(2018). *Psychological Research*, *82*, 371-384.

Trial-by-Trial Switching Between Procedural and Declarative Categorization Systems

Matthew J. Crossley
SRI International

Jessica L. Roeder
University of California, Santa Barbara

Sebastien Helie
Purdue University

F. Gregory Ashby
University of California, Santa Barbara

Considerable evidence suggests that human category learning recruits multiple memory systems. A popular assumption is that procedural memory is used to form stimulus-to-response mappings, whereas declarative memory is used to form and test explicit rules about category membership. The multiple systems framework has been successful in motivating and accounting for a broad array of empirical observations over the last 20 years. Even so, only a couple of studies have examined how the different categorization systems interact. Both previous studies suggest that switching between explicit and procedural responding is extremely difficult. But they leave unanswered the critical questions of whether trial-by-trial system switching is possible, and if so, whether it is qualitatively different than trial-by-trial switching between two explicit tasks. The experiment described in this article addressed these questions. The results 1) confirm that effective trial-by-trial system switching, although difficult, is possible; 2) suggest that switching between tasks mediated by different memory systems is more difficult than switching between two declarative memory tasks; and 3) point to a serious shortcoming of current category-learning theories.

Introduction

Evidence that humans have multiple memory systems (Eichenbaum & Cohen, 2001; Squire, 2004; Tulving & Craik, 2000) inspired the development of theories that category learning is also mediated by multiple qualitatively distinct systems. (Ashby, Alfonso-Reese, et al., 1998; Ashby & O'Brien, 2005). According to this view, procedural memory is used to form many-to-one stimulus-to-response mappings (i.e., S-R associations), whereas declarative memory is used to apply rules and test explicit hypotheses about category membership. This arrangement raises a number of important questions as to how these putative systems resolve their competition for access to the motor systems that they must share. For example, given a daily need to perform a variety of tasks – some best served by declarative systems, and others best served by procedural systems – can control be flexibly passed between systems on a moment-by-moment basis?

ATRIUM (Erickson & Kruschke, 1998) and COVIS (Ashby et al., 1998), the two dominant multiple system category-learning theories, each assume that trial-by-trial switching is a routine and common occurrence. However, both theories were formulated in the absence of *any* data on this important issue. Unfortunately, during the ensuing 18 years, the landscape has only marginally changed. We know of only two studies that directly address this issue (Ashby & Crossley, 2010; Erickson, 2008). Both studies used experiments that required participants to switch between procedural and declarative categorization strategies on a trial-by-trial basis in order to achieve optimal performance. Ashby and Crossley (2010) reported that only 2 of 53 participants ($\sim$ 4%) showed any evidence of trial-by-trial switching, whereas Erickson (2008), using a design that included more switching cues, reported that only 51 of 170 participants ($\sim$ 30%) successfully switched between systems on a trial-by-trial basis.

The poor success rates reported by Erickson (2008) and Ashby and Crossley (2010) suggest that the current theories might be much too optimistic about the ability of people to system switch, and therefore that a more valid and conservative theory of system switching is badly needed. Constructing such a theory on the basis of these two studies seems fruitless however, because too many critical questions remain unanswered. For example, why did the two studies find such different success rates? Can trial-by-trial switching between systems ever be reliably achieved? If so, what conditions trigger a system switch? Is switching between declarative and procedural systems qualitatively different than switching between two tasks both mediated by declarative systems?

The primary goal of this article is to address these questions. The experiment described below included two conditions. In one, participants were required to switch between declarative and procedural strategies on a trial-by-trial basis following a training procedure that was similar to the one used by Erickson (2008). Both prior studies estimated the

number of participants that successfully switched between systems by using decision bound model fits to count how many participants were able to adopt strategies of the optimal type. The present experiment extends this method by adding an additional behavioral probe at the end of the experiment to test whether switching was successful. A second condition replicated the first, except participants were instead required to switch between two different declarative strategies on a trial-by-trial basis. Our results suggest that trial-by-trial switching between declarative and procedural systems is possible given enough training and under optimal conditions, and that switching between declarative and procedural strategies is more difficult than switching between different declarative strategies.

A secondary goal of this article is to relate system switching to the large task-switching literature, which has been primarily concerned with switching back and forth between different declarative-memory-based tasks (e.g., Kiesel et al., 2010; Monsell, 2003). Many such studies have established that switch trials reliably increase response times (RTs) and often decrease accuracy. The properties of the component tasks that determine switch costs are of increasing interest in this field. For example, some of the factors that have been explored include the number and identity of response effectors (Philipp, Weidner, Koch, & Fink, 2013), the complexity of the stimuli (Witt & Stevens, 2013), the abstractness of the rules (Stelzel, Basten, & Fiebach, 2011), and the perceptual and attentional demands of the component tasks (Chiu & Yantis, 2009; Nagahama et al., 2001; Ravizza & Carter, 2008; Rushworth, Hadland, Paus, & Sipila, 2002). This article is the first to compare task switching (i.e., between two declarative-memory tasks) and system switching (between a declarative- and a procedural-memory task), and therefore makes an important contribution to the task-switching and cognitive-control literatures, in addition to the category-learning literature.

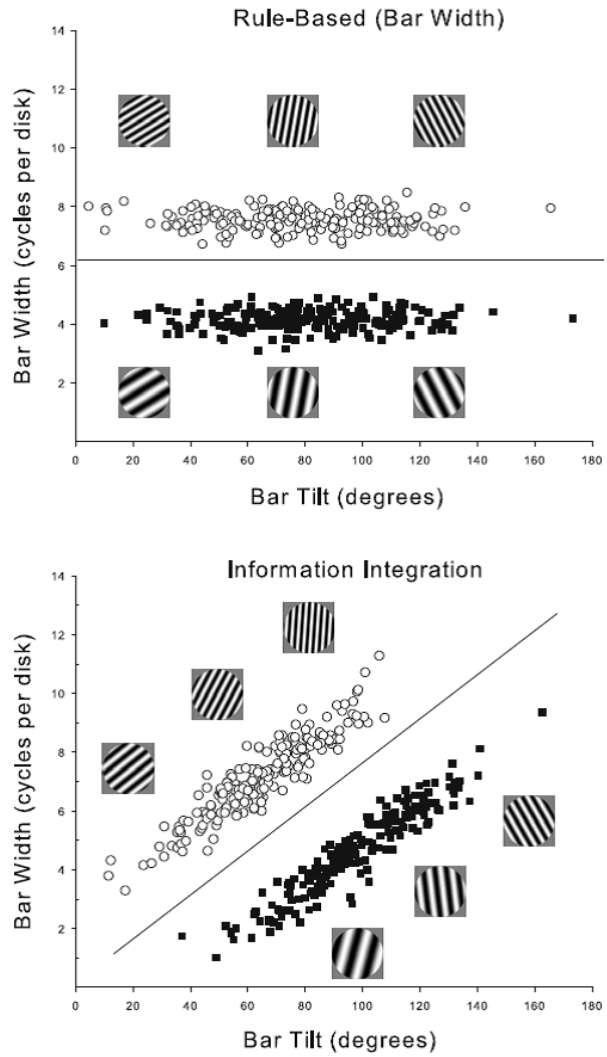

Figure 1. Examples of one-dimensional RB (top) and II (bottom) stimuli and category structures.

Rule-Based and Information-Integration Category Learning

The current and previous research on system switching during categorization depends strongly on prior research with rule-based (RB) and information-integration (II) category-learning tasks. Example trials from the present experiment and example RB and II category structures are shown in Figure 1. 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 categories. However, there is no requirement that RB tasks be one-dimensional (1D). For example, a conjunction rule (e.g., respond 'A' if the bars are thick and the orientation is shallow) is an RB task because a conjunction is a pair of logical conditionals, and thus, separate 1D rules are first made about each relevant dimension and then these separate decisions are combined. In II tasks, accuracy is maximized only if information from two or more incommensurable stimulus dimensions is integrated perceptually at a pre-decisional stage (Ashby & 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 because they produce a maladaptive focus on only one stimulus dimension.

A variety of evidence suggests that success in RB tasks depends on working memory and executive attention (Ashby et al., 1998; Maddox, Ashby, Ing, & Pickering, 2004; Waldron & Ashby, 2001; Zeithamova & Maddox, 2006), and is

supported by a broad neural network that includes the prefrontal cortex, anterior cingulate, the head of the caudate nucleus, and medial temporal lobe structures (Brown & Marsden, 1988; Filoteo, Maddox, Song, et al., 2007; Muhammad, Wallis, & Miller, 2006; Seger & Cincotta, 2006). In contrast, evidence suggests that success in II tasks depends on procedural learning that is mediated largely within the striatum (Ashby & Ennis, 2006; Filoteo, Maddox, Salmon, & Song, 2005; Knowlton, Mangels, & Squire, 1996; Nomura et al., 2007). For example, switching the locations of the response keys has no effect on RB categorization, but as in more traditional procedural-learning tasks, switching response keys interferes with II categorization (Ashby, Ell, & Waldron, 2003; Maddox, Ashby, et al., 2004; Maddox, Glass, O'Brien, Filoteo, & Ashby, 2010).

The stimuli and category structures used in our experiment are illustrated in Figure 2. Note that there are two conditions. The RB/II condition required trial-by-trial switching between II and 1D RB categories, whereas the RB/RB condition required switching between two RB category structures – one that requires a conjunction rule for optimal performance and one that requires a 1D rule.

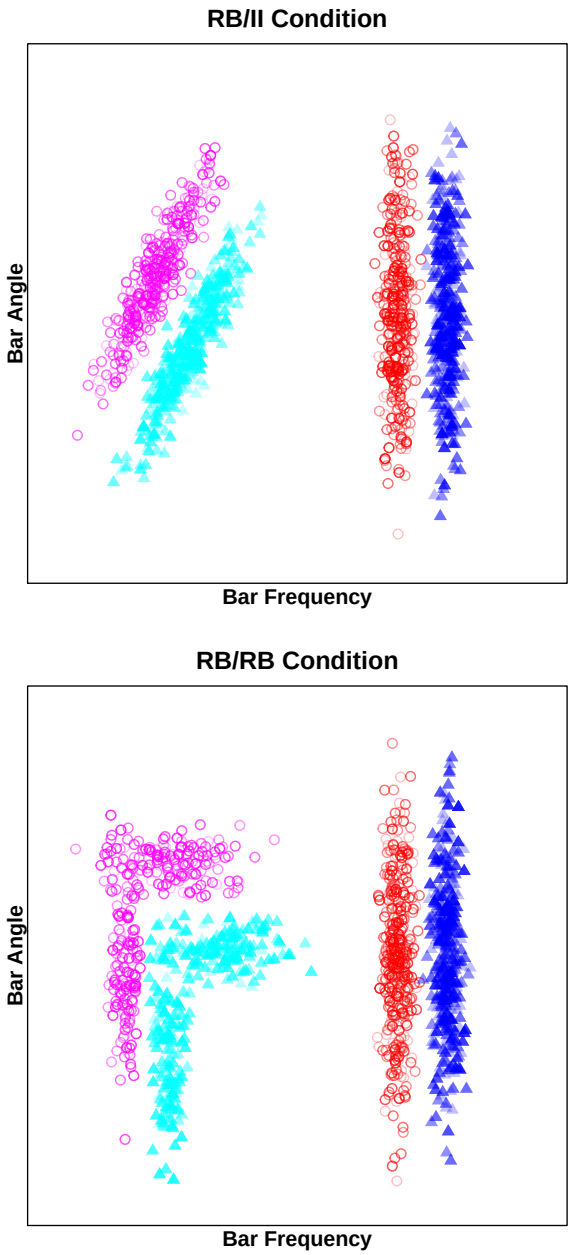

Figure 2. Stimuli and Category Structures used in the RB/II (top panel) and RB/RB Conditions (bottom panel).

A comparison of Ashby and Crossley (2010) and Erickson (2008)

As mentioned previously, the only two behavioral studies to examine system switching during categorization reported somewhat discrepant results. Ashby and Crossley (2010) reported an almost complete failure to find any evidence of trial-by-trial switching, whereas Erickson (2008) reported that 30% of his participants appeared to switch successfully between systems on a trial-by-trial basis.

Ashby and Crossley (2010) used circular sine-wave gratings like those shown in Figure 1 with a hybrid category structure that required a procedural strategy for half the stimuli, and a 1D rule for the other half. A 1D rule was optimal when the bars had a steep orientation and a procedural strategy was optimal when the orientation was shallow. Thus, the only cue that signaled which type of strategy to use was bar orientation. In contrast, Erickson (2008) included three cues that signaled whether a declarative or procedural strategy was required. First, the stimuli requiring a procedural strategy were perceptually distinct from the stimuli requiring an explicit rule. Second, stimuli requiring a procedural strategy were presented in one color, whereas stimuli requiring a rule were presented in a different color. Third, the II categories required different responses than the RB categories (i.e., A and B versus C and D).

One possibility is that Erickson (2008) observed more trial-by-trial switching because of the extra cues that he used. Another possibility, however, is that Erickson's participants did not actually switch between different memory systems. Instead, perhaps they were able to perform well by switching between two different declarative strategies. This possibility is difficult to rule out because the stimuli used by Erickson (2008) were constructed from commensurable stimulus dimensions (height of a rectangle and the horizontal position of an internal vertical line segment). When two stimulus dimensions are in the same units, then diagonal decision bounds are often easy to describe verbally and therefore easy to discover through an explicit, logical reasoning process. For example, consider rectangles that vary in height and width. In this

case, a diagonal bound with slope +1 defines a shape rule. When the bound has an intercept of zero, then all rectangles above the boundary are taller than they are wide, and all rectangles below the boundary are wider than they are tall. Following this example, it is possible that Erickson's (2008) participants used a rule based on the difference between the height of the rectangle and the distance from the internal vertical line to the left edge (for example) of the rectangle. The appropriate category response would be chosen depending on whether or not this difference exceeded a criterion. In general, when stimuli are constructed from commensurable stimulus dimensions it is often difficult to determine whether explicit or procedural strategies are used from an accuracy or model-fit analysis alone.

Testing for Successful Switching Between Systems

Ashby and Crossley (2010) and Erickson (2008) attempted to diagnose successful system switching by analyzing block-by-block accuracy and decision-bound model fits. While each of these techniques makes an important contribution, neither is sufficient to prove system switching conclusively. In the experiments described below, we added a test block after training that reversed the locations of the response keys (which we henceforth refer to as a button-switch). Previous research suggests that a button switch impairs procedural strategies more than declarative strategies (Ashby et al., 2003; Maddox, Bohil, & Ing, 2004; Maddox et al., 2010). Theoretically, this is because procedural learning is mediated by S-R associations that were gradually strengthened through trial and error. Reversing the buttons then requires unlearning of the original S-R associations, and relearning the new reversed associations[1]. Declarative strategies, on the other hand, can be quickly adapted to accommodate a button switch because performance in this case is driven by explicitly applied rules. Therefore, if participants are successfully switching between declarative and procedural strategies, then the button switch should impair trials that require a procedural strategy, but not trials that require a declarative strategy.

There is, however, evidence that button switches can also impair sufficiently complex declarative strategies (Nosofsky, Stanton, & Zaki, 2005). Thus, any impairment that occurs as a result of a button switch could conceivably be due to the use of a complex declarative strategy, rather than a procedural strategy. We therefore ran a control condition (the RB/RB condition) in which the II structures were replaced with complex RB structures in order to specifically examine this possibility. If button switch impairments in our switching task are due to the use of complex declarative strategies, then they should also be present in this condition. If not, then any button switch impairment observed in the RB/II condition is likely due to procedural learning.

Methods

Participants & Conditions

Thirty-four undergraduates at UCSB served as participants in the RB/II condition, and 22 served as participants in the RB/RB condition. All participants were given course credit for their participation, and they all had normal or corrected to normal vision.

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 either a blue or a green background using routines from the Psychophysics toolbox (Brainard, 1997).

Stimuli were sampled from one of four possible distributions (illustrated in the top panel of Figure 2 for the RB/II condition, and the bottom panel of Figure 2 for the RB/RB condition) following the randomization technique developed by Ashby and Gott (1988). To control for statistical outliers, the random sample was discarded if its Mahalanobis distance (Fukunaga, 1990) was greater than 3.0. 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. After each sample was collected, the coordinates of all stimuli were linearly transformed so that the sample statistics exactly equaled the population parameter values. Each random sample (x, y) was converted to a stimulus according to the nonlinear transformations defined by Treutwein, Rentschler, and Caelli (1989), which roughly equates the salience of each dimension (see Appendix B for details).

Procedure

The procedures were identical in both conditions. Each condition consisted of one session lasting approximately 50 minutes in duration that included 9 blocks of 100 trials each. Participants were free to rest as long as they wished between blocks. Participants were required to classify a stimulus into one of four categories on every trial. Stimuli sampled from the 1D RB categories were displayed against a blue background, and stimuli sampled from the II categories or the conjunction-rule RB categories were displayed against a green background. Participants were informed that the background colors indicated that different categorization strategies would be necessary for optimal performance. They were

[1]This account actually predicts that a button switch should cause catastrophic interference in procedural strategies, which has never been observed in an II task. We recently proposed a revision to this classic account that correctly predicts a more moderate button-switch interference (Cantwell, Crossley, & Ashby, 2015). The key point, however, is simply that button switches reliably interfere with II tasks more robustly than they interfere with RB tasks

Table 1
Category Distribution Parameters.

	μ_x	μ_y	σ_x	σ_y	cov_{xy}
RB/II Condition					
II A	43	57	167.91	119.0	59.36
II B	57	43	167.91	119.0	59.36
RB C	140	50	10	200	0
RB D	160	50	10	200	0
RB/RB Condition					
Conjunction A1	31.33	48.67	11.4	171.8	0
Conjunction B1	48.67	31.33	11.4	171.8	0
Conjunction A2	51.33	68.67	171.8	11.4	0
Conjunction B2	68.67	51.33	171.8	11.4	0
1-D C	140	50	10	200	0
1-D D	160	50	10	200	0

further informed that stimuli displayed against a blue background (1D RB trials) only required attention to one dimension and that stimuli displayed against a green background (II and conjunction-rule trials) required attention to both dimensions. They were instructed to press the 's' key with the second finger of their left hand for category 'A', to press the 'd' key with the first finger on their left hand for category 'B', to press the 'k' key with the first finger on their right hand for category 'C', and to press the 'l' key with the second finger on their right hand for category 'D'. Participants were further informed that all stimuli displayed against a blue background belonged to either category 'A' or category 'B' and that stimuli displayed against a green background belonged to either category 'C' or category 'D'.

Each trial began with a fixation cross lasting 750 ms. A stimulus was then presented for a maximum duration of 5000 ms. If the participant responded within 5000 ms the stimulus disappeared, and 500 ms later a feedback tone was presented for 1000 ms. Correct responses were indicated by a pure sine tone (500 Hz, .73 seconds in duration), and incorrect feedback was indicated by a saw-tooth tone (200 Hz Hz, 1.22 seconds in duration).

Participants were first trained on the 1D RB categories for 100 trials, then on the II (RB/II condition) or conjunction-rule RB categories (RB/RB condition) for 400 trials, and then on randomly intermixed (with equal probability) RB and II categories for 300 trials in the RB/II condition or on randomly intermixed 1D RB and conjunction-rule RB categories for 300 trials in the RB/RB condition. Each condition concluded with 100 trials of intermixed RB and II categories (RB/II condition) or 1D rule and conjunction-rule categories (RB/RB condition) with the response key-category label mappings switched. Specifically, the category A and B response keys switched locations, and so did the category C and D response keys. 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 correct keyboard key - category label mapping. Thus, when the button locations were switched, so were the labels.

Decision Bound Modeling

To identify participants most likely to have switched successfully between declarative and procedural strategies, we partitioned the data from each participant into blocks of 100 trials, isolated and grouped the trials according to their respective category substructure (i.e., II or RB) and fit different decision bound models to the responses from each substructure (Ashby & Gott, 1988; Maddox & Ashby, 1993). Three different kinds of models were fit to each of these data sets. Rule-learning models assumed either a 1D rule (on either orientation or bar width) or a conjunction rule (respond 'B' if the bars are wide and the orientation is shallow; otherwise respond 'B'). The 1D rule models have two free parameters (a decision criterion on the relevant perceptual dimension, and a perceptual noise variance), and the conjunction rule model has three free parameters (a separate decision criterion on each dimension, and a perceptual noise variance). Procedural-learning models assumed a linear decision bound of arbitrary slope and intercept. These models are consistent with a procedural strategy since they integrate perceptual information from the two stimulus dimensions pre-decisionally. Procedural-learning models have three free parameters (the slope and intercept of the linear decision bound, and a perceptual noise variance). The third model class assumed a guessing strategy. One version assumed unbiased guessing (no free parameters), and another version (with one free parameter) assumed biased guessing (guess A with probability p and guess B with probability $1 - p$, where p is a free parameter).

We estimated best-fitting parameters via maximum likelihood, and used the 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 data given the model. 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. As in Erickson (2008), only participants whose responses during the last block of intermixed trials (i.e., trials 701-800) were best fit by a model that assumed a strategy of the optimal type were classified as 'switchers.'

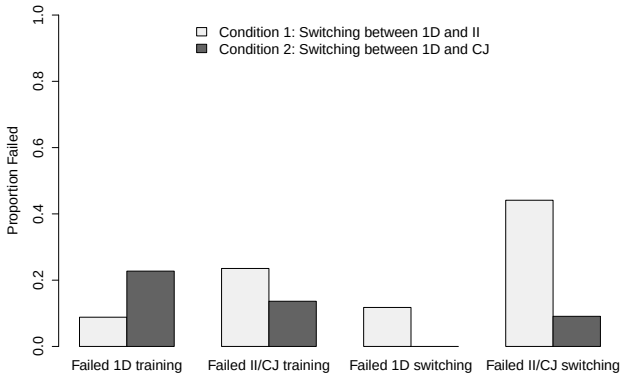

Figure 3. Proportion of participants failing to avoid the exclusion criteria during each phase of the experiment. (1D = one-dimensional RB, CJ = conjunction-rule RB)

Results

Exclusion Criteria

Since we are interested in system switching, it is essential that we identify and remove participants who failed to learn during any of the single category-structure training phases. Obviously, it is only possible to study the ability of people to switch between categorization strategies in participants who can learn each strategy separately. So the statistic we used to screen participants was mean accuracy during the last block of single category-structure training. We separately analyzed our data with exclusion criteria of 55%, 60%, 65%, and 70% correct. The results were qualitatively identical for each of these criteria, although some of the statistics that were significant for the more stringent criteria were nonsignificant for the more lenient criteria. Here we report results based on an exclusion criterion of 65% correct because this value reflected a fairly natural break point that seemed to best separate learners from nonlearners. Figure 3 shows the proportion of participants in both conditions that failed to reach this criterion level of accuracy (i.e., at least 65% during the single category-structure training, or during the blocks where the different category structures were intermixed).

The proportion of participants who failed on II trials in the RB/II condition during the switching phase was significantly greater than the proportion of participants who failed on conjunction-rule RB trials in the RB/RB condition [$\chi^2(1) = 13.09, p = 0.00, h = 1.64$]. None of the other differences between conditions shown in Figure 3 are significant [1D training: $\chi^2(1) = 0.74, p = 0.39, h = -0.47$; II / CJ training: $\chi^2(1) = 0.94, p = 0.33, h = 0.51$; 1D switching: $\chi^2(1) = 1.79, p = 0.18, h = 1.01$]. All participants that failed any task element were excluded from further analyses. This left 17 of the 34 participants in the RB/II condition and 13 of the 22 participants in the RB/RB condition.

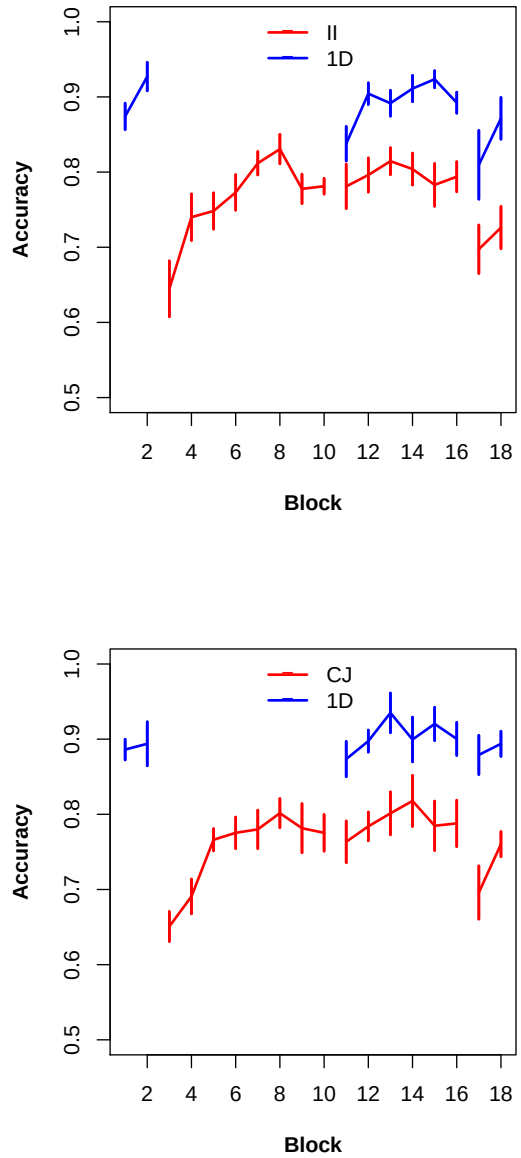

Figure 4. Mean accuracy of non-excluded participants in each block of 50 trials. Error bars are SEMs. (1D = one-dimensional, CJ = conjunction rule)

Accuracy-Based Analyses

Figure 4 shows mean accuracy for every block of 50 trials in both conditions. Recall that participants were first trained for 100 trials on the 1D RB categories, followed by 400 trials either on II categories (RB/II condition) or conjunction-rule RB categories (RB/RB condition). In both conditions, the

single category-structure training was followed by 300 trials where stimuli from the two category structures were intermixed. Finally, the experiment concluded with 100 more intermixed trials, with the response keys switched within each category structure.

One-Dimensional RB Training. The 1D RB categories were learned well within the first training block as indicated by a non-significant effect of block [$F(1,28) = 2.74, p = 0.11, \Omega = 0.61$], and equally well in both conditions as indicated by a non-significant effect of condition [$F(1,28) = 0.22, p = 0.64, \Omega = 0.05$], and a non-significant condition × block interaction [$F(1,28) = 1.52, p = 0.23, \Omega = 0.34$].

II and Conjunction-Rule Training. The II and conjunction-rule categories were learned with practice, as indicated by a significant effect of block [$F(7,196) = 12.36, p = 0.00, \Omega = 0.95$]. They were matched in difficulty as indicated by a non-significant effect of condition [$F(1,28) = 0.32, p = 0.58, \Omega < 0.01$], and a non-significant interaction [$F(7,196) = 0.59, p = 0.77, \Omega = 0.05$].

Intermixed Performance. Performance on 1D RB trials remained considerably better than performance on either the II or conjunction-rule trials during the intermixed phase, as indicated by a significant main effect of trial type [$F(1,308) = 149.25, p = 0.00, \Omega = 0.88$]. Performance on both trial types improved equally well with practice, as indicated by a significant main effect of block [$F(5,308) = 2.54, p = 0.03, \Omega = 0.07$], and non-significant interactions [condition × block: $F(5,308) = 0.16, p = 0.98, \Omega < 0.01$; condition × trial type: $F(1,308) = 0.89, p = 0.34, \Omega = 0.01$; block × cue: $F(5,308) = 0.75, p = 0.59, \Omega = 0.02$; condition × block × cue: $F(5,308) = 0.59, p = 0.71, \Omega = 0.02$].

Button Switch Performance. Figure 5 shows button-switch costs for all trial types and conditions. In the RB/II condition, the cost on 1D trials was significant during the first button-switch block [$t(16) = 2.23, p < 0.05 >, d = 1.24$], but not during the second button-switch block [$t(16) = 1.27, p = 0.22, d = 0.41$]. The cost on II trials was significant during the first [$t(16) = 3.67, p < 0.01, d = 3.37$], and the second [$t(16) = 2.95, p < 0.05, d = 2.18$] button-switch blocks. The cost on II trials was not significantly greater than the cost on 1D trials during the first button-switch block [$t(16) = 0.21, p = 0.42, d = 0.01$], but was marginally greater during the second button-switch block [$t(16) = 1.51, p = 0.08, d = 0.57$].

In the RB/RB condition, the cost on 1D trials was not significant during the first [$t(12) = 1.62, p = 0.13, d = 0.76$], or the second [$t(12) = 1.17, p = 0.26, d = 0.40$] button-switch block. The cost on conjunction-rule trials was significant during the first button-switch block [$t(12) = 2.41, p = 0.03, d = 1.68$], but not during the second button-switch block [$t(12) = 1.23, p = 0.24, d = 0.43$]. The cost on conjunction trials was marginally significantly greater than the cost on 1D trials during the first [$t(12) = 1.45, p = 0.09, d = 0.60$], but not the second [$t(12) = 0.46, p = 0.33, d = 0.06$] button-switch block.

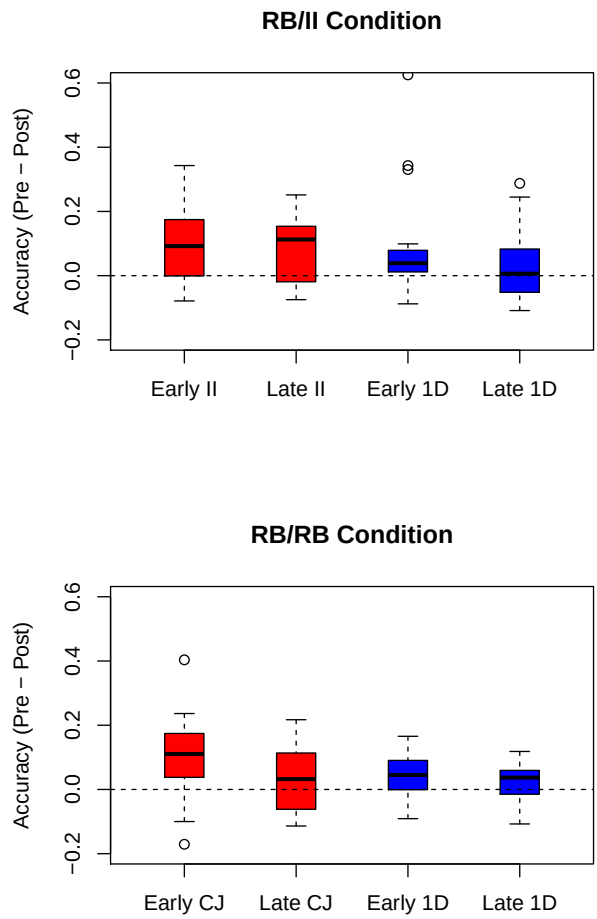

Figure 5. Button-switch costs in non-excluded participants. The solid black lines inside each box represent the median, and each box extends from the 25th to the 75th percentile. Each whisker extends to the most extreme data point that is within 1.5 times the interquartile range from the median. Circles represent outliers that are further away from the median than this. Top) RB/II Condition. Bottom) RB/RB Condition. 'Early' refers to the difference in accuracy between the last 50 trials of the intermixed phase and the first 50 trials of the button-switch phase. 'Late' refers to the accuracy difference between the last 50 trials of the intermixed phase and the last 50 trials of the button-switch phase.

Note that the II button-switch cost in the RB/II condition and the conjunction-rule button-switch cost in the RB/RB condition were similar during the first button-switch block, and the cost to each decreased during the second button-switch block. Even so, the recovery on II trials (RB/II condi-

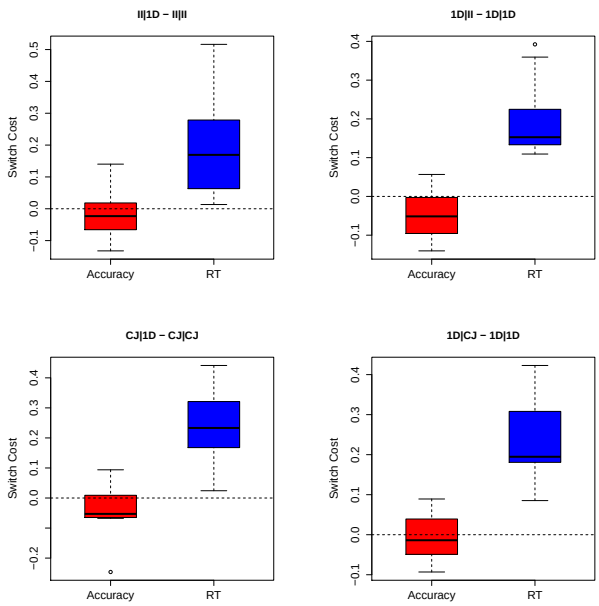

Figure 6. Trial-by-trial switch costs for accuracy (percent correct) and mean RT (seconds). The solid black lines inside each box represent the median, and each box extends from the 25th to the 75th percentile. Each whisker extends to the most extreme data point that is within 1.5 times the interquartile range from the median. Circles represent outliers that are further away from the median than this.

tion) was only partial, whereas complete recovery occurred on conjunction-rule trials (RB/RB condition). However, if participants are using a procedural system to respond to II trials and a declarative system to respond to conjunction-rule trials, then we would expect the cost incurred on II trials to be significantly greater than the cost incurred on conjunction-rule trials. Our data displayed this pattern qualitatively, but failed to reach significance: the recovery during conjunction-rule trials was not significantly greater than the recovery during II trials [$t(20) = -0.74, p = 0.23, d = 0.12$].

Trial-By-Trial Switch Cost. The task switching literature has more or less ubiquitously reported switch costs in the form of decreased accuracy and / or increased response times (RTs) on switch trials relative to stay trials (Monsell, 2003; Wylie & Allport, 2000). Here, we examine whether the switch cost incurred when switching to a procedural system from a declarative system differs from the switch cost incurred from switching the opposite direction.

Every stimulus was either from 1D RB categories, conjunction-rule RB categories, or II categories. Therefore, let J|K denote the event in which the stimulus on the current trial is from type J catgories and the stimulus from the preceding trial was from type K categories, for J and K = 1D, CJ (for conjunction-rule RB), or II. In the RB/II condition, the four trial types are II|II, II|1D, 1D|II, and 1D|1D (corresponding to II stay, II switch, RB switch, and RB stay trials, respectively), whereas in the RB/RB condition, the four trial types are CJ|CJ CJ|1D, 1D|CJ, and 1D|1D. The trial-by-trial switch costs are therefore defined as II|1D − II|II and 1D|II − 1D|1D in the RB/II condition and CJ|1D − CJ|CJ and 1D|CJ − 1D|1D in the RB/RB condition.

Figure 6 shows the trial-by-trial accuracy and mean RT switch costs of all four types. There was a reliable RT switch cost for every switch type in both conditions [RB/II condition: II|1D − II|II: $t(16) = 5.42, p < 0.001, d = 7.34$; 1D|II − 1D|1D: $t(16) = 9.06, p < 0.001, d = 20.51$; RB/RB condition: CJ|1D − CJ|CJ: $t(12) = 6.76, p < 0.001, d = 13.19$; 1D|CJ − 1D|1D: $t(12) = 7.74, p < 0.001, d = 17.28$]. The accuracy switch cost was highly significant when switching to 1D from II [1D|II − 1D|1D: $t(16) = -3.48, p < 0.001, d = 3.03$], and it was marginally significant when switching to a conjunction rule from 1D [CJ|1D − CJ|CJ: $t(12) = -1.92, p = 0.08, d = 1.06$]. The other two types of switch costs were not significant [II|1D − II|II: $t(16) = -0.86, p = 0.40, d = 0.19$; 1D|CJ − 1D|1D: $t(12) = -0.66, p = 0.52, d = 0.13$].

All prior evidence has indicated that system switching is difficult (Ashby & Crossley, 2010; Erickson, 2008), which might seem to suggest that the between-system switch costs in the RB/II condition should be greater than the within-system switch costs in the RB/RB condition. Our results provided only weak support for this prediction. The accuracy cost of switching to a 1D rule was greater in the RB/II condition than in the RB/RB condition [$t(26) = -1.76, p = 0.04, d = 0.61$], but the RT cost was not [$t(22) = -1.18, p = 0.88, d = 0.30$], and the costs of switching from a 1D rule to an II strategy were not significantly different from the costs of switching from a 1D rule to a conjunction rule [Accuracy: $t(26) = 0.86, p = 0.80, d = 0.14$; RT: $t(28) = -0.63, p = 0.73, d = 0.08$].

Model-Based Analyses

Figure 7 shows the number of participants whose responses were best fit by each type of decision bound model for every block in both conditions. Recall that during the first (100 trial) block in both conditions, participants exclusively practiced the 1D categories, during blocks 2 – 5 they exclusively practiced either the II categories (in the RB/II condition) or the conjunction-rule categories (in the RB/RB condition), during blocks 6 – 8 they switched back and forth between 1D and II (RB/II condition) or between 1D and a conjunction rule (RB/RB condition), and in block 9 they continued to switch back and forth except with the response keys switched.

Visual inspection of Figure 7 shows clearly that during the exclusive training blocks the vast majority of participants responded in a manner consistent with the optimal strategy for

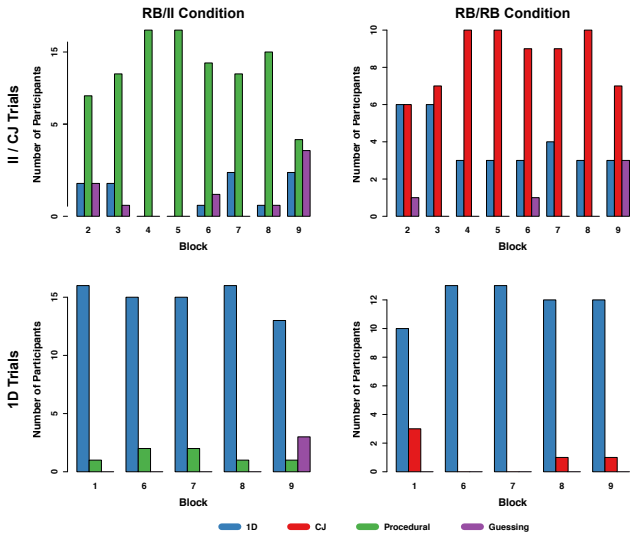

Figure 7. The number of participants best fit by each model type during each block in both conditions. The left column shows RB/II results and the right column shows RB/RB results. The top row shows performance on II and conjunction-rule trials, and the bottom row shows performance on 1D trials.

all category structures in both conditions (i.e., 1D rule use dominates during 1D training, procedural strategies dominate during II training, and conjunction rule use dominates during CJ training). For 3 of the 4 category structures, optimal strategy use was unaffected by trial-by-trial switching (i.e., during the intermixed blocks 6 – 8). The one exception was on II trials in the RB/II condition. During the early intermixed blocks (6 and 7), a few participants abandoned their procedural strategies to either guess or use a 1D rule. By the last intermixed block however (i.e., block 8), all but 2 were using a procedural strategy again.

The button switch (during block 9) had no effect on strategy use during 1D trials in the RB/RB condition. Three participants abandoned an optimal-type strategy in favor of guessing both during the conjunction rule trials in the RB/RB condition and during 1D trials in the RB/II condition. However, neither of these reductions was significant [1D users RB/II condition: $\chi^2(1) = 0.94, p = 0.17, h = 0.52$; conjunction rule users RB/RB condition: $\chi^2(1) = 0.68, p = 0.20, h = 0.49$]. On the other hand, the button switch had a more serious effect on categorization strategies during II trials in the RB/II condition. In fact, the number of participants using a procedural strategy dropped by more than half, which is a significant reduction [$\chi^2(1) = 6.31, p = 0.01, h = 1.05$]. Of the 8 participants who abandoned procedural strategies, 5 resorted to guessing and 3 switched to a 1D rule.

Discussion

Ashby and Crossley (2010) reported an almost complete failure of system switching in a straightforward categorization task in which perfect accuracy was possible if participants used a simple 1D categorization rule for disks with steep orientations and a procedural strategy for disks with shallow orientations. This abysmal performance was unexpected given that a number of studies have shown that participants readily learn a variety of nonlinear decision bounds that are at least as complex as the decision bound in the Ashby and Crossley (2010) experiment (Ashby, Waldron, Lee, & Berkman, 2001; Maddox & Ashby, 1993). One possible key difference though is that in the previous studies the complex bound had no 1D component (horizontal or vertical). Thus, participants were never consistently rewarded for using an explicit strategy on a significant subset of trials. As a result, the best interpretation of those earlier studies may be that participants responded via the procedural system on every trial.

Erickson (2008) reported a higher success rate at trial-by-trial system switching than Ashby and Crossley (2010), using a design that included a number of cues that signaled whether each stimulus required a declarative or procedural strategy. Even so, only about 30% of Erickson's participants showed evidence of successful switching, and even this value may be an over-estimation because the stimuli used by Erickson (2008) were constructed from commensurable stimulus dimensions, which sometimes make it difficult to identify procedural responding.

Thus, in summary, only a couple of prior studies have investigated trial-by-trial system switching, and those studies paint a bleak picture. Both studies suggest that switching between explicit and procedural responding is extremely difficult. But they leave unanswered a number of critical questions. Is effective trial-by-trial system switching possible? If so, is it qualitatively different than trial-by-trial switching between two explicit tasks? The experiment described in this article addressed these questions. In the RB/II condition, participants attempted to trial-by-trial switch between an explicit 1D rule and a nonverbalizable similarity-based strategy that depends on procedural learning and memory. In the RB/RB condition, different participants trial-by-trial switched between two different explicit rules – the same 1D rule as in the RB/II condition and a conjunction rule that was approximately equal in difficulty to the procedural strategy required in the RB/II condition.

Half of our RB/II participants performed well, and they did so in a manner consistent with system switching – that is, their performance was consistent with the hypothesis that declarative systems mediated performance on 1D trials and procedural systems mediated performance on II trials. First, the responses of almost all of these participants were best accounted for by a 1D explicit rule on RB trials and by a

model assuming a procedural strategy on II trials. Second, accuracy on II trials was initially impaired more than accuracy on 1D trials during the button-switch phase. Third, the button-switch impairment on 1D trials fully recovered during the second button-switch block, whereas the impairment on II trials never recovered. Fourth, in the RB/RB condition, there was no button-switch impairment at all on the 1D trials, and the initial impairment on conjunction-rule trials fully recovered during the second button-switch block. These latter two points are consistent with the hypothesis that declarative systems mediated performance on all trials in the the RB/RB condition and control was passed back and forth between procedural and declarative systems in the II/RB condition. The key idea here is that initial button-switch costs may reflect a plethora of processes indicative of either declarative or procedural processes. For example, working memory demands and procedural interference will both be high soon after a button-switch. However, we suggest that these working memory demands should ease off with relative ease as participants get used to the reverse mappings. Procedural interference, on the other hand, requires the gradual rewiring of associations formed through trial-and-error, and should therefore be considerably harder to adapt to the reversed mappings.

The task switching literature has been primarily concerned with switching back and forth between different declarative-memory-based tasks (e.g., Kiesel et al., 2010; Monsell, 2003), and has now examined a variety of factors that influence this process (see our introduction for some of these factors). This literature indicates that switch trials reliably increase response times (RTs) and often decrease accuracy. Our article is the first to compare task switching (i.e., between two declarative-memory tasks) and system switching (between a declarative- and a procedural-memory task). Our results suggest that switching between tasks mediated by different memory systems is more difficult than switching between two declarative-memory tasks. Several results support this conclusion. First, more RB/II than RB/RB participants failed to meet the meager accuracy criterion of 65% correct during the intermixed training phase (see Figure 3). Second, more RB/II than RB/RB participants abandoned a strategy of the optimal type during intermixed training (see Figure 7). Third, the trial-by-trial switch costs were slightly though significantly greater in the RB/II condition than in the RB/RB condition (i.e., the accuracy cost of switching to a 1D rule was significantly greater in the RB/II condition).

Our results have important theoretical implications. All current category-learning models that include multiple systems assume that trial-by-trial system switching is a routine and common occurrence. For example, COVIS (Ashby et al., 1998) assumes that control is passed back and forth between systems depending on which system is most confident on each trial. Similarly, ATRIUM (Erickson & Kruschke, 1998, p. 119) assumes that 'each module learns to classify those stimuli for which it is best suited'. Our results, together with those of Erickson (2008) and Ashby and Crossley (2010), suggest that system switching is much more difficult that assumed by such models, and therefore that some significant revisions of existing multiple systems models are in order.

In hindsight, the assumption of effortless trial-by-trial system switching made by models such as COVIS and ATRIUM might now seem unrealistic. Even so, at the time these theories were proposed, no relevant data existed that would allow a more accurate model of system switching to be constructed, and the assumption of effortless switching was easy to implement computationally. In 1998, the primary focus was on establishing that humans have multiple category-learning systems, not on building an accurate model of how control is passed back-and-forth between those putative and at that time, hypothetical systems. After nearly two decades of research directed at this primary focus, the time finally seems propitious to direct attention at the second-generation question of system switching. Building a more accurate model of system switching, however, requires an empirical database. We believe that our results represent a significant step in this direction, and for this reason, that the present article fills a critical void in the literature.

References

Ashby, F. G., Alfonso-Reese, L. A., et al. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *105*(3), 442.

Ashby, F. G., & Crossley, M. J. (2010). Interactions between declarative and procedural-learning categorization systems. *Neurobiology of learning and memory*, *94*(1), 1–12.

Ashby, F. G., Ell, S. W., & Waldron, E. M. (2003). Procedural learning in perceptual categorization. *Memory & Cognition*, *31*(7), 1114–1125.

Ashby, F. G., & Ennis, J. M. (2006). The role of the basal ganglia in category learning. *Psychology of Learning and Motivation*, *46*, 1.

Ashby, F. G., & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(1), 33.

Ashby, F. G., & O'Brien, J. B. (2005). Category learning and multiple memory systems. *TRENDS in Cognitive Science*, *2*, 83-89.

Ashby, F. G., Waldron, E. M., Lee, W. W., & Berkman, A. (2001). Suboptimality in human categorization and identification. *Journal of Experimental Psychology: General*, *130*(1), 77.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial vision*, *10*, 433–436.

Brown, R., & Marsden, C. (1988). Internal versus external cues and the control of attention in parkinson's disease. *Brain*, *111*(2), 323–345.

Cantwell, G., Crossley, M. J., & Ashby, F. G. (2015). Multiple stages of learning in perceptual categorization: evidence and neurocomputational theory. *Psychonomic Bulletin & Review*, 1–16.

Chiu, Y.-C., & Yantis, S. (2009). A domain-independent source of cognitive control for task sets: shifting spatial attention and switching categorization rules. *The Journal of Neuroscience*, *29*(12), 3930–3938.

Eichenbaum, H., & Cohen, N. J. (2001). *From conditioning to conscious recollection: Memory systems of the brain.* Oxford University Press.

Erickson, M. A. (2008). Executive attention and task switching in category learning: Evidence for stimulus-dependent representation. *Memory & Cognition*, *36*(4), 749–761.

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General*, *127*(2), 107.

Filoteo, J. V., Maddox, W. T., Salmon, D. P., & Song, D. D. (2005). Information-integration category learning in patients with striatal dysfunction. *Neuropsychology*, *19*(2), 212.

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*(2), 305–320.

Fukunaga, K. (1990). *Introduction to statistical pattern recognition*. Academic press.

Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., & Koch, I. (2010). Control and interference in task switchingâĂŤa review. *Psychological Bulletin*, *136*(5), 849–874.

Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science*, *273*(5280), 1399–1402.

Maddox, W. T., & Ashby, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception & Psychophysics*, *53*(1), 49–70.

Maddox, W. T., Ashby, F. G., Ing, A. D., & Pickering, A. D. (2004). Disrupting feedback processing interferes with rule-based but not information-integration category learning. *Memory & Cognition*, *32*(4), 582–591.

Maddox, W. T., Bohil, C. J., & Ing, A. D. (2004). Evidence for a procedural-learning-based system in perceptual category learning. *Psychonomic Bulletin & Review*, *11*(5), 945–952.

Maddox, W. T., Glass, B. D., O'Brien, J. B., Filoteo, J. V., & Ashby, F. G. (2010). Category label and response location shifts in category learning. *Psychological Research PRPF*, *74*(2), 219–236.

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, *7*(3), 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. *Journal of cognitive neuroscience*, *18*(6), 974–989.

Nagahama, Y., Okada, T., Katsumi, Y., Hayashi, T., Yamauchi, H., Oyanagi, C., … Shibasaki, H. (2001). Dissociable mechanisms of attentional control within the human prefrontal cortex. *Cerebral Cortex*, *11*(1), 85–92.

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

Nosofsky, R. M., Stanton, R. D., & Zaki, S. R. (2005). Procedural interference in perceptual classification: Implicit learning or cognitive complexity? *Memory & Cognition*, *33*(7), 1256–1271.

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 Mapping*, *34*(8), 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*(12), 2924–2935.

Rushworth, M., Hadland, K., Paus, T., & Sipila, P. (2002). Role of the human medial frontal cortex in task switching: a combined fmri and tms study. *Journal of Neurophysiology*, *87*(5), 2577–2592.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, *6*(2), 461–464.

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

Squire, L. R. (2004). Memory systems of the brain: a brief history and current perspective. *Neurobiology of learning and memory*, *82*(3), 171–177.

Stelzel, C., Basten, U., & Fiebach, C. J. (2011). Functional connectivity separates switching operations in the posterior lateral frontal cortex. *Journal of Cognitive Neuroscience*, *23*(11), 3529–3539.

Treutwein, B., Rentschler, I., & Caelli, T. (1989). Perceptual spatial frequency—orientation surface: psychophysics and line element theory. *Biological cybernetics*, *60*(4), 285–295.

Tulving, E., & Craik, F. I. (2000). *The oxford handbook of memory*. Oxford University Press.

Waldron, E. M., & Ashby, F. G. (2001). The effects of concurrent task interference on category learning: Evidence for multiple category learning systems. *Psychonomic Bulletin & Review*, *8*(1), 168–176.

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

Wylie, G., & Allport, A. (2000). Task switching and the measurement of âĂIJswitch costsâĂİ. *Psychological Research*, *63*(3-4), 212–233.

Zeithamova, D., & Maddox, W. T. (2006). Dual-task interference in perceptual category learning. *Memory & cognition*, *34*(2), 387–398.

Appendix: Nonlinear Stimulus Transform

This appendix describes the method we used to generate spatial frequency, orientation (f, o) pairs that define our stimuli. Spatial frequency, f, values carry units of cycles per degree of visual angle, and orientation values carry units of radians. First, (x, y) pairs with arbitrary units were generated from a bivariate uniform distribution on the interval $(0, 100)$ for each dimension. Next, these (x, y) pairs were linearly transformed into (x_T, y_T) pairs to span the interval $(-1, 2)$ on

dimension x_T and $(\frac{\pi}{11}, \frac{3\pi}{8} + \frac{\pi}{11})$ on dimension y_T via

$$x_T = \frac{3x}{100} - 1 \tag{A.1}$$

$$y_T = \frac{3\pi}{8}\frac{y}{100} + \frac{\pi}{11} \tag{A.2}$$

Next, x_T values were mapped to spatial frequency values, f, via

$$f = 2^{x_T} \tag{A.3}$$

We used a multistep procedure to convert y_T values in o values. First, we collected and sorted in ascending order all y_T values into a vector $\mathbf{y}_s$. From $\mathbf{y}_s$, we defined new vectors

$$\mathbf{z} = 4.7 \sin^2 \mathbf{y}_s \tag{A.4}$$

and

$$\mathbf{y}_{T_2} = \frac{\mathbf{y}_{T_2}(n-1) + \sqrt{(\mathbf{y}_s^2(n) - \mathbf{y}_s^2(n))^2 + (\mathbf{z}^2(n) - \mathbf{z}^2(n))^2}}{(\max \mathbf{y}_{T_2} - \min \mathbf{y}_{T_2})(\max \mathbf{y}_s - \min \mathbf{y}_s)} - \min \mathbf{y}_{T_2} + \min \mathbf{y}_s \tag{A.5}$$

where the n terms reference the nth element of the corresponding vector, and

$$\mathbf{y}_{T_2}(1) = \sqrt{\mathbf{y}_s^2(1) + \mathbf{z}^2(1)} \tag{A.6}$$

Finally, the elements of $\mathbf{y}_{T_2}$ were returned to their original sort order and recombined into (f, o) pairs.

Author Notes

The research described in this article was supported in part by grant number 2R01MH063760 from NIMH and by AFOSR grant FA9550-12-1-0355.