (19), creative synthesis of mental imagery (20), and visuospatial reasoning (21). Thus, the human brain appears to support a mental space similar to the physical world in which mental representations can be constructed, manipulated, and tested in a flexible manner. Following Logie (1), I will refer to this cognitive system as the mental workspace.

A dominant model of the architecture of the mental workspace is Baddeley's conception of working memory (7, 22), in which a central executive system controls the maintenance and manipulation of representations in subsystems such as the visuospatial sketchpad (for visual representations) or the phonological loop (for verbal/auditory representations). Working memory is often treated as a cognitive system responsible for maintaining mental representations of limited size for short periods of time. Canonical tests of working memory capacity such as the n-back and memory span tasks (23–25) reflect this viewpoint: In general, participants must hold a variable number of items online (numbers, letters, words, images, etc.) for some amount of time, often accompanied by competing distractor stimuli. Working memory abilities are closely linked to control of attention (26) and intelligence (27). Recent work has established that working memory capacity can be improved with training and that this improvement can transfer to other abilities such as fluid intelligence (28–31).

While behavioral work on the mental workspace is well established, relatively little is understood about the neural mechanisms that make it possible. Neuroimaging studies have implicated widely-distributed regions of the cortex in working memory (28, 32), with both lateral-frontal and parietal cortical activity commonly co-occurring in working memory tasks. This fronto-parietal coupling has been proposed as the core of a network that mediates many higher order mental functions (33–36). According to several models, this network consists of a frontal executive system that directs attention over contents located in parietal and surrounding modality-specific regions (7, 37–41). Supporting this view, Harrison and Tong (42) could decode the contents of visual working memory in early visual areas, and Oh and colleagues (43) found that auditory imagery recruits frequency-specific regions of auditory cortex. In a meta-analysis of neuroimaging studies of mental rotation, Zacks (2008) found that regions throughout the cortex and cerebellum were involved but that many studies placed at least part of the machinery of mental rotation in the intraparietal sulcus and adjacent regions along with the medial superior precentral cortex. Zacks suggested that parietal regions maintain representations of the objects being rotated and that precentral motor cortex executes the motor simulations.

However, empirical support for such anatomically-modular models of the mental workspace derives in many cases from a failure to find (or look for) relevant information in regions outside those proposed by such models (7, 38, 40, 44, 45). Other models and mounting empirical evidence derived from new, network- and information-based analytical techniques paint a more complex picture, suggesting that many high-level cognitive processes occur at a level of organization that transcends any single neural structure (46–51). These recent advances suggest that processing in the brain may be much more distributed in nature than suspected previously. Therefore, one hypothesis of the studies presented in this thesis is that the mental workspace emerges out of the distribution and sharing of informational processes throughout the cortex. However, this emergent organization would only become apparent if studied using analytical methods that are sensitive to informational connections between widely distributed network nodes. A model of the brain as fundamentally distributed is not new: Rumelhart's and McClelland's (48) Parallel Distributed Processing proposal gained wide interest in the late 1980s and sparked a revolution in artificial intelligence based on neural networks and connectionist models of cognition. However, evaluating this model empirically has been severely limited by technical limitations in the measurement and analysis of brain activity. While functional magnetic resonance imaging (fMRI) enables the indirect measure of neural activity over the entire brain, interpreting fMRI data in a connectionist framework requires methods that have only recently begun to gain a foothold in the neuroimaging community. Early fMRI research was dominated by univariate analyses that limited inference to isolated voxels and encouraged efforts to localize function to such a degree that the field's early work has been termed "neo-phrenology" (52). When multivariate pattern analysis (MVPA) and related techniques such as representational similarity analysis (RSA) were introduced, they drove the field to reframe questions in terms of the informational roles and relationships of and between networks of brain regions (53–56). A parallel line of methodological research has sought to use existing ideas in network analysis to characterize the large-scale network structure of the brain (46, 49, 57). Recent advances in the analysis of structural and functional connectivity have allowed both undirected (58, 59) and directed (60, 61) relationships between widespread cortical and subcortical regions to be investigated. These methods have set the stage for a new approach to the brain as a densely interconnected, fundamentally distributed information processing network.

The studies in this thesis build on the above techniques to investigate the neural architecture of the mental workspace. In order to study the mental workspace as a distributed network, a combination of existing and novel methods is used. In particular, we develop and exploit new methods in the following areas:

- ***Relating classifier confusion to task structure.*** Multivariate classification techniques usually rely on classification accuracy—the degree to which a machine classifier can distinguish between experimental conditions above chance levels— as the measure of classifier performance (62). While this measure allows the investigator to probe whether brain activity supports information that distinguishes between experimental conditions, its interpretation can be uncertain or misleading when potential differences between conditions exist in addition to the difference of interest. For instance, differences in difficulty or attentional demands between two tasks could result in above chance classification accuracy, but failing to account for this possibility could lead the investigator to conclude that the successful classification was due to the designed contrast between conditions. However, classification analyses also yield confusion matrices that record the particular patterns of confusion between conditions. In designs with more than two conditions, these confusion matrices can be compared to the expected pattern of confusion that would occur due to the structure of informational relationships between conditions. Significant correlation between the confusion matrices and the expected or model similarity structure between conditions can potentially provide evidence for task-specific evidence that goes beyond potential confounding factors such as difficulty or attention. Each of the three studies presented in this thesis use this technique in order to provide robust evidence for task-specific processing in the investigated regions of interest.

- ***Shared vs. distinct informational formats.*** Multivariate techniques have revealed that information pertaining to many cognitive processes is distributed widely in the cortex (45, 63, 64). However, current techniques do not distinguish whether this information occurs in a common or distinct format between network nodes. This is an important distinction, since anatomically modular theories of the mental workspace rely in part on information being distributed yet specialized throughout the cortex (40). Study 2 (Ch. 3) develops a novel method to investigate the relationship between the informational formats of network nodes.
- ***Patterns of information flow.*** Understanding the flow of information between a network's nodes is necessary in order to understand how that network functions. Existing methods for assessing directed information flow are concerned primarily with quantifying the degree to which processing in one region influences later processing in another region (60, 61). In this sense, these methods are similar to univariate analyses in that they can detect increases or decreases in directed connectivity, but are insensitive to information that may be carried via connectivity patterns. In this dissertation we are not concerned directly with whether information flows between nodes, since in a densely connected, distributed network each node will likely exert some degree of control over all other nodes. Rather, we are interested in whether patterns of information flow between underlying processes that are

distributed among these nodes are informative about mental workspace functions. Study 2 develops a novel method to investigate the information carried by patterns of information flow between network nodes.

The following chapters present the results of three studies that investigate the distributed neural network underlying the mental workspace, focusing specifically on visual imagery because of its extensive existing literature. Study 1 (Ch. 2) investigates the structure and dynamics of the neural network that supports the mental manipulation of visual imagery. Study 2 (Ch. 3) investigates the distribution and flow of information in this network that supports the representation and manipulation of visual imagery. Study 3 (Ch. 4) uses mental rotation as a case study to investigate how the mental workspace recruits specialized subnetworks for specific functions.

Ch. 2: Network structure and dynamics of the mental workspace[1]

Abstract. 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 magnetic resonance imaging 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.

[1] This chapter was originally published as ref. (95).

Introduction

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” (12). Creative thought in science as well as in other domains such as the visual arts, mathematics, music, and dance requires the capacity to flexibly manipulate mental representations. Cognitive scientists refer to this capacity as a “mental workspace” and suggest that it is a key function of consciousness (65) involving the distribution of information among widespread, specialized subdomains (66).

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 (16), have found that many mental operations closely resemble their corresponding physical operations. This 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 (53), visual imagery (42), and even dreams (67) can be decoded from activity in visual cortex. These results suggest that the same regions that mediate representations in sensory perception are also involved in mental imagery. Yet, 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 (68, 69) including mental rotation (33), analogical reasoning (34), working memory (35), and fluid intelligence (36). 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 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.

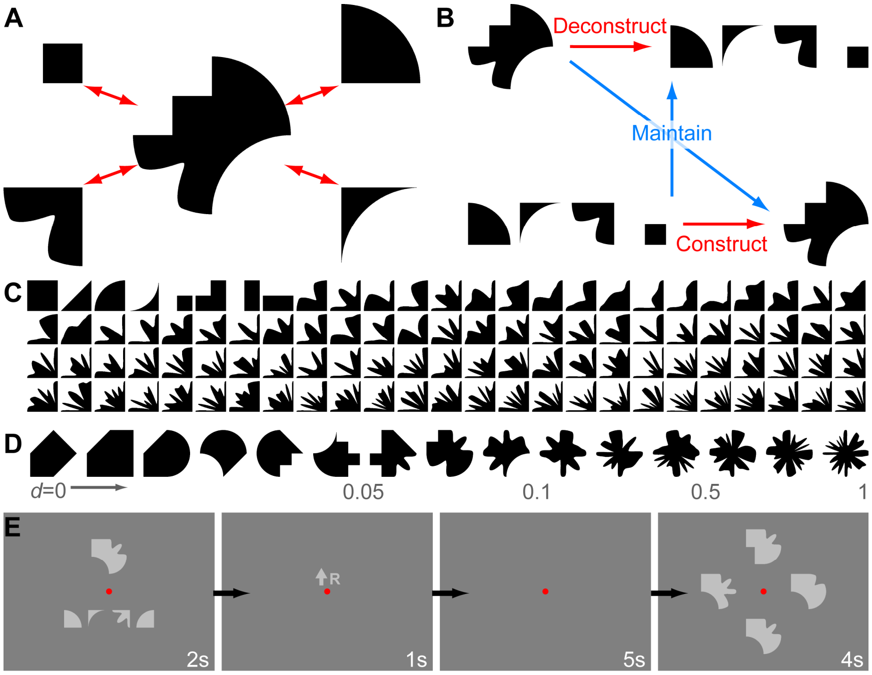


Figure 2.1. Experimental design

A. Parts could be constructed into 2 × 2 figures and figures 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 figures. Parts and figures ranged from simple to complex according to an index d. This allowed the task difficulty to be equated across conditions. **E.** Trial schematic. Trials begin with a figure and four unrelated parts presented for 2s, followed by a task prompt for 1s 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 5s delay period during which no stimulus is shown and the participant performs the indicated operation. Finally, a test screen appears for 2.5s. Either four figures or four sets of parts (depending on the task) are presented, and the participant indicates the correct output of the operation.

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 magnetic resonance imaging (fMRI) measurements of their neural activity. As stimuli, we developed 100 abstract parts that could be combined into 2 × 2 figures (Figure 2.1A & 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 (Figure 2.1B). Stimuli were presented briefly at the beginning of each trial, followed by a task prompt and a 6s 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 (Figure 2.1D). Adjusting the complexity of the stimuli allowed us to equate for task difficulty by maintaining 2/3 accuracy for each participant in each of the four conditions (chance would be 1/4 correct; Figure 2.1E).

Results

As an initial region of interest (ROI) selection procedure 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 (Figure 2.2), suggesting that a widespread network mediated the manipulation tasks. All but two of the ROIs (medial temporal lobe and medial frontal cortex) showed greater activation in manipulation than in maintenance conditions. 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 is expected, given that visual stimuli were equated across the 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. This gave us 12 ROIs to investigate for informational content relevant to the mental operations.

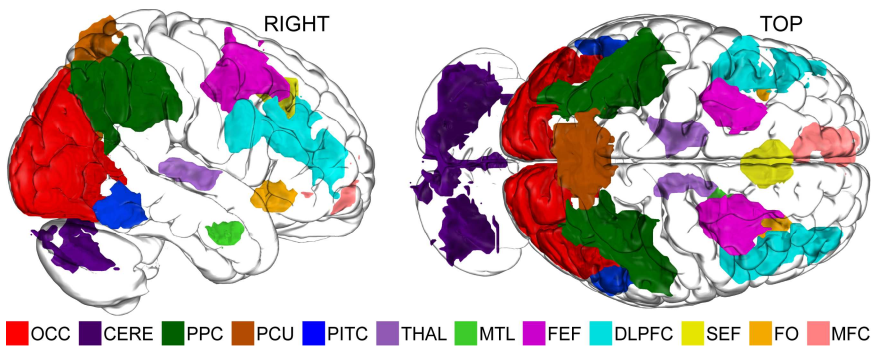


Figure 2.2. ROIs

11 ROIs showing differential activity levels between manipulation and maintenance conditions. An additional occipital cortex ROI was defined anatomically. ***Abbreviations.*** OCC: occipital cortex; CERE: cerebellum; PPC: posterior parietal cortex; PCU: precuneus; PITC: posterior inferior temporal cortex; THAL: thalamus; MTL: medial temporal lobe; FEF: frontal eye fields; DLPFC: dorsolateral prefrontal cortex; SEF: supplementary eye field; FO: frontal operculum; MFC: medial frontal cortex.

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 (53) within each ROI.

In this analysis, a classifier algorithm is first 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 is likely 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 Figure 2.3A. In order 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 (Figure 2.3B). 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 manipulation or 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 (Figure 2.3C). 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 due to ROI size or differences in response times between conditions (see Figure S2.1 & Table S2.1).

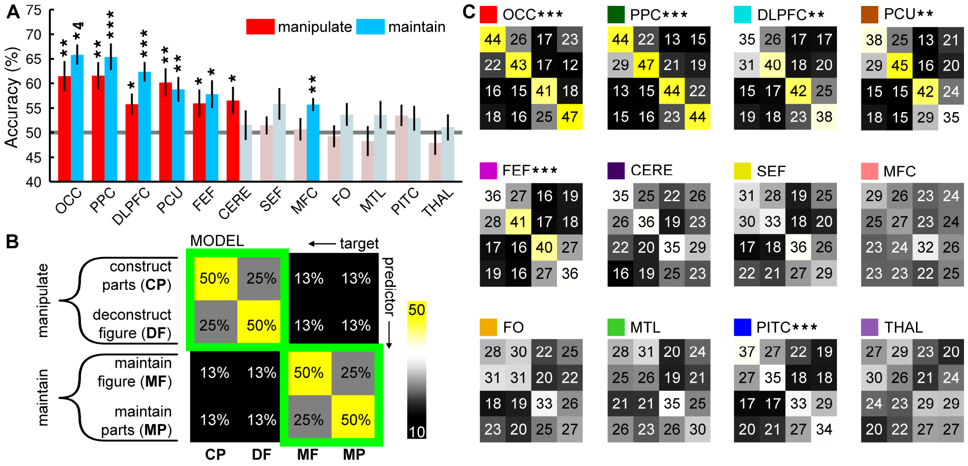
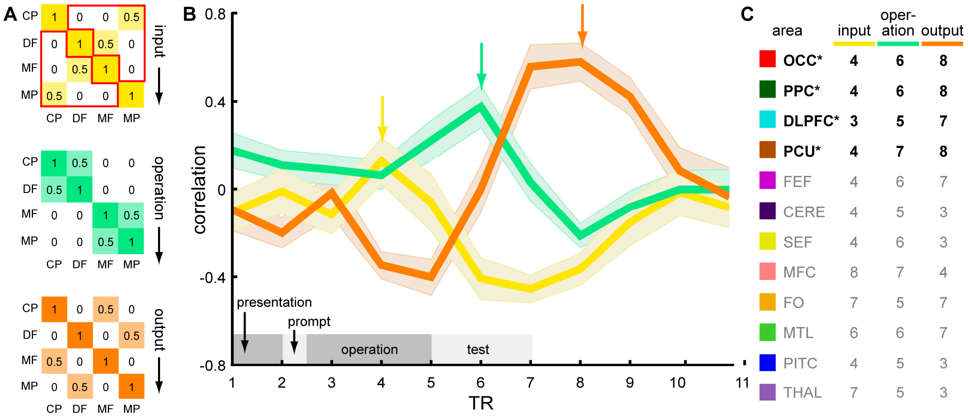


Figure 2.3. Multivariate pattern classification results

A. Results for two-way classifications in each ROI between manipulation conditions and between maintenance conditions. Bar plot shows classification accuracies, in descending order. Error bars are standard errors of the mean. Asterisks indicate accuracies significantly above chance ($p \leq 0.05$, false discovery rate [FDR] corrected across the 24 comparisons). Table S2.2 shows full statistical results. **B.** Model similarity structure for regions that mediate the mental operations. Manipulation and maintenance conditions should be more similar within 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 only considered values within manipulation and maintenance conditions in the correlation (within the green squares in part B). Table S2.3 shows full statistical results.

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 on that representation (construct, deconstruct, or maintain), and produced an output mental representation. Each of these stages entailed a unique relationship among the mental states associated with the four conditions (Figure 2.4A). 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 this, 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 Figure 2.4. 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. Figure 2.4B shows the time course of correlations with each model in occipital cortex. In Figure 2.4C, we report peak correlation times in each of the 12 ROIs. In the four regions with highest classification accuracies in Figure 2.3A, 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 (orange) correlation time course since 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.



	CP	DF	MF	MP
CP	1	0	0	0.5
DF	0	1	0.5	0
MF	0	0.5	1	0
MP	0.5	0	0	1

	CP	DF	MF	MP
CP	1	0.5	0	0
DF	0.5	1	0	0
MF	0	0	1	0.5
MP	0	0	0.5	1

	CP	DF	MF	MP
CP	1	0	0.5	0
DF	0	1	0	0.5
MF	0.5	0	1	0
MP	0	0.5	0	1

area	input	oper-ation	output
OCC*	**4**	**6**	**8**
PPC*	**4**	**6**	**8**
DLPFC*	**3**	**5**	**7**
PCU*	**4**	**7**	**8**
FEF	4	6	7
CERE	4	5	3
SEF	4	6	3
MFC	8	7	4
FO	7	5	7
MTL	6	6	7
PITC	4	5	3
THAL	7	5	3

Figure 2.4. Temporal progression of 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 while constructing parts and maintaining figures have similar output representations. 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. Error bars are standard errors of the mean. Schematic at bottom shows the trial stages. **c.** Peak correlation times for the 12 ROIs. In the four ROIs with highest classification accuracies in Figure 2.3, the peaks in the correlation time courses followed a significant sequence from input mental representation, through operation, to output representation (significant ROIs indicated with asterisks). Table S2.4 shows full statistical results.

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 evidence about how these regions communicate within the mental workspace network. We investigated this 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 (58). This 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 [$t(14) = 2.4$, $p = 0.029$]. This indicates 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 (Figure 2.5A). Connectivity increases between pairs with positive weights drove the classifier toward the manipulation conditions, while 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 medial temporal lobe indicated maintenance conditions. In Figure 2.5B, we plot the difference in functional connectivity between conditions. The precuneus and posterior inferior temporal cortex showed stronger connectivity with several frontal and parietal regions during manipulation conditions while connectivity between the medial temporal lobe 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.

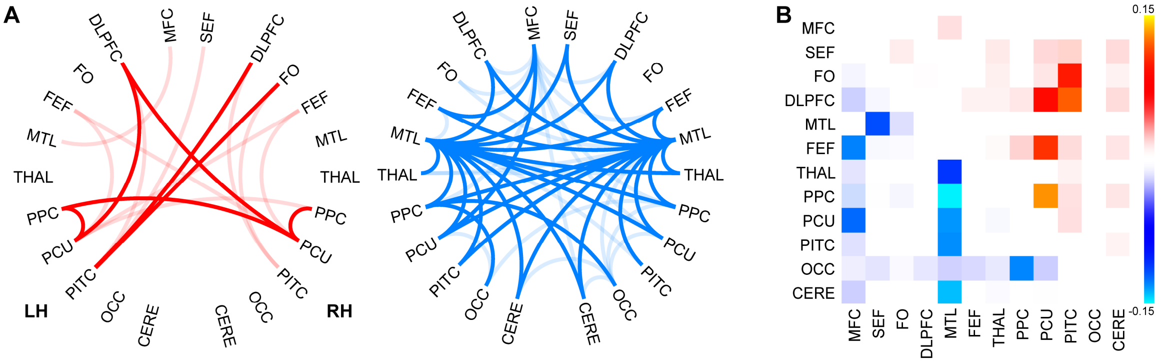


Figure 2.5. Multivariate pattern analysis of functional connectivities

A. Sensitivities for each pair of ROIs in a between-subject classification of functional connectivity between manipulation and maintenance conditions. Red sensitivities are positive, driving the classifier toward choosing “manipulate.” Blue sensitivities are negative, driving it toward “maintain.” Only significant non-zero sensitivities are shown [$p \leq 0.05$, corrected for similarity between folds (70)]. Saturated colors indicate sensitivities that survived FDR correction across the 231 comparisons. **B.** Difference in functional connectivity between manipulation and maintenance conditions. Positive and negative differences are separated into the upper and lower diagonals, respectively. Only significant connectivity differences are shown ($p \leq 0.05$), and differences surviving FDR correction are shown saturated.

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 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. An extended network of regions is therefore likely 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 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 (33–36). Miller and colleagues 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 (71). Thus, the DLPFC appears to be part of a network that maintains representations in working memory via attention. Human neuroimaging studies have shown that DLPFC and PPC are both activated regardless of the type of information that is held in working memory (72, 73). Selectivity for task rather than representation distinguishes this system from subsidiary systems that are capable only of maintaining particular classes of information (74). These findings support the view that the frontoparietal network is an executive system that recruits subsidiary systems, as proposed in Baddeley’s (75) model of working memory. Modeling work by O’Reilly and colleagues (68, 69) has shown how prefrontal cortex may be able to flexibly self-organize abstract rules and later apply them to specific representations. This ability is common to many flexible cognitive processes in humans such as analogical reasoning, creativity (34), and fluid intelligence (36). Our data provide empirical support for this model by showing that the DLPFC and PPC mediate not just 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 occipital cortex.

Occipital cortex. Several studies have found that the occipital cortex processes information relevant to internally-generated visual experience. Harrison and Tong (42) used patterns of activity in early visual cortex to decode the orientation of gratings that participants maintained in working memory. Recently, Horikawa and colleagues (67) decoded the contents of participants' visual experience while 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 are not only formed but also operated on in visual cortex. This may generalize to other sensory domains, such that the brain mediates perceptual processes and operates on the corresponding mental representations in the same regions.

Precuneus. Margulies and colleagues reported that the precuneus in humans is functionally connected to lateral frontal, posterior parietal, and occipital cortices (76). 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 (77) cite a body of evidence that the precuneus is involved in visuospatial imagery, is relatively larger in humans than in non-human primates and other animals, and is one of the last regions to myelinate during development. Consistent with these findings, Vogt and Laureys (78) 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 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 exclusively involved in motor coordination, is now known to connect strongly to prefrontal and posterior parietal cortices and to mediate attentional processes (79). Posterior regions of the inferotemporal cortex are involved in visual object processing (80). The thalamus is a hub for interaction between cortical areas and may play a critical role in consciousness (47). The medial temporal lobe (MTL) is a hub in memory formation and retrieval (81). This 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 (82). Recently, Higo and colleagues (83) showed that the frontal operculum controls attention toward occipito-temporal representations of stimuli held in memory. And the medial frontal cortex is a hub in the default mode network that plays a role in self-directed attentional processes (84). Thus, all of these regions are likely involved in the mental operations performed by participants.

A significant new 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, while manipulation of those representations recruits a sparse, slightly left-lateralized network with a hub in the posterior precuneus. Whereas 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 suggests that these hubs serve distinct functions across the tasks. The MTL appears to bind network components together, while 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 or have been able to classify between maintenance and manipulation in certain areas (85–87) but have not shown that these areas are responsible for the manipulations themselves. An alternative explanation of these findings could merely be that attentional allocation is increased during manipulation over maintenance tasks. A major advance of the current study is the investigation of neural activity in two qualitatively distinct types of manipulations. We showed that a subset of areas in the mental workspace network contains information specific to particular manipulations. We additionally showed that the task-related informational structure of these areas evolves over time in accordance with the manipulations performed. This provides novel and 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 (65) 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. While 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. 16 participants (6 females, aged 19-30 years) gave informed written consent according to the Institutional Review Board guidelines of Dartmouth College prior to 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. 100 abstract parts served as the stimulus set (Figure 2.1C). The first eight parts were manually defined. Each subsequent part was generated by randomly perturbing a quarter circle while fixing the endpoints. For each part, 1000 shapes were randomly generated 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 (Figure 2.1A). A difficulty index *d* that scaled from 0 to 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 were always displayed in a horizontal row, rotated into the correct orientation such that, if constructed into a figure, they would be ordered clockwise starting with the upper right quadrant. During each 12s 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 2s, the stimulus disappeared and was replaced for 1s by a task prompt consisting of either an upward or downward facing arrow and the letter “C”, “D”, or “R”. The arrow indicated which of either the figure or parts would be used in the task, and the letter indicated the operation to perform on the stimulus (C: construct; D: deconstruct; R: remember). The participant then had 5s to perform the operation, during which 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 2.5s. 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 4s 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 2/3 of 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 (2000ms TR, 20ms TE; 90° flip angle, 240 × 240mm FOV; 3 × 3 × 3.5mm voxels; 0mm slice gap; 35 slices). Structural images were acquired using a T1-weighted magnetization-prepared rapid acquisition gradient echo sequence (8.176ms TR; 3.72ms TE; 8° flip angle; 240 × 220mm FOV; 188 sagittal slices; 0.9375 × 0.9375 × 1mm voxels; 3.12min duration). Participants completed 10 functional runs. Each run consisted of 16 trials interleaved with 10s blanks, giving 40 trials for each condition. The *d* value was updated on each trial so that participants achieved 2/3 accuracy for each trial type.

MRI preprocessing. fMRI data were preprocessed using FSL (88). Data were motion and slice-time corrected, high-pass filtered with a 100s cutoff, and spatially smoothed with a 6mm FWHM Gaussian kernel. Structural images were processed using the FreeSurfer image analysis suite (89).

ROI selection procedure. A whole brain GLM 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 MNI 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 were then thresholded at $p \leq 0.05$ according to Gaussian Random Field theory (88). This analysis yielded 11 bilateral ROIs that were then 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 (MVPA): Classification. MVPA was carried out using PyMVPA (90). Spatiotemporal 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 4s 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 to chance (50%) and false discovery rate (FDR) corrected across the 24 comparisons (one for each ROI and classification). A four-way classification was also carried out in each ROI to produce the confusion matrices in Figure 2.3C. Correlation analyses were carried out between each of these confusion matrices and the model similarity structure in Figure 2.3B. 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 produced a confusion matrix for each time point, and these were correlated with each of the model similarity structures in Figure 2.4A. 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 Figure 2.4C. These calculations were restricted to TRs 3-8, representing the pre-test portion of the trial, HRF shifted by 4s. 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, Patterson, and Ulrich for latency analyses (70). In a jackknifed analysis with N subjects, N grand means of the data are calculated, each with one subject left out. The analysis is then performed on these grand means with corrections applied for the jackknife-induced decrease in variance. In the case of noisy estimates such as occurs when calculating latencies from single-subject time courses, this procedure provides cleaner results while not biasing estimates of significance. For each ANOVA we defined two orthogonal linear contrasts (C1 = -1/-1/2 [input/operation/output]; 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 (58), 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 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 Figure 5A are significantly different from zero in a one-sample *t*-test, corrected for the artificially low variance due to similarity between folds (70) and thresholded at $p \leq 0.05$. Saturated colors indicate sensitivities that survived FDR correction across the 231 comparisons (one for each connectivity). Differences are thresholded at $p \leq 0.05$ in a one-sample *t*-test. Saturated colors again show differences that survived FDR correction.

Ch. 3: Information processing in the mental workspace is fundamentally distributed[2]

Abstract. The brain is a complex, interconnected information processing network. In humans, this network supports a mental workspace that enables high-level abilities such as scientific and artistic creativity. Do the component processes underlying these abilities occur in discrete anatomical modules or are they distributed widely throughout the brain? How might the flow of information within such a network support specific cognitive functions? Current approaches have limited ability to answer such questions. Here we report novel multivariate methods to analyze information flow within the mental workspace during visual imagery manipulation. We find that mental imagery entails distributed information flow and shared representations throughout the cortex. These findings challenge existing, anatomically modular models of the neural basis of higher-order mental functions, suggesting instead that such processes may occur at a fundamentally distributed level of organization. The novel methods we report may be useful in studying other similarly complex, high-level informational processes.

[2] This chapter is currently under review for publication (108).

Introduction

A hallmark of human cognition is the ability to volitionally construct and flexibly manipulate mental representations. Such abilities have been studied using several overlapping psychological constructs including working memory (7), mental imagery (8, 91), visuospatial ability (9), mental models (10), analogical reasoning (11), and the mental workspace (1). In general, these terms denote the ability to work flexibly with mental representations, a skill that underlies much of human life from mundane tasks such as planning seating arrangements at family get-togethers to our species' greatest artistic and scientific achievements. For instance, Albert Einstein wrote that his scientific thought process consisted primarily of "certain signs and more or less clear images which can be 'voluntarily' reproduced and combined" (12). Here we will use Logie's term "mental workspace" to refer to the mental space in which these flexible cognitive processes occur.

How does the human brain support the mental workspace underlying flexible and creative mental phenomena such as mathematical, scientific, and artistic thought (1)? Understanding how the brain enables the imaginative abilities of the mental workspace is an important goal for many fields (92, 93), and several models have proposed potential mechanisms (1, 7, 38, 47, 48, 94). Previous research has shown that manipulating visual imagery in the mental workspace recruits a neural network extending throughout the cerebral cortex and associated structures (95). An important question to answer of such a network is whether the component processes underlying the network's function occur in anatomical modules or via a fundamentally distributed level of organization that transcends anatomical boundaries. However, our ability to measure and analyze complex informational processes that are distributed widely in the human brain remains underdeveloped, and thus such questions are currently difficult to answer (39, 60, 61).

Manipulation of visual imagery requires multiple component processes including a) forming a mental representation of an image and b) performing an operation to manipulate that representation. Several current, standard models propose that the different functional units of this network correspond to anatomically distinct neural structures. For instance, the ‘central executive’ in Baddeley’s model of working memory has been proposed to reside in dorsal lateral prefrontal cortex (DLPFC) and direct the formation of mental representations in modality specific regions such as visual cortex for the ‘visuospatial sketchpad’ or auditory cortex for the ‘phonological loop’ (7, 39–41, 44). Likewise, Postle argues that prefrontal cortex is not involved in the representation of working memory contents; instead, his model states that mental representations are processed exclusively by domain-specific sensory- or action-related regions (38). Thus, while these models hold that working memory and related abilities may recruit a “distributed” neural network in the sense that the complex functions of the network are mediated collectively by anatomically widespread regions, the component processes that constitute those complex functions are relegated to anatomically distinct modules. In many cases, empirical support for the anatomical modularity of these models derives from a failure to find (i.e. acceptance of null results) or often even look for relevant information in regions outside of those that the models propose (7, 38, 40, 44, 45). For instance, both Ishai and colleagues (44) and Lee and colleagues (40) found information pertaining to the visual but not the non-visual aspects of working memory representations in extrastriate visual cortex and found the opposite for lateral prefrontal cortex. Both groups interpreted this to mean that extrastriate visual cortex processes visual aspects of working memory tasks but not non-visual aspects, and vice versa for lateral prefrontal cortex. While such conclusions are a common practice in the field, they amount to acceptance of null results regarding the information that was not detected in each respective area and are thus in danger of failing to account for information that may have been present but that was not detected by their methods. Baddeley's anatomically localized model of working memory similarly relies on studies that either did not find or did not look for relevant information outside of hypothesized regions (7).

However, mounting empirical evidence derived from new, network- and information-based analytical techniques paints a more complex picture, suggesting that many high-level cognitive processes occur at a level of organization that transcends any single neural structure (46–51, 96). We therefore hypothesized that the mental workspace emerges out of a fundamentally distributed sharing of informational processes throughout the cortex. This hypothesis runs contrary to modular accounts that claim information is segregated to specific anatomical regions, such as visual information occurring only in visual cortex or executive processing occurring only in prefrontal cortex (7, 38, 40, 44). However, we predicted that this emergent organization would only become apparent if studied using analytical methods that are sensitive to informational connections between widely distributed network nodes.

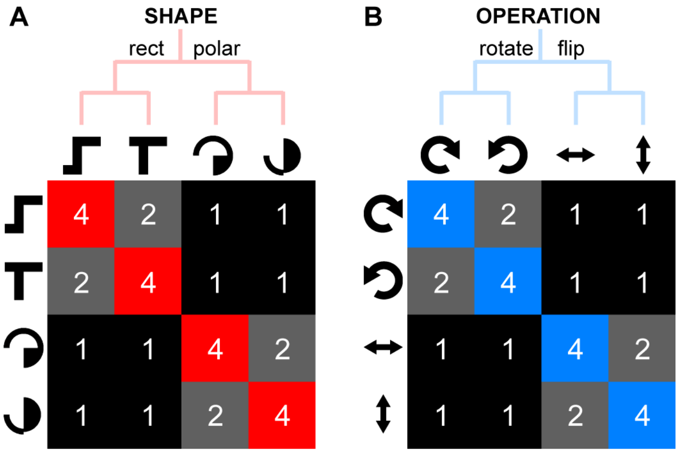


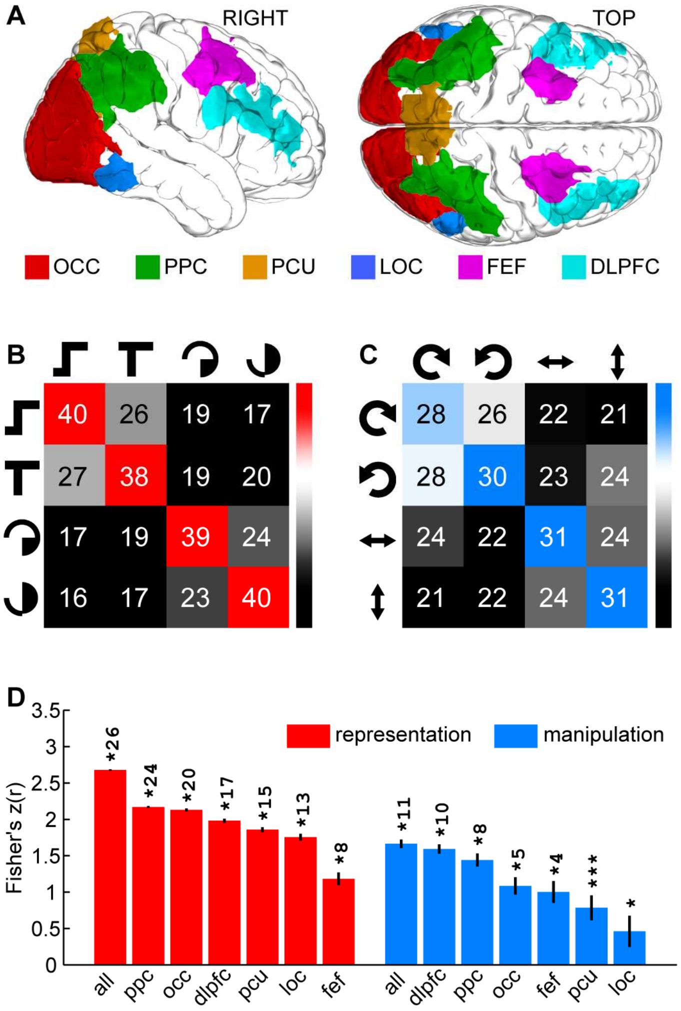
Figure 3.1. Experimental design

A. The four shapes used in the experiment, related in a two-level hierarchical structure. Two shapes were derived from a 4 × 4 rectangular grid, and two shapes were derived from an analogous polar grid. At bottom is a similarity structure that represents the matrix form of the hierarchy. Each shape is more similar to itself than it is to any other shape, and each rectangular shape is more similar to the other rectangular shape than it is to either polar shape (and vice-versa). See ref. (95) for details on the particular values used in the similarity structure. **B.** The four mental operations used in the experiment, also related in a two-level hierarchy: 90° clockwise rotation, 90° counterclockwise rotation, horizontal flip, and vertical flip. The similarity structure for operations is constructed in the same manner as for shapes.

To evaluate our hypothesis and investigate how the mental workspace network functions in both the representation and manipulation of visual imagery, we used functional magnetic resonance imaging (fMRI) to record cortical activity as participants performed a series of trials involving the mental manipulation of shapes maintained in working memory. During each trial, participants recalled one of four abstract shapes memorized previously (Figure 3.1A) and performed one of four mental operations on that shape (clockwise 90° rotation, counter-clockwise 90° rotation, horizontal flip, or vertical flip; Figure 3.1B). To support the functional analyses described below, the shapes were related in a two-level hierarchy of similarity (see Figure 3.1A). The operations shared an analogous relationship (see Figure 3.1B). In order to ensure that neural activity associated with the shapes and operations was due to visual imagery rather than the presented visual stimuli, we constructed a unique mapping for each participant from shapes to letters and from operations to numbers. Each trial occurred as follows: At the start of a trial, four letter/number pairs (e.g. “C3”) appeared for 2s, with an arrow pointing to a single pair to indicate the shape and operation for the current trial. The other three pairs were shown as a visual control to ensure that any successful classification analyses were due to mental imagery rather than the visual stimuli. After a 6s period during which the participant performed the indicated mental operation, four shapes at various orientations appeared on the screen for 2s. One of these was the shape indicated previously, while the other three shapes again served as a visual control. The participant indicated whether the displayed shape was at the orientation that would result from the indicated operation and was then given feedback regarding whether their choice was correct or incorrect (see Figure S3.1 for a trial schematic).

Our analyses of the task-related fMRI data used a combination of existing and novel multivariate methods to investigate the informational structure of the network underlying the mental workspace. First, we performed ROI classification analyses with trials labeled based on either shape or operation, to determine the regions in which cortical activity supported information about mental representations and/or mental manipulations. Second, we developed a novel ROI cross-classification analysis to determine whether this information was shared between regions. Third, we developed a novel classification analysis on patterns of information flow between cortical regions to determine how information related to the task was transferred between regions. Detailed descriptions of our analytical methods are presented in Materials and Methods and in Figure S3.2 and Figure S3.3, and summaries are given below.

Figure 3.2. ROI classification results



B

40	26	19	17
27	38	19	20
17	19	39	24
16	17	23	40

C

28	26	22	21
28	30	23	24
24	22	31	24
21	22	24	31

A. The six bilateral ROIs used in the current experiment, derived from the results of a previous study (see Materials and Methods). OCC: occipital cortex; PPC: posterior parietal cortex; PCU: precuneus; LOC: lateral occipital cortex; FEF: frontal eye fields; DLPFC: dorsolateral prefrontal cortex. **B.** Mean confusion matrix from a four-way classification among mental representations across the entire mental workspace network (compare to Figure 3.2A). **C.** Analogous confusion matrix from a classification among mental manipulations (compare to Figure 3.1B). **D.** Results of four-way classification analyses in each ROI. Correlations between resulting confusion matrices and similarity structures in Figure 3.1 are Fisher's Z-transformed. Error bars are jackknife-corrected standard errors of the mean (see Methods). Asterisks indicate significance in a one-tailed jackknifed t-test comparing Fisher's *Z*-transformed correlations to zero across subjects (*: *p* <= 0.05; ***: *p* <= 0.001; $*^{(n)}$: *p* <= 1 x 10^{-n}). Results are false discovery rate (FDR) corrected for multiple comparisons across the seven ROIs.

Results

Performance accuracy was high after an initial training session during which participants memorized the shapes, operations, and corresponding letter and number mappings (responses were correct in 95.8% of trials across participants and conditions). One-way between-subjects analyses of variance showed no significant differences in accuracy across conditions, confirming that the difficulties of shapes and operations were well matched [for shapes: $F(3,72) = 1.65$, $p = 0.185$; for operations: $F(3,72) = 0.369$, $p = 0.775$; see Table S2.1 for behavioral results].

ROI Classification Analysis. Our regions of interest (ROIs) for analysis of the fMRI data were the six bilateral cortical regions that contained information pertaining to the transformation of visual imagery in a previous study that used data independent from those of the current study (Figure 3.2A; see Materials and Methods for details on how these ROIs were defined) (95). Each area has been shown to play a role in neural processing related to the current task (33, 39, 42, 76, 80, 82). We used multivariate decoding methods to determine whether each ROI supported information about mental representations and/or mental manipulations of visual imagery, i.e. whether patterns of neural activity in each ROI could be used to classify either the shape that was represented in visual imagery during each trial or the operation that was used to manipulate that representation.

Because of the hierarchical relationship that we introduced among shapes and operations, we measured classifier performance using a representational similarity analysis in which we correlated the confusion matrix resulting from each four-way classification with the matrix form of this hierarchical similarity structure (Figure 3.1A &

B) (56, 95). This measure allowed us to use information from both correct classifications (classification ‘hits’) and specific patterns of confusion (classification ‘misses’) between conditions that resulted from the relationships among shapes and among operations. Thus, classification was only “successful” if the classifier performed according to our hypothesized pattern of correct-classification and confusion, allowing us to verify that our results were not due to task-irrelevant factors such as the letters or numbers used in the task mapping.

Initial classifications using the union of all ROIs confirmed that the information processing structure of this network matched precisely the similarity structures of both shape and operation sets [Figure 3.2B & 2C; for shapes: t(18) = 106., p = 8.59 x 10-26; for operations: t(18) = 16.0, p = 4.54 x 10-12; results are false discovery rate (FDR) corrected for multiple comparisons]. This result also held true for classification analyses performed on each ROI separately (Figure 3.2D; FDR corrected for multiple comparisons across the seven total analyses for each classification scheme). Because all of our results were significant, we verified the specificity of our analysis by conducting control classifications using two additional masks. The first was a functionally-defined, bilateral thalamus ROI from our previous study that showed increased but not task specific activity during mental manipulation of imagery compared to maintenance of imagery; the second was an anatomically-defined ventricle mask. None of the four control classification analyses using these masks reached significance, confirming that our original analyses detected information about the shapes and operations specifically within our six ROIs (see Table S3.2 for ROI control analysis results). As a further control to confirm that our analysis was valid and unbiased, we shuffled the labels randomly in each classification and found that the correlations between confusion matrices and model similarity structures were no longer significant (Table S3.3). Thus, neural activity in each ROI supported information about both representation and manipulation of visual imagery. This result provides evidence that processing of both representations and manipulations is distributed throughout the mental workspace network, running counter to models such as Baddeley's or Postle's that propose that its component processes are segregated to particular cortical regions (7, 38, 40, 44, 45). The large effect sizes and specificity of our results underscore the sensitivity of our experimental design and RSA-based analysis for uncovering information that other techniques such as univariate analyses or two-way classifications may have missed.

ROI Cross-classification Analysis. Information about both representations and manipulations thus appears to be distributed throughout the mental workspace network, but what format does this information take? Our hypothesis states that information is shared commonly throughout the network. However, alternative proposals state that each network node specializes in a unique informational aspect of representation and manipulation. For instance, Lee and colleagues (40) suggest that whereas visual cortex represents image-level information, information in prefrontal cortex is conceptual in nature. To evaluate these alternatives, we developed a novel multivariate cross-classification analysis to investigate whether information is shared among the nodes of the mental workspace network. In this analysis, we trained a classifier on data from one ROI and tested it on data from another ROI. A successful classification using this procedure would provide strong evidence for a shared informational format between regions, rather than the alternative possibility that both ROIs support information about the task but in independent formats. However, ROIs are incompatible as voxel-based features spaces, presenting a technical hurdle to cross-classifying because cross-classification requires patterns to share the same feature space. In other words, cross-classification would require the feature space of each ROI to have identical dimensionality and each feature of one ROI to carry the same meaning as the corresponding feature in the other ROI. Thus, we first needed to transform each ROI's data into a common feature space before we could perform the cross-classification analysis.

Conceptually, we hypothesized that the functional data for a given ROI were a set of signals in voxel-space that represented a mixture of a number of underlying informational subprocesses that were shared in a distributed manner between the ROIs. If this characterization is valid, then principal component analysis (PCA) would allow us to transform our voxel-based data independently for each ROI in order to recover a set of principal component signals that represented the underlying subprocesses that were mixed between the voxel-space signals that we actually measured. We therefore used PCA to convert the voxel-based data from each ROI into 50 principal component signals. We chose the number 50 in order to construct classification patterns of sufficient size while remaining smaller than the size of our smallest ROIs (e.g. the lateral occipital ROI); however, we did not test whether this was the optimum dimensionality to use. This step allowed us to establish feature spaces for the ROIs that had uniform dimensionality. The second step required to construct a common feature space for cross-classification was to rearrange the dimensions of these feature spaces such that corresponding features carried the same informational meaning across ROIs. To achieve this for each cross-classification between two ROIs, we performed a pairwise matching of component signals between the two ROIs in order to maximize the total correlation between matched component signal pairs (i.e. so that each component signal from the first ROI was matched to a maximally similar signal from the second ROI). We performed this matching step independently for each fold of the cross-classification, leaving out data from the testing set in order to avoid artificially inflating the similarity of test patterns across the two ROIs.

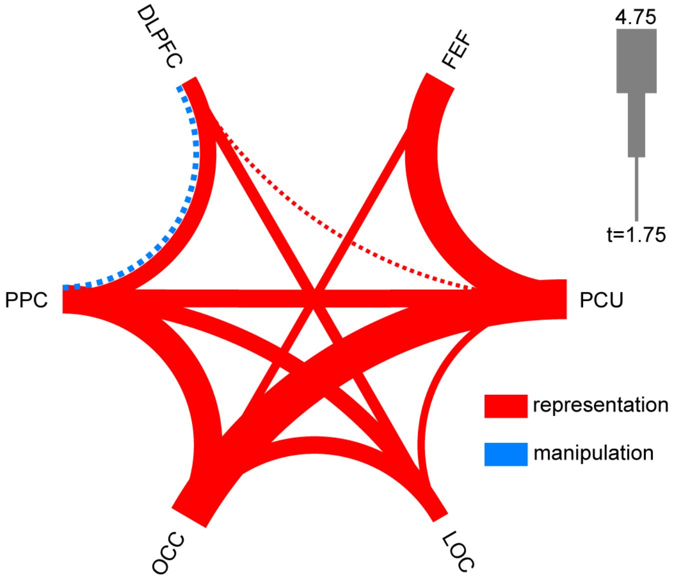


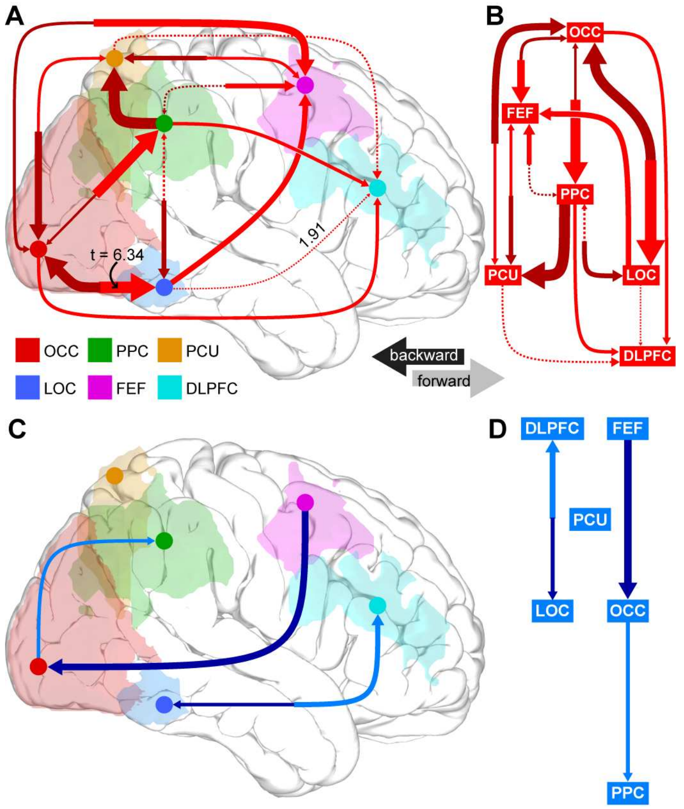
Figure 3.3. ROI cross-classification results

Arcs indicate pairs of ROIs in which cross-classification was successful. Dotted arcs indicate classifications that were significant but did not pass FDR correction across the 15 ROI pairs. All other displayed classifications passed FDR correction. Arc thickness indicates t-statistic values in a one-tailed, jackknifed t-test of Fisher’s Z-transformed correlations between confusion matrices and model similarity structures, compared to zero (see text). Abbreviations are as in Figure 3.2.

This two-step process yielded a common 50-dimensional feature space for each fold of each cross-classification analysis (see Figure S3.2 for a visual schematic of the procedure). Classification then proceeded exactly as in the previous ROI-classification analysis. We cross-classified between each pair of ROIs, with results presented in Figure 3.3 (all results FDR corrected across the 15 ROI pairs). We could successfully cross-classify mental representations between most pairs of ROIs, providing evidence that information about mental representations held in visual imagery is shared widely in a common format throughout the mental workspace network. The cross-classification of mental manipulation was significant only between DLPFC and PPC [$t(18) = 1.93$, $p = 0.0346$ (uncorrected)]. However, this result did not hold after FDR correction. This result suggests that information about manipulations of visual imagery is distributed but may be more compartmentalized in the network, with DLPFC and PPC possibly sharing some information. As in the ROI classification above, we confirmed the validity of the analysis by performing control analyses in which labels were shuffled randomly. In this case, the cross-classifications were no longer significant, ruling out the possibility that our cross-classification results occurred due to unknown biases introduced by our analysis pipeline (Table S3.4). Thus, information about mental representations is not only distributed throughout the network, but is also shared in a common format between many network nodes, while information about mental manipulations may be more compartmentalized but shared between some nodes.

Information Flow Classification Analysis. In order to investigate how this information becomes shared, we developed a new method to analyze whether information pertaining to visual imagery representations and manipulations was carried in condition-specific patterns of directed connectivity between pairs of network nodes. In other words, this analysis abstracted away from information contained in patterns of activity *within* neural regions, seeking instead to probe the informational content of patterns of information flow *between* neural regions. Established methods for assessing directed connectivity are concerned primarily with determining whether processing in one region is predictive of later processing in another region (60, 61, 97). In this sense, these methods are analogous to univariate analyses in that they can detect increases or decreases in directed connectivity, but are insensitive to information that may be carried via patterns of such connectivity. Because of this limitation, two processes (e.g. clockwise and counterclockwise mental rotation) may entail distinct patterns of directed connectivity without involving differing overall magnitudes of directed connectivity, and would thus be indistinguishable by current methods. Furthermore, in the present analysis we were not concerned directly with *whether* information flowed between nodes, since in a densely connected, distributed network each node will likely exert a complex pattern of control over all other nodes. Rather, we wanted to know whether the condition-specific *patterns* of directed connectivity between the underlying informational processes that were distributed among these nodes supported information about specific mental representations and manipulations. If so, then the current analysis would provide further evidence for the findings of the previous two analyses that the information processing underlying the manipulation of mental imagery occurs at a fundamentally distributed level of organization in the cortex.

Figure 3.4. Information flow classification results



A. Graph indicating directed ROI pairs in which patterns of information flow could be used successfully to classify mental representation. Dotted arrows indicate classifications that were significant but did not pass FDR correction across the 30 directed ROI pairs. All other displayed classifications passed FDR correction. Arrow thickness indicates *t*-statistic values in a one-tailed, jackknifed t-test of Fisher's *Z*-transformed correlations between confusion matrices and model similarity structures, compared to zero (see text). Light red arrows indicate posterior to anterior connections and dark red arrows indicate anterior to posterior connections. The greatest effect occurred for the OCC to LOC connection and the smallest significant effect occurred for the LOC to DLPFC connection. Both effect sizes for these two connections are indicated on the graph for reference. Abbreviations are as in Table S3.2. **B.** A topological sorting of the graph from panel A reveals that the OCC resides at the top of a bottom-up hierarchy of information flow for mental visual representations. **C.** Graph indicating directed ROI pairs in which patterns of information flow could be used successfully to classify mental manipulation. Arrow properties are as in panel A. **D.** A topological sorting of the graph from panel C reveals that the DLPFC and FEF reside at the top of a top-down hierarchy of information flow for mental manipulations of visual imagery.

As directed connectivity patterns we used Granger-causal graphs (GC-graph) constructed independently for each unique task condition (61). Granger-causality is a statistical method for evaluating the ability of a source signal to predict the future of a destination signal beyond the predictive power provided by the destination signal's own past. While the validity of Granger-causality for fMRI data has come under scrutiny, computational and empirical work has shown that it is a viable technique when proper precautions such as those used in the present study are taken (61, 98, 99). Specifically, we investigated differences in patterns of Granger-causality between conditions rather than attempting to establish "ground-truth" connectivity between regions. Our GC-graphs were constructed as follows: First, voxel-based data from each ROI were transformed individually using PCA into 10 principal component signals, with the same rationale as described above for the cross-classification analysis. We used 10 components here instead of 50 so that our resulting GC-graphs would have a reasonable dimensionality for classification, but we again did not evaluate the optimal dimensionality to use. Next, we constructed a 10 × 10 GC-graph for each of the 16 unique task conditions (e.g. shape 1 + clockwise rotation), each participant, and each directed pair of ROIs (e.g. from PPC to DLPFC). Each GC-graph was constructed by computing the Granger-causality from each of the 10 principal components in the source ROI to each of the 10 principal components in the destination ROI, using only data from the single task condition. For each participant, task condition, and directed pair of ROIs this process yielded a pattern of directed connectivity (the GC-graph) that represented a task-specific, directed pattern of information flow between regions. We then used these GC-graphs as inputs to classification analyses as described above, with results presented in Figure 3.4 (Figure S3.3 provides a visual schematic of this analysis). Complementing our ROI classification and ROI cross-classification results, we found that frequently bidirectional patterns of directed information flow between many nodes of the mental workspace network could be used to classify shape representations. A topological sorting of the resulting directed graph of significant classification results revealed a posterior to anterior hierarchy for mental representations, with the OCC at the top and connectivity cascading down to the DLPFC (i.e. a bottom-up hierarchy; Figure 3.4B). The pattern of results for the manipulation classification shows a sparser graph, with the DLPFC and FEF at the top of an anterior to posterior hierarchy (i.e. a top-down hierarchy; Figure 3.4D). Here, being placed at the top of the hierarchy indicates dominance in the sense that a higher node supports more information in outward flowing rather than inward flowing directed connectivity patterns. As in the previous two analyses, we performed control classifications with shuffled labels, confirming the validity of the analysis (Table S3.5).

Discussion

The mental workspace is a cognitive system that enables the volitional, flexible mental operations underlying the mathematical, scientific, and artistic creativity that distinguish humans as a species (1, 65). Here we applied novel network-level pattern analysis methods to reveal the structure of information flow in the neural network that supports the mental workspace. We find that the component processes of representing and manipulating visual imagery entail a level of informational organization that transcends the anatomically-modular structures that standard models of working memory and related processes have regarded as functionally encapsulated modules. Instead, our data imply that such processes emerge out of the fundamentally distributed sharing and flow of information between the nodes of a cortex-wide network. We found that representations entail the sharing and flow of information between all of the ROIs we tested. Mental manipulations showed patterns of information flow between all but one of our ROIs, but we did not find significant sharing of information at the scale of our fMRI data after correcting for multiple comparisons. It is important to note, however, that further information sharing and flow could have occurred at spatial or temporal scales or levels of information processing to which our data or analyses were insensitive. Because fMRI data are temporally low-pass filtered by the hemodynamic response function, our data can only address information flow that occurs on the scale of seconds. Nonetheless, our findings call into question ‘textbook’ anatomically-modular models of the neural basis of working memory and other higher order mental functions (7, 38, 40, 41, 45).

Existing neural models of working memory and related processes could be described as “distributed” in the sense that they assign the component functions of working memory to anatomical modules that are distributed across the brain. However, a key advance in the present study is to suggest that even these component processes that underlie the more complex functions we studied are distributed in the brain. Thus, contrary to models such as Baddeley’s that localizes executive functions to lateral prefrontal cortex and the storage of visual representations to occipital cortex, our data suggest that informational processing in the mental workspace is fundamentally distributed. In other words, anatomy may be incidental for the high-level mental functions studied here, with the actual functional separation of processes occurring at a higher level of informational organization.

Our work advances recently developed analytical techniques that approach the brain as an information processing network. Multivariate classification and representational similarity analyses allow the informational structure of processes at many levels of organization to be probed (55, 56). Directed connectivity measures enable the investigation of effective functional coupling between network nodes (60, 61, 100–102). Here, we adapted these techniques to answer two new kinds of question. First, our ROI cross-classification analysis was able to evaluate whether information is shared between multiple network nodes. Note that traditional RSA analyses as proposed by Kreigeskorte (56) are not able to answer this question generally. For instance, it could have been the case that visual cortex represents mental images only at a stimulus level (e.g. edges, corners, contrast) while prefrontal cortex represents those images only at a conceptual level (e.g. “the S-shape” or “the tadpole shape”). In this case, the dissimilarity structures derived from each ROI could still be highly correlated with each other (e.g. shape 1 and shape 2 are similar at the stimulus level because they are both derived from a rectilinear grid, and are also “conceptually” similar because they both look like letters). However, these matching dissimilarity structures would have derived from very different underlying informational spaces, and thus it would be erroneous to conclude that the correlation between those dissimilarity structures indicated sharing of information between the ROIs. The second question our new techniques allowed us to address was whether patterns of information flow between network nodes carry information about the functional significance of the connections between those nodes. These questions and the techniques described here to investigate them are generally applicable across a range of topics both within neuroscience—for instance learning (50), intelligence (36), language (2), and attention (103)—and in other fields that study similar informational networks in biology and beyond (104).

It should be noted that using fMRI restricted our sensitivity to functional interactions occurring at millimeter or larger spatial scales and on the order of seconds. It is likely that we missed the distribution and sharing of information occurring in more local small-scale neural circuits and on shorter timescales than we could measure. For instance, the reduced sharing of information and connectivity we found for manipulations of visual imagery may not be an indication that such sharing and connectivity do not occur in the brain, if such processes happen at finer spatial or temporal scales than fMRI can measure. Additionally, focusing on the six ROIs that had previously shown information pertaining to visual imagery increased the power of our analyses within this restricted network. However, this statistical power was gained at the expense of potentially missing a larger scope for the mental workspace network. Indeed, we previously found six additional bilateral neural regions in the cerebellum, thalamus, medial temporal lobe, supplementary eye field, frontal operculum, and medial frontal cortex with activity that differed depending on whether visual imagery was manipulated or maintained, but we could not classify between different mental operations in these regions (95). Presumably, then, the mental workspace network is even larger and more distributed than we report here, with the contribution of these additional nodes yet to be determined.

While we found shared information pertaining to representations in each of the ROIs we studied, an alternative explanation for this finding could be that information about representations merely spreads passively from a single area such as visual cortex that actually processes that information. However, the widespread bidirectional information flow between many network nodes suggests that this is an unlikely possibility. The bidirectionality, density, and hierarchical nature of the connectivity between these nodes lead more parsimoniously to an interpretation that the brain processes mental visual representations in a fundamentally distributed manner.

Finally, connectivity analyses such as those presented here are vulnerable to the lurking variable problem, in which two network nodes appear to support a direct informational connection when in fact each supports independent yet parallel processes or is mutually driven by a third, unknown process. Our information flow results may be affected by this situation, since our network showed a dense pattern of connectivity and we did not test each connection for mediating variables. Because of this, we suggest that these findings be interpreted more holistically as providing evidence for fundamentally distributed information processing in the brain, rather than as having deduced a precise wiring diagram of the mental workspace network.

Our results provide new evidence that high-level cognitive processes such as the representation and manipulation of visual imagery are mediated via the complex, fundamentally distributed flow and sharing of information throughout the cerebral cortex. While much work in cognitive neuroscience has been concerned with reducing the brain’s functions to discrete, localized regions, our results provide evidence that the component processes of at least some forms of high-level cognition occur in a manner that transcends any single neural structure, emerging fundamentally from the interaction between several levels of organization (104, 105). The field has found studying such interactions vital yet difficult (92, 93, 105, 106), and the new methods reported here to investigate the structure, sharing, and flow of information in the brain may prove useful in understanding many other complex cognitive processes (2, 6, 36, 50, 103). Future work should investigate how precisely the distributed flow of information in the cortex supports high-level cognitive abilities and whether this mode of information processing is unique to certain forms of cognition or common across many cortical functions.

Materials and Methods

Participants. 19 participants (6 females, aged 18-51 years) with normal or corrected-to-normal vision gave informed written consent according to the guidelines of the Committee for the Protection of Human Subjects (CPHS) at Dartmouth College prior to participating. All experimental protocols were approved by CPHS (IRB #15822). Participation consisted of two experimental sessions: one behavioral session in which participants practiced the task until they reached criterion (described below) and a subsequent 1.75 hour fMRI scanning session.

Experimental Design. During each of a series of trials, participants performed one of four mental operations on one of four abstract visual shapes. The four mental operations were: 90° clockwise rotation, 90° counterclockwise rotation, horizontal flip, and vertical flip. The four abstract shapes are shown in Figure 3.1: two shapes were constructed from a 4 × 4 rectangular grid, and two were constructed from an analogous polar grid. All shapes were matched for area. To equate the visual presentation between conditions, we did not display the shape or operation to use in a given trial. Instead, each shape was mapped to one of the letters A, B, C, or D, and each operation was mapped to one of the numbers 1, 2, 3, or 4. Each participant was assigned a unique mapping and spent the practice session committing the shapes, operations, and mapping to memory. The practice session concluded once the participant responded correctly on 10 consecutive trials. At the start of each trial, a 2-second-long prompt screen displayed four letter/number pairs (e.g. “C3”). An arrow pointed to one of these pairs to indicate the shape and operation to use for the current trial. This screen was replaced by a fixation dot for 6-s during which the participant performed the indicated mental operation on the indicated shape. After this period, a 2-second-long test screen displayed each of the four shapes at various orientations relative to the starting orientations learned by the participants. The participant was instructed to identify the current trial’s shape on the screen and indicate via a button press within that 2s period whether it was in the orientation that would result from the trial’s indicated operation. In half of the trials the shape was in the correct orientation, and in the other half it was in a random, incorrect orientation. During the fMRI session, the operations and shapes were counterbalanced across all trials, and correct/incorrect trials and display positions were randomized. In order to encourage attentiveness, participants were paid based on their performance (receiving money for correct responses and losing money for incorrect responses, with a minimum base rate of reimbursement). Participants completed 15 fMRI runs, each of which consisted of 16 trials interleaved with 8s of rest to ensure that the BOLD response for a given trial was not influenced by activity from the previous trial (5’28” per run). Thus, each stimulus and operation occurred four times per run (60 times in total during the experiment), and 240 trials were administered over the scanning session.

MRI acquisition. MRI data were collected using a 3.0-Tesla Philips Achieva Intera scanner with a 32-channel sense head coil located at the Dartmouth Brain Imaging Center. One T1-weighted structural image was collected using a magnetization-prepared rapid acquisition gradient echo sequence (8.176ms TR; 3.72ms TE; 8° flip angle; 240 ×

220mm FOV; 188 sagittal slices; 0.9375 × 0.9375 × 1mm voxel size; 3.12 min acquisition time). T2*-weighted gradient echo planar imaging scans were used to acquire functional images covering the whole brain (2000ms TR, 20ms TE; 90° flip angle, 240 × 240mm FOV; 3 × 3 × 3.5mm voxel size; 0mm slice gap; 35 slices).

MRI data preprocessing. High-resolution anatomical images were processed using the FreeSurfer image analysis suite (89). Standard preprocessing of fMRI data was carried out: data were motion and slice-time corrected, high pass filtered temporally with a 100s cutoff, and smoothed spatially with a 6mm full-width-at-half-maximum Gaussian kernel, all using FSL (88). Data from each run were concatenated temporally for each participant after aligning each run using FSL’s FLIRT tool and demeaning each voxel’s time course. For the ROI classification (described below), data were prewhitened using FSL’s MELODIC tool (i.e. principal components were extracted using MELODIC’s default dimensionality estimation method with a minimum of 10 components per ROI).

ROI Classification Analysis. Each trial could be labeled based on either the shape that was represented in visual imagery or on the operation that was performed to manipulate that representation. For each of these two labeling schemes, we used PyMVPA (90) to perform a four-way spatiotemporal multivariate classification analysis in each of the six ROIs that showed information pertaining to manipulation of visual imagery in a previous study (see Figure 3.2A) (95). Five of these (LOC, PPC, PCU, DLPFC, and FEF) were bilateral ROIs that showed greater activity during visual imagery manipulation than visual imagery maintenance in a whole-brain, group level general linear model analysis (see Ch. 2 for details). These ROIs were transformed separately for each participant from MNI space to that participant’s native functional space for use in the current study. The remaining mask (OCC) was defined anatomically in each participant's native anatomical space using the following labels from FreeSurfer's cortical parcellation: 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 (all bilateral). For the control ROI analysis, the thalamus was defined functionally as above, and the ventricle mask was defined anatomically from the following FreeSurfer cortical parcellation masks: left and right lateral ventricles, left and right inferior lateral ventricles, 3rd ventricle, 4th ventricle, and 5th ventricle.

For the spatiotemporal multivariate classification we used a linear support vector machine classifier and leave-one-trial-out cross validation. Because we only considered correct-response trials, a non-uniform number of trials existed for each condition and participant (57.4 trials per condition on average [SEM: 0.203]; see Table S3.1 for details). Even though the difference in number of trials was small, we ensured that they could not affect the classification results by including a target balancing step in our cross-validation procedure. In this step, each classification fold was performed 10 times using random, balanced samples of the data, and the results for that fold were averaged across the 10 bootstrapped folds. For each classification we used the spatiotemporal pattern of prewhitened BOLD data from the first 3 TRs of each correct response trial, shifted by 1 TR to account for the hemodynamic response function (HRF) delay inherent in fMRI data. We shifted by 1 TR only in order to include as much trial data as possible without also including data that could have been influenced by the test display. Pre-whitening reduced each ROI's voxel-based pattern to an average of 93.6 data features (SEM: 4.83).

Thus each classification used spatiotemporal patterns of, on average, 280.8 dimensions (SEM: 14.5). Each feature dimension was z-scored by run prior to classification to reduce between-run differences in signal that may have occurred due to scanner or physiological noise.

Our measure of classifier performance was the correlation between the confusion matrix resulting from the classification and the matrix form of either the shape or operation similarity structure (see Figure 2.1B & C). This measure is more sensitive than classification accuracy because it also takes into account confusions between conditions that result from the hierarchical relationship between the shapes and between the operations. We used a jackknife procedure to perform random-effects analyses evaluating the significance of the correlations (70). In the case of noisy estimates such as individual subject confusion matrices, jackknifed analyses can provide cleaner results without biasing statistical significance (see ref. (70) for more details on this method). In a jackknifed analysis with N subjects, N grand means of the data (in this case, confusion matrices) are calculated, each with one subject left out. The correlation between each of these grand mean confusion matrices and the model similarity structure was then calculated, and a one-tailed t-test evaluated whether the Fisher's Z-transformed correlations were positive (i.e. whether there was a significant correlation between confusion matrices and the model similarity structure across participants). Because the jackknife procedure reduces the variance between subjects artificially, a correction must be applied to the t-statistic calculation; specifically, the sample standard deviation between correlations is multiplied by the square root of (N-1).

ROI Cross-classification Analysis. To assess whether information about mental representations or mental manipulations was shared in a common format between areas, we performed a cross-classification analysis in which a classifier was trained on data from one ROI and tested on data from a second ROI. This analysis used the same procedures as the ROI classification analysis described above. However, because the voxel-based feature space of each ROI differed, data from pairs of ROIs needed to be transformed into a common feature space prior to classification. In order to do this, we first used FSL's MELODIC tool to transform each ROI's data from voxel space to 50 principal component signals using PCA. After this step, each ROI's pattern had the same dimensionality, but those patterns' features would be unlikely to correspond. Therefore, for each pair of ROIs these component signals were matched pairwise as follows in order to maximize the total similarity between component signals. First, the correlation distance ($1 - |r|$) between each pair of components was calculated, yielding a 50 × 50 correlation distance matrix. Next, the rows and columns of this matrix were reordered using the Hungarian algorithm to minimize the matrix trace (107). The components meeting along the diagonal of this reordered, trace-minimized matrix defined the pairwise matching. If two components were matched by this procedure but were anti-correlated, one component was negated in order to produce positively-correlated component pairs. We performed this matching procedure for each fold of the cross validation independently, excluding test data in order to avoid inflating the similarity between training and testing patterns artificially. Once this procedure was complete, data from the two ROIs shared a common feature space, i.e. the two feature spaces had the same dimensionality and corresponding features in the two spaces were maximally similar.

Cross-classification could then proceed by training the classifier on data from one ROI and testing it on data from the other ROI. Each ROI served both as the training set and as the testing set, with results averaged between the two cases. Figure S3.2 provides a visual schematic of the cross-classification analysis procedure.

Information Flow Classification Analysis. The goal of this analysis was to determine whether patterns of directed connectivity between processes occurring in pairs of ROIs could be used to classify either mental representations or mental manipulations. To this end, we transformed the functional data using PCA as above, but with dimensionality fixed at 10 components. For each participant, task condition (i.e. unique combination of shape and operation), and directed pair of areas (e.g. from PPC to DLPFC), we then calculated the Granger-causality with a lag of 1 TR between each directed pair of principal component signals (e.g. between component *i* of PPC and component *j* of DLPFC). As input data for each component we used the temporal concatenation of data from the first 5 TRs of each correct-response trial of that condition, shifted by 1 TR to account for the HRF delay. For each participant and directed pair of ROIs, this procedure yielded 16 10 × 10 Granger-causal graphs which were used as the patterns for classification. Each pattern was labeled based on either shape or operation and analyzed using a multivariate classification as in the ROI classifications described above. Because these patterns were defined for each task condition rather than for each trial, we used leave-one-operation-out cross validation for the representation analysis and leave-one-shape-out cross validation for the manipulation analysis. Directed connections with classification results that passed FDR correction for multiple comparisons across the 30 directed pairs in each analysis were used to construct directed graphs which were then sorted topologically (see Figure 3.4B & D). Figure S3.3 provides a visual schematic of the information flow classification analysis procedure.

Ch. 4: Widespread information sharing integrates the motor network into the mental workspace during mental rotation[3]

Abstract. Studies of mental rotation and similar cognitive abilities suggest that the manipulation of mental representations in the human brain resembles the physical manipulation of real-world objects. Neuroscientific research has revealed that the representations and operations underlying such mental manipulations are implemented in distributed information processing networks. In particular, some neuroimaging studies have found increased activity in motor regions during mental rotation, suggesting that mental and physical operations may involve overlapping neural implementations. Does the motor network contribute information processing to mental rotation? If so, does it play a similar computational role in both mental and manual rotation, and how does it communicate with the wider network of areas involved in the mental workspace? Here we use a variation of a classic mental rotation paradigm along with multivariate decoding methods to investigate these questions. We find that information about mental rotations is shared robustly throughout and beyond the motor network, and that this information only partially resembles that involved in manual rotation. These findings establish that the motor network is recruited for mental rotation in a manner that both resembles and differs from analogous manual rotations. Additionally, these findings provide evidence that the mental workspace is organized as a distributed core network that dynamically recruits existing subnetworks for specific tasks.

[3] This chapter is currently under review for publication (138).

Introduction

In a seminal experiment on the mental manipulation of visual imagery, Shepard and Metzler (16) asked participants to mentally rotate visually presented three-dimensional objects to determine whether they matched other similar objects. Participants' response times correlated tightly with the angle of rotation that would be necessary in order to align the two objects, suggesting that they had mentally rotated endogenous mental models of the objects in a continuous manner as if manually rotating a physical object through space. Subsequent behavioral research has explored other operations such as mental paper folding (17), the generation and analysis of mental analog clocks (18), and mental simulations of mechanical systems (10), a primary result being that volitional mental operations appear in many respects to resemble their corresponding physical operations. Other work has documented similar processes in domains such as mental time travel (19), creative synthesis of mental imagery (20), and visuospatial reasoning (21). Thus, the human brain appears to support a mental space analogous to the physical world in which mental models can be constructed, manipulated, and tested in a flexible manner.

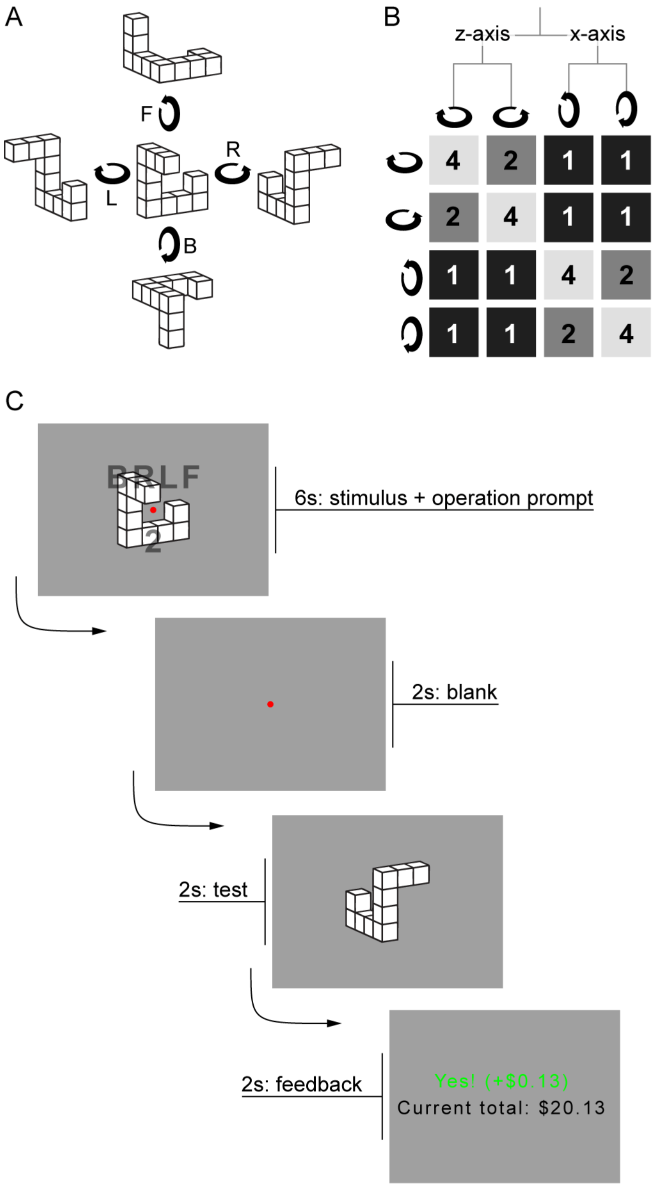
Such abilities have been studied using several overlapping psychological constructs including working memory (7), mental imagery (8, 91), visuospatial ability (9), mental models (10), analogical reasoning (11), and the mental workspace (1). Following Logie (1), we will refer to the mental space in which these flexible cognitive processes occur as the mental workspace.

What is the neural basis of this mental workspace that appears to be so central to the human capacity for imagination? While traditional neural models of working memory and related processes posit an anatomically-modular organization in which physically segregated regions implement component functions such as a “central executive” or a “visuospatial sketchpad” (7, 38, 40, 44, 53), recently developed information- and network-based neuroscientific methods suggest instead that the mental workspace and its component processes may be implemented in a fundamentally distributed manner across the cortex and related regions (46–51, 95, 108). In particular, in a recent study we showed that information about both visual mental imagery and mental manipulations of that imagery is distributed among several regions across the cortex and that this information is shared in a common format via complex, hierarchical patterns of information flow (108). If, as these studies suggest, information is fundamentally distributed across the cortex during such high-level mental activity, then how and where does this information originate? Cognitive work such as that of Shepard and Metzler suggests the possibility that, in order to direct actions within the mental workspace, the brain may recruit existing neural circuitry that evolved for interactions with the physical world.

In fact, several neuroimaging studies have reported activation in various motor areas during mental rotation tasks (33, 109–113). In addition, Kosslyn and colleagues (114) found evidence that participants can be trained to simulate the mental rotation of objects as if they were rotated manually by the hand. These findings support the idea that the mental workspace permits mental operations on endogenously constructed models as if they existed physically. However, these neuroimaging studies have given inconsistent accounts of the motor regions involved in mental rotation. Moreover, the functional role of the increases in cortical activation found in earlier studies is difficult to interpret.

Given the ambiguity and diversity of past findings, it is still unclear precisely what role motor processing may play, if any, in mental rotation.

Figure 4.1. Experimental design



A. The four 90° mental rotations used in the experiment. Left ("L") and right ("R") rotations occurred along the z-axis, while forward ("F") and backward ("B") rotations occurred along the x-axis. **B.** The four rotation directions are related in a two-level hierarchy, such that trials involving a given rotation are most similar to other trials involving the same rotation, moderately similar to trials involving opposite rotations along the same axis, and least similar to trials involving rotations along a different axis. This similarity structure is encoded in the model similarity matrix presented. **C.** A schematic of the trial design. In screen 1 the stimulus figure and mental rotation for the current trial are presented. In screen 2 only a fixation dot is shown. During these first 8s of the trial, participants perform the indicated mental rotation and a concurrent fixation task to ensure that they do not move their hands in accordance with the mental rotation. In screen 3 a test figure appears, and participants indicate whether this figure matches the result of the mental rotation. In screen 4 participants are given feedback about their response.

In light of the above findings and an emerging view of the mental workspace as both highly flexible and fundamentally distributed, we hypothesized that the network underlying the core functionality of the mental workspace would recruit the motor network into a larger, dynamically constructed network in order to carry out mental rotation. Here, we define the motor network as the set of brain regions that are responsible for the planning, production, and monitoring of movements. In order to test the hypothesis that the role of the motor network in mental rotation is to simulate the execution of physical rotations on imagined mental representations, we additionally investigated the relationship between information processing in the motor network during mental rotations and information processing during corresponding physical hand ("manual") rotations.

In a variation of Shepard and Metzler's classic paradigm, we recruited 24 right-handed participants for an initial behavioral session and a subsequent functional MRI (fMRI) scanning session in which they completed a series of trials involving either the mental rotation of three-dimensional cube assemblages (see Figure S4.1) or corresponding manual rotations. Figure 4.1C provides a visual schematic of the experimental trial design. In each mental rotation trial, participants mentally rotated a presented stimulus figure by 90° in one of four hierarchically related rotation directions (Figure 4.1A & B). In manual rotation trials, participants merely rotated their empty right hand in analogous directions. Because of the hierarchical relationship among the rotation directions, we could use multivariate decoding methods to evaluate whether rotation-specific information processing in a set of cortical and subcortical regions of interest (ROIs) matched the informational structure of the rotation operations themselves, thus providing a strong test of the functional role of each ROI in the network during mental rotation. We additionally used a newly developed ROI cross-classification analysis to evaluate whether the information carried by each network node was shared among all nodes in a common format (108), as would be expected if information processing in the mental workspace is fundamentally distributed.

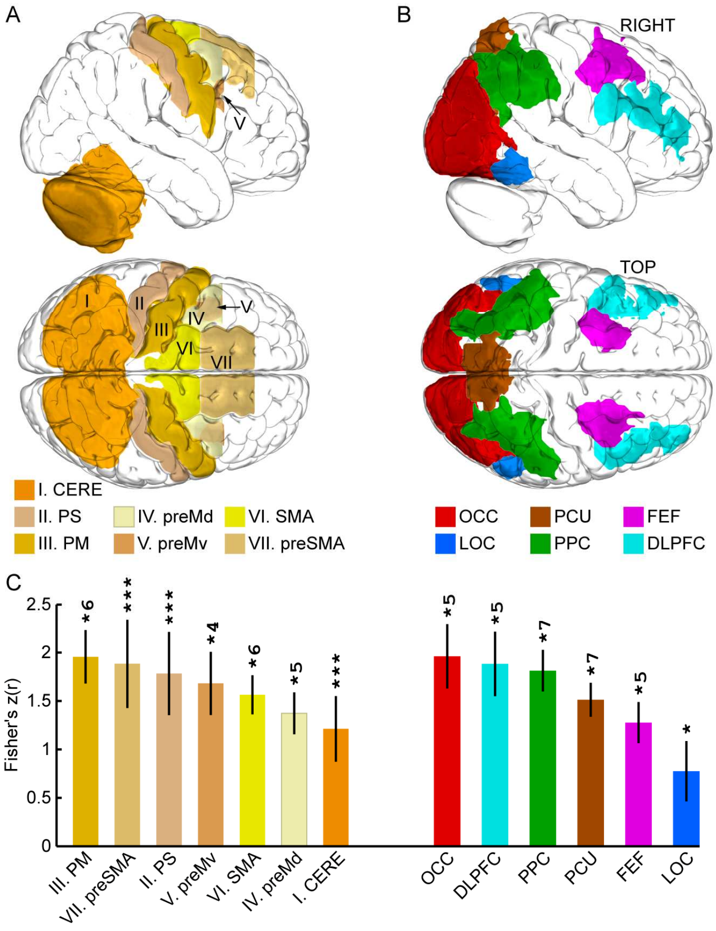
We used two strategies to evaluate the hypothesis that the motor network's role in mental rotation is related to its function during physical motor actions. First, we used a two-group training design similar to that used by Kosslyn and colleagues (114) to evaluate whether the neural similarity of mental and manual rotations could be manipulated. In an initial behavioral session, participants were randomly assigned to one of two training groups without their knowledge and subsequently completed 100 training trials. Interleaved on half of the training session trials, participants in the first "non-motoric" training group were shown an animation of the stimulus figure being rotated; in the subsequent fMRI session they were told to "imagine the mental rotations as an internal movie playing in your head." Instead of the animations, participants in the second "motoric" training group were provided with physical wooden replicas of the stimulus figures that they could rotate manually; in the fMRI session they were told to "imagine rotating your mental image as you did the physical model." Our second strategy to evaluate the role of the motor network was to perform a cross-classification analysis comparing the fMRI data from mental rotation and manual rotation trials in order to assess whether information processing in the motor network during mental rotation trials resembles information processing during analogous manual rotation trials.

Results

Performance accuracy was high during the fMRI session (mean correct response rate was 88.5% [S.E.M. 1.02%] across all participants and conditions), indicating that participants had little difficulty carrying out the instructed mental rotations. A one-way analysis of variance showed no significant differences in the correct response rate between conditions [$F(3,92) = 1.33$, $p = 0.270$], confirming that the difficulty was well matched between rotation conditions (see Table S4.1 for behavioral results separated by rotation direction). We additionally found no significant behavioral differences between the two training groups [correct response rate for non-motoric training group: 86.2% (S.E.M 2.87%); for motoric training group: 90.9% (S.E.M 1.64%); $t(22) = -1.42$, $p = 0.170$].

ROI Classification Analysis. We defined 13 ROIs for each subject that we evaluated for mental rotation-specific information processing using a multivariate classification analysis (Figure 4.2A & B). Seven of these ROIs were anatomically defined regions of the motor network, and six were previously shown to form part of a cortex-wide network of regions that mediate mental workspace processes (95, 108) (see Materials and Methods for details on how each ROI was defined). The classification analysis used a standard cross-validation procedure (62). Briefly, in each fold of the cross-validation, a linear support vector machine classifier was initially trained by presenting it with a set of brain activity patterns derived from individual correct-response mental rotation trials along with the directions of the rotations performed on those trials. In a subsequent testing step the classifier was presented with a holdout sample activity pattern without a rotation-direction label and its ability to correctly label the pattern based on the previous training step was evaluated. This procedure was performed individually for each ROI and participant.

Figure 4.2. ROI classification results



A. Seven bilateral motor network ROIs used in the analyses, shown on an MNI template brain. **CERE**: cerebellar cortex; **PS**: primary somatosensory cortex; **PM**: primary motor cortex; **preMd**: dorsal premotor cortex; **preMv**: ventral premotor cortex; **SMA**: supplementary motor area; **preSMA**: pre-supplementary motor area. **B.** Six bilateral ROIs found in previous studies (95, 108) to mediate processing in the mental workspace and used in the analyses here. **OCC**: occipital cortex; **LOC**: lateral occipital cortex; **PCU**: precuneus; **PPC**: posterior parietal cortex; **FEF**: frontal eye fields; **DLPFC**: dorsolateral prefrontal cortex. **C.** Results of four-way classification analyses in each ROI. Correlations between resulting confusion matrices and the similarity structure in Figure 4.1B are Fisher's *Z*-transformed. Error bars are jackknife-corrected standard errors of the mean (see Materials and Methods). Asterisks indicate significance in a one-tailed jackknifed t-test comparing Fisher's *Z*-transformed correlations to zero across subjects (*: $p <= 0.05$; ***: $p <= 0.001$; $*^{(n)}$: $p <= 1 \times 10^{-n}$). Results are false discovery rate (FDR) corrected for multiple comparisons across the 13 ROIs.

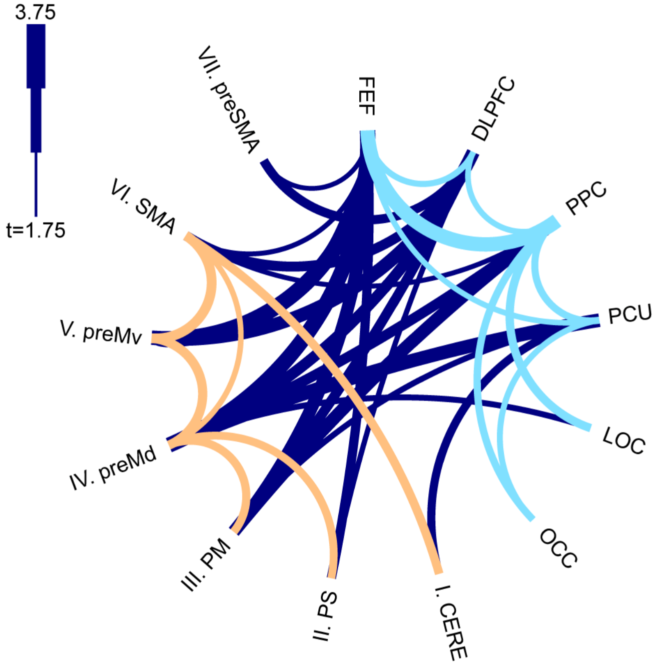
The result of each cross-validation was a 4 × 4 confusion matrix that represented a summary record of the classifier’s predicted labels relative to the true target labels across all cross-validation folds. A perfect classifier would yield a confusion matrix with non-zero values only along the diagonal, since cells along the diagonal represent instances in which the target and predicted labels were the same. However, because we used mental rotations that shared a specific hierarchical similarity relationship (see Figure 4.1B), we expected the classifier to make a specific pattern of confusions among the brain activity patterns in ROIs that were involved in carrying out those mental rotations. For example, we expected that the classifier would confuse a left rotation with a right rotation (both along the z-axis) more often than it would confuse a left rotation (z-axis) with a forward rotation (x-axis), but only if the information processing underlying the brain activity patterns was related specifically to mental rotation. Thus, our measure of classifier performance was the correlation between the confusion matrix resulting from the cross-validation and the matrix form of the rotation-direction similarity structure shown in Figure 4.1B. Note that because we used correlation as our measure, the specific numerical values of this model similarity matrix are irrelevant. Only the relative magnitudes of values matter for the correlation calculation, in this case signifying that a trial involving a particular rotation direction is most highly related to trials with the same rotation, moderately related to trials with opposite rotations along the same axis, and least related to trials with rotations along a different axis. We have successfully used this confusion matrix correlation measure in previous studies to probe the complex structure of information processing in the mental workspace (95, 108).

We conducted this procedure for each ROI and participant individually, and then assessed the information content within each ROI by performing an across-subject random effects analysis to determine whether a significant correlation existed between that ROI's confusion matrices and the model rotation-direction similarity matrix. Results of this analysis are presented in Figure 4.2C, showing that each of our 13 ROIs supported robust information processing related to mental rotation (all results are false discovery rate [FDR] corrected for multiple comparisons across the 13 ROIs; see Figure S4.2 for confusion matrices for each ROI). This result may seem surprising according to traditional models of functional localization, since it indicates that areas as seemingly unrelated to the rotation directions as occipital cortex and primary somatosensory cortex carry information about specific mental rotations. However, this finding is consistent with previous results suggesting that information processing in the mental workspace is fundamentally distributed in the sense that traditional anatomical boundaries of functionality break down in these high level mental processes. In particular, this analysis establishes robustly that information processing related directly to mental rotation occurs throughout the motor network.

ROI Cross-classification Analysis. We next sought to assess whether the processing of mental rotations is truly distributed across the 13 regions of this network. An alternative possibility is that each of the 13 regions plays a role in mental rotation, but that processing in each area is functionally isolated as would be expected in the case of anatomically-modular functional localization. Investigating this question also allowed us to evaluate whether the motor network plays a separate role or becomes tightly integrated into the larger mental workspace network during mental rotation. We used a recently developed ROI cross-classification analysis to assess these alternatives. In this analysis a classifier is trained on data from one ROI and tested on data from a different ROI (108). A successful cross-classification would provide evidence that information is shared in a common format between the two ROIs. An unsuccessful cross-classification would leave open the possibility that the two ROIs represent information in separate formats.

A technical challenge to cross-classifying between ROIs is that each ROI exists initially as an incompatible voxel-based feature space (i.e. each ROI consists of a different number of voxels [feature dimensions], and there is no meaningful mapping between the voxels of each ROI). Thus, cross-classification between two ROIs first requires data from the ROIs to be transformed into a common feature space. See ref. (108), Materials and Methods, and Figure S3.2 for details of this method. Briefly, we conceptualize our data as reflecting a set of high-level cognitive processes that are mixed between the voxels of the ROIs. A principal components analysis (PCA) rotation performed independently on the voxel-based data from each ROI allows us to transform our data from voxel-space to process/component-space, and additionally to control the dimensionality of each of the two feature spaces. We set the dimensionality of each ROI’s feature space to a fixed value (in this case 50-dimensions), and then pair up feature dimensions between the two ROIs in order to maximize the total correlation between component signals. This procedure is performed independently for each fold of each cross-validation, leaving out data from the testing set in order to avoid artificially inflating the similarity of test patterns across the two ROIs.

Figure 4.3. ROI cross-classification results



Arcs indicate pairs of ROIs in which cross-classification was successful. Results are FDR corrected across the 78 ROI pairs. Arc thickness indicates t-statistic values in a one-tailed, jackknifed t-test of Fisher's *Z*-transformed correlations between confusion matrices and the model similarity structure in Figure 4.1B, compared to zero. Connections within the motor network are colored light orange, those within the core mental workspace network are colored light blue, and those crossing between the subnetworks are colored dark blue. Abbreviations are as in Figure 4.2.

Other than the feature-space transformation and difference in training and testing data sets, the ROI cross-classification was conducted exactly as described for the ROI classification above. We performed this cross-classification analysis for each ROI pair, with results shown in Figure 4.3 (all results FDR-corrected across the 78 ROI pairs). Each arc represents a successful cross-classification, indicating that information associated with mental rotations is shared between that pair of ROIs. Connections within the motor or core mental workspace subnetworks are shown in light orange and light blue, respectively, while connections across these two subnetworks are shown in dark blue. We could successfully cross-classify between most pairs of ROIs, suggesting that information processing related to mental rotations is shared in a distributed manner across the network. In particular, a robust set of connections exist both within and between the motor network and other mental workspace regions, suggesting that the motor network becomes tightly integrated into the greater mental workspace network during mental rotation.

Mental/manual rotation cross-classification. What role does the motor network play in mental rotation? To assess the possibility that the mental workspace recruits the motor network to simulate mental rotations as if they were manual rotations of physical objects, we performed a cross-classification analysis within each ROI in which we trained a classifier on data from mental rotation trials and tested it on data from manual rotation trials, and vice-versa. A successful cross-classification in a given ROI would imply that mental and manual rotations share overlapping neural implementations within that ROI. Other than the difference in training and testing data sets and the different number of cross-validation folds (data in this analysis were partitioned by mental/manual rotation condition rather than by trial), the classification analysis was performed and evaluated exactly as in the ROI classification analysis. Results of the cross-classification for each ROI are presented in Figure 4 (all results FDR-corrected across the 13 ROIs). Three ROIs showed significant informational similarity between mental and manual rotations. One of these ROIs, primary motor cortex, was in the motor network, and two ROIs, posterior parietal cortex and precuneus, were in the core mental workspace network. Two additional ROIs showed significant cross-classification results that did not pass multiple comparisons correction (supplementary motor area and dorsolateral prefrontal cortex). Thus, some of the tested ROIs appear to share overlapping implementations of mental and manual rotations, while others may implement each process in a distinct manner.

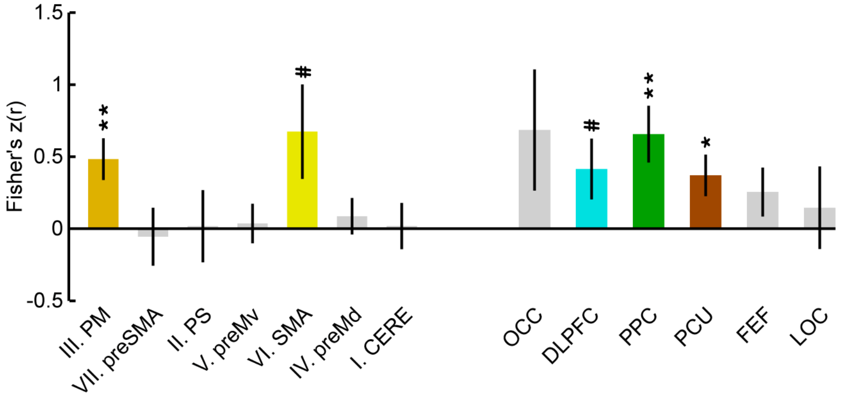


Figure 4.4. Manual/mental cross-classification results

Results of four-way classification analyses in which the classifier was trained using data from mental rotation trials and tested using data from manual rotation trials, and vice versa. Correlations between resulting confusion matrices and the similarity structure in Figure 4.1B are Fisher's *Z*-transformed. Error bars are jackknife-corrected standard errors of the mean (see Materials and Methods). Asterisks indicate significance in a one-tailed jackknifed t-test comparing Fisher's *Z*-transformed correlations to zero across subjects (#: $p <= 0.05$ before FDR correction; *: $p <= 0.05$; **: $p <= 0.01$). Results are false discovery rate (FDR) corrected for multiple comparisons across the 13 ROIs. Abbreviations and ordering are the same as in Figure 4.2.

Between-group differences in mental rotation. We reported above that the non-motoric and motoric training groups did not show significant differences in behavioral performance. However, as Kosslyn and colleagues (114) suggest, participants in different training groups may still have employed different cognitive strategies that would lead to differences in information processing when performing mental rotation. We conducted several analyses to evaluate this possibility.

First, we conducted a univariate analysis similar to that used by Kosslyn and colleagues (114) to assess whether training induced differences in mental rotation-related brain activity. We initially restricted our analysis to the 13 ROIs from the previous analyses. For each ROI and participant we calculated the mean blood-oxygenation-level dependent (BOLD) activity change during mental rotation trials. For each ROI we then performed a two-tailed, unpaired t-test to assess whether these mean mental rotation-related activity levels differed between the two groups. No ROI showed a significant difference in activity after FDR-correction across the 13 ROIs (see Table S4.2 for results). We next conducted an analogous but more exploratory whole-brain analysis to identify regions of the cortex that showed differences in activity between the two groups. No voxels were significant in this analysis after FDR-correction.

While we found no univariate differences in brain activity between the two training groups, our more sensitive multivariate analysis might still show that information processing differed between the groups. To assess this possibility, we performed two-tailed, unpaired t-tests as above but compared the ROI classification results, the ROI cross-classification results, and the mental/manual cross-classification results between the two groups. Each set of t-tests was FDR-corrected independently, and none of these tests showed significant differences between the two groups after correction (see Table S4.3, Table S4.4, Table S4.5). Thus, we failed to replicate the findings of Kosslyn and colleagues (114), since none of our multiple analyses found a behavioral or neuronal difference between the two groups due to the training manipulation.

Discussion

Here we investigated the role of the motor network during mental rotation and its integration into the wider mental workspace. We found that the motor network supported robust information processing related directly to mental rotation and that this processing became dynamically integrated with the distributed, cortex-wide neural network underlying the mental workspace. These findings support a model of the mental workspace as consisting of a flexible core network that can dynamically recruit domain-specific subnetworks for specific functions, much like a general contractor would employ specialists as needed for specific jobs.

Each of the seven motor network ROIs that we tested carried information about specific mental rotations. This result held even in primary somatosensory cortex, a region better known for its role in mediating peripheral sensation. While perhaps surprising, several previous studies of mental rotation have found increases in activity during mental rotation in this and several other areas of the cortex (33). The present results move beyond this previous work by showing that activity in each of these regions is specific to the mental rotations that participants performed. Thus many areas in and beyond the motor network appear to play a functional role in carrying out mental rotations.

Not only do regions throughout the cerebral and cerebellar cortex support information specific to mental rotation, but this information additionally appears to be shared in a common format throughout a widely distributed network. Our ROI cross-classification analysis found that many pairs of ROIs in the network that we studied shared information in the sense that a classifier could use information from one ROI to make a mental rotation-related prediction based on information from a different ROI.

This information sharing held true both within the motor network, suggesting that several subregions of the motor network become tightly integrated during mental rotation, and also between the motor network and a core network of regions underlying the mental workspace. Such widely distributed information about mental rotations and the associated dense pattern of information sharing suggest that information processing in mental rotation entails a breakdown in the anatomical modularity argued for by models of the cortex that are based on functional localization. In support of this view, a recent neurophysiological study by Siegel and colleagues (115) suggests that anatomically segregated regions may only show functional specialization in the early stages of processing, whereas later stages of information processing occur in a much more distributed manner. Our ROI cross-classification results suggest that a common representational format may underlie the inter-regional communication and coordination that would be required within such a distributed system.

Our findings are consistent with a model of the mental workspace that involves a domain general core network that can recruit other specialized subnetworks (e.g. the visual cortex or motor network) for specific tasks as needed. In particular, we found that the motor network was recruited and tightly integrated into a wider network during mental rotation. Consistent with the proposal that the motor network's role is to simulate rotations of imagined objects as if they existed physically, we found that information processing in some regions of the network resembled information processing that occurred during actual physical hand rotations. However, in other regions both within and outside the motor network we found no similarity between mental and manual rotations. We also did not find that training participants to think of mental rotations as simulations of manual manipulations of physical objects had any affect on subsequent neural activity. The reason for our failure to replicate the effect reported by Kosslyn and colleagues (114) is unclear. However, this difference in results may underscore the flexibility of the mental workspace that could allow different participant groups to implement the same functions (e.g. mental rotation) using widely different strategies. In sum, our results suggest that, while the motor network may contribute specialized action-related functionality to the mental workspace during mental rotation, its constituent nodes are also recruited in novel ways for processing that is unique to purely mental simulations.

Much of the last two decades of cognitive neuroscience research has been concerned with assigning functions to localized regions of the cortex in what has been described as a kind of “neophrenology” (52). However, recent studies such as ours and that of Siegel and colleagues (95, 108, 115) and recent work focusing on the brain as a densely connected network (46, 104) suggest instead that high level cognition and possibly cognition generally may entail fundamentally distributed processing and the breakdown of local specialization of function. Furthermore, these findings suggest that distributed informational processing may coexist with functionally localized processing, either on different timescales or at different levels of informational organization. These new models may hint at a level of neural information processing that could form the basis of conscious activity similar to that of the Global Workspace Theory proposed by researchers such as Baars and Dehaene (65, 116), while remaining consistent with localized accounts proposed by Zeki and others (117). Future work should investigate the range of cognitive processes that entail dynamically distributed processing such as that described here. Is this kind of fundamentally distributed information processing unique to high level mental functions, or might new methodological advances reveal that distributed processing is the rule rather than the exception for the brain?

Materials and Methods

Participants. 24 participants (11 females, aged 18-24 years) with normal or corrected-to-normal vision gave informed written consent according to the guidelines of the Committee for the Protection of Human Subjects at Dartmouth College prior to participating. All were right-handed according to the Edinburgh Handedness Inventory (118). Participation consisted of two sessions: one behavioral session in which participants were trained in the task and a subsequent fMRI scanning session.

Task. During each trial, participants performed one of four mental rotations on one of eight figures derived from Shepard and Metzler's (16) original stimulus set (Figure S4.1). All rotations were 90°; two rotations were along the x-axis (called "forward" and "backward" rotations) and two rotations were along the z-axis (called "left" and "right" rotations) (see Figure 4.1A). Each trial lasted 12s and consisted of three phases: the task prompt and operation phase (8s), the test phase (2s), and the feedback phase (2s) (see Figure 4.1C for a visual schematic of the following trial description).

At the beginning of the prompt/operation phase, a randomly chosen figure from the stimulus set, 8° of visual angle in size, was shown centrally. The figure was shown either as depicted in Figure S4.1 or flipped across the y-axis, and additionally either un-rotated or rotated 180° along either the x- or z-axes. Superimposed on this figure in partially transparent text were two prompts: above, a randomly permuted sequence of the letters L, R, F, and B, and below, an integer from 1 to 4. The integer indicated the position of the letter in the above sequence that denoted the mental rotation to carry out on the current trial (e.g. the integer "3" shown below the sequence "BLFR" would refer to the "F" and indicate that the current trial called for a forward rotation). The trial's rotation was indicated in this way in order to equate the visual stimuli across the four mental rotation conditions. Had each rotation's corresponding prompt letter appeared alone on each trial, the visual stimulus would then have differed systematically between conditions and created a possible visual confound in the subsequent multivariate classification analyses (described below). One could argue that increased attention was still directed to the indicated letter and thus may have led to systematic differences in visual representational processing between conditions. However, our confusion matrix-based classifier performance measure (described below) served as a control for this possibility, as it was sensitive to a particular structure of relationships between the rotation directions that did not occur between the stimulus letters.

The figure and rotation direction stimuli remained on screen for 6s and were replaced by a blank screen for 2s. The participant was instructed to perform the indicated mental rotation on the presented figure and to construct as vivid a mental image of the output as possible during this 8s period. Additionally, a red fixation dot appeared centrally during this phase of the trial. The fixation dot blinked blue on average once every 2s, and the participant was instructed to press the "up" button on a four-button box held in the right hand whenever this color change occurred. This fixation task was used in order to minimize the chance that participants might move their hands to mimic the mental rotation being performed.

After the prompt/operation phase, a test figure appeared on the screen for 2s. On half of trials, the test figure was the initial prompt figure after having undergone the indicated rotation (“correct” figure); on the other half of trials, the test figure was a y-axis-flipped version of the initial prompt figure that had undergone the same rotation (“mirror image” figure). Participants were instructed to indicate within the 2s that the test figure was present on screen whether it was the correct figure (“left” button) or the mirror image figure (“right” button).

Finally, a feedback screen indicated whether the participant made the correct response and, during the fMRI session, the current reimbursement amount. As an incentive to attend carefully during the approximately 1.5 hour fMRI session, participants gained $0.125 for each correct response and lost $0.625 for each incorrect response, with a baseline, minimum reimbursement of $20 and a maximum of $40.

Each 5 minute, 28 second run of the fMRI session consisted of 16 trials (4 trials of each rotation type in counterbalanced order), with 8s of rest in between each trial. The fMRI session consisted of 10 runs of mental rotation trials followed by 3 runs of analogous hand rotation trials, in which physical rotations of the right hand were performed instead of mental rotations. Hand rotation trials matched the design of the mental rotation trials, except that no figures were shown, no fixation task or test response was required, and participants merely rotated their right hand continuously according to the prompt until the word “Stop” appeared at the time at which the feedback screen appeared during mental rotation trials. Before the hand rotation runs, videos were shown to the participant to demonstrate proper hand rotation in each of the four directions. Hand rotations resembled the motor actions that would be performed if a physical object was rotated in the same manner as the mentally rotated figures. Even though left and right rotations and forward and back rotations, respectively, involved back and forth hand rotations along the same axis, participants were instructed and the videos demonstrated that more emphasis should be placed on motion in the indicated direction (e.g. more emphasis on the forward phase of rotation during forward rotation trials). Participants were not told about the hand rotation runs until they occurred, in order to avoid biasing participants to imagine their hands as playing a role in the mental rotations.

Training. During the initial behavioral session, participants were instructed in the task and completed 100 practice trials. The prompt/operation phase of practice trials was self-paced: participants viewed the prompt stimulus for as long as desired and indicated with a key press when they were ready for the test phase. In half of the practice trials, the prompt stimulus was accompanied by a guide to assist participants in performing the mental rotation, and in the other half of the trials the prompt occurred without a guide. Guide and no-guide trials were interleaved. Without their knowledge, participants were divided randomly into two training groups (12 participants in each group). In the non-motoric training group, the guide was a looping animation shown below the prompt stimulus that depicted the figure undergoing the indicated rotation. In the motoric training group, the guide was a physical, wooden model that matched the prompted figure and that participants held and rotated manually.

MRI acquisition. MRI data were collected using a 3.0-Tesla Philips Achieva Intera scanner with a 32-channel sense head coil located at the Dartmouth Brain Imaging Center. One T1-weighted structural image was collected using a magnetization-prepared rapid acquisition gradient echo sequence (8.176ms TR; 3.72ms TE; 8° flip angle; 240 ×

220mm FOV; 188 sagittal slices; 0.9375 × 0.9375 × 1mm voxel size; 3.12 min acquisition time). T2*-weighted gradient echo planar imaging scans were used to acquire functional images covering the whole brain (2000ms TR, 20ms TE; 90° flip angle, 240×240mm FOV; 3 × 3 × 3.5mm voxel size; 0mm slice gap; 35 slices).

MRI data preprocessing. High-resolution anatomical images were processed using the FreeSurfer image analysis suite (89). fMRI data were motion and slice-time corrected, temporally high pass filtered with a 100s cutoff, and spatially smoothed with a 6mm full-width-at-half-maximum Gaussian kernel, all using FSL (88). Data from each run were concatenated temporally for each participant after aligning each run using FSL's FLIRT tool and demeaning each voxel's timecourse. For the ROI classification (described below), data were prewhitened for each ROI separately using FSL's MELODIC tool (i.e. principal components were extracted using MELODIC's default dimensionality estimation method with a minimum of 10 components per ROI).

ROI classification. For each of the 13 ROIs, we used PyMVPA (90) to perform a spatiotemporal multivariate classification analysis between the four mental rotation directions. Five of these ROIs (LOC, PPC, PCU, DLPFC, and FEF) were functionally-defined, bilateral masks in MNI space that were then transformed into each participant's native functional space. In a previous study these five ROIs, along with an occipital (OCC) ROI that was defined anatomically for each participant, supported information about the manipulation of visual imagery (95). The OCC ROI was defined in each participant's native anatomical space using the following labels from FreeSurfer's cortical parcellation: 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 (all bilateral). The remaining seven motor network ROIs were defined anatomically using the following FreeSurfer labels (again all bilateral): CERE (cerebellar cortex); PS (postcentral gyrus); PM (precentral gyrus, central sulcus, precentral sulcus [inferior and superior parts]); preMd (posterior third of the middle frontal gyrus, lateral half of the posterior third of the superior frontal sulcus); preMv (inferior frontal sulcus, opercular part of the inferior frontal gyrus); SMA (posterior third of the superior frontal gyrus, medial half of the posterior third of the superior frontal sulcus); preSMA (middle third [in the posterior-anterior direction] of the superior frontal gyrus). In a post-processing step for each participant, voxels that were initially shared between multiple ROIs were assigned to only one ROI using the following, descending order of preference: preMv, preMd, SMA, preSMA, PM, PS, CERE, DLPFC, FEF, PPC, PCU, LOC, OCC. The ROIs shown in Figure 4.2A & B were created as described above but for the MNI template brain. For the spatiotemporal multivariate classification we used a linear support vector machine classifier and leave-one-trial-out cross validation. Because we only considered correct-response trials, a non-uniform number of trials existed for each condition and participant (35.4 trials per condition on average; see Table S4.1 for details). Even though these differences were small, we ensured that they could not affect the classification results by including a target balancing step in our cross-validation procedure. In this step, each classification fold was performed 10 times using random, balanced samples of the training data, and the results for that fold were averaged across the 10 bootstrapped folds. For each classification we used the spatiotemporal pattern of prewhitened BOLD data from the first 5 TRs of each correct response trial, shifted by 1 TR to account for the hemodynamic response function

(HRF) delay inherent in fMRI data. We shifted by only 1 TR in order to include as much trial data as possible. Pre-whitening reduced each ROI's voxel-based pattern to an average of 72.3 data features (SEM: 3.69). Thus each classification used spatiotemporal patterns of, on average, 361 dimensions (SEM: 18.5). Each feature dimension was z-scored by run prior to classification to reduce between-run differences in signal that may have occurred due to scanner or physiological noise.

Our measure of classifier performance was the correlation between the confusion matrix resulting from the four-way classification and the matrix form of the rotation similarity structure (see Figure 4.1B). This measure is more sensitive than classification accuracy because it also takes into account confusions between conditions that result from the hierarchical relationship between the rotations. We used a jackknife procedure to perform random-effects analyses evaluating the significance of the correlations (70). In the case of noisy estimates such as individual subject confusion matrices, jackknifed analyses can provide cleaner results without biasing statistical significance (see ref. (70) for more details on this method). In a jackknifed analysis with *N* subjects, *N* grand means of the data (in this case, confusion matrices) are calculated, each with one subject left out. The correlation between each of these grand mean confusion matrices and the model similarity structure was then calculated, and a one-tailed t-test evaluated whether the Fisher's *Z*-transformed correlations were positive (i.e. whether there was a significant correlation between confusion matrices and the model similarity structure across participants). Because the jackknife procedure reduces the variance between subjects artificially, a correction must be applied to the *t*-statistic calculation; specifically, the sample standard deviation between correlations is multiplied by the square root of (*N*-1).

ROI Cross-classification Analysis. To assess whether information about mental rotations was shared in a common format between areas, we performed a cross-classification analysis in which a classifier was trained on data from one ROI and tested on data from a second ROI. This analysis used the same procedures as the ROI classification analysis described above. However, because the voxel-based feature space of each ROI differed, data from pairs of ROIs needed to be transformed into a common feature space prior to classification. In order to do this, we first used FSL's MELODIC tool to transform each ROI's data from voxel space to 50 principal component signals using PCA. After this step, each ROI's pattern had the same dimensionality, but those patterns' features would be unlikely to correspond. Therefore, for each pair of ROIs these component signals were matched pairwise as follows in order to maximize the total similarity between component signals. First, the correlation distance (1 - $|r|$) between each pair of components was calculated, yielding a 50 × 50 correlation distance matrix. Next, the rows and columns of this matrix were reordered using the Hungarian algorithm to minimize the matrix trace (107). The components meeting along the diagonal of this reordered, trace-minimized matrix defined the pairwise matching. If two components were matched by this procedure but were anti-correlated, one component was negated in order to produce positively-correlated component pairs. We performed this matching procedure for each fold of the cross validation independently, excluding test data in order to avoid inflating the similarity between training and testing patterns artificially. Once this procedure was completed, data from the two ROIs shared a common feature space, i.e. the two feature spaces had the same dimensionality and corresponding features in the two spaces were maximally similar. Cross-classification could then proceed by training the classifier on data from one ROI and testing it on data from the other ROI. Each ROI served both as the training set and as the testing set, with results averaged between the two cases. Figure S3.2 provides a visual schematic of the cross-classification analysis procedure.

Mental/manual rotation cross-classification. To assess whether motor involvement in mental rotation resembled motor activity during physical rotation of the hands, we performed a cross-classification analysis for each ROI in which we trained a classifier on data from the mental rotation trials and tested the classifier on data from the manual rotation trials, and vice-versa. Mental rotation trials were given the same labels as the corresponding manual rotation trials (e.g. trials in which forward mental rotations were carried out were given the same label as trials in which a forward hand rotation was prompted). The classification analysis was performed and evaluated identically to the ROI classification analysis described above except for the difference between training and testing datasets. Note that each cross-validation involved only two folds in this analysis (train on mental rotation and test on manual rotation, train on manual rotation and test on mental rotation), but the same 10-subfold target balancing procedure was used to ensure that training data were balanced.

ROI BOLD comparison of training groups. To assess whether the two different training procedures induced differential brain activity that reflected different cognitive strategies employed during mental rotation, for each ROI we conducted a two-tailed unpaired t-test across participants comparing trial-related mean blood-oxygenation-level dependent (BOLD) activity between the training groups.

We initially used FSL's FEAT tool to perform a first-level whole-brain GLM for each participant in which we defined boxcar predictors for correct response and incorrect response trials. The resulting voxel-wise beta-weights for the correct response predictor, representing the average change in BOLD signal in each voxel during correct response mental rotation trials compared to rest, were then averaged across each ROI. This procedure yielded a single mean trial-related activity estimate for each participant and ROI. For each ROI these values were then partitioned by training group and used to perform a two-tailed unpaired t-test.

Whole brain BOLD comparison of training groups. In a more exploratory variant of the ROI-based BOLD comparison of training groups described above, we performed a whole-brain gray-matter only BOLD comparison using FSL's permutation-based randomise tool with 5000 permutations. The gray matter mask used to restrict the analysis was derived from FreeSurfer's gray matter segmentation of the MNI template brain. The input data to randomise were the correct response beta-weight volumes resulting from the first-level GLM analysis described above (one volume for each participant). The design matrix supplied to randomise defined a single predictor that differentiated between non-motoric and motoric participants. T-contrasts were defined for non-motoric > motoric and motoric > non-motoric.

ROI classification comparison of training groups. To assess whether patterns of mental rotation-related activity differed between the two training groups, we used a procedure similar to that used in the ROI-based BOLD comparison described above to compare the results of the ROI classification analyses between the groups. In this case, our inputs to the unpaired t-tests were the classification results for each participant and

ROI, specifically the Fisher's *Z*-transformed correlations between each participant's confusion matrix resulting from the four-way classification and the model mental rotation similarity structure.

Ch. 5: Discussion

The preceding three chapters presented evidence that the human mental workspace is supported by a fundamentally distributed neural network that spans cortical and subcortical structures throughout the brain. The results and implications of these findings are summarized below, and the reader is referred to the Discussion section of each chapter for a more in depth treatment.

Study 1 (Ch. 2) found a network of bilateral regions throughout the cortex and subcortical regions that supports information specific to mental manipulations of visual imagery. Time point by time point classifications revealed that at least some of these regions tracked the timecourse of the task that participants performed, showing an evolving pattern of information processing from input representation through mental operation to output mental representation. The network switched between two connectivity profiles depending on whether mental representations were maintained or manipulated.

Study 2 (Ch. 3) found that information about the component processes underlying the mental workspace is distributed fundamentally in the brain. All regions studied supported information about both the mental representations held in visual working memory and the mental operations used to manipulate those representations, running directly counter to dominant models of the neural basis of working memory such as that proposed by Baddeley (7). Furthermore, information about mental representations was shared in a common format between many regions, and dense, bidirectional, hierarchically organized patterns of information flow between regions supported information about both mental representations and mental manipulations. These results provide strong evidence that the component processes mediated by this network are fundamentally distributed rather than being segregated to anatomical modules.

Study 3 (Ch. 4) found evidence that the mental workspace is implemented via a domain general core network that dynamically recruits existing subnetworks for specific tasks. Specifically, during mental rotation, widespread information sharing among and between core mental workspace nodes and several motor-related regions integrates the motor network into a larger, transient and task-specific network. Processing in the nodes of this network during mental rotation partially resemble processing involved in physical rotations of the hand, suggesting that the motor network is recruited specifically to simulate the rotation of mentally imagined stimuli as if they are physical objects. However, differences in motor network processing between mental and manual rotation also suggest that the motor network is subsumed into the larger mental workspace network to participate in purely mental phenomena.

A primary conclusion from these three studies is that the investigation of high-order mental phenomena requires a shift in focus away from a currently dominant paradigm that seeks to localize cognitive functions to particular, fixed anatomical substrates. Instead, the field should seek out and develop new methods and conceptual models that explain how higher levels of functional organization in the brain might emerge from and operate on top of the lower-order, automatic, domain-specific, anatomically-modular levels of processing that have so far dominated the field’s inquiry.

Note that anatomically modular and fundamentally distributed modes of information processing are not necessarily mutually exclusive. A fundamental insight into the nervous system has been that conceptually similar functions may be implemented at multiple levels of organization simultaneously. For instance, the cerebral cortex re-implements many of the functions of the brain stem, but at a conscious level that allows for finer, more flexible control over behavior. Likewise, it should not be inconceivable that even within the cortex similar functions are implemented multiple times at different levels of organization. In fact, a recent study by Siegel and colleagues (115) has found evidence supporting the co-existence of both modular and distributed processing. Their data suggest that during initial stages of cortical processing, information is generated by and exists locally within anatomical modules. At later stages, however, this information becomes distributed widely throughout the cortex such that functional localization breaks down. Thus, the development of methods to study the brain as a distributed network may complement, rather than replace, existing insights gained by traditional methods such as lesion studies that have provided evidence for the functional localization.

The studies presented in this thesis focused on the neural basis of the mental workspace and on the manipulation of visual imagery specifically. Thus, several questions regarding the generality of the results remain open for further investigation. For instance:

- Do other domains of processing in the mental workspace entail processes that are as fundamentally distributed as visual imagery? A future study could compare visual imagery to auditory or tactile imagery. The model proposed in this thesis predicts that, much like the occipital cortex and motor network were integrated into a larger network for the spatial manipulation of mental visual images, auditory imagery would entail the integration of information from auditory cortex into the same core network nodes. A test of the extent of the distribution of processing in the mental workspace would be whether information specific to auditory mental images occurs in visual cortex and/or visual imagery information occurs in auditory cortex.

- Is the fundamentally distributed processing revealed by the present studies unique to high-order cognitive functions such as those of the mental workspace, or might such processing occur much more generally in cognition? One possibility is that the methods developed in this dissertation have revealed a general property of cortical activity to which previous methods were insensitive. A future study could investigate this possibility by using the present methods to compare visually imagined representations with physically perceived stimuli. Do straightforward visual perceptual processes such as object categorization (56) also entail the widespread sharing of information in the same regions as those studied here?
- Are mental workspace-like abilities and/or the fundamentally distributed neural processes underlying them unique to humans (119–128)? Future work could investigate the extent to which chimpanzees or other non-human animals can volitionally manipulate mental representations, and also differences in the functional and structural connectivity between species that may underlie the uniqueness of human cognition.
- How do differences in the organization of the mental workspace network account for differences in cognitive style, such as those employed by “visual” or

"propositional" thinkers (129–132)? A future study could compare differences in the sharing and distribution of visual and propositional mental representations between participants with either visual or propositional cognitive tendencies.

- This thesis claims that the abilities studied here lie at the root of the human imagination that enables creative abilities such as scientific and artistic thought. However, the link between the ability to mentally manipulate imagery and creative ability was not tested directly. Future studies could investigate whether mental manipulation ability relates to creativity (20) and whether information is integrated between the regions revealed by the current studies and other networks known to play a role in creative cognition (6, 133, 134).
- Could the methods developed here be used to investigate the neural bases of other complex cognitive processes such as intelligence (36), learning (50), development (135), attention (103), language (2), art (6), or social cognition (136)?

Appendix I: Supplemental Tables

Table S2.1. Response time v. classification accuracy control analysis results

To verify that our ROI classification results were not influenced by response time (RT) differences between the construct parts and the deconstruct figure conditions, we performed a cross-subject correlation analysis for each ROI between classification accuracy and RT difference as in ref. (137). In no ROI was there a significant correlation between RT difference and accuracy (all *p*'s uncorrected). In fact, for our four primary areas of interest there are non-significant *inverse* correlations between the two, suggesting that, if anything, larger reaction time differences were associated with *lower* classification accuracies.

ROI	*r*	*p*
OCC	-0.161	0.566
PPC	-0.265	0.340
DLPFC	-0.341	0.214
PCU	-0.221	0.428
FEF	-0.142	0.613
CERE	0.00320	0.991
SEF	-0.225	0.421
MFC	0.300	0.278
FO	0.230	0.411
MTL	0.267	0.335
PITC	-0.305	0.268
THAL	0.347	0.205

Table S2.2. ROI two-way classification results

Statistical results of two-way classification analyses in each ROI. *t*-tests are one-tailed, compared to 50%. p_{corr} values are false discovery rate corrected *p* values. ***Abbreviations:*** CP = construct parts; DF = deconstruct figure; MP = maintain parts; MF = maintain figure. ROI abbreviations are as in Figure 2.2.

ROI	accuracy (%)		*t*		*p*		p_{corr}	
	CP v. DF	MP v MF	CP v. DF	MP v MF	CP v. DF	MP v MF	CP v. DF	MP v MF
OCC	61.4	65.8	3.75	7.54	1.07e-3	1.35e-6	4.28e-3	3.24e-5
PPC	61.5	65.4	4.25	5.62	4.08e-4	3.16e-5	1.96e-3	2.53e-4
DLPFC	55.8	62.3	2.71	6.18	8.52e-3	1.19e-5	0.0205	1.43e-4
PCU	60.2	58.8	3.58	3.51	1.51e-3	1.74e-3	5.17e-3	5.21e-3
FEF	56.0	57.8	2.16	2.78	0.0244	7.33e-3	0.0489	0.0195
CERE	56.5	51.5	2.38	0.505	0.0160	0.311	0.0350	0.393
SEF	51.5	55.8	0.778	1.79	0.225	0.0471	0.300	0.0870
MFC	50.6	55.7	0.283	4.36	0.391	3.28e-4	0.446	1.96e-3
FO	49.2	53.7	-0.345	1.56	0.632	0.0700	0.690	0.112
MTL	48.3	53.6	-0.577	1.28	0.713	0.110	0.744	0.166
PITC	53.5	52.9	1.59	1.14	0.0675	0.137	0.112	0.193
THAL	47.9	51.1	-0.864	-0.397	0.799	0.349	0.799	0.418

Table S2.3. Correlation-based classification results

Statistical results of correlation analyses between the model similarity structure from Figure 2.3B and confusion matrices from four-way classifications in each ROI.

ROI	*r*	*p*	p_{corr}
OCC	0.970	6.48e-5	3.89e-4
PPC	0.977	2.91e-5	3.49e-4
DLPFC	0.921	1.18e-3	2.82e-3
PCU	0.911	1.66e-3	3.33e-3
FEF	0.955	2.22e-4	6.66e-4
CERE	0.583	0.129	0.165
SEF	0.734	0.0381	0.0654
MFC	0.638	0.0887	0.133
FO	0.355	0.388	0.388
MTL	0.444	0.27	0.295
PITC	0.957	1.90e-4	6.66e-4
THAL	-0.573	0.137	0.165

Table S2.4. Peak correlation time analysis results

Statistical results of linear contrast analyses on peak correlation times from analysis shown in Figure 2.4. LC = linear contrast result. C1 = contrast 1 (input: -1, operation: -1, output: 2). C2 = contrast 2 (input: -1, operation: 1, output: 0). *p*-values for negative contrast results are not shown.

ROI	LC		*F*		*p*	
	C1	C2	C1	C2	C1	C2
OCC	5.65	1.90	178.	60.0	5.19e-12	1.00e-7
PPC	5.40	2.19	9.24	4.56	5.82e-3	0.0435
DLPFC	6.06	1.81	46.5	12.4	7.51e-7	1.92e-3
PCU	4.81	2.93	48.9	54.4	3.98e-7	1.66e-7
FEF	3.66	2.27	0.343	0.396	0.564	0.536
CERE	-1.46	1.31	0.0237	0.0574	-	0.813
SEF	-3.87	2.19	1080.	1030.	-	<1e-12
MFC	-6.69	-0.381	0.894	8.68e-3	-	-
FO	3.36	-1.13	0.156	0.0532	0.697	-
MTL	3.14	0.136	7.70	0.0432	0.0111	0.837
PITC	-2.84	1.27	9.22	5.54	-	0.0280
THAL	-5.76	-2.01	3.79	1.38	-	-

Table S3.1. Behavioral results

Mean number of correct trials per condition across all subjects.

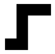
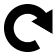
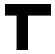
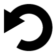
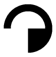
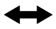
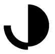
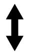
Shape	Mean	S.E.M.	Operation	Mean	S.E.M
	57.2 (95.3%)	0.642		57.5 (95.8%)	0.393
	58.4 (97.3%)	0.520		57.8 (96.3%)	0.308
	56.6 (94.3%)	0.578		57.4 (95.7%)	0.698
	57.5 (95.8%)	0.579		57.0 (95.0%)	0.757

Table S3.2. ROI control analysis results

Control ROI classification analyses with thalamus and ventricle masks. ***z***: Mean of jackknifed Fisher's *Z*-transformed correlations between the classification confusion matrices and the model similarity structures from Figure 3.1. ***t*(18)**: Statistical results of one-tailed t-tests on the jackknifed Fisher's *Z*-transformed correlations, compared to zero. ***p***: p-values from the t-tests.

	Representation			**Manipulation**		
ROI	***z***	***t*(18)**	***p***	***z***	***t*(18)**	***p***
THALAMUS	0.516	1.000	0.165	-0.607	-2.79	0.994
VENTRICLE	0.270	0.620	0.271	-0.133	-0.450	0.671

Table S3.3. Shuffled-label control classification results

z: Mean of jackknifed Fisher's *Z*-transformed correlations between the classification confusion matrices and the model similarity structures from Figure 3.1. ***t*(18)**: Statistical results of one-tailed t-tests on the jackknifed Fisher's *Z*-transformed correlations, compared to zero. ***p***: *p*-values from the t-tests.

	Representation			**Manipulation**		
ROI	***z***	***t*(18)**	***p***	***z***	***t*(18)**	***p***
all	-0.275	-0.917	0.814	-0.226	-0.836	0.793
OCC	-0.200	-0.592	0.719	-0.740	-3.49	0.999
PPC	-0.132	-0.682	0.748	-0.014	-0.044	0.517
PCU	0.148	0.519	0.305	-0.645	-3.04	0.996
LOC	0.293	0.962	0.174	-0.616	-1.77	0.953
FEF	-0.342	-1.21	0.880	-0.256	-1.21	0.879
DLPFC	-0.472	-1.19	0.875	0.185	0.457	0.327

Table S3.4. Shuffled-label control cross-classification results

z: Mean of jackknifed Fisher's *Z*-transformed correlations between the classification confusion matrices and the model similarity structures from Figure 3.1. ***t*(18)**: Statistical results of one-tailed t-tests on the jackknifed Fisher's *Z*-transformed correlations, compared to zero. ***p***: *p*-values from the t-tests. ***p_{corr}***: False discovery rate corrected p-values across the 30 comparisons.

		Representation				**Manipulation**			
ROI1	**ROI2**	***z***	***t*(18)**	***p***	***p_{corr}***	***z***	***t*(18)**	***p***	***p_{corr}***
DLPFC	*FEF*	0.309	1.285	0.107	0.403	0.404	2.108	0.025	0.370
DLPFC	*OCC*	-0.106	-0.391	0.650	0.843	-0.045	-0.165	0.565	0.814
DLPFC	*PCU*	-0.068	-0.494	0.687	0.843	0.097	0.710	0.243	0.814
DLPFC	*LOC*	0.258	1.863	0.039	0.403	0.037	0.185	0.428	0.814
DLPFC	*PPC*	-0.292	-1.595	0.936	0.962	-0.166	-0.744	0.767	0.822
FEF	*OCC*	0.000	0.001	0.500	0.836	0.011	0.080	0.469	0.814
FEF	*PCU*	0.368	1.645	0.059	0.403	0.065	0.332	0.372	0.814
FEF	*LOC*	0.243	1.002	0.165	0.494	-0.065	-0.472	0.679	0.814
FEF	*PPC*	-0.001	-0.004	0.501	0.836	-0.048	-0.259	0.601	0.814
OCC	*PCU*	-0.526	-1.876	0.962	0.962	-0.124	-0.550	0.705	0.814
OCC	*LOC*	0.277	1.411	0.088	0.403	-0.010	-0.032	0.513	0.814
OCC	*PPC*	-0.050	-0.235	0.591	0.843	-0.366	-1.208	0.879	0.879
PCU	*LOC*	0.208	0.594	0.280	0.700	0.100	0.382	0.353	0.814
PCU	*PPC*	0.055	0.264	0.397	0.836	-0.027	-0.189	0.574	0.814
LOC	*PPC*	-0.133	-0.626	0.730	0.843	-0.099	-0.472	0.679	0.814

Table S3.5. Shuffled-label control information flow classification results

z: Mean of jackknifed Fisher's *Z*-transformed correlations between the classification confusion matrices and the model similarity structures from Figure 3.1. ***t*(18)**: Statistical results of one-tailed t-tests on the jackknifed Fisher's *Z*-transformed correlations, compared to zero. ***p***: *p*-values from the t-tests. $\boldsymbol{p_{corr}}$: False discovery rate corrected *p*-values across the 60 comparisons.

		Representation				**Manipulation**			
ROI1	**ROI2**	***z***	***t*(18)**	***p***	$\boldsymbol{p_{corr}}$	***z***	***t*(18)**	***p***	$\boldsymbol{p_{corr}}$
DLPFC	*FEF*	0.128	0.378	0.355	1.000	-0.445	-1.370	0.906	1.000
DLPFC	*OCC*	0.242	0.531	0.301	1.000	-0.216	-1.106	0.858	1.000
DLPFC	*PCU*	-0.198	-0.758	0.771	1.000	0.377	1.038	0.157	1.000
DLPFC	*LOC*	0.556	2.308	0.017	0.331	0.550	1.490	0.077	0.921
DLPFC	*PPC*	-0.289	-0.813	0.787	1.000	-0.439	-1.651	0.942	1.000
FEF	*DLPFC*	0.002	0.005	0.498	1.000	-0.026	-0.065	0.526	1.000
FEF	*OCC*	-0.168	-1.073	0.851	1.000	-0.412	-1.144	0.866	1.000
FEF	*PCU*	-0.179	-0.541	0.702	1.000	-0.192	-0.546	0.704	1.000
FEF	*LOC*	-0.143	-0.45	0.671	1.000	0.158	0.540	0.298	1.000
FEF	*PPC*	0.244	0.816	0.213	1.000	-0.560	-2.090	0.974	1.000
OCC	*DLPFC*	-0.461	-1.476	0.921	1.000	0.235	0.698	0.247	1.000
OCC	*FEF*	-0.226	-0.669	0.744	1.000	0.241	0.491	0.315	1.000
OCC	*PCU*	0.287	0.825	0.21	1.000	0.879	3.742	0.001	0.045
OCC	*LOC*	-0.118	-0.401	0.654	1.000	-0.549	-1.658	0.943	1.000
OCC	*PPC*	0.178	0.437	0.333	1.000	-0.251	-1.082	0.853	1.000
PCU	*DLPFC*	-0.162	-0.582	0.716	1.000	0.136	0.482	0.318	1.000
PCU	*FEF*	-0.117	-0.233	0.591	1.000	-0.261	-1.372	0.907	1.000
PCU	*OCC*	0.189	0.720	0.240	1.000	0.418	1.819	0.043	0.642
PCU	*LOC*	0.717	2.602	0.009	0.270	0.301	0.945	0.178	1.000
PCU	*PPC*	-0.038	-0.140	0.555	1.000	-0.564	-1.585	0.935	1.000
LOC	*DLPFC*	-0.040	-0.125	0.549	1.000	-0.791	-3.860	0.999	1.000
LOC	*FEF*	0.028	0.100	0.461	1.000	-0.559	-2.976	0.996	1.000
LOC	*OCC*	-0.320	-1.243	0.885	1.000	-0.473	-1.801	0.956	1.000
LOC	*PCU*	-0.471	-1.832	0.958	1.000	-0.541	-2.509	0.989	1.000
LOC	*PPC*	-0.025	-0.074	0.529	1.000	-0.422	-1.557	0.932	1.000
PPC	*DLPFC*	-0.021	-0.066	0.526	1.000	-0.953	-4.081	1.000	1.000
PPC	*FEF*	0.146	0.374	0.356	1.000	-0.124	-0.802	0.783	1.000
PPC	*OCC*	0.154	0.457	0.326	1.000	-0.394	-1.285	0.892	1.000
PPC	*PCU*	-0.073	-0.263	0.602	1.000	-0.283	-0.798	0.782	1.000
PPC	*LOC*	0.301	1.092	0.145	1.000	-0.359	-1.592	0.936	1.000

Table S4.1. Behavioral results

Mean number of correct trials per condition across all subjects.

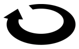
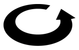
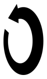
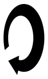
Rotation	Mean	S.E.M.	% Correct
	34.2	0.909	85.5
	35.2	1.04	87.9
	36.	0.689	90.
	36.3	0.509	90.7

Table S4.2. Comparison of univariate activity between training groups

Comparison of ROI-based univariate mental rotation-related activity between training groups. For each ROI, a two-tailed, unpaired t-test compared the mean activity level between the two groups. See Figure 4.2 for abbreviations. $\boldsymbol{\beta}$: GLM beta-weight representing mean mental rotation-related change in brain activity in the specified ROI. **S.E.M.**: standard error of the mean of the beta-weights across participants. $\boldsymbol{p_{corr}}$: FDR-corrected *p*-values across the 13 ROIs.

	non-motoric group		**motoric group**		**unpaired t-test**		
ROI	$\boldsymbol{\beta}$	**S.E.M.**	$\boldsymbol{\beta}$	**S.E.M.**	***t*(22)**	***P***	$\boldsymbol{p_{corr}}$
CERE	11.7	3.64	12.0	4.83	-0.0548	0.957	0.957
PS	19.8	5.13	12.0	6.52	0.937	0.359	0.916
PM	29.5	3.70	27.5	5.89	0.290	0.775	0.916
preMd	11.9	7.97	13.6	8.26	-0.147	0.885	0.957
preMv	-0.224	4.85	16.5	10.3	-1.47	0.156	0.916
SMA	30.5	4.99	22.0	7.30	0.951	0.352	0.916
preSMA	-9.79	4.36	-2.72	7.57	-0.809	0.427	0.916
OCC	48.6	3.63	54.6	8.34	-0.654	0.520	0.916
LOC	26.3	9.16	20.5	8.39	0.465	0.646	0.916
PCU	38.5	10.3	52.4	16.0	-0.729	0.474	0.916
PPC	48.2	6.34	42.0	7.12	0.658	0.518	0.916
FEF	31.2	5.63	27.9	6.83	0.373	0.713	0.916
DLPFC	22.2	4.44	19.0	5.94	0.433	0.669	0.916

Table S4.3. Comparison of ROI classification between training groups

For each ROI, a two-tailed, unpaired t-test compared the classification results between the two groups. See Figure 4.2 for abbreviations. ***z***: mean jackknifed Fisher's *Z*-transformed correlation between the classification confusion matrices and the model similarity structure from Figure 4.1B. **S.E.M.**: standard error of the mean of the *Z*-transformed correlations across participants. ***p_{corr}***: FDR-corrected *p*-values across the 13 ROIs.

	non-motoric group		**motoric group**		**unpaired t-test**		
ROI	***z***	**S.E.M.**	***z***	**S.E.M.**	***t*(22)**	***p***	***p_{corr}***
CERE	0.429	0.110	0.270	0.102	1.06	0.300	0.885
PS	0.300	0.079	0.482	0.122	-1.25	0.224	0.885
PM	0.944	0.145	0.764	0.143	0.881	0.388	0.885
preMd	0.561	0.100	0.405	0.132	0.946	0.355	0.885
preMv	0.338	0.110	0.613	0.0724	-2.09	0.0489	0.635
SMA	0.639	0.120	0.518	0.156	0.616	0.544	0.885
preSMA	0.393	0.0662	0.391	0.188	0.0127	0.990	0.990
OCC	0.758	0.0972	0.725	0.0751	0.266	0.793	0.990
LOC	0.221	0.103	0.211	0.102	0.0676	0.947	0.990
PCU	0.645	0.119	0.662	0.146	-0.0921	0.927	0.990
PPC	0.687	0.143	0.788	0.0667	-0.644	0.526	0.885
FEF	0.426	0.132	0.560	0.0932	-0.827	0.417	0.885
DLPFC	0.584	0.109	0.658	0.109	-0.478	0.637	0.921

Table S4.4. Comparison of ROI cross-classification between training groups

For each significant pair of ROIs in the analysis presented in Figure 4.3, a two-tailed, unpaired t-test compared the cross-classification results between the two groups. See Figure 4.2 for abbreviations. ***z***: mean jackknifed Fisher's *Z*-transformed correlation between the cross-classification confusion matrices and the model similarity structure from Figure 4.1B. **S.E.M.**: standard error of the mean of the *Z*-transformed correlations across participants. $\boldsymbol{p_{corr}}$: FDR-corrected *p*-values across the 33 significant ROI pairs.

		non-motoric group		motoric group		unpaired t-test		
ROI 1	ROI 2	*z*	S.E.M.	*z*	S.E.M.	*t*(22)	*p*	p_{corr}
CERE	*SMA*	0.131	0.0596	0.203	0.0706	-0.780	0.444	0.829
CERE	*PCU*	0.00383	0.0560	0.138	0.0606	-1.62	0.119	0.576
PS	*preMd*	0.0768	0.0461	0.131	0.0666	-0.666	0.513	0.829
PS	*FEF*	0.0639	0.0416	0.186	0.0679	-1.53	0.140	0.576
PS	*DLPFC*	0.0508	0.0434	0.127	0.0648	-0.979	0.338	0.798
PM	*preMd*	0.187	0.0804	0.183	0.0536	0.0467	0.963	0.963
PM	*PCU*	0.114	0.0711	0.199	0.0735	-0.834	0.413	0.829
PM	*PPC*	0.130	0.0326	0.175	0.0728	-0.557	0.583	0.837
PM	*FEF*	0.207	0.0481	0.261	0.0679	-0.642	0.527	0.829
PM	*DLPFC*	0.135	0.0565	0.162	0.0661	-0.310	0.759	0.921
preMd	*preMv*	0.0975	0.0469	0.229	0.0821	-1.39	0.179	0.588
preMd	*SMA*	0.167	0.0570	0.0517	0.0656	1.33	0.197	0.588
preMd	*LOC*	0.115	0.0484	0.0853	0.0627	0.372	0.713	0.910
preMd	*PCU*	0.0753	0.0501	0.227	0.0463	-2.23	0.0362	0.576
preMd	*PPC*	0.126	0.0500	0.220	0.0528	-1.28	0.214	0.588
preMd	*FEF*	0.141	0.0652	0.190	0.0544	-0.576	0.570	0.837
preMd	*DLPFC*	0.146	0.0318	0.135	0.0579	0.174	0.863	0.946
preMv	*SMA*	0.131	0.0550	0.168	0.0723	-0.41	0.686	0.910
preMv	*FEF*	0.0901	0.0451	0.261	0.0833	-1.80	0.0855	0.576
preMv	*DLPFC*	0.101	0.0366	0.336	0.0826	-2.60	0.0164	0.543
SMA	*PPC*	0.160	0.0496	0.154	0.0619	0.074	0.942	0.963
SMA	*FEF*	0.0715	0.0740	0.145	0.0505	-0.818	0.422	0.829
SMA	*DLPFC*	0.142	0.0568	0.210	0.0716	-0.749	0.462	0.829
preSMA	*FEF*	0.0865	0.0498	0.141	0.0648	-0.668	0.511	0.829
preSMA	*DLPFC*	0.163	0.0677	0.134	0.0412	0.367	0.717	0.910
OCC	*PCU*	0.103	0.0596	0.125	0.0870	-0.208	0.837	0.946
OCC	*PPC*	0.0642	0.0281	0.207	0.0836	-1.62	0.119	0.576
LOC	*PPC*	0.0777	0.0387	0.205	0.0836	-1.38	0.181	0.588
PCU	*PPC*	0.142	0.0537	0.119	0.0631	0.281	0.782	0.921
PCU	*FEF*	0.125	0.0630	0.110	0.0840	0.142	0.889	0.946
PPC	*FEF*	0.0839	0.0595	0.180	0.0699	-1.05	0.307	0.779
PPC	*DLPFC*	0.0851	0.0648	0.258	0.0558	-2.02	0.0553	0.576
FEF	*DLPFC*	0.126	0.0422	0.233	0.0525	-1.59	0.127	0.576

Table S4.5. Comparison of mental/manual cross-classification between training groups

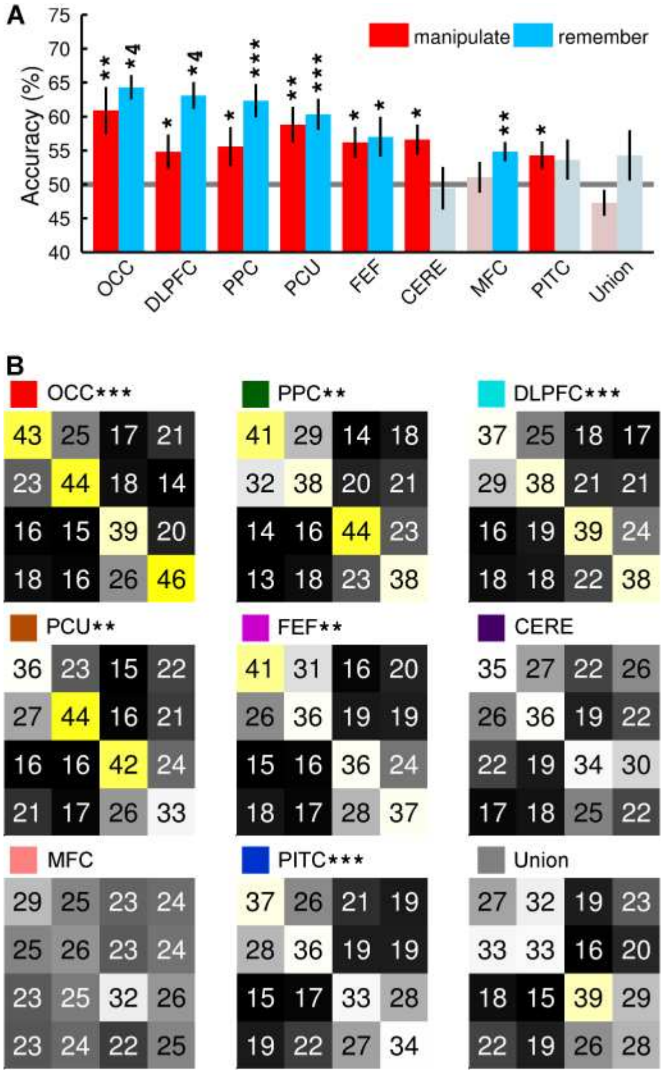
For each significant ROI in the analysis presented in Figure 4.4, a two-tailed, unpaired t-test compared the cross-classification results between the two groups. See Figure 4.2 for abbreviations. ***z***: mean jackknifed Fisher's *Z*-transformed correlation between the cross-classification confusion matrices and the model similarity structure from Figure 4.1B. **S.E.M.**: standard error of the mean of the *Z*-transformed correlations across participants. $\boldsymbol{p_{corr}}$: FDR-corrected *p*-values across the 5 significant ROIs.

	non-motoric group		**motoric group**		**unpaired t-test**		
ROI	***z***	**S.E.M.**	***z***	**S.E.M.**	***t*(22)**	***p***	$\boldsymbol{p_{corr}}$
PM	0.174	0.0721	0.150	0.0660	0.241	0.811	0.811
preMv	-0.0690	0.0519	0.0953	0.0495	-2.29	0.0319	0.160
OCC	0.117	0.0836	0.167	0.0619	-0.487	0.631	0.789
PCU	0.159	0.0824	0.0843	0.0820	0.643	0.527	0.789
PPC	0.159	0.116	0.279	0.0527	-0.945	0.355	0.789

Appendix II: Supplemental Figures

Figure S2.1. ROI-classification to control for ROI size

A. Classification results using the same procedure as described in Figure 2.3, with two modifications. First, to verify that the inability to classify between conditions in SEF, FO, MTL, and THAL was not due to ROI size, the union of these ROIs was constructed and the classification performed within this "Union" ROI. The average size of this ROI across participants was 360 voxels. Second, to verify that classification results 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). B. Confusion matrices and correlation analysis results as in Figure 3C but using the ROIs described above.



OCC***

43	25	17	21
23	44	18	14
16	15	39	20
18	16	26	46

PPC**

41	29	14	18
32	38	20	21
14	16	44	23
13	18	23	38

DLPFC***

37	25	18	17
29	38	21	21
16	19	39	24
18	18	22	38

PCU**

36	23	15	22
27	44	16	21
16	16	42	24
21	17	26	33

FEF**

41	31	16	20
26	36	19	19
15	16	36	24
18	17	28	37

CERE

35	27	22	26
26	36	19	22
22	19	34	30
17	18	25	22

MFC

29	25	23	24
25	26	23	24
23	25	32	26
23	24	22	25

PITC***

37	26	21	19
28	36	19	19
15	17	33	28
19	22	27	34

Union

27	32	19	23
33	33	16	20
18	15	39	29
22	19	26	28

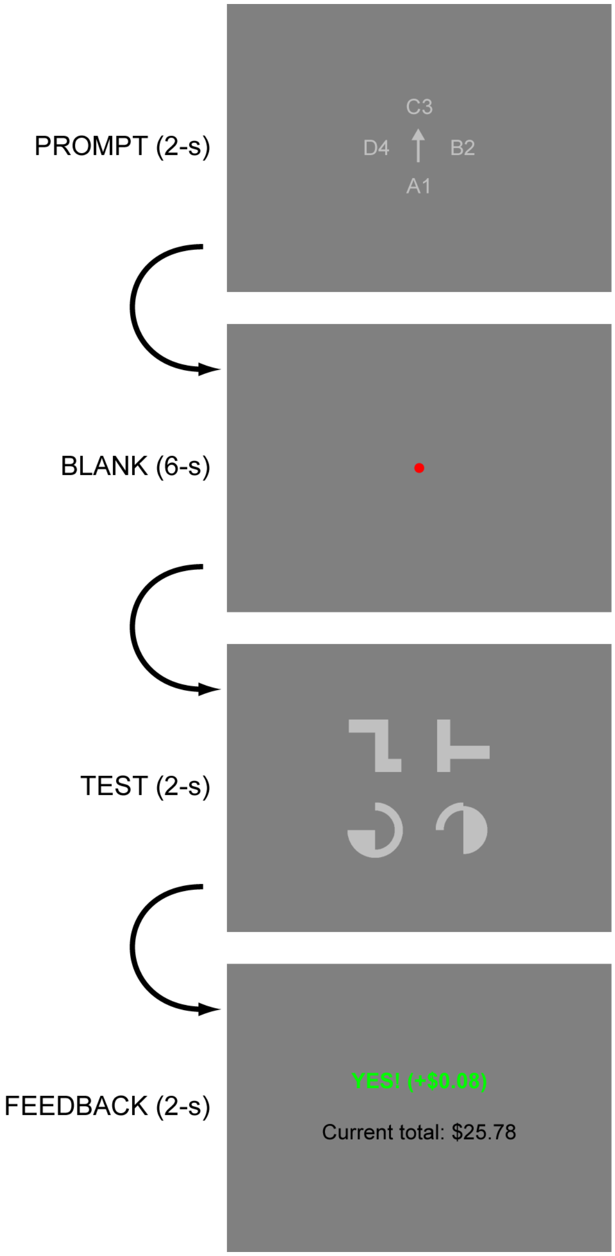
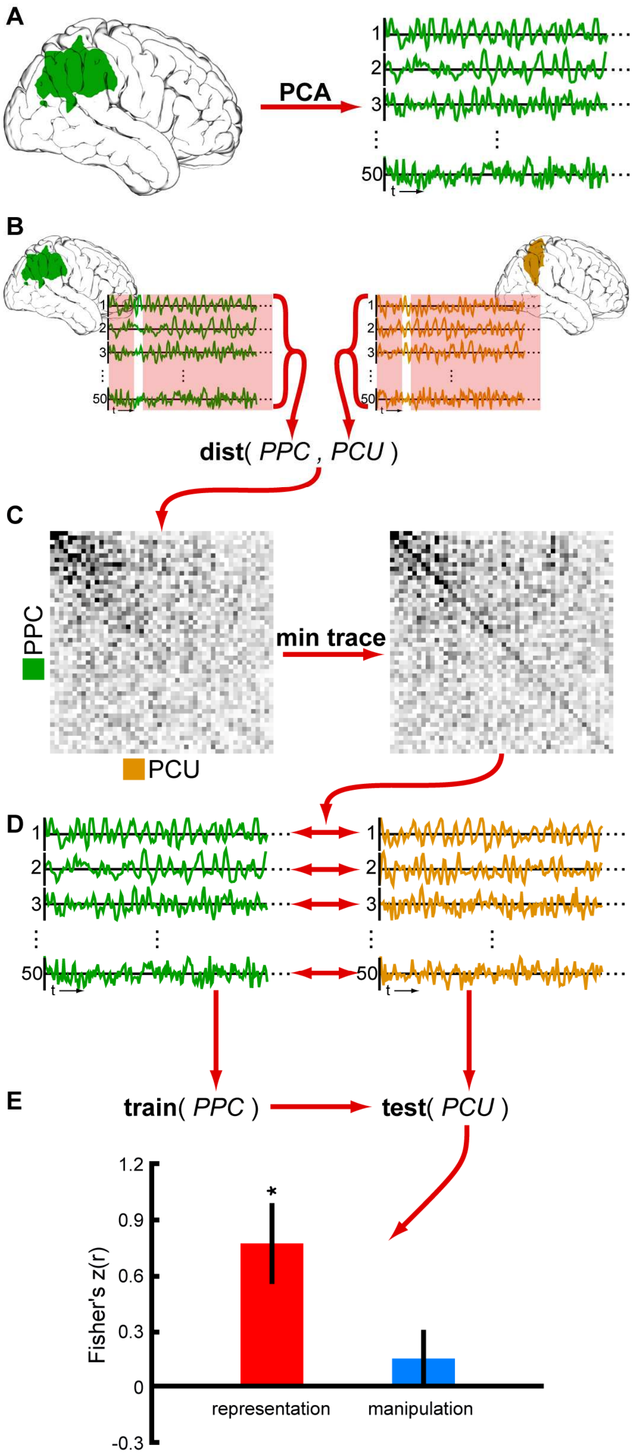


Figure S3.1. Trial schematic

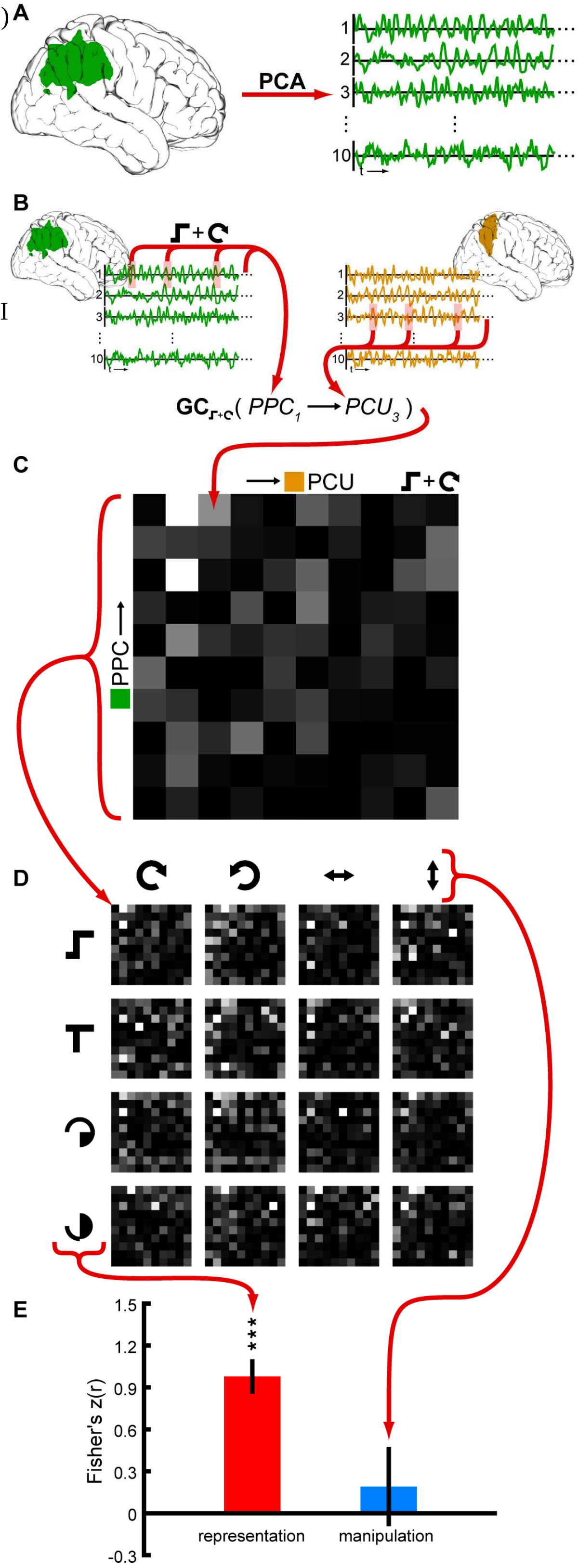
A 2s prompt screen indicated the shape and operation for the current trial. This was followed by a 6s blank screen during which the participant performed the indicated operation on the indicated shape. Next, a 2s test screen appeared, during which the participant indicated whether a displayed shape matched the output of the indicated operation. Finally, the participant was given feedback on their response.

Figure S3.2. Visual schematic of cross-classification analysis procedure



A. Functional data from each ROI (PPC shown here) were transformed from voxel space to 50 principal component signals using PCA. **B**. For a single cross-classification analysis between two ROIs (PPC and PCU here), the correlation distance between each pair of principal component signals was calculated. This calculation was performed independently for each classification fold, leaving out the test data from that fold (visualized here as a gap in the data that was used to calculate distances). This resulted in a 50 × 50 correlation distance matrix. **C**. The trace of this correlation distance matrix was minimized using the Hungarian algorithm in order to compute a matching of component signals between the two ROIs that maximized their pairwise similarity (i.e. minimized their correlation distance). **D**. This procedure resulted in a common 50-dimensional feature space shared between the two ROIs. Matched principal component signals between ROIs were maximally similar to each other. **E**. A cross-classification analysis was performed using these transformed functional data. The classifier was trained on data from one ROI (PPC in this case) and tested on data from the other ROI (PCU in this case). Other than the difference between training and testing data, the classification was carried out identically to that of the ROI-based analyses in Figure 3.2.

Figure S3.3. Visual schematic of information flow classification analysis procedure



A. Functional data from each ROI (PPC shown here) were transformed from voxel space to 10 principal component signals using PCA. **B**. For a given directed ROI pair (PPC to PCU here) and condition (Shape 1 and clockwise rotation here), the Granger causality from the source ROI to the destination ROI was calculated for each pair of principal component signals (PPC component 1 and PCU component 3 here), using only data from trials of that condition. **C**. This resulted in a 10 × 10 Granger-causal graph for each participant, directed ROI pair, and condition. **D**. The resulting 16 Granger-causal graphs for a given participant and directed ROI pair could be labeled based on either shape or operation. **E**. A classification analysis then proceeded as in the other analyses, except that either leave-one-shape-out or leave-one-operation-out cross validation was performed.

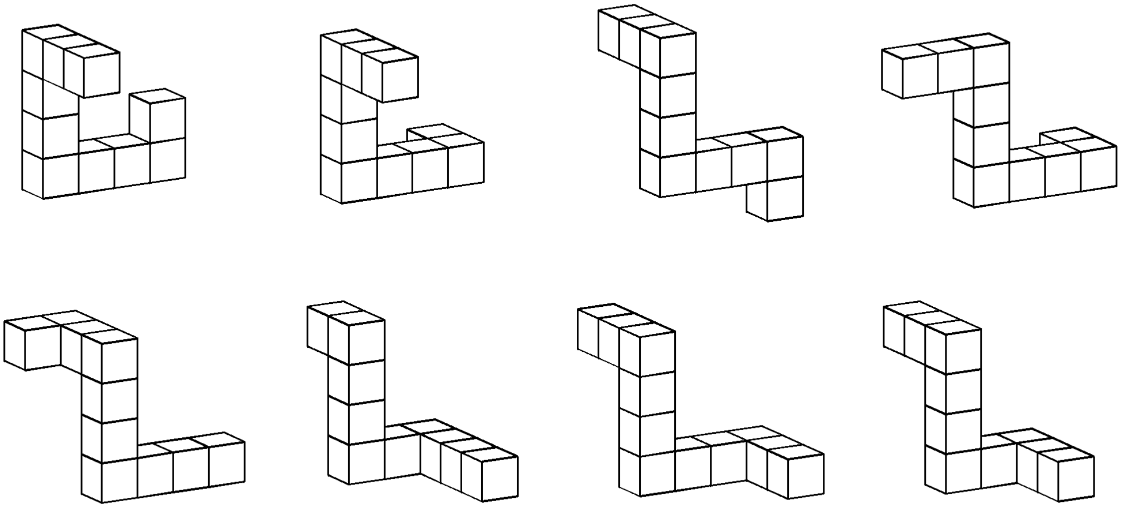
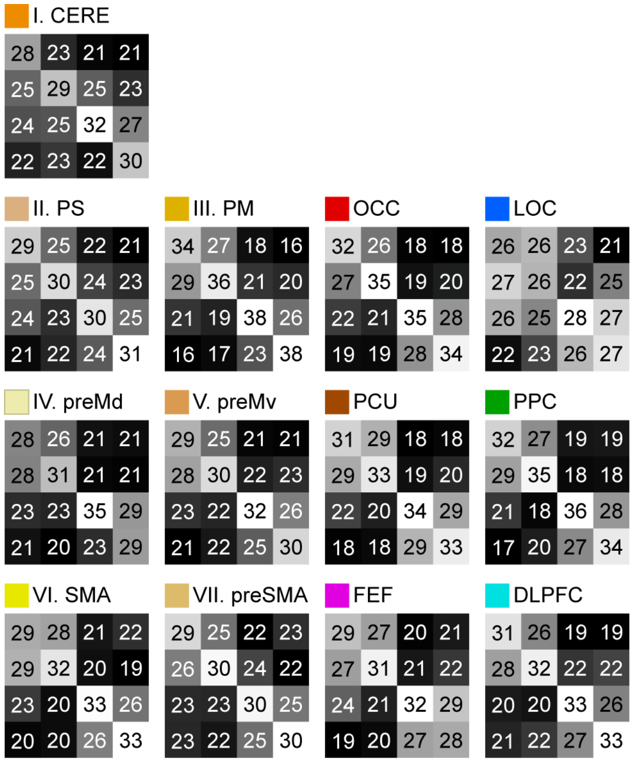


Figure S4.1. The eight figures used for mental rotation

Figures were shown during the prompt phase of the trial either as above or flipped across the y-axis, and unrotated, rotated 180° around the x-axis, or rotated 180° around the z-axis.



I. CERE

28	23	21	21
25	29	25	23
24	25	32	27
22	23	22	30

II. PS

29	25	22	21
25	30	24	23
24	23	30	25
21	22	24	31

III. PM

34	27	18	16
29	36	21	20
21	19	38	26
16	17	23	38

OCC

32	26	18	18
27	35	19	20
22	21	35	28
19	19	28	34

LOC

26	26	23	21
27	26	22	25
26	25	28	27
22	23	26	27

IV. preMd

28	26	21	21
28	31	21	21
23	23	35	29
21	20	23	29

V. preMv

29	25	21	21
28	30	22	23
23	22	32	26
21	22	25	30

PCU

31	29	18	18
29	33	19	20
22	20	34	29
18	18	29	33

PPC

32	27	19	19
29	35	18	18
21	18	36	28
17	20	27	34

VI. SMA

29	28	21	22
29	32	20	19
23	20	33	26
20	20	26	33

VII. preSMA

29	25	22	23
26	30	24	22
23	23	30	25
23	22	25	30

FEF

29	27	20	21
27	31	21	22
24	21	32	29
19	20	27	28

DLPFC

31	26	19	19
28	32	22	22
20	20	33	26
21	22	27	33

Figure S4.2. Individual mean confusion matrices from the ROI classification analysis

Compare to model similarity structure in Figure 1B. Values represent percent of cross-validation folds in which each target/prediction combination occurred. Color scaling for visualization was performed separately for each confusion matrix, because only relative values matter for the correlation analysis. Matrix elements are ordered as in Figure 1B, and abbreviations are as in Figure 2.

References

1. Logie RH (2003) Spatial and visual working memory: A mental workspace. *Psychol Learn Motiv* 42:37–78.

2. Schlegel A, Rudelson JJ, Tse PU (2012) White matter structure changes as adults learn a second language. *J Cogn Neurosci* 24:1664–70.

3. Schlegel A et al. (2013) Barking up the wrong free: readiness potentials reflect processes independent of conscious will. *Exp Brain Res* 229:329–335.

4. Alexander P et al. (2014) in *Surrounding Free Will: Philosophy, Psychology, Neuroscience*, ed Mele A (Oxford University Press, Oxford), pp 203–230.

5. Schlegel A et al. (2015) Hypnotizing Libet: Readiness potentials with non-conscious volition. *Conscious Cogn* 33:196–203.

6. Schlegel A et al. (2015) The artist emerges: Visual art learning alters neural structure and function. *Neuroimage* 105:440–51.

7. Baddeley AD (2003) Working memory: looking back and looking forward. *Nat Rev Neurosci* 4:829–39.

8. Kosslyn SM, Behrmann M, Jeannerod M (1995) The cognitive neuroscience of mental imagery. *Neuropsychologia* 33:1335–1344.

9. Uttal DH et al. (2013) The malleability of spatial skills: a meta-analysis of training studies. *Psychol Bull* 139:352–402.

10. Hegarty M (2004) Mechanical reasoning by mental simulation. *Trends Cogn Sci* 8:280–5.

11. Bassok M, Dunbar KN, Holyoak KJ (2012) Introduction to the special section on the neural substrate of analogical reasoning and metaphor comprehension. *J Exp Psychol Learn Mem Cogn* 38:261–263.

12. Hadamard J (1954) *The psychology of invention in the mathematical field* (Dover Publications, Inc., New York).

13. Penn DC, Holyoak KJ, Povinelli DJ (2008) Darwin’s mistake: explaining the discontinuity between human and nonhuman minds. *Behav Brain Sci* 31:109–30; discussion 130–178.

14. Matsuzawa T (2013) *Imagination: Human mind viewed from chimpanzee mind: Tetsuro Matsuzawa at TEDxYouth@Kyoto 2013*.

15. Matsuzawa T (2009) The chimpanzee mind: in search of the evolutionary roots of the human mind. *Anim Cogn* 12:S1–9.

16. Shepard RN, Metzler J (1971) Mental rotation of three-dimensional objects. *Science (80-)* 171:701–703.

17. Shepard RN, Feng C (1972) A chronometric study of mental paper folding. *Cogn Psychol* 3:228–243.

18. Paivio A (1978) Comparisons of mental clocks. *J Exp Psychol Hum Percept Perform* 4:61–71.

19. Addis DR, Wong AT, Schacter DL (2007) Remembering the past and imagining the future: common and distinct neural substrates during event construction and elaboration. *Neuropsychologia* 45:1363–77.

20. Finke RA, Slayton K (1988) Explorations of creative visual synthesis in mental imagery. *Mem Cognit* 16:252–257.

21. Blazhenkova O, Kozhevnikov M (2010) Visual-object ability: a new dimension of non-verbal intelligence. *Cognition* 117:276–301.

22. Baddeley AD, Hitch GJL (1974) in *Psychology of Learning and Motivation: Advances in Research and Theory, Vol. 8,* ed Bower GA (Academic Press, New York), pp 47–89.

23. Kirchner WK (1958) Age differences in short-term retention of rapidly changing information. *J Exp Psychol* 55:352–358.

24. Wechsler D (1939) *The Measurement of Adult Intelligence* (Williams & Witkins, Baltimore, MD).

25. Daneman M, Carpenter PA (1980) Individual differences in working memory and reading. *J Verbal Learning Verbal Behav* 19:450–466.

26. Baluch F, Itti L (2011) Mechanisms of top-down attention. *Trends Neurosci* 34:210–224.

27. Kyllonen PC, Christal RE (1990) Reasoning ability is (little more than) working-memory capacity?! *Intelligence* 14:389–433.

28. Klingberg T (2010) Training and plasticity of working memory. *Trends Cogn Sci* 14:317–24.

29. Jaeggi SM, Buschkuehl M, Jonides J, Perrig WJ (2008) Improving fluid intelligence with training on working memory. *Proc Natl Acad Sci* 105:6829–33.

30. Jaeggi SM, Buschkuehl M, Shah P, Jonides J (2013) The role of individual differences in cognitive training and transfer. *Mem Cognit*.

31. Redick TS et al. (2012) No evidence of intelligence improvement after working memory training: a randomized, placebo-controlled study. *J Exp Psychol Gen*.

32. Smith EE (1999) Storage and executive processes in the frontal lobes. *Science (80-)* 283:1657–1661.

33. Zacks JM (2008) Neuroimaging studies of mental rotation: a meta-analysis and review. *J Cogn Neurosci* 20:1–19.

34. Green AE, Kraemer DJM, Fugelsang JA, Gray JR, Dunbar KN (2012) Neural correlates of creativity in analogical reasoning. *J Exp Psychol Learn Mem Cogn* 38:264–72.

35. Salazar RF, Dotson NM, Bressler SL, Gray CM (2012) Content-Specific Fronto-Parietal Synchronization During Visual Working Memory. *Science (80-)* 338:1097–1100.

36. Jung RE, Haier RJ (2007) The Parieto-Frontal Integration Theory (P-FIT) of intelligence: converging neuroimaging evidence. *Behav Brain Sci* 30:135–87.

37. Christophel TB, Hebart MN, Haynes J-D (2012) Decoding the contents of visual short-term memory from human visual and parietal cortex. *J Neurosci* 32:12983–9.

38. Postle BR (2006) Working memory as an emergent property of the mind and brain. *Neuroscience* 139:23–38.

39. Crowe DA et al. (2013) Prefrontal neurons transmit signals to parietal neurons that reflect executive control of cognition. *Nat Neurosci* 16:1484–1491.

40. Lee S-H, Kravitz DJ, Baker CI (2013) Goal-dependent dissociation of visual and prefrontal cortices during working memory. *Nat Neurosci* 16:997–9.

41. Kane MJ, Engle RW (2002) The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: an individual-differences perspective. *Psychon Bull Rev* 9:637–71.

42. Harrison SA, Tong F (2009) Decoding reveals the contents of visual working memory in early visual areas. *Nature* 458:632–5.

43. Oh J, Kwon JH, Yang PS, Jeong J (2013) Auditory imagery modulates frequency-specific areas in the human auditory cortex. *J Cogn Neurosci* 25:175–87.

44. Ishai A, Ungerleider LG, Haxby J V (2000) Distributed neural systems for the generation of visual images. *Neuron* 28:979–990.

45. Sreenivasan KK, Vytlacil J, D’Esposito M (2014) Distributed and Dynamic Storage of Working Memory Stimulus Information in Extrastriate Cortex. *J Cogn Neurosci* 26:1141–1153.

46. Sporns O (2014) Contributions and challenges for network models in cognitive neuroscience. *Nat Neurosci* 17:652–60.

47. Tononi G (2008) Consciousness as integrated information: a provisional manifesto. *Biol Bull* 215:216–42.

48. Rumelhart DE, McClelland JL (1986) *Parallel Distributed Processing* (MIT Press, Cambridge, MA).

49. Turk-Browne NB (2013) Functional interactions as big data in the human brain. *Science (80-)* 342:580–4.

50. Bassett DS et al. (2010) Dynamic reconfiguration of human brain networks during learning. *Proc Natl Acad Sci* 1010.3775.

51. Van den Heuvel MP, Stam CJ, Kahn RS, Hulshoff Pol HE (2009) Efficiency of functional brain networks and intellectual performance. *J Neurosci* 29:7619–24.

52. Uttal WR (2003) *The new phrenology: the limits of localizing cognitive processes in the brain* (MIT Press, Cambridge, MA).

53. Haxby J V et al. (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science (80-)* 293:2425–30.

54. Haxby J V et al. (2011) A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron* 72:404–16.

55. Haxby J V, Connolly AC, Guntupalli JS (2014) Decoding Neural Representational Spaces Using Multivariate Pattern Analysis. *Annu Rev Neurosci*.

56. Kriegeskorte N, Mur M, Bandettini P (2008) Representational similarity analysis - connecting the branches of systems neuroscience. *Front Syst Neurosci* 2:1–28.

57. Van Essen DC (2013) Cartography and connectomes. *Neuron* 80:775–90.

58. Fox MD et al. (2005) The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc Natl Acad Sci* 102:9673–8.

59. Jones DK (2010) *Diffusion MRI: Theory, Methods, and Applications* (Oxford University Press, USA).

60. Lizier JT, Heinzle J, Horstmann A, Haynes J-D, Prokopenko M (2011) Multivariate information-theoretic measures reveal directed information structure and task relevant changes in fMRI connectivity. *J Comput Neurosci* 30:85–107.

61. Barnett L, Seth AK (2014) The MVGC multivariate Granger causality toolbox: a new approach to Granger-causal inference. *J Neurosci Methods* 223:50–68.

62. Norman KA, Polyn SM, Detre GJ, Haxby J V (2006) Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn Sci* 10:424–30.

63. Yourganov G et al. (2014) Pattern classification of fMRI data: applications for analysis of spatially distributed cortical networks. *Neuroimage* 96:117–32.

64. Friston KJ (2002) Beyond phrenology: What can neuroimaging tell us about distributed circuitry? *Annu Rev Neurosci* 25:221–50.

65. Dehaene S, Naccache L (2001) Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition* 79:1–37.

66. Baars BJ, Franklin S (2003) How conscious experience and working memory interact. *Trends Cogn Sci* 7:166–172.

67. Horikawa T, Tamaki M, Miyawaki Y, Kamitani Y (2013) Neural decoding of visual imagery during sleep. *Science (80-)*.

68. 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* 102:7338–43.

69. O'Reilly RC (2006) Biologically based computational models of high-level cognition. *Science (80-)* 314:91–4.

70. Miller J, Patterson T, Ulrich R (1998) Jackknife-based method for measuring LRP onset latency differences. *Psychophysiology* 35:99–115.

71. Miller EK, Erickson CA, Desimone R (1996) Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *J Neurosci* 16:5154–67.

72. Todd JJ, Marois R (2004) Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature* 428:751–4.

73. Xu Y, Chun MM (2006) Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature* 440:91–5.

74. Druzgal TJ, D'Esposito M (2003) Dissecting contributions of prefrontal cortex and fusiform face area to face working memory. *J Cogn Neurosci* 15:771–84.

75. Baddeley AD (1986) *Working Memory* (Oxford University Press, New York).

76. Margulies DS et al. (2009) Precuneus shares intrinsic functional architecture in humans and monkeys. *Proc Natl Acad Sci* 106:20069–74.

77. Cavanna AE, Trimble MR (2006) The precuneus: a review of its functional anatomy and behavioural correlates. *Brain* 129:564–83.

78. Vogt BA, Laureys S (2009) Posterior cingulate, precuneal & retrosplenial cortices: Cytology & components of the neural network correlates of consciousness. *Prog Brain Res* 6123:205–217.

79. Bostan AC, Dum RP, Strick PL (2013) Cerebellar networks with the cerebral cortex and basal ganglia. *Trends Cogn Sci*:1–14.

80. Tanaka K (1996) Inferotemporal cortex and object vision. *Annu Rev Neurosci* 19:109–39.

81. 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* 20:11–8.

82. Schall JD (2004) On the role of frontal eye field in guiding attention and saccades. *Vision Res* 44:1453–67.

83. 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* 108:4230–5.

84. 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.

85. 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* 24:3917–25.

86. 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* 21:890–904.

87. Formisano E, Linden DEJ, Salle F Di, Trojano L (2002) Tracking the mind's image in the brain I: time-resolved fMRI during visuospatial mental imagery. *Neuron* 35:185–194.

88. Smith SM et al. (2004) Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* 23 Suppl 1:S208–19.

89. Dale AM, Fischl B, Sereno MI (1999) Cortical surface-based analysis: I. segmentation and surface reconstruction. *Neuroimage* 9:179–194.

90. Hanke M et al. (2009) PyMVPA: A python toolbox for multivariate pattern analysis of fMRI data. *Neuroinformatics* 7:37–53.

91. Tong F (2013) Imagery and visual working memory: one and the same? *Trends Cogn Sci*:4–5.

92. Insel TR, Landis SC, Collins FS (2013) The NIH brain initiative. *Science (80-)* 340:687–688.

93. Markram H (2012) The Human Brain Project. *Sci Am* 306:50–55.

94. Graham DJ, Rockmore D (2011) The packet switching brain. *J Cogn Neurosci* 23:267–76.

95. Schlegel A et al. (2013) Network structure and dynamics of the mental workspace. *Proc Natl Acad Sci* 110:16277–16282.

96. Ester EF, Serences JT, Awh E (2009) Spatially global representations in human primary visual cortex during working memory maintenance. *J Neurosci* 29:15258–65.

97. Friston KJ (2011) Functional and effective connectivity: a review. *Brain Connect* 1:13–36.

98. Friston K, Moran R, Seth AK (2013) Analysing connectivity with Granger causality and dynamic causal modelling. *Curr Opin Neurobiol* 23:172–178.

99. Wen X, Rangarajan G, Ding M (2013) Is Granger Causality a Viable Technique for Analyzing fMRI Data? *PLoS One* 8.

100. Seghier ML, Friston KJ (2013) Network discovery with large DCMs. *Neuroimage* 68:181–91.

101. Xue G et al. (2010) Greater neural pattern similarity across repetitions is associated with better memory. *Science (80-)* 330:97–101.

102. Schurger A, Pereira F, Treisman A, Cohen JD (2010) Reproducibility distinguishes conscious from nonconscious neural representations. *Science (80-)* 327:97–9.

103. Baldauf D, Desimone R (2014) Neural mechanisms of object-based attention. *Science (80-)* 344:424–7.

104. Bassett DS, Gazzaniga MS (2011) Understanding complexity in the human brain. *Trends Cogn Sci*:1–10.

105. Bressler SL, Menon V (2010) Large-scale brain networks in cognition: emerging methods and principles. *Trends Cogn Sci* 14:277–290.

106. Bullmore E, Sporns O (2009) Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat Rev Neurosci* 10:186–98.

107. Kuhn HW (1955) The Hungarian method for the assignment problem. *Nav Res Logist Q* 2:83–97.

108. Schlegel A, Alexander P, Tse PU (2015) Information processing in the mental workspace is fundamentally distributed. *Under Rev*.

109. Cohen MS et al. (1996) Changes in cortical activity during mental rotation. A mapping study using functional MRI. *Brain* 119:89–100.

110. Kosslyn SM, DiGirolamo GJ, Thompson WL, Alpert NM (1998) Mental rotation of objects versus hands: neural mechanisms revealed by positron emission tomography. *Psychophysiology* 35:151–61.

111. Langner R et al. (2013) Translating working memory into action: Behavioral and neural evidence for using motor representations in encoding visuo-spatial sequences. *Hum Brain Mapp* 00:1–20.

112. Michelon P, Vettel JM, Zacks JM (2006) Lateral somatotopic organization during imagined and prepared movements. *J Neurophysiol* 95:811–22.

113. Sack AT, Lindner M, Linden DEJ (2007) Object-and direction-specific interference between manual and mental rotation. *Percept Psychophys* 69:1435–1449.

114. Kosslyn SM, Thompson WL, Wraga M, Alpert NM (2001) Imagining rotation by endogenous versus exogenous forces: distinct neural mechanisms. *Neuroreport* 12:2519–2525.

115. Siegel M, Buschman TJ, Miller EK (2015) Cortical information flow during flexible sensorimotor decisions. *Science (80-)* 348:1352–1355.

116. Baars BJ (2002) The conscious access hypothesis: origins and recent evidence. *Trends Cogn Sci* 6:47–52.

117. Zeki S, Bartels A (1999) Toward a theory of visual consciousness. *Conscious Cogn* 8:225–59.

118. Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9:97–113.

119. Hayashi M, Takeshita H (2009) Stacking of irregularly shaped blocks in chimpanzees (Pan troglodytes) and young humans (Homo sapiens). *Anim Cogn* 12 Suppl 1:S49–58.

120. Povinelli DJ, Dunphy-Lelii S (2001) Do chimpanzees seek explanations? Preliminary comparative investigations. *Can J Exp Psychol* 55:187–95.

121. Fujita K, Matsuzawa T (1989) Comparison of the Representational Abilities of Chimpanzees and Humans: Short-term Memory Reproduction and Mental Rotation. *Primate Res* 5:58–74.

122. Potì P, Hayashi M, Matsuzawa T (2009) Spatial construction skills of chimpanzees (Pan troglodytes) and young human children (Homo sapiens sapiens). *Dev Sci* 12:536–48.

123. Inoue S, Matsuzawa T (2007) Working memory of numerals in chimpanzees. *Curr Biol* 17:R1004–5.

124. Ludwig VU, Adachi I, Matsuzawa T (2011) Visuoauditory mappings between high luminance and high pitch are shared by chimpanzees (Pan troglodytes) and humans. *Proc Natl Acad Sci* 108:20661–5.

125. Gillan DJ, Premack D, Woodruff G (1981) Reasoning in the chimpanzee: I. Analogical reasoning. *J Exp Psychol Anim Behav Process* 7:1–17.

126. Vauclair J, Fagot J, Hopkins WD (1993) Rotation of mental images in baboons when the visual input is directed to the left cerebral hemisphere. *Psychol Sci* 4:99–103.

127. Hopkins WD, Fagot J, Vauclair J (1993) Mirror-image matching and mental rotation problem solving by baboons (Papio papio): unilateral input enhances performance. *J Exp Psychol Gen* 122:61–72.

128. Hollard VD, Delius JD (1982) Rotational invariance in visual pattern recognition by pigeons and humans. *Science (80-)* 218:3–5.

129. Pearson DG, Logie RH, Gilhooly KJ (1999) Verbal representations and spatial manipulation during mental synthesis. *Eur J Cogn Psychol* 11:295–314.

130. Brandimonte MA, Hitch GJL, Bishop DVM (1992) Verbal recoding of visual stimuli impairs mental image transformations. *Mem Cognit* 20:449–55.

131. MacLeod CM, Hunt EB, Mathews NN (1978) Individual differences in the verification of sentence—picture relationships. *J Verbal Learning Verbal Behav* 17:493–507.

132. Clark HH, Chase WG (1972) On the process of comparing sentences against pictures. *Cogn Psychol* 3:472–517.

133. Arden R, Chavez RS, Grazioplene R, Jung RE (2010) Neuroimaging creativity: a psychometric view. *Behav Brain Res* 214:143–156.

134. Dietrich A, Kanso R (2010) A review of EEG, ERP, and neuroimaging studies of creativity and insight. *Psychol Bull* 136:822–48.

135. Diamond A (2007) Interrelated and interdependent. *Dev Sci* 10:152–8.

136. Herrmann E, Call J, Hernàndez-Lloreda MV, Hare B, Tomasello M (2007) Humans have evolved specialized skills of social cognition: the cultural intelligence hypothesis. *Science (80-)* 317:1360–6.

137. Maus GW, Fischer J, Whitney D (2013) Motion-dependent representation of space in area MT+. *Neuron* 78:554–62.

138. Schlegel A, Konuthula D, Alexander P, Blackwood E, Tse PU (2015) Widespread information sharing integrates the motor network into the mental workspace during mental rotation. *Under Rev*.