The Integrative Role of Frontopolar Cortex in Rule-Based Category Learning

Dmitrii Paniukov & Tyler Davis

Department of Psychological Sciences, Texas Tech University

Address:

Texas Tech University, Department of Psychological Sciences, Box 42051 Lubbock, TX, 79409-2051, United States of America

Email addresses of authors: dmitrii.paniukov@gmail.com, tyler.h.davis@ttu.edu

Corresponding author is Dmitrii Paniukov. Preferred method of communication is email: dmitrii.paniukov@gmail.com

Abstract

Category learning is a critical neurobiological function that allows organisms to simplify a complex world. Frontopolar cortex (FPC) is often active in neurobiological studies of category learning. FPC has been associated in previous literature with either switching between representations or with representational integration. The goal of the present study was to clarify the role of FPC in category learning and dissociate switching and integration accounts. Two common types of category learning tasks, matching and classification, were utilized. The matching task involved matching a reference stimulus to one of four target stimuli. In the classification task, participants were shown a single stimulus and learned to classify it into one of two categories. Although nearly identical, matching and classification place differential demands on switching and integration. In matching, a rule can be known with certainty after a single correct answer. In classification, participants may need to integrate evidence for a rule even after an initial correct response. This critical difference allows isolation of integrative functions from switching functions. If the FPC is primarily involved in switching between representations, it should cease to be active once participants settle on a given rule in both tasks. If the FPC is involved in integration, its activation should persist in the classification task, but not matching. The results revealed that FPC activation persisted into correct trials in classification, but not matching, suggesting that it continues to integrate information even after subjects have arrived at the correct rule.

Keywords: category learning, frontopolar cortex, representational integration, representational switching, fMRI

Highlights

- Differences between rule-based matching and classification tasks were highlighted

- Frontopolar cortex is involved in representational integration in rule-based category learning tasks.
- Frontopolar cortex is functionally connected with caudate nucleus, a region that may also play an integrative role in rule learning.

1. Introduction

Friend or foe? Predatory or prey? Edible or poisonous? Category learning is a fundamental cognitive capacity that is critical for survival. Grouping objects into categories allows organisms to generalize information to novel examples and make inferences about future events. As a complex cognitive function, many brain regions are involved in category learning including the prefrontal cortex (PFC), medial temporal lobes, striatum, and visual cortex (Ashby & Maddox, 2005, 2011; Poldrack & Foerde, 2008; Seger & Miller, 2010; Smith & Grossman, 2008).

Although there are many types of category learning, one of the most studied types in cognitive neuroscience is rule-based category learning. In rule-based category learning, people learn a logical rule that can be used to determine whether items are members of the category or not. Many real world categories are associated with logical, albeit imperfect, rules. For example, members of the category mammals can often be categorized based on whether they have fur and nurse their young. Broadly speaking, rule-based categorization is thought to depend upon executive cortico-striatal loops that connect the PFC and the head of the caudate nucleus (Alexander et al., 1986; Seger & Miller, 2010). Although early work with patients and fMRI tended to treat the PFC as being involved in executive functions as a whole (Konishi et al., 1998, Robinson et al., 1980), recent work in cognitive neuroscience has begun to test whether specific subregions of the PFC serve distinct mechanisms (Seger, 2008; Seger & Miller, 2010; Ma et al., 2016).

One critical PFC region that has thus far eluded a thorough explanation in terms of its role in category learning is the lateral frontopolar cortex (FPC). The lateral FPC is known to be involved in a broad array of higher-level cognitive functions including abstract symbolic reasoning and analogical problem solving (Green et al., 2006; Specht et al., 2009), relational category learning (Davis et al., 2016), and goal-directed reward learning (Spreng et al., 2010). It is often described as a seat of human reasoning powers as it is significantly larger in humans than other primates, and its development across childhood tracks the development of fluid reasoning capacities (Gogtay et al., 2004; Semendeferi et al., 2011).

In humans, increased FPC activation is often observed for rule-based tasks involving abstract symbolic (Specht et al., 2009) and/or relational reasoning (Davis et al., 2016; Gray et al., 2003). However, the FPC's precise mechanistic role in these tasks has been described in a number of different ways. One characterization of the FPC focuses on switching between representations. For example, the FPC has been found to be more active when participants switch between cognitive sets in rule-based tasks (Konishi et al., 1998, 2002; Monchi et al., 2001; Strange et al., 2001; Liu et al., 2015), in tasks requiring switching between internally focused and externally focused attention (Burgess et al., 2007), and in reward learning tasks when exploring the values of different choices (Daw et al., 2006). An alternative characterization of the FPC focuses on its role in integrating information into abstract symbolic representations. For example, Badre, Kayser and D'Esposito (2010) found that the FPC tracks levels of rule abstraction in reinforcement learning. Nee, Jahn & Brown (2014) further observed FPC being associated with relational abstraction and cue integration. The goal of the present study was to investigate whether the FPC's role in rule-based category learning is best characterized by switching or integration accounts.

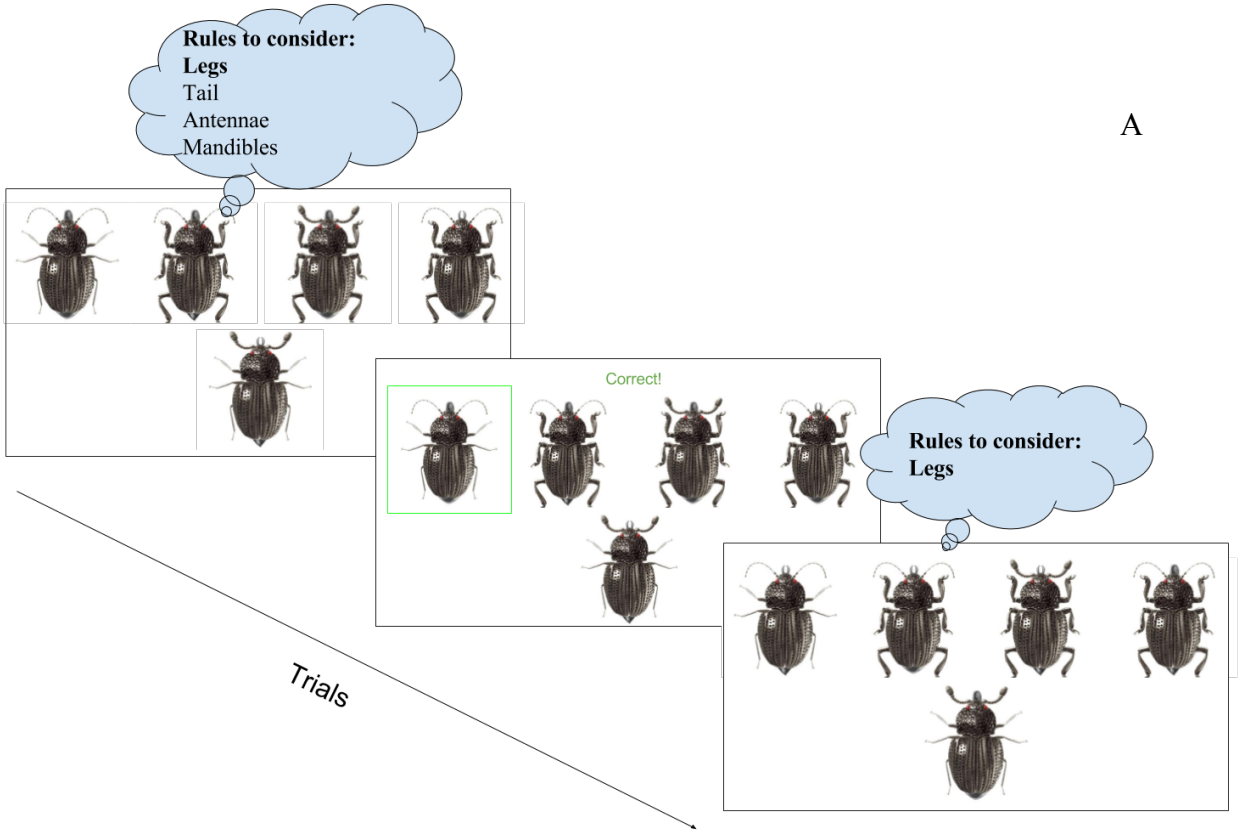
To test the distinction between rule-switching and integration accounts of FPC function, two commonly used rule-based tasks in the category learning literature — matching and classification tasks — were compared. Matching tasks, like the Wisconsin Card Sorting Test (Heaton, 1993), involve matching a multidimensional reference stimulus to a set of target stimuli that each match the reference on a single dimension. There is a rule that determines which target to match the reference to that is based on the stimulus dimensions. Participants learn the matching rule through trial and error. For example, in the matching task that we used in the present study, reference stimuli were schematic beetles that differed in terms of their legs, tail, antennae, and mandibles. Participants learned to match these reference stimuli to target beetles by choosing different candidate targets and receiving feedback. Often matching tasks will cycle through a number of rules forcing subjects to abandon rules and shift to a new rule when old rules cease being productive.

As a neuropsychological measure, the primary process of interest in matching tasks like the Wisconsin Card Sorting Test is the process of shifting between cognitive sets to accommodate novel rules and suppress the previously correct rules. Consistent with the theory that the FPC governs representational switching, results from a number of neuroimaging studies have identified activation in FPC during trials in which the rule is switched (Konishi et al., 1998, 2002; Monchi et al., 2001; Strange et al., 2001; Liu et al., 2015). However, switch trials as well as the immediately following trials in which the novel rule is being acquired also place demands upon integration mechanisms. Participants must not only switch between the previous and new candidate rule representations, but also accumulate and integrate information about the current candidate rules when feedback is received. One critical aspect of the matching tasks is that participants gain some conclusive information on every trial: if a choice is wrong, a candidate

rule can be eliminated; if a choice is correct, the rule is known with certainty. This aspect of matching tasks means that integration of evidence for a rule and switching are strongly intertwined because participants will switch rules whenever they are wrong (Konishi et al., 1998; Monchi et al., 2001) and will quit integrating new information once they are correct and proceed with applying the rule. Thus, for participants performing rationally, every possible switch trial is also a trial in which new information is being obtained and integrated. Likewise, every trial that provides new information about the rule is a trial in which the participant has switched the rule they are using.

We contrast matching tasks with another common type of rule-based category learning — classification tasks. Many neurobiological theories of category learning assume that matching and classification tasks have strong overlap in terms of the systems involved in acquiring and storing new rules and use results from both types of tasks interchangeably (e.g., Ashby & Maddox, 2005). However, classification tasks differ from matching tasks critically in requiring more extensive cross-trial integration of information about a rule, while otherwise keeping the demands on representational switching constant. In classification tasks, participants are shown single examples of stimuli and learn, using trial-and-error, to classify the stimuli into one or more categories based on the stimulus dimensions. Psychologically, as in matching tasks, participants must integrate information for the correct rule by switching their attention between different stimulus dimensions to test various candidate rules. Also, like matching tasks, it is possible to eliminate a particular rule based on feedback in a single trial — if a participant tries the rule “thick legs = Category A”, incorrect feedback will tell them that they need to switch the rule they are using and eliminate this rule. However, individual correct trials in classification tasks contain less information about the correct rule than in matching tasks. Whereas a single

trial in a matching task can tell whether a rule is correct, a correct trial in a classification tasks only tells the participant that some feature of the stimulus is connected to the category they chose. In order to solve a rule with full confidence in classification tasks, a participant often will need to complete several correct trials in a row. This greater integration demand on correct trials is not accompanied by an equivalent increase in switching demands, because subjects will not tend to switch from a rule when it is eliciting correct feedback. This critical difference whereby correct trials in matching tasks indicate perfect knowledge of the rule whereas correct classification trials do not, suggests that comparing matching and classification may help to elucidate whether the FPC's role in category learning is more consistent with integration or switching accounts (see Figure 1 for illustration of integrative demands in the tasks).



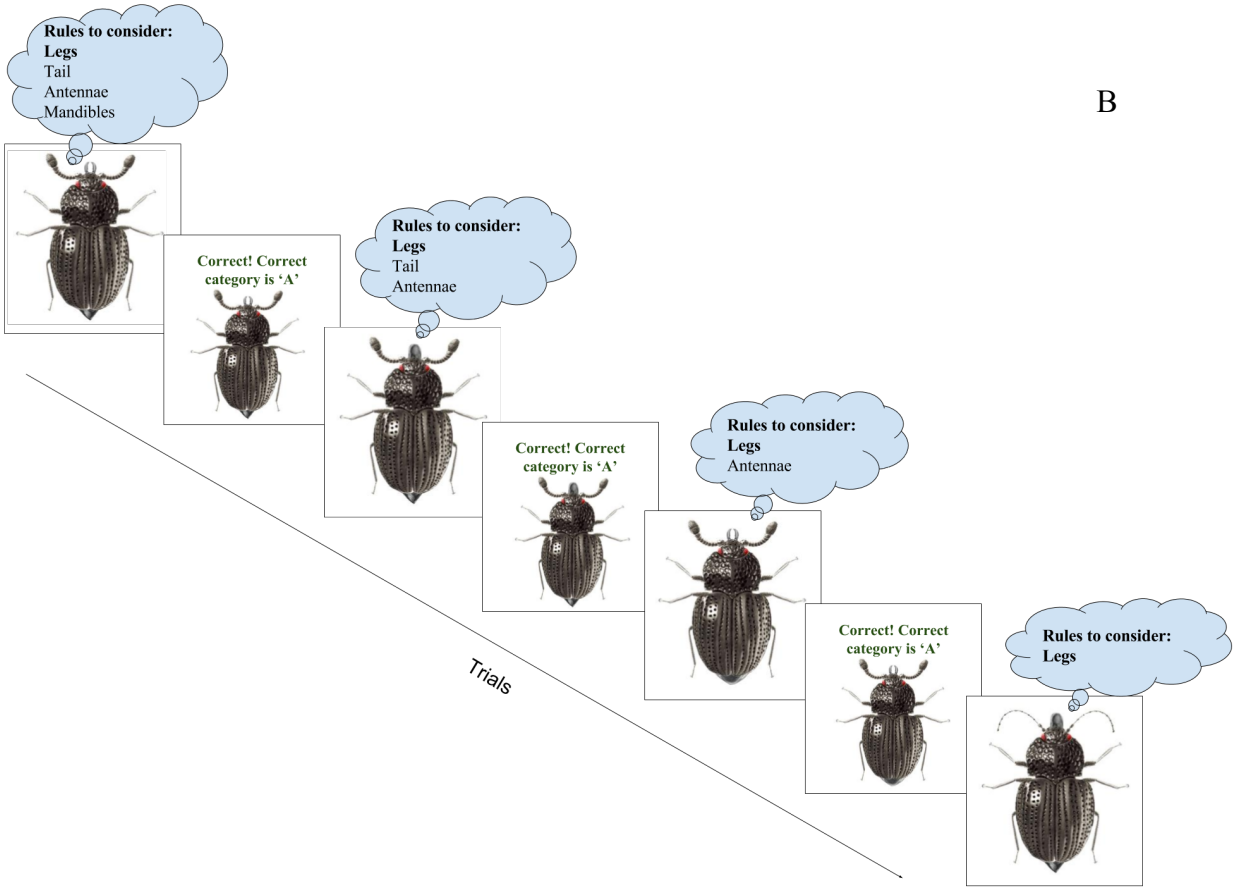


Figure 1. Examples of matching (A) and classification tasks (B) and how rules are eliminated in both tasks. In matching tasks, when the rule is correctly selected, all other rules are eliminated. Contrastingly, in classification tasks, even if a correct rule is selected, several additional trials may be necessary to rule out other rules that are possible given the history of stimulus-category pairings participants have seen up to that point. In both of these cases, participants have arrived at the correct rule dimension ("legs" in bold) on the first trial of the depicted sequence. In matching, a single instance of correct feedback establishes that "legs" are the rule-dimension. In classification, although the participant may start out using the legs rule, they may only fully eliminate other possible rules that are consistent with the stimulus history after a number of correct trials. For example, if in the first trial the participants chooses category A for the thin

legged beetle and gets correct feedback, the rules "Pointy tail = 'A'; Thick antennae = 'A'; and bisected mandibles = 'A'" are all still active. They only become eliminated after the participants encounters variations of the "thin legs = 'A'" rule in which these candidate rules do not also hold.

To test how differences in integration demands between matching and classification tasks impact FPC activation, we had participants complete the matching and classification tasks using schematic beetle stimuli. In both tasks, participants would iteratively learn rules via trial and error. Once rules were learned (defined by four correct trials in a row) the rule would switch and the participant would begin learning a new rule. For each rule there was therefore a *rule learning* phase, when participants were narrowing down and switching between rules, and a brief *rule application* phase, in which participants were using the final rule that they have arrived at (Figure 2).

	Rule learning				Rule application			
Trial #	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6	Trial 7	Trial 8
Feedback	Incorrect	Correct	Incorrect	Incorrect	Correct	Correct	Correct	Correct

Figure 2. An example of how the rule learning and rule application phases were separated for analysis. The last four or more contiguous correct trials were labeled as rule application. In this phase participants applied a final rule they arrived at. All prior to the last correct trials were labeled as rule learning. Participants were narrowing down and switching between rules trying to find out the correct rule. All trials when participants did not respond were excluded from the phases.

The different accounts of FPC function can be used to make predictions for the rule learning and application phases. If the FPC is primarily involved in switching, it should only be active during the rule learning phases of both matching and classification tasks, when participants are trying different rules and switching between them. Contrastingly, if the FPC is involved in integration, it should be active not only during the rule learning phase of both tasks, but also during the rule application phase for classification when participants have arrived at the correct rule, but are still integrating evidence in support of it. Because integration is not necessary during matching, once the correct rule is found, FPC should not be active during the application phase of the matching task.

In addition to the FPC, we will also examine whether functional connectivity between the FPC and caudate nucleus is more consistent with the integration or switching account. Like FPC, the caudate nucleus had been implicated in integration and switching (Ashby & Maddox, 2005; Ashby et al., 1998; Filoteo et al., 2005; Liu et al., 2015; Monchi et al. 2001; Seger & Cincotta, 2006). Thus, to complement our main analyses, we used a Psychophysiological interaction (PPI) analysis that modeled connectivity between these regions during the rule-learning and application phases of the matching and classification trials. As with above, the integration account predicts that FPC and caudate connectivity should remain higher during the application phase of classification rules compared to matching.

To foreshadow our results, we found evidence consistent with the integration hypothesis. FPC activation during rule application for the classification task was significantly greater than rule application for the matching task. Greater FPC activation was also present for classification compared to matching during the rule learning phases. Likewise, there was significantly greater connectivity between an FPC seed and the caudate during both the rule application and learning

phased for the classification task compared to matching. These results suggest that the FPC's role in category learning may be primarily integrative and serve to accumulate information when acquiring novel rules.

2. Method

2.1. Participants

Twenty-seven participants were recruited from the Texas Tech University community via an electronically posted announcement. Participants were required to be at least eighteen years old, right-handed, have a minimum of an eighth grade education, speak English fluently, and not have any contraindications for MRI research. Participants were compensated with $35 for the study. Two participants fell asleep during their scanning session, and therefore were removed from the further analyses. One participant opted out from the last two scanning runs, but the rest of his/her data were used in the analyses. The study was approved by Human Research Protection Program at Texas Tech University.

2.2. Stimuli and Procedure

Participants were asked to sign a consent form, MRI-safety checklist and complete a computer-based tutorial, where they were shown example trials and stimuli outside of the scanner. Upon completion of the screening forms and the tutorial, participants were placed into the scanner. In the scanner, participants completed four runs of the matching task and four runs of the classification task in an order that was balanced across participants.

Two tasks, matching and classification, utilized sixteen images of schematic beetles. The beetles varied across four binary feature dimensions: legs, mandibles, antennae, and tail (see Figure 3 for two examples with opposite features). For both tasks, the sixteen stimuli were presented in blocks in which each stimulus was shown once in a random order.

Figure 3. Stimuli of schematic beetles with completely distinct feature dimensions. The feature dimensions were legs (thick or thin), mandibles (closed or bisected), antennae (fuzzy or dotted), and tail (pointy or round).

In the matching task, each screen contained a reference beetle and four target beetles, each of which matched the reference beetle on a single dimension (see Figure 4). The position of the beetles on the screen was randomized to minimize effects of feature salience and balance motor responses.

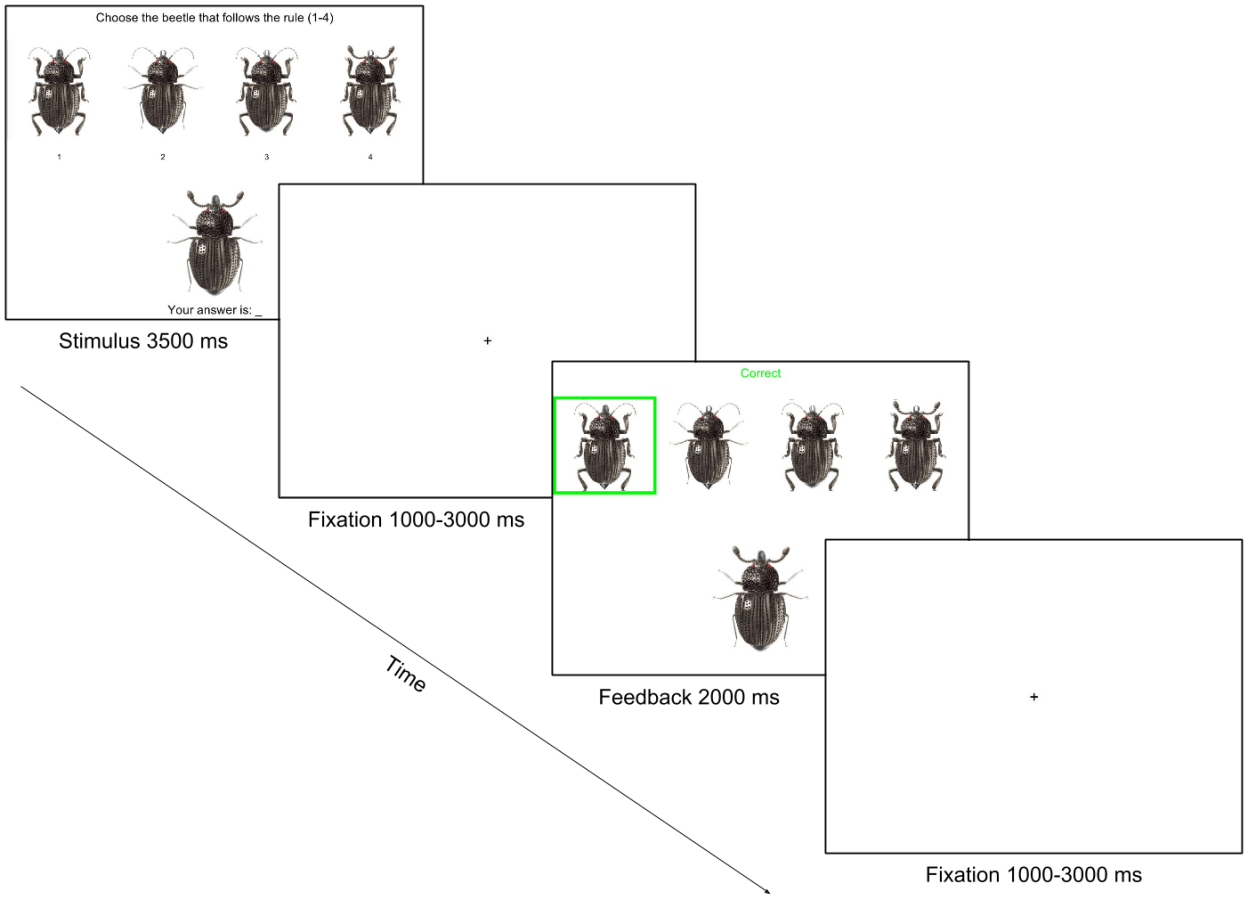


Figure 4. Example of the matching task. Participants saw a reference beetle on the bottom and four target beetles on top. The target beetles each matched the reference beetle on a single feature. Participants would select a target beetle then receive feedback about whether their choice was correct or incorrect.

In the classification task, each screen contained a single beetle (see Figure 5). The beetles were assigned into category A or B, based on a random rule, defined by a single feature.

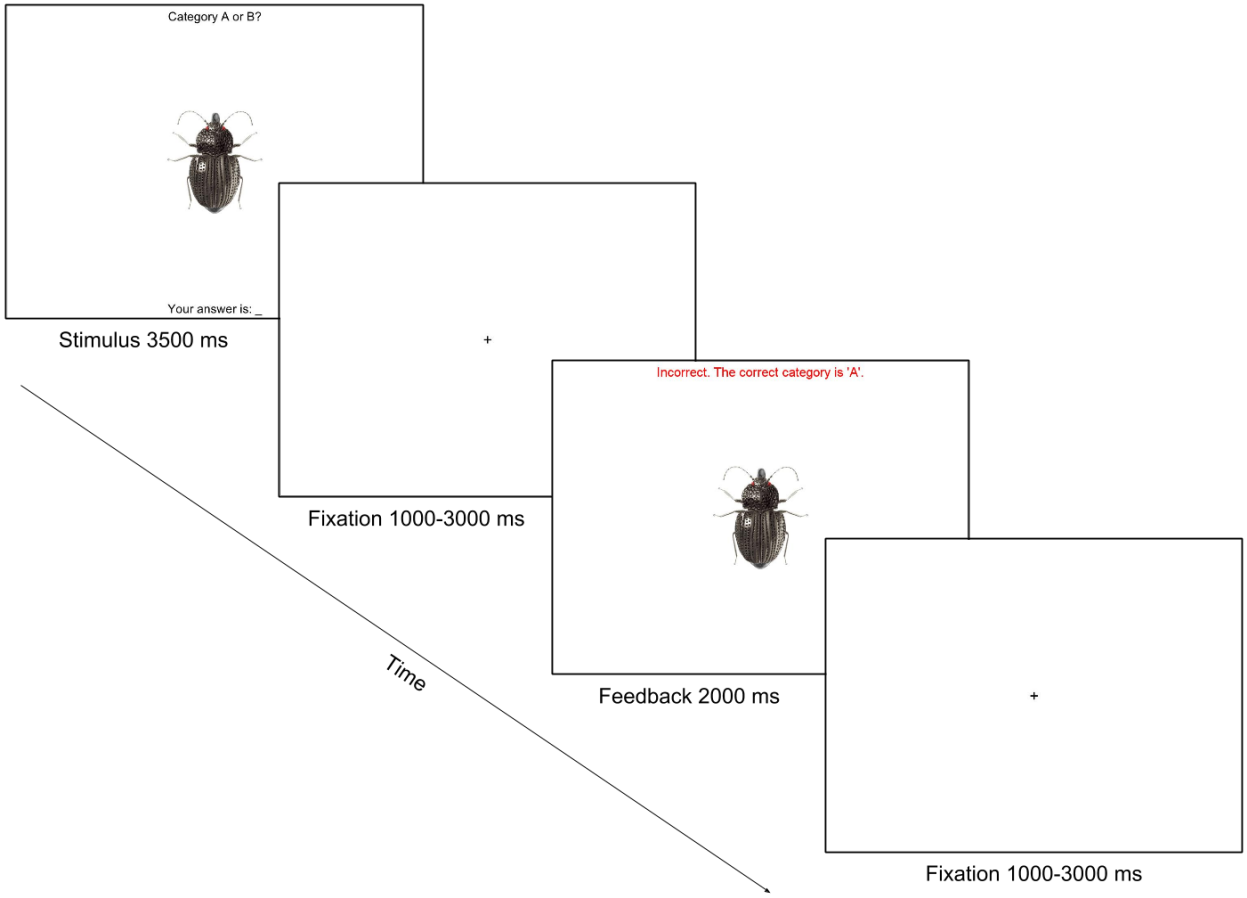


Figure 5. Example of the classification task. Participants saw a single beetle on the screen and where asked to categorize it as a member of Category A or B. After participants made the choice, they were given feedback about whether their choice was correct and the correct category label.

In both tasks, participants had 3.5 seconds to categorize a stimulus using a button box held in their right hand. After a brief fixation (1, 2, or 3s; mean = 2s), participants were provided with feedback for 2s, during which the beetle was presented again along with a message "Correct/Incorrect/Failed to respond" in the matching task or "Correct/Incorrect/Failed to respond." and "The correct category is A/B" in the classification task. After the feedback, a brief fixation was presented (1, 2, or 3s; mean = 2s). Upon achieving a termination criterion of four correct answers in the row, the rule was switched to a new randomly selected rule. When a run

was completed, participants were presented with a number of correctly solved rules during the run (i.e., "You successfully learned [number] categories! Good job!"). Participants had thirty-two trials in each run and eight runs total (four runs in a block for each task), followed by an anatomical scan. After the scan, participants were thanked, compensated and dismissed from the study.

2.3. fMRI Data Acquisition

The data were collected at Texas Tech Neuroimaging Institute using a Siemens Skyra 3T scanner with a 20-channel head coil. A T1-weighted sagittal MPRAGE was obtained with TR = 1.9s, TE = 2.49, flip angle = 9, matrix of 256x256, field of view = 240, slice thickness= 0.9 mm with a gap of 50%, 1 slice. T2-weighted BOLD echo planar images (EPI) were obtained with TR=2s, TE=30ms, flip angle = 90, matrix = 64x64, field of view = 192, slice thickness = 3x3x4 mm, 35 ascending axial slices, 156 volumes in each scanning run. The slice prescription was tilted off of parallel with AC/PC to reduce susceptibility artifact in orbital frontal cortex (Deichmann et al., 2003).

2.4. fMRI Data Preprocessing and Analysis

MRI data preprocessing included the following steps: DICOM imaging files were converted to NifTI files using dcm2nii from the Mricron software package (Rorden & Brett, 2000). Skulls were removed using the BET tool (Smith, 2002) from FSL software package (Jenkinson et al., 2012; Woolrich et al., 2009) for the BOLD EPI images and ANTs (Avants et al., 2009) with OASIS template (Avants & Tustison, 2014) for the T1 anatomical images. Motion correction was carried out using a 6-DOF affine transformation implemented in FSL's MCFLIRT tool (Jenkinson et al., 2002). Data were smoothed with an 8 mm FWHM Gaussian kernel and highpass filtered (100s cut off). Finally, data were manually checked for quality

issues such as visual spikes produced by the scanner, incorrect brain extraction, and excessive motion. The quality check revealed visual artifacts in the first run for four participants and thus these runs were excluded from further analysis.

Functional MRI data analysis was carried out using a standard three-level analysis in FSL's FEAT, as implemented in Nipype (Gorgolewski et al., 2011). The first-level analysis consisted of prewhitening using FSL's FILM (Woolrich et al., 2001) to account for temporal autocorrelation, task based-regressors (see below) convolved with double-gamma hemodynamic response, and their temporal derivatives. Additional confound regressors of no interest included six motion parameters, their temporal derivatives and regressors for scrubbing high motion volumes exceeding a framewise displacement of 0.9mm (Siegel, et al., 2014). The same high pass filter setting that was used to process the fMRI data was used to process the design matrix. First-level statistical maps were registered to a standard space in a two-stage registration consisting of (1) registration of each BOLD timeseries to respective participants' T1 MPRAGE using the BBR algorithm (Greve & Fischl, 2009) and (2) registration of the T1 to the standard space (MNI-152 brain template) using nonlinear ANTs registration (Avants et al., 2009). Second-level analysis combined across runs within a participant, and was carried out using a fixed effects model in FLAME (Beckmann et al., 2003; Woolrich et al., 2004; Woolrich, 2008). Second-level regressors included task type (matching vs. classification). Third-level analyses examined the effects of first and second-level contrasts across participants. Final thresholding of statistical maps at $p < 0.05$ was done via a cluster mass correction (with primary/cluster-forming threshold of $t(24)=2.49$) (Bullmore et al., 2000) for multiple comparisons implemented in FSL's non-parametric permutation tool Randomise (Winkler et al., 2014) with variance smoothing set at 8 mm.

All trials in both tasks were split into rule learning phase in which participants were acquiring the correct rule, and a rule application phase in which subjects had acquired the rule and were now applying it (see Figure 2 for an example of the phases for the classification task). In the matching task, all correct trials were considered as rule application trials due to the nature of the task, and all incorrect trials as rule learning. In the classification task only last four or more continuous correct trials were considered as rule application, and the rest of the trials as rule learning.

Two univariate models were used to model the fMRI data. Model 1 used six explanatory variables (EVs): (1) stimulus presentation (onsets) for rule application trials, (2) onsets for rule learning trials, (3) onsets for trials when participants did not answer, (4) correct feedback, (5) incorrect feedback, and (6) feedback for trials when participants did not answer. These EVs were compared in three contrasts: (1) rule application > baseline, (2) rule learning > baseline, (3) rule application > rule learning. Statistical maps describing the results of these contrasts were thresholded to correct for multiple comparisons at the whole-brain level and within an a priori FPC mask (ROI), which was defined as the max-probability frontal pole region from the Harvard-Oxford atlas included in FSLView (Desikan, et al., 2006).

A second model (Model 2) tested whether the task variables modulated the connectivity between FPC and other brain regions using a psychophysiological interaction (PPI) analysis. FPC signal was extracted from an 8mm sphere around the highest peak found in the rule learning > baseline contrast for classification trials (MNI coordinates: x=-40, y=54, z=4 mm). The EVs in Model 2 included all six EVs from Model 1 in addition to three new EVs, namely (7) the FPC timeseries, (8) the interaction of rule application onsets with FPC timeseries, and (9) interaction of rule learning onsets with FPC timeseries. No convolution, temporal filtering or temporal

derivatives were used for the FPC timeseries EV and its interactions. Three contrasts examined the (1) interaction of rule application onsets with FPC timeseries vs. baseline, (2) interaction of rule learning onsets with FPC timeseries vs. baseline, and (3) interaction of rule application onsets with FPC timeseries vs. interaction of rule learning onsets with FPC timeseries. The Harvard-Oxford max-probability caudate nucleus mask was used as an a-priori ROI in Model 2.

3. Results

3.1. Behavioral results

For the behavioral results we used a poisson mixed effects model from LME4 package (Bates et al., 2015) in R (R Core Team, 2014) and a mixed effects model Cox regression survival model from Coxme package (Therneau, 2015) in R to examine how number of rules solved and trials to the rule termination criterion differed between task types. Consistent with the assumption that the classification task placed more demands on integrating over trials than matching, participants solved more rules during matching (M=15.6, SD=4.14) than during the classification task (M=9.92, SD=3.22), z=5.58, p<.001. Likewise, participants took fewer trials to reach a rule termination criterion in matching (M=6.56, SD=4.67) than in the classification task (M=11.73, SD=7.56), z=12.11, p<.001.

3.2. Imaging results

To examine the neuroimaging results, all trials were sorted into either rule learning or rule application phases. In the matching task all correct trials were considered as rule application trials due to the nature of the task, and all incorrect trials -- as rule learning. In the classification task only last four or more contiguous correct trials were considered as rule application, and the rest of the trials as rule learning. Trials when participants failed to respond were not included into any phase.

3.2.1. Rule learning vs. rule application in the matching task.

Both the matching and integration accounts of FPC function predict that there should be more FPC activation during the learning phase of the matching task. Rule learning should require more integration demands than application because rules are known with full certainty once a correct answer is found, and thus participants will likely cease integrating additional evidence for a rule. Likewise, prior to arriving at the correct rule, participants should switch between rules when they are presented with incorrect feedback. Consistent with these accounts, we found a cluster (659 voxels) in the FPC ROI that activated greater for learning than application trials.

Additional areas revealed to be more active during learning than application in the matching task were parietal and lateral occipital regions (see Table 1 for detailed list of regions and Figure 6). These results are consistent with other studies examining activation during learning phases of similar rule-based category learning tasks (Seger & Cincotta, 2006; Liu et al., 2015). Whole-brain results for the opposite contrasts (rule application > rule learning) were also consistent with previous findings on rule application (Seger & Cincotta, 2006; Nomura et al., 2007) with greater activation in hippocampus during rule application. Additional regions more active for rule application included the ventromedial prefrontal cortex (VMPFC), a region we have recently identified as being associated with stronger decision evidence during categorization (Davis et al. 2016), putamen, parietal cortex, occipital cortex, temporal cortex, insular cortex, parahippocampal gyrus (see Table 1 for details; all brain maps and other project details are posted at https://osf.io/ge8vf/)

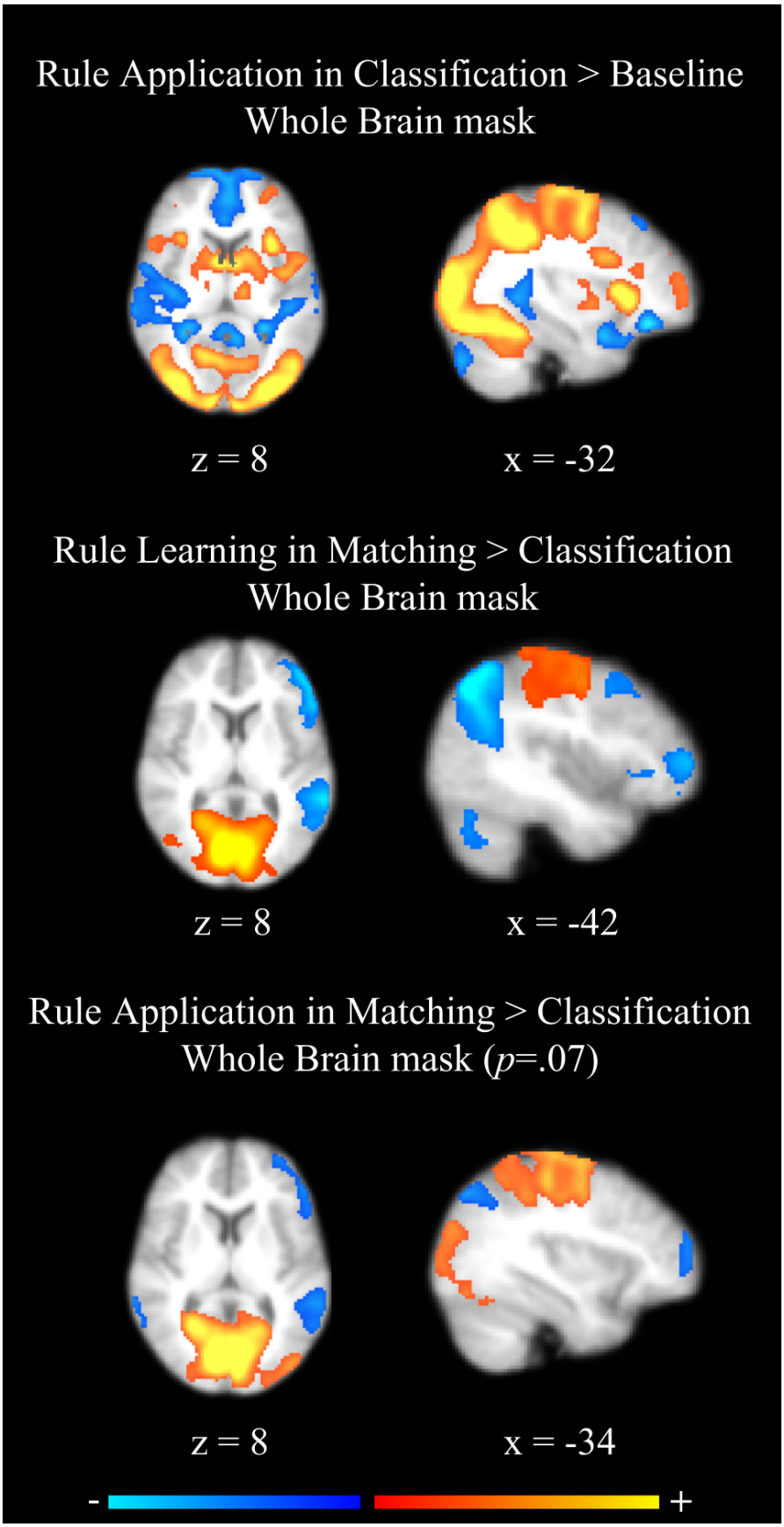

Rule Application in Classification > Baseline
Whole Brain mask
z = 8
x = -32
Rule Learning in Matching > Classification
Whole Brain mask
z = 8
x = -42
Rule Application in Matching > Classification
Whole Brain mask (p=.07)
z = 8
x = -34
-
+

Figure 6. Brain activations for the main contrasts. Matching > classification contrasts are in red-yellow and classification > matching in blue.

Table 1

Rule learning and rule application in the matching task

Cluster Region(s)	**Cluster Size (voxels), Cluster p-value**	**Local Maximum Region**	**t-value**	**MNI coordinates (x,y,z)**		
Rule Learning > Rule Application						
Superior Parietal Cortex Lateral Occipital Cortex	3709, p=0.029	Superior Parietal Lobule	5.13	-32	-54	42
		Lateral Occipital Cortex, superior division	4.84	-24	-64	40
		Lateral Occipital Cortex, superior division	4.42	12	-76	56
		Lateral Occipital Cortex, superior division	4.25	30	-70	56
Rule Application > Rule Learning						
Medial and Lateral Parietal Cortex Ventral Occipital Cortex Temporal Cortex Insular Cortex Putamen Hippocampus Parahippocampal Gyrus	30173, p=0.001	Supramarginal Gyrus, anterior division	6.83	62	-28	30
		Precuneous Cortex	6.55	-10	-58	12
		Parietal Operculum Cortex	6.51	60	-24	18
		Lingual Gyrus	6.44	-28	-40	-8
Medial Prefrontal Cortex Frontal Pole	2365, p=0.0466	Frontal Pole	6.46	0	58	-10
		Frontal Medial Cortex	5.48	0	48	-16
		Paracingulate Gyrus	4.43	-2	56	6
		Frontal Pole	3.93	0	56	20
Rule Learning > Baseline						
Superior Parietal Cortex Lateral Occipital Cortex Lateral and Medial Prefrontal Cortex Putamen Caudate	66716, p=0.0002	Occipital Fusiform Gyrus	18.5	26	-70	-10
		Occipital Fusiform Gyrus	17.4	-24	-72	-10
		Lateral Occipital Cortex, superior division	16.9	32	-68	32

Talamus Insular Cortex Cerebellum		Lateral Occipital Cortex, superior division	16	32	-78	22
Baseline > Rule Learning						
Lateral Occipital Cortex Lateral Parietal Cortex Temporal Cortex Hippocampus Parahippocampal Gyrus	22230, p=0.0018	Lateral Occipital Cortex, superior division	9.03	-46	-76	32
		Lateral Occipital Cortex, superior division	8.62	-52	-62	28
		Lateral Occipital Cortex, superior division	8.43	-46	-68	44
		Cingulate Gyrus, posterior division	8.17	-10	-50	32
Lateral Parietal Cortex Insular Cortex Temporal Cortex Hippocampus Parahippocampal Gyrus	12367, p=0.0094	Superior Temporal Gyrus, posterior division	7.98	62	-24	-2
		Insular Cortex	7.33	40	-16	2
		Supramarginal Gyrus, anterior division	7.21	66	-28	34
		Angular Gyrus	7.21	46	-54	24
Lateral and Medial Prefrontal Cortex	7832, p=0.0162	Frontal Pole	7.42	0	58	18
		Frontal Pole	7.0	-20	42	48
		Frontal Medial Cortex	6.63	0	48	-10
		Superior Frontal Gyrus	6.34	-26	28	54
Rule Application > Baseline						
Occipital Cortex Superior Parietal Cortex Lateral Prefrontal Cortex Medial Prefrontal Cortex Frontal Pole Caudate Putamen Thalamus Insular Cortex Cerebellum	64831, p=0.0002	Occipital Fusiform Gyrus	12	-24	-66	-8
		Postcentral Gyrus	11.9	-46	-32	46
		Occipital Fusiform Gyrus	11.8	24	-74	-12
		Juxtapositional Lobule Cortex	11.6	-4	2	54
Baseline > Rule Application						
Posterior Parietal Cortex Lateral Occipital Cortex Temporal Cortex	7256, p=0.0162	Lateral Occipital Cortex, superior division	7.35	-46	-76	32
		Lateral Occipital Cortex, superior division	6.83	-46	-68	44
		Angular Gyrus	6.59	-50	-60	30

		Angular Gyrus	5.84	-58	-54	38
Medial and Lateral Prefrontal Cortex	6713, p=0.0198	Middle Frontal Gyrus	7.05	30	32	54
		Superior Frontal Gyrus	6.71	0	38	58
		Frontal Pole	5.99	-2	60	36
		Middle Frontal Gyrus	5.94	-28	28	54
Posterior Parietal Cortex Lateral Occipital Cortex Temporal Cortex	6506, p=0.0198	Angular Gyrus	6.58	60	-50	38
		Lateral Occipital Cortex, superior division	6.55	50	-68	36
		Superior Temporal Gyrus, posterior division	6.39	64	-24	-2
		Angular Gyrus	6.35	46	-54	26
Motor Cortex Superior Parietal Cortex	3076, p=0.0332	Postcentral Gyrus	7.32	0	-44	72
		Precuneous Cortex	6.36	0	-56	66
		Precentral Gyrus	6.02	-2	-30	76
		Precuneous Cortex	5.9	-2	-58	34

Note: Local maximum regions are based on the Harvard-Oxford atlas. The coordinates are in the standard MNI space. The p-values are taken from non-parametric tests produced by Randomise tool in FSL.

3.2.2. Rule learning and rule application in the classification task.

Similar to the matching task, parietal and occipital regions were activated in the rule learning > rule application contrast for the classification task. Additional areas that were more active for rule learning than application included the, dorsolateral PFC and cerebellum. See Table 2 for the regions found in the classification task contrasts. No regions were significantly more active in rule application compared to rule learning (rule application > rule learning) in the whole brain results.

Table 2

Rule learning and rule application in the classification task

Cluster Region(s)	**Cluster Size**	**Local Maximum Region**	**t-value**	**MNI coordinates (x,y,z)**

	(voxels), Cluster p-value					
Rule Learning > Rule Application						
Superior Parietal Cortex Lateral Occipital Cortex	14340, p=0.0006	Lateral Occipital Cortex, superior division	6.4	32	-68	32
		Supramarginal Gyrus, posterior division	6.04	38	-38	40
		Lateral Occipital Cortex, superior division	5.95	28	-70	54
		Superior Parietal Lobule	5.73	-32	-54	42
Lateral Prefrontal Cortex	1911, p=0.0498	Middle Frontal Gyrus	5.47	-32	2	66
		Middle Frontal Gyrus	4.58	-36	0	52
		Middle Frontal Gyrus	4.39	-46	32	26
		Middle Frontal Gyrus	4.1	-50	14	38
Rule Learning > Baseline						
Lateral Occipital Cortex Superior Parietal Cortex Lateral Prefrontal Cortex Medial Prefrontal Cortex Frontal Pole Caudate Putamen Thalamus Insular Cortex Cerebellum	66143, p=0.0002	Lateral Occipital Cortex, superior division	17.4	32	-68	32
		Superior Parietal Lobule	17.1	-32	-48	42
		Lateral Occipital Cortex, superior division	16.9	-32	-88	8
		Lateral Occipital Cortex, superior division	15.9	28	-68	54
Baseline > Rule Learning						
Medial Prefrontal Cortex Lateral Prefrontal Cortex Frontal Pole Temporal Cortex Posterior Parietal Cortex Occipital Cortex	21313, p=0.0018	Frontal Medial Cortex	9	0	48	-12
		Paracingulate Gyrus	8.53	2	54	-2
		Frontal Pole	8.19	0	60	18
		Planum Polare	7.29	42	-16	0
Medial Parietal Cortex Hippocampus	6926, p=0.011	Precuneous Cortex	8.23	2	-46	72
		Cingulate Gyrus, posterior division	7.25	-6	-52	26

		Cuneal Cortex	6.98	4	-82	44
		Cingulate Gyrus, posterior division	6.97	10	-42	0
Rule Application > Baseline						
Occipital Cortex Superior Parietal Cortex Lateral Prefrontal Cortex Medial Prefrontal Cortex Frontal Pole Caudate Putamen Thalamus Insular Cortex Cerebellum	49757, p=0.0002	Occipital Fusiform Gyrus	10.7	28	-68	-12
		Postcentral Gyrus	10.7	-48	-32	48
		Superior Parietal Lobule	10.7	-32	-48	42
		Temporal Occipital Fusiform Cortex	10.6	32	-46	-20
Baseline > Rule Application						
Medial Parietal Cortex Temporal Cortex Insular Cortex Frontal Pole Lateral Prefrontal Cortex Medial Prefrontal Cortex	12940, p=0.006	Frontal Medial Cortex	6.95	0	52	-12
		Frontal Pole	6.83	0	64	20
		Superior Frontal Gyrus	6.64	0	54	42
		Frontal Orbital Cortex	6.55	-32	36	-14
Superior Parietal Cortex Occipital Cortex Hippocampus Parahippocampal Gyrus	5713, p=0.013	Precuneous Cortex	9.19	0	-46	70
		Postcentral Gyrus	6.8	30	-36	74
		Postcentral Gyrus	6.5	14	-30	80
		Precentral Gyrus	6.49	34	-26	72
Superior Parietal Cortex Temporal Cortex Insular Cortex	4548, p=0.0198	Middle Temporal Gyrus, posterior division	6.32	62	-14	-8
		Heschl's Gyrus (includes H1 and H2)	6.12	42	-16	2
		Lingual Gyrus	6.04	32	-46	4
		Superior Temporal Gyrus, posterior division	4.73	62	-34	4
Cerebellum	2184, p=0.0438	Cerebellum	5.38	26	-86	-36
		Cerebellum	4.83	-32	-82	-38
		Cerebellum	4.71	24	-86	-26

	Cerebellum	4.33	38	-80	-36

Note: Local maximum regions are based on the Harvard-Oxford atlas. The coordinates are in the standard MNI space. The p-values are taken from non-parametric tests produced by Randomise tool in FSL.

Our primary region of interest, FPC, was significantly activated relative to baseline in both rule learning > baseline and rule application > baseline contrasts in both the whole brain and ROI analyses (see Table 2). Further, there was no difference between FPC activation during rule learning and application for the classification task within either the ROI or whole brain analysis, even at liberal uncorrected thresholds. Together, these results suggest that the FPC is engaged in a more integrative role in the classification task. This result is inconsistent with switching accounts because participants should not be switching rules once they arrive at the appropriate rule. However, they may continue integrating confirmatory evidence for their chosen rule because the rule is rarely known with certainty after the first correct trial in the classification phase.

3.2.3. Comparing rule learning in the matching and classification tasks.

Occipital and parietal cortices as well as thalamus were more active during the learning phase for matching than for classification. The classification task activated more FPC (in both whole brain and ROI analyses), in addition to occipital, parietal, temporal, lateral prefrontal cortices and cerebellum (Table 3).

Table 3

Rule learning in the matching and classification tasks

Cluster Region(s)	**Cluster Size (voxels), Cluster p-value**	**Local Maximum Region**	**t-value**	**MNI coordinates (x,y,z)**

Matching > Classification						
Occipital Cortex Superior Parietal Cortex Thalamus	20367, p=0.0002	Intracalcarine Cortex	17.4	10	-78	4
		Lingual Gyrus	15.1	-12	-76	-10
		Intracalcarine Cortex	14	-8	-82	2
		Lateral Occipital Cortex, superior division	10.9	22	-88	24
Classification > Matching						
Occipital Cortex Superior Parietal Cortex Temporal Cortex Cerebellum	6663, p=0.0052	Angular Gyrus	7.41	-46	-54	56
		Lateral Occipital Cortex, superior division	7.2	-44	-68	46
		Supramarginal Gyrus, posterior division	6.27	-54	-48	50
		Middle Temporal Gyrus, posterior division	5.82	-56	-30	-4
Frontal Pole Lateral Prefrontal Cortex	2725, p=0.039	Frontal Pole	5.23	-50	42	4
		Inferior Frontal Gyrus, pars triangularis	5.03	-56	24	16
		Middle Frontal Gyrus	4.62	-54	24	26
		Inferior Frontal Gyrus, pars triangularis	4.61	-50	20	0

Note: Local maximum regions are based on the Harvard-Oxford atlas. The coordinates are in the standard MNI space. The p-values are taken from non-parametric tests produced by Randomise tool in FSL.

3.2.4. Comparing rule application in the matching and classification tasks.

The primary goal of the current study was to compare rule integration versus switching accounts of FPC function. Because participants, by definition, quit switching rules during rule application in both matching and classification tasks, but the classification task often involves continued uncertainty and integration of evidence for a the rule, contrasting rule application during matching and classification allows us to isolate such integration mechanisms. Consistent with the integration account, FPC was more active in the classification task compared to the matching task. This difference was significant for the a priori ROI analysis and marginal in the

whole-brain analysis ($p = 0.07$). In addition to FPC, there was greater activation in superior parietal, temporal, and lateral occipital regions, in the classification task, and inferior parietal and medial occipital regions in the matching task (Table 4). The superior parietal regions have been associated with measures of decisional uncertainty in the broader categorization literature (Davis et al., 2016; Grinband et al. 2006; Degutis and D'Esposito 2007; Braunlich and Seger 2016).

Table 4

Rule application in the matching and classification tasks

Cluster Region(s)	**Cluster Size (voxels), Cluster p-value**	**Local Maximum Region**	**t-value**	**MNI coordinates (x,y,z)**		
Matching > Classification						
Occipital Cortex Superior Parietal Cortex Cerebellum	29803, p=0.0002	Lingual Gyrus	12.7	12	-74	-4
		Intracalcarine Cortex	12.4	10	-80	4
		Lingual Gyrus	11.9	-12	-76	-10
		Intracalcarine Cortex	11.2	-8	-84	2
Classification > Matching						
Occipital Cortex Posterior Parietal Cortex Temporal Cortex	3768, p=0.017	Supramarginal Gyrus, posterior division	6.22	-48	-50	56
		Middle Temporal Gyrus, posterior division	5.69	-62	-28	-4
		Lateral Occipital Cortex, superior division	5.27	-42	-68	48
		Angular Gyrus	4.87	-60	-58	12
Occipital Cortex Posterior Parietal Cortex Temporal Cortex	2033, p=0.0366	Angular Gyrus	5.66	54	-50	50
		Angular Gyrus	5.42	60	-48	38
		Middle Temporal Gyrus, temporooccipital part	5.19	66	-46	-4
		Lateral Occipital Cortex,	4.53	52	-62	40

	superior division

Note: Local maximum regions are based on the Harvard-Oxford atlas. The coordinates are in the standard MNI space. The p-values are taken from non-parametric tests produced by Randomise tool in FSL.

3.2.5. FPC Connectivity analysis.

Although our primary hypotheses about the FPC were about its activation during rule application for classification and matching, it is also useful to examine connectivity between the FPC and other regions. One a priori region we were interested in due to its prominence in theories of rule-based categorization is the caudate nucleus. The caudate nucleus is another brain region that is known to be involved in rule learning and has been ascribed a number of different roles including hypothesis testing and rule switching. (Ashby & Maddox, 2005; Ashby et al., 1998; Filoteo et al., 2005; Liu et al., 2015; Monchi et al. 2001; Seger & Cincotta, 2006). Hypothesis testing itself is a complex process that involves both testing of different rules and integration of evidence for a rule. In the broader literature, the caudate is involved with many tasks, including procedural task chunking and habit formation (Graybiel, 2008), so one possibility is that it interacts with the FPC to facilitate rule integration during learning. To test this hypothesis, we ran a psychophysiological interaction analysis (PPI) using FPC as a seed region. Consistent with the idea that the caudate interacts with FPC to facilitate rule integration, we found stronger connectivity between the FPC and caudate during both rule learning and application for the classification task compared to the matching task. This clarifies the role of the caudate in hypothesis testing as being involved specifically with the integration of evidence for a rule.

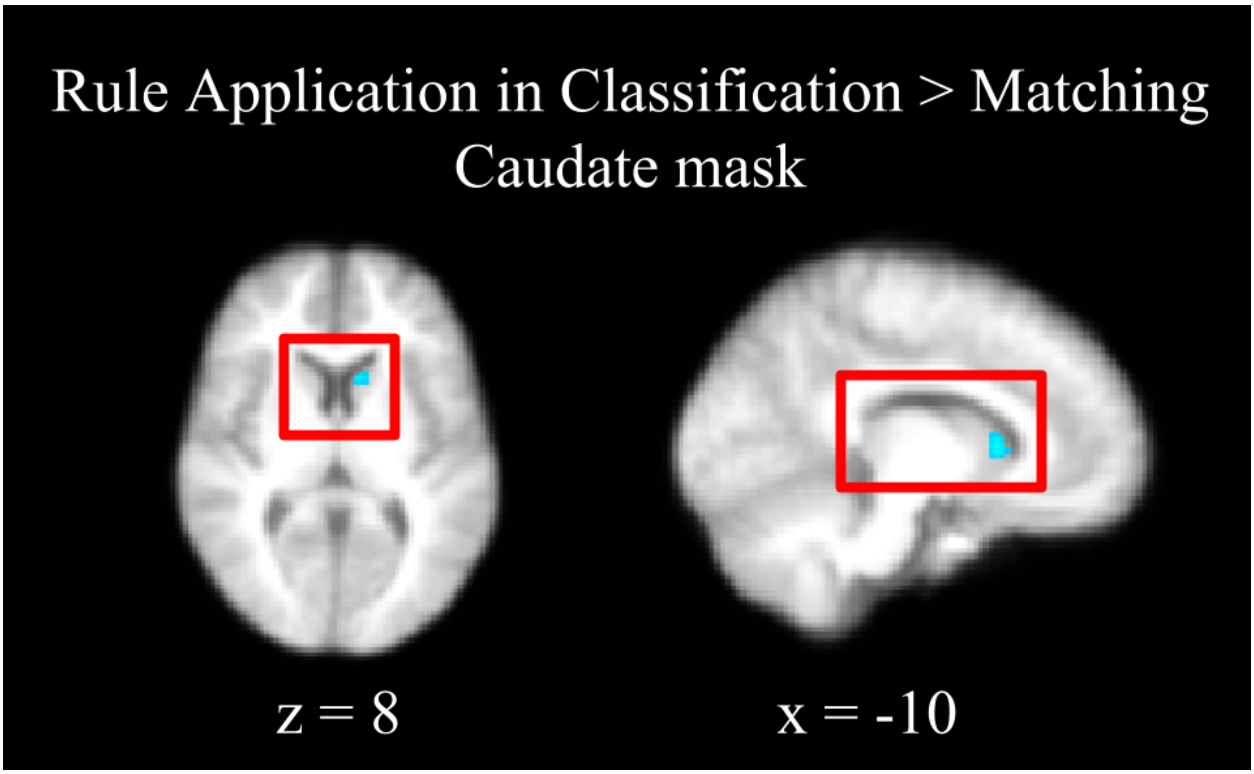


Figure 7. The caudate nucleus was active in rule application for classification vs. matching contrast. The result is constrained to an anatomically defined caudate mask.

Table 5

Connectivity analysis.
A seed from the Frontopolar cortex was used. Caudate nucleus was a region of interest.

Cluster Region(s)	**Cluster Size (voxels), Cluster p-value**	**t-value**	**MNI coordinates (x,y,z)**		
Classification > Matching in Rule Application					
Left Caudate	58, p=0.0416	3.22	-8	16	4
Classification > Matching in Rule Learning					
Left Caudate	385, p=0.0004	5.25	-8	0	10
Right Caudate	282, p=0.0016	6.36	8	2	8

Note: Cluster regions are based on the Harvard-Oxford atlas. The coordinates are in the standard MNI space. The p-values are taken from non-parametric tests produced by Randomise tool in FSL.

4. Discussion

The goal of the current study was to compare integration and switching accounts of FPC function in rule-based category learning. The FPC has been established as a critical brain region for many higher-order cognitive capacities, yet so far theories of category learning have not fully established a role for the FPC. Based on the broader literature, we developed two contrasting accounts suggesting that the FPC is involved in rule switching or in the integration of evidence for rule-based representations. We tested these contrasting accounts by comparing FPC activation during rule-based categorization tasks requiring either matching or classification learning. Although often treated as the same, these two tasks place different demands on integration mechanisms. In matching tasks, a rule can be known with certainty after the participant answers a single question correctly and can be eliminated with certainty with every incorrect answer. Therefore participants will tend to integrate evidence for a rule on the same trials in which they switch between rules. This is not the case in classification tasks, where eliminating a rule can be accomplished in a single trial, but it may be necessary to integrate evidence over several correct trials before the participant fully learns a rule. Given this asymmetry in switching and integration between matching and classification tasks, we were able to isolate integration mechanisms by comparing activation on trials in which participants were applying a rule in matching and classification tasks. Consistent with the hypothesis that the FPC is involved in integration of evidence for a rule, we found that the FPC remained active during rule application in classification learning, but was not active during rule application in matching.

The current results are consistent with other recent studies revealing FPC involvement in category learning. Seger & Cincotta (2006) and Liu et al. (2015) found the FPC was more active

during rule learning compared to rule application in classification-like tasks, and suggested the FPC is a part of a "cognitive" cortico-striatal loop involved in rule-based categorization. The current findings are consistent with this idea but go further in suggesting that there may be critical differences between rule-based learning tasks in terms of the demands they place on these integration mechanisms. This is an important finding as results from matching and classification tasks have often been used interchangeably in the literature on the neural basis of categorization. Our findings suggest that some fundamental contrasts (e.g., rule application vs. learning) can differ markedly between the types of rule-based task employed.

By revealing critical differences between types of rule-based tasks, our results fit well within the recent multiple systems literature that examines whether there are differences in the neural systems recruited between types of categorization tasks (rule-based vs. procedural learning; Ell et al., 2006; Nomura et al., 2007; See Ashby & Ell, 2001; Ashby & Maddox, 2005; Schnyer et al., 2009; Seger and Miller, 2010, for reviews) or even amongst tasks of the same basic type (e.g., multiple prototype learning systems; Ziethamova et al., 2008). Although it is important to note here that we are not suggesting that matching and classification involve wholly separate systems, only that they differentially load on these systems, leading to system-level differences in BOLD activation. Future research attempting to build a comprehensive account of the neural basis of categorization will need to take more care in considering commonalities and differences between types of rule-based task. Consideration of the demands of particular rule-based tasks may also be critical for neuropsychological assessment where matching tasks, like the Wisconsin Card Sorting Test, are currently the most popular type of task, but may not be as diagnostic of frontal integration processes as classification learning.

The present results may also inspire future research on FPC function in animal models. Recent work on FPC function in animal models has demonstrated a critical role for the FPC in rapid, one-shot learning of rules, but not in application of well-learned rules (Boschin et al., 2015). Our matching results are consistent with this role of rapid, one-shot rule learning in that the FPC was engaged up until participants learned the rule, but then activation dropped to baseline during even early application. However, the classification results suggest that the FPC may be engaged even as participants begin to apply the rule, if they need to integrate additional evidence for the correct rule (e.g., rule out remaining alternatives). Putting these results together suggests that FPC can allow one-shot learning, but whether it is engaged for longer-term rule acquisition depends on the demands of the task; one-shot learning is not a necessary characteristic of the system's involvement in learning. Future work with animal models on FPC function would benefit from examining tasks, like our classification task, where evidence for a rule must be integrated across trials.

Comparing to Boschin et al.'s (2015) results with rule-learning in macaques, one limitation of our study is that we did not have any extended application trials in which participants were applying very-well learned rules for which they had already achieved automaticity. Thus it is not possible, within the current data, to establish whether the FPC would remain more engaged for classification learning relative to matching, or if eventually FPC involvement would decrease. Given the rapid shift in FPC involvement during matching, where it was active during learning but declined in activation as soon as participants began applying the rule, it is likely that FPC activation would also decline rather rapidly in classification once the rule was known with certainty. This very brief role for the FPC would be consistent with Boschin's findings from rule learning in macaques as well as other recent studies examining FPC

involvement in category learning. For example, we recently found, in a relational category learning task (Davis et al., 2016), that FPC activation was high early in learning, but in later test phases was only re-engaged when participants needed to generalize the rule to novel relational examples, not for applying the rule to well-learned examples. In the future, it will be important to test whether the FPC exhibits similar trajectories in basic rule-based classification tasks.

Outside of the immediate importance of this work for research on rule-based category learning, the present study adds to a growing literature on FPC function in higher-level cognition. Just as it has in category learning, ascribing a single cognitive function to FPC has been difficult due to its activation in a wide range of tasks (Duncan & Owen, 2000; Gilbert et al., 2006). These roles include hypotheses about switching between representations, whether between rules (Konishi et al., 1998, 2002; Monchi et al., 2001; Strange et al., 2001) or between representations of the internal and external world (Burgess et al., 2007), and representational integration (Nee et al., 2014), but also extend into more distant domains such as meta-cognition (Fleming et al., 2012; 2014) and exploratory behavior in reinforcement learning (Daw, et al., 2006). Perhaps the best current candidate for integrating across literatures comes from the representational literature suggesting that the FPC sits on top of a rostral-caudal hierarchy of representational abstraction (Badre, & D'Esposito, 2009, Badre et al., 2010). So far, this framework has not specified a representational format for FPC representations, but one possibility is that the FPC is involved in forming and evaluating evidence for structured representations that bind object features with logical rules in a predicate-argument structure (e.g., in this case “Category A has thick legs”; see Ramnani & Owen, 2004 for a related argument). This type of representational building and evaluation may help to connect external features or objects symbolically with the internal representations of a task, accounting for external-internal

switching results (Burgess et al., 2007), and be involved in within and between item representational integration accounting for the present results and results from relational integration (Christoff, et al., 2001; Green et al., 2006). This structured representation building hypothesis might also connect across disparate observations within our lab. In previous research (Davis et al., 2016), we found that the FPC was engaged for learning relational categories and integrating new observations into relational categories, but not for simple feature-based categories that were defined by very basic visual features (e.g., angle of Gabor patches). It is possible that some feature-based rules that are based on very elementary perceptual features can be solved with simple perceptual accumulation mechanisms, and thus do not harness FPC, whereas categories defined by novel combinations of higher-level features encourage participants to use a more symbolic, structured representation building strategy. At the top of this hierarchy may be relational tasks that can only be solved by building structured representations that abstract across concrete features. Future research will need to test this hypothesis by examining FPC function across a number of different rule types, and possibly by combining categorization tasks with other measures known to influence FPC function such as metacognitive judgments and reward/payout schedules that encourage exploration.

In conclusion, the present study examines the role of the FPC in category learning and differentiates between two accounts of FPC function that have been discussed in this literature, representational switching and representational integration. To test these accounts, we compared activation during different phases of two types of rule-based tasks, matching and categorization, that differ in their demands on integration. Consistent with the integration hypothesis, the FPC was active for tasks requiring more integration demands, even when switching between rules was not necessary. These results are critical because they help to establish a role for the FPC in

category learning and because they highlight novel differences in the systems engaged for two rule-based tasks that have been used interchangeably in much of the neurobiological literature on category learning thus far. Future research can build on these results by investigating how different types of rule-based representations and task demands impact involvement of FPC and its connection with memory systems involved with rule learning, such as the striatum.

5. References

Alexander, G. E., DeLong, M. R., & Strick, P. L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual Review of Neuroscience*, *9*(1), 357-381.

Ashby, F. G., Alfonso-Reese, L. A., Turken, U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *105*(3), 442-481.

Ashby, F. G., & Ell, S. W. (2001). The neurobiology of human category learning. *Trends in cognitive sciences*, *5*(5), 204-210.

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology*, *56*, 149-178.

Ashby, F. G., & Maddox, W. T. (2011). Human category learning 2.0. *Annals of the New York Academy of Sciences*, *1224*(1), 147-161.

Avants, B. B., Tustison, N., & Song, G. (2009). Advanced normalization tools (ANTS). *Insight Journal*, *2*, 1-35.

Avants, B. B.& Tustison, N. (2014). *ANTs/ANTsR Brain Templates*. Retrieved: 13 34, Dec 27, 2016 (GMT) from: https://dx.doi.org/10.6084/m9.figshare.915436.v1

Badre, D., Kayser, A. S., & D'Esposito, M. (2010). Frontal cortex and the discovery of abstract action rules. *Neuron*, *66*(2), 315-326.

Badre, D., & D'Esposito, M. (2009). Is the rostro-caudal axis of the frontal lobe hierarchical?. *Nature Reviews Neuroscience*, *10*(9), 659-669.

Bates, D., Maechler, M., Bolker, B., Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48. doi:10.18637/jss.v067.i01.

Beckmann, C. F., Jenkinson, M., & Smith, S. M. (2003). General multilevel linear modeling for group analysis in FMRI. *Neuroimage*, *20*(2), 1052-1063.

Boschin, E. A., Piekema, C., & Buckley, M. J. (2015). Essential functions of primate frontopolar cortex in cognition. *Proceedings of the National Academy of Sciences*, *112*(9), E1020-E1027.

Braunlich, K., & Seger, C. A. (2016). Categorical evidence, confidence, and urgency during probabilistic categorization. *NeuroImage*, *125*, 941-952.

Bullmore, E., Horwitz, B., Honey, G., Brammer, M., Williams, S., & Sharma, T. (2000). How good is good enough in path analysis of fMRI data? *NeuroImage*, *11*(4), 289-301.

Bunge, S. A., Helskog, E. H., & Wendelken, C. (2009). Left, but not right, rostrolateral prefrontal cortex meets a stringent test of the relational integration hypothesis. *Neuroimage*, *46*(1), 338-342.

Burgess, P. W., Dumontheil, I., & Gilbert, S. J. (2007). The gateway hypothesis of rostral prefrontal cortex (area 10) function. *Trends in Cognitive Sciences*, *11*(7), 290-298.

Christoff, K., Prabhakaran, V., Dorfman, J., Zhao, Z., Kroger, J. K., Holyoak, K. J., & Gabrieli, J. D. (2001). Rostrolateral prefrontal cortex involvement in relational integration during reasoning. *Neuroimage*, *14*(5), 1136-1149.

Dale, A.M., Fischl, B., Sereno, M.I. (1999). Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage, 9*, 179-194.

Davis, T., Goldwater, M., & Giron, J. (2016). From concrete examples to abstract relations: the rostrolateral prefrontal cortex integrates novel examples into relational categories. *Cerebral Cortex*, bhw099.

Davis, T., Love, B. C., & Preston, A. R. (2012). Striatal and hippocampal entropy and recognition signals in category learning: simultaneous processes revealed by model-based fMRI. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(4), 821.

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, *441*(7095), 876-879.

DeGutis, J., & D'Esposito, M. (2007). Distinct mechanisms in visual category learning. *Cognitive, Affective, & Behavioral Neuroscience*, *7*(3), 251-259.

Deichmann, R., Gottfried, J. A., Hutton, C., & Turner, R. (2003). Optimized EPI for fMRI studies of the orbitofrontal cortex. *Neuroimage*, *19*(2), 430-441.

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., ... & Albert, M. S. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage*, *31*(3), 968-980.

Duncan, J., & Owen, A. M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends in Neurosciences, 23*(10), 475-483.

Ell, S. W., Marchant, N. L., & Ivry, R. B. (2006). Focal putamen lesions impair learning in rule-based, but not information-integration categorization tasks. *Neuropsychologia*, *44*(10), 1737-1751.

Gilbert, S. J., Spengler, S., Simons, J. S., Steele, J. D., Lawrie, S. M., Frith, C. D., & Burgess, P. W. (2006). Functional specialization within rostral prefrontal cortex (area 10): a meta-analysis. *Journal of cognitive neuroscience, 18*(6), 932-948.

Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, A. C., ... & Rapoport, J. L. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proceedings of the National academy of Sciences of the United States of America*, *101*(21), 8174-8179.

Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., & Ghosh, S. S. (2011). Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. *Frontiers in Neuroinformatics*, *5*, 13.

Gray, J. R., Chabris, C. F., & Braver, T. S. (2003). Neural mechanisms of general fluid intelligence. *Nature Neuroscience*, *6*(3), 316-322.

Graybiel, A. M. (2008). Habits, rituals, and the evaluative brain. *Annual. Review of Neuroscience, 31*, 359-387.

Green, A. E., Fugelsang, J. A., Kraemer, D. J., Shamosh, N. A., & Dunbar, K. N. (2006). Frontopolar cortex mediates abstract integration in analogy. *Brain Research*, *1096*(1), 125-137.

Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *Neuroimage*, *48*(1), 63-72.

Grinband, J., Hirsch, J., & Ferrera, V. P. (2006). A neural representation of categorization uncertainty in the human brain. *Neuron*, *49*(5), 757-763.

Heaton, R. K., Chelune, G. J., Curtiss, G., Kay, G. G., & Talley, J. L. (1993). *Wisconsin card sorting test*. Psychological Assessment Resources.

Filoteo, J. V., Maddox, W. T., Simmons, A. N., Ing, A. D., Cagigas, X. E., Matthews, S., & Paulus, M. P. (2005). Cortical and subcortical brain regions involved in rule-based category learning. *Neuroreport*, *16*(2), 111-115.

Fleming, S. M., Huijgen, J., & Dolan, R. J. (2012). Prefrontal contributions to metacognition in perceptual decision making. *Journal of Neuroscience*, *32*(18), 6117-6125.

Fleming, S. M., Ryu, J., Golfinos, J. G., & Blackmon, K. E. (2014). Domain-specific impairment in metacognitive accuracy following anterior prefrontal lesions. *Brain*, *137*(10), 2811-2822.

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, *17*(2), 825-841.

Jenkinson, M., Beckmann, C. F., Behrens, T.E., Woolrich, M.W., Smith, S.M (2012). FSL. *NeuroImage, 62*, 782-790.

Kruschke, J. K. (1992). ALCOVE: an exemplar-based connectionist model of category learning. *Psychological Review*, *99*(1), 22.

Konishi, S., Nakajima, K., Uchida, I., Kameyama, M., Nakahara, K., Sekihara, K., & Miyashita, Y. (1998). Transient activation of inferior prefrontal cortex during cognitive set shifting. *Nature Neuroscience*, *1*(1), 80-84.

Konishi, S., Hayashi, T., Uchida, I., Kikyo, H., Takahashi, E., & Miyashita, Y. (2002). Hemispheric asymmetry in human lateral prefrontal cortex during cognitive set shifting. *Proceedings of the National Academy of Sciences*, *99*(11), 7803-7808.

Liu, Z., Braunlich, K., Wehe, H. S., & Seger, C. A. (2015). Neural networks supporting switching, hypothesis testing, and rule application. *Neuropsychologia*, *77*, 19-34.

Ma, Q., Huang, B., Wang, J., Seger, C., Yang, W., Li, C., ... & Huang, R. (2016). Altered modular organization of intrinsic brain functional networks in patients with Parkinson's disease. *Brain imaging and behavior*, 1-14.

Monchi, O., Petrides, M., Petre, V., Worsley, K., & Dagher, A. (2001). Wisconsin Card Sorting revisited: distinct neural circuits participating in different stages of the task identified by event-

related functional magnetic resonance imaging. *The Journal of Neuroscience*, *21*(19), 7733-7741.

Nee, D. E., Jahn, A., & Brown, J. W. (2014). Prefrontal cortex organization: dissociating effects of temporal abstraction, relational abstraction, and integration with fMRI. *Cerebral Cortex*, *24*(9), 2377-2387.

Nomura, E. M., Maddox, W. T., Filoteo, J. V., Ing, A. D., Gitelman, D. R., Parrish, T. B., ... & Reber, P. J. (2007). Neural correlates of rule-based and information-integration visual category learning. *Cerebral Cortex, 17*(1), 37-43.

Peirce, J.W. (2007). PsychoPy - Psychophysics software in Python. *Journal of Neuroscience Methods, 162*(1-2), 8-13.

Peirce, J.W. (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics, 2*:10. doi: 10.3389/neuro.11.010.2008

Poldrack, R. A., Clark, J., Pare-Blagoev, E. J., Shohamy, D., Moyano, J. C., Myers, C., & Gluck, M. A. (2001). Interactive memory systems in the human brain. *Nature*, *414*(6863), 546-550.

Poldrack, R. A., & Foerde, K. (2008). Category learning and the memory systems debate. *Neuroscience & Biobehavioral Reviews*, *32*(2), 197-205.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ramnani, N., & Owen, A. M. (2004). Anterior prefrontal cortex: insights into function from anatomy and neuroimaging. *Nature Reviews Neuroscience*, *5*(3), 184-194.

Robinson, A. L., Heaton, R. K., Lehman, R. A., & Stilson, D. W. (1980). The utility of the Wisconsin Card Sorting Test in detecting and localizing frontal lobe lesions. Journal of consulting and clinical psychology, 48(5), 605.

Rorden, C., Brett, M. (2000). Stereotaxic display of brain lesions. *Behavioural Neurology, 12*, 191-200.

Schnyer, D. M., Maddox, W. T., Ell, S., Davis, S., Pacheco, J., & Verfaellie, M. (2009). Prefrontal contributions to rule-based and information-integration category learning. *Neuropsychologia*, *47*(13), 2995-3006.

Seger, C. A. (2008). How do the basal ganglia contribute to categorization? Their roles in generalization, response selection, and learning via feedback. *Neuroscience & Biobehavioral Reviews*, *32*(2), 265-278.

Seger, C. A., & Cincotta, C. M. (2006). Dynamics of frontal, striatal, and hippocampal systems during rule learning. *Cerebral cortex*, *16*(11), 1546-1555.

Seger, C. A., & Miller, E. K. (2010). Category learning in the brain. *Annual Review of Neuroscience*, *33*, 203.

Semendeferi, K., Teffer, K., Buxhoeveden, D. P., Park, M. S., Bludau, S., Amunts, K., ... & Buckwalter, J. (2011). Spatial organization of neurons in the frontal pole sets humans apart from great apes. *Cerebral Cortex*, *21*(7), 1485-1497.

Siegel, J. S., Power, J. D., Dubis, J. W., Vogel, A. C., Church, J. A., Schlaggar, B. L., & Petersen, S. E. (2014). Statistical improvements in functional magnetic resonance imaging analyses produced by censoring high-motion data points. *Human Brain Mapping*, *35*(5), 1981-1996.

Smith, E. E., & Grossman, M. (2008). Multiple systems of category learning. *Neuroscience & Biobehavioral Reviews*, *32*(2), 249-264.

Smith, S. M. (2002). Fast robust automated brain extraction. *Human Brain Mapping*, *17*(3), 143-155.

Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage*, *44*(1), 83-98.

Specht, K., Lie, C. H., Shah, N., & Fink, G. R. (2009). Disentangling the prefrontal network for rule selection by means of a non-verbal variant of the Wisconsin Card Sorting Test. *Human Brain Mapping*, *30*(5), 1734-1743.

Spreng, R. N., Stevens, W. D., Chamberlain, J. P., Gilmore, A. W., & Schacter, D. L. (2010). Default network activity, coupled with the frontoparietal control network, supports goal-directed cognition. *Neuroimage*, *53*(1), 303-317.

Strange, B. A., Henson, R. N. A., Friston, K. J., & Dolan, R. J. (2001). Anterior prefrontal cortex mediates rule learning in humans. *Cerebral Cortex*, *11*(11), 1040-1046.

Therneau, T. (2015). coxme: Mixed Effects Cox Models. R package version 2.2-5. https://CRAN.R-project.org/package=coxme

Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M., & Nichols, T. E. (2014). Permutation inference for the general linear model. *Neuroimage*, *92*, 381-397.

Woolrich, M. (2008). Robust group analysis using outlier inference. *Neuroimage*, *41*(2), 286-301.

Woolrich, M. W., Behrens, T. E., Beckmann, C. F., Jenkinson, M., & Smith, S. M. (2004). Multilevel linear modelling for FMRI group analysis using Bayesian inference. *Neuroimage*, *21*(4), 1732-1747.

Woolrich, M. W., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., ... & Smith, S. M. (2009). Bayesian analysis of neuroimaging data in FSL. *Neuroimage*, *45*(1), S173-S186.

Woolrich, M. W., Ripley, B. D., Brady, M., & Smith, S. M. (2001). Temporal autocorrelation in univariate linear modeling of FMRI data. *Neuroimage*, *14*(6), 1370-1386.

Zeithamova, D., Maddox, W. T., & Schnyer, D. M. (2008). Dissociable prototype learning systems: evidence from brain imaging and behavior. *Journal of Neuroscience*, *28*(49), 13194-13201.