

UC Berkeley

UC Berkeley Previously Published Works

Title

Dissociable Neural Systems Support the Learning and Transfer of Hierarchical Control Structure

Permalink

<https://escholarship.org/uc/item/38z41184>

Journal

Journal of Neuroscience, 40(34)

ISSN

0270-6474

Authors

Eichenbaum, Adam
Scimeca, Jason M
D'Esposito, Mark

Publication Date




2020-08-19

DOI

10.1523/jneurosci.0847-20.2020

Peer reviewed

Dissociable Neural Systems Support the Learning and Transfer of Hierarchical Control Structure

 Adam Eichenbaum,  Jason M. Scimeca, and  Mark D'Esposito

Helen Wills Neuroscience Institute, University of California, Berkeley, California 94720

Humans can draw insight from previous experiences to quickly adapt to novel environments that share a common underlying structure. Here we combine functional imaging and computational modeling to identify the neural systems that support the discovery and transfer of hierarchical task structure. Human subjects (male and female) completed multiple blocks of a reinforcement learning task that contained a global hierarchical structure governing stimulus–response action mapping. First, behavioral and computational evidence showed that humans successfully discover and transfer the hierarchical rule structure embedded within the task. Next, analysis of fMRI BOLD data revealed activity across a frontoparietal network that was specifically associated with the discovery of this embedded structure. Finally, activity throughout a cingulo-opercular network supported the transfer and implementation of this discovered structure. Together, these results reveal a division of labor in which dissociable neural systems support the learning and transfer of abstract control structures.

Key words: fMRI; hierarchy; learning; learning to learn; reinforcement learning; transfer

Significance Statement

A fundamental and defining feature of human behavior is the ability to generalize knowledge from the past to support future action. Although the neural circuits underlying more direct forms of learning have been well established over the last century, we still lack a solid framework from which to investigate more abstract, higher-order human learning and knowledge generalization. We designed a novel behavioral paradigm to specifically isolate a learning process in which previous knowledge, rather than directly indicating the correct action, instead guides the search for the correct action. Moreover, we identify that this learning process is achieved via the coordinated and temporally specific activity of two prominent cognitive control brain networks.

Introduction

Whether it is learning how to drive a new car or interacting with an unfamiliar social group, humans show remarkable adaptability inferring the correct action given minimal information. Such learning usually occurs via trial and error where feedback works to guide future behavior. These problem-solving approaches are routinely accelerated by generalizing previous knowledge (Woodworth and Thorndike, 1901). When simple stimulus–response mappings are learned in experimental settings, responses learned in one context can be directly transferred to a subsequent context, leading to an immediately observable benefit (Behrens et al., 2007; Collins et al., 2014; Collins and Frank, 2016). Although humans can encounter

scenarios such as these (e.g., opening computer applications on a Windows vs Apple operating system), humans also encounter settings where this approach leads to failure (e.g., starting computer programs on Windows/Apple vs Linux). In these cases, it is advantageous to instead leverage prior knowledge to guide the learning of the correct behavior, a process known as “learning to learn” (Harlow, 1949; Kemp et al., 2010; Bavelier et al., 2012; Botvinick et al., 2019). While the behavioral and neurobiological underpinnings of more direct types of transfer have been relatively well characterized (Collins et al., 2014; Collins and Frank, 2016), the neural systems and mechanisms underlying this more abstract form of transfer remain poorly understood.

Everyday experiences are often structured hierarchically where actions and experiences are influenced by superordinate contexts and rules. For example, when traveling away from home it is common to pack a bag with clothes and overnight necessities. However, the rule that restricts packing small-volume liquids is only relevant in certain contexts: when traveling by airplane, not by car. By grouping these sets of behaviors and experiences hierarchically, one is able to easily generalize rules from one context to another, and even to contexts that have not yet been personally experienced. One way in which learned

Received Apr. 7, 2020; revised May 15, 2020; accepted July 8, 2020.

Author contributions: A.E. and J.M.S. designed research; A.E. performed research; A.E. contributed unpublished reagents/analytic tools; A.E. analyzed data; A.E., J.M.S., and M.D. wrote the paper.

The authors declare no competing financial interests.

This work was funded by a from the National Institutes of Health Grant MH-63901. We thank Michael Frank and Anne Collins for sharing the original, and discussing appropriate revisions to, the mixture of experts code.

Correspondence should be addressed to Adam Eichenbaum at eichenbaum@berkeley.edu.

<https://doi.org/10.1523/JNEUROSCI.0847-20.2020>

Copyright © 2020 the authors

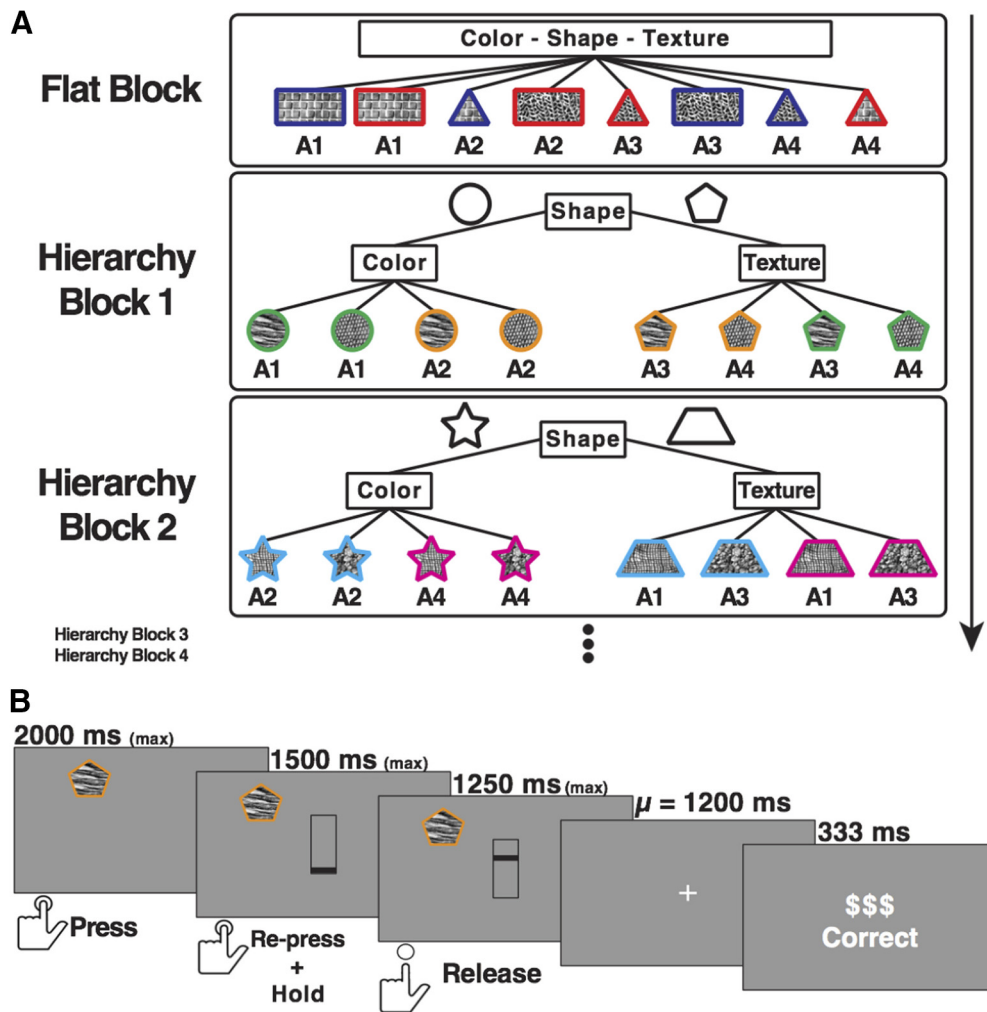


Figure 1. Schematic depiction of experimental logic and trial sequence. **A**, Schematic of task design showing example stimulus-to-action mappings. Subjects completed five blocks in total throughout the experiment. The stimuli in each block varied along the following three dimensions: shape, color, and texture. Each block contained two stimulus features for each dimension (e.g., two shapes) and the specific features changed for each block. The first block contained a flat policy structure such that the mapping between stimuli and actions (e.g., A1, A2) was randomly assigned. The remaining four blocks all shared the same global second-order policy structure: the shape of the stimulus indicated whether first-order rules were determined by color or texture on the current trial. In the example shown for hierarchical block 1, a circular stimulus indicated that color determined the correct action (i.e., green pairs with A1, orange pairs with A2). Hierarchical blocks included an irrelevant fourth dimension (stimulus position on screen) that is not shown here. **B**, Schematic of trial design. Trials began with stimulus presentation, after which subjects had up to 2 s to respond by pressing one of four buttons mapped to their right index, middle, ring, and pinky fingers. Subjects then indicated their confidence in their answer by positioning a black bar along the screen in a one-shot manner. Subjects received auditory and visual feedback following a jittered interstimulus interval.

hierarchical structures may be generalized to novel contexts is the creation of task sets or task structures that span across related contexts regardless of low-level features (Collins and Frank, 2013).

Although the combination of task sets and hierarchical processing provides a natural candidate solution for how learned hierarchical structure is generalized, the neural basis of these cognitive processes has typically been studied in isolation. Growing neurobiological and computational evidence suggests that the frontal cortex facilitates hierarchically structured behavior (Koechlin et al., 2003; Badre and D'Esposito, 2007; Badre et al., 2010; Frank and Badre, 2012; Collins and Frank, 2013; Nee and D'Esposito, 2016; Wang et al., 2018). Specifically, left lateral frontal cortex is organized along a rostrocaudal gradient wherein more rostral regions support the learning and execution of increasingly higher-order hierarchically structured rules. It remains undetermined whether these regions, likely those more rostrally, additionally support the transfer of learned structure (Badre and Nee, 2018). In addition, the processing of task sets

has generally been related to activity in frontal cortex, as well as to a distributed “cingulo-opercular” (CO) network of regions (Dosenbach et al., 2008; Sakai, 2008). As generalization of hierarchical knowledge involves the integration of information across multiple sources, it is likely that a network of regions spanning beyond frontal cortex will be involved.

To investigate the discovery and transfer of abstract hierarchical structure, we designed a hierarchical reinforcement learning task that promotes the creation and transfer of a superordinate structure (Fig. 1). Critically, although each block contained entirely new stimulus features, a global second-order hierarchical rule remained. Therefore, successful performance of a previous block conveyed no immediate advantage on subsequent blocks. However, knowledge of the correct hierarchical structure instead facilitated a more rapid learning of the correct response mappings. We leveraged converging computational modeling approaches to confirm (1) when subjects first discovered the global hierarchical structure, and (2) that rapid learning occurred thereafter, indicating transfer of learned structure. Last, we used fMRI to investigate

the left lateral frontal regions along the predefined rostrocaudal gradient, as well as broader neural systems, that support these two processes.

Materials and Methods

Human subject details

Thirty-two healthy right-handed subjects (age range, 18–29 years; mean = 19.63; SD = 2.54; 20 females) with normal or corrected-to-normal vision participated in the study at the University of California, Berkeley. Target sample size was based on prior relevant literature (Badre et al., 2010; Collins and Frank, 2016; Nee and D'Esposito, 2016). Eight subjects were excluded from all behavioral analyses (four subjects failed to complete the entire session, two subjects did not follow the instructions, and two subjects exhibited subthreshold behavioral performance; no above-chance performance in any hierarchical block [i.e., state-space model outcomes of the distribution around the probability to produce a correct response always included the chance-level performance value]). Five additional subjects were excluded from all fMRI analyses [one subject because of above-threshold in-scanner motion (>2.5 mm in X, Y, or Z across all blocks), one subject for atypical anatomic data, and three subjects because of scanner image reconstruction failure].

All behavioral analyses presented here include data from the 24 subjects for whom we obtained a complete behavioral dataset (age range, 18–24 years; mean = 19.25; SD = 1.75; 16 females). All fMRI analyses presented here include data from the 19 subjects for whom we obtained a complete behavioral and fMRI dataset (age range, 18–24 years; mean = 19.26; SD = 1.88; 13 females). Behavioral analyses restricted to these 19 subjects show the same results as the 24-subject group. All research protocols were approved by the Committee for Protection of Human Subjects at the University of California, Berkeley. Informed and written consent was obtained from all subjects before participation.

Experimental design and statistical analyses

Task design. In the current experiment, we designed a reinforcement learning task (inspired by Badre et al., 2010) that required learning multiple distinct second-order hierarchical rules (hereafter referred to as second-order policy) that shared a global hierarchical policy structure (hierarchical blocks). Specifically, a second-order hierarchical policy determined that the shape of the stimulus cued first-order rules defined by other stimulus dimensions (e.g., if the stimulus is a square, perform action 1 for red squares, and action 2 for blue squares; however, if the stimulus is a circle, perform action 3 for striped circles, and action 4 for checkered circles, regardless of other stimulus features). Thus, subjects who learn the block-specific hierarchical policy in successive blocks can discover the existence of the global hierarchical structure. By transferring their knowledge of the global hierarchical structure to subsequent blocks, subjects can more rapidly learn the block-specific hierarchical policy.

Subjects completed one block containing a rule set in which there was no higher-order structure [flat block (Flat)] and four hierarchical blocks (Hiers) while inside the scanner (Fig. 1). Subjects viewed stimuli that varied along three or four dimensions, as follows: shape, color, black-and-white image pattern (referred to as “texture”), and stimulus position on the screen (hierarchical blocks only; Fig. 1A). For each block, stimulus dimensions could vary between two features (e.g., color: red/blue; shape: square/circle), resulting in 8 unique stimuli in the flat block and 16 unique stimuli in each hierarchical block. All blocks contained unique features, and thus subjects had to learn entirely new stimulus–response mappings for each block. We assigned stimulus features to blocks by random assignment.

Stimuli. Stimuli were generated using PsychoPy (Peirce, 2007, 2008). Colors included red, green, blue, yellow, magenta, cyan, white, maroon, black, and orange. Shapes included a circle, square, rectangle, triangle, pentagon, rhombus, trapezoid, six-sided star, oval, and tear drop. Texture images were sourced from the Normalized Brodatz Texture Database (Abdelmounaime and Dong-Chen, 2013). These images included close-up photographs of various real-world textures, such as tree rings, sand dunes, snakeskin, and bubbles. Subjects did not report

difficulty in discriminating between textures (Fig. 1A). The stimuli generally subtended $\sim 7.5^\circ$ of visual angle. The stimulus position in the hierarchical blocks was computed along an invisible circle positioned at the center of the screen with a radius subtending $\sim 7.5^\circ$ of visual angle. The eight locations along this circle began at 27.5° clockwise from the vertical meridian and were equally spaced by 45° increments.

Flat block. The flat block consisted of 20 repetitions of each stimulus for a total of 160 trials. Stimulus order was randomized within each set of eight trials so as to restrict the range of the number of trials between stimulus repetitions. On average, each stimulus was viewed once every eight trials, ranging from 0 to 15. Before the start of the block, subjects had the opportunity to view all eight stimuli created for the upcoming block. All stimuli were presented on screen in a 2×4 array and remained on screen until the subject chose to proceed. No additional instructions were provided regarding the viewing of the stimuli.

Trials began with the presentation of the stimulus slightly offset left of the center of the screen for a maximum (max) of 2000 ms (Fig. 1B). Stimulus composition included a black-and-white image cropped into a specific shape with a colored border. Subjects were instructed to respond to the presentation of the stimulus by pressing one of four buttons mapped to their right index, middle, ring, and pinky fingers. Responding within 2000 ms advanced the trial to the confidence response phase. This phase began with the appearance of a vertical rectangle offset right of center, with a horizontal black bar appearing either on the bottom or top of the rectangle. To indicate their confidence that their most recent response to the stimulus was correct, subjects had 1500 ms to re-press and hold down the button they had just pressed. By re-pressing the button, the black bar began to move away from its starting position at a constant rate until it reached the other side of the rectangle, a process that lasted up to 1250 ms. Regardless of the starting position of the bar, the top of the rectangle indicated 100% confidence in their answer being correct, while the bottom of the rectangle indicated 0%. Subjects were instructed to be as precise as possible with their confidence rating. Following the release of the held-down button, or after 1250 ms, both the stimulus and confidence probe disappeared from the screen. Following a pseudorandom interstimulus interval (200, 1200, or 2200 ms), subjects received audiovisual feedback. Correct feedback involved the presentation of the word “Correct” and “\$\$\$\$” stacked vertically in the center of the screen, as well as a pleasant tone. Incorrect feedback contained the word “Incorrect” and an unpleasant tone. The feedback stimulus persisted for 333 ms. Feedback was 100% valid. Following feedback, a fixation cross was displayed for the remainder of the trial duration (5283, 6283, or 7283 ms, depending on the duration of the interstimulus interval on that trial). The next trial then began after a variable intertrial interval (ITI) with a mean of 1500 ms (range, 500–4500 ms). The order of ITIs within a block was optimized to permit estimation of the event-related response using optseq2 (Dale, 1999).

We assigned two stimuli to each response option so that each button had a 25% chance of being correct on any given trial. Stimulus–response mappings were independent from one another, such that no higher-order structure was present, thus requiring each response to be learned individually. Following the final trial, mean block accuracy was presented on screen.

Hierarchical block. We designed the hierarchical blocks identically to the flat block with the following exceptions. (1) Stimuli now included a fourth dimension: position on screen. In each of the four hierarchical blocks, the stimulus could appear in one of two locations on screen. These locations were semirandomly selected from eight possible equidistant positions along an invisible aperture around the center of the screen. We assigned the positions in each block in pairs, such that each pair was offset in both the x -axis and y -axis so as to create as large a separation and difference as possible. Position was not included in the flat block as pilot testing indicated subjects were unable to learn above chance 16 independent stimuli across four button responses in an appropriate amount of time. (2) The number of stimulus repetitions decreased from 20 to 6, resulting in a decrease in the number of total trials from 160 to 96 per block. (3) Given the new position dimension, the confidence probe was moved to the center of the screen so as not to interfere with the stimulus. (4) The position-on-screen dimension was not included in

the preblock stimulus presentation screen in which all eight stimuli were shown.

Last, and most critically, all hierarchical blocks contained a second-order policy relationship that subjects could discover and transfer across blocks so as to facilitate their learning, instead of learning 16 independent stimulus–response mappings. Specifically, the shape dimension cued first-order rules dependent on either the colors or textures and, as a result, screen position was irrelevant. By learning and exploiting this structure, the number of rules to be learned decreased to four (i.e., two rules for color, two rules for texture). The same second-order policy relationship was maintained across blocks, in that the shape dimension (shape) always cued rules based on either color or texture dimensions.

Instructions and training protocol. Before performing the task inside the MRI scanner, all subjects completed a training session on a desktop computer to make sure they understood the task and could perform it adequately. After obtaining experimental consent, and confirming both study and MRI scanner eligibility, subjects reviewed the instructions of the task. Along with visual aids on the computer, the experimenter described the task such that subjects knew they had to learn stimulus–response mappings across multiple task blocks; however, no information was provided that could cue subjects to the hierarchical structure of the task. Subjects then practiced the confidence-reporting component of the task in a guided environment using stimuli not present in the real experiment. Subjects received guided instructions indicating which button to press and how confident they should report feeling for each practice trial. Instructed confidence levels included 0, 15, 35, 50, 65, 85, and 100%. Subjects needed to place the confidence bar at the appropriate location along the vertical rectangle to match the instructed confidence level across 21 practice trials (three repetitions of each level). A 93% accuracy criterion was required to progress. Subjects had to repeat the 21-trial practice block until they met the criterion. The timing of all events matched that of the real experiment.

Following completion of the confidence reporting practice, subjects then performed 24 practice trials of a flat block, using the same stimuli as before. Just as in the real task, subjects had to learn eight independent stimulus response mappings across four buttons using the feedback provided at the end of each trial. No performance criterion was included, as the goal of this practice session was to familiarize subjects with the components of the task in real time.

Upon completion of the practice session, subjects were then escorted to the MRI scanner suite and placed inside the scanner. During the acquisition of an anatomic scan (details below), subjects went through the practice instructions and confidence-reporting session again so as to both become accustomed the MRI-compatible four-button response box and to being inside the active scanner. Subjects received compensation at a rate of \$20/h and could earn a bonus of up to \$10 based on their overall trial accuracy.

Statistical analyses of behavioral data. Analyses of behavioral data included the use of paired *t* tests with one exception. When analyzing the number of learned second-order rules across blocks, we used Wilcoxon sign-ranked tests because of the nonparametric nature of the data (i.e., subjects could learn either zero, one, or two second-order rules per block) and the within-subjects design of the study. In addition, the stimulus dimension of position-on-screen was fully ignored in all analyses of the data.

Statistical analyses of fMRI data. Whole-brain analyses were performed in SPM (Statistical Parametric Mapping; www.fil.ion.ucl.ac.uk/spm), and cluster correction was performed at the familywise error rate of $p = 0.05$, using $p = 0.001$ as the cluster defining threshold. Correlations between fMRI data and behavioral data were performed using standard parametric linear regression, as well as nonparametric rank-ordered regression to better control for potential outliers in the dataset. Results for each assessment are presented in tandem throughout the article.

fMRI data acquisition

Whole-brain imaging was performed at the Henry H. Wheeler Jr. Brain Imaging Center at the University of California, Berkeley, using a Siemens 3 T Trio MRI scanner using a 32-channel head coil. Functional imaging data were acquired with a gradient echo echoplanar pulse

sequence using a multiband acceleration factor of 4 (TR = 1000 ms; TE = 33 ms; flip angle = 40°; array = 84 × 84; 52 slices; voxel size = 2.5 mm isotropic). T1-weighted (T1w) MP-RAGE anatomic images were collected as well (TR = 2300 ms; TE = 2.98 ms; flip angle = 9°; array = 256 × 256; 160 slices; voxel size = 1 mm isotropic). The subject's head movement was restricted using foam padding. Auditory feedback was presented through in-ear headphones connected to the stimulus presentation computer. The flat block consisted of a single run of 1290 TRs, while each hierarchical block consisted of 760 TRs.

fMRI data preprocessing

Preprocessing was performed using FM RIPREP version 1.0.2 (Esteban et al., 2018), a Nipype-based tool (Gorgolewski et al., 2011). Each T1w volume was corrected for intensity nonuniformity using N4BiasFieldCorrection version 2.1.0 (Tustison et al., 2010) and skull stripped using ANTs BrainExtraction. Spatial normalization to the ICBM 152 Nonlinear Asymmetrical template version 2009c was performed through nonlinear registration with the antsRegistration tool of ANTs version 2.1.0 (Avants et al., 2008), using brain-extracted versions of both T1w volume and template. Brain tissue segmentation of CSF, white matter, and gray matter was performed on the brain-extracted T1w using FSL fast (Zhang et al., 2001). Functional data were motion corrected using FSL mcflirt (Jenkinson et al., 2002). This was followed by coregistration to the corresponding T1w using boundary-based registration (Greve and Fischl, 2009) with 9 df, using flirt (FSL). Motion-correcting transformations, BOLD-to-T1w transformation, and T1w-to-template (MNI) warp were concatenated and applied in a single step using ANTs ApplyTransforms using Lanczos interpolation. Slice-timing correction was not performed. Preprocessed data were spatially smoothed with an 8 mm FWHM isotropic Gaussian kernel. Motion estimates used for subject exclusion were calculated using the SPM realign function.

Computational modeling: state-space model

Trial responses were modeled with a state-space modeling approach (Smith et al., 2004) to produce learning curves. The model outputs trial-by-trial estimates of the probability of a correct response on each trial, as well as a 90% confidence interval around each estimate. Similar to the study by Badre et al. (2010), our analyses focused on the following metrics derived from the learning curve: (1) the trial for which the 90% confidence interval no longer included chance performance, referred to as the “learning trial”; (2) the maximal first derivative of the learning curve, which indexes the rate of learning; and (3) the maximal second derivative, which indexes the rate of change in one's learning rate.

Computational modeling: mixture of experts model

We make use of a hybrid Bayesian-reinforcement learning mixture of experts (MoE) model previously used by Frank and Badre (2012) to estimate subjects' attention to various hypothesis states that we assume are being tested while subjects perform the task. Given the observed stimuli and responses, the MoE model estimates individual subjects' attention to likely hypotheses about the relationship between context (i.e., the features of the stimulus) and action (i.e., the available button responses) in each task block. Each expert in the model represents a prediction about how a stimulus feature, or a combination of features, relates to the likelihood of obtaining a reward given the motor actions available to the subject. For example, the “shape expert” could learn the likelihood of obtaining a reward based only on the shape of the stimulus. For each trial, the expert makes its prediction about what action is likely to be correct given its assigned feature, and experts who contribute accurate predictions are rewarded while experts providing unreliable predictions are not. For hierarchical experts, the model makes predictions about subordinate stimulus dimensions (i.e., color or texture) contingent on the identity of a third, superordinate dimension (i.e., shape), such that weights assigned to predictions about each subordinate dimension are dynamically gated based on the feature of the superordinate dimension (e.g., circle vs square). The MoE model also assigns attentional weights to experts that learn the overall reliability of hierarchical versus flat predictions based on the reliability of all the hierarchical and flat experts, respectively.

For the current study, we adapted the model to allow for individual fits to each hierarchical expert. As the original version used a single

hyperparameter across all three hierarchical experts, thus preventing the ability to estimate different initial weights, we instead modeled each hierarchical expert with a separate parameter. We also removed the decay parameter originally used to model the degree to which the attentional weights of the current block carried over into the next block. Instead, we modeled a separate set of parameters for the various experts in each block. By removing the decay parameter, and modeling each block independently, we ensure that the model is incapable of being biased by the previous block. As a result, any differences between blocks in the parameter values, as well as the computed attentional weights at the beginning of the block, are the result of that data of the block alone.

Specifically, subjects' beliefs about reward probability for each of the four available response options (per expert) were modeled as a beta distribution and updated via Bayes' rule. For example, the color expert was updated by the following:

$$p(\theta_{R,C}|r_1 \dots r_n) \propto p(r_1 \dots r_n|\theta_{R,C})p(\theta_{R,C}),$$

where $\theta_{R,C}$ reflects the parameters determining the belief distribution about rewards given the presence of color C and the choice of response R , with $r_1 \dots r_n$ being the rewards seen so far when this specific R was chosen. Next, the probability of selecting each response is calculated by comparing the means μ of their reward distributions using a softmax function. For example, the probability of the color expert selecting R_i on trial t was as follows:

$$p_{R_i}^C(t) = \frac{e^{\frac{\mu_{R_i}^C(t)}{\kappa}}}{\sum_j e^{\frac{\mu_{R_j}^C(t)}{\kappa}}},$$

where κ governs the choice stochasticity, with lower (higher) values reflecting less (more) noise. The same computations were performed for each expert e , including a shape expert and a texture expert, as well as all two-way conjunctions, and finally the full three-way conjunction. The model represents subjects' beliefs about the reliability of each expert with another beta distribution, and again uses Bayes' rule to learn the probability that the expert is reliable. For example, the color expert is updated as follows:

$$p(\theta_C|r'_1 \dots r'_n) \propto p(r'_1 \dots r'_n|\theta_C)p(\theta_C),$$

where r' are the rewards indicating whether the expert contributed to the outcome. Specifically, if R_i is the chosen action, rewards are delivered as follows:

$$r = \begin{cases} r, & \text{if } \mu_{R_i} > \mu_{R_j}, \forall j \neq i \\ 1 - r, & \text{otherwise} \end{cases}.$$

Thus, experts were rewarded when a reward was received and that expert assigned the highest probability to the chosen response. If, on the other hand, the expert predicted one of the unselected options it would not be rewarded (i.e., $r=0$). Moreover, if the chosen action was not correct and the expert assigned the largest probability to that action, then it was not rewarded. However, it was rewarded if the outcome was not correct (i.e., it did not contribute to the incorrect action). We can assign an attentional weight to each expert that reflects its history of contributing to successful outcomes. To do so, we use another softmax function to assign weights to each expert, relative to all other experts. For the color expert, we can determine its weight with the following equation:

$$w_C(t) = \frac{e^{\frac{\mu^C(t)}{\xi}}}{\sum_E e^{\frac{\mu^E(t)}{\xi}}},$$

where w_C on trial t is the attentional weight, based on its expected reward probability μ^C relative to all other experts. Last, ξ acts as a gain parameter that discriminates between the separate experts (similar to the

κ parameter in the action selection softmax). Thus, the probability of selecting response R_i is the sum of the experts E in proportion to their weight, as follows:

$$p_{R_i}^f(t) = \sum_E w_E p_{R_i}^E(t),$$

where p^f refers to the probability of generating responses for a superordinate "flat expert" (the combination of all subordinate experts so far mentioned).

At this point, the model is incapable of detecting any hierarchical structure that may be present in the task. To afford the model this ability, we now discuss the inclusion of a set of "hierarchical experts." These experts learn about two of the stimulus dimensions contingent on the identity of another, higher-order dimension. For example, the hierarchical texture expert $h_{CS|T}$ would learn reward probabilities for color and shape separately for each texture option in T . This manner of learning is accomplished by having two subordinate experts learn the reward probability for selecting a response for color C (shape S) given texture T , as follows:

$$p(\theta_{R,C|T}|r_1 \dots r_n) \propto p(r_1 \dots r_n|\theta_{R,C|T})p(\theta_{R,C|T})$$

$$p(\theta_{R,S|T}|r_1 \dots r_n) \propto p(r_1 \dots r_n|\theta_{R,S|T})p(\theta_{R,S|T}).$$

Credit assignment works as it did with the flat experts, but now across the subordinate experts within the hierarchical expert framework. For the $h_{CS|T}$ hierarchical texture expert, attentional weights are dynamically assigned to the color or shape dependent on the texture, as follows:

$$w_{C|T}(t) = \frac{e^{\frac{\mu^{C|T}(t)}{\xi}}}{e^{\frac{\mu^{C|T}(t)}{\xi}} + e^{\frac{\mu^{S|T}(t)}{\xi}}},$$

where $w_{C|T}(t)$ is the attentional weight to the color expert relative to the shape expert when texture T is present. The probability of selecting a response, R_i , according to this hierarchical texture expert is the result of mixing the subordinate experts on each trial:

$$p_{R_i}^{h_{CS|T}}(t) = w_{C|T} p_{R_i}^{C|T}(t) + w_{S|T} p_{R_i}^{S|T}(t).$$

In addition to the texture expert, we also included a hierarchical shape and hierarchical color expert. Similar to the overall flat expert, a superordinate hierarchical expert assigned attention weights to the hierarchical experts via the following:

$$p_{R_i}^h(t) = w_{CS|T} p_{R_i}^{CS|T}(t) + w_{CT|S} p_{R_i}^{CT|S}(t) + w_{TS|C} p_{R_i}^{TS|C}(t).$$

Last, a second-level attentional selection step was included to arbitrate between the two overall experts (flat, hierarchical), as follows:

$$w_H(t) = \frac{e^{\frac{\mu^H(t)}{\xi}}}{e^{\frac{\mu^H(t)}{\xi}} + e^{\frac{\mu^C(t)}{\xi}}},$$

where ξ determines the gain of the discrimination between the hierarchical and flat expert. The ultimate response is then selected as follows:

$$p_{R_i}(t) = w_H p_{R_i}^H(t) + w_F p_{R_i}^F(t).$$

In total, the model included 11 free parameters to be estimated, with each block being fit independently. Three of these consisted of the α -parameters from each one-way flat experts' initial beta distribution (the mean of which is represented by, in the example of the flat color expert, μ^C). Another two came from the β -parameter of the beta distribution for the two-way and three-way flat experts, where in the case of the three

two-way experts, the value acted as a hyperparameter over each expert. Another three consisted of the β -parameter of the beta distribution for each hierarchical expert. The last three included the noise/gain parameters in each of the three softmax functions (i.e., action selection, attentional weight assignment, and superordinate attention to hierarchy).

To obtain the best fit for the data, we first modeled all subjects together (pseudo- $R^2 = 0.25$ and 0.12 for the mean hierarchical block and flat block, respectively) to generate appropriate initial starting parameter values to be used as our initialization point for the model when fitting each subject individually (mean pseudo- $R^2 = 0.33$ and 0.15 for the mean hierarchical block and flat block, respectively). Model fitting occurred via maximum likelihood estimation. These pseudo- R^2 values are similar to those reported in the study by Frank and Badre (2012). Validation of the revised MoE model involved simulating datasets across each of the five task blocks. We used the parameter values obtained from fitting the model to the real subject data to generate simulated responses to the task. To draw comparisons to the human data, the simulated data were then fit to the state-space model so as to produce learning curves, which allowed for calculation of learning metrics (i.e., maximum second derivative). Overall, the revised MoE model was successfully able to recreate the qualitative patterns of behavior and attentional weight recovery across blocks seen in the human data.

Univariate fMRI analysis

Statistical models were constructed for each subject under the assumptions of the general linear model using SPM 12. Each trial was modeled by one of two sets of the following five boxcar regressors: (1) a regressor for the stimulus response phase (beginning with stimulus onset and ending when a response was made); (2) a regressor with the same onset and duration as the stimulus response phase, but whose value was parametrically modulated by the subject's reaction time to the stimulus; (3) a regressor for the confidence response phase (beginning and ending with the onset and offset, respectively, of the confidence probe); (4) a regressor with the same onset and duration as the confidence response phase, but whose value was parametrically modulated by the reported confidence level; and (5) a regressor for the feedback phase (beginning and ending with the onset and offset, respectively, of the audiovisual feedback). To match the analysis approach of Badre et al. (2010), one set of regressors exclusively modeled correct trials, while the other set exclusively modeled incorrect trials. To ensure the parametrically modulated regressors only explained the variance unique to processes associated with the modulatory values (i.e., stimulus reaction time and confidence level), we orthogonalized both the modulated stimulus response phase and modulated confidence response phase regressors with respect to their respective unmodulated regressors. Next, we included three additional regressors to remove variance associated with events related to the subject failing to make a required response. Two regressors modeled stimulus and confidence response phases where no stimulus or confidence response, respectively, was made. The third regressor modeled feedback phases where "No Response" was presented. Although trials where subjects failed to indicate their level of confidence could be separated by whether the subject's stimulus response was correct or incorrect, we chose to model these events together because we considered both events to be of no interest and thus nuisance signals. Last, five block regressors were included to account for run-to-run variance. In total, each block contained a theoretical maximum of 14 regressors: some subjects had blocks where all required responses were made, and thus no regressors could be made that modeled events related to a failure to respond. Low-frequency signals were removed with a 1/128 Hz high-pass filter. This first-level regression thus yielded standardized regression coefficients ("betas") for each voxel in the brain for each regressor included in the model. Linear contrasts were used to obtain subject-specific effects, which were then entered into a second-level analysis treating subjects as a random effect and comparing voxel effects against a value of zero. Cluster correction was performed on all whole-brain, voxelwise analyses using an initial height threshold of $p < 0.001$ to then define a familywise error rate threshold of $p = 0.05$. The first voxelwise analysis of stimulus response phase activity compared with baseline (see Fig. 3) resulted in an extent threshold of 29,516 voxels. The voxelwise map

revealing the contrast of stimulus response phase activity in Hier 2 greater than the average of Hier 1 and Hier 3 resulted in an extent threshold of 106 voxels (see Fig. 4A), while the voxelwise map assessing the behavioral metric of transfer in Hier 3 and Hier 4 resulted in an extent threshold of 107 voxels (see Fig. 4B).

Region of interest (ROI) analyses supplemented the whole-brain search. ROIs were constructed with the Marsbars (Brett et al., 2002) and wfpickatlas (Maldjian et al., 2003) toolboxes in SPM12. Coordinates and sphere size for frontal cortex nodes [i.e., dorsal premotor cortex (PMd), pre-dorsal premotor cortex (pre-PMd), mid inferior frontal sulcus (Mid-IFS), and frontal polar cortex (FPC)] were taken from the study by Badre et al. (2010). Cingulo-opercular and frontoparietal (FP) coordinates and size [i.e., CO: bilateral anterior prefrontal cortex, bilateral anterior insula/frontal operculum, bilateral thalamus, and dorsal anterior cingulate cortex (ACC)/mid-superior frontal cortex; FP: bilateral intraparietal sulcus (IPS), bilateral frontal cortex (approximately BA 6), bilateral precuneus, bilateral inferior parietal lobule (IPL), bilateral dorsolateral prefrontal cortex (approximately BA 9/46), and midcingulate cortex] were taken from the study by Dosenbach et al. (2007).

Behavioral metrics of transfer

To test for brain–behavior correlations that relate individual differences in transfer performance to fMRI activity, we calculated the behavioral metric of transfer based on the state-space model we used. We computed a difference score between the fourth and the first hierarchical block so as to assess the maximum impact that hierarchical structure transfer could have on behavioral performance. Specifically, our metric of transfer came from computing the change in the state-space maximum second-derivative measure of the model. We chose to focus on the maximum second derivative as it should best capture the degree to which learning accelerates once the subject determines the appropriate first-order rules associated with the known second-order policy. Defining transfer in this manner allowed us to contrast subjects' performance when learning a hierarchically structured task with no ability to transfer knowledge of a second-order policy to when subjects have the greatest likelihood of transferring learned second-order policy.

For the whole-brain analysis, we defined a contrast for each subject that contrasted mean stimulus response phase activity for the third and fourth hierarchical blocks against baseline. At the second level, the transfer metric was used as a covariate and regressed against this contrast to identify univariate activity across individuals that was associated with differences in transfer.

Data availability

The custom Python and MATLAB code used for model fitting and data analysis is available on request.

Results

State-space model reveals discovery and transfer of global hierarchical structure

Trial outcomes from each block were fit with a state-space model (Fig. 2A; Smith et al., 2004), and the following metrics were computed from the learning curves in each block: the (1) maximal first derivative (mean \pm within-subjects SEM: Flat, 0.006 ± 0.003 ; Hier 1, 0.019 ± 0.004 ; Hier 2, 0.024 ± 0.003 ; Hier 3, 0.037 ± 0.004 ; Hier 4, 0.037 ± 0.004 ; Fig. 2B); (2) maximal second derivative (mean \pm within-subjects SEM: Flat, 0.002 ± 0.0011 ; Hier 1, 0.0052 ± 0.0012 ; Hier 2, 0.0067 ± 0.0010 ; Hier 3, 0.0125 ± 0.0019 ; Hier 4, 0.0113 ± 0.0015 ; Fig. 2B), and (3) the "learning trial" (mean \pm within-subjects SEM: Flat, 79.29 ± 8.82 ; Hier 1, 48.08 ± 5.12 ; Hier 2, 36.83 ± 5.60 ; Hier 3, 33.67 ± 4.92 ; Hier 4, 24.12 ± 5.53 ; Fig. 2B; see Materials and Methods for definitions).

We first tested whether subjects acquired second-order hierarchical rules in blocks that contained a hierarchical policy structure, which should be reflected in differences in the learning curve metrics. Compared with the flat block, learning in the first

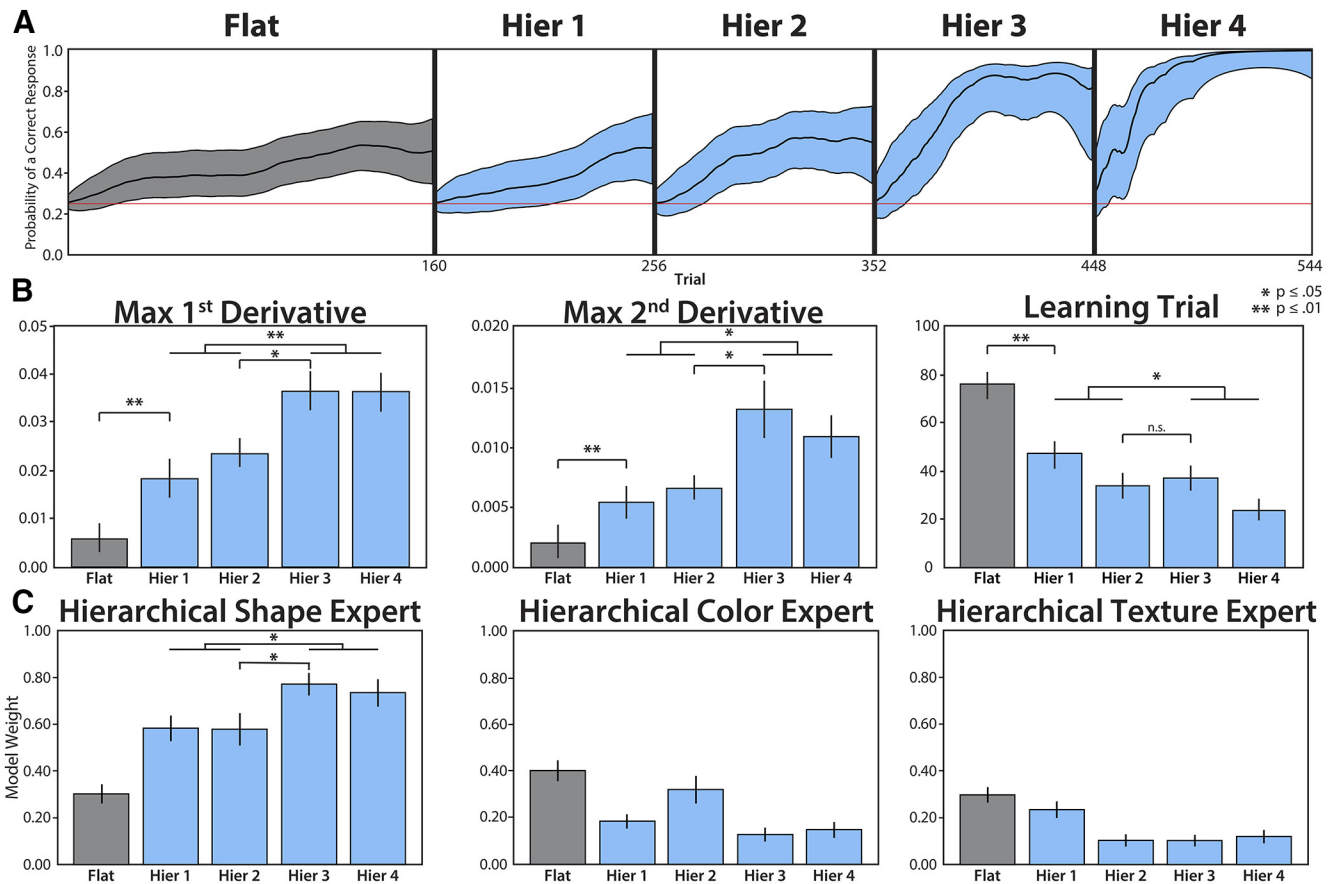


Figure 2. Learning curve and mixture of experts results reveal the discovery and transfer of the global hierarchical policy structure. **A**, Output of the state-space model (Smith et al., 2004) for a representative subject. For each trial within a block, the model computes the probability of a correct response given the trial outcomes of the block. The 90% confidence interval around the estimated probability of each trial is shown in gray (Flat block) and blue (Hierarchical blocks). The red line indicates chance-level performance. **B**, State-space model estimates for maximal first and second derivatives and learning trial, averaged across subjects. The first and second derivative metrics reveal a significant increase in learning following the second hierarchical block, while the mean learning trial improves more gradually across hierarchical blocks. **C**, Mixture of experts model weights for attention to the hierarchical shape, color, and texture experts at the beginning of the flat (gray) and four hierarchical (blue) blocks. Each expert corresponds to a latent hypothesis regarding the hierarchical task structure that a subject might hold at the beginning of each block. Following the second hierarchical block, there is a significant increase in attention for the expert that corresponds to the global second-order policy: shape cues color or texture (hierarchical shape expert). Error bars represent within-subjects SEM. Significance is assessed at $p < 0.05$. Statistical significance values of $p > 0.05$ are reported as n.s.

hierarchical block was more efficient (earlier learning trial: $t_{(23)} = 3.22$, $p = 0.004$; Fig. 2B) and showed the abrupt gains in accuracy expected from the generalization of learned second-order policy to unknown first-order rules (greater max first derivative: $t_{(23)} = 3.30$, $p = 0.003$; max second derivative: $t_{(23)} = 3.20$, $p = 0.004$; Fig. 2B). This pattern was also present when comparing learning curve metrics from the flat block to the average metrics across all hierarchical blocks (max first derivative: $t_{(23)} = 5.74$, $p < 0.001$; max second derivative: $t_{(23)} = 4.68$, $p < 0.001$; learning trial: $t_{(23)} = 3.95$, $p < 0.001$; Fig. 2B).

We next sought to investigate the role of hierarchical structure transfer. In the first hierarchical block, subjects acquired and exploited the block-specific second-order policy to facilitate learning relative to the flat block. Subsequently, the second hierarchical block provides the opportunity for subjects to discover the global second-order policy structure: after acquiring the block-specific second-order policy in the second hierarchical block, subjects can discover that the same abstract second-order policy (i.e., shape cues color or texture) has been shared across the first two hierarchical blocks. Subjects can then transfer their learned knowledge of a global second-order policy structure to subsequent blocks, which should greatly facilitate the acquisition of a block-specific second-order policy (e.g., star cues color, trapezoid cues texture) and subsequently allow the subject to more

rapidly resolve first-order rules within the known hierarchical structure. Thus, we predicted that successful structure transfer would result in markedly more efficient and abrupt learning following the second hierarchical block.

To test for behavioral evidence of hierarchical structure transfer, performance in hierarchical block 3—where subjects can implement learned structure knowledge from the start of the block—was compared with hierarchical block 2—where subjects can initially discover the global second-order policy structure (Fig. 2B). As predicted, there is a significant improvement in hierarchical learning as measured by the max first derivative ($t_{(23)} = 2.25$, $p = 0.035$) and max second derivative ($t_{(23)} = 2.23$, $p = 0.036$). However, the learning trial metric does not show the same pattern ($t_{(23)} = 0.41$, $p = 0.688$). This improvement is not easily explained by general practice effects: there is not a reliable change in performance metrics from the first to the second hierarchical block—when subjects can take advantage of task practice and general familiarity with the trial procedure—but must still discover the global second-order policy structure (as assessed by all three metrics: max $t = 1.28$, $p = 0.21$). Instead, the evidence of transfer is only observed after subjects have had the opportunity to discover the global structure in the second hierarchy block.

Following discovery of the global second-order policy structure, hierarchical knowledge transfer can facilitate learning for all

subsequent blocks. Therefore, the learning metrics averaged across hierarchical blocks 3 and 4 (when the knowledge can be implemented to support learning) were compared with the average across hierarchical blocks 1 and 2 (when the knowledge has not yet been acquired). Subjects showed evidence of improved hierarchical learning in the last two hierarchical blocks versus the first two across all behavioral metrics (max first derivative: $t_{(23)} = 2.99$, $p = 0.007$; max second derivative: $t_{(23)} = 2.76$, $p = 0.011$; learning trial: $t_{(23)} = 2.50$, $p = 0.020$).

Last, we used a method previously developed to assess hierarchical learning (Badre et al., 2010) to analyze hierarchical structure learning and transfer. Instead of modeling all trials together within a block, responses to each unique stimulus were individually analyzed to obtain separate learning trials. Moreover, in tasks with hierarchically structured second-order policy, one can conclude that a second-order rule is completely learned if all of its subordinate first-order rules are learned above chance. Then, evidence of hierarchical structure transfer can be assessed, which should allow for faster and more complete learning of second-order rules (mean \pm within-subjects SEM: Flat, 0.17 ± 0.08 ; Hier 1, 0.25 ± 0.10 ; Hier 2, 0.50 ± 0.09 ; Hier 3, 1.00 ± 0.12 ; Hier 4, 0.92 ± 0.11). Subjects learned more second-order rules in the hierarchical blocks than in the flat block ($Z = 15.5$, $p < 0.001$). Moreover, there was a significant increase in learned second-order rules from the second to the third hierarchical block ($Z = 4.5$, $p = 0.008$). Last, subjects also learned more second-order rules in the last two hierarchical blocks than in the first two ($Z = 12.0$, $p < 0.001$). Together, these results provide evidence that learning and subsequently transferring the global second-order policy structure supports more efficient hierarchical learning, over and above the expected level of hierarchical learning if the hierarchical policy must be relearned on every block.

Mixture of experts model confirms transfer of specific hierarchical structure

Although learning rate metrics derived from the state-space model allow us to characterize how learning changes across blocks, they do not provide information about why learning may have changed. We theorized that subjects discovered the specific second-order policy that was globally persistent across blocks. When learning the rules for a new block, this knowledge should encourage subjects to test the hypothesis that shape determines second-order policy. In turn, this would enhance learning by biasing their attention toward the relevance of the shape dimension, and away from the color and texture dimensions. As an alternative explanation, subjects might have discovered that the presence of hierarchical policy, in general, was persistent across blocks: one dimension cues the relevant first-order dimensions. When learning the rules for a new block, this knowledge should encourage subjects to test the hypothesis that a second-order policy exists. This knowledge could enhance learning by biasing their attention toward the relevance of second-order policies, in general, versus a flat policy. Because the state-space model cannot distinguish these two explanations, we used a hybrid Bayesian-reinforcement learning MoE model to infer the latent hypothesis states of each subject during the learning process (Frank and Badre, 2012). This approach allows us to probe the underlying cognitive mechanisms that support transfer by estimating how specific hypotheses regarding hierarchical task structure were being attended and transferred across blocks (for details, see Materials and Methods).

The MoE model was used to derive attention measures for four modeled “experts,” each associated with a specific

hypothesis. The first measure indexes the attention subjects place on the specific hypothesis that the shape dimension forms the top of the second-order policy and cues subordinate first-order rules based on either color or texture (referred to as “attention to the hierarchical shape expert”). The second and third measures index the attention placed on the specific hypotheses that the color or texture dimensions, respectively, form the top of the hierarchy. The fourth measure indexes the attention subjects place on the general hypothesis that hierarchical structure, in the form of any second-order policy, exists in the block compared with a flat policy (referred to as “attention to hierarchy”). The attention to hierarchy measure does not discern among which dimension sits atop the hierarchy, in contrast to the other three measures. To characterize what knowledge is being transferred from the previous block, we focus on the model estimates for these measures that capture the state of the subject before encountering the first trial of the block. These estimates of the subject’s latent state before the block begins are inferred by fitting the model to each individual’s trial-by-trial sequence of choices and rewards. Therefore, a discrimination can be made between whether a subject is transferring a hypothesis regarding a specific second-order policy (attention to the hierarchical shape expert), compared with a general hypothesis regarding the presence of second-order policy (attention to hierarchy), at the start of the block.

First, to determine whether subjects discover the global second-order policy that is persistent across blocks and then test the hypothesis that this policy applies to subsequent blocks, the attention to the hierarchical shape expert was analyzed across blocks (mean \pm within-subjects SEM: Flat, 0.30 ± 0.04 ; Hier 1, 0.58 ± 0.05 ; Hier 2, 0.58 ± 0.07 ; Hier 3, 0.77 ± 0.05 ; Hier 4, 0.73 ± 0.06 ; Fig. 2C). In line with our predictions, subjects’ attention to the hierarchical shape expert at the start of the block increases from the second to the third hierarchical block ($t_{(23)} = 2.08$, $p = 0.049$; Fig. 2C), after they have had the opportunity to discover the global second-order policy structure. Moreover, because this knowledge can inform the hypotheses for all subsequent blocks, attention to the hierarchical shape expert is greater at the start of hierarchical blocks 3 and 4 than at the start of the first two hierarchical blocks ($t_{(23)} = 2.64$, $p = 0.015$). Although specific statistical predictions regarding attention to the hierarchical color and texture experts were not made, attention to these experts should generally be diminished when attention is biased in favor of the hierarchical shape expert. Indeed, attention to the color and texture experts is qualitatively low in the hierarchical blocks (Fig. 2C).

Next, we analyzed whether subjects test the hypothesis that a hierarchical policy, in general, is persistent across blocks (mean \pm within-subjects SEM: Flat, 0.22 ± 0.01 ; Hier 1, 0.22 ± 0.01 ; Hier 2, 0.23 ± 0.01 ; Hier 3, 0.26 ± 0.01 ; Hier 4, 0.26 ± 0.01). Subjects’ attention to hierarchy does not increase from the second to the third hierarchical block ($t_{(23)} = 1.70$, $p = 0.103$). However, there is a more gradual change in attention to hierarchy such that the measure increases from the first two hierarchical blocks to the last two ($t_{(23)} = 2.51$, $p = 0.019$). Together, these results show that the improvement in hierarchical learning observed after the second hierarchical block can be explained by subjects discovering and then transferring their knowledge of the appropriate global second-order policy structure that is persistent across all hierarchical blocks.

Lateral frontal regions linked to discovery of global hierarchical structure

First, a whole-brain univariate contrast of activity during the stimulus response phase on correct trials across all blocks

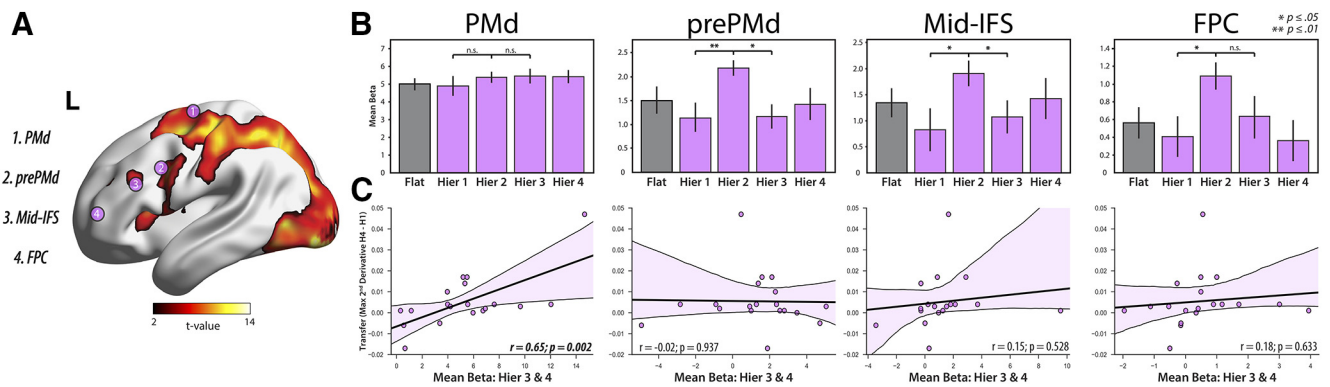


Figure 3. Lateral frontal regions linked to the discovery of global hierarchical policy structure and behavioral transfer. **A**, Group-level activity across all blocks during the stimulus response phase on correct trials only. The overlaid numbered pink circles indicate the position of each of the four lateral frontal cortex ROIs. Map is cluster corrected to a familywise error rate of $p < 0.05$. **B**, ROI analyses for the regions shown in **A**. The mean β -coefficients from the stimulus response phase show elevated activity during the second hierarchical block (vs the first and third hierarchical blocks) in all regions except PMd. Error bars indicate the within-subject SEM. **C**, Correlations between behavioral transfer and activity in the left lateral frontal cortex ROIs following discovery of the global hierarchical policy structure. Only activity in PMd is tentatively correlated with individual differences in transfer. Statistical significance values of $p > 0.05$ are reported as n.s.

compared with baseline was performed ($p = 0.05$ cluster corrected; Fig. 3A). The resultant map is consistent with those seen in previous hierarchical reinforcement learning studies (Badre et al., 2010). The task recruited regions along the lateral frontal cortex associated with hierarchical task performance (Koechlin et al., 2003; Badre and D'Esposito, 2007), as well as parietal cortex, and more specifically the intraparietal sulcus, anterior insula, mid-cingulate cortex, occipital lobe, thalamus, and medial temporal lobe.

To address which brain regions supported searching for and discovering the global second-order policy structure, we first focused on the following regions in left lateral frontal cortex that support the learning and execution of hierarchical control policies: the PMd, pre-PMd, Mid-IFS, and FPC (Badre et al., 2010; Fig. 3A). In their original work, Badre and D'Esposito (2007) discovered that PMd resolved competition between first-order rules regarding motor response options, pre-PMd resolved competition between second-order rules relating one stimulus feature to another (e.g., for squares, red cues action 1 while blue cues action 2), Mid-IFS resolved competition between third-order rules, and FPC resolved competition between fourth-order task contexts. Moreover, activity in these regions has been associated with the search for a specific hierarchical policy within a task block (Badre et al., 2010). However, it remains unknown whether these same regions also support the learning of a more abstract, global hierarchical structure that facilitates learning the specific hierarchical policies within each block.

The behavioral results demonstrate that subjects were able to learn block-specific hierarchical policies, as well as search for and discover the global hierarchical policy structure during the second hierarchical block. To identify activity in the frontal cortex that is related to discovering the global structure, over and above activity associated with learning a block-specific hierarchical policy, activity in the second hierarchical block relative to the first hierarchical block was assessed (mean \pm within-subjects SEM; PMd: Hier 1, 4.90 ± 0.56 ; Hier 2, 5.39 ± 0.31 ; pre-PMd: Hier 1, 1.14 ± 0.31 ; Hier 2, 2.17 ± 0.17 ; Mid-IFS: Hier 1, 0.94 ± 0.48 ; Hier 2, 2.16 ± 0.25 ; FPC: Hier 1, 0.40 ± 0.24 ; Hier 2, 1.09 ± 0.15 ; Fig. 3B). With the exception of PMd ($t_{(18)} = 0.72$, $p = 0.483$), activity across the lateral frontal cortex regions is greater in the second hierarchical block compared with the first (pre-PMd: $t_{(18)} = 3.48$, $p = 0.003$; Mid-IFS: $t_{(18)} = 2.10$, $p = 0.050$; FPC: $t_{(18)} = 2.19$, $p = 0.042$). Next, activity in the second hierarchical

block was compared with that in the third hierarchical block, where subjects no longer need to search for structure and can instead implement their transferred structure knowledge from the second hierarchical block (mean \pm within-subjects SEM; PMd: Hier 2, 5.39 ± 0.31 ; Hier 3, 5.45 ± 0.41 ; pre-PMd: Hier 2, 2.17 ± 0.17 ; Hier 3, 1.16 ± 0.25 ; Mid-IFS: Hier 2, 2.16 ± 0.25 ; Hier 3, 1.21 ± 0.36 ; FPC: Hier 2, 1.09 ± 0.15 ; Hier 3, 0.62 ± 0.26 ; Fig. 3B). Again, activity in pre-PMd ($t_{(18)} = 2.80$, $p = 0.012$) and Mid-IFS ($t_{(18)} = 2.12$, $p = 0.048$) is greater in the second hierarchical block. Activity is also numerically greater in FPC ($t_{(18)} = 1.52$, $p = 0.147$), but not statistically significant. Last, activity in PMd did not differ across the blocks ($t_{(18)} = 0.12$, $p = 0.904$). Because the activity in rostral regions of frontal cortex is elevated in the second hierarchical block relative to both the preceding and proceeding blocks, the observed results are likely because of a process that is preferentially engaged in the second hierarchical block, as opposed to a process that continuously evolves over time such as effects related to time on task or practice.

Lateral frontal regions linked to transfer of global hierarchical structure

Next, we determined whether activity in the lateral frontal ROIs predicts behavioral transfer, which was indexed by more abrupt hierarchical learning in the blocks that follow discovery of the global hierarchical structure (for definitions and details, see Materials and Methods). Different lateral frontal cortex regions could support transfer of the global hierarchical policy structure. For example, pre-PMd could support transfer of second-order policy by means of a more efficient resolution of competition between competing within-block second-order rules. Alternatively, if transfer is an additional third level in the policy hierarchy (i.e., the task block contextualizes second-order rules associated with the shape dimension), then Mid-IFS (e.g., the region associated with policy abstraction one level greater than that being transferred) could support transferring learned structure. Last, FPC activity could support transfer, as structure transfer may be a form of extended temporal contextualization, or episodic control, that biases task representations across multiple blocks. Knowledge of the position of the shape dimension in the hierarchy may take the role of a schema and thus recruit FPC to support the accommodation and contextualization of new information within this framework.

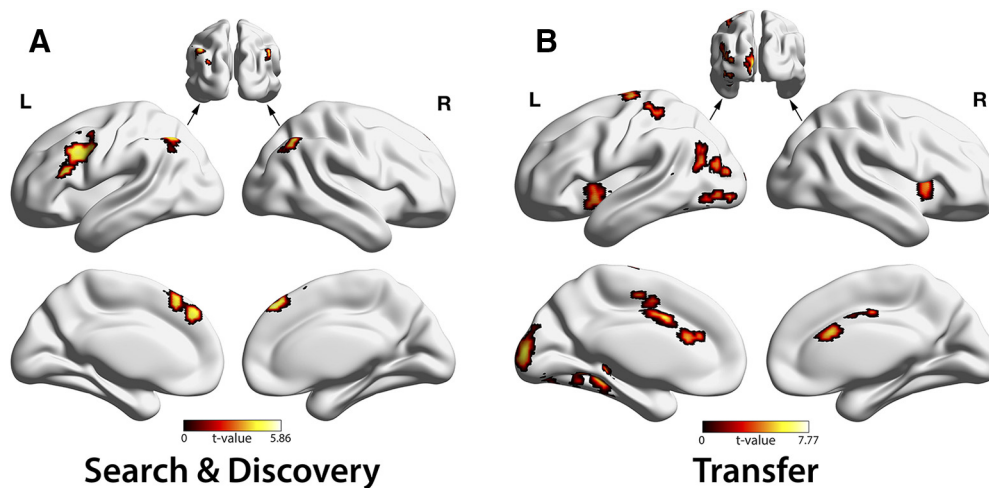


Figure 4. Voxelwise analyses reveal regions linked to unique behavioral roles. **A**, Activity during the search and discovery of the global hierarchical structure during the second hierarchical block shown by the contrast of Hier 2 > Hier 1 + Hier 3. **B**, Whole-brain analysis of regions for which stimulus response phase activity following discovery of the global hierarchical structure correlates with behavioral transfer. All activity maps are cluster corrected to a familywise error rate of $p < 0.05$.

To test these predictions, correlations between the mean activity in each lateral frontal ROI from blocks where behavioral transfer could occur (i.e., hierarchical blocks 3 and 4), and the behavioral metric of transfer for each subject were performed (Fig. 3C). Activity in pre-PMd ($r = -0.02$, $p = 0.937$), Mid-IFS ($r = 0.15$, $p = 0.528$), and FPC ($r = 0.18$, $p = 0.633$) did not reliably correlate with behavioral transfer. However, activity in the most caudal frontal region, PMd, appeared to reliably correlate with behavioral transfer ($r = 0.65$, $p = 0.002$). To test the robustness of these individual differences results, we also performed the analyses using a nonparametric rank-ordered regression test. In line with the previous results, PMd was significantly correlated with behavioral transfer (Spearman $\rho = 0.48$, $p = 0.037$), while pre-PMd, Mid-IFS, and FPC were not statistically significant (absolute value of all Spearman ρ values < 0.43 , all p values > 0.065). However, one high-leverage subject who showed substantial behavioral transfer also had the highest activity in PMd (Fig. 3C). When this subject is removed from the analysis, the positive correlation no longer reaches statistical significance [PMd: $r = 0.37$, $p = 0.12$ ($\rho = 0.39$, $p = 0.11$); all other ROI p values > 0.38 (nonparametric p values > 0.09)]. Thus, these results suggest that PMd is the most likely lateral frontal region to relate to transfer, although this relationship may be modest and awaits confirmation in future studies.

Whole-brain analyses: regions linked to discovery of global hierarchical structure

To identify regions recruited by the search and discovery of the global hierarchical policy structure, a whole-brain voxelwise analysis was performed by contrasting activity in the second hierarchical block to the average of the first and third hierarchical blocks (Fig. 4A). This contrast revealed activity that overlapped with the left pre-PMd and Mid-IFS ROIs. However, activity was also found in medial superior frontal gyrus, the left IPL and IPS, and the right IPL. The locations of these lateral frontal and parietal regions overlap with a set of regions referred to as the “FP network” that have been previously implicated in cognitive control functions (Dosenbach et al., 2007, 2008).

Whole-brain analyses: regions linked to transfer of global hierarchical structure

To further identify which regions support behavioral transfer in blocks following the discovery of the global hierarchical policy

structure, a whole-brain analysis was performed using the degree of behavioral transfer as a parametric modulator of the mean stimulus response phase activity in the third and fourth hierarchical blocks ($p = 0.05$ cluster-corrected; Fig. 4B). PMd activity (overlapping with our ROI)—in accord with the previous ROI analyses—as well as bilateral anterior insula/frontal operculum, anterior cingulate cortex, left lateral occipital cortex, and left medial temporal cortex correlated with behavioral transfer. Anterior insula and dorsal anterior cingulate cortex correspond to the “core” regions of the putative cingulo-opercular network commonly found in tasks requiring cognitive control (“CO network,” Dosenbach et al., 2007, 2008; Sadaghiani and Kleinschmidt, 2016).

Dissociation of behavioral roles for FP and CO networks

The FP and CO networks have been proposed as two components of a dual-network architecture of cognitive control (Dosenbach et al., 2008), and regions in both the FP and CO networks were active during performance of our hierarchical learning task. However, these regions may support task performance by making separable behavioral contributions. To test this hypothesis, we directly compared the relationship between activity across the networks’ respective regions and (1) discovering the global hierarchical policy structure versus (2) the transferring of hierarchical structure knowledge across blocks.

First, we assessed the relationship between activity in these networks and the search and discovery of hierarchical structure that occurs during the second hierarchical block. The canonical FP and CO networks were defined based on a previous meta-analysis of cognitive control tasks (FP: bilateral frontal cortex, bilateral dorsolateral prefrontal cortex, bilateral intraparietal sulcus, bilateral inferior parietal lobule, bilateral precuneus, and midcingulate cortex; CO: bilateral anterior insula/frontal operculum, bilateral anterior prefrontal cortex, bilateral thalamus, and dorsal anterior cingulate cortex/mid-superior frontal cortex; Fig. 5A; coordinates are from Dosenbach et al., 2007). Separately for the FP network and CO network ROIs, the activity during each block was estimated and a contrast was performed for the activity in the second hierarchical block versus the first and third hierarchical blocks (mean \pm within-subjects SEM; FP: Hier 1, 0.74 ± 0.23 ; Hier 2, 1.57 ± 0.16 ; Hier 3, 1.01 ± 0.25 ; CO: Hier 1,

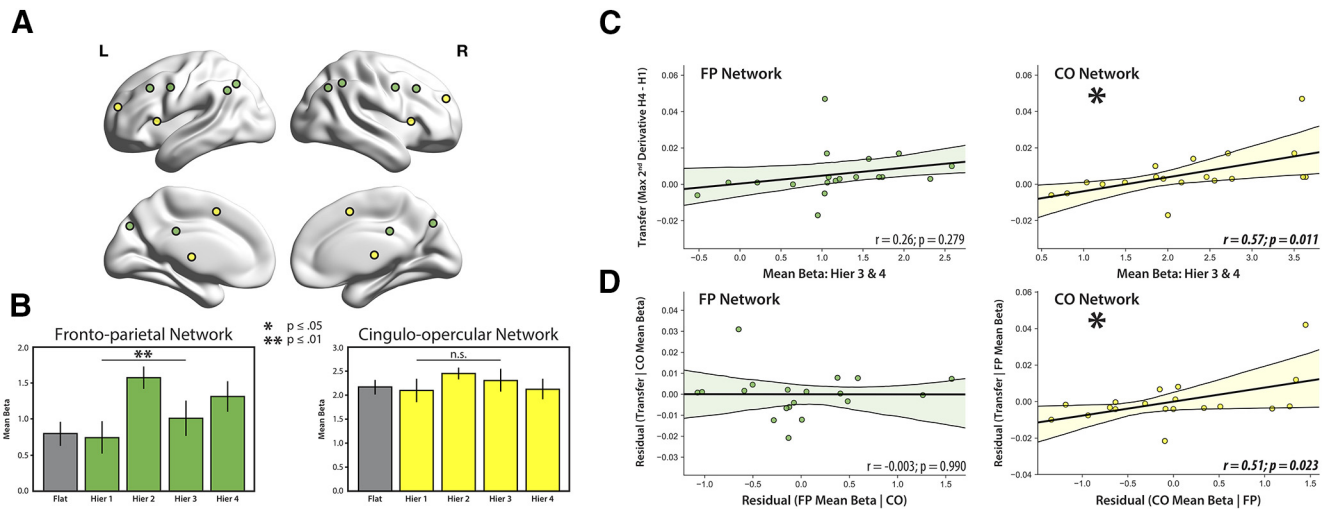


Figure 5. fMRI analyses reveal dissociation of behavioral roles for FP and CO networks. **A**, Locations of regions that define the FP (green) and CO (yellow) networks are from the study by Dosenbach et al. (2007). FP: bilateral frontal cortex, dorsolateral prefrontal cortex, IPL, IPS, precuneus, and midcingulate. CO: bilateral anterior insula/frontal operculum, anterior prefrontal cortex, thalamus, and dorsal ACC/mid-superior frontal cortex. **B**, The contrast of the mean β -coefficients from the stimulus response phase across all respective regions in the second hierarchical block compared with the first and third blocks reveals an increase in activity during the search and discovery phase only in the FP regions. Moreover, there is a significant interaction such that the difference in activity between these blocks is greater in the FP network than in the CO network. Error bars indicate within-subject SEM. **C**, The correlation of transfer with the mean β -coefficient of each network from the third and fourth hierarchical blocks. **D**, Regression analyses for the FP and CO networks against behavioral transfer reveal a unique role of the CO network in structure transfer. Shown are the partial correlation coefficients from a multiple regression that accounts for the effects of both networks. Statistical significance values of $p > 0.05$ are reported as n.s.

2.09 ± 0.25 ; Hier 2, 2.45 ± 0.12 ; Hier 3, 2.31 ± 0.24 ; Fig. 5B). FP activity was significantly increased during the second hierarchical block ($t_{(18)} = 3.49$, $p = 0.003$), as expected based on the whole-brain results, whereas CO activity was not significantly different ($t_{(18)} = 1.46$, $p = 0.162$). Since the FP network was chosen for further analysis based on the observation of left lateral frontal and bilateral parietal activity in our previous whole-brain contrast, any ROI analyses that include these regions may be biased by circularity (Vul et al., 2009). To address this possibility, a separate analysis was performed that included only the FP network ROIs that were not observed in the original whole-brain results (i.e., right frontal cortex, right dorsolateral frontal cortex, right intraparietal sulcus, bilateral precuneus, and midcingulate cortex), which also found a significant result for the contrast ($t_{(18)} = 2.80$, $p = 0.012$). Next, to formally dissociate the patterns observed across the FP and CO networks (Henson, 2006), we tested the interaction between block (second hierarchical block; average of first and third hierarchical blocks) and region (FP; CO) and found that the difference in activity between the second hierarchical block compared with the first and third blocks is significantly greater in the FP regions than in the CO regions ($t = 3.37$, $p = 0.003$).

We next assessed the relationship between activity in these networks and the transfer of hierarchical structure knowledge. As before, we sought to confirm the relationship between behavioral transfer and the canonically defined CO network, while additionally ruling out the potential for circularity in our analyses. Our first analysis confirmed a significant relationship between activity averaged across all CO regions and behavioral transfer ($r = 0.57$, $p = 0.011$, Spearman $\rho = 0.67$, $p = 0.002$; Fig. 5C). Moreover, to control for circularity in this analysis, we ran a separate test of the relationship between the CO network and behavioral transfer by excluding the insular and anterior cingulate ROIs that were present in the original whole-brain regression. This new analysis, which only included activity from bilateral thalamus and bilateral anterior prefrontal cortex (referred to as

the “periphery” of the CO network; Dosenbach et al., 2008), found a significant correlation between mean ROI activity and the behavioral transfer metric ($r = 0.66$, $p = 0.002$, Spearman $\rho = 0.77$, $p < 0.001$). In contrast to the robust correlation between the CO network and behavioral transfer, activity in the FP network in the last two hierarchical blocks is only modestly correlated with the behavioral transfer metric ($r = 0.26$, $p = 0.279$, Spearman $\rho = 0.61$, $p = 0.006$; Fig. 5C).

To test whether the CO network is uniquely related to transfer, both the CO network and FP network activity were included in a multiple regression with behavioral transfer as the dependent variable, as this approach controls for any shared contribution made by both networks. This analysis revealed that activity in the CO network selectively predicts transfer (CO network: $r = 0.52$, $p = 0.023$; Spearman $\rho = 0.47$, $p = 0.045$; FP network: $r = -0.003$, $p = 0.990$; Spearman $\rho = 0.04$, $p = 0.881$; Fig. 5D). Collectively, these findings demonstrate a clear dissociation: the regions of the FP network are specifically involved in the search and discovery of hierarchical structure, whereas the regions of the CO network are selectively involved in the transfer of hierarchical structure knowledge across blocks.

Discussion

Subjects were able to efficiently discover and exploit abstract structure during a hierarchical reinforcement learning task. During the task, subjects rapidly discovered and generalized an embedded global task structure to subsequent novel task blocks. Moreover, this generalization was supported by an increase in subjects’ awareness of the specific global hierarchical structure at the start of a new block. The fMRI data revealed that multiple left lateral frontal regions were involved during task performance (pre-PMd, Mid-IFS, and FPC). In addition, regions within a frontoparietal network were involved in the initial discovery of the global hierarchical structure, while regions within a cingulo-opercular network, and potentially PMd, were involved in the transfer of this structure.

Previous work on structure learning in the context of hierarchical reinforcement learning (Collins and Frank, 2013, 2016; Collins et al., 2014) has shown that subjects tend to build generalizable structures that allow for components of the stimulus (e.g., shape) to act as a higher-order context that cues rules based on other stimulus features (e.g., color). However, in contrast to previous work where stimulus–response groupings could be directly transferred, our task design prevented direct block-to-block transfer of action mappings. Instead of discovering structure that immediately informed action, such as learning one of the block-specific hierarchical policies, subjects discovered structure that informed subordinate task-set policies, as evidenced by more rapid learning in hierarchical blocks following discovery. Moreover, when a MoE model was used to derive an estimate of subjects' attention to the hierarchical shape rule at the start of the third hierarchical block, the model-derived estimate was greater than at the start of the second hierarchical block, indicating that subjects transferred and immediately applied their structural knowledge following discovery in the second hierarchical block. This demonstrates that subjects are capable of learning a higher-order representation between stimulus dimensions that can abstract away from the groupings of specific response pairings, and can then transfer this knowledge to new contexts.

Our brain imaging findings have implications regarding the functional organization of the frontal cortex in support of hierarchical learning. The lateral frontal cortex is recruited for both the learning and execution of hierarchical rules (Koechlin et al., 2003; Badre and D'Esposito, 2007; Badre et al., 2010; Collins et al., 2014; Nee and D'Esposito, 2016; Badre and Nee, 2018), with recruitment of more rostral regions during processing of higher levels of policy abstraction. In addition, patients with lateral frontal cortex lesions exhibit the following asymmetric behavioral impairments: caudal lesions impair both concrete and abstract cognitive control task performance, while rostral lesions only impair abstract task performance (Badre et al., 2009). In tasks where hierarchical rules had to be implicitly learned, different lateral frontal regions are simultaneously involved in the search for hierarchical policy within a block (Badre et al., 2010). However, patients with pre-PMd lesions are impaired at learning the full second-order policy, but not the subordinate first-order rules (Kayser and D'Esposito, 2013). This asymmetric functional deficit is evidence of the hierarchical organization of functions associated with these regions. Our study extends these findings by demonstrating that frontal cortex is involved in the search for a global hierarchical structure, beyond that of the block-specific second-order policies, when evidence of its presence is first available. We conclude that the same hierarchical frontal cortex organization used to execute policy rules, as well as search for hierarchical relationships of varying complexity within the moment (i.e., block-specific policies), is also involved in the search for hierarchical relationships across contexts.

There existed a potential relationship between activity in the most caudal region (PMd) and the measure of transfer and implementation of global hierarchical structure, defined as the change in the maximum second-derivative across blocks. The maximum second-derivative captures the initial rise of the learning curve, indicating the transition from searching for higher-order rules to the resolution of first-order rules. Subjects are transitioning from a phase of the task where the search space of

possible structures is large to one where it has become well defined and narrow. With conflict of the second-order policy resolved, all that remains is the resolution of first-order rules, a process linked to PMd function. It is likely that subjects who resolve the second-order conflict more rapidly can then rely primarily on processes associated with PMd (i.e., linking specific colors and textures to motor responses) for the remainder of the block, therefore facilitating performance.

Together with previous work, the current findings suggest a sophisticated coordination among motor control, rule implementation, rule discovery, and rule generalization in the service of hierarchical control, where each function incorporates knowledge of both the immediate setting (i.e., task block) and overall environment (i.e., global hierarchical structure). In simple tasks lacking contextual elements, caudal premotor regions likely resolve response competition without influence from superordinate rostral frontal regions. However, in tasks for which contextual information must be considered (e.g., abstracted hierarchical policy), rostral premotor and mid-dorsolateral regions are likely recruited to exert control over sensory–motor conflict in more caudal premotor regions (Badre et al., 2009; Kayser and D'Esposito, 2013). In settings where actions and rules are being learned, these contextual influences are likely being tested and updated via cortical–striatal interactions in response to task-based feedback signals (Badre and Frank, 2012; Frank and Badre, 2012). Thus, when a subject discovers and transfers global structure, knowledge of this structure works to restrict the search space of potential hypotheses, resulting in selective recruitment along the rostrocaudal gradient to those involved in representing the generalized known structure. Thus, multiple regions can be involved in the process of structure transfer, but specifically only those regions along the gradient necessary for the resolution of the remaining unresolved block-specific rules.

Several cortical and subcortical regions outside the lateral frontal cortex associated with behavioral transfer were identified. Subjects with greater levels of activity in regions comprising the CO network learned the block-specific hierarchical policies faster following discovery of the global hierarchical structure. Critically, this association was not found in regions comprising the FP network, suggesting that CO network activity is specifically related to the manner in which subjects maintain and implement the learned structure. Alternatively, it is possible that CO activity is increased in subjects who are more engaged and attentive to the task (Sadaghiani and Kleinschmidt, 2016). We favor the former interpretation because our transfer metric indexes a difference between performance in the first, compared with the final, hierarchical block, and is thus insensitive to differences between subjects who perform poorly in both phases (when it could be assumed that subjects are failing to pay attention to the current task), and those who perform exceedingly well in both phases (when it is likely that attentional engagement is greatest).

Previous work has implicated the CO network in both “task-set maintenance” (Dosenbach et al., 2006, 2007, 2008), broadly defined as the configuration of control signals required to perform any type of task, and “tonic alertness” (Sadaghiani et al., 2010; Sadaghiani and D'Esposito, 2015), or the user-driven sustained control necessary to remain prepared to process incoming information. Task-set maintenance requires that a specific structure be known to the individual—that which defines successful

performance of the task—whereas tonic alertness precludes any need for a specific structural representation of the task as alertness takes the role of “nonselective disengagement” (Sadaghiani and Kleinschmidt, 2016). Thus, our findings are more consistent with a role of the CO network in task-set maintenance, although a role in tonic alertness during our task cannot be ruled out.

Whereas the CO network was uniquely related to transfer, the FP network was selectively involved in the search and discovery of the global hierarchical structure. Our findings suggest that the FP network is not only involved in the representation and integration of current task rules and response mappings, but also in the integration of previous task-relevant components. The integration of this information would likely allow for complex structured relationships to be discovered across blocks. Although the component processes of searching for and discovering abstract hierarchical structure overlap with behaviors associated with learning and navigating the explore-exploit dilemma—classically linked to regions along ACC—it is unlikely that ACC would be uniquely linked to search and discovery as additional roles associated with ACC likely occurred during the preceding and proceeding phases of the task (e.g., exploring and evaluating individual hierarchical policies in the first hierarchical block, representing exploitative behaviors in the third hierarchical block; Walton et al., 2003; Quilodran et al., 2008; Stoll et al., 2016). Recent studies have discovered that tasks requiring varying levels of cognitive control recruit regions along a caudal–rostral gradient in parietal cortex in a fashion similar to that found in lateral frontal cortex (Choi et al., 2018). Moreover, regions along both gradients showed mirroring patterns of functional connectivity with striatal sites, which is in line with previous work (Badre and Frank, 2012; Collins and Frank, 2013). Accordingly, the present results implicate a system of parallel and distributed hierarchical gradients across frontal and parietal cortex that supports the search and discovery of structure of varying complexity within and across task blocks.

References

- Abdelmounaime S, Dong-Chen H (2013) New Brodatz-based image databases for grayscale color and multiband texture analysis. *ISRN Mach Vis* 2013:1–14.
- Avants BB, Epstein CL, Grossman M, Gee JC (2008) Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med Image Anal* 12:26–41.
- Badre D, D’Esposito M (2007) Functional magnetic resonance imaging evidence for a hierarchical organization of the prefrontal cortex. *J Cogn Neurosci* 19:2082–2099.
- Badre D, Frank MJ (2012) Mechanisms of hierarchical reinforcement learning in cortico-striatal circuits 2: evidence from fMRI. *Cereb Cortex* 22:527–536.
- Badre D, Hoffman J, Cooney JW, D’Esposito M (2009) Hierarchical cognitive control deficits following damage to the human frontal lobe. *Nat Neurosci* 12:515–522.
- Badre D, Kayser AS, D’Esposito M (2010) Frontal cortex and the discovery of abstract action rules. *Neuron* 66:315–326.
- Badre D, Nee DE (2018) Frontal cortex and the hierarchical control of behavior. *Trends Cogn Sci* 22:170–188.
- Bavelier D, Green CS, Pouget A, Schrater P (2012) Brain plasticity through the life span: learning to learn and action video games. *Annu Rev Neurosci* 35:391–416.
- Behrens TEJ, Woolrich MW, Walton ME, Rushworth MFS (2007) Learning the value of information in an uncertain world. *Nat Neurosci* 10:1214–1221.
- Botvinick M, Ritter S, Wang JX, Kurth-nelson Z, Blundell C, Hassabis D (2019) Reinforcement learning, fast and slow. *Trends Cogn Sci* 23:408–422.
- Brett M, Anton J-L, Valabregue R, Poline J-B (2002) Region of interest analysis using an SPM toolbox. Paper presented at 8th International Conference on Functional Mapping of the Human Brain, Sendai, Japan, June.
- Choi EY, Drayna GK, Badre D (2018) Evidence for a functional hierarchy of association networks. *J Cogn Neurosci* 30:722–736.
- Collins AGE, Frank MJ (2013) Cognitive control over learning: creating, clustering, and generalizing task-set structure. *Psychol Rev* 120:190–229.
- Collins AGE, Frank MJ (2016) Neural signature of hierarchically structured expectations predicts clustering and transfer of rule sets in reinforcement learning. *Cognition* 152:160–169.
- Collins AGE, Cavanagh JF, Frank MJ (2014) Human EEG uncovers latent generalizable rule structure during learning. *J Neurosci* 34:4677–4685.
- Dale AM (1999) Optimal experimental design for event-related fMRI. *Hum Brain Mapp* 8:109–114.
- Dosenbach NUF, Visscher KM, Palmer ED, Miezin FM, Wenger KK, Kang HC, Burgund ED, Grimes AL, Schlaggar BL, Petersen SE (2006) A core system for the implementation of task sets. *Neuron* 50:799–812.
- Dosenbach NUF, Fair DA, Miezin FM, Cohen AL, Wenger KK, Dosenbach RAT, Fox MD, Snyder AZ, Vincent JL, Raichle ME, Schlaggar BL, Petersen SE (2007) Distinct brain networks for adaptive and stable task control in humans. *Proc Natl Acad Sci U S A* 104:11073–11078.
- Dosenbach NUF, Fair DA, Cohen AL, Schlaggar BL, Petersen SE (2008) A dual-networks architecture of top-down control. *Trends Cogn Sci* 12:99–105.
- Esteban O, Blair R, Markiewicz CJ, Berleant SL, Moodie C, Ma F, Isik AI, Erramuzpe A, Kent JD, Goncalves M, DuPre E, Sitek KR, Poldrack RA, Gorgolewski KJ (2018) poldracklab/fmriprep: 1.0.10. Meyrin, Switzerland: Zenodo, CERN.
- Frank MJ, Badre D (2012) Mechanisms of hierarchical reinforcement learning in corticostriatal circuits 1: computational analysis. *Cereb Cortex* 22:509–526.
- Gorgolewski K, Burns CD, Madison C, Clark D, Halchenko YO, Waskom ML, Ghosh SS (2011) Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in Python. *Front Neuroinform* 5:13.
- Greve DN, Fischl B (2009) Accurate and robust brain image alignment using boundary-based registration. *Neuroimage* 48:63–72.
- Harlow HF (1949) The formation of learning sets. *Psychol Rev* 56:51–65.
- Henson R (2006) Forward inference using functional neuroimaging: dissociations versus associations. *Trends Cogn Sci* 10:64–69.
- Jenkinson M, Bannister P, Brady M, Smith S (2002) Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* 17:825–841.
- Kayser AS, D’Esposito M (2013) Abstract rule learning: the differential effects of lesions in frontal cortex. *Cereb Cortex* 23:230–240.
- Kemp C, Goodman ND, Tenenbaum JB (2010) Learning to learn causal models. *Cogn Sci* 34:1185–1243.
- Koechlin E, Ody C, Kouneiher F (2003) The architecture of cognitive control in the human prefrontal cortex. *Science* 302:1181–1185.
- Maldjian JA, Laurienti PJ, Kraft RA, Burdette JH (2003) An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *NeuroImage* 19:1233–1239.
- Nee DE, D’Esposito M (2016) The hierarchical organization of the lateral prefrontal cortex. *Elife* 5:e12112.
- Peirce JW (2007) PsychoPy-psychophysics software in Python. *J Neurosci Methods* 162:8–13.
- Peirce JW (2008) Generating stimuli for neuroscience using PsychoPy. *Front Neuroinform* 2:10.
- Quilodran R, Rothé M, Procyk E (2008) Behavioral shifts and action valuation in the anterior cingulate cortex. *Neuron* 57:314–325.
- Sadaghiani S, D’Esposito M (2015) Functional characterization of the cingulo-opercular network in the maintenance of tonic alertness. *Cereb Cortex* 25:2763–2773.

- Sadaghiani S, Kleinschmidt A (2016) Brain networks and α -oscillations: structural and functional foundations of cognitive control. *Trends Cogn Sci* 20:805–817.
- Sadaghiani S, Scheeringa R, Lehongre K, Morillon B, Giraud A-L, Kleinschmidt A (2010) Intrinsic connectivity networks, alpha oscillations, and tonic alertness: a simultaneous electroencephalography/functional magnetic resonance imaging study. *J Neurosci* 30:10243–10250.
- Sakai K (2008) Task set and prefrontal cortex. *Annu Rev Neurosci* 31:219–245.
- Smith AC, Frank LM, Wirth S, Yanike M, Hu D, Kubota Y, Graybiel AM, Suzuki WA, Brown EN (2004) Dynamic analysis of learning in behavioral experiments. *J Neurosci* 24:447–461.
- Stoll FM, Fontanier V, Procyk E (2016) Specific frontal neural dynamics contribute to decisions to check. *Nat Commun* 7:11990.
- Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, Gee JC (2010) N4ITK: improved N3 bias correction. *IEEE Trans Med Imaging* 29:1310–1320.
- Vul E, Harris C, Winkielman P, Pashler H (2009) Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspect Psychol Sci* 4:274–290.
- Walton ME, Bannerman DM, Alterescu K, Rushworth MFS (2003) Functional specialization within medial frontal cortex of the anterior cingulate for evaluating effort-related decisions. *J Neurosci* 23:6475–6479.
- Wang JX, Kurth-Nelson Z, Kumaran D, Tirumala D, Soyer H, Leibo JZ, Hassabis D, Botvinick M (2018) Prefrontal cortex as a meta-reinforcement learning system. *Nat Neurosci* 21:860–868.
- Woodworth RS, Thorndike EL (1901) The influence of improvement in one mental function upon the efficiency of other functions. (I). *Psychological Review* 8:247–261.
- Zhang Y, Brady M, Smith S (2001) Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans Med Imaging* 20:45–57.