

# Human EEG Uncovers Latent Generalizable Rule Structure during Learning

Anne G. E. Collins,<sup>1</sup> James F. Cavanagh,<sup>1,2</sup> and Michael J. Frank<sup>1</sup>

<sup>1</sup>Cognitive, Linguistic, and Psychological Sciences, Brown University, Providence, Rhode Island 02912, and <sup>2</sup>Department of Psychology, University of New Mexico, Albuquerque, New Mexico 87131

Human cognition is flexible and adaptive, affording the ability to detect and leverage complex structure inherent in the environment and generalize this structure to novel situations. Behavioral studies show that humans impute structure into simple learning problems, even when this tendency affords no behavioral advantage. Here we used electroencephalography to investigate the neural dynamics indicative of such incidental latent structure. Event-related potentials over lateral prefrontal cortex, typically observed for instructed task rules, were stratified according to individual participants' constructed rule sets. Moreover, this individualized latent rule structure could be independently decoded from multielectrode pattern classification. Both neural markers were predictive of participants' ability to subsequently generalize rule structure to new contexts. These EEG dynamics reveal that the human brain spontaneously constructs hierarchically structured representations during learning of simple task rules.

**Key words:** prefrontal cortex; task-set; reinforcement learning; rules; EEG

## Introduction

A hallmark of human cognitive flexibility is our ability to derive abstract rules from experience. We can even generalize such rules from one context to another yet also allow for some rules to be context-dependent (Miller, 2000). A wealth of evidence supports the notion that the prefrontal cortex is particularly important for linking environmental contexts to latent rule representations, which can then be used to guide behavior (i.e., which actions to select given the rule and sensory features) (Miller, 2000; Frank et al., 2001; Bunge et al., 2005; Badre and Frank, 2012). Such a model is implicitly hierarchical: it assumes that some features of the environment act as higher-order contexts indicative of the valid rule or "task-set" (TS), whereas others act as lower-order stimuli indicative of the valid action given the rule (Koechlin et al., 2003; Badre and D'Esposito, 2007).

Most studies of rule-based behavior assume that the cues indicative of the context, and hence the valid TS rule, are explicit. However in many real-world circumstances, individuals have to simultaneously learn which features are indicative of the rule, which features are not, and what action to take for each sensory event. For example, one may learn to play various arpeggios on the piano as well as on the violin. Although the repertoire of piano skills will be

directly generalizable when switching to the harpsichord (positive transfer), neither of the learned skills will apply when learning to play the banjo. Indeed, zealous reapplication of learned violin rules might even induce some negative transfer: biasing one to play incorrect arpeggios until a specific set of rules can be separately constructed and contextualized for the banjo, thus preventing interference. Which neural mechanisms allow us to create such abstract TS structure in a form amenable to future generalization?

We hypothesized that, when learning new associations between stimuli and actions, humans would spontaneously adopt a hierarchically structured organization that would permit subsequent generalization of task rules (Collins and Frank, 2013). Previous behavioral and computational modeling studies have provided support for this conjecture (Collins and Frank, 2013). Given that TS representations are not directly observable, these studies have relied on indirect behavioral measures, including switch costs, error patterns, and positive and negative transfer of rules to new contexts. In this report, we use EEG to directly measure prefrontal cortical activities while participants learned both context-TS structure and stimulus-action mappings. We show that latent hierarchical TS structure can be decoded from brain activity, that these signals preferentially respond to the dimension of the valid TS, and that they are predictive of the ability to capitalize on this structure when generalizing rules to new contexts.

## Materials and Methods

**Experimental protocol.** The experiment consisted of three separate blocks, each comprising a learning phase and transfer phase (labeled as such only for exposition; to participants, these proceeded seamlessly). Both phases required learning the correct action (of four possible actions) in response to four 2D visual input patterns (colored shapes) from reinforcement feedback (Fig. 1). The learning phase was designed such that there would be no overt advantage to representing structure in the learning problem. The transfer phase was designed such that any structure built during

Received Sept. 11, 2013; revised Feb. 8, 2014; accepted Feb. 26, 2014.

Author contributions: A.G.E.C. and M.J.F. designed research; A.G.E.C. and J.F.C. performed research; A.G.E.C. analyzed data; A.G.E.C., J.F.C., and M.J.F. wrote the paper.

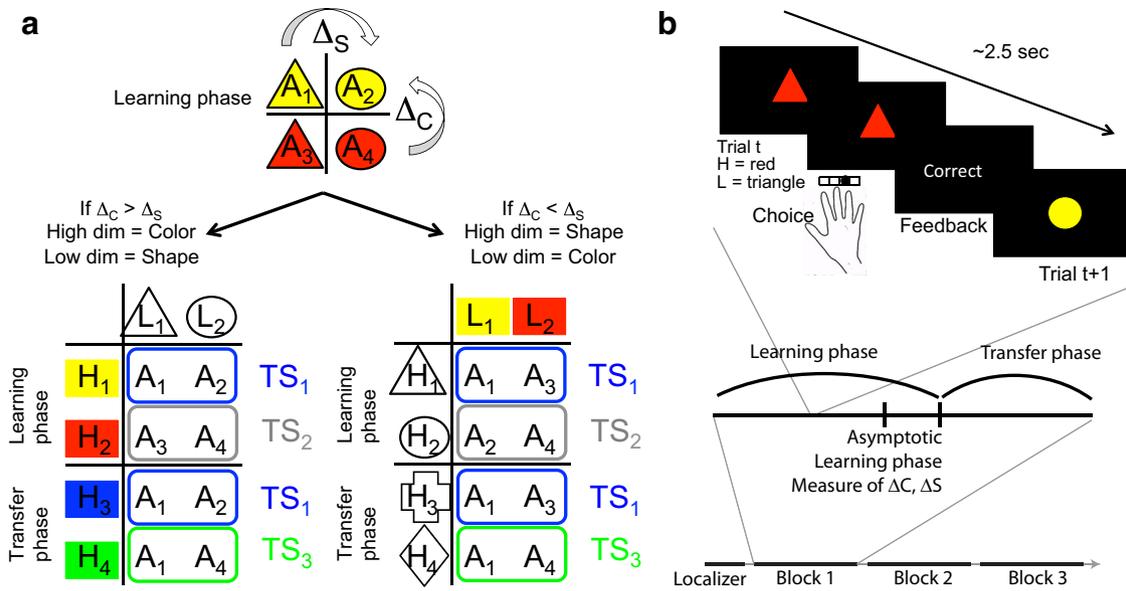
This work was supported by National Institutes of Health Grants 5T32MH019118-21 and R01 MH080066-01, and National Science Foundation Grant 1125788. We thank Jerome Sanes for use of the EGI system, Sean Masters for running the participants, and Catherine Hegarty for help in data cleaning.

The authors declare no competing financial interests.

Correspondence should be addressed to Dr. Michael J. Frank, Cognitive, Linguistic, and Psychological Sciences, Brown University, 190 Thayer Street, Providence, RI 02912. E-mail: Michael\_frank@brown.edu.

DOI:10.1523/JNEUROSCI.3900-13.2014

Copyright © 2014 the authors 0270-6474/14/344677-09\$15.00/0



**Figure 1.** Experimental protocol. **a**, Task design. In the learning phase, subjects learned to pick one of 4 actions for 4 different colored shapes. During the asymptotic learning phase, RT switch costs (switch minus stay RTs) were computed online for the color or shape dimension ( $\Delta_C$  and  $\Delta_S$ , respectively). If  $\Delta_C > \Delta_S$ , color was treated as context (high dimension [H]), and shape was stimulus (low dimension [L]); otherwise, their roles were reversed. In the transfer phase, subjects learned four new inputs from the previous stimuli (shapes or color), but with new contexts (rules) H3 and H4. H3 tested positive transfer of an old TS ( $TS_1$ ), whereas H4 tested negative transfer for learning a new TS ( $TS_3$ ). **b**, Temporal structure of the task. Subjects saw a colored shape and were required to press one key within 1.5 s. Responses were followed by deterministic audiovisual feedback. Three identically structured blocks of the task were completed. Each block comprised a learning phase followed (without interruption) by the transfer phase. Asymptotic learning phases were defined as the last 40 trials of the learning phase, occurring after subjects reached a learning criterion.

learning would facilitate positive transfer for one new context (H3) but negative transfer for another (H4).

**Rule structure.** More precisely, the pattern of input–action associations to be learned was chosen to test the incidental structure hypothesis. In principle, input–action associations could be efficiently learned in a flat conjunctive way, with no privileged status for color or shape and thus no separable context versus stimulus (high vs low) dimensions. An example of a conjunctive rule would be as follows: respond action 1 for yellow triangles, action 2 for yellow circles, action 3 for red triangles, and action 4 for red circles. Alternatively, if subjects incidentally treated one of the dimensions as higher order context, they could learn a rule such as follows: “if the color is yellow, then the following rule applies: respond action 1 for triangles, and action 2 for circles. If the color is red, then a different rule is applicable, namely, triangles and squares should be responded with actions 3 and 4.” We reasoned that, if subjects spontaneously adopt such an inherently hierarchically structured TS representation, they should exhibit greater TS switch costs when the higher dimension (in this case, color) changed from one trial to the next, compared with the RT cost associated with switches of the other dimension (Monsell, 2003; Hyafil et al., 2009; Collins and Frank, 2013). A variety of converging behavioral findings suggested that most subjects adopted such structure, with the higher-order dimension signified as that which produces a greater switch cost (Collins and Frank, 2013).

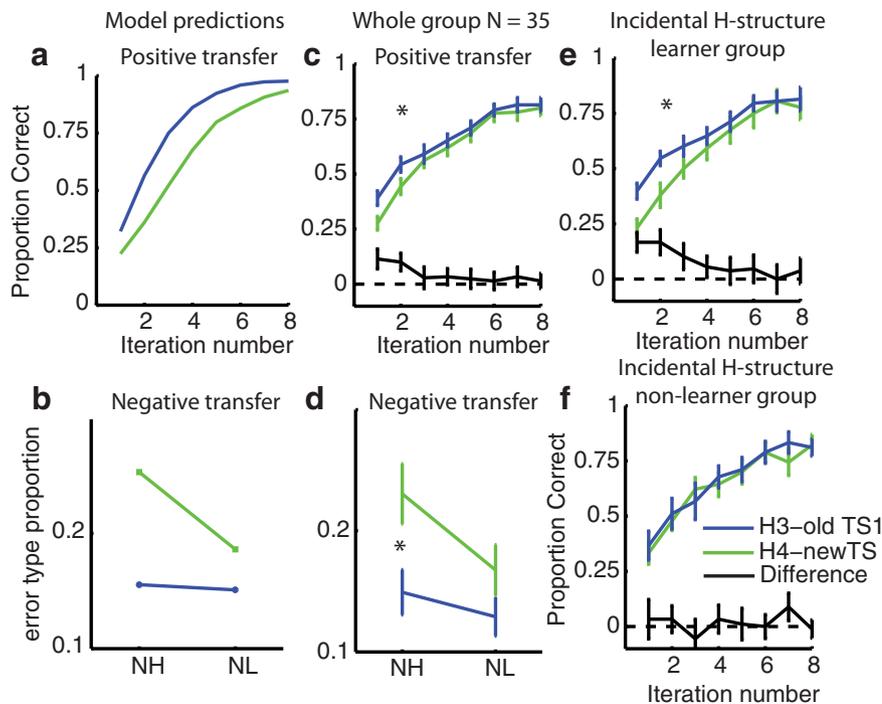
**Structure identification.** We thus relied on this same asymmetric switch cost to infer the identity of the higher and lower dimensions during the asymptotic performance period of the learning phase. The transfer phase then provided an opportunity for participants to demonstrate positive transfer in the H3 condition (which was designed to indicate that a previously learned TS was valid), and negative transfer in the H4 condition (which was designed such that the valid TS would overlap with a previously learned one, but because this overlap is not perfect, adopting the previous TS would lead to suboptimal performance). All participants were given the opportunity to generalize their (inferred) TS structure: for those inferred to assign color as the high dimension, the transfer phase presented two new colors with old shapes, whereas for those assigning shape as high dimension, the transfer phase presented two new shapes with old colors. This procedure allowed us to test whether subjects would generalize their knowledge to new contexts regardless of which one they chose to be “higher” dimension during learning. For simplicity, we here-

after refer to the context dimension as the “High” dimension (H) and stimulus dimension as “Low” dimension (L).

**Temporal structure within each trial.** After input presentation, subjects had to respond before stimulus offset at 1.5 s by selecting one of four keys with index or middle fingers of either hand. Deterministic audiovisual feedback was provided indicating whether the choice was correct (ascending tone, increment to a cumulative bar) or incorrect (descending tone, decrement to a cumulative bar) 0–300 ms after stimulus offset. If subjects did not respond in time, no feedback was provided. Subjects were encouraged not to miss trials, and to respond as fast and accurately as possible. After feedback, a fixation point appeared until next trial for 700–800 ms.

**Sequence of stimuli within each block.** The learning phase comprised a minimum of 10 and a maximum of 30 trials for each input (for a total of 40–120 trials), or up to a criterion of at least 8 of the last 10 trials correct for each input. An asymptotic performance period in which we assessed switch costs (resulting from changes in color or shape from one trial to the next) ensued at the end of this learning phase, comprised of 10 additional trials per input (40 trials total). Sequence order was pseudo-randomized to ensure identical number of trials in which color (or shape) remained identical, or changed across successive inputs. We used this period to assess online for each subject and block which dimension (color or shape) was more likely to serve as a higher-order context (vs lower-order stimulus) if subjects had effectively built structure.

We controlled the task sequence in the transfer phase such that the first correct response of the two new contexts associated with stimulus S1 was defined as context H3. This allowed us to test transfer without regard for a possible higher-level strategy participants could apply. Specifically, some subjects may assume a one-to-one mapping between the 4 possible actions and the 4 different inputs during each experimental phase. This strategy can cause subjects to be less likely to repeat action A1, even though it actually applies to both contexts H3 and H4, which would reduce the likelihood of observing transfer if they happened to respond correctly to C4 first. Of course, if analyzed as such, there would be a bias favoring transfer because H3-S1 performance is by definition better early during learning than H4-S1. To avoid this bias, we limit all assessment of transfer to the S2 stimuli. Even with this restriction, we still obtained a significantly better performance for H3 than H4 ( $p = 0.037, t = 2.16$ ). TS



**Figure 2.** Behavioral results. *a, b*, Model predictions from computational model described previously (Collins and Frank, 2013). *c, d*, Behavioral data, whole group ( $N = 35$ ). *a, c*, Positive transfer is evident by enhanced learning curves for H3 context in which it was possible to generalize the TS T51 learned previously in context H1, compared with H4, associated with a new TS. Black lines indicate within-subject performance differences between contexts H3 and H4. *b, d*, Negative transfer is evident by the distribution of error types. Errors in H4 were more often associated with neglecting the high dimension (NH) rather than low dimension (NL), indicating that participants selected actions that would have been correct in H1 or H2, even though these rules were not applicable here. *e, f*, Individual differences were evident such that some participants exhibited very robust positive transfer ( $N = 19$ ), whereas others showed none at all ( $N = 16$ ). Error bars indicate SEM. \* indicates  $p < 0.05$ .

generalization is thus expected to improve performance on S2 for H3 but not H4 without being influenced by S1 stimuli. Although the pattern of correct actions across contexts and stimuli was identical for all subjects, the mapping between actions and physical key-presses was randomized across subjects, such that the correct finger patterns were different across subjects.

**Participants.** Forty subjects participated in this experiment. Technical software problems occurred for 2 subjects, and 3 subjects failed to attend to the task (as indicated by a large number of nonresponses) and were thus excluded from analysis. Final sample size for behavioral analysis was  $N = 35$  subjects (16 female and 19 males). Behavioral results for the first block, during which subjects are naive to task structure and thus incidental structure learning can be tested, were reported by Collins and Frank, 2013. Here we report results across all three blocks and the novel integration with EEG for the first time.

**EEG methods.** EEG was recorded using a 128 channel EGI system. Four anterior facial electrodes inferior to the brim of the cap were removed because of poor reliability, leaving 125 total electrodes after reconstruction of the vertex site. We used previously identified data cleaning and preprocessing methods (Cavanagh et al., 2009) facilitated by the EEGlab toolbox (Delorme and Makeig, 2004).

**Preprocessing.** EEG was recorded continuously with hardware filters set from 0.1 to 100 Hz, a sampling rate of 250 Hz, and an online vertex reference. Continuous EEG was epoched around the cues (−2500 to 4500 ms). Data were then visually inspected to identify bad channels to be interpolated and bad epochs to be rejected. Blinks were removed using independent component analysis from EEGlab. Of the 35 subjects included in the behavioral analysis, 2 presented strong artifacts and had to be excluded from the EEG analysis.

**ROIs.** Given the electrode density, we identified dorsolateral prefrontal *a priori* regions of interest (see list below). EEG was averaged across these electrodes for the event-related analyses: rdIPFC electrodes, 2, 3, 4, 116,

117, 118, 122, 123, and 124; and ldIPFC electrodes, 19, 20, 23, 24, 26, 27, 28, 33, and 34.

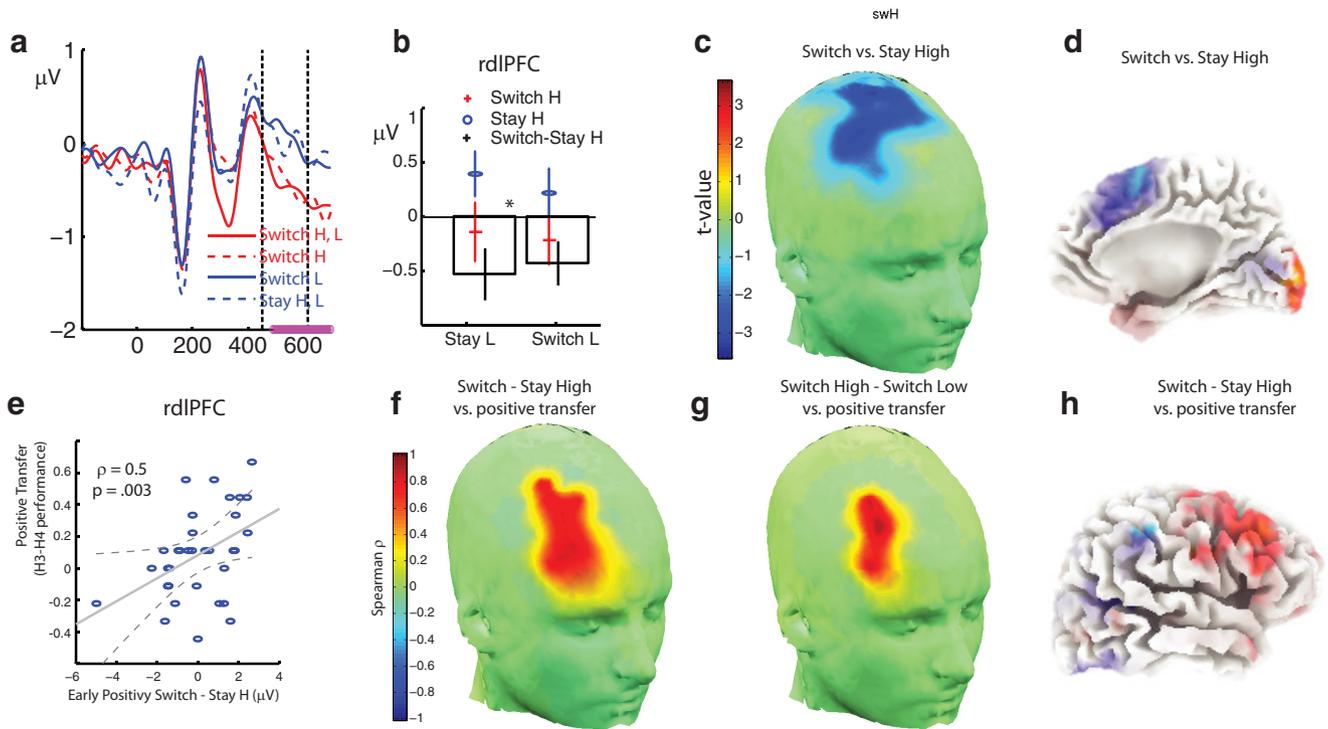
For a *a posteriori* comparison with other brain regions, we defined a further mdIPFC ROI. For further investigation of hierarchical gradient in right dorsolateral PFC, we defined more precise anterior and posterior subregions as such: mdIPFC electrodes, 5, 6, 7, 11, 12, 13, 106, and 112; anterior rdIPFC electrodes, 1, 2, 3, 4, 8, 9 and 10; and posterior rdIPFC electrodes, 116, 117, 118, 122, 123, and 124.

**ERP.** For event-related potentials (ERP), data were bandpass filtered from 0.5 to 20 Hz. ERPs were baseline corrected to the average activity from −300 to 0 ms before the visual input. Components were identified on grand-average stimulus-locked ERPs and quantified within each subject by finding local minima/maxima within time windows of interest. Specifically, we identified candidate temporal windows for the P2 peak and N1 trough based on the grand average (over all participants); we then algorithmically identified the P2 peak and N1 trough for each participant within these windows. The early ERP positivity was quantified as the difference between the P2 peak (224–236 ms) and the preceding N1 trough (152–172 ms). For the late ERP, negativity was quantified as the difference from activity in the 450–609 ms time window from prestimulus baseline activity (−100 to −0 ms precue period). This time window was identified from 3 factors. First, published data found late negativity in task switching paradigm with cue-target interval 0 (Nicholson et al., 2005), which were 406–698 ms. Second, the onset of a true late negativity would need to occur after the last phase-locked component of the obligatory peak-trough cycles of the canonical ERP, occurring here at ~450 ms. Last, to ensure minimal overlap with reaction times (RTs), the window for the late negativity was constrained to end at the time of mean RT of the fastest condition (Stay-HL, 609 ms).

The number of trials for ERP analyses ranged between 17 and 34 for all conditions and subjects.

**RT effects on late negativity.** To confirm that the results of the stronger late negativity in switch in high dimension H compared with stay H were not the result of differences in RTs, by which dimensions H and L are identified, we performed a within-subjects analysis of RTs across two factors (condition: Switch-H/Stay-L vs Stay-H/Switch-L; RT: slow vs fast, as identified by a median split within each condition). It should be noted in particular that slow Stay-H/Switch-L trials were slower ( $p < 0.001$ ) than fast Switch-L/Stay-L trials. We found a marginal effect of condition coherent with late negativity ( $t = 1.8, p = 0.08$ ), but no effect of RT ( $t = 0.57, p = 0.57$ ).

**Source localization methods.** To investigate the likely sources of the topographical results shown in results (see Fig. 3), we applied an EEG source localization method known as standardized low resolution electrical tomography analysis (sLORETA) (Pascual-Marqui, 2002). Unlike some methods of source localization, sLORETA represents the smoothest distribution of source activities without *a priori* user specification. No additional baselines or transformations were applied, and multiple comparisons were controlled following 5000 iterations of permutation testing. We specifically investigated two contrasts: (1) the difference between switch high and stay high conditions in the time window of the late negativity, and (2) the correlation between behavior and the early positivity for switch high versus stay high conditions. Time windows for sLORETA analyses were the same as ERP analyses with one exception. The source estimate of the early positivity effect was tested only at the P2 component, in contrast to the ERP analyses, which used the P2-N1 dif-



**Figure 3.** Cue-locked ERP effects ( $N = 33$ ). Top row, Late negativity. **a**, ERP over right dorsolateral prefrontal (rdIPFC) electrodes for switches and stays on the high and low dimensions. Pink circles represent significant effect of Switch-H versus Stay-H ( $p < 0.05$ , uncorrected). **b**, Average value over late window 480–580 ms shows main effect of Switch-H versus Stay-H ( $p = 0.008$ ; black bars represent within-subject differences), but no effect of Switch-L versus Stay-L ( $p > 0.7$ ). **c**,  $t$  values for contrasts of switching on high dimension ( $p < 0.01$ , uncorrected), revealing that these late ERP effects occur over mid and right anterior sites. **d**, Minimum norm source-localization of the late negativity effect implicated the right superior/medial frontal gyri. Bottom row, Early positivity effects. **e**, The amplitude difference of Switch-H versus Stay-H, a marker of hierarchical task structure during learning, was correlated with subsequent positive transfer. This effect remained significant when removing one low early-positivity effect outlier ( $p < 0.05$ ). **f, g**, Topographic maps of significant ( $p < 0.01$ ) correlations between ERP indices and positive transfer, revealing that these correlations are selective to Switch-H, compared with both Stay-H and to Switch-L, and are localized to right anterior electrodes. **h**, Minimum norm source-localization of the early positivity effect implicated the right middle frontal gyrus.

ference. We chose this restriction on both methodological and practical considerations. Empirically, the topographical relationships between P2 and performance were highly similar but less statistically powerful than the P2-N1 difference. Although the P2-N1 difference is likely to be a more accurate assessment of the temporal nature of the phase-locked oscillations that are modulated by task-specific effects, this contrast includes a difference between varying topographical effects (i.e., quantification of the discrete N1 and P2 stages), compromising accurate spatial estimation. Therefore, we used a more conservative approach to source localization by keeping the data as close to a natural brain state (i.e., the P2 peak) as possible.

**Multivariate methods.** For the purpose of EEG multivariate pattern classification, subjects underwent a color and shape attention localizer task before the experimental task described above (see Fig. 5*a*). The task consisted of 8 blocks, each with 16 colored shapes. In four blocks, a single fixed shape changed color every 1.5 s. In the other four blocks, shapes changed every 1.5 s but remained the same color. Participants were told to count the number of occurrences of a pseudo-randomly chosen shape or color for each block, and to report it at the end of the block. This task lasted  $< 5$  min. The EEG data from one subject during this task were contaminated by artifacts; therefore, the subsequent sample size for classification analyses was  $N = 32$ .

**Classification.** The data from this localizer task were used to train classifiers to distinguish brain states associated with attention to color versus attention to shape. All classification sets were sampled with an equal number of color and shape trials. Each classifier was trained on the voltage from 125 electrodes and 12 samples (48 ms) beginning  $-100$  ms to 500 ms peristimulus onset, with overlapping temporal windows of 50%. Classification was performed for each subject separately. We used the LASSO algorithm, a penalized logistic regression method (Sjöstrand et al., 2012). Extensive pretask development using six different classifica-

tion algorithms revealed the LASSO to be the strongest and most reliable classification algorithm for the current preprocessing strategy. Training was performed  $> 60\%$  randomly chosen localizer task trials. Regularization weights that constrained sparsity of the logistic regression were selected to optimize generalization on a separate validation set (20% of trials). Classification performance was assessed on the test set, containing the final unbiased 20% of remaining trials. Training, validation, and test sets, as well as TS task trial features, were all mean and SD normalized on training set features. Each temporospatial classifier was performed 50 times on the same dataset, with random selection of the training, validation, and test sets. Weights of these 50 classifications were averaged for each participant for application to the EEG data gathered during the TS experiment. Results described were similar across different algorithms (elastic net, linear support vector machines, and discriminant analysis) and cross-validation methods.

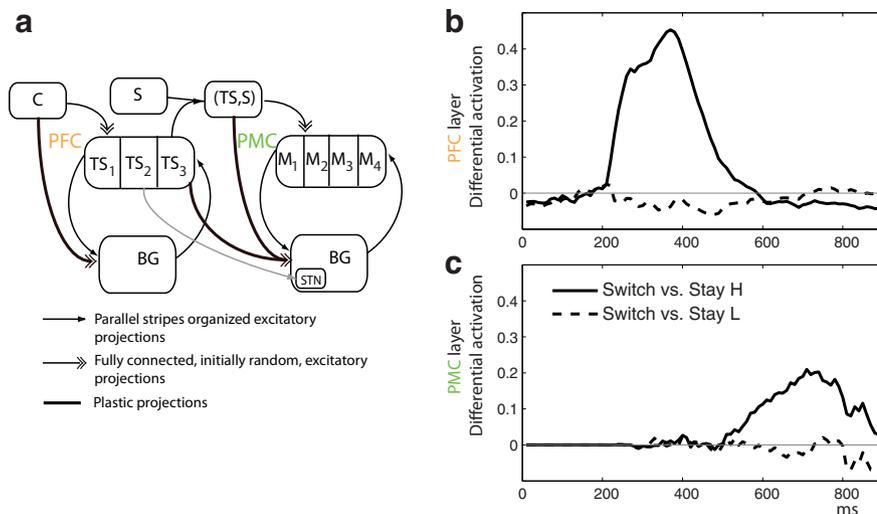
**Modeling methods.** Behavioral predictions in Figure 2*a, b* come from simulation of the Context-TS model, as reported previously (Collins and Frank, 2013). Parameters were  $\beta = 10$ ,  $\alpha = 2$ . A total of 1000 simulations were averaged.

Dynamic activations reported in Figure 4 were obtained from averaging  $> 1000$  trials per condition from 10 different network simulations of the hierarchical network described previously (Collins and Frank, 2013). All parameters were identical to those reported earlier; reported activations were averaged over the entire “PFC” (prefrontal) and “PMC” (premotor) layers.

## Results

### Behavioral results

Half of the subjects in the first block ( $N = 18$  vs  $N = 17$ ) used color as the context dimension as assessed by the switch-cost comparison (whereas the other half used shape), thus confirming



**Figure 4.** Neural network activation shows differential early and late switch H effects and no switch L effects. **a**, Sketch of the neural network described by Collins and Frank (2013). TSs are selected in the more anterior part of prefrontal cortex (labeled here PFC)–basal ganglia (BG) loop in response to a context (C). Motor actions (M) are chosen in a more posterior PFC (or PMC)–BG loop in response to TS-contextualized stimuli S. **b**, Anterior PFC layer shows early, H-dimension-specific switch-related increase in activation. This layer is related to the cue-TS selection process. **c**, Posterior prefrontal (PMC) layer shows late, H-dimension-specific switch-related increase in activation. This layer only becomes active in the presence of a target (or stimulus) after the target is presented.

that there was no overall bias to treat one dimension or another as context or higher order.

Moreover, the behavior across all three blocks replicated our previous findings, exhibiting the same pattern predicted by computational models of structure building and generalization (Collins and Frank, 2013). Positive transfer was measured as the difference in accuracy for the first 3 trials of the transfer test condition H3 compared with new test condition H4 (Fig. 2*a,c*). This early performance improvement was significantly positive ( $t = 2.3$ ,  $p = 0.029$ ). As described previously, the task design was such that hierarchical structure is evident not only by positive transfer in H3, but also by the pattern of errors in H4. Specifically, negative transfer results from abusive generalization of a known TS partially overlapping with the new TS. Thus, negative transfer would be evident by a greater preponderance of errors in the H4 condition in which responses would have been correct for the other context. These errors, characterized as “neglecting the high dimension” (NH) were more frequent in the H4 condition than the H3 condition ( $t = 2.64$ ,  $p = 0.01$ ; Fig. 2*b,d*). This pattern was present even in the first block of the experiment, in which subjects were necessarily naive of task dynamics (NH vs NL in H4:  $t = 2.6$ ,  $p = 0.01$ ; H4 vs H3 in NH:  $t = 3$ ,  $p = 0.004$ ; interaction error type vs context:  $t = 2.45$ ,  $p = 0.02$ ).

There were substantial individual differences in this measure of incidental structure, which were independent of overall initial learning speed (Spearman’s  $\rho < 0.16$ ,  $p > 0.35$ ). For display purposes, we median split the sample into individuals associated with substantial positive transfer, who were putatively more likely to structure the task hierarchically for potential generalization (Fig 2*e*), whereas the other participants showed no evidence of transfer (Fig 2*f*). Several other behavioral markers predicted by our model of learning hierarchical structure were also observed, including within-task set errors after context switches, and patterns of RTs for each error type (Collins and Frank, 2013).

In sum, the pattern of behavioral results confirmed that subjects spontaneously constructed hierarchically structured TSs

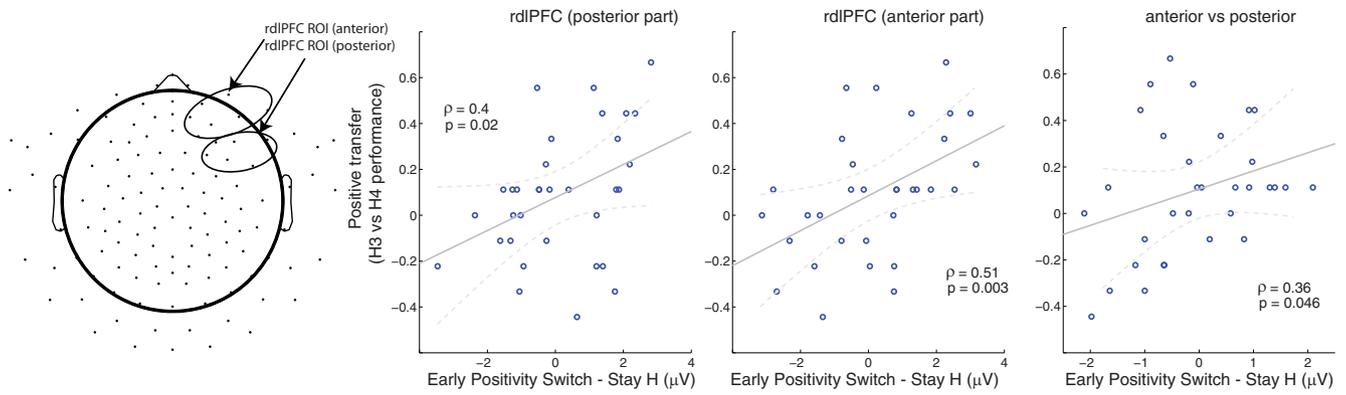
without incentive to do so and that we could reasonably infer the nature of their structure (i.e., their assignment of higher and lower dimensions via the proxy measure of differential RT switch-cost). However, substantial individual differences in the ability to transfer such structure were evident. We next sought to investigate whether such individual differences might be related to the tendency to structure the task hierarchically during the initial learning process. To do so, we investigated whether neural indices of latent structure were observable during the learning process, and if so, whether these could predict transfer.

### EEG results

We reasoned that, if participants treated the task hierarchically, then a switch of the higher dimension from one trial to the next (e.g., from one color to another) would be akin to a task-switch, and thus high (but not low) dimension changes would be accompanied by EEG markers of task switching. Our neural model of hierarchically structured learning exhibits

prefrontal activity dynamics at two distinct time points associated with a task-switch: early, when a new TS is updated in anterior PFC; followed by a later signal related to contextualizing the sensory stimulus in terms of the new TS in a more posterior PFC region. Similarly, the (instructed) task switching literature has identified two commonly observed ERP indices of task switching: an early positivity observed as early as 100–200 ms after TS cue onset (Nicholson et al., 2005; Rushworth et al., 2005; Karayanidis et al., 2009, 2010; Wylie et al., 2009; Nessler et al., 2012), thought to support proactive inhibition of the previous task set and a late (300–700 ms after target onset, depending on preparation time) negativity indicating a more stimulus-dependent TS reconfiguration (Swainson et al., 2003; Nicholson et al., 2005, 2006; Gladwin et al., 2006; Jamadar et al., 2010; Karayanidis et al., 2010; Li et al., 2012; Elchlepp et al., 2013). We thus assessed these ERPs during the asymptotic learning phase to provide independent evidence of structured learning.

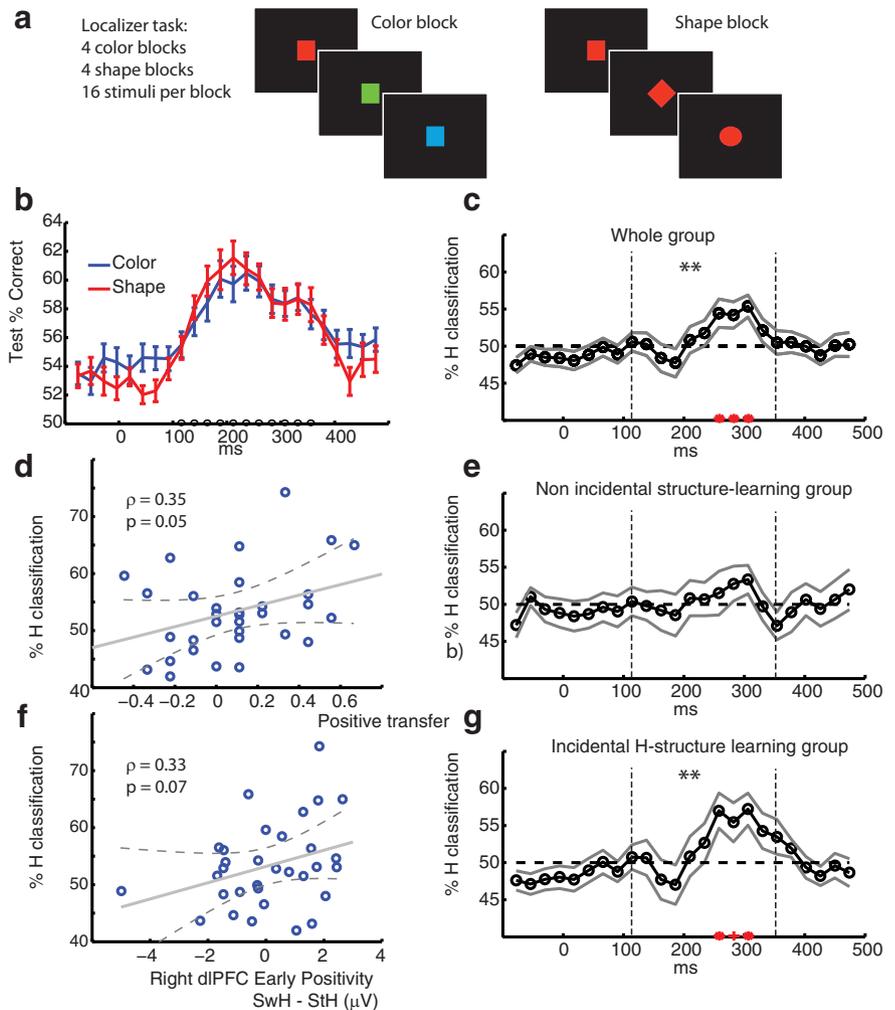
We first investigated the switch-related late negativity, focusing on ROIs in electrodes over lateral prefrontal cortex (PFC). Figure 3*c* displays the topographical map of  $t$  values for Switch-H versus Switch-L, revealing effects over the *a priori* right dorsolateral PFC ROI. Within this ROI, there was an expected main effect of switch vs stay on the high dimension ( $t < -2.8$ ,  $p < 0.008$ ), but no main effect of switch versus stay on the low dimension and no interaction ( $p > 0.7$ ). *Post hoc* tests showed that the Switch-H effect was present on both Switch-L ( $t < -2.3$ ,  $p < 0.025$ ) and Stay-L trials ( $t < -2.1$ ,  $p < 0.05$ ). Thus, this neural signature is sensitive only to change in high-level TS and not to the stimulus or action. The robustness of this effect was further confirmed by separating slow from fast trials within each condition, demonstrating that this effect was not related to response time (see Materials and Methods). The estimated source of this late negativity contrast was tested as the paired difference between conditions. Figure 3*d* displays the log of ratio of averages (similar to the log of the  $F$  ratio), demonstrating that the right superior and medial frontal gyri were most clearly modulated by the Switch-H versus Stay-H contrast.



**Figure 5.** Early positivity effects in anterior versus posterior rdIPFC electrodes. Topoplot on left, Position of the *a posteriori* rdIPFC ROIs. The scatter plots show that the amplitude difference of Switch-H versus Stay-H, a marker of hierarchical task structure during learning, is correlated with subsequent positive transfer in the posterior part of rdIPFC ROI (left) and anterior part rdIPFC ROI (middle). The correlation is stronger in anterior than posterior electrodes (right).

The other candidate index of TS switching, the early switch-related positivity (quantified as the difference between the P2 peak [224–236 ms] and the preceding N1 trough [152–172 ms]; see Materials and Methods), was not evident across the entire group of subjects within lateral prefrontal cortex ROI but was observable selectively in those participants most likely to structure the task hierarchically and hence exhibit transfer. Indeed, the magnitude of the early positivity Switch-H/Stay-H difference wave over right anterior lateral prefrontal cortex was correlated with subsequent positive transfer (Fig. 3*e*; Spearman  $\rho = 0.5$ ,  $p = 0.003$ ). Critically, transfer was only predicted by ERP to switches of the high dimension: we did not find any electrode where early Switch-L versus Stay-L trials activity predicted transfer. Follow-up analysis further confirmed the specificity of this effect, where the contrast of Switch-H/Stay-L against Stay-H/Switch-L (controlling for the overall amount of visual change in a trial) predicted positive transfer. The performance-dependent modulation of condition-specific P2 effects were assessed in sLORETA using a regression model with the paired contrast of Switch-H and Stay-H conditions. Figure 3*h* displays the correlation coefficient, showing that the right middle frontal gyrus was clearly indicated as being preferentially modulated by Switch-H vs Stay-H in relation to behavioral positive transfer.

These two spatially and temporally dissociated electrophysiological findings lend further support the hypothesis that participants spontaneously built TSs and that we can identify which TS structure they built from the RT switch-cost measure. Furthermore, they support our proposed implementation of structure building.



**Figure 6.** Classification results ( $N = 32$ ). **a**, Localizer task. Subjects saw 8 blocks of stimuli changing on a single dimension, either color or shape. These localizers were used to train the pattern classifier to identify neural indices indicative of attention to color or shape and were subsequently applied to classify activity during the asymptotic learning phase of the structure learning task. **b**, Classifier performance for localizer color and shape trials. **c, e, g**, Percentage classification as high dimension across the whole group of subjects (**c**), and separately in those who exhibited little evidence for positive transfer (**e**), and those showing substantial evidence for positive transfer (**g**). Groups are the same as in Figure 2*c, d, f*. Scatterplots represent the correlation between classification performance and positive transfer (**d**) or Switch-H-related early positivity (bottom). Error bars indicate SEM. **b**, Open circles represent significantly higher classification than prebaseline ( $p = 0.05$ , uncorrected). **c, g**, \*\*Significant classification toward H over the relevant time interval ( $p < 0.01$ ). Red asterisk indicates *post hoc* tests of significance ( $p < 0.05$ ); +  $p = 0.07$ .

### Computational model

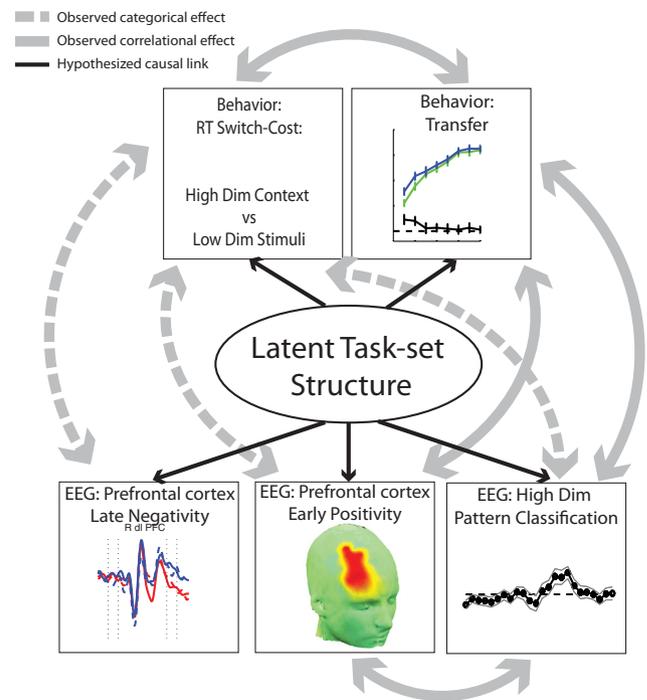
We simulated the task using our previously described neural network model of hierarchical frontostriatal circuits in learning hierarchical rule structure (Collins and Frank, 2013). The model includes an anterior prefrontal cortex layer (labeled “PFC”) that selects TS representations given a contextual cue and a more posterior prefrontal layer (labeled here PMC) layer that selects motor actions given the selected PFC representation and sensory stimulus (Fig. 4*a*) (for detailed model description, see Collins and Frank, 2013). Figure 4*b, c* illustrates the effects of task-switch dynamics on activity in prefrontal cortical layers of the model, as a hypothesized generative mechanism for the early and late ERP switch effects. Simulations showed in both prefrontal cortex layers of the network an increased activation for Switch-H versus Stay-H (but not for Switch-L vs Stay-L). However, this effect occurred early in the anterior layer responsible for cue TS selection (Fig. 3*b*) but late in the posterior layer responsible for engaging in response action selection contextualized by selected PFC TS (Fig. 3*c*). This model prediction is qualitatively similar to the dissociation in early cue-locked EEG activity to TS switches versus a late target locked EEG dynamic to response selection. It is tantalizing to note that the sources of the early and late EEG effects were more anterior and posterior of the frontal cortex (Fig. 3), as predicted by the model, although spatial limitations of EEG technique preclude more accurate estimation of true generative sources.

Furthermore, as predicted by the neural model of this task and these ERP components during task switches, the early positivity effect, as a putative indicator of TS update, was significantly stronger in anterior than posterior right dorsolateral PFC electrodes ( $p < 0.05$ ; Fig. 5), although it was significant in both ( $p = 0.003$ , and  $p = 0.02$ , respectively). Conversely, the late negativity effect, as a putative indicator of action selection contextualized by TS, was significant only in posterior ( $p = 0.004$ ) but not anterior electrodes ( $p = 0.09$ ), although the difference between them was not significant. Thus, although with EEG we cannot precisely identify the source of the individual components, we can precisely identify their relative temporal dynamics, and these findings, together with the source localization results, are broadly consistent with the hierarchical anterior–posterior gradient of action selection within lateral prefrontal cortex (Koechlin et al., 2003; Badre and D’Esposito, 2007; Badre et al., 2010) and in our model (Collins and Frank, 2013).

In contrast to these LPFC task-switching effects specific to the high dimension, we found that ERPs over medial PFC responded significantly but nondiscriminately and additively to changes in either the L or the H dimension ( $p < 0.01$  for both dimensions, over N2 and P3 ERPs; data not shown). This finding is consistent with various evidence implicating mPFC ERPs as reflective of unexpected outcomes and response conflict (Yeung et al., 2004; Cohen, 2011) and further reifies the specificity of the LPFC effects to higher-order structure.

### Multivariate pattern analysis

In addition to these traditional ERP markers of task switching, we next sought to assess whether a decomposition of EEG activities across all electrodes could be leveraged to decode whether participants exhibited preferential attention to higher-order dimensions for signifying the valid TS and guiding response selection, as predicted by the hierarchical structure model. Before the main experiment, we administered a “localizer task” (Fig. 6*a*) to train a multivariate pattern classifier (using LASSO; see Materials and Methods) to decode from EEG activity whether participants



**Figure 7.** Indirect observations of latent TS structure. Although hierarchical TS structure is latent, and hence directly unobservable, converging indicators of such structure are evident in behavior and dynamic EEG activity. These indices are correlated with each other (full gray arrows), suggesting that they reflect the same underlying construct. Similarly, these measures exhibit an asymmetric role of high and low dimensions (dotted gray arrows).

were paying attention to color or to shape, when only one of the features varied from one trial to the next (see Materials and Methods). The classifier was successful at doing so, with cross-validated accuracy reaching a maximum of  $\sim 60\%$  at 200 ms after stimulus onset, significantly above chance ( $p < 0.001$ ; Fig. 6*b*). There was no bias toward color or shape in classification. We identified the time window (116–356 ms) during which classification was better than baseline for further analysis on the learning dataset.

We then used LASSO weights from this initial localizer task to classify EEG activity recorded when both feature dimensions could change from trial to trial during the structured learning task, during the asymptotic learning phase. Notably, based on EEG signals alone, classifier weights from the separate localizer task reliably predicted that participants were preferentially attending to the high dimension in the reinforcement-learning task. Specifically, the classifier was more likely to predict attention to shape in those participants for whom shape was the high dimension, and was more likely to predict attention to color in those participants for whom color was the high dimension, despite no bias in the initial localizer classification toward color or shape in either group (GLM with time within time window as a covariate; fixed effect different from chance:  $p = 0.007$ ). This finding suggests that EEG patterns associated with simple attention to shapes versus attention to color can be leveraged to determine the latent aspects of imperative learning cues that participants are using to inform their hierarchical decision strategies. *Post hoc* analyses showed that the time course of these effects was such that classification was significantly above chance, between 258 and 358 ms ( $t = 2.53$ ,  $p = 0.01$ ). Strikingly, the degree to which the classifier was predictive of attention to the high dimension was significantly related to the extent of positive

transfer (Spearman  $\rho = 0.35$ ,  $p < 0.05$ ), supporting the notion that attending to hierarchical task structure facilitates generalization of this structure to new contexts. Classification of high dimension was also marginally correlated with the early positivity switch-H effect (Spearman  $\rho = 0.33$ ,  $p = 0.068$ ), supporting the interpretation that both of these EEG measures are indicators of task structure, even though their signatures were observed during distinct time points.

Collectively, these findings indicate that it is possible to uncover the construction of latent structure is constructed in brain states through a conjunction of previously identified behavioral and EEG benchmarks, theoretically informed models, and data-driven approaches (Figs. 6 and 7).

## Discussion

These findings provide strong support for the idea that healthy adult human subjects create latent TS structure during reinforcement learning of arbitrary new skills, which provides potential advantages for future transfer of such structure. Because most reinforcement learning problems can be handled by a multiplicity of mechanisms that support good asymptotic performance (Collins and Frank, 2012), it is important to identify the strategies that subjects use. The findings reported here exposed an array of such independent observables that are indicative of the specific hierarchical strategy of building TS structure.

Our structure-learning hypothesis relies strongly on the theory that action selection policies are hierarchically nested, with a higher context-rule (or TS) selection level constraining a lower stimulus-action selection level (Miller and Cohen, 2001). Although much evidence supports this basic tenet during instructed task switching and cognitive control tasks, it is unclear how the brain spontaneously constructs such structure during learning. Computational models highlight the potential utility of such strategies for long-term optimality, even when there is some cost in using them during initial learning of tasks that have no obvious structure (Collins and Koechlin, 2012; Collins and Frank, 2013). We posit that these strategies arise from hierarchically nested frontostriatal circuits (Frank and Badre, 2012; Collins and Frank, 2013) and that prefrontal cortex facilitates the association of contexts to latent TSs. Electrophysiological findings confirmed model predictions that the late negativity effect was linked to more posterior regions of the prefrontal cortex, whereas the early effect was associated with more anterior regions. These results support the predictions of the aforementioned computational models, with a more anterior, earlier switching signal corresponding to TS selection followed by a more posterior, later signal associated with TS controlled action selection. Although EEG source localization techniques provide complementary spatial information to ERP scalp topographies and these findings are consistent with the literature, further research using more precise spatial tools, such as fMRI, should be used to confirm the precise localization of spatially dissociated neural systems involved in spontaneous structure building and contextualizing of stimulus-action-outcome learning.

Electrophysiological data provided additional evidence for spontaneous rule generation, as attention for the stimulus features decoded from EEG activity was preferentially guided toward the dimension indicative of the valid TS. This classification finding was particularly notable given that the classifier was trained based on EEG activity during a separate localizer task in which there was no bias for one feature or another and that TS structure was not given by the task but inferred from individual subjects' behavior.

Another key aspect of structure learning is the necessity of representing rules abstractly. This abstract representation, not tied to the context in which the rule is learned or applied, enables the building of a repertoire of rules that can be generalized to new contexts (Collins and Koechlin, 2012; Collins and Frank, 2013). Recent neuroimaging research has provided evidence for neural indicators of abstract rules in instructed situations (Haynes et al., 2007; Cole et al., 2011; Reverberi et al., 2011; Woolgar et al., 2011a; b). Here we provide support for the existence of abstract TS representations that spontaneously develop during the learning of simple rules. Uncovering the mechanisms governing this spontaneous adoption of hierarchical structure during learning may be fruitful for understanding the flexibility of human cognition and how this flexibility may be challenged in developmental learning disabilities.

## References

- 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. [CrossRef Medline](#)
- Badre D, Kayser AS, D'Esposito M (2010) Frontal cortex and the discovery of abstract action rules. *Neuron* 66:315–326. [CrossRef Medline](#)
- Badre D, Frank MJ (2012) Mechanisms of hierarchical reinforcement learning in cortico-striatal circuits: 2. Evidence from fMRI. *Cereb Cortex* 22:527–536. [CrossRef Medline](#)
- Bunge S, Wallis JD, Parker A, Brass M, Crone EA, Hoshi E, Sakai K (2005) Neural circuitry underlying rule use in humans and nonhuman primates. *J Neurosci* 25:10347–10350. [CrossRef Medline](#)
- Cavanagh JF, Cohen MX, Allen JJ (2009) Prelude to and resolution of an error: EEG phase synchrony reveals cognitive control dynamics during action monitoring. *J Neurosci* 29:98–105. [CrossRef Medline](#)
- Cohen MX (2011) Error-related medial frontal theta activity predicts cingulate-related structural connectivity. *Neuroimage* 55:1373–1383. [CrossRef Medline](#)
- Cole MW, Etzel JA, Zacks JM, Schneider W, Braver TS (2011) Rapid transfer of abstract rules to novel contexts in human lateral prefrontal cortex. *Front Hum Neurosci* 5:142. [Medline](#)
- Collins A, Koechlin E (2012) Reasoning, learning, and creativity: frontal lobe function and human decision-making. *PLoS Biol* 10:e1001293. [CrossRef Medline](#)
- Collins AG, Frank MJ (2012) How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. *Eur J Neurosci* 35:1024–1035. [CrossRef Medline](#)
- Collins AG, Frank MJ (2013) Cognitive control over learning: creating, clustering, and generalizing task-set structure. *Psychol Rev* 120:190–229. [CrossRef Medline](#)
- Delorme A, Makeig S (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods* 134:9–21. [CrossRef Medline](#)
- Elchlepp H, Rumball F, Lavric A (2013) A brain-potential correlate of task-set conflict. *Psychophysiology* 50:314–323. [CrossRef Medline](#)
- Frank MJ, Loughry B, O'Reilly RC (2001) Interactions between frontal cortex and basal ganglia in working memory: a computational model. *Cogn Affect Behav Neurosci* 1:137–160. [CrossRef Medline](#)
- Frank MJ, Badre D (2012) Mechanisms of hierarchical reinforcement learning in corticostriatal circuits: 1. Computational analysis. *Cereb Cortex* 22:509–526. [CrossRef Medline](#)
- Gladwin TE, Lindsen JP, de Jong R (2006) Pre-stimulus EEG effects related to response speed, task switching and upcoming response hand. *Biol Psychol* 72:15–34. [CrossRef Medline](#)
- Haynes JD, Sakai K, Rees G, Gilbert S, Frith C, Passingham RE (2007) Reading hidden intentions in the human brain. *Curr Biol* 17:323–328. [CrossRef Medline](#)
- Hyafil A, Summerfield C, Koechlin E (2009) Two mechanisms for task switching in the prefrontal cortex. *J Neurosci* 29:5135–5142. [CrossRef Medline](#)
- Jamadar S, Michie PT, Karayanidis F (2010) Sequence effects in cued task switching modulate response preparedness and repetition priming processes. *Psychophysiology* 47:365–386. [CrossRef Medline](#)

- Karayanidis F, Jamadar S, Ruge H, Phillips N, Heathcote A, Forstmann BU (2010) Advance preparation in task switching: converging evidence from behavioral, brain activation, and model-based approaches. *Front Psychol* 1:25. [CrossRef Medline](#)
- Karayanidis F, Mansfield EL, Galloway KL, Smith JL, Provost A, Heathcote A (2009) Anticipatory reconfiguration elicited by fully and partially informative cues that validly predict a switch in task. *Cogn Affect Behav Neurosci* 9:202–215. [CrossRef Medline](#)
- Koechlin E, Ody C, Kouneiher F (2003) The architecture of cognitive control in the human prefrontal cortex. *Science* 302:1181–1185. [CrossRef Medline](#)
- Li L, Wang M, Zhao QJ, Fogelson N (2012) Neural mechanisms underlying the cost of task switching: an ERP study. *PloS One* 7:e42233. [CrossRef Medline](#)
- Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex. *Annu Rev Neurosci* 24:167–202. [CrossRef Medline](#)
- Miller EK (2000) The prefrontal cortex and cognitive control. *Nat Rev Neurosci* 1:59–65. [CrossRef Medline](#)
- Monsell S (2003) Task switching. *Trends Cogn Sci* 7:134–140. [CrossRef Medline](#)
- Nessler D, Friedman D, Johnson R Jr (2012) A new account of the effect of probability on task switching: ERP evidence following the manipulation of switch probability, cue informativeness and predictability. *Biol Psychol* 91:245–262. [CrossRef Medline](#)
- Nicholson R, Karayanidis F, Poboka D, Heathcote A, Michie PT (2005) Electrophysiological correlates of anticipatory task switching processes. *Psychophysiology* 42:540–554. [CrossRef Medline](#)
- Nicholson R, Karayanidis F, Bumak E, Poboka D, Michie PT (2006) ERPs dissociate the effects of switching task sets and task cues. *Brain Res* 1095:107–123. [CrossRef Medline](#)
- Pascual-Marqui RD (2002) Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details. *Methods Find Exp Clin Pharmacol* 24 [Suppl D]:5–12.
- Reverberi C, G6rgen K, Haynes JD (2012) Compositionality of rule representations in human prefrontal cortex. *Cereb Cortex* 22:1237–1246. [CrossRef Medline](#)
- Rushworth MF, Passingham RE, Nobre AC (2005) Components of attentional set-switching. *Exp Psychol* 52:83–98. [CrossRef Medline](#)
- Sj6rstrand K, Clemmensen LH, Larsen R, Ersb6ll B (2012) SpaSM: a Matlab toolbox for sparse statistical modeling. *J Statist Software*.
- Swainson R, Cunnington R, Jackson GM, Rorden C, Peters AM, Morris PG, Jackson SR (2003) Cognitive control mechanisms revealed by ERP and fMRI: evidence from repeated task switching. *J Cogn Neurosci* 15:785–799. [CrossRef Medline](#)
- Woolgar A, Hampshire A, Thompson R, Duncan J (2011a) Adaptive coding of task-relevant information in human frontoparietal cortex. *J Neurosci* 31:14592–14599. [CrossRef Medline](#)
- Woolgar A, Thompson R, Bor D, Duncan J (2011b) Multi-voxel coding of stimuli, rules, and responses in human frontoparietal cortex. *Neuroimage* 56:744–752. [CrossRef Medline](#)
- Wylie GR, Murray MM, Javitt DC, Foxe JJ (2009) Distinct neurophysiological mechanisms mediate mixing costs and switch costs. *J Cogn Neurosci* 21:105–118. [CrossRef Medline](#)
- Yeung N, Botvinick MM, Cohen JD (2004) The neural basis of error detection: conflict monitoring and the error-related negativity. *Psychol Rev* 111:931–959. [CrossRef Medline](#)