NeuroImage 150 (2017) 150–161

Contents lists available at ScienceDirect

NeuroImage

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

Hierarchical control of procedural and declarative category-learning systems

Benjamin O. Turner[a,*], Matthew J. Crossley[b], F. Gregory Ashby[a]

[a] University of California, Santa Barbara, United States
[b] SRI International, United States

ARTICLE INFO

Keywords:
Cognitive control
Functional connectivity
Dual systems

ABSTRACT

Substantial evidence suggests that human category learning is governed by the interaction of multiple qualitatively distinct neural systems. In this view, procedural memory is used to learn stimulus-response associations, and declarative memory is used to apply explicit rules and test hypotheses about category membership. However, much less is known about the interaction between these systems: how is control passed between systems as they interact to influence motor resources? Here, we used fMRI to elucidate the neural correlates of switching between procedural and declarative categorization systems. We identified a key region of the cerebellum (left Crus I) whose activity was bidirectionally modulated depending on switch direction. We also identified regions of the default mode network (DMN) that were selectively connected to left Crus I during switching. We propose that the cerebellum—in coordination with the DMN—serves a critical role in passing control between procedural and declarative memory systems.

1. Introduction

Evidence that humans have multiple memory systems (Eichenbaum and Cohen, 2001; Squire, 2004; Tulving and Craik, 2000) inspired theories that humans also have multiple, qualitatively distinct category-learning systems (Ashby et al., 1998; Erickson and Kruschke, 1998). According to this view, procedural memory is used to form many-to-one stimulus-to-response mappings (S-R associations), whereas declarative memory is used to apply rules and test explicit hypotheses about category membership. Much of the neuroimaging evidence supporting these distinctions depends on prior research with rule-based (RB) and information-integration (II) category-learning tasks (Hélie et al., 2010; Nomura et al., 2007; Soto et al., 2013; Waldschmidt and Ashby, 2011). In RB tasks, the categories can be learned via an explicit hypothesis-testing procedure (Ashby et al., 1998). In the simplest variant, only one dimension is relevant (e.g., bar width), and the task is to discover this dimension and then map the different dimensional values to the relevant category responses. In II tasks, accuracy is maximized only if information from two or more stimulus dimensions is integrated perceptually at a pre-decisional stage (Ashby and Gott, 1988). In most cases, the optimal strategy in II tasks is difficult or impossible to describe verbally (Ashby et al., 1998). Verbal rules may be (and sometimes are) applied, but they lead to suboptimal performance. Example II and RB categories are illustrated in panels A and B of Fig. 1.

Much evidence suggests that II tasks recruit procedural memory, whereas RB tasks recruit declarative mechanisms. Even so, a natural question for readers unfamiliar with the category-learning literature is how any classification task can be a good choice for studying procedural behaviors. For instance, how can a task with such simple motor demands (e.g., “push a button”) possibly recruit procedural networks that are strongly tied to motor processes? In fact, the empirical evidence is strong that performance improvements in some types of classification tasks are mediated via procedural learning and memory. At least 25 different behavioral dissociations tie II learning to procedural memory and RB learning to declarative memory (for reviews, see Ashby and Maddox, 2005, 2010; Ashby and Valentin, 2017). This hypothesis is further supported by a variety of investigations into the neural underpinnings of successful II and RB learning. Specifically, success in RB tasks depends on a broad neural network that includes the prefrontal cortex (PFC), anterior cingulate, the head of the caudate nucleus, and medial temporal lobe structures—regions that are also frequently associated with declarative memory and executive attention (Brown and Marsden, 1988; Filoteo et al., 2007; Seger and Cincotta, 2006). Arguably, the most important region in this network is the PFC, where rules are thought to be initially represented (Miller and Cohen, 2001; Wallis et al., 2001). Success in II tasks, on the other hand, depends on regions that have been implicated in proce-

[*] Corresponding author.
E-mail address: ben.turner@psych.ucsb.edu (B.O. Turner).

http://dx.doi.org/10.1016/j.neuroimage.2017.02.039
Received 15 December 2016; Received in revised form 30 January 2017; Accepted 13 February 2017
Available online 16 February 2017
1053-8119/ Published by Elsevier Inc.

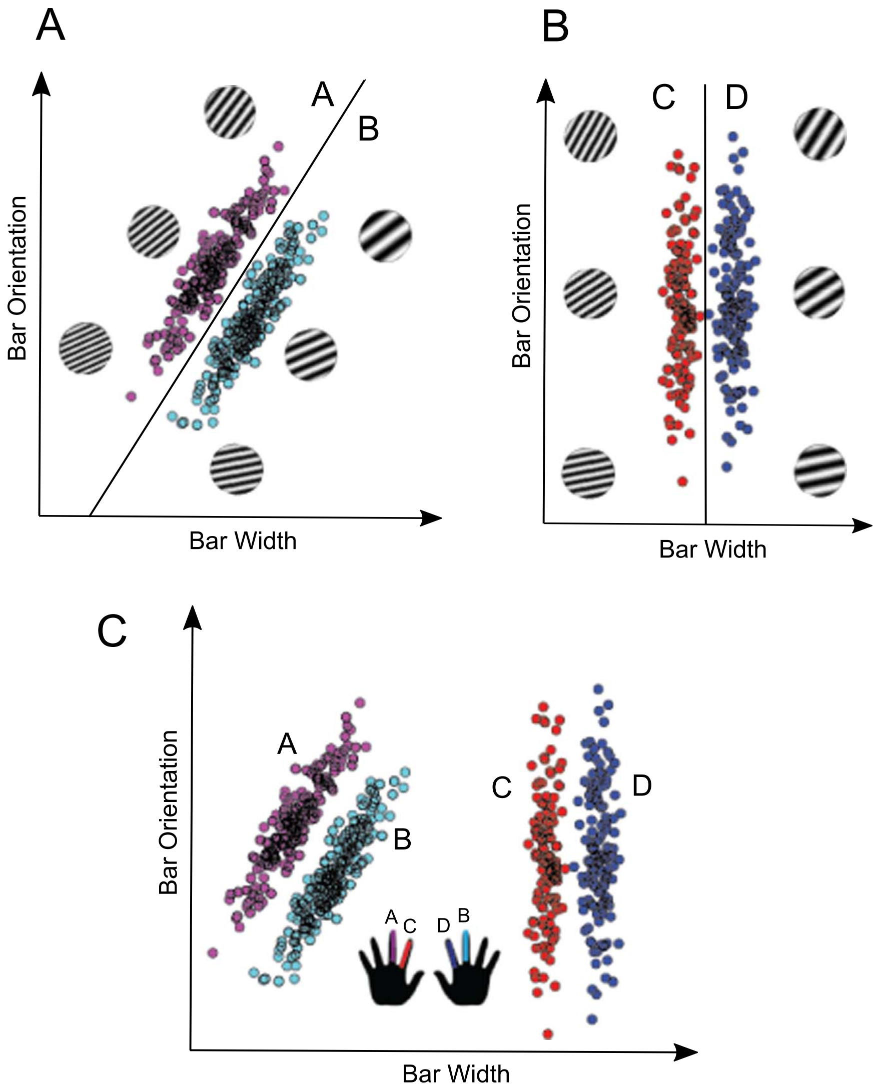

Fig. 1. A: Example II categories. B: Example RB categories. C: Stimuli and categories used in the present experiment. Finger colors correspond to category label.

dural memory, including the striatum, premotor cortex, and the associated sensorimotor basal ganglia loop (Ashby and Ennis, 2006; Filoteo et al., 2005; Knowlton et al., 1996; Nomura et al., 2007). This network is consistent with the idea that S-R associations are built at cortical-striatal synapses via dopamine-dependent reinforcement learning (Ashby and Crossley, 2011; Houk et al., 1995; Joel et al., 2002).

This article asks what brain networks mediate switching between procedural and declarative control. This question is important because the survival demands of daily life are not laboratory-tuned in favor of one system or the other. Rather, some tasks are best solved by declarative systems and others by procedural systems. Control must therefore be passed back-and-forth between the two systems. To address this question, we used a task that required participants to switch between previously-learned RB and II categorization tasks in a trial-by-trial manner. This paradigm allowed us to investigate the activity associated exclusively with switching between systems, rather than with simple performance of one task or the other. Because the task was fairly difficult, our analyses are limited to the subset of participants who were invited back because of their proficiency at the task. Our results therefore apply to members of the population who demonstrate aptitude in such system switching.

Note that our aims are distinct from typical studies of “task-switching,” which usually focus on networks that mediate the switching between separate declarative memory-based tasks. For example, Monsell (2003) describes a popular variation in which participants switch back and forth between deciding whether a single digit is odd or even and whether it is high or low in value. Thus, the focus of the task-switching literature has been on switching between tasks within the same memory system, whereas our focus is on switching between tasks mediated by anatomically and functionally distinct memory systems. This difference may anticipate our main findings. In particular, much of the existing work on task-switching has focused either on task-general switch-related activity (“switch” contrasted with “stay,” collapsing across tasks), or else on task-specific switching activity (e.g., isolating “switch > stay” activity for a particular task, possibly contrasted with another task).

In contrast, we are interested in identifying networks that show modulated activity during switching for both tasks, regardless of the direction of modulation for each task. There are a variety of theories for how control might be passed between systems, which make a variety of predictions for how activity will change depending on the direction of the switch. In order to adjudicate among these theories, we used an analysis method that could identify any common regions, irrespective of the direction of modulation. To preview our key result, we identified a region that showed bidirectional activity: increased activity when switching one direction, and decreased activity in the other.

2. Methods

2.1. Participants

Participants for the fMRI experiment were pre-screened in order to identify those who were able to switch successfully between category structures (see below). A total of 179 participants were pre-screened, of whom 46 were invited back for scanning. Our criteria for inviting participants back were based on the results of decision bound modeling (see 'Decision Bound Models' section below). Specifically, we invited participants back if their data were best fit by a model that assumed a decision strategy of the optimal type on each category structure during any two of the three 100-trial blocks of switch training. A total of 28 participants responded to our invitation, were eligible for scanning, and participated in the fMRI experiment. Of these, seven were excluded (due to technical problems or claustrophobia, excess motion, or problems with the analysis/too few trials of certain types). Our final sample included 21 participants (mean age: 19 years; 16 female; 18 right-handed). All participants were given course credit (during pre-screening) and paid (for the fMRI portion) for their participation, and all had normal or corrected to normal vision. Participants were scanned between 7–23 days after completing pre-screening (mean interval =12.29 days).

2.2. Stimuli and categories

Stimuli were gray-scale, circular sine-wave gratings that varied across trials in spatial frequency (cycles per degree, CPD) and orientation (radians, rad). Each stimulus subtended approximately 5 degrees of visual angle and was displayed against a gray background using routines from the Psychophysics toolbox (Brainard, 1997).

Stimuli were sampled from one of four possible distributions (illustrated in panel C of Fig. 1), following the randomization technique developed by Ashby and Gott (1988). To control for statistical outliers, any sample whose Mahalanobis distance (Fukunaga, 1990) was greater than 3.0 was removed and resampled. This process was repeated until 400 Category A, 400 Category B, 400 Category C and 400 Category D exemplars had been generated. Parameters for these category distributions are reported in Table 1. Each random sample (x, y) was converted to a stimulus according to the nonlinear transformations proposed by Treutwein et al. (1989). This transformation is important, for example, to ensure that all changes of say 10 units in bar width are equally perceptually salient.

2.3. Decision bound modeling

Perfect performance requires a procedural strategy (i.e., perceptual integration) for the II categories and an explicit strategy (one-dimensional rule) for the RB categories. Even so, a one-dimensional explicit rule can achieve accuracy well above chance on the II categories. A participant using an explicit rule on the II categories would be switching between two different declarative memory strategies, rather than between declarative and procedural memory strategies, and so it is critical that we are able to identify such participants before analyzing their fMRI data. We used decision bound modeling to solve this problem (Maddox and Ashby, 1993; Ashby and Gott, 1988).

Decision bound models assume that classification is achieved by dividing the perceptual space into separate response regions via a decision bound, deciding on which side of the bound the current stimulus lies, and then emitting the associated response (Maddox and Ashby, 1993; Ashby and Gott, 1988). We fit three different types of decision bound models—one type that assumed an explicit rule strategy, one type that assumed a procedural strategy, and one type that assumed random guessing. For details, see Ashby and Valentin (2017).

Briefly, explicit rule models assume either a horizontal or vertical decision bound, which is equivalent to classifying stimuli according to whether their orientation is shallow or steep, or whether their bars are thin or thick. Explicit rule models have two free parameters (a criterion on the relevant dimension and a perceptual noise variance). Procedural models assume a general linear decision bound of any arbitrary slope and intercept. These models, which assume linear integration across the two stimulus dimensions, have three free parameters (the slope and intercept of the linear decision bound and a perceptual noise variance). The guessing models assume random guessing that is either unbiased (both responses guessed with equal probability; so no free parameters) or biased (guess one response with probability p and the other with probability $1 - p$, where p is a free parameter).

We estimated best-fitting parameters via maximum likelihood, and used the Bayesian information criterion (BIC; Schwarz, 1978) for model selection. BIC is defined as $\text{BIC} = r \ln N - 2 \ln L$, where r is the number of free parameters, N is the sample size, and L is the likelihood of the model given the data. The BIC statistic penalizes models for extra free parameters. To determine the best-fitting model, the BIC statistic is computed for each model, and the model with the smallest BIC value is the winning model.

Table 1
Category distribution parameters.

	μ_x	μ_y	σ_x	σ_y	cov_{xy}
Category					
II A	43	57	167.91	167.91	59.36
II B	57	43	167.91	167.91	59.36
RB C	140	50	227.27	108.55	0
RB D	160	50	227.27	108.55	0

2.4. Procedure

The experiment consisted of one pre-screening session lasting approximately 50 minutes in duration that included 7 blocks of 100 trials each, and one scanning session that took place at a later date.

2.4.1. Pre-screening

Participants were required to classify a stimulus into one of four categories on every trial. Stimuli sampled from the RB categories were framed by a blue box, and stimuli sampled from the II categories were framed by a green box (see Fig. 2). Participants were informed that the frame colors indicated that different categorization strategies would be necessary for optimal performance. They were further informed that stimuli displayed with a blue frame (RB trials) only required attention to one dimension and that stimuli displayed against a green frame (II trials) required attention to both dimensions. They were instructed to press the 's' key with the middle finger of their left hand for category 'A', to press the 'l' key with the middle finger on their right hand for category 'B', to press the 'd' key with the index finger on their left hand for category 'C', and to press the 'k' key with the index finger on their right hand for category 'D' (see Fig. 1). Participants were further informed that all stimuli displayed with a blue frame belonged to either category 'A' or 'B' and that stimuli displayed with a green frame belonged to either category 'C' or 'D'. Throughout the entire experiment the category labels 'A', 'B', 'C', and 'D' appeared along the bottom of the screen in a spatial position and order that corresponded to the response key – category label mapping.

Each trial began with a fixation cross lasting 750 ms. A stimulus was then presented for a maximum of 5 s. If the participant responded within 5 s the stimulus disappeared, and 500 ms later a feedback tone was presented for 1 s. Correct responses were indicated by a pure tone (500 Hz, 730 ms in duration), and incorrect responses were indicated by a saw-tooth tone (200 Hz, 1220 ms in duration). Participants were first trained on the RB categories for 50 trials, then on the II categories

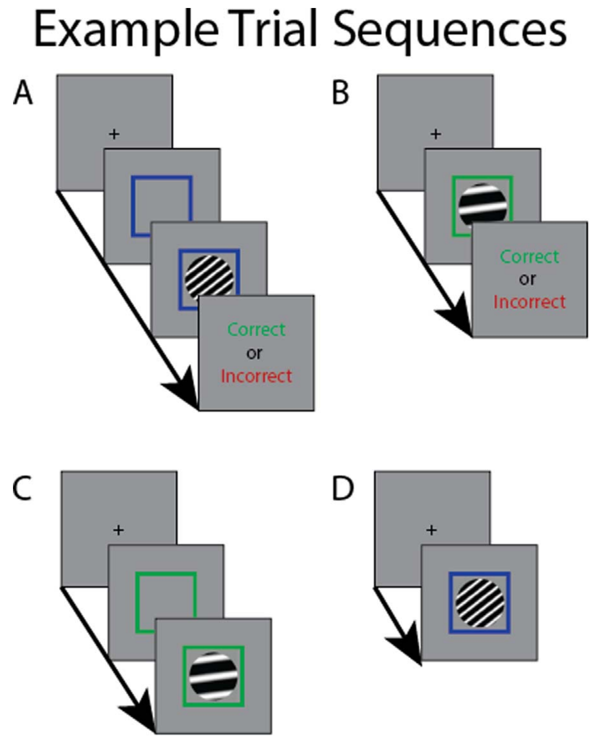

Fig. 2. Possible trial types during scanning. Time between trial events in the scanner was always 1 TR (i.e., 2 s). Time between trial events during pre-screening is described in the text. A: Example RB trial in which the cue, stimulus, and feedback are all presented. B: Example II trial in which the pre-cue is omitted. C: Example II trial in which feedback is omitted. D: Example RB trial in which pre-cue and feedback are omitted.

for 300 trials, and then on randomly intermixed RB and II categories for 350 trials. Participants were free to rest as long as they wished between each block of 50 trials.

2.4.2. Scanning

The task used in the scanning session was identical to that used during pre-screening, except that II and RB trials were intermixed the entire time, and the timing was modified to ensure estimability. Additionally, in order to provide feedback compatible with the loud fMRI environment, correct responses were indicated with a large green check-mark, and incorrect responses were indicated with a large red X, rather than via auditory tones. In particular, we used a partial trials design (Ollinger et al., 2001; Serences, 2004). The ‘canonical’ trial included a cue (that is, the colored box, but without a stimulus) for two seconds, followed by a stimulus plus cue for two seconds, and finishing with visual feedback for two seconds. However, on 50% of trials, the cue was omitted, and on 25% of trials, feedback was omitted. Previous work has demonstrated that a partial feedback rate of 75% is sufficient to maintain performance, especially after performance has stabilized (Ashby et al., 1999; Ashby, 2007). Fig. 2 illustrates all of these possible trial types.

The exact sequence of events (that is, whether each trial presented an II or RB stimulus and included a cue or feedback) was generated using custom scripts that simulated behavior based on each participant's observed II and RB accuracies during pre-screening, and then searched through design space for the order that maximized power for the contrasts of interest (described below), with the constraints that each block contained 50 RB and 50 II trials, and that the cue and feedback probabilities were 50% and 75%, respectively. Six such blocks, comprising 600 trials total, were generated for each participant. Simulations confirmed that neither this design optimization nor our use of a partial trials design resulted in confounds.

2.5. Data acquisition

fMRI data were acquired at the UCSB Brain Imaging Center on a 3 T Siemens Tim Trio MRI scanner with an 8-channel phased array head coil. Head motion was minimized using foam cushions placed around the participant's head. Functional runs used a T2*-weighted single shot gradient echo, echo-planar sequence (TR: 2000 ms; TE: 30 ms; FA: 90° DC4; FOV: 192 mm) with generalized auto- calibrating partially parallel acquisitions (GRAPPA). Each volume consisted of 33 slices (interleaved acquisition, 3 mm thick with.5 mm gap; 3 mm×3 mm in-plane resolution) acquired at an angle manually adjusted to maximize cortical and cerebellar coverage and minimize orbitofrontal distortions. There were a total of six functional runs per participant; the length of each run was determined by the specific sequence created for that run, and ranged from 214–249 TRs (7:08–8:18), with a mean of 233 TRs (7:46). A high-resolution anatomical scan was acquired using a T1-flash sequence (TR: 15 ms; TE: 4.2 ms; FA: 20°; 192 0.89 mm thick sagittal slices;.9 mm×.9 mm in-plane resolution; 256×256 matrix). Additional scans included a localizer, a GRE field map, and a DTI scan, none of which were used in the analyses presented here.

The experiment was run using Psychophysics Toolbox (Brainard, 1997). During scanning, participants responded using the Lumina Response Pad System (model LU400-Pair), with the same finger–category mapping as during the in-lab pre-screening. Stimuli and feedback were presented on a digital projector and screen viewed through a head-coil-mounted mirror during scanning sessions.

2.6. Data analyses

This section describes the fMRI analyses employed in this article. Note that all reported results are qualitatively unchanged when a parametric regressor encoding response time, demeaned with respect to the mean of all trials in a given functional run, is included in the GLM model or the LS-S activity estimation model.

2.6.1. Preprocessing

Data were preprocessed using FEAT v5.98, part of FSL (Jenkinson et al., 2012). Preprocessing steps included motion correction, brain extraction, spatial smoothing (kernel full-width at half-max of 5 mm), and temporal filtering (high pass filter with 120 second cutoff). The data were registered to the MNI atlas using nonlinear registration (warp resolution of 10 mm). The data were prewhitened with respect to the model described in the following section.

2.6.2. GLM analyses

The GLM analyses were carried out in FEAT v5.98. Events of interest were all defined based on the identity of the current stimulus, conditioned on the identity of the stimulus from the preceding trial. Every stimulus was either from the RB categories or from the II categories. Therefore, let J | K denote the event in which the current stimulus is of type J and the stimulus from the preceding trial was of type K, for J and K = RB or II. Then the four events of primary interest are II | II, II | RB, RB | II, and RB | RB (corresponding to II stay, II switch, RB switch, and RB stay trials, respectively). These events were defined only for trials in which responses to both members of the pair were correct. There were 12 other event types of non-interest: II incorrect stimulus, RB incorrect stimulus, post-error stimulus, cue for each of II | II, II | RB, RB | II, RB | RB, post-error cue, II positive feedback, II negative feedback, RB positive feedback, and RB negative feedback. Each of these events was convolved with a gamma HRF (phase=0 s; sigma=3 s; delay =6 s), and included with its temporal derivative. Additionally, the six relative motion parameters returned by MCFLIRT and their six temporal derivatives were included without convolution as nuisance parameters. All events were temporally filtered using the same temporal filtering that was applied to the data.

The primary contrasts of interest were between II | RB and II | II, and between RB | II and RB | RB—that is, II switch versus stay and RB switch versus stay. These contrasts were designed to identify activity associated exclusively with switching to an II or RB task, while removing activity associated with performing II or RB categorization *per se*. These events were defined at the low-level analyses in FEAT. The results from each functional run within participants were combined using a “mid-level” analysis, treating runs as a fixed effect. Finally, results across participants were combined using FEAT's FLAME 1.

Regions that show common switching activity—i.e., increased or decreased activity for one type of switch and increased or decreased activity for the other type of switch—may be involved in passing control from one system to the other (or equivalently, in switching between suppressing and enhancing a single system). In order to identify such regions, we carried out conjunction analyses (Nichols et al., 2005) of each pair of switch versus stay SPMs described above (e.g., II | RB > II | II with RB | RB > RB | II, or RB | II > RB | RB with II | II > II | RB). That is, we constructed the minimum statistic map for each pair, and thresholded the resulting image using FEAT's cluster-based thresholding tool (voxelwise z threshold =2.33; cluster p threshold =0.05).

2.6.3. *Functional connectivity analyses*

The data for the functional connectivity analysis were preprocessed as described above. The connectivity analysis we performed attempts to estimate BOLD responses for individual trials, and is consequently potentially heavily impacted by motion artifacts. To mitigate this effect, we took the betas associated with the motion regressors from the above GLM, and used the residuals with respect to the predicted timeseries derived from these betas as the (motion-scrubbed) timeseries in the subsequent connectivity analysis. With the estimated effect of motion removed, motion parameters were excluded from the models for the connectivity analysis. These steps improve the stability of the activity estimation procedure described below. Nevertheless, the results reported here are qualitatively unchanged if this motion regression step is skipped.

Functional connectivity analyses were carried out using the “beta series regression” approach first employed by Rissman et al. (2004) (see also Cisler et al. (2014)). This approach requires a separate estimate of activity for each occurrence of each event of interest. We generated these estimates using a variant of the least-squares separate (LS-S) approach of Mumford et al. (2012) and Turner et al. (2012). Briefly, the activity due to event i at instance j, $B_{i,j}$, is uniquely estimated by entering it into a GLM as a unique event, while combining all other occurrences of event i into a single regressor, alongside regressors for all occurrences of all other events. The events for which we generated trial-unique estimates are defined in the section “GLM analysis”. However, several events of non-interest were combined for the sake of model stability. In particular, error and post-error stimulus events were combined; positive feedback trials were combined; and negative feedback trials were combined. The resulting design matrix contained 11 events, of which all instances of the correct stimulus events (II | II, II | RB, RB | II, RB | RB) were held out in turn for unique estimation.

In addition to applying this estimation procedure across the whole brain in a voxel-wise manner, we generated an ROI-average timeseries, for the ROI identified in the conjunction analysis described above. We did this by transforming and resampling the group-level cluster in functional space (using nearest-neighbor interpolation) and taking a weighted average of the timeseries of all voxels within this cluster. LS-S was then applied to this ROI-averaged timeseries to generate vectors of ROI-specific activity estimates. This ROI serves as our seed region in the subsequent whole-brain connectivity analysis. After generating event-by-event activity estimates (whole brain as well as for our functionally-defined ROI), we further processed the estimates as follows. Separately for each participant and each functional run (and within voxel, where applicable), the vectors of estimates for each class of event (that is, II | II, II | RB, etc.) were demeaned relative to the mean for that class within that functional run. Then, the demeaned vectors of estimates were aggregated across runs. Note that this procedure is conservative, insofar as it prevents variations across runs, which may be caused by true signal, from driving connectivity.

After aggregating across runs, the vector of estimates from our functional ROI for each event was correlated with the vector of estimates from the same event in every voxel in the whole brain using a Spearman correlation. This yields four correlation maps per participant: one each for II | II, II | RB, RB | II, and RB | RB. Finally, each of these maps was normalized by subtracting the weighted average correlation in the functional ROI (weighted according to the conjunction statistic values) from every voxel. This step ensures that trivial differences, for example in sample size or overall noise level, do not drive the results. All reported results are qualitatively unchanged if this normalizing step is skipped.

These normalized connectivity maps were combined across participants into two contrast SPMs: II | RB versus II | II, and RB | II versus RB | RB. In particular, for each participant we created a difference image for each of these pairs between the raw Spearman correlation values (that is, II | RB-II | II and RB | II-RB | RB) and carried out a Wilcoxon signed-rank test across participants in each voxel; the resulting statistic was converted to an equivalent z statistic, and the resulting maps were cluster thresholded (voxelwise z threshold=1.96; cluster p threshold=0.05). Note that the raw correlation value, rather than a z-transformed version thereof, was used to prevent differences in degrees of freedom from driving the results; the Wilcoxon signed-rank test was used to accommodate the non-normality inherent in correlation values.

2.6.4. *Difficulty analyses*

The goal of this research was to study switching between procedural and declarative systems. However, there is one major alternative explanation which may confound our interpretation: healthy humans find the RB categorization task used here to be much easier than the II task[1]. It is therefore possible that our results reflect switching between “easy” and “difficult” tasks. Although we know of no theories that bear on issues of task difficulty and predict the patterns observed here, we nonetheless feel it is important to address this potential confound.

To begin, we note that according to this counter hypothesis, our labels of “II” and “RB” are really just synonyms for “difficult” and “easy” respectively. Luckily, the paradigm we employed has another built-in “easy”/“difficult” distinction that is orthogonal to the II/RB distinction: distance to the categorization decision bound. Previous research has confirmed the theoretical prediction that classification decisions for stimuli closer to the category boundary are more difficult than for those stimuli far from the boundary (e.g., Spiering and Ashby, 2008). By doing a within-distribution median split of stimuli on the basis of distance-to-bound, separately for each functional run, we classified every pair of successive categorization stimuli as near | far, near | near, far | far, and far | near. Note that this classification scheme is completely orthogonal to the original classification based on the II/RB distinction. As a result, the two analyses are logically unrelated. In other words, any cross-event pattern in activity observed in one analysis can exist without conflicting with any pattern observed for the other analysis.

With these new labels based on distance-to-bound, we reran the above GLM analysis, substituting “near” in place of “II” (because both are, according to the difficulty hypothesis, reducible to “difficult”), and

[1] This difference is in sharp contrast to pigeons, who learn both category structures equally well and at exactly the same rate (Smith et al., 2011). These pigeon data show that there is no inherent complexity difference between the tasks, and that the human difference must therefore be due to the fact that humans learn the two tasks in qualitatively different ways.

"far" in place of "RB." So long as the difficulty (i.e., accuracy) difference observed for II versus RB replicates for near versus far, the difficulty hypothesis must predict that results on the basis of the near/far distinction have the same form as those based on the II/RB distinction, although the magnitudes may differ due to differences in power, separability, etc. Therefore, rather than focus on the whole-brain results for each of these distance-to-bound analogs of the previous analyses, we will restrict our discussion of the distance-to-bound results to the question of whether the key patterns observed in our main analyses are duplicated in this distance-to-bound analysis.

3. Results

We present our results in four parts. We focus first on the behavioral results, which support our contention that participants are in fact switching between systems on a trial-by-trial basis. Second, we describe results of a standard GLM analysis that was designed to isolate regions where activity differs between switch and stay trials—and in particular, where activity was modulated differentially across tasks when switching to one system versus to the other. Third, we present the results of an analysis based on the alternative hypothesis that our GLM results are driven by differences in task difficulty, rather than system identity. Finally, we describe the results of a functional connectivity analysis designed to identify the networks associated with a key region of interest (ROI) discovered in the GLM analysis—again, with a particular focus on networks in which connectivity with this key ROI differs between switch and stay trials.

3.1. Behavioral results

Due to our pre-screening procedure, our scanning participants generally performed well on the task, despite its difficulty. Mean accuracies on each of the key trial types were as follows (with standard deviations in parentheses): II | II: 83.7% (3.9%); II | RB: 82.7% (6.1%); RB | II: 93.7% (2.6%); RB | RB: 95.9% (3.0%). Accuracy was significantly higher on RB trials than on II trials, both on switch trials (difference =11.0%; $t(20) = 7.39$, $p < .001$) and stay trials (difference=12.1%; $t(20) = 17.27$, $p < .001$). Following standard methods, we define the switch cost as the accuracy on stay trials minus the accuracy on switch trials. So, for example, the II switch cost equals the mean accuracy on II | II trials minus the mean accuracy on II | RB trials. Switch cost was not significant on II trials (1.0%, $t(20) = 1.00$, $p > .3$), but was significant on RB trials (2.2%, $t(20) = 3.03$, $p < .01$).

Distance-to-bound results. According to our alternative labeling scheme based on distance-to-bound, we observed a similar accuracy difference to that just described. "Near" accuracy =83.4%, "far" accuracy =93.6%; this difference is significant, $t(20) = 11.48$ ($p < .001$).

3.2. fMRI results

3.2.1. GLM results

The results of our GLM analysis are presented in Figs. 3–4 and Table 2. The regions associated with greater activity on II stay trials relative to II switch trials include L cerebellum (principally Crus I), L inferotemporal cortex, bilateral BA10 (spanning the anteriormost aspects of middle and superior frontal gyri), and R BA44/BA9 (spanning inferior frontal gyrus pars opercularis and pars triangularis, and middle frontal gyrus). Conversely, RB switch trials were associated with greater activity than RB stay trials in L cerebellum, L inferotemporal cortex, and R lateral posterior parietal areas (including midline superior lateral occipital cortex/precuneus, superior parietal lobule, and supramarginal gyrus). There were no significant clusters in either of the complementary contrasts (i.e., II | RB > II | II or RB | RB > RB | II).

Table 2
Results for GLM analyses. All coordinates are in mm in MNI space. Volume is in mm^3.

Contrast	Region	Centroid (*x*,*y*,*z*)			Peak *z*	Volume
II ∣II >II ∣RB	R BA10	39	37	11	4.00	16824
	L Crus I/II	−19	−77	−29	3.93	14152
	L BA10	−38	48	−2	3.58	4392
RB ∣II >RB ∣RB	L VI/Crus I	−28	−65	−31	3.57	5160
	b/l precuneus/sLOC	−6	−66	50	3.29	4648
	L SPL/SMG	−37	−46	45	3.51	3984

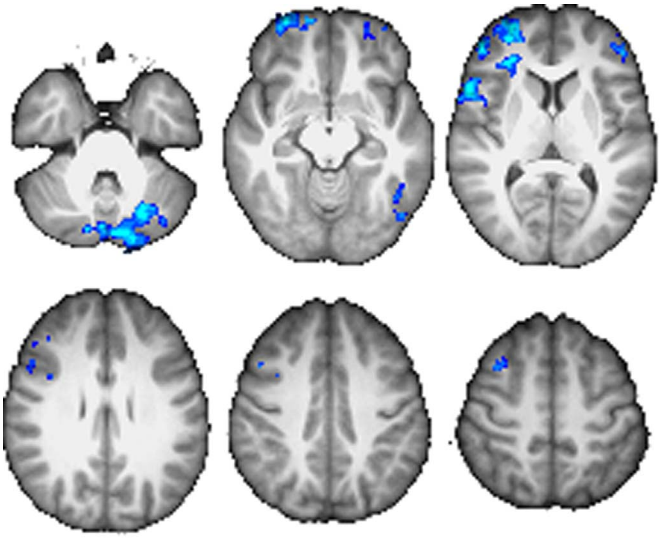

Fig. 3. GLM results for II | RB > II | II contrast. Throughout the manuscript, coordinates are in MNI space, images are displayed according to radiologic convention, and warm colors denote positive *z* values while cool colors denote negative *z* values. Unless otherwise noted, *z*-axis slice coordinates correspond to {−34, −12, 8, 28, 40, 52} from top-left to bottom-right. Color scale peak at $z = \pm 3.5$.

In an effort to identify regions whose activity is modulated on switch trials across both tasks, we carried out an additional conjunction analysis (Nichols et al., 2005) across the maps shown in Figs. 3–4. The results are shown in Fig. 5, and reveal a single cluster in left Crus I of the cerebellum. Examining the event-specific parameter estimates within the resulting ROI (taking a weighted average across all voxels in the ROI based on the conjunction statistic map) reveals a straightforward pattern of event-specific activity, as shown in Fig. 6. As expected (based on the way in which this cluster was defined), the

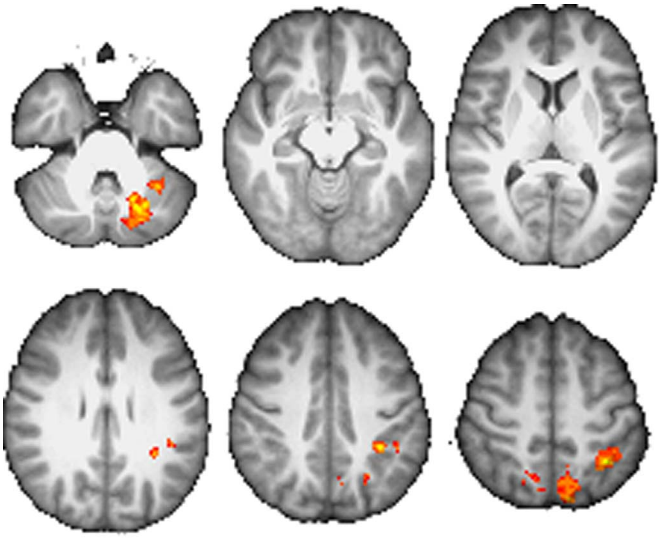

Fig. 4. GLM results for RB | II > RB | RB contrast. Color scale peak at $z = \pm 3.5$.

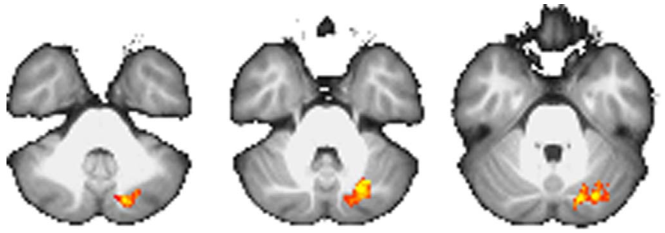

Fig. 5. Results from conjunction analysis of Figs. 3–4. z-axis slices at $\{-38, -34, -30\}$. Color scale peak at $z = \pm 3.3$.

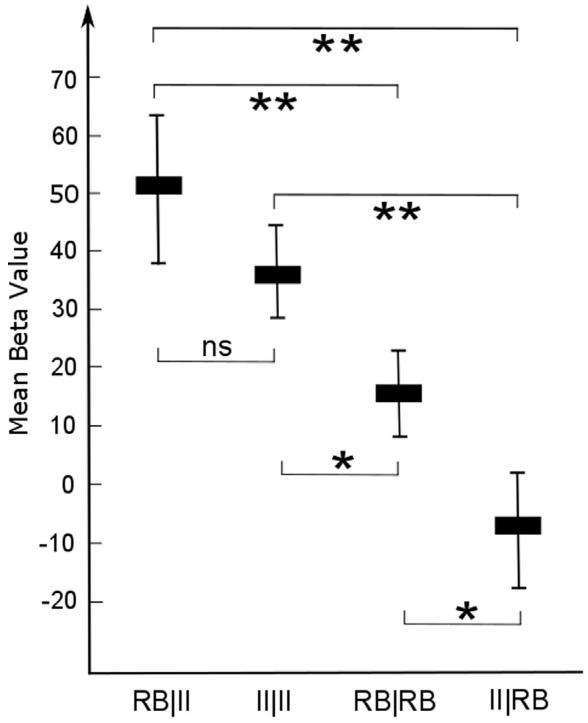

Fig. 6. Trial-type-specific beta values from the cluster shown in Fig. 5. Error bars show ± 1 standard error of the mean.

weighted average activity within this cluster is significantly lower on II | RB trials than on II | II trials [$t(20) = -4.77$, $p < .001$], and significantly higher on RB | II trials than on RB | RB trials [$t(20) = 3.70$, $p < .01$]. Comparing across systems, we see that average activity on RB | II trials is significantly greater than on II | RB trials [$t(20) = 3.68$, $p < .01$], while activity on II | II trials is significantly greater than on RB | RB trials [$t(20) = 2.76$, $p < .05$]. Finally, activity on RB | RB trials is significantly greater than on II | RB trials [$t(20) = 2.09$, $p < .05$], while activity does not differ significantly between II | II and RB | II trials [$t(20) = 1.23$, $p > .2$].

These post-hoc tests are important in understanding the role this region plays during different events. In particular, the observed pattern of results suggests that this region is moderately active on II | II and RB | RB trials (though significantly more active on the former), and that this activity increases on RB | II trials and decreases on II | RB trials.

Distance-to-bound results. The analogous result to Fig. 6 from the difficulty control analysis is described here. Recall that according to the difficulty hypothesis, "II" ≡"near" because both are difficult, and likewise "RB" ≡"far" because both are easy. With that in mind, here are the means (standard errors) in our Crus I ROI from this analysis: far | near =-13.72 (11.07); near | near =3.95 (6.06); far | far =-1.38 (7.12); near | far =24.69 (7.72). Of the planned *post hoc* comparisons, three were significant (two-tailed, uncorrected for multiple comparisons)—near | far was greater than each of the other three: far | near [$t(20) = 2.68$, $p < .05$]; near | near [$t(20) = 2.89$, $p < .01$]; and far | far [$t(20) = 2.85$, $p < .01$].

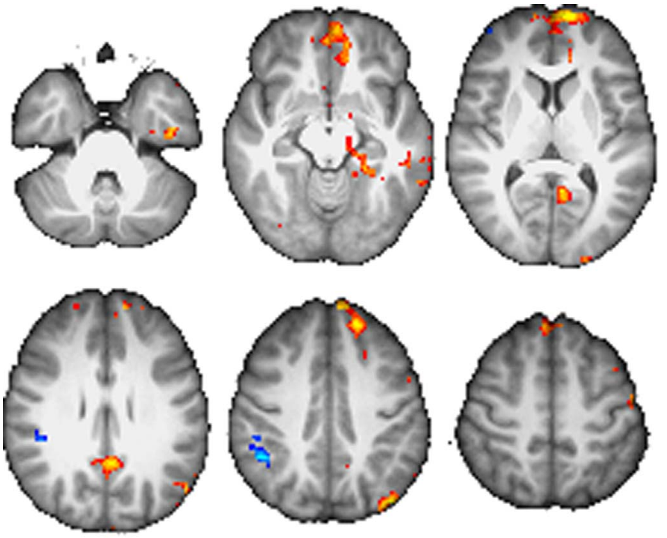

Fig. 7. Results of connectivity analysis with ROI shown in Fig. 5, showing regions of greater connectivity for II | RB than II | II (warm) and vice versa (cool). Color scale peak at $z = \pm 3.5$.

3.2.2. Connectivity results

Having identified a region whose pattern of activity matches the signature of a bidirectionally modulated system-switching region, we next sought to identify what networks this region was associated with during different event types. In particular, the maps presented in Figs. 7–8 and Table 3 show how whole-brain connectivity with our cerebellar ROI differed between II | RB and II | II trials, and between RB | II and RB | RB trials.

Regions with greater connectivity with our cerebellar ROI during II | II trials compared to II | RB trials include right ventrolateral PFC and right posterior parietal areas spanning supramarginal and angular gyri. Likewise, we see greater connectivity in left cerebellum, left insula, and left precentral gyrus during RB | RB trials compared to RB | II trials.

Consistent with this region's proposed central role in switching, more extensive networks are revealed when looking at the converses of those two contrasts: for II | RB > II | II, we observe clusters in medial prefrontal (mPFC), posterior cingulate (PCC), left temporal pole, left lateral prefrontal and posterior parietal, and right cerebellar areas. For RB | II > RB | RB, we again observe clusters in medial prefrontal and

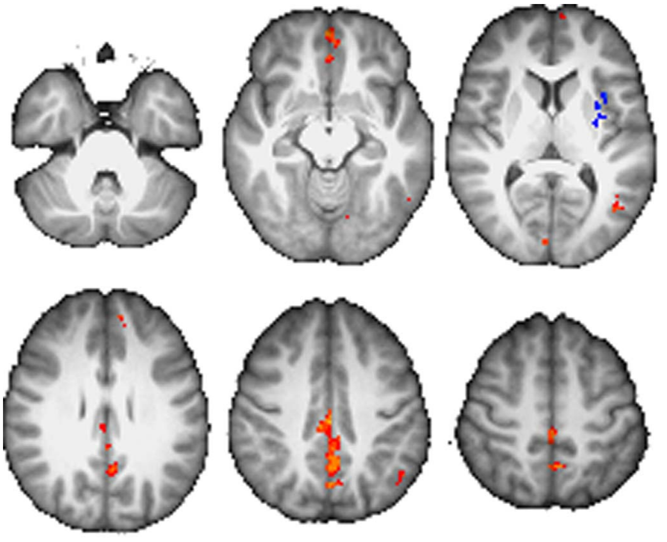

Fig. 8. As in Fig. 7, for contrast RB | II > RB | RB. Color scale peak at $z = \pm 3.5$.

Table 3
Results for functional connectivity analyses. All coordinates are in mm in MNI space. Volume is in mm^3.

Contrast	Region	Centroid (*x*,*y*,*z*)			Peak *z*	Volume
II∣RB > II∣II	R BA10	39	54	2	3.53	6016
	R SPL/Parietal Operculum	43	−36	30	3.11	2804
II∣II > II∣RB	b/l mPFC	−2	51	12	3.53	25294
	b/l PCC/Precuneus	−4	−54	15	3.46	7623
	b/l BA18	−3	−100	5	3.67	6930
	R Crus II	30	−87	−39	3.32	4756
	R BA20/21	68	−15	−19	3.04	3465
RB∣II > RB∣RB	b/l posteromedial occipital/parietal	0	−49	34	3.53	20853
	b/l vmPFC	1	42	−7	3.56	9009
	L LOC	−46	−71	12	3.67	8284
	L dmPFC	−8	45	30	3.08	2520
RB∣RB > RB∣II	L VIIb/Crus I/II	−27	−77	−53	3.39	8127
	R VIIb/Crus II	10	−79	−48	3.18	5418
	R occipital pole	15	−99	−18	3.84	4725

posterior cingulate areas, along with left occipitoparietal, ventral midline occipital, and bilateral somatomotor areas.

Of particular interest is the obvious overlap between these latter two contrasts: posterior cingulate and medial prefrontal cortex are canonical regions in the default mode network (DMN). In order to investigate this phenomenon further, we used an *a priori* DMN mask (Laird et al., 2009) to isolate the intersection of the positive clusters in Figs. 7 and 8 that fell within the DMN, and examined the weighted average connectivity in the resulting clusters across trial types. The results of this follow-up testing reveal that connectivity between both of these DMN regions and left Crus I is higher during switch trials (PCC: II | RB r=0.306, RB | II r=0.284; mPFC: II | RB r=0.275, RB | II r=0.262), and lower during stay trials (PCC: II | II r=0.231, RB | RB r=0.229; mPFC: II | II r=0.199, RB | RB r=0.199).

4. Discussion

We performed fMRI during a categorization task in which participants switched on a trial-by-trial basis between category structures that have been shown previously to recruit either procedural or declarative memory systems. We identified a key region of the cerebellum known to be interconnected with the PFC (left Crus I) that was maximally activated when switching to the declarative system, was deactivated when switching to the procedural system, and was intermediately activated when no system switch was required. This result represents the first reported evidence of cerebellar involvement in category learning, and ties our work to a literature refining our understanding of the cerebellum in non-motor cognition. By examining task-modulated connectivity with this region, we also identified several regions traditionally associated with the DMN that are selectively more strongly connected with this cerebellar ROI during system switching, along with several other regions that are connected in a more ad hoc fashion.

4.1. Task difficulty effects

To address the possibility that our results were driven by changes in task difficulty, rather than by system switching, we conducted a follow-up analysis using an alternative labeling scheme in which ‘difficult’ and ‘easy’ trials were defined according to whether the categorization stimulus was near or far from the optimal categorization boundary. This bifurcation of the data produced a similar behavioral difficulty difference as dividing the data according to whether the stimulus was from the RB or II categories. When difficulty is defined by distance-to-bound, then ‘difficult’ and ‘easy’ trials each include equal numbers of RB and II trials, so this alternative labeling scheme is orthogonal to the labeling scheme used in our system-switching analysis.

The distance-to-bound difficulty-based GLM analysis produced a pattern of Crus I activity that was almost exactly opposite to the result from our original analysis. In particular, in our original analysis, Crus I activity followed the pattern (substituting ‘difficult’ for ‘II’ and ‘easy’ for ‘RB’ to ease cross-analysis comparisons): easy | difficult > difficult | difficult > easy | easy > difficult | easy. In the distance-to-bound analysis, the pattern was instead: easy | difficult < difficult | difficult > easy | easy < difficult | easy. Moreover, this numerical near-reversal was borne out in the statistical comparisons of the two end points of each pattern: RB | II > II | RB, $t(20) = 3.68$, $p < .01$, while far | near < near | far, $t(20) = 2.68$, $p < .05$. Because our near/far distinction produced an almost identical accuracy difference to the II/RB distinction, this pattern reversal strongly contradicts a difficulty-based account of our primary GLM results.

There are also other reasons to suspect that difficulty is not driving our results. First, we found no significant correlation between the between-task differences in individual participants’ accuracies (a plausible index of subjective difficulty) and the magnitude of the differential BOLD response between tasks in the cerebellar region that we identified. Second, the pattern of BOLD responses on the various trial types obeyed the ordinal relationship RB | II > II | II > RB | RB > II | RB, a pattern that is inconsistent with task difficulty (ordering trial types by difficulty should produce II | RB > II | II > RB | II > RB | RB). Third, the hypothesis that RB versus II behavioral differences are driven by task difficulty has been tested and rejected many times before (e.g., Ashby and Maddox, 2005, 2010). Nevertheless, despite all this indirect evidence that task difficulty was not driving our results, our experiment was not designed to compare task difficulty and system switching directly, and as a result, more research is needed to understand fully the role that task difficulty might be playing in our overall results.

4.2. Role of the cerebellum

Our finding that the associative cerebellum facilitates system switching is in line with previous work seeking to understand the role of the cerebellum in non-motor cognition (for excellent reviews, see Koziol et al., 2014, Baumann et al., 2015). Though many theories of cerebellar function have been proposed, no singular grand unifying theory of the cerebellum has yet been embraced. Of the proposed theories, three seem applicable to the current findings:

1. *Contextual tuning of sensory acquisition* (Bower, 1997). This view roughly holds that the activation of the cerebellum will reflect the need for sensory vigilance, and consequently, predicts that cerebellar activation may be modulated by task difficulty (e.g., because more difficult tasks will typically require more sensory vigilance). This idea is appealing given our current data, since it is intuitive to suppose that more sensory vigilance is required for the II categories (which require attention to two stimulus dimensions), than for the RB categories (which require attention only to one stimulus dimension). Moreover, Fig. 5 does show a task difficulty effect in that II | II trials evoked a greater cerebellar response than RB | RB trials. However, as noted in the previous section, task difficulty does a poor job of accounting for our results, and it is unclear how this theory would account for the observed bidirectional modulation on switch trials, especially since RB | II trials evoked the greatest cerebellar response (i.e., and RB trials should require the least sensory vigilance).
2. *Bayesian state estimation* (Paulin, 2005). This theory posits that the cerebellum evolved to approximate Bayesian posterior distributions of prey locations and that these origins endow the cerebellum with general Bayesian state estimation abilities. This is a very general

theory that could be applied to almost any domain, depending on how one defines 'state'. In the context of our current data, the states being estimated could be the attentional demands of the stimuli (bar thickness for the RB categories, and both bar thickness and bar angle for II categories). As with the contextual-tuning-of-sensory-acquisition hypothesis, this theory also struggles to account for the bidirectional profile we observed on switch trials.

3. *Transitions from controlled to automatic behavior* (Balsters and Ramnani, 2011). This idea is that the cerebellum is critical for transitions from 'controlled' to 'automatic' responding. This hypothesis is consistent with our current results, but only if one assumes that declarative- and procedural-mediated behaviors can be classified as controlled and automatic, respectively. Declarative-mediated behaviors meet standard controlled criteria, but procedural-mediated behaviors meet only some of the standard automaticity criteria (Ashby and Crossley, 2012; Ashby et al., 2010). For example, unlike automatic behaviors, initial II performance depends strongly on the presence of immediate feedback. Thus, one interpretation is that our results generalize the theory that the cerebellum facilitates the transition from controlled to automatic responding by suggesting that it also facilitates the transition from declarative- to procedural-mediated behaviors.

In summary, none of these theories capture the entire pattern of results. Moreover, it is obvious that they are themselves in disagreement. Considering all the results we reported—the system switching-related activity, along with the system-nonspecific connectivity between Crus I and canonical DMN regions, and the distance-to-bound results showing the highest activity for near | far trials and the lowest for far | near trials—we can speculatively put forth an alternative theory of Crus I function. In this view, Crus I plays a key role in boosting or suppressing prefrontal, cognitive control-supporting regions. Because these regions overlap, or are interconnected with, the regions that underlie declarative categorization, Crus I activity tracks the exertion (on RB switch trials) or release (on II switch trials) of prefrontal control over categorization performance. Similarly, difficult (i.e., near-to-bound) trials might require greater cognitive effort, in the form of increased activity in these prefrontal areas, whereas easy (i.e., far-from-bound) trials allow for a relaxing of that effort. This hypothesis appears to bridge several existing theories, and also accounts for all of the patterns of activity we observed in Crus I in the present study. It also may begin to explain some of the other activity and connectivity results we observed—for example, the deactivation and simultaneous increased Crus I connectivity of a region that has been previously implicated in cognitive control, BA10 (Roca et al., 2011). Of course, much more work is required to test the validity of this hypothesis.

Within the category-learning literature, the cerebellum has not featured prominently, though a few investigations have asked whether or not patients with cerebellar damage or degeneration are impaired in category-learning tasks (Ell and Ivry, 2008; Maddox et al., 2005; Witt et al., 2002). These studies all reported essentially normal performance in cerebellar patients, although it is important to note that none of the tasks used in these studies placed heavy demands on system switching—at most, an ideal participant in an II task is hypothesized to make a single switch, from an initial suboptimal declarative strategy to the more-optimal procedural strategy. Cerebellar activation has also been reported in a few neuroimaging studies of category learning (Seger and Miller, 2010). Even so, the cerebellar area identified in those studies was in the anterior lobe (the regions connected with motor cortex), not the posterior lobe (Crus I and II; regions connected with the PFC) as found here. Furthermore, none of these studies attempted to assign any functional role to the cerebellum during category learning or categorization performance. Note, however, that many studies using standard field of view parameters may not have included the full cerebellum, so earlier null results, particularly in the more posterior aspects of the cerebellum, may be due to its not having been scanned at all.

The cerebellum has long been associated with various forms of procedural learning (Gomez-Beldarrain et al., 1998; Shin and Ivry, 2003; Torriero et al., 2004), and also with motor control (Kawato and Wolpert, 1998; Wolpert and Kawato, 1998; Wolpert et al., 1998). Because of this latter association, the failure of previous studies to find a role for the cerebellum in category learning might be expected (i.e., because motor learning demands are typically minimal in category-learning tasks). Even so, the cerebellum is interconnected with all major cortical nodes associated with successful II and RB learning. For example, evidence suggests that the critical cortical node for RB learning is within the PFC, and the critical cortical node for II learning is in supplementary motor area (or dorsal premotor cortex). Both of these are interconnected with the cerebellum. The cerebellum is organized into functionally segregated multi-synaptic loops between the cerebellar and cerebral cortices, with the anterior lobes of the cerebellum interconnected with the motor cortex, and the posterior lobes interconnected with the PFC (Barbas et al., 1991; Goldman-Rakic and Porrino, 1985; Kelly and Strick, 2003**?**; Middleton and Strick, 2001). Thus the cerebellum is well situated to interact with the cortical substrates of both procedural and declarative learning.

A similar story holds true at the subcortical level. Procedural learning in II tasks is thought to be implemented via dopamine-dependent reinforcement learning at synapses between cortex and dorsal-lateral striatal medium-spiny neurons. Declarative learning, on the other hand, is thought to depend on interactions between PFC and dorsal-medial striatum (e.g., head of caudate nucleus) for rule selection and switching. The output from cerebellar deep nuclei is disynaptically relayed to the striatum via the intralaminar nuclei of the thalamus, that is, the centromedian- parafascicular nuclei (CM-Pf; Hoshi et al., 2005). Thus, the cerebellum is also well situated to interact with the subcortical substrates of both procedural and declarative learning.

In light of the widespread anatomical connectivity between cerebellum and regions associated with categorization, there are at least two plausible (and novel) explanations for the role that Crus I could play in passing control between the procedural and declarative systems: the first involves the thalamic pathway from cerebellum, while the second focuses on cerebellum's connections with subthalamus. We discuss each of these in turn below.

The subcortical path from cerebellum to thalamus presents one possible mechanism by which cerebellum may be influencing the passing of control between systems, given that the regions of the thalamus most innervated by cerebellum afferents are the intralaminar nuclei including CM-Pf (Hoshi et al., 2005). These regions of the thalamus are known to project heavily to striatal cholinergic interneurons (called Tonically Active Neurons or TANs). There are essentially two current theories of CM-Pf–TAN projections. One assigns this pathway a foundational role in attention and arousal regulation (Kimura et al., 2004). Another suggests a much finer-grained role in context recognition and behavioral switching among contexts (Ashby and Crossley, 2011; Bradfield et al., 2013; Crossley et al., 2014). This latter view resonates well with our current behavioral paradigm, as well as with a body of evidence demonstrating the sensitivity of TANs to contextual features (Apicella, 2007; Shimo and Hikosaka, 2001; Yamada et al., 2004). Furthermore, the current literature suggesting such a contextual role for the CM-Pf–TAN pathway is dominated by tasks focusing on behavioral flexibility (place learning, n-arm maze navigation, etc.). These tasks are mediated by largely the same cortical and subcortical associative territories that are interconnected via the cerebellar ROI (Crus I) found here.

A second possible mechanism via which cerebellum may guide or facilitate system transitions involves the subthalamic nucleus. We have previously speculated that system switching may be mediated via the hyperdirect pathway of the basal ganglia (Ashby and Crossley, 2010; Crossley and Ashby, 2015; Crossley and Roeder,). The hyperdirect pathway begins with direct excitatory glutamate projections from frontal cortex to the subthalamic nucleus. The subthalamic nucleus then sends excitatory glutamate projections directly to the internal segment of the globus pallidus (Joel and Weiner, 1997; Parent and Hazrati, 1995a, 1995b). This extra excitatory input to the globus pallidus tends to offset inhibitory input from the striatum, making it more difficult for striatal activity to affect cortex. Hence, the hyperdirect pathway is hypothesized to control whether striatal outputs influence cortex: when subthalamic activity is reduced, striatal output can pass through, whereas when subthalamic activity is increased, striatal output is prevented from influencing cortex.

Our imaging data did not provide direct evidence for hyperdirect pathway involvement, which is unsurprising given the well-documented difficulties involved in imaging the STN (de Hollander et al., 2015). In general, special imaging protocols are required to reliably observe STN activation (e.g., Aron and Poldrack, 2006). One interesting possibility, however, is that the cerebellar activation we report may reflect a driving influence of the STN, which recently has been found to project to the pontine nuclei, and from there onward to the cerebellar cortex (Bostan et al., 2010, 2013; Bostan and Strick, 2010; Hoshi et al., 2005). In other words, the activity we observed in Crus I may in effect be a distal indicator of STN activity, since it showed precisely the activity profile predicted by the hyperdirect pathway hypothesis.

4.3. Role of the default mode network

Although left Crus I emerged as a critical region in our investigation, there were a number of other brain regions that showed either switching-related activity (although not bidirectionally, as in the cerebellum), or else were differentially connected with Crus I across trial types. The most interesting of these other regions were the PCC and mPFC, two areas strongly associated with the DMN. Our results show that these regions are more strongly connected with cerebellum on switch trials (regardless of direction of switch) than on stay trials. There are at least two possibilities (which are not mutually exclusive) that explain this pattern. The first possibility is somewhat trivial: the DMN tends to be un- (or anti-) correlated with task-positive areas (though note that we observed positive correlations between cerebellum and these DMN regions across all trial types). This decreased correlation may be especially pronounced when participants are disengaged from the task, as they might be to a greater degree during the relatively less cognitively demanding stay trials.

However, recent work has suggested a second, far more intriguing possibility. The DMN has been suggested to have high controllability, such that it is well-situated to facilitate certain types of transitions between brain states (Gu et al., 2015). This complements research demonstrating that the DMN coordinates with parts of the cognitive control network (Leech et al., 2011), particularly during tasks that require complex cognition and divergent thinking (Beaty et al., 2015). It is therefore possible that the DMN regions identified in our analyses, in part, reflect or directly aid the disruption of the network that controlled behavior on the previous trial, and thereby facilitate the transfer of control from one system to the other.

4.4. System switching versus task switching

There is a rich literature devoted to identifying the regions involved in task switching, with most researchers reporting frontal and parietal regions as important nodes (e.g., see Ruge et al., 2013, for a review). The precise anatomical location of these nodes varies considerably across studies, and seems to depend on a variety of factors, including: stimulus complexity (Witt and Stevens, 2013); the degree of abstractness of the response rule (Kim et al., 2011); and whether switching entails changes to the rules governing correct responding, changes to the effectors used to emit a response (Philipp et al., 2013; Stelzel et al., 2011), changes to spatial attention (Chiu and Yantis, 2009), or changes to perceptual demands (Ravizza and Carter, 2008; Nagahama et al., 2001).

At a glance, our results are consistent with this literature, in that we also identified frontal and parietal regions. However, our major novel result is the bidirectional pattern of activity in Crus I of the cerebellum, and its connectivity with the DMN. Few of these studies reported cerebellar activity, and no study (to our knowledge) emphasized a cerebellar result. We propose that this may be because nearly all of this earlier work examined switching between two tasks that both relied on declarative memory systems[2]. These previous studies are therefore largely studies of 'within-system' task switching, in contrast to our experiment, which examined switching between memory systems.

4.5. Caveats and limitations

Another alternative to a system-switching account is that our results reflect switching between tasks that require attention to different stimulus dimensions. The RB task only requires attention to spatial frequency, whereas the II task requires attention to both spatial frequency and orientation. In line with this possibility, Le et al. (1998) reported bilateral cerebellar activation similar to our own in a task-switching paradigm that required participants to switch attention between stimulus shape and color dimensions. However, their cerebellar activation was ipsilateral to the finger used to make a response, raising the possibility that it was linked to motor preparation. Furthermore, they collapsed across switch types, and therefore the region they identified was uniformly more active during attentional switches, while we observed bidirectional modulation of activation. For these reasons, it seems unlikely that the cerebellar activation we observed was driven by attentional switching.

A second concern has to do with the possibility that our results reflect the development of automaticity. Prior work from our lab has shown that although performance early in training is controlled by the expected networks for II and RB tasks, control gradually migrates to a common cortical circuit in both tasks (Hélie et al., 2010, 2010; Soto et al., 2013). This earlier work gave an approximate time-course for the amount of training required before this cortical circuit took over, which was roughly after 13 training sessions of approximately 600 trials each, or in other words after approximately 7800 training trials (Hélie et al., 2010). In the present experiment, each participant completed a total of 525 RB trials and 775 II trials. So each participant received less than 10% of the training that previous work showed was required for automaticity to develop in our RB and II tasks. As a result, it seems highly unlikely that any of our results reflect automatic responding.

One final caveat relates to our pre-screening procedure. Logically, we needed to select participants able to perform the task in order to study task performance. We expect that our results implicating Crus I in system switching would generalize to other system-switching contexts, even in those participants who failed to meet our pre-screening criteria. For instance, an individual who is carrying on a conversation while driving might be switching between explicit (engaging in conversation) and procedural (driving) processes. Even if this person would be unable to perform our difficult task, the cerebellum may nonetheless be responsible for such switching. However, because our design precludes scanning such individuals, it is an unanswered question whether the switching mechanisms identified here are universal.

4.6. Conclusion

Categorization is among the most important cognitive skills that humans possess. It allows us to navigate in a dangerous world, and to find food, shelter, and friends. There is now substantial evidence that

[2] However, see Jimura et al. (2014) for one counterexample, in that case switching between a novel skill and a highly overtrained one, which has previously been proposed to engage distinct systems (Ashby et al., 2010).

humans have multiple category-learning systems, which are largely neuroanatomically separate, learn by qualitatively different rules, and have adapted to learning different types of category structures. This article describes the results of one of the first studies to examine the natural next question of how these various systems interact. This is an important problem because during daily life we must often switch between different categorization systems (e.g., declarative and procedural). For example, many components of driving are procedural, but at the same time some explicit decisions are required. Following an explicit decision, a failure to quickly switch back to a procedural strategy could greatly increase the risk of an accident. Toward this end, we identified a key region of the cerebellum (left Crus I) that, within our sample of task-proficient participants, was maximally activated when switching to the declarative system, was minimally activated when switching to the procedural system, and was intermediately activated when no system switch was required. We propose that this region, perhaps in conjunction with the DMN, facilitates the successful transfer of control between procedural and declarative systems.

Acknowledgements

This work was partially supported by U.S. Army Research Office through the Institute for Collaborative Biotechnologies under Grant W911NF-07-1-0072 and by grant 2R01MH063760 from the National Institutes of Health. Thanks to Dani Bassett for her helpful feedback.

References

Apicella, P., 2007. Leading tonically active neurons of the striatum from reward detection to context recognition. Trends Neurosci. 30, 299–306.

Aron, A.R., Poldrack, R.A., 2006. Cortical and subcortical contributions to stop signal response inhibition: role of the subthalamic nucleus. J. Neurosci. 26, 2424–2433.

Ashby, F.G., Alfonso-Reese, L.A., Turken, A.U., Waldron, E.M., 1998. A neuropsychological theory of multiple systems in category learning. Psychol. Rev. 105, 442–481.

Ashby, F.G., Crossley, M.J., 2010. Interactions between declarative and procedural-learning categorization systems. Neurobiol. Learn. Mem. 94, 1–12.

Ashby, F.G., Crossley, M.J., 2011. A computational model of how cholinergic interneurons protect striatal-dependent learning. J. Cogn. Neurosci. 23, 1549–1566.

Ashby, F.G., Crossley, M.J., 2012. Automaticity and multiple memory systems. Wiley Interdiscip. Rev.: Cogn. Sci. 3, 363–376.

Ashby, F.G., Ennis, J.M., 2006. The role of the basal ganglia in category learning. Psychol. Learn. Motiv. 46, 1–36.

Ashby, F.G., Gott, R.E., 1988. Decision rules in the perception and categorization of multidimensional stimuli. J. Exp. Psychol.: Learn., Mem., Cogn. 14, 33–53.

Ashby, F.G., Maddox, W.T., 2005. Human category learning. Annu. Rev. Psychol. 56, 149–178.

Ashby, F.G., Maddox, W.T., 2010. Human category learning 2.0. Ann. New Y. Acad. Sci. 1224, 147–161.

Ashby, F.G., O'Brien, J.B., 2007. The effects of positive versus negative feedback on information-integration category learning. Percept. Psychophys. 69, 865–878.

Ashby, F.G., Queller, S., Berretty, P.M., 1999. On the dominance of unidimensional rules in unsupervised categorization. Percept. Psychophys. 61, 1178–1199.

Ashby, F.G., Turner, B.O., Horvitz, J.C., 2010. Cortical and basal ganglia contributions to habit learning and automaticity. Trends Cogn. Sci. 14, 208–215.

Ashby, F.G., Valentin, V.V., 2017a. The categorization experiment: Experimental design and data analysis. In: Wagenmakers, E.J., Wixted, J.T., (Eds.), Stevens' handbook of experimental psychology and cognitive neuroscience, Fourth Edition, Volume Five: Methodology (p. in press). New York: Wiley.

Ashby, F.G., Valentin, V.V., 2017b. Multiple systems of perceptual category learning: Theory and cognitive tests. In: Cohen, H., Lefebvre, C., (Eds.), Handbook of categorization in cognitive science, Second Edition (p. in press). New York: Elsevier.

Balsters, J.H., Ramnani, N., 2011. Cerebellar plasticity and the automation of first-order rules. J. Neurosci. 31, 2305–2312.

Barbas, H., Henion, T., Dermon, C., 1991. Diverse thalamic projections to the prefrontal cortex in the rhesus monkey. J. Comp. Neurol. 313, 65–94.

Baumann, O., Borra, R.J., Bower, J.M., Cullen, K.E., Habas, C., Ivry, R.B., Leggio, M., Mattingley, J.B., Molinari, M., Moulton, E.A., et al., 2015. Consensus paper: the role of the cerebellum in perceptual processes. Cerebellum 14, 197–220.

Beaty, R.E., Benedek, M., Silvia, P.J., and Schacter, D.L., 2015. Creative cognition and brain network dynamics. Trends in cognitive sciences.

Bostan, A.C., Dum, R.P., Strick, P.L., 2010. The basal ganglia communicate with the cerebellum. Proc. Natl. Acad. Sci. 107, 8452–8456.

Bostan, A.C., Dum, R.P., Strick, P.L., 2013. Cerebellar networks with the cerebral cortex and basal ganglia. Trends Cogn. Sci. 17, 241–254.

Bostan, A.C., Strick, P.L., 2010. The cerebellum and basal ganglia are interconnected. Neuropsychol. Rev. 20, 261–270.

Bower, J.M., 1997. Is the cerebellum sensory for motor's sake, or motor for sensory's sake: the view from the whiskers of a rat? Progress. Brain Res. 114, 463–496.

Bradfield, L.A., Bertran-Gonzalez, J., Chieng, B., Balleine, B.W., 2013. The thalamostriatal pathway and cholinergic control of goal-directed action: interlacing new with existing learning in the striatum. Neuron 79, 153–166.

Brainard, D.H., 1997. The psychophysics toolbox. Spat. Vision. 10, 433–436.

Brown, R.G., Marsden, C.D., 1988. Internal versus external cues and the control of attention in parkinson's disease. Brain 111, 323–345.

Cisler, J.M., Bush, K., Steele, J.S., 2014. A comparison of statistical methods for detecting context-modulated functional connectivity in fMRI. NeuroImage 84, 1042–1052.

Chiu, Y.-C., Yantis, S., 2009. A domain-independent source of cognitive control for task sets: shifting spatial attention and switching categorization rules. J. Neurosci. 29, 3930–3938.

Crossley, M.J., Ashby, F.G., 2015. Procedural learning during declarative control. J. Exp. Psychol.: Learn. Mem. Cogn. 41, 1388–1403.

Crossley, M.J., Ashby, F.G., Maddox, W.T., 2014. Context-dependent savings in procedural category learning. Brain Cogn. 92, 1–10.

Crossley, M.J., Roeder, J.L., Hélie, S., Ashby, F.G., 2017 (in press). Trial-by-trial switching between procedural and declarative categorization systems. Psychological Research.

Eichenbaum, H., Cohen, N.J., 2001. From Conditioning to Conscious Recollection: Memory Systems of the Brain. Oxford University Press.

Ell, S.W., Ivry, R.B., 2008. Cerebellar pathology does not impair performance on identification or categorization tasks. J. Int. Neuropsychol. Soc. 14, 760–770.

Erickson, M.A., Kruschke, J.K., 1998. Rules and exemplars in category learning. J. Exp. Psychol.: General. 127, 107–140.

Filoteo, J.V., Maddox, W.T., Salmon, D.P., Song, D.D., 2005. Information-integration category learning in patients with striatal dysfunction. Neuropsychology 19, 212–222.

Filoteo, J.V., Maddox, W.T., Song, D.D., et al., 2007. Characterizing rule-based category learning deficits in patients with parkinson's disease. Neuropsychologia 45, 305–320.

Fukunaga, K., 1990. Introduction to Statistical Pattern Recognition. Academic Press, New York.

Goldman-Rakic, P.S., Porrino, L.J., 1985. The primate mediodorsal (md) nucleus and its projection to the frontal lobe. J. Comp. Neurol. 242, 535–560.

Gomez-Beldarrain, M., Garcia-Monco, J., Rubio, B., Pascual-Leone, A., 1998. Effect of focal cerebellar lesions on procedural learning in the serial reaction time task. Exp. Brain Res. 120, 25–30.

Gu, S., Pasqualetti, F., Cieslak, M., Telesford, Q.K., Alfred, B.U., Kahn, A.E., Medaglia, J.D., Vettel, J.M., Miller, M.B., Grafton, S.T., Bassett, D.S., 2015. Controllability of structural brain networks. Nat. Commun. 6.

Hélie, S., Roeder, J.L., Ashby, F.G., 2010. Evidence for cortical automaticity in rule-based categorization. J. Neurosci. 30, 14225–14234.

Hélie, S., Waldschmidt, J.G., Ashby, F.G., 2010. Automaticity in rule-based and information-integration categorization. Atten., Perception Psychophys. 72, 1013–1031.

de Hollander, G., Keuken, M.C., Forstmann, B.U., 2015. The subcortical cocktail problem; mixed signals from the subthalamic nucleus and substantia nigra. PloS One 1, 18.

Hoshi, E., Tremblay, L., Féger, J., Carras, P.L., Strick, P.L., 2005. The cerebellum communicates with the basal ganglia. Nat. Neurosci. 8, 1491–1493.

Houk, J.C., Adams, J.L, Barto, A.G., 1995. A model of how the basal ganglia generate and use neural signals that predict reinforcement. In: Houk, J.C., Davis, J.L., Beiser, D.G. (Eds.), Models of Information Processing in the Basal Ganglia. MIT Press, Cambridge, MA, 249–270.

Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M., 2012. Fsl. Neuroimage 62, 782–790.

Jimura, K., Cazalis, F., Stover, E.R., Poldrack, R.A., 2014. The neural basis of task switching changes with skill acquisition. Front. Human. Neurosci., 8.

Joel, D., Niv, Y., Ruppin, E., 2002. Actor-critic models of the basal ganglia: new anatomical and computational perspectives. Neural Netw. 15, 535–547.

Joel, D., Weiner, I., 1997. The connections of the primate subthalamic nucleus: indirect pathways and the open-interconnected scheme of basal ganglia-thalamocortical circuitry. Brain Res. Rev. 23, 62–78.

Kawato, M., Wolpert, D., 1998. Intern. Model. Mot. control. Sens. Guid. Mov. 218, 291–307.

Kelly, R.M., Strick, P.L., 2003. Cerebellar loops with motor cortex and prefrontal cortex of a nonhuman primate. J. Neurosci. 23, 8432–8444.

Kim, C., Johnson, N.F., Cilles, S.E., Gold, B.T., 2011. Common and distinct mechanisms of cognitive flexibility in prefrontal cortex. J. Neurosci. 31, 4771–4779.

Kimura, M., Minamimoto, T., Matsumoto, N., Hori, Y., 2004. Monitoring and switching of cortico-basal ganglia loop functions by the thalamo-striatal system. Neurosci. Res. 48, 355–360.

Knowlton, B.J., Mangels, J.A., Squire, L.R., 1996. A neostriatal habit learning system in humans. Science 273, 1399–1402.

Koziol, L.F., Budding, D., Andreasen, N., D'Arrigo, S., Bulgheroni, S., Imamizu, H., Ito, M., Manto, M., Marvel, C., Parker, K., et al., 2014. Consensus paper: the cerebellum's role in movement and cognition. Cerebellum 13, 151–177.

Laird, A.R., Eickhoff, S.B., Li, K., Robin, D.A., Glahn, D.C., Fox, P.T., 2009. Investigating the functional heterogeneity of the default mode network using coordinate-based meta-analytic modeling. J. Neurosci. 29, 14496–14505.

Le, T.H., Pardo, J.V., Hu, X., 1998. 4 t-fmri study of nonspatial shifting of selective attention: Cerebellar and parietal contributions. J. Neurophysiol. 79, 1535–1548.

Leech, R., Kamourieh, S., Beckmann, C.F., Sharp, D.J., 2011. Fractionating the default mode network: distinct contributions of the ventral and dorsal posterior cingulate cortex to cognitive control. J. Neurosci. 31, 3217–3224.

Maddox, W.T., Aparicio, P., Marchant, N.L., Ivry, R.B., 2005. Rule-based category learning is impaired in patients with parkinson's disease but not in patients with cerebellar disorders. J. Cogn. Neurosci. 17, 707–723.

Maddox, W.T., Ashby, F.G., 1993. Comparing decision bound and exemplar models of categorization. Percept. Psychophys. 53, 49–70.

Middleton, F.A., Strick, P.L., 2000. Basal ganglia and cerebellar loops: motor and cognitive circuits. Brain Res. Rev. 31, 236–250.

Middleton, F.A., Strick, P.L., 2001. Cerebellar projections to the prefrontal cortex of the primate. J. Neurosci. 21, 700–712.

Miller, E.K., Cohen, J.D., 2001. An integrative theory of prefrontal cortex function. Annu. Rev. Neurosci. 24, 167–202.

Monsell, S., 2003. Task switching. Trends Cogn. Sci. 7, 134–140.

Muhammad, R., Wallis, J.D., Miller, E.K., 2006. A comparison of abstract rules in the prefrontal cortex, premotor cortex, inferior temporal cortex, and striatum. J. Cogn. Neurosci. 18, 974–989.

Mumford, J.A., Turner, B.O., Ashby, F.G., Poldrack, R.A., 2012. Deconvolving bold activation in event-related designs for multivoxel pattern classification analyses. NeuroImage 59, 2636–2643.

Nagahama, Y., Okada, T., Katsumi, Y., Hayashi, T., Yamauchi, H., Oyanagi, C., Konishi, J., Fukuyama, H., Shibasaki, H., 2001. Dissociable mechanisms of attentional control within the human prefrontal cortex. Cereb. Cortex 11, 85–92.

Nichols, T., Brett, M., Andersson, J., Wager, T., Poline, J.-B., 2005. Valid conjunction inference with the minimum statistic. Neuroimage 25, 653–660.

Nomura, E., Maddox, W., Filoteo, J., Ing, A., Gitelman, D., Parrish, T., Mesulam, M., Reber, P., 2007. Neural correlates of rule-based and information-integration visual category learning. Cereb. Cortex 17, 37–43.

Ollinger, J., Shulman, G.L., Corbetta, M., 2001. Separating processes within a trial in event-related functional mri: i. the method. Neuroimage 13, 210–217.

Parent, A., Hazrati, L.-N., 1995. Functional anatomy of the basal ganglia. i. the cortico-basal ganglia-thalamo-cortical loop. Brain Res. Rev. 20, 91–127.

Parent, A., Hazrati, L.-N., 1995. Functional anatomy of the basal ganglia. ii. the place of subthalamic nucleus and external pallidium in basal ganglia circuitry. Brain Res. Rev. 20, 128–154.

Paulin, M., 2005. Evolution of the cerebellum as a neuronal machine for bayesian state estimation. J. Neural Eng. 2, S219–S234.

Philipp, A.M., Weidner, R., Koch, I., Fink, G.R., 2013. Differential roles of inferior frontal and inferior parietal cortex in task switching: evidence from stimulus-categorization switching and response-modality switching. Human. Brain Mapp. 34, 1910–1920.

Ravizza, S.M., Carter, C.S., 2008. Shifting set about task switching: behavioral and neural evidence for distinct forms of cognitive flexibility. Neuropsychologia 46, 2924–2935.

Rissman, J., Gazzaley, A., D'Esposito, M., 2004. Measuring functional connectivity during distinct stages of a cognitive task. Neuroimage 23, 752–763.

Roca, M., Torralva, T., Gleichgerrcht, E., Woolgar, A., Thompson, R., Duncan, J., Manes, F., 2011. The role of area 10 (ba10) in human multitasking and in social cognition: a lesion study. Neuropsychologia 49, 3525–3531.

Ruge, H., Jamadar, S., Zimmermann, U., Karayanidis, F., 2013. The many faces of preparatory control in task switching: reviewing a decade of fmri research. Human. brain Mapp. 34, 12–35.

Schwarz, G., 1978. Estimating the dimension of a model. Ann. Stat. 6, 461–464.

Seger, C.A., Cincotta, C.M., 2006. Dynamics of frontal, striatal, and hippocampal systems during rule learning. Cereb. Cortex 16, 1546–1555.

Seger, C.A., Dennison, C.S., Lopez-Paniagua, D., Peterson, E.J., Roark, A.A., 2011. Dissociating hippocampal and basal ganglia contributions to category learning using stimulus novelty and subjective judgments. Neuroimage 55, 1739–1753.

Seger, C.A., Miller, E.K., 2010. Category learning in the brain. Annu. Rev. Neurosci. 33, 203.

Serences, J.T., 2004. A comparison of methods for characterizing the event-related bold timeseries in rapid fmri. Neuroimage 21, 1690–1700.

Shimo, Y., Hikosaka, O., 2001. Role of tonically active neurons in primate caudate in reward-oriented saccadic eye movement. J. Neurosci. 21, 7804–7814.

Shin, J., Ivry, R.B., 2003. Spatial and temporal sequence learning in patients with parkinson's disease or cerebellar lesions. J. Cogn. Neurosci. 15, 1232–1243.

Smith, J.D., Ashby, F.G., Berg, M.E., Murphy, M.S., Spiering, B., Cook, R.G., Grace, R.C., 2011. Pigeons' categorization may be exclusively nonanalytic. Psychon. Bull. Rev. 18, 414–421.

Soto, F.A., Waldschmidt, J.G., Hélie, S., and Ashby, F.G., 2013. Brain activity across the development of automatic categorization: A comparison of categorization tasks using multi-voxel pattern analysis. Neuroimage, 71, pp. 284–897. 10.1016/j.neuroimage.2013.01.008.

Spiering, B.J., Ashby, F.G., 2008. Initial training with difficult items facilitates information integration, but not rule-based category learning. Psychol. Sci. 19, 1169–1177.

Squire, L.R., 2004. Memory systems of the brain: A brief history and current perspective. Neurobiol. Learn. Mem. 82, 171–177.

Stelzel, C., Basten, U., Fiebach, C.J., 2011. Functional connectivity separates switching operations in the posterior lateral frontal cortex. J. Cogn. Neurosci. 23, 3529–3539.

Torriero, S., Oliveri, M., Koch, G., Caltagirone, C., Petrosini, L., 2004. Interference of left and right cerebellar rtms with procedural learning. J. Cogn. Neurosci. 16, 1605–1611.

Treutwein, B., Rentschler, I., Caelli, T., 1989. Perceptual spatial frequency–orientation surface: psychophysics and line element theory. Biol. Cybern. 60, 285–295.

Tulving, E., Craik, F.I., 2000. The Oxford Handbook of Memory. Oxford University Press.

Turner, B.O., Mumford, J.A., Poldrack, R.A., Ashby, F.G., 2012. Spatiotemporal activity estimation for multivoxel pattern analysis with rapid event-related designs. NeuroImage 62, 1429–1438.

Waldschmidt, J.G., Ashby, F.G., 2011. Cortical and striatal contributions to automaticity in information-integration categorization. Neuroimage 56, 1791–1802.

Wallis, J.D., Anderson, K.C., Miller, E.K., 2001. Single neurons in prefrontal cortex encode abstract rules. Nature 411, 953–956.

Witt, K., Nühsman, A., Deuschl, G., 2002. Intact artificial grammar learning in patients with cerebellar degeneration and advanced parkinson's disease. Neuropsychologia 40, 1534–1540.

Witt, S.T., Stevens, M.C., 2013. fmri task parameters influence hemodynamic activity in regions implicated in mental set switching. NeuroImage 65, 139–151.

Wolpert, D.M., Kawato, M., 1998. Multiple paired forward and inverse models for motor control. Neural Netw. 11, 1317–1329.

Wolpert, D.M., Miall, R.C., Kawato, M., 1998. Internal models in the cerebellum. Trends Cogn. Sci. 2, 338–347.

Yamada, H., Matsumoto, N., Kimura, M., 2004. Tonically active neurons in the primate caudate nucleus and putamen differentially encode instructed motivational outcomes of action. J. Neurosci. 24, 3500–3510.