(2016). *Cognition*, *154*, 22-33.

What is Automatized during Perceptual Categorization?

Jessica L. Roeder & F. Gregory Ashby
University of California, Santa Barbara

An experiment is described that tested whether stimulus-response associations or an abstract rule are automatized during extensive practice at perceptual categorization. Twenty-seven participants each completed 12,300 trials of perceptual categorization, either on rule-based (RB) categories that could be learned explicitly or information-integration (II) categories that required procedural learning. Each participant practiced predominantly on a primary category structure, but every third session they switched to a secondary structure that used the same stimuli and responses. Half the stimuli retained their same response on the primary and secondary categories (the congruent stimuli) and half switched responses (the incongruent stimuli). Several results stood out. First, performance on the primary categories met the standard criteria of automaticity by the end of training. Second, for the primary categories in the RB condition, accuracy and response time (RT) were identical on congruent and incongruent stimuli. In contrast, for the primary II categories, accuracy was higher and RT was lower for congruent than for incongruent stimuli. These results are consistent with the hypothesis that rules are automatized in RB tasks, whereas stimulus-response associations are automatized in II tasks. A cognitive neuroscience theory is proposed that accounts for these results.

Introduction

There is now abundant evidence that declarative and procedural memory systems both contribute to perceptual category learning (e.g., Ashby & Maddox, 2005, 2010; Love, Medin, & Gureckis, 2004; Reber, Gitelman, Parrish, & Mesulam, 2003). Much of this evidence comes from rule-based (RB) and information-integration (II) category-learning tasks. In RB tasks, the categories can be learned via some explicit reasoning process (Ashby, Alfonso-Reese, Turken, & Waldron, 1998). In the most common applications, only one stimulus dimension is relevant, and the participant's task is to discover this relevant dimension and then to map the different dimensional values to the relevant categories. In II tasks, accuracy is maximized only if information from two or more incommensurable stimulus components is integrated at some predecisional stage (Ashby & Gott, 1988; Ashby et al., 1998).

Figure 1 shows typical examples of RB and II tasks. In both cases, the two categories are composed of circular sine-wave gratings that vary in the width and orientation of the dark and light bars. The solid lines denote the category boundaries. Note that a simple verbal rule perfectly partitions the categories in the RB task, but no verbal rule correctly separates the two categories in the II task. A variety of evidence suggests that success in RB tasks depends on declarative memory systems and especially on working memory and executive attention (Ashby et al., 1998; Maddox, Ashby, Ing, & Pickering, 2004; Waldron & Ashby, 2001; Zeithamova & Maddox, 2006), whereas 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).

Although many studies have reported empirical dissociations between RB and II learning, much less is known about the automatic performance of RB and II categorization decisions. The few available studies have failed to find any qualitative differences between automatic RB and II categorization (Hélie, Roeder, & Ashby, 2010; Hélie, Waldschmidt, & Ashby, 2010; Waldschmidt & Ashby, 2011), and as a result Ashby and Crossley (2012) tentatively proposed that there may be only one neural system for mediating automatic behaviors. This article describes an extensive behavioral experiment that reports the first known difference between automatic RB and II categorization. In particular, we report evidence that stimulus-response (SR) associations are automatized in II tasks, whereas rules are automatized in RB tasks.

In previous studies of RB and II automaticity conducted in our lab, every participant was trained either on an II category structure or on one of two different RB structures for almost 14,000 trials each, distributed over 23 sessions (Hélie, Roeder, & Ashby, 2010; Hélie, Waldschmidt, & Ashby, 2010; Waldschmidt & Ashby, 2011). Many of these participants completed four of these sessions inside an MRI scanner (sessions 1, 4, 10, and 20 for RB participants and sessions 2, 4, 10, and 20 for II participants). Although differences between the RB and II tasks were apparent during the early sessions of these experiments, by session 13, almost all of differences had disappeared.

Helie et al. (2010) reported that after the third session, al-

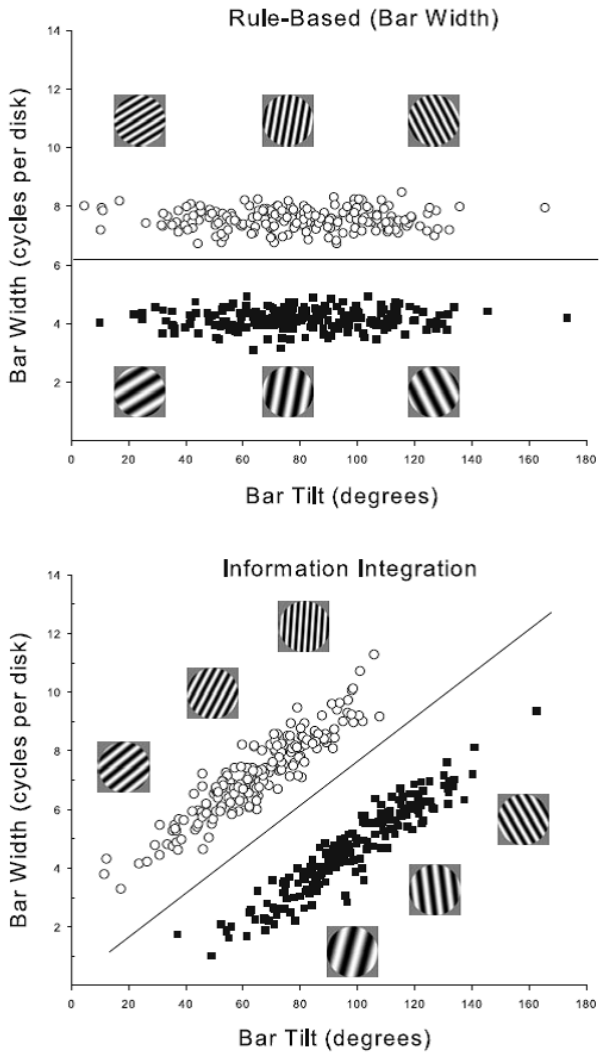

Figure 1. Examples of rule-based and information-integration category structures. Each stimulus is a sine-wave disk that varies across trials in bar tilt (or orientation) and bar width. For each task, three illustrative Category A and Category B stimuli are shown. The small rectangles and open circles denote the specific values of all stimuli used in each task. In the rule-based task, only bar width carries diagnostic category information, so the optimal strategy is to respond with a one-dimensional bar width rule (thin versus thick). In the information-integration task, both dimensions carry useful but insufficient category information. The optimal strategy requires integrating information from both dimensions in a way that is impossible to describe verbally.

most no significant behavioral differences could be discerned among any of the groups. More specifically, all groups showed similar accuracy levels and similar response times (RTs), and the performance of all participants in every condition was best described by a model emulating an optimal decision strategy. In addition, after more than 10,000 trials of practice, switching the location of the response keys produced interference in all conditions (on both accuracy and RT), and there was almost no recovery from this interference over the course of 600 trials. Similarly, after RB and II categorization became automatic, there was no dual-task interference in either task. Thus, although switching the response keys interferes much more with early II than RB performance (Ashby, Ell, & Waldron, 2003; Crossley, Paul, Roeder, & Ashby, in press; Maddox, Bohil, & Ing, 2004; Spiering & Ashby, 2008) and a dual task interferes much more with early RB than II performance (Waldron & Ashby, 2001; Zeithamova & Maddox, 2006), these differences disappear after automaticity has been achieved.

The neuroimaging results also showed convergence. In early training sessions, activation patterns for the RB and II tasks were qualitatively different. For example, RB performance was correlated with activation in PFC, the hippocampus, and the head of the caudate nucleus (Helie et al., 2010), whereas early II training depended heavily on the putamen (Waldschmidt & Ashby, 2011). By session 20 however, activation in all of these areas no longer correlated with performance. Instead, only cortical activation (e.g., in premotor cortex) was positively correlated with response accuracy.

Thus, although much behavioral and neuroscience evidence suggests that early RB and II learning are mediated by different cognitive and neural systems, there is less data on automatic RB and II categorization, and the data that is available has failed to find any qualitative differences between the two. But what sort of differences should be expected? Although we know of no theory that offers an answer to this question, one possibility is that different types of information are automatized in RB and II tasks. For example, the evidence is good that initial II learning is of SR associations, whereas initial learning in RB tasks is of abstract rules (Ashby & Waldron, 1999; Casale, Roeder, & Ashby, 2012; Smith et al., 2015). These data suggest that one plausible hypothesis is that SR associations are automatized in II tasks, whereas the rule is automatized in RB tasks.

Even so, a number of results in the literature suggest that the development of RB automaticity could plausibly be characterized by a transition from rule learning to SR learning. First, comparative neuroimaging analyses suggest that as automaticity develops, task-related activation in RB and II tasks becomes more similar (Hélie, Roeder, & Ashby, 2010; Soto, Waldschmidt, Helie, & Ashby, 2013; Waldschmidt & Ashby, 2011). Second, early in learning, switching the locations of the response keys interferes with II categorization much

more than with RB categorization (Ashby et al., 2003; Crossley et al., in press; Maddox, Bohil, & Ing, 2004; Spiering & Ashby, 2008), which is consistent with the hypothesis that early RB learning is of abstract rules, whereas early II learning is of SR associations. However, after automaticity develops, RB and II categorization are equally disrupted by a switch of the response keys (Hélie, Waldschmidt, & Ashby, 2010). Because of these results, it seems almost equally plausible that the development of automaticity in RB tasks is characterized by a gradual switch from rules to SR associations.

In summary, the existing data seem to suggest that SR associations are automatized in II tasks, although to our knowledge this prediction has never been tested. For RB tasks, no strong prediction is possible – either rules or SR associations could be automatized. This article describes the results of an extensive experiment that provides the first known examination of these issues. As we will see, our results provide the first known evidence that SR associations are automatized during II categorization and rules are automatized during RB categorization. Thus, this article is also the first to report a qualitative difference in automatic RB and II categorization.

In the experiment described below, 27 participants each completed 12,300 categorization trials distributed over 21 separate II or RB training sessions. Each participant learned the primary and secondary category structures shown in Figure 2. There were 7 three-day cycles. During the first 6 of these cycles, participants practiced the primary category structure for the first two sessions and the secondary structure during the third session. During the 7th cycle, they again practiced the primary category structures for the first two sessions. During the third and final session (i.e., day 21), they repeated categorization on the primary category structure, but this time they also completed a simultaneous dual task known to recruit executive attention and working memory (a numerical Stroop task described below). This session was used to assess whether participants were categorizing the primary categories automatically. Note that prior to the dual-task session, each participant had completed 14 training sessions on the primary structures, which is more training than Helie et al. (2010) indicated was necessary for automaticity to develop.

It is important to note that the stimuli that were used in each of the four category structures shown in Figure 2 are identical, and thus, the only difference among any of the conditions is the boundary that separates the two categories. Therefore, the RB and II tasks are exactly equated on category size, instance variability, category separation and a priori perceptual difficulty. In addition, note that the RB category structures switch back and forth between the two stimulus dimensions on primary and secondary days, so as in the II condition, both dimensions are relevant in the RB condition, albeit at different times. Therefore, any difference between

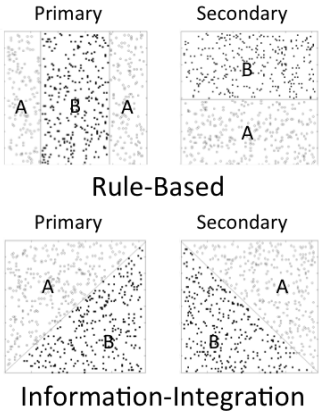

Figure 2. Rule-based and information-integration category structures used in the current experiment. Each subject alternated between two training sessions on the primary structures and one session on the secondary structures. This pattern was repeated 7 times for a total of 21 separate training sessions.

the conditions should be due to the nature of learning, rather than to a confound inherent to the category structures.

Note from Figure 2 that in both the RB and II conditions, exactly half of the stimuli maintain their same SR mapping on primary and secondary days and the other half switch responses. Thus, if SR associations are automatized then training on the secondary structures should delay automaticity for stimuli in the incongruent regions of stimulus space (i.e., the regions where the responses switch on secondary days) more than in the congruent regions. However, if the categorization rule (or decision boundary) is automatized rather than SR associations, then the secondary training should slow the development of automaticity on the primary structures equally for all stimuli.

In the II condition, we expect SR associations to be automatized, so training on the secondary structures should delay automaticity for stimuli in the incongruent regions of stimulus space. If SR associations are also automatized in RB tasks then a similar delaying of automaticity should occur for stimuli in the incongruent regions of stimulus space. However, if rule are automatized then the secondary structures should not interfere with the development of automaticity for any stimuli.

Methods

Participants

Twenty participants were recruited from the Santa Barbara community to participate in the RB condition, and thirteen were recruited to participate in the II condition. Three participants in the RB condition were excluded from final analyses: one person because he or she had not learned the

primary categories by session 10, and two people who did not complete the experiment due to the extended nature of the study. Three people were also excluded from analyses in the II condition because they failed to learn the primary categories by session 10. Thus, the data analyses were based on 17 participants in the RB condition and 10 participants in the II condition. These sample sizes were selected based on similar previous automaticity experiments (i.e., Waldschmidt & Ashby, 2011; Hélie, Roeder, & Ashby, 2010). Participants were paid $10 a session for their participation, for a total of $210 for those participants who completed the experiment.

Stimuli and Apparatus

The stimuli were circular sine-wave gratings presented on 21-inch monitors (1280 × 1024 resolution). All stimuli had the same size, shape and contrast, and differed only in bar width (as measured by cycles per degree of visual angle or cpd) and bar tilt (measured in degrees counterclockwise rotation from horizontal). The bar width and tilt values for each category were generated from sets of points (x_1, x_2) sampled from a 100×100 stimulus space, then converted to perceptual space using the equations $x_1^* = 0.1x_1 + 0.25$ and $x_2^* = \frac{\pi}{200}x_2$. The stimuli were generated with MATLAB using Brainard's (1997) Psychophysics Toolbox, and subtended an approximate visual angle of 5°. The order in which the stimuli were presented was randomized across participants and sessions.

For the primary disjunctive categories used in the RB condition (described in Figure 2), Category A was defined as $x_1 < 2.75$ or $x_1 > 7.75$ cpd. Category B was defined as the rest of the space, or $2.75 < x_1 < 7.75$ cpd. For the secondary one-dimensional categories (also described in Figure 2), Category A stimuli were defined by $x_2 > 0.8$, and Category B stimuli satisfied $x_2 < 0.8$. The relevant dimension was switched for the one-dimensional condition, as this neatly allows for 50% overlap of the category structures.

The stimuli in the II condition were identical to those in the RB condition, only the category boundaries were different. In the primary II condition, the categories were separated by the decision bound $x_2 = x_1$, with stimuli belonging to category A if $x_2 < x_1$, and category B if $x_2 > x_1$. The secondary categories were separated by the orthogonal decision bound $x_2 = -x_1$. See Figure 2 for a description of both category structures.

Stimulus presentation, feedback, response recording and RT measurement were acquired and controlled using MATLAB on a Macintosh computer. Responses were given on a standard QWERTY keyboard; the "d" and "k" keys had large "A" and "B" labels placed on them, respectively. The response keys were not counter-balanced across subjects; category "A" was always the "d" key and category "B" was always the "k" key. Auditory feedback was given for correct and incorrect responses made within a 5-second time limit. If a key was pressed that was not one of the marked response keys, a distinct tone played and the screen displayed the words "wrong key." If the response was made outside of the time limit, the words "too slow" appeared on the screen. In both the wrong key and the too slow cases, the trial was terminated with no category-related feedback, and these trials were excluded from analysis.

Procedure

The experiment lasted for 21 sessions over as close to 21 consecutive workdays as possible (participants were encouraged to participate in sessions over the weekend, and were discouraged from going more than three days without participating). There were two types of sessions: for the majority of the sessions, participants were trained on the primary categories (either the disjunctive or the $y = x$ boundaries, depending on the condition), but every third session of the experiment participants were trained on the secondary categories (either the one-dimensional or the $y = -x$ boundary). All of the sessions except for the 21^{st} consisted of 600 trials with rest opportunities every 50 trials.

Sessions 1-20: At the beginning of the first session, participants were informed they would be participating in a computer-based experiment in which the goal was to learn to categorize objects. They were then informed that they would view a series of disks one at a time, and that they would have to determine whether each disk belonged to category A or category B based on the disk's visual appearance. They were also informed that the disks varied across trials on two visual features: the thickness and angle of the black and white bars.

At the beginning of the third session participants were instructed that the disks were now grouped into categories A and B in a new way, and that they were still to use the "A" and "B" response keys. From session three on they received a prompt at the beginning of the day telling them whether they were seeing the "usual" categories or the "other" categories.

Session 21 was described to participants as a "test day." This session included a numerical Stroop task on each trial, which roughly doubled the amount of time each trial took to complete. Therefore session 21 consisted of 300 trials divided into six 50-trial blocks, instead of the usual 600 trials. The same numerical Stroop dual task described in Waldron and Ashby (2001) was used. In this task, two different digits were displayed on either side of the monitor screen for 200 msecs, then the screen went blank for 100 msecs before the categorization stimulus appeared. One of the digits was displayed in a larger font than the other. After the subject had categorized the disk and received feedback on their categorization judgment, on a random half of the trials a prompt appeared on the screen that asked which side of the screen the digit was on that was larger in numerical value. On the other half of trials the prompt asked which side of the screen the digit was on that was larger in font size. The response keys

for the Stroop task were the "r" and "u" keys marked L and R, respectively, in the same manner as the A and B response keys. Visual feedback was given. Participants were given five practice trials before the beginning of the experiment, and these were not recorded. Participants were explicitly instructed to be as accurate as possible on the number task, and to categorize the "usual" categories with whatever cognitive resources they had left. Feedback was given on every trial for the Stroop task and for the categorization task separately.

Assessing Automaticity

As almost any cognitive or motor skill is practiced, accuracy increases, execution time decreases, and less cognitive control is needed. These changes can occur over very long time scales. For example, Crossman (1959) reported that factory workers were still improving their cigar-rolling performance after a million trials of practice. Thus, the transition from newly learned behavior to automatic skill is a continuum. As practice moves a behavior along this continuum, it is common at some point to label it 'automatic'. Many different criteria have been proposed for identifying 'automatic' behaviors (Ashby & Crossley, 2012; Moors & De Houwer, 2006; Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). Unfortunately, these various criteria do not all necessarily agree, and many can yield discrepant results – especially when applied to behaviors learned via different memory systems (Ashby & Crossley, 2012). As a result, we do not take a strong theoretical position on the point at which a behavior should be considered automatic. It seems certain that the categorization behaviors studied in this article would continue to change if we had continued training our participants past the 21 sessions in our study.

As a result, we define automaticity operationally. First, we assumed that any manipulation that slows learning will slow the development of automaticity. All researchers agree that overall accuracy and response time (RT) should asymptote before a behavior is considered automatic, so any manipulation that delays that asymptote should also delay the development of automaticity. Second, we assumed that the 14 sessions of training that our participants received on the primary category structures was enough to meet automaticity criteria commonly adopted by other researchers.

Our choice to include 14 sessions depended heavily on the extensive behavioral and neuroimaging results of Hélie, Waldschmidt, and Ashby (2010), Hélie, Roeder, and Ashby (2010), and Waldschmidt and Ashby (2011), who examined the development of automaticity in tasks almost identical to the RB and II tasks studied here. As mentioned earlier, their conclusion was that by standard criteria, RB and II categorization were 'automatic' after 13 sessions of training. This is the reason that we included 14 sessions of training on the primary category structures in both the RB and II conditions. Even so, we can test the validity of this assumption using several other well-known criteria. First, we can assess whether overall accuracy and RT have asymptoted by the end of training. This test is more important for accuracy than for RT. Our participants were given instructions that encouraged them to maximize accuracy and no mention was made of RT. In the absence of explicit RT instructions, observed RTs are often highly variable and therefore difficult to interpret (e.g., Luce, 1986). Second, perhaps the most widely used criterion in cognitive science is that a behavior should be considered automatic if it can be executed successfully while the participant is simultaneously engaged in some other secondary task (i.e., so its execution requires little cognitive control; Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). The present experiment tested this criterion by adding a dual task on session 21. If the categorization behaviors are automatic then the drop in accuracy caused by the dual should be minimal in both the RB and II conditions.

In summary, the development of automaticity is a gradual process that can take as many as 10,000 hours of deliberate practice (Ericsson, Krampe, & Tesch-Römer, 1993). The present experiment studies categorization behaviors after much less practice, so our behaviors – like virtually all other behaviors that are acquired in laboratory settings – are not fully automatized. Even so, as we will see, the categorization behaviors we study meet multiple criteria that are currently accepted as sufficient to classify the behavior operationally as 'automatic'.

Decision Bound Modeling

We used decision bound modeling to determine what type of strategy participants used to categorize the stimuli. Decision bound models assume that participants partition the perceptual space into response regions (Maddox & Ashby, 1993). On every trial, the participant determines which region the percept is in and then gives the associated response. Three different types of models were fit to each participant's responses: models assuming an explicit rule-learning strategy, models assuming a procedural-learning strategy, and models that assume random guessing.

There were three different types of rule-learning models: 1) models that assumed a one-dimensional rule (i.e., either on bar width or bar orientation); 2) models that assumed a conjunction rule; and 3) models that assumed a disjunction rule. Models assuming a one-dimensional rule had two free parameters (a decision criterion on the single relevant dimension and a perceptual noise variance), whereas the conjunction and disjunction models had three free parameters (two decision criteria and a perceptual noise variance). The procedural-learning models, which assumed a decision bound of arbitrary slope and intercept, had three free parameters (slope and intercept of the decision bound and perceptual noise variance). Two different guessing models were fit – one that assumed the probability of responding A equaled $\frac{1}{2}$ on

every trial (zero free parameters) and one that assumed the probability of responding A equaled p on every trial (where p was a free parameter). This latter model was included to detect participants who just responded A (or B) on almost every trial. For all models, all parameter estimates were free to take any values, and the models were independently fit to each 300-trial block both between and within participants.

All parameters were estimated using the method of maximum likelihood and the statistic used for model selection was the Bayesian information criterion (BIC; Schwarz, 1978), which is defined as $BIC = r \ln N - 2 \ln L$, where r is the number of free parameters, N is the sample size, and L is the likelihood of the model given the data. The BIC statistic penalizes models for extra free parameters. To determine the best-fitting model within a group of competing models, the BIC statistic is computed for each model, and the model with the smallest BIC value is the winning model.

The relationship between the BIC score of the best-fitting model and the participant's accuracy is typically negative, with BIC scores decreasing as performance improves. For example, in the modeling reported below, the correlations between the BIC score of the best-fitting model and the participant's accuracy during that 300-trial block of responding was $r = -0.934$ in the RB condition [$t(668) = -67.34, p < 0.001$], and $r = -0.793$ in the II condition [$t(398) = -25.99, p < 0.001$]. Perfect accuracy would mean that the decision bound model assuming an optimal bound would fit the data perfectly, and the BIC score would be as low as possible. With each additional response on the wrong side of the optimal bound, the likelihood of the model given the data decreases, increasing the negative log likelihood (the second term in the BIC equation), thus increasing the BIC score.

The BIC values identify which model provides the best account of the participant's responses, but this fact alone does not indicate whether the fit was good or bad. It is possible that all models provided poor fits and the best-fitting model just happened to provide the *least* poor fit. Unfortunately, the numerical value of the raw BIC score does not help with this problem because BIC scores increase with sample size, regardless of the quality of fit.

Any model that assumes either a rule or procedural decision strategy will provide a poor fit to randomly generated data. With random data, the guessing model will provide the best fit. So one way to assess how well a decision bound model (DBM; either rule or procedural) fits the data is to compare its fit to the fit of the guessing model. Bayesian statistics allows a method to make such comparisons (via the so-called Bayes factor). If the prior probability that the DBM model M_{DBM} is correct is equal to the prior probability that the guessing model M_{G} is correct, then under certain technical conditions (e.g., Raftery, 1995), it can be shown that

$$P(M_{\mathrm{DBM}}|\mathrm{Data}) \doteq \frac{1}{1 + \exp\left[-\frac{1}{2}(BIC_{\mathrm{G}} - BIC_{\mathrm{DBM}})\right]}, \quad (1)$$

where $P(M_{\mathrm{DBM}}|\mathrm{Data})$ is the probability that the DBM is correct, assuming that either the DBM or guessing model is correct, and $\doteq$ means 'is approximately equal to'. Thus, for example, if the DBM model is favored over the guessing model by a BIC difference of 2, the probability that the DBM model is correct is approximately .73. In other words, even though the DBM fits better than the guessing model, the fit is not very good because there is better than 1 chance in 4 that the data were just generated by random coin tossing. In contrast, if the BIC difference is 10, then the probability that the DBM model is correct is approximately .99, which means that we can be very confident that this participant was consistently using a single decision strategy that is well described by our DBM. In this case, the DBM provides an excellent fit to the data. Thus, in addition to fitting all the different decision bound models to each data set, we will also compute the Eq. 1 values.

Results

Accuracy-based Analyses

The mean accuracy and the mean of the median RTs for all 21 sessions of both conditions are shown in Figure 3. Figures 4 and 5 show the RB accuracy and RT, respectively, for training sessions 3 – 20. In both figures, performance is plotted separately for congruent and incongruent stimuli – that is, for stimuli that maintained their SR association on every day of training (i.e., congruent stimuli) and stimuli that switched responses when the category structures changed from primary to secondary (i.e., incongruent stimuli). The breaks in the curves occur at each transition between primary and secondary category structures, and vice versa. Data from the first two days of training on the primary categories are not shown in these (or later) figures because before the first secondary day, there are no incongruent stimuli.

Figures 4 and 5 suggest a number of important results. First, note that accuracy asymptotes on the primary categories early in training. In contrast, on the secondary categories, accuracy continues to improve throughout training. On the other hand, RT initially decreases for both category structures (e.g., see Figure 3) and then slowly increases across the later training sessions. This increase is inconsistent with standard automaticity criteria. However, recall that participants were given no instructions regarding the speed of their responses (except that there was a 5-second response deadline), and they were never told that their RTs were being recorded. Instead, they were told that their only task was to maximize accuracy. For this reason, the accuracy results are much more important than the RT results. During the

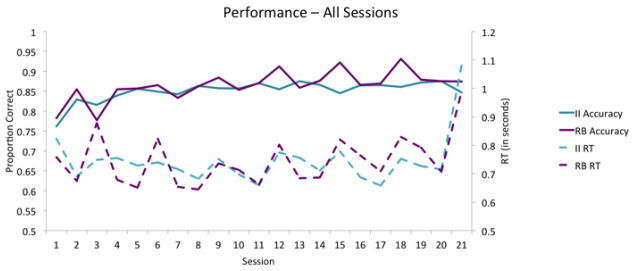

Figure 3. Accuracy and mean of the median RTs per session averaged across participants for all 21 sessions of the experiment. Sessions 1-20 of the experiment were 600 trials each, and session 21 was 300 trials. The solid lines denote accuracy, and are plotted against the primary (left) axis. The dotted lines denote RT, and are plotted against the secondary (right) axis.

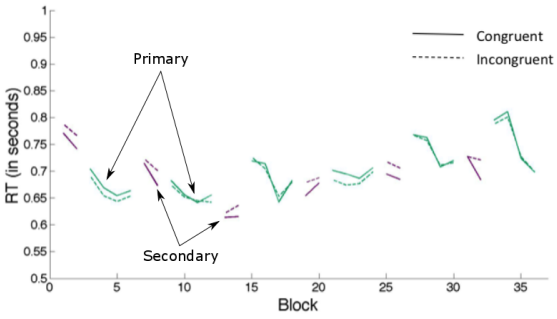

Figure 5. Mean of the median RB RTs per block averaged across participants for sessions 3 – 20. Each block included 300 trials, so there were two blocks per session. The solid lines denote the learning curves for stimuli from the congruent regions of stimulus space, and the dashed lines denote learning curves for stimuli from the incongruent regions of stimulus space. Data from primary structures are shown in green and data from secondary structures are shown in purple. The breaks in the lines are at transitions between sessions with primary and secondary category structures.

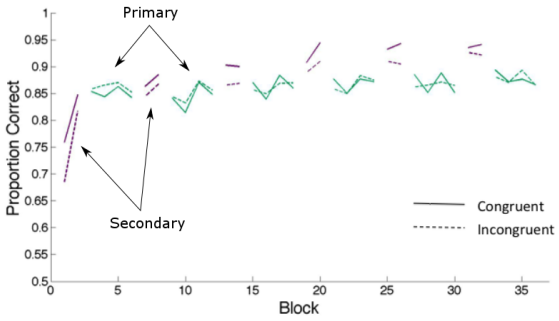

Figure 4. RB accuracy per block averaged across participants for sessions 3 – 20. The first two sessions are omitted in this figure, as congruent and incongruent regions of stimulus space had not been established before session 3. Each block includes 300 trials, so there are two data points (blocks) per session. The solid lines denote the learning curves for stimuli from the congruent regions of stimulus space, and the dashed lines denote learning curves for stimuli from the incongruent regions of stimulus space. Data from primary structures are shown in green and data from secondary structures are shown in purple. The breaks in the lines are at transitions between sessions with primary and secondary category structures.

sessions when the RTs are increasing, note that accuracy remains high. Therefore, we believe that task fatigue and a decrease in motivation are the primary factors contributing to the gradual increase in RTs that occurred during the second half of the experiment.

Second, note that performance on the primary structures seems identical for congruent and incongruent stimuli. By midway through training, the accuracies and RTs are virtually superimposed. Thus, it appears that training on the secondary categories did not disrupt the knowledge acquired about the primary categories in any stimulus-specific way. In contrast, for the secondary categories, performance is consistently better for congruent stimuli than for incongruent stimuli. Accuracy is consistently higher and RT is consistently lower.

These conclusions were tested statistically using a variety of repeated measures ANOVAs. First, for the data from the primary category structures only, we ran 2 stimulus types (congruent versus incongruent) $\times$ 17 sessions ANOVAs separately on both accuracy and RT. The results showed a significant effect of session on both accuracy [$F(16, 23) = 2.99, p < 0.001$] and RT [$F(16, 23) = 2.07, p < 0.01$], but no significant difference between the accuracy [$F(1, 16) = 0.06, p = 0.81$] or RT [$F(1, 16) = 1.47, p = 0.24$] for congruent and incongruent stimuli. There was, however, a significant interaction between stimulus type and session on accuracy [$F(1, 23) = 1.62, p < 0.05$]. This means that accu-

racy increased with practice, although at significantly different rates for congruent and incongruent stimuli. Likewise, RTs changed with practice, but at a similar rate for congruent and incongruent stimuli. A simple main effects analysis showed that the accuracy interaction was driven by congruent-incongruent performance differences early in training (i.e., before block 10).

A similar ANOVA was conducted on the data from the secondary categories. Results showed a significant effect of session on accuracy [$F(16, 11) = 12.90, p < 0.001$] and RT [$F(16, 11) = 2.01, p < 0.05$]. This time, however, there was a main effect of stimulus type on both accuracy [$F(1, 16) = 31.01, p < 0.001$] and RT [$F(1, 16) = 19.17, p < 0.001$], as well as a session by trial type interaction on accuracy [$F(1, 11) = 3.17, p < 0.001$]. Thus, accuracy increased with practice (and RTs changed), performance was significantly better on congruent stimuli than on incongruent stimuli, and accuracy changed across training differently on congruent and incongruent trials.

Figures 6 and 7 show the mean accuracy and the mean of the median RTs during each block in the II condition for sessions 3 – 20. Note first that accuracy asymptoted by midway through training, especially for the primary categories. The RTs remained roughly the same across training although within each set of blocks on the same structure, there was a rapid RT decrease. Second, note that after the first few sessions, accuracy was higher and RT was lower for congruent stimuli than for incongruent stimuli on both the primary and secondary category structures. These conclusions were tested via the same set of repeated measures ANOVAs that were used on the RB data. For the primary category structures, results showed a significant effect of session on accuracy [$F(1, 23) = 2.42$, $p < 0.001$], but not on RT [$F(1, 23) = 1.35, p = 0.14$]. There was also a significant effect of congruency on both accuracy [$F(1, 9) = 17.22, p < 0.01$] and RT [$F(1, 9) = 8.17, p < 0.05$]. The stimulus type by session interaction was not significant, either for accuracy [$F(1, 23) = 1.05, p = 0.41$] or RT [$F(1, 23) = 1.08, p = 0.37$]. Thus, accuracy improved with practice and RT did not significantly change with practice. More importantly though, accuracy was significantly higher and RTs were significantly lower in congruent regions of stimulus space than in incongruent regions.

The difference in the effect of the congruency of stimuli on the two conditions was tested in a three-way mixed ANOVA [i.e., stimulus type (congruent versus incongruent) × condition (RB versus II) × session]. The interaction between condition and session was not significant, [$F(1, 11) = 0.32, p = 0.98$], nor was the three-way interaction between condition, stimulus type, and session [$F(1, 11) = 0.50, p = 0.91$], but the interaction between stimulus type and condition [$F(1, 1) = 16.944, p < 0.001$] was significant, supporting the conclusion that there was a congruent/incongruent

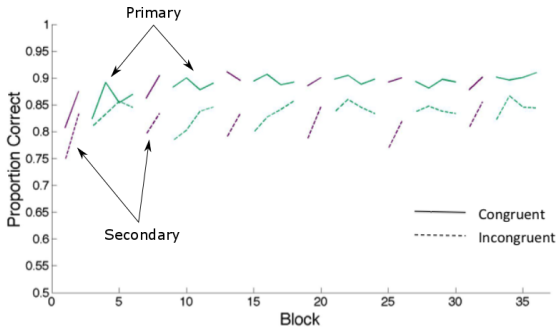

Figure 6. II accuracy per block averaged across participants for sessions 3 – 20. Each block included 300 trials, so there were two blocks per session. The solid lines denote the learning curves for stimuli from the congruent regions of stimulus space, and the dashed lines denote learning curves for stimuli from the incongruent regions of stimulus space. Data from primary structures are shown in green and data from secondary structures are shown in purple. The breaks in the lines are at transitions between primary and secondary category structures.

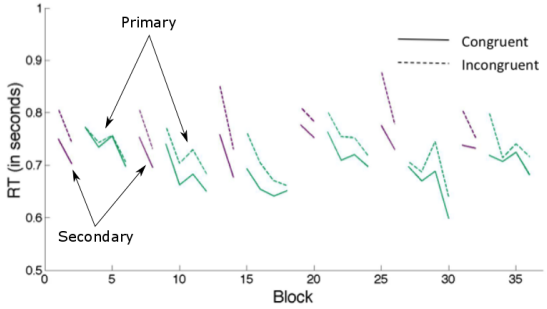

Figure 7. Mean of the median II response times per block averaged across participants for sessions 3 – 20. Each block included 300 trials, so there were two blocks per session. The solid lines denote the learning curves for stimuli from the congruent regions of stimulus space, and the dashed lines denote learning curves for stimuli from the incongruent regions of stimulus space. Data from primary structures are shown in green and data from secondary structures are shown in purple. The breaks in the lines are at transitions between primary and secondary category structures.

difference in the II condition, but not the RB condition.

With respect to automaticity, the most important session is the last training session of the experiment (i.e., session 20), since this is the session for which categorizing stimuli from the primary categories should be most automatic. In the II condition, during these 600 trials participants were significantly more accurate on congruent stimuli than on incongruent stimuli [Congruent accuracy = 89.3%, Incongruent accuracy = 82.1%, $t(9) = 3.74, p = 0.005$], whereas in the RB condition this difference was nonsignificant [Congruent accuracy = 87.1%, Incongruent accuracy = 88.0%, $t(16) = -0.82, p = 0.43$]. Computing the Bayes factors on these t tests (Rouder, Speckman, Sun, Morey, & Iverson, 2009) suggests that in the II condition, the alternative hypothesis (i.e., that congruent and incongruent accuracies are different) is 11.5 times more probable than the null hypothesis (i.e., that congruent and incongruent accuracies are equal), whereas in the RB condition the null hypothesis is 2.99 times more probable than the alternative. This latter value, although impressive, actually underestimates the evidence in favor of the null hypothesis because during session 20, RB participants were slightly more accurate on the incongruent stimuli than on the congruent stimuli. Thus, if the congruent and incongruent accuracies really are unequal, then our results would favor the possibility that participants were actually more accurate on incongruent stimuli than on congruent stimuli (an outcome that in Bayesian terms would have a very low prior probability). Thus, a Bayesian analysis strongly reinforces the conclusion that when their responding was most automatic, II participants were more accurate on congruent stimuli than on incongruent stimuli, whereas RB participants were equally accurate on both types.

Finally, note that a comparison of Figures 4 and 6 shows that the primary disjunction categories in the RB condition proved to be about equally difficult for participants to learn as the primary II categories. II performance was better on congruent than incongruent stimuli, but note that across the whole experiment, mean accuracy on these two stimulus types is approximately equal to the accuracy on the RB disjunction categories (and worse than accuracy on the RB one-dimensional categories). Thus, the primary category structures used in the RB and II conditions were approximately equal in difficulty.

Model-based Analyses

The accuracy and RT results suggest that practicing on the secondary category structures selectively slowed the development of automaticity for stimuli in the incongruent regions of stimulus space compared to stimuli in congruent regions in the II condition, but not in the RB condition. However, before interpreting these results it is important to determine the decision strategies that participants adopted, and especially whether practicing on the secondary structures caused participants to change decision strategies on the primary structures. To answer these questions, we fit decision bound models to the responses of each participant in every 300-trial block of the experiment (see the Methods section for details).

The model fitting results are shown in Figure 8. In the II condition, the overwhelming majority of data sets were best fit by a model that assumed a strategy of the optimal type (i.e., a procedural-learning model). This was equally true for the primary and secondary category structures.

In the RB condition, it can be seen that participants overwhelmingly used the optimal strategy on the primary categories during all sessions (i.e., a disjunction rule). In contrast, there is more suboptimality on the secondary one-dimensional categories, although this trend gradually decreases throughout training. Even so, at first glance it might be surprising that suboptimal strategies are used more frequently on the easier one-dimensional categories than on the more difficult disjunction-rule categories. Recall that participants received twice as much training on the primary categories as on the secondary categories. Therefore, one possibility is that the extra training on the disjunction rule led to occasional failures of the constant cognitive control needed on secondary days to inhibit inappropriate use of the more heavily practiced disjunction rule. Accuracy was high during secondary sessions, but an occasional incorrect response far from the category boundary can bias the model fitting in favor of models that assume a procedural strategy. This hypothesis is supported by several facts. First, of those participants whose responses were best fit by a model assuming a suboptimal strategy, the most common best-fitting model assumed a procedural strategy. Second, on secondary days, accuracy was lower in the incongruent regions of stimulus space than in the congruent regions (see Figure 4).

So far our analyses indicate that a model assuming a decision strategy of the optimal type provided the best overall account of the responses in both conditions across participants. But these analyses do not indicate how well or how consistently participants used that strategy. To address this question, we used Eq. 1 (see Methods section) to estimate the probability that the decision bound model is correct, assuming that either the decision bound model is correct or participants randomly guessed. In the RB condition, the mean estimate of this probability (i.e., across participants) was 0.99 on the primary sessions and 0.93 on the secondary sessions. Thus, the data are strongly consistent with the use of an optimal-type strategy, especially on the primary structures. In the II condition, the means were .957 on the primary sessions and .943 on the secondary sessions. Again, the data are strongly consistent with a strategy of the optimal type (i.e., procedural). In summary, in both conditions, the models that fit best did so because they accurately reproduced the decision strategies used by participants, and not because they have more free parameters or model flexibility than alterna-

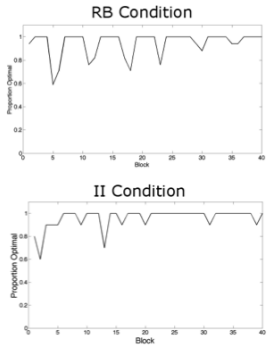

Figure 8. Proportion of participants in each block whose responses were best fit by a decision bound model that assumed a decision strategy of the optimal type. The dips in the RB figure occur on the secondary training days (see text for discussion).

tive models (e.g., the guessing models).

Dual-Task Performance

Recall that during session 21, all participants categorized the primary categories while simultaneously performing a dual task that recruits working memory and executive attention (i.e., a numerical Stroop task). This session was included to test one of the most classic automaticity criteria – namely, that automatic behaviors can be executed successfully while the participant is simultaneously engaged in some other task (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977).

Several studies have reported that during initial training, a simultaneous dual task of the type used here interferes significantly with RB learning, but not with II learning (Waldron & Ashby, 2001; Zeithamova & Maddox, 2006). These results have two important implications for the present study. First, of course, we should expect the dual task to cause only minimal interference in the II condition. Second, and more importantly, the early learning data suggest that we can use the II results as a baseline for the RB results. Of course, the ideal result would be that performance during the dual task is absolutely identical to session 20 performance (last training day before the dual task). But this might be unrealistic, especially since the transition from initial learning to fully automatized behavior is gradual (see Methods section). Thus, a lower and perhaps more realistic bar for automaticity is that the qualitative difference that exists during early RB and II dual-task performance has disappeared.

In the II condition, mean accuracy on the numerical Stroop task 94%, with a standard deviation of 3.9% (range: 86%-98%). Thus, participants devoted sufficient attentional resources to the numerical Stroop task to perform at a high level. The II accuracy and RT results are shown in Figure 9. As in all of our other analyses, the accuracy results are far more important than the RT results. Because participants were not informed that RT was being recorded, all RT results should be interpreted with caution.

To assess whether the dual task affected categorization performance, repeated measures t-tests were conducted to test for performance differences between the last session before the dual task (i.e., Session 20) and the dual task session, both for congruent and incongruent trials. Because null results are predicted, we also report the Bayes factor (BF) for each test (Rouder et al., 2009). This is an estimate of the probability that the null hypothesis is true divided by the probability that the alternative hypothesis is true. Thus, a BF of 3 means that the data suggest that the null hypothesis is 3 times more likely than the alternative hypothesis.

In the II condition, as expected, the dual task did not significantly increase accuracy (i.e., from session 20) [$t(9) = 1.21, p = 0.26, BF = 1.80$]. This was true for both congruent [$t(9) = 0.72, p = 0.49, BF = 2.60$] and incongruent stimuli [$t(9) = 1.25$, $p = 0.24$, $BF = 1.74$]. On the other hand, there was a significant increase in overall RT [$t(9) = -3.56, p < 0.01, BF = 9.21$], and for both trial types [Congruent trials: $t(9) = -3.20, p < 0.01$; Incongruent trials: $t(9) = -3.56, p < 0.01$]. There are several reasons to believe that these RT increases had little or nothing to do with the level of automaticity of the II categorization behaviors. First, the congruent/incongruent accuracy difference during the last training session (session 20) suggests that categorization of the incongruent stimuli was less automatic than categorization of the congruent stimuli. Yet the RT increases during the dual task were nearly identical for congruent and incongruent stimuli. Second, our results are consistent with Helie et al. (2010), who reported that after 20 sessions of practice on a single category structure, the same dual task used here did not significantly reduce RB or II categorization accuracy, but it did increase categorization RT in both conditions. Helie et al. (2010) also did not give participants any instructions about RT, so their results show that under the tasks instructions used here, RT increases should be expected during a dual task even when participants are given an extra 7 sessions of practice.

In the RB condition, mean accuracy on the numerical Stroop task was 93%, with a standard deviation of 5.4% (range: 81%-99%). Thus, as in the II condition, participants devoted sufficient attentional resources to the numerical Stroop task to perform at a high level. The accuracy and RT results are shown in Figure 10. The repeated measures t-tests showed that the dual tasks caused no significant change in overall accuracy [$t(16) = 0.09, p = 0.40, BF = 4.01$], nor did it increase accuracy on congruent [$t(16) = 0.10, p = 0.92, BF = 3.90$] or incongruent trials [$t(16) = 0.06, p = 0.95, BF = 3.91$]. Note that the Bayesian analysis provides

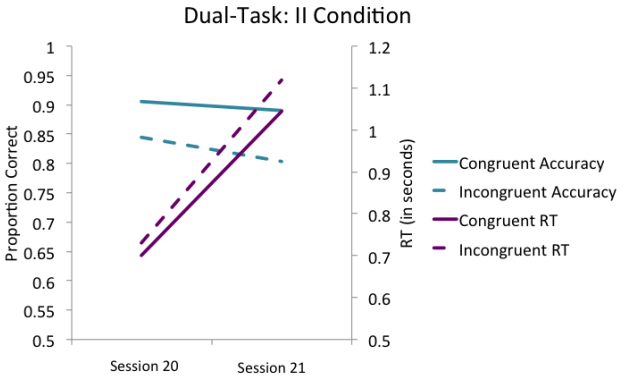

Figure 9. Proportion correct and mean of the median RTs for the last session of training (session 20) and the session with the secondary task (session 21) for the II condition. The solid lines denote the learning curves for stimuli from the congruent regions of stimulus space, and the dashed lines denote learning curves for stimuli from the incongruent regions of stimulus space. Data from primary structures are shown in green and data from secondary structures are shown in purple.

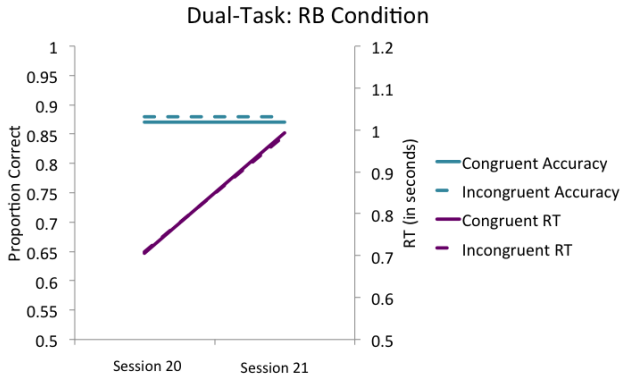

Figure 10. Proportion correct and mean of the median RTs for the last session of training (session 20) and the session with the secondary task (session 21) for the RB condition. The solid lines denote the learning curves for stimuli from the congruent regions of stimulus space, and the dashed lines denote learning curves for stimuli from the incongruent regions of stimulus space. Data from primary structures are shown in green and data from secondary structures are shown in purple.

reasonably strong support for this inference. In particular, the BFs suggest that the hypothesis that the dual task did not affect accuracy is 4 times more likely than the hypothesis that it did affect accuracy.

On the other hand, as in the II condition, the dual task was associated with a significant increase in overall RT [$t(16) = -6.19, p < 0.001$], and in RT on both trial types [Congruent trials: $t(16) = -6.19, p < 0.001$; Incongruent trials: $t(16) = -6.86, p < 0.001$]. But note that these increases are almost identical to the increases in the II condition. As a result, for the same reasons as in the II condition, there is no reason to suspect that these RT increases are due to a failure of automaticity.

Discussion

This article reports the results of an extensive experiment in which each of 27 participants completed more than 12,000 trials of perceptual categorization distributed across 21 different training sessions, either on RB or II category structures. Overall, our results are based on more than 330,000 total trials. Each participant practiced predominantly on a primary category structure, but every third session they switched to a secondary structure that used the same stimuli and responses. Importantly, half of the stimuli retained their same SR association when the secondary structures were practiced and half switched associations. The last session included a final test of automaticity in which participants categorized stimuli in the primary structures while simultaneously performing a dual task known to recruit working memory and executive attention.

A number of results stood out. First, the primary RB and II category structures were approximately equal in difficulty (i.e., the learning curves were roughly the same). Second, in both conditions, the performance of participants on the primary categories met a number of standard automaticity criteria – 1) training included 14 sessions; 2) accuracy was at asymptote for many sessions; and 3) a dual-task did not reduce categorization accuracy. Third, all participants reached high accuracy on both the primary and secondary category structures, and decision-bound modeling showed that participants overwhelmingly adopted decision strategies of the optimal type. Fourth, and most importantly, for the primary categories in the RB condition, accuracy and RT were identical on congruent stimuli that maintained their category label on every day and incongruent stimuli that switched labels on secondary category-structure days. In contrast, for the primary II categories, accuracy was higher and RT was lower for congruent than for incongruent stimuli.

As mentioned earlier, the evidence is good that in II tasks, initial learning is of SR associations, not of an abstract rule or a decision bound (Ashby & Waldron, 1999; Casale et al., 2012). Our results strongly suggest that these SR associations become automatized after thousands of trials of practice. In our experiment, practicing the secondary categoriza-

tion task reversed half of the SR assignments and preserved the other half. Our results clearly showed impaired learning for the stimuli that switched responses – exactly as predicted by the SR-learning hypothesis.

In contrast, in the RB condition, the congruent and incongruent stimuli in the primary category structures were learned equally well. Thus, practicing the opposite SR associations for some stimuli caused no interference at all. This finding is inconsistent with the hypothesis that SR associations become automatized in RB tasks. Instead, it is consistent with the hypothesis that the categorization rule becomes automatized. To our knowledge, this is the first empirical difference between automatic RB and II categorization that has been reported.

Many studies have reported that behavioral and neural changes still occur long after the 21 sessions of training included in our experiment (Crossman, 1959; Ericsson et al., 1993; Matsuzaka, Picard, & Strick, 2007). For example, Matsuzaka et al. (2007) reported that functional changes were still occurring in the primary motor cortex of monkeys even after the animals had practiced the same motor sequence for more than two years. As a result, it is impossible to rule out the possibility that if we had provided more than 21 training sessions, our results might have been different. The most plausible difference perhaps, is that eventually SR associations would become automatized in RB tasks. Although we can not rule out this hypothesis, there are several reasons why this possibility should not negate our results. First, in the RB condition, there was absolutely no evidence that SR associations were even beginning to develop by the end of training, since performance on congruent and incongruent stimuli was identical during the last training session. Thus, if significant SR associations do eventually develop, they must require far more than 14 sessions of training. Second, the amount of training we provided was enough to meet a number of standard automaticity criteria. Thus, the behaviors studied here were automatized to a similar extent as in other laboratory studies of automaticity. Third, if there are important qualitative changes that occur after hundreds or thousands of hours of training, then new methods will be required to identify these because few traditional laboratory studies will have the resources required for such extensive training. Finally, the possibility that changes might occur after such extreme overtraining does not diminish the importance of the present results. For example, a dual task interferes with initial RB performance (Waldron & Ashby, 2001; Zeithamova & Maddox, 2006), so the absence of a session 21 dual-task interference suggests that a qualitative change in RB performance occurred between our sessions 1 and 21. Our finding that after such a qualitative change, RB categorization depends on rules and not on SR associations is novel and must be accounted for by any complete theory of categorization automaticity.

Ashby and Crossley (2012) reviewed behavioral, neuroimaging, neuropsychological, and single-cell recording studies that all provide evidence for a single system of automatic categorization. Do the present results suggest more than one automaticity system? Of course it is far too early to rule this possibility out, but a more conservative question might be to ask whether any single system could accommodate our results. In the remainder of this section, we outline a candidate single-system theory of how automaticity develops in RB and II categorization.

As mentioned previously, the evidence is good that early II learning depends critically on the basal ganglia, and especially on the striatum (e.g., Ashby & Ennis, 2006; Seger & Miller, 2010). Ashby, Ennis, and Spiering (2007) proposed that in contrast, automatic II categorization is mediated entirely within cortex and that the development of II automaticity is associated with a gradual transfer of control from the striatum to cortical-cortical projections from the relevant sensory areas directly to the premotor areas that initiate the behavior. According to this account, a critical function of the basal ganglia is to train purely cortical representations of automatic behaviors. The idea is that, via dopamine-mediated reinforcement learning, the basal ganglia learns to activate the correct post-synaptic target in premotor cortex (Cantwell, Crossley, & Ashby, in press), which allows the appropriate cortical-cortical synapses to be strengthened via Hebbian learning[1]. Once the cortical-cortical synapses have been built, the basal ganglia are no longer required to produce the automatic behavior. This model easily accounts for button-switch impairments in II categorization, both during early learning and after automaticity has developed (i.e., because of the prominent role played by premotor cortex throughout learning). And it explains why II learning is of SR associations (i.e., because of the direct projections from sensory areas to premotor cortex).

A variety of evidence supports a similar model of automaticity in rule-guided tasks such as RB categorization (e.g., for a review, see Hélie, Ell, & Ashby, 2015). First, many studies have reported the existence of rule-sensitive neurons in prefrontal cortex (PFC) (Hoshi, Shima, & Tanji, 2000; Wallis, Anderson, & Miller, 2001; White & Wise, 1999), which is consistent with the evidence that early RB learning depends critically on the PFC (Anderson, Damasio, Jones, & Tranel, 1991; Hélie, Roeder, & Ashby, 2010; Konishi et al., 1999; Monchi, Petrides, Petre, Worsley, & Dagher, 2001; Rogers, Andrews, Grasby, Brooks, & Robbins, 2000). However, rule-sensitive neurons have also been found in premotor

[1]According to this account, cortical-cortical synaptic plasticity follows Hebbian learning rules because the low levels of cortical dopamine active transporter (DAT) prevents the rapid fluctuations in cortical dopamine levels needed for true reinforcement learning. In contrast, the basal ganglia are rich in DAT, so there, synaptic plasticity follows reinforcement learning rules

cortex (Muhammad, Wallis, & Miller, 2006; Wallis & Miller, 2003), and there is some evidence that during extended training, behavioral control might gradually pass from the PFC to premotor cortex. First, the neuroimaging data collected by Helie et al. (2010) over the course of 20 sessions of RB categorization were consistent with this hypothesis. Second, Muhammad et al. (2006) recorded from single neurons in the PFC and premotor cortex while monkeys were making rule-based categorization responses. In agreement with Helie et al. (2010), they found many neurons in the PFC that fired selectively to a particular rule. However, after training the animals for a year, they also found many premotor neurons that were rule selective, and even more importantly, these neurons responded on average about 100 ms before the PFC rule-selective cells. Thus, after categorization had become automatic, the PFC, although still active, was not mediating response selection. Instead, the single-unit data suggested that the automatic representation had moved to regions that included the premotor cortex. Third, within the PFC, several studies have reported that the more concrete the rule, the more caudal the representation (Badre, Kayser, & D'Esposito, 2010; Bunge & Zelazo, 2006; Christoff, Keramatian, Gordon, Smith, & Mädler, 2009). Based on evidence such as this, Helie et al. (2010) proposed that as rules become more concrete with more extensive training, they are progressively re-coded more caudally in the PFC until eventually reaching the premotor cortex, at which time they become automatic.

Thus, according to this view, the primary goal of rule-learning circuits centered in PFC and procedural-learning circuits centered in the basal ganglia is to train automatic representations between sensory cortex and premotor cortex. If so, then the only difference between automaticity in RB and II tasks is that the terminal projection in RB tasks is onto premotor rule-sensitive neurons, whereas in II tasks the terminal projection is onto premotor response-sensitive neurons. In other words, after extensive training, in RB tasks the sight of a familiar stimulus automatically triggers the appropriate rule, whereas in II tasks the sight of a familiar stimulus automatically triggers the appropriate motor response.

References

Anderson, S. W., Damasio, H., Jones, R. D., & Tranel, D. (1991). Wisconsin card sorting test performance as a measure of frontal lobe damage. *Journal of Clinical and Experimental Neuropsychology*, *13*(6), 909-922.

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

Ashby, F. G., & Crossley, M. J. (2012). Automaticity and multiple memory systems. *Wiley Interdisciplinary Reviews: Cognitive Science*, *3*(3), 363-376.

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*, 1-36.

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*, 33-53.

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

Ashby, F. G., & Maddox, W. T. (2010). Human category learning 2.0. *Annals of the New York Academy of Sciences*, *1224*, 147-161.

Ashby, F. G., & Waldron, E. M. (1999). On the nature of implicit categorization. *Psychonomic Bulletin & Review*, *6*(3), 363-378.

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

Bunge, S. A., & Zelazo, P. D. (2006). A brain-based account of the development of rule use in childhood. *Current Directions in Psychological Science*, *15*(3), 118-121.

Cantwell, G., Crossley, M. J., & Ashby, F. G. (in press). Multiple stages of learning in perceptual categorization: Evidence and neurocomputational theory. *Psychonomic Bulletin & Review*.

Casale, M. B., Roeder, J. L., & Ashby, F. G. (2012). Analogical transfer in perceptual categorization. *Memory & Cognition*, *40*(3), 434-449.

Christoff, K., Keramatian, K., Gordon, A. M., Smith, R., & Mädler, B. (2009). Prefrontal organization of cognitive control according to levels of abstraction. *Brain Research*, *1286*, 94-105.

Crossley, M. J., Paul, E. J., Roeder, J. L., & Ashby, F. G. (in press). Declarative strategies persist under increased cognitive load. *Psychonomic Bulletin & Review*.

Crossman, E. R. F. W. (1959). A theory of the acquisition of speed-skill. *Ergonomics*, *2*(2), 153-166.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*(3), 363-406.

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), 212-222.

Hélie, S., Ell, S. W., & Ashby, F. G. (2015). Learning robust cortico-cortical associations with the basal ganglia: An integrative review. *Cortex*, *64*, 123-135.

Hélie, S., Roeder, J. L., & Ashby, F. G. (2010). Evidence for cortical automaticity in rule-based categorization. *The Journal of Neuroscience*, *30*(42), 14225-14234.

Hélie, S., Waldschmidt, J. G., & Ashby, F. G. (2010). Automaticity in rule-based and information-integration categorization. *Attention, Perception, & Psychophysics*, *72*(4), 1013-1031.

Hoshi, E., Shima, K., & Tanji, J. (2000). Neuronal activity in the primate prefrontal cortex in the process of motor selection based on two behavioral rules. *Journal of Neurophysiology*, *83*(4), 2355-2373.

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

1402.

Konishi, S., Kawazu, M., Uchida, I., Kikyo, H., Asakura, I., & Miyashita, Y. (1999). Contribution of working memory to transient activation in human inferior prefrontal cortex during performance of the wisconsin card sorting test. *Cerebral Cortex*, *9*(7), 745-753.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). Sustain: a network model of category learning. *Psychological Review*, *111*(2), 309-332.

Luce, R. D. (1986). *Response times* (No. 8). Oxford University Press.

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.

Matsuzaka, Y., Picard, N., & Strick, P. L. (2007). Skill representation in the primary motor cortex after long-term practice. *Journal of Neurophysiology*, *97*(2), 1819–1832.

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.

Moors, A., & De Houwer, J. (2006). Automaticity: a theoretical and conceptual analysis. *Psychological Bulletin*, *132*(2), 297–326.

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.

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.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, *25*, 111–164.

Reber, P. J., Gitelman, D. R., Parrish, T. B., & Mesulam, M. (2003). Dissociating explicit and implicit category knowledge with fmri. *Cognitive Neuroscience, Journal of*, *15*(4), 574–583.

Rogers, R. L., Andrews, T. K., Grasby, P., Brooks, D., & Robbins, T. (2000). Contrasting cortical and subcortical activations produced by attentional-set shifting and reversal learning in humans. *Cognitive Neuroscience, Journal of*, *12*(1), 142-162.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. detection, search, and attention. *Psychological Review*, *84*(1), 1-66.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461-464.

Seger, C. A., & Miller, E. K. (2010). Category learning in the brain. *Annual Review of Neuroscience*, *33*, 203-219.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: Ii. perceptual learning, automatic attending and a general theory. *Psychological Review*, *84*(2), 127-190.

Smith, J. D., Zakrzewski, A., Johnston, J. J. R., Roeder, J., Boomer, J., Ashby, F. G., & Church, B. A. (2015). Generalization of category knowledge and dimensional categorization in humans (homo sapiens) and nonhuman primates (macaca mulatta). *Journal of Experimental Psychology: Animal Behavior Processes*, in press.

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

Spiering, B. J., & Ashby, F. G. (2008). Response processes in information–integration category learning. *Neurobiology of Learning and Memory*, *90*(2), 330-338.

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.

Waldschmidt, J. G., & Ashby, F. G. (2011). Cortical and striatal contributions to automaticity in information-integration categorization. *Neuroimage*, *56*(3), 1791-1802.

Wallis, J. D., Anderson, K. C., & Miller, E. K. (2001). Single neurons in prefrontal cortex encode abstract rules. *Nature*, *411*(6840), 953-956.

Wallis, J. D., & Miller, E. K. (2003). From rule to response: neuronal processes in the premotor and prefrontal cortex. *Journal of neurophysiology*, *90*(3), 1790-1806.

White, I. M., & Wise, S. P. (1999). Rule-dependent neuronal activity in the prefrontal cortex. *Experimental brain research*, *126*(3), 315-335.

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

Author Notes

This research was supported by NIH grant 2R01MH063760.

For all of the experiments described here, all measures, conditions, and exclusions are reported.