Cognition 226 (2022) 105168

Contents lists available at ScienceDirect

Cognition

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

On what it means to automatize a rule

Paul Kovacs, F. Gregory Ashby *

University of California, Santa Barbara, United States of America

ARTICLE INFO

Keywords:
Rule-guided behavior
Rule-based categorization
Automaticity
Selective attention

ABSTRACT

The results of two experiments are reported that included a combined total of approximately 633,000 categorization trials. The experiments investigated the nature of what is automatized after lengthy practice with a rule-guided behavior. The results of both experiments suggest that an abstract rule, if interpreted as a verbal-based strategy, was not automatized during training, but rather the automatization linked a set of stimuli with similar values on one visual dimension to a common motor response. The experiments were designed to test and refine a recent neurocomputational model of how rule-guided behaviors become automatic (Kovacs, Hélie, Tran, & Ashby, 2021). The model assumes that rule-guided behaviors are initially controlled by a distributed neural network centered on rule units in prefrontal cortex, and that in addition to initiating behavior, this network also trains a faster and more direct network that includes projections from visual cortex directly to the rule-sensitive neurons in premotor cortex. The present results support this model and suggest that the projections from visual cortex to prefrontal and premotor cortex are restricted to visual representations of the relevant stimulus dimension only.

1. Introduction

Repeatedly practicing a skill eventually causes it to be executed automatically. The study of this process has a long history – dating back at least to Charles Sherrington (1906) and William James (1914). During much of this time, however, the focus was on behavioral signatures of automatic behaviors. This work was important because, once identified, these signatures could then be used to test whether a behavior had been practiced long enough to become automatic. The classic work on this problem was by Shiffrin and Schneider (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977), who identified many features of automatic behaviors that are still used today to identify automaticity. Among other examples, they noted that automatic behaviors require few attentional resources, and as a result, they can be performed fluidly at the same time as other simple behaviors. In other words, automatic behaviors are resistant to dual-task interference. As another example, automatic skills are behaviorally inflexible, in the sense that changing the response requirements – for example, by switching the locations of the response keys – interferes with the execution of automatic skills.

In contrast, relatively little work has studied exactly what is automatized during the long period of practice that is required for automaticity. Among the first studies to examine this issue reported evidence that the nature of the knowledge that is automatized depends on the learning system used to acquire the behavior. In particular, Roeder and Ashby (2016) reported evidence that rules are automatized with rule-guided behaviors, whereas stimulus-response associations are automatized with skills that are acquired via procedural learning. Stimulus-response associations seem unambiguous, but a rule could be instantiated in many different ways. For example, is the automatized rule an abstract set of instructions that can be applied with equal facility to any relevant stimulus, or is it highly stimulus specific? And does it require selective attention to individual stimulus features or components, or can it operate on the stimulus gestalt?

This article describes the results of two extensive experiments that investigated the nature of what is automatized after lengthy practice with a rule-guided behavior. The experiments were designed to test novel predictions of a recent neurocomputational model of how rule-guided behaviors become automatized (Kovacs, Hélie, Tran, & Ashby, 2021). The results of both experiments support the predictions of the model and suggest that an abstract rule, if interpreted as a verbal-based strategy, was not automatized during training, but rather the automatization linked a set of stimuli with similar values on one visual dimension to a common motor response.

* Corresponding author at: Department of Psychological & Brain Sciences, University of California, Santa Barbara, Santa Barbara, CA 93106, United States of America.
E-mail address: fgashby@ucsb.edu (F.G. Ashby).

https://doi.org/10.1016/j.cognition.2022.105168
Received 24 June 2021; Received in revised form 10 May 2022; Accepted 10 May 2022
Available online 26 May 2022

1.1. Experiment 1 overview

Experiment 1 trained 29 naive participants on novel categories of unfamiliar visual stimuli long enough so that their responses became automatic (i.e., 8400 trials each). Next, each participant completed a final transfer session in which they categorized novel stimuli that they had never seen before. Our analyses focused on how well their categorization training prepared them to categorize these novel stimuli. All of the novel stimuli presented during this transfer session could be categorized perfectly using the same strategy that was automatized during training. As a result, we expected transfer accuracy to be high. Our main goal therefore, was to assess whether automaticity transferred to the novel stimuli. Specifically, the aim of the experiment was to determine whether participants categorized the transfer stimuli automatically or whether they appealed back to the more effortful categorization strategy they used during the early training sessions.

The stimuli were circular sine-wave gratings that varied across trials in bar width (spatial frequency) and bar orientation. Fig. 1 illustrates the stimuli and categories used during training and transfer in both of our experimental conditions. There were two training categories and perfect performance could be achieved via the simple one-dimensional rule: "respond A if the orientation of the bars is shallow; otherwise respond B". Participants were given no instructions about the optimal strategy. They were simply told that there were two categories of disks, A and B, and their job was to use the trial-by-trial feedback to learn to assign each presented disk to its correct category.

The experiment included 15 sessions of 600 categorization trials each. Therefore, each participant completed a total of 9000 categorization trials. The first 14 sessions were identical for all participants. Each of these 8400 trials (i.e., 14 × 600) were standard categorization trials. The stimuli and categories used during training are denoted in Fig. 1 by the open squares. The goal of the training sessions was to train participants on the categorization task long enough that their responses became automatic. Previous research with the same stimuli indicated that 8400 trials of training was sufficient for automaticity to develop (Hélie, Waldschmidt, & Ashby, 2010).

The nature of the knowledge that participants acquired during training was assessed during the final transfer session (i.e., session 15). There were two conditions, with separate participants in each condition. In the Relevant-Dimension Transfer (RDT) condition, the stimuli presented to participants changed values on the relevant dimension (i.e., orientation of the bars), but not on the irrelevant dimension (bar width). The transfer stimuli in the RDT condition are denoted in Fig. 1 by the light gray dots. Note that the separation between the category A and B exemplars in the RDT condition is greater during transfer than during training, and as a result, the transfer task is objectively easier than the training task. In the Irrelevant-Dimension Transfer (IDT) condition, the stimuli changed values on the irrelevant dimension (i.e., bar width), but not on the relevant dimension. The transfer stimuli in the IDT condition are denoted in Fig. 1 by the black dots. Note that the separation between the category A and B exemplars in the IDT condition is the same as during training, so the IDT transfer task is objectively equal in difficulty to the training task.

Note that the simple one-dimensional rule that perfectly categorizes the training stimuli also works perfectly in both transfer conditions. As a result, based on previous research, we expected transfer accuracy to be high in both conditions (Casale, Roeder, & Ashby, 2012). For this reason, our primary goal was to determine whether automaticity transferred to the novel stimuli that participants categorized during the final session. To answer this question, we used two classic tests for assessing automaticity – the performance of automatic behaviors should be: 1) unaffected by having to perform a simultaneous dual task, and 2) impaired if the location of the response buttons is reversed (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977).

To implement these tests, the final session was divided into three separate blocks of 200 trials each. These are described in Fig. 2. During the first 200 trials (block 1), participants categorized the novel transfer stimuli while simultaneously performing a dual task that required working memory and executive attention (i.e., a numerical Stroop task). During the third block of 200 trials, participants categorized the transfer stimuli using the same procedures as during training, except that the locations of the response buttons were switched. Participants were informed of this switch before the block began and cues were presented on the screen on every trial that signaled the new button locations. Therefore, no new learning was required. Finally, during the second block of 200 trials, participants categorized the transfer stimuli using the same procedures as during training. The data from these trials served as a baseline or control that was used to assess the effects of the dual task and button switch on performance. Therefore, in summary, the final session followed a 2 × 3 factorial design, in which 2 conditions (RDT, IDT) were crossed with 3 block types (categorization only, dual task, button switch).

1.2. A neurocomputational model of automatic rule-guided behaviors

Kovacs et al. (2021) recently proposed the first neurocomputational model of how rule-guided behaviors become automatic. Fig. 3 shows the model as it would look at the end of the 14 training sessions of Experiment 1. The model builds on the many reports that there are rule-sensitive neurons in both prefrontal cortex (PFC) and premotor cortex (PMC; e.g., Muhammad, Wallis, & Miller, 2006; Wallis & Miller, 2003; Vallentin, Bongard, & Nieder, 2012). The model assumes that rule-guided behaviors are initially controlled by a distributed neural network centered on the PFC rule units, and that in addition to initiating behavior, this network also trains a faster and more direct network that includes projections from visual cortex directly to the rule-sensitive neurons in PMC.

Each rule unit includes two simulated neurons. In the case of the Experiment 1 training rule, one neuron signals if the stimulus has a small orientation (the S unit), and one signals if the orientation is large (the L unit). The idea is that orientation-sensitive units in visual cortex that respond to shallow orientations project to the PFC-S neuron, whereas visual cortical units that respond to steep orientations project to the PFC-L neuron. In this way, the S neuron responds to any shallow orientation and the L neuron responds to any steep orientation. The model assumes that these PFC rule units develop as a result of life-long practice with a rule. During early training sessions, the stimulus activates the appropriate PFC rule unit (i.e., S or L), which then activates the analogous PMC rule unit, which then activates the appropriate unit in motor cortex that causes the model to respond A on trials when the orientation is shallow and B when the orientation is steep. The same visual cortical units that project to the S and L PFC rule units also project to the S and L PMC rule units. Initially, however, these visual cortex-to-PMC synapses are not strong enough to activate the appropriate unit in motor cortex. Instead, PFC activation is also required.[1]

The model also assumes that Hebbian learning will strengthen all active synapses in the Fig. 3 network. The most critical of these for behavioral predictions are highlighted in the figure by the thicker projections. First consider the synapses between visual cortex and PMC. In Hebbian learning, synaptic strengthening is proportional to the product of the pre- and post-synaptic activations. During early training, much of the post-synaptic activation (i.e., the activation within the PMC units) is driven by input from PFC. As the visual cortex-to-PMC synaptic strength increases, it eventually becomes strong enough so that visual input alone is enough to cause the PMC unit to activate the appropriate target in motor cortex. The pathway through PFC is still active, but because it is longer, it no longer controls behavior. At this point, the behavior has

[1] This is because the experimental participants all presumably have a life history of making judgments about orientation, but not about pressing an A or B button to signal the outcomes of these judgments.

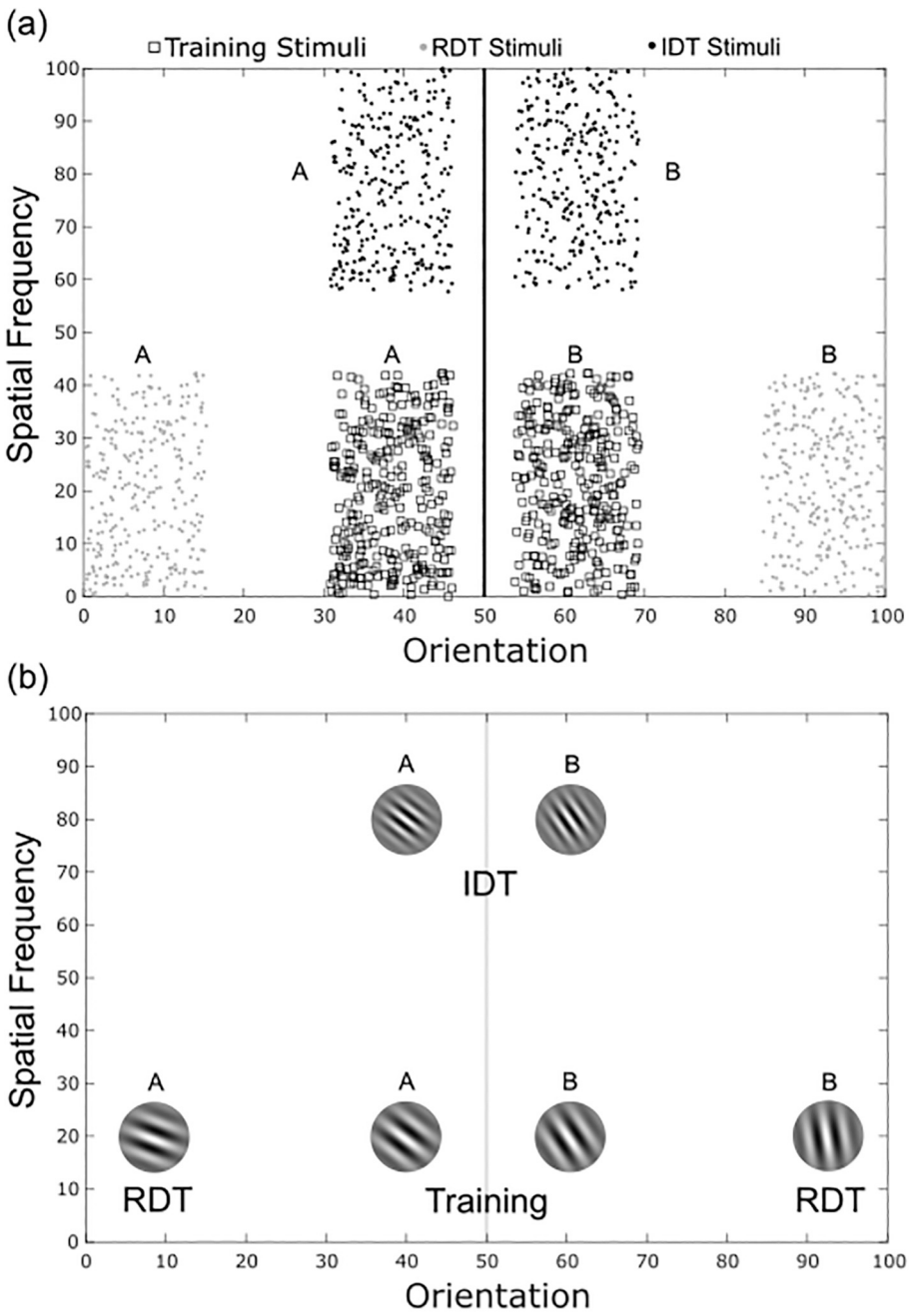

Fig. 1. Stimuli and category structures used in Experiment 1. The optimal bound for all category structures is $x_1 = 50$. Panel (a) shows coordinate values of all stimuli used and panel (b) shows some example stimuli.

become automatic.

Second, consider the synapses between PMC and primary motor cortex. Initially these are weak because participants have no prior association between shallow or steep orientations and A or B button presses. But after thousands of practice trials, Hebbian learning will strengthen these associations. The model therefore predicts that both transfer conditions will be susceptible to a button-switch interference. This is because the transfer conditions introduce novel stimuli, but the categorization rule and motor responses remain the same as during training.

The model successfully accounts for single-unit recordings and human behavioral data that are problematic for other models of automaticity. For example, it accounts for resistance to dual-task interference because the working memory circuits centered in PFC are not needed to initiate automatic behaviors, and it accounts for an interference when the response button locations are switched because Hebbian learning between PMC and primary motor cortex strengthens the motor associations during training so much that top-down executive attention is unable to reverse them completely after the switch occurs.

This model predicts that automatic rule-guided behaviors are stimulus specific, but initial rule-guided behaviors are not. In particular, the model predicts that early rule-guided behaviors are mediated by abstract rules that are represented in PFC, whereas automatic rule-guided behaviors are mediated by direct projections from visual cortex to PMC units that control the behavior. Because of Hebbian learning, the associations between the stimulus representations in visual cortex and the motor associations in PMC eventually become strong enough to trigger the behavior without assistance from the abstract rule representations in PFC.

Now consider the predictions of the model for the RDT and IDT

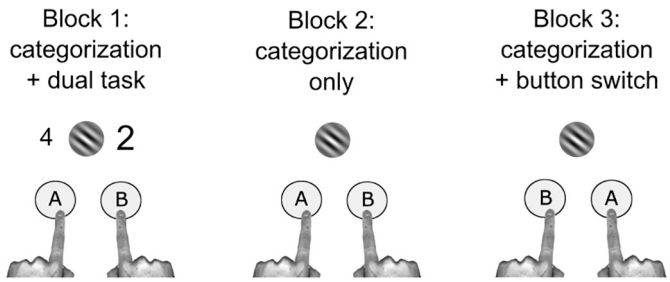

Fig. 2. Description of the three 200-trials blocks of the 15th and final (transfer) session of Experiment 1. All stimuli during this session were either from the IDT or RDT categories shown in Fig. 1. During the first 200 trials, participants categorized the novel stimuli while completing a simultaneous numerical Stroop dual task. During the second block of 200 trials, participants categorized the stimuli under the same procedures as during the first 14 training sessions. Finally, during the last block of 200 trials, participants categorized the stimuli in the usual manner, except the locations of the response buttons were reversed, and participants were explicitly instructed of this change.

conditions of Experiment 1. In the RDT condition, the model predicts that the novel orientations of the transfer stimuli will activate visual cortical neurons that were never activated during training. As a result, their synapses into PMC will be weak (i.e., untrained), dropping the PMC response to visual input below the threshold needed to activate motor cortex. In this case, PFC input is needed to cause enough PMC activation to trigger a motor response. Accuracy should remain high, however, because the PFC retains the representation of the correct rule. Even so, because application of that rule now depends on working memory and executive attention (unlike automatic behaviors), transfer performance should be susceptible to dual-task interference. This is a strong prediction because the transfer categories are more widely separated in the RDT condition than the training categories (see Fig. 1), and therefore the transfer task is objectively easier than the training task. Thus, the model predicts that even though the transfer categories are easier, participants should lose the ability to respond automatically to the RDT transfer stimuli.

Somewhat counterintuitively, however, the model also predicts that transfer performance during the button-switch block of the RDT condition should appear automatic, in the sense that it should be susceptible to button-switch interference. This is because the model predicts that no matter how the response is selected, response execution is mediated by the same PMC-to-primary motor cortex projections during both training and transfer. Therefore, even if control is passed back to PFC during the RDT blocks, the same PMC-to-primary motor projections must be used to initiate the motor response as during training, and therefore a button-switch interference should still occur. In summary then, the model makes a set of strong and novel predictions about transfer performance in the RDT condition: 1) accuracy should remain high, 2) performance during the simultaneous dual-task should appear non-automatic (i.e., susceptible to interference), and 3) performance after the button switch should appear automatic (also susceptible to interference).

Next, consider the IDT condition. The only difference between the IDT and RDT conditions is in the transfer stimuli. In both conditions, the categorization rule remains the same during training and transfer, and so do the response buttons. As a result, the model predicts that transfer accuracy should be high in both conditions and both conditions should be susceptible to a button-switch interference. But what about a dual-task interference? The IDT transfer stimuli differ from the training stimuli, but only on the irrelevant dimension. So the model predictions depend on what type of visual representation projects to PFC and PMC. If the projections from visual cortex to PMC are of the stimulus gestalt, then the visual inputs to PFC and PMC change in both conditions, so the model makes identical predictions in the RDT and IDT conditions. Abstract rule representations in PFC would be needed to initiate motor behaviors in both conditions, so IDT transfer responding should be susceptible to a dual-task interference. In contrast, if the projections from visual cortex to PMC are only of values on the relevant dimension, then the model predicts that dual-task and button-switch performance should both remain automatic because from the perspective of PMC, the visual representations received during transfer would be identical to the visual representations received during training (since the stimuli do not change on the relevant dimension).

Kovacs et al. (2021) made no assumptions about whether the visual representations used by the model were of stimulus gestalts or were restricted to the relevant stimulus dimensions only. Even so, there is reason to favor the hypothesis that the representations are of single dimensions. For example, humans learn categories like the ones used during the Experiment 1 training – in which the optimal strategy is a simple one-dimensional rule – much more quickly than categories that are identical except the stimulus space is rotated 45°, so that the optimal decision boundary is diagonal (e.g., Ashby, Smith, & Rosedahl, 2020). In contrast, pigeons and rats learn both types of categories at exactly the same rate (Broschard, Kim, Love, Wasserman, & Freeman, 2019; Qadri,

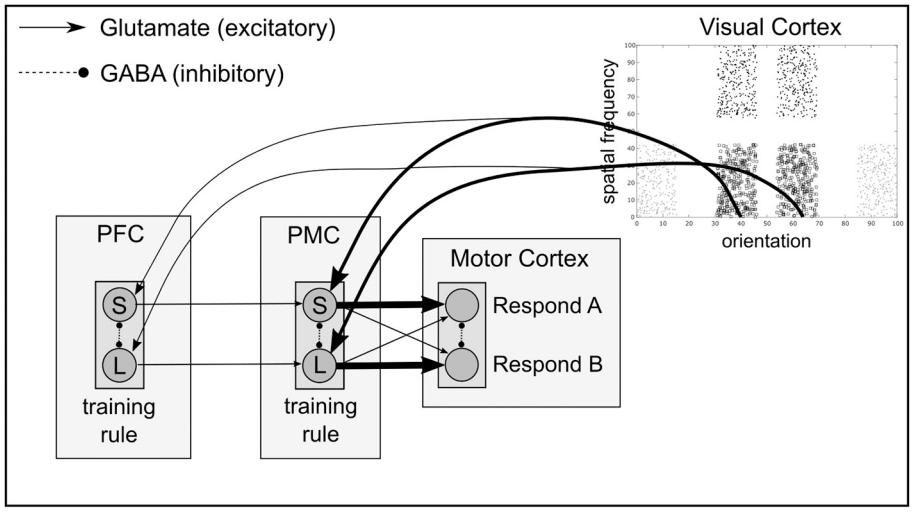

Fig. 3. A schematic of the Kovacs et al. (2021) model as it would look at the end of the training sessions of Experiment 1. The thicker projections represent increases in synaptic strength that result from Hebbian learning. PFC = prefrontal cortex, PMC = premotor cortex, S and L refer to units that respond to stimuli with small and large orientations, respectively.

Ashby, Smith, & Cook, 2019; Smith et al., 2011). This across-species difference supports the hypothesis that the human one-dimensional advantage is due to their ability to apply explicit rules with one-dimensional categories, and that pigeons and rats lack this ability. Critically though, both macaque and capuchin monkeys show a similar advantage to humans in the one-dimensional task, relative to the rotated diagonal-bound task (Smith, Beran, Crossley, Boomer and Ashby, 2010; Smith et al., 2012; Smith et al., 2015). This result suggests that the human one-dimensional learning advantage is not necessarily language based, and instead may be due to an ability to attend selectively to the single relevant dimension – a skill that is closely tied to PFC (e.g., Miller & Cohen, 2001). If so, then it seems natural that the visual representations used by the PFC rule units would exploit this selective attention ability.

2. Experiment 1

Experiment 1 tests some highly non-intuitive predictions of the Kovacs et al. (2021) theory – for example, that transfer performance in the RDT condition should appear automatic during the button-switch trials but non-automatic during the dual-task trials, and that this loss of automaticity in the presence of a dual task should occur even though the RDT transfer stimuli are objectively easier to categorize than the training stimuli (i.e., the RDT transfer categories are more widely separated than the training categories). In addition, it also tests whether the visual representations supporting explicit rule use are of gestalts or limited only to the relevant stimulus dimension.

2.1. Methods

2.1.1. Participants

Twenty-nine healthy undergraduate students at the University of California, Santa Barbara, participated in this experiment in exchange for class credit. Fourteen participants were randomly assigned to the RDT condition, and the remaining 15 participants were assigned to the IDT condition.

2.1.2. Stimuli and apparatus

All stimuli were circular sine-wave gratings of constant contrast and size presented on a 21-in. monitor (1280 × 1024 resolution). Each stimulus was defined by a set of points (x_1, x_2) sampled from a 100 × 100 stimulus space and converted to a disk using the following equations: spatial frequency $= 2^{(x_1/28)}$ cycles per disk and orientation $= 9x_2/10 +$ 15 degrees counterclockwise rotation from horizontal(Treutwein, Rentschler, & Caelli, 1989).

During training, stimuli in category A were uniformly distributed (in the 100 × 100 space) in the interval [30.77, 46.15] on the orientation dimension and [0, 42.31] on the spatial frequency dimension. Stimuli in category B were also uniformly distributed, over the intervals [53.85, 69.23] and [0, 42.31] for orientation and spatial frequency, respectively. The stimuli were generated with PsychoPy (Peirce, 2007), and subtended an approximate visual angle of 13°. Note that perfect accuracy is possible if participants use the simple one-dimensional decision rule: Respond A if the orientation is less than 50°; otherwise respond B.

During the transfer session, the stimulus values were the same as during training, except in the RDT condition, the stimulus values were shifted on the relevant dimension – that is, orientation – whereas in the IDT condition they were shifted on the irrelevant dimension (i.e., spatial frequency). In the RDT condition, the category A stimuli were uniformly distributed over the intervals [0, 15.35] and [0, 42.31] for orientation and spatial frequency, respectively, and the category B stimuli were uniformly distributed over the intervals [84.62, 100] and [0, 42.31] for orientation and spatial frequency, respectively. In the IDT condition, the category A stimuli were uniformly distributed over the intervals [30.77, 46.15] and [57.7, 100] for orientation and spatial frequency, respectively, and the category B stimuli were uniformly distributed over the intervals [53.85, 69.23] and [57.7, 100] for orientation and spatial frequency, respectively.

Stimulus presentation, feedback, response recording, and response time (RT) measurement were acquired and controlled using PsychoPy on a Macintosh computer. Responses were given on a standard Macintosh keyboard: the "D" key for an A categorization and the "K" key for a B categorization (sticker-labeled as either A or B). Each correct response was followed by the word "Correct" on the screen in green letters, and each incorrect response was followed by the word "Incorrect" in red letters.

2.1.3. Procedure

The experiment lasted for 15 sessions over 15 consecutive workdays. The first 14 sessions were training, and the last session was transfer. Each session included 600 categorization trials. All together, each participant completed 8400 trials of training and 600 trials of transfer.

On training days, participants were informed that they were taking part in a categorization experiment and were instructed to assign each stimulus to one of two categories, either A or B. A single trial proceeded as follows: The stimulus appeared in the center of the screen and remained on the screen until the participant responded, after which correct or incorrect visual feedback was immediately displayed for 2 s.

During the transfer session, participants performed a total of 600 trials split into three blocks: 1) 200 trials of categorization with a concurrent numerical Stroop task, 2) 200 trials of categorization only, and 3) 200 trials of categorization with the locations of the response buttons switched.

During the 200 dual-task trials of the transfer session (block 1), two different digits were randomly chosen on every trial (ranging from 2 to 8), and displayed for 1 s on the left and right of the center of the screen, with each offset by approximately 2° of visual angle. One of the digits was displayed in a larger font at 6 cm in height. The other digit was 3 cm in height. A "congruent" trial in the numerical Stroop task was defined as a trial in which the digit with the larger value was displayed in a larger font, whereas an "incongruent" trial was defined as a trial where the digit with the smaller value was displayed in the larger font. Incongruent trials produce a Stroop-like interference (Waldron & Ashby, 2001). The response keys and feedback for the numerical Stroop task were the same as for the categorization task. The D key (labeled A) was used to indicate left, and the K key (labeled B) was used to indicate right (matching their locations on a regular keyboard).

Participants were instructed to memorize the numerical value and physical size of the two digits. The digits then disappeared and were followed by a blank screen for 300 msec, and then followed by the categorization stimulus. The categorization stimulus stayed on the screen until a categorization response was made. Categorization feedback was given after 300 msec and stayed on the screen for 700 msec. After the feedback, the screen went blank for 300 msec followed by a cue, either the word "Size" or the word "Value" If the cue was "Size," the participant indicated whether the number presented in the larger font was on the right or the left of the screen. If the cue was "Value," the participant indicated whether the number with the larger value was on the right or the left of the screen. The cue remained on the screen until the participant responded. Feedback was given in the same way as in the categorization task. As in the training sessions, half the categorization stimuli were from category A and half were from category B. In the numerical Stroop task, 170 trials were incongruent (85%), and the remaining 30 trials were congruent (15%). This manipulation aimed at drawing the analogy with the original Stroop task – that is, by opposing the natural bias of associating digit size with digit value. Half the correct responses were located on the left, and half on the right. Also, the digit with the larger value was located on the left for half the trials, and half the digits with the larger size were located on the left. Participants were instructed to focus on the numerical Stroop task and to perform the categorization task with the attentional resources they had left. Additionally, participants were instructed to respond as quickly as they could

without sacrificing accuracy.

The trial-by-trial procedures for the 200 categorization-only trials of the transfer session (block 2) were identical to the training sessions. During the break between blocks 1 and 2, participants were again instructed to respond as quickly as they could without sacrificing accuracy.

During the 200 button-switch trials of the transfer session (block 3), categorization trials were identical to training trials except the categorization response key locations were switched. The letters "A" and "B" were displayed on the left and right side of the bottom of the screen in positions corresponding to the new locations of the response keys. Participants were instructed at the end of block two that everything in the next 200 trials would be the same except that the response keys would switch positions. They were also instructed to refer to the letters "A" and "B" displayed at the bottom of the screen to remind them of the new button locations. Additionally, participants were again instructed to respond as quickly as they could without sacrificing accuracy.

2.2. Results

Fig. 4 shows the mean proportion correct averaged over participants during each session of training. As expected, accuracy increased quickly and plateaued at a high level of performance (above 90% correct). The means of each participant's median RTs are shown in Fig. 5. Also as expected, note that RT gradually decreased over sessions, beginning at about 700 ms on session 1 and ending at 580 ms during the last training session (i.e., Session 14).

2.2.1. Standard statistical analysis

Results from the final transfer session are summarized in Fig. 6. The data from the categorization-only trials (i.e., block 2) were used as controls.

As a first analysis, we analyzed the transfer session data using a series of generalized linear mixed models (GLMM). The accuracy analysis assumed a logistic link function, whereas the link function for the RT analysis was the identity. The main advantage of using a GLMM analysis instead of ANOVA is that, in the case of the trial-by-trial Bernoulli distributed accuracy data, the ANOVA assumption of normality is violated. However, we also analyzed the RTs using a standard ANOVA and the results were qualitatively identical.

Recall that the final transfer session followed a 2×3 factorial design, in which 2 conditions (RDT, IDT) were crossed with 3 block types (categorization only, dual task, button switch). Therefore, the GLMM analysis included all of the models that would be tested in a standard ANOVA. This includes a null model in which there are no main effects or interaction, a model that only includes a main effect of condition (model Cond), a model that only includes a main effect of block (model Block), a model that includes main effects of condition and block (model CondBlock), and a full model that includes both main effects and an interaction. Separate GLMM analyses were performed for accuracy and RT. The accuracy results are described in Table 1 and the RT results are shown in Table 2.

For the accuracy analysis, the best-fitting model was CondBlock, suggesting both main effects were significant, but not the interaction. The Bayes factors (BF) suggest that the evidence for both main effects is extreme, and the evidence that there is no interaction is also extreme (Lee & Wagenmakers, 2014). An examination of Fig. 6 suggests that the main effect of condition is driven by the higher accuracy in the RDT condition than in the IDT condition. This is not surprising since the RDT transfer stimuli were objectively easier to categorize than the IDT transfer stimuli (i.e., see Fig. 1). The main effect of block is driven by the lower accuracy during the button-switch block compared to the control or dual-task blocks, and the lack of an interaction suggests that the lower button-switch accuracy was similar in both conditions.

The RT analysis led to different conclusions. The evidence for both main effects was again extreme, but now the evidence for an interaction was also extreme. In particular, the Full model was, by far, the best-fitting model, and a comparison of the Bayes factors for the Full and CondBlock models suggests the evidence for an interaction was extreme.[2] Fig. 6 suggests that the main effect of condition is driven by the faster RTs in the RDT condition and the main effect of block is largely due to the faster RTs during the control block. The difference between the control and button-switch RTs is approximately the same in the two conditions, so the highly significant interaction is driven by the much larger difference between the control and dual-task RTs in the RDT condition than in the IDT condition.

We also assessed all pairwise differences in Fig. 6 for significance via standard *t*-tests. These largely confirmed the GLMM analyses. In the IDT condition, the difference between control and dual-task accuracy was not significant [$t\,(14) = 0.70$, $p = .49$], nor was the RT difference [$t\,(14) = 1.73$, $p = .11$]. However, the differences between control and button-switch performance were significant – both for accuracy [$t\,(14) = -6.35$, $p < .005$] and RT [$t\,(14) = 4.88$, $p < .005$]. In the RDT condition, the difference between control and dual-task accuracy was not significant [$t\,(13) = 1.25$, $p = .23$], but the RT difference was significant [$t\,(13) = 4.46$, $p < .005$]. Finally, the control versus button-switch differences were both significant in the RDT condition [accuracy: $t\,(13) = -5.19$, $p < .005$; RT: $t\,(13) = 6.35$, $p < .005$].

The *t*-tests suggest that both conditions exhibited a button-switch interference that was characterized by a decrease in accuracy and an increase in RT (relative to control) when the response buttons switched locations. On the other hand, these tests also suggest no effect on accuracy of the dual task in either condition, but a significant increase in RT in the RDT condition only. To examine this RT difference more closely, Fig. 7 shows the median RTs (averaged across participants) during each 40-trial block of the dual-task trials. Also shown for comparison are these mean RTs during the categorization-only trials. Note that in both conditions, responding is slower in block 1 than in any subsequent blocks – presumably because there was a settling-in period as participants adjusted to the sudden demand to perform two tasks at once. Furthermore, RT dropped about equally from blocks 1 to 2 in both conditions. Therefore, this figure suggests that the most appropriate comparison is between performance on blocks 2–5. When dual-task RTs are compared to control RTs over these blocks, t-tests indicate that the effect of the dual task on RT was not significant in the IDT condition [$t\,(14) = 1.30$, $p = 0.22$], and highly significant in the RDT condition [$t\,(13) = 4.17$, $p = 0.001$].

2.2.2. Decision-bound modeling analysis

Before attempting to interpret these results, it is important to assess the type of decision strategy that participants were using. This is because a variety of different strategies could lead to approximately equal accuracies, and one group could have higher accuracy than another, not because they were more likely to use a strategy of the optimal type, but for some other reason (e.g., better criterial learning; less criterial noise). To examine this issue, we fit a variety of different decision-bound models (Ashby & Valentin, 2018; Maddox & Ashby, 1993) to the responses of individual participants separately during each of their 15 experimental sessions. The models assumed a procedural strategy, a rule-based strategy, or random guessing. These models are described in the Appendix, but briefly, the rule-based models assumed a single vertical or horizontal decision bound. The procedural-strategy model assumed that the decision bound was a single line of arbitrary slope and intercept, and the guessing models assumed that participants guessed randomly on each trial. The procedural and rule-based models all included a noise variance parameter, and either one (in the case of the

[2] The Bayes factors in Tables 1 and 2 estimate the likelihood of the model relative to the likelihood of the null model. The ratio of the Bayes factors for the Full and CondBlock models estimates the likelihood of the Full model relative to the CondBlock model.

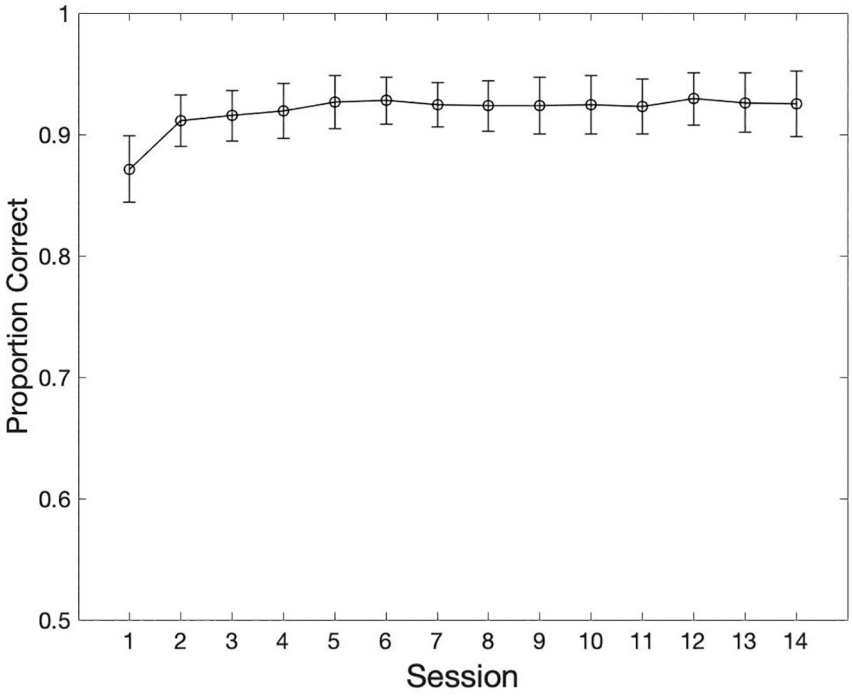

Fig. 4. Mean proportion correct for all training sessions of Experiment 1 averaged across participants. The error bars are 95% confidence intervals.

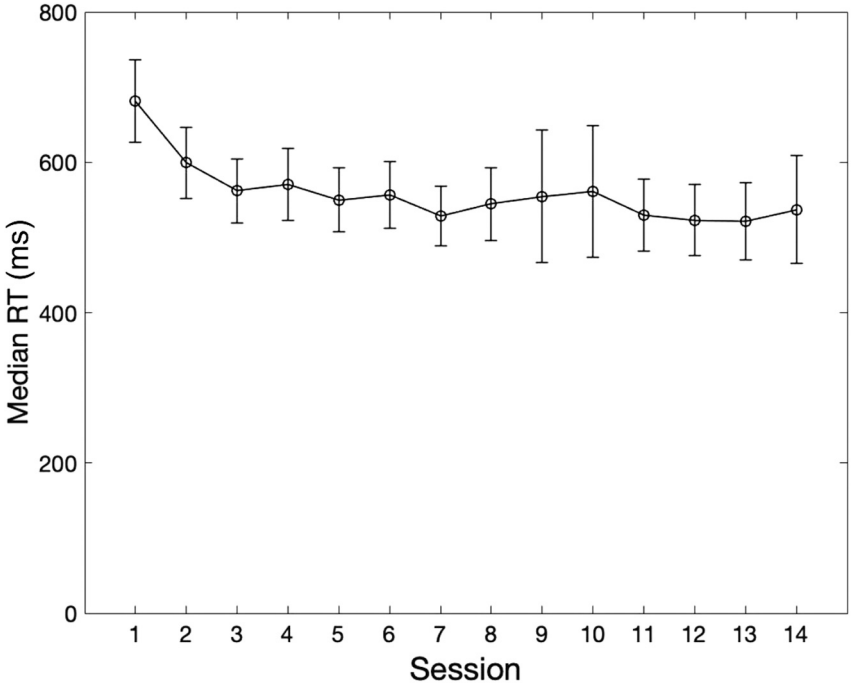

Fig. 5. Median RTs for all training sessions of Experiment 1 averaged across participants. The error bars are 95% confidence intervals.

rule models), or two (in the case of the procedural model) free parameters that described the decision bound. For every participant, each of these different models was fit separately to responses from each of the 14 training sessions, and to each of the three 200-trial blocks of the transfer session and in each case, the best-fitting model was recorded (i.e., the model with the lowest value of the BIC goodness-of-fit statistic).

During the first session, 86% of the participants' responses were best accounted for by a model of the optimal type – that is, a model that assumed a vertical line decision bound. During the other training sessions, this percentage ranged from 72% to 100%. In all cases that a vertical-bound rule model did not fit best, the best fit was provided by a model that assumed a procedural strategy. However, in all cases, visual examination of the decision bounds predicted by these models indicated a bound that was nearly vertical – suggesting that there were only a few trials in these data sets that included responses that were inconsistent with a vertical-bound rule. Overall, this analysis suggests that participants clearly learned the optimal categorization strategy early in training and used this strategy consistently throughout the 13

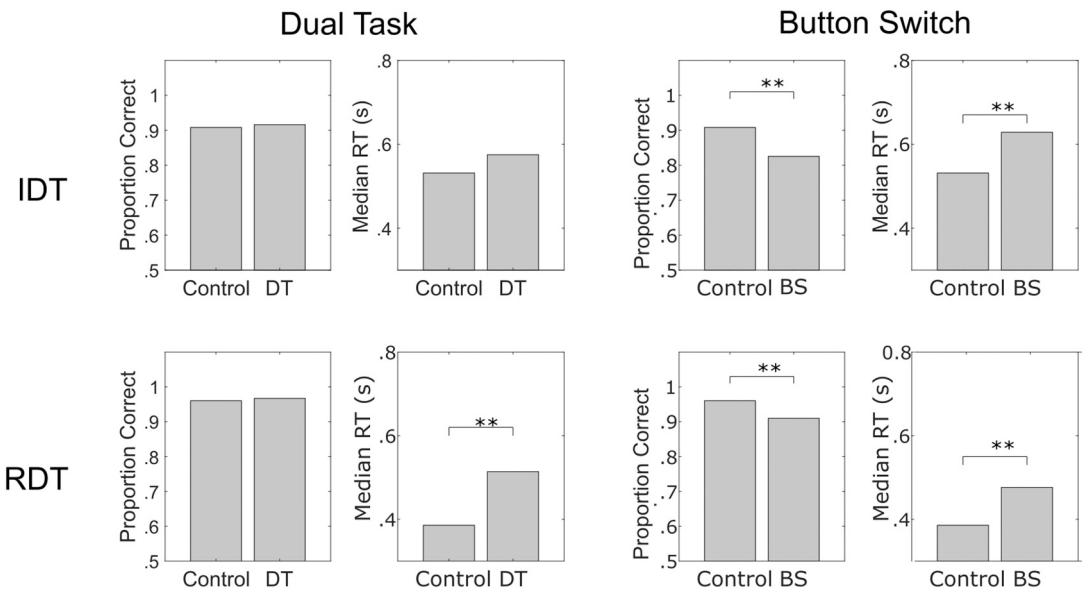

Fig. 6. Results from the final transfer session of Experiment 1. Control results are from the categorization-only block. DT = data from the dual-task block; BS = data from the button-switch block. Accuracy values are computed as a mean of each participant's proportion correct. RTs are the mean of each participant's median RT. Comparisons were performed with *t*-tests (** indicates $p < 0.005$).

Table 1
GLMM results for the accuracy data from the Experiment 1 transfer session.

Model	Terms	Log L	BIC	BF
Null	β_0	5126	10,263	1
Cond	$\beta_0 + C$	5015	10,050	1.5e46
Block	$\beta_0 + B$	5009	10,048	4.0e46
CondBlock	$\beta_0 + C + B$	4897	9832	2.9e93
Full	$\beta_0 + C + B + (C \times B)$	4895	9849	6.2e89

Table 2
GLMM results for the RTs from the Experiment 1 transfer session.

Model	Terms	Log L	BIC	BF
Null	β_0	13,634	27,288	1
Cond	$\beta_0 + C$	13,388	26,806	5.6e104
Block	$\beta_0 + B$	13,441	26,921	6.2e79
CondBlock	$\beta_0 + C + B$	13,189	26,427	9.5e186
Full	$\beta_0 + C + B + (C \times B)$	13,149	26,366	1.7e200

subsequent training sessions.

The results for the transfer session are shown in Table 3. Note that in both conditions, use of the optimal strategy was high in all three blocks. Therefore, the appearance of novel stimuli did not cause participants to switch strategies, nor did the presence of a dual task. Even the button switch had only a minor effect on strategy – confusing a few RDT participants enough to cause them to resort to guessing.

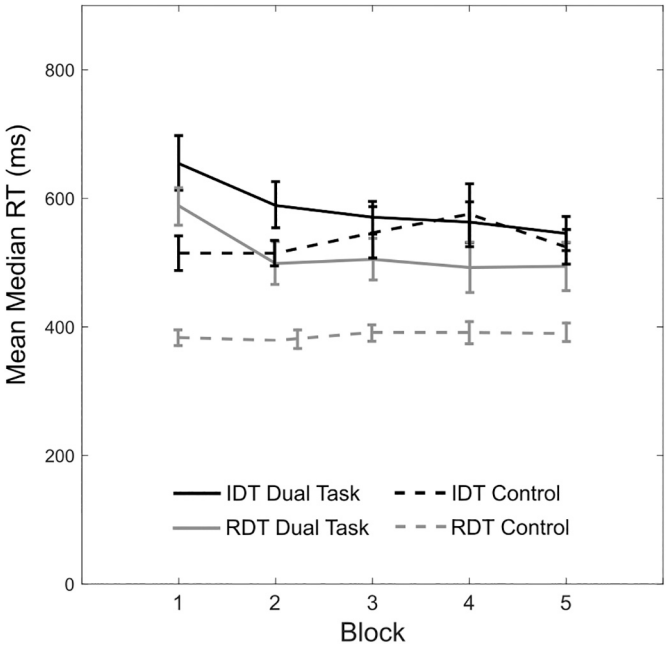

Fig. 7. The mean of all participants' median RTs for each 40-trial block during the Experiment 1 transfer-session dual-task trials. The dotted lines show the mean RTs from the categorization-only trials of the transfer session. The error bars denote standard errors.

2.3. Discussion

Twenty-nine participants each completed 8400 categorization training trials distributed over 14 experimental sessions. During this time they repeatedly practiced a simple one-dimensional categorization rule. Previous research suggests that after this amount of training, their responses were automatic. The participants were then divided into two groups and both groups completed one final session of 600 trials. During this last session, all participants saw new stimuli that could be categorized using the same rule that they had automatized during training. In the IDT condition, the new stimuli had identical values as the training stimuli on the relevant dimension and unique values on the irrelevant dimension. In the RDT group, the opposite occurred – that is, the new stimuli had novel values on the relevant dimension, but the values on the irrelevant dimension were the same as in training. We then assessed whether automaticity persisted for these novel stimuli by examining performance in the presence of a dual task, and following a switch of the response buttons.

Accuracy was universally high in both conditions, suggesting that participants had no trouble transferring the rule they had been practicing to the novel stimuli. Similar results have been reported after only one session of training (Casale et al., 2012), so this result is not unexpected.

Table 3
Decision-bound modeling results of the Experiment 1 transfer data. Number and percentage (in parentheses) of participants whose responses were best accounted for by each type of decision bound model.

Block	IDT	RDT
Single-Task Control		
Optimal 1D Rule	13 (87%)	13 (93%)
Procedural Strategy	2 (13%)	0 (0%)
Guessing	0 (0%)	1 (7%)
Dual Task		
Optimal 1D Rule	15 (100%)	13 (93%)
Procedural Strategy	0 (0%)	0 (0%)
Guessing	0 (0%)	1 (7%)
Button Switch		
Optimal 1D Rule	14 (93%)	9 (64%)
Procedural Strategy	1 (7%)	0 (0%)
Guessing	0 (0%)	5 (36%)

The more interesting results concern our tests of whether automaticity transferred to the novel stimuli that participants categorized during the transfer session. First, consider the IDT condition. Our results strongly suggest that automaticity transferred in this condition. In particular, there was no effect of the dual task on either accuracy or RT, whereas switching the locations of the response buttons decreased accuracy and increased RT. Both of these results are classic criteria of automatic responding (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977).

Next consider the RDT condition. Switching the response buttons decreased accuracy and increased RT, which is symptomatic of automatic responding. However, the dual-task results suggest a contradictory conclusion. Although the dual task had no effect on accuracy, it did significantly increase RT – by more than 100 ms. At first glance, it might seem that this interference could have been caused by a surprise effect – that is, that the surprise of seeing stimuli with novel values on the relevant dimension caused participants to respond more slowly. However, closer examination makes this hypothesis easy to reject. Most critically, Nosofsky (1991) reported that surprise effects of this type disappear after only two stimulus presentations. In Nosofsky's experiment, participants learned a one-dimensional categorization rule similar to the one used here. The stimuli were circles that varied in size and the orientation of a radial line. The single relevant dimension was size. After a training period, participants completed several transfer blocks in which a few trials included stimuli that were much larger than any seen during training. On the first two such trials, RT was significantly greater than on trials when the largest training stimuli were presented. But on the third and fourth such trials, responding was faster to these novel transfer stimuli than to any other stimuli. Therefore, the surprise effect persisted for only two trials. The RDT dual-task block included 200 trials, and Fig. 7 shows that the dual-task interference persisted for all 200 trials – far longer than any documented surprise effect. Fig. 7 does show that the dual-task interference was largest during the first 40 trials, and the Nosofsky (1991) results suggest that surprise might have contributed to this effect. Even so, Fig. 3 shows that after 180 trials of practice and long after there was any possibility that participants were still surprised by the stimuli, there was still a dual-task interference in the RDT condition of around 100 ms.

The classical interpretation of the dual-task interference that we observed in the RDT condition is that categorization was dependent on working memory and executive attention during the RDT dual-task trials, and therefore was no longer automatic (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). In fact, there is direct evidence linking dual-task interference to the "overloaded recruitment" of PFC working memory units (Watanabe & Funahashi, 2014).

On the other hand, the conclusion that automaticity did not transfer in the RDT condition requires more careful analysis because Hélie et al. (2010) concluded that the same qualitative pattern of results supported automaticity. Specifically, they reported that after 20 sessions of training on essentially the same category structure used here, and with the same stimuli, a similar simultaneous dual task had no effect on accuracy but significantly increased RT. They concluded from this result that, despite the RT interference, responding was automatic. What justifies a different conclusion here?

We believe that a number of results suggest that automaticity did not transfer in our RDT condition. First, if the dual-task interference on RT in the RDT condition occurred despite automatic responding, then the same interference should have been apparent in both conditions. However, we found no effect of the dual task on RT (or accuracy) in the IDT condition. This is especially noteworthy because the RDT categories were more widely separated than the IDT categories (i.e., see Fig. 1). Because of this greater separation, the stimuli in the RDT categories were objectively easier to categorize than the stimuli in the IDT categories. Despite this difficulty difference, the simultaneous dual task interfered more with the easier RDT categories than with the more difficult IDT categories, which strongly suggests that RDT responding was not automatic.

Second, the absence of a dual-task interference on accuracy can not be taken as evidence of automatic responding. When a dual task is introduced on the very first trial of initial training, it significantly impairs learning, in the sense that accuracy is lower at every point of training than in a single-task control group (Waldron & Ashby, 2001; Zeithamova & Maddox, 2006). However, in the present experiment, there is nothing left to learn during the dual-task transfer blocks. Rather than learn a rule, participants only have to apply a well-learned and highly practiced rule. The Kovacs et al. (2021) model predicts that participants will be able to do this accurately, regardless of whether they respond automatically, or whether they respond by appealing back to the learned rule.

Third, there are a number of reasons that the dual-task RT interference reported by Hélie et al. (2010) is more consistent with automaticity than with controlled rule application. First, Hélie et al. (2010) gave no RT instructions to their participants, and as a result there is no reason to believe they were responding as quickly as possible. In contrast, in the present experiment, participants were instructed to respond as quickly as possible without sacrificing accuracy. Second, Hélie et al. (2010) found an identical RT interference in an information-integration (II) categorization condition that is known to recruit procedural learning and memory, rather than rule learning. This is important because a dual task does not interfere with II category learning, even during the first session of training (Waldron & Ashby, 2001; Zeithamova & Maddox, 2006). Therefore, the RT interference in the II condition is inconsistent with either automatic or controlled responding, and instead suggests that the identical RT interference that Hélie et al. (2010) observed in all conditions might have been an artifact caused by some unrelated design feature. One possibility is that participants were given no RT instructions, but another possibility concerns the slightly different timing used in the two studies. In both studies, the Stroop digits were displayed first, followed by a blank screen, followed by the categorization stimulus. Participants then made their categorization response, followed by their dual-task response. In the current experiment, the digits were displayed for 1 s and the blank screen lasted for 300 ms. Therefore, participants had 1300 ms to encode the sizes and values of the Stroop digits before responding to the categorization stimulus. In the Hélie et al. (2010) experiment, the digits were displayed for 200 ms and the blank screen lasted for 100 ms, so participants only had 300 ms to encode the Stroop digits. Therefore, one hypothesis is that 300 ms was insufficient to complete this encoding and as a result, dual-task encoding persisted after the categorization stimulus was presented, thereby delaying the

categorization RT.[3]

In summary, we believe that the best account of our RDT results is that the button-switch results are consistent with automaticity, whereas the dual-task results are consistent with controlled responding, and therefore a loss of automaticity. Interestingly, this is exactly the pattern of results predicted by the Kovacs et al. (2021) model. Recall that this model predicts that rule-guided behaviors are initially triggered by the application of explicit rules, which are represented primarily in PFC, but after the behaviors become automatic they are initiated by projections from the stimulus representations in visual cortex directly to the relevant motor representations in PMC. Therefore, a change in the values of the relevant stimulus dimension should activate representations in visual cortex that project to untrained synapses in PMC. As a result, automatic responding is lost. Even so, the correct rule representation remains in PFC, so accuracy remains high. The cost though, is that suddenly relying on PFC makes the categorization susceptible to dual-task interference. On the other hand, the model also predicts that the projections from PMC to primary motor cortex are activated anytime a response is triggered, regardless of whether the PMC units are activated by direct projections from visual cortex (after automaticity) or by rule units in PFC (before automaticity and during transfer). Therefore, the model predicts a button-switch interference because of the 8400 previous button presses that participants made in this task.

The model does not make strong predictions about the results of the IDT condition – primarily because it does not completely describe the nature of the stimulus representations that are used to activate units in PMC. Certainly a change in values on the relevant stimulus dimension would cause the stimulus representations to change. But the model makes no predictions about whether a change in values of the irrelevant dimension will cause the stimulus representations to change. There are two clear alternatives. First, the stimulus representations used to select responses in one-dimensional categorization tasks could be gestalts. In this case, the model makes the same predictions in both conditions, because the stimuli changed between training and transfer in both conditions. The second possibility though, is that selective attention filters out irrelevant stimulus information, in which case the stimulus representations used to select responses depend only on values on the relevant stimulus dimension. In this case, the stimulus representations that were projected to PFC and PMC in the IDT condition were identical during training and transfer, so the model predicts that automatic responding will transfer to the novel stimuli. Our results strongly support this latter hypothesis. In the IDT condition, the dual-task and button-switch results were both consistent with automaticity – that is, there was no dual-task interference on either accuracy or RT, and the button-switch interference was significant for both dependent measures.

The sample sizes in the RDT and IDT conditions were relatively modest (14 and 15, respectively), which raises the question of whether Experiment 1 was sufficiently powered. Unfortunately, computing power for the appropriate GLMMs is statistically challenging, not only because of the multiple factors included in the experiment, but also because accurate power estimation requires knowledge of both the within- and between-participant variability. As a result, the standard approach is to estimate power from thousands of simulated data sets (e.g., Kumle, Võ, & Draschkow, 2021), and even then, these estimates are only valid if all the sources of variance are correctly specified. Because we know of no prior literature that could be used to estimate between-participant variability, we did not attempt these simulations. However, there are several reasons why we believe that Experiment 1 was sufficiently powered. First, although the most critical statistical analyses were restricted to data collected during the final transfer session, each participant completed 14 prior sessions that included a total of 8400 trials. This extensive training strongly decreases within-participant variability in both accuracy and RT (e.g., Hélie et al., 2010), which means that our design should be more powerful than the typical categorization experiment with the same number of participants that excludes the extensive prior training. Second, the Bayes factors show that the evidence supporting the critical RT interaction was extreme, and power analyses are most critical when interpreting nonsignificant effects.[4] Third, Experiment 2 tests a prediction that follows directly from our interpretation of the Experiment 1 results. As we will see, that prediction was strongly confirmed, which increases confidence in our interpretation of the Experiment 1 results.

3. Experiment 2

The results of Experiment 1 suggest that automatic rule-guided behaviors are not initiated by some abstract verbal rule, but rather directly by the visual stimulus – and more specifically, only by the relevant dimension(s) of the visual stimulus. This conclusion seems to conflict with results reported by Roeder and Ashby (2016), who concluded that abstract rules are automatized in RB categorization tasks. The experimental design used by Roeder and Ashby (2016) and a summary of their results are shown in Fig. 8. Each participant in this experiment completed 21 sessions that included 7 consecutive 3-day cycles. During days 1 and 2 of each cycle, participants practiced on the primary categories shown in panel (a) of Fig. 8, whereas on the third day of each cycle they practiced the secondary categories. At the beginning of each session, participants were told whether the categories that day were primary or secondary, although they were never given any other instructions about the category structures or about what categorization strategy they should use. Note that the optimal strategy on the primary categories is a logical disjunction: "Respond A if the stimulus has a small value on dimension 1 or if the stimulus has a large value on dimension 1; otherwise respond B." In contrast, for the secondary categories the optimal strategy is a simple one-dimensional rule.

An examination of panel (a) of Fig. 8 shows that half the stimuli changed category membership on days when the secondary categories were practiced and half the stimuli retained their primary category assignments. The stimuli that retained the same category assignment on all days, called congruent stimuli, are denoted in Fig. 8 by black symbols, whereas stimuli that switched assignments, called incongruent stimuli, are denoted by gray symbols.

The key data-analysis question was whether performance differed on congruent and incongruent stimuli. If an abstract rule is automatized then there should be no difference because the rules on primary and secondary days are different. However, if stimulus-response associations are automatized then performance should be worse on incongruent stimuli, which is exactly what Roeder and Ashby (2016) observed in a separate group of participants who practiced on II categories that are known to recruit procedural learning and memory systems. The RB results are shown in the bottom panel of Fig. 8. Note that on primary days, there was no difference in accuracy or RT between congruent and incongruent stimuli, and on this basis, Roeder and Ashby (2016) concluded that abstract rules are automatized in RB tasks.

However, on further reflection, the Roeder and Ashby (2016) results do not necessarily conflict with the results of our Experiment 1. The Experiment 1 results suggest a refinement of the Kovacs et al. (2021) Fig. 3 model in which the projections from visual cortex to PFC and PMC are restricted to visual representations of the relevant stimulus dimension(s) only. The Roeder and Ashby (2016) primary and secondary categories had different relevant dimensions. Therefore, this hypothesis predicts that the visual projections on primary and secondary days will be from different visual units onto different synapses in PFC and PMC and therefore practicing different stimulus-response

[3] We thank Sebastien Hélie (personal communication) for suggesting this account.

[4] If an effect is nonsignificant, then the only possible error is a type 2 error, and power is one minus this probability.

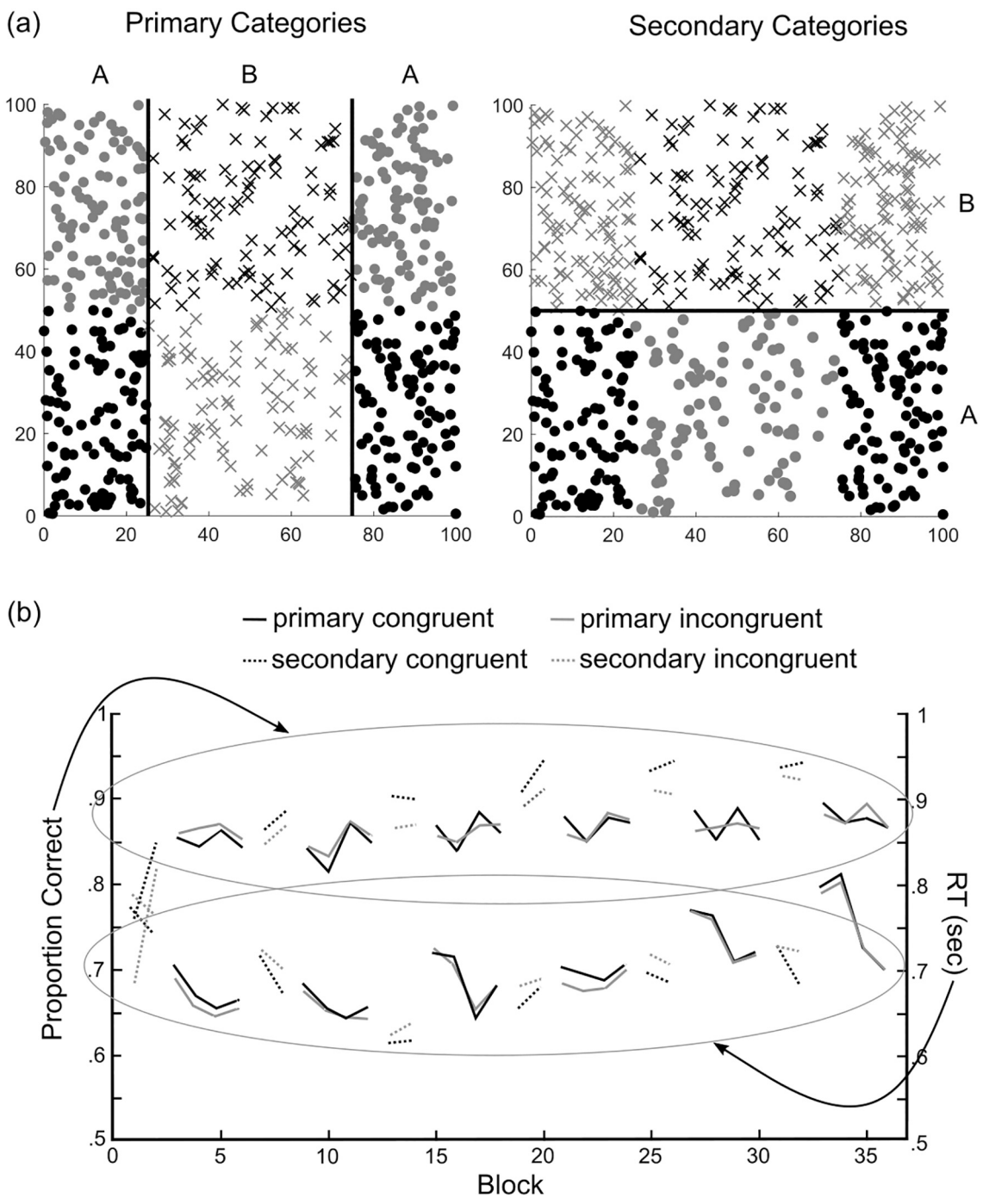

Fig. 8. (a) Categories used in the rule-based condition of the experiment reported by Roeder and Ashby (2016). Congruent stimuli that maintained their same category assignment on primary and secondary days are shown in black, whereas incongruent stimuli that switched assignments are shown in gray. (b) Proportion corrects and RTs over the first 20 experimental sessions of the experiment.

associations on incongruent stimuli during secondary days will not interfere with associations formed on primary days. Our hypothesis is that, from the perspective of PMC, completely different stimuli were used on primary and secondary days and therefore, there were no stimuli in the Roeder and Ashby (2016) study that switched response assignments.

Experiment 2 tests this prediction by replicating the design of Roeder and Ashby (2016), except with category structures for which the revised Kovacs et al. (2021) model predicts that the incongruent stimuli should cause interference. The stimuli and categories we used in Experiment 2 are shown in Fig. 9. As in Roeder and Ashby (2016), Experiment 2 included seven consecutive 3-day cycles. On the first two days of each cycle, participants practiced the primary categories shown in Fig. 9. On the third day of each cycle, they practiced the secondary categories. At the beginning of each day, participants were instructed about whether they would be practicing the primary or secondary categories during that session, but they were never given any instructions about the nature of the categories.

Note that, as in the Roeder and Ashby (2016) experiment, half the stimuli in Experiment 2 switch their category assignments on primary and secondary days, and half maintain their same assignment on all days. Also note that the primary categories are identical in the two

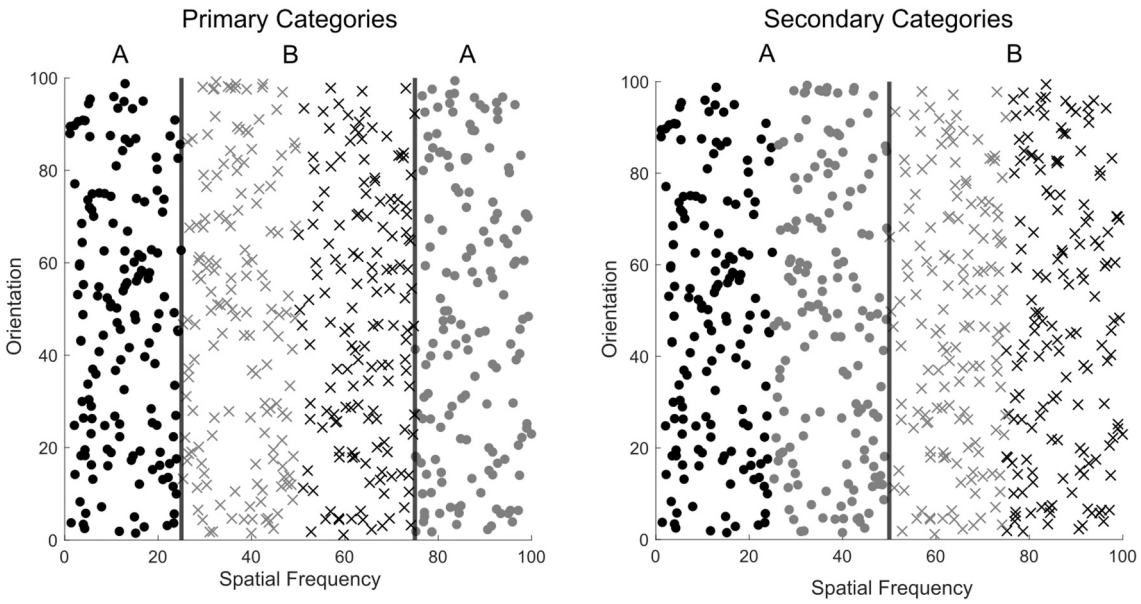

Fig. 9. Stimuli and category structures used in Experiment 2. Congruent stimuli that maintained their same category assignment on primary and secondary days are shown in black, whereas incongruent stimuli that switched assignments are shown in gray.

experiments, and in both experiments the secondary categories are separated by a simple one-dimensional rule. However, unlike Roeder and Ashby (2016), the same stimulus dimension is relevant on all days in our Experiment 2. Therefore, the revised Kovacs et al. (2021) model predicts that, in contrast to the results of Roeder and Ashby (2016), performance should be worse on incongruent stimuli than on congruent stimuli.

3.1. Methods

3.1.1. Participants

Thirty-one undergraduate students at the University of California, Santa Barbara participated in this experiment in exchange for course credit.

3.1.2. Stimuli and apparatus

Due to COVID restrictions, participants performed the experiment at home on their own home computers. As in Experiment 1, all stimuli were circular sine-wave gratings that varied across trials in spatial frequency (i.e., bar width) and bar orientation. Each stimulus was defined by a set of points (x_1, x_2) sampled from a 100 × 100 stimulus space and converted to a disk using the following equations: spatial frequency = $.1x_1$ + 0.25 cycles per disk and orientation = $.9x_2$ degrees counterclockwise rotation from horizontal.[5]

There were two different kinds of sessions: primary and secondary. The experiment included seven 3-day blocks, during which participants practiced the primary categories on the first two days and the secondary categories on the third day. The secondary session was omitted from the last cycle, so the entire experiment included 20 sessions over 20 nearly consecutive days.

The stimuli were the same as in Experiment 1, as were the events that occurred on each trial, and their timing. The category structures are shown in Fig. 9. On primary days, the optimal rule was a 1D disjunctive rule. On secondary days, the optimal rule was a simple 1D rule. In both sessions, the single relevant stimulus dimension was spatial frequency.

During primary sessions, stimuli in category A were uniformly distributed (in the 100 × 100 space) in two distinct intervals [0, 25] and [75, 100] on the spatial frequency dimension and [0,100] on the orientation dimension. Stimuli in category B were uniformly distributed (in the 100 × 100 space) in the interval [25,75] on the spatial frequency dimension and [0, 100] on the orientation dimension. During secondary sessions, stimuli in category A were uniformly distributed (in the 100 × 100 space) in the interval [0, 50] on the spatial frequency dimension and [0, 100] on the orientation dimension. Stimuli in category B were uniformly distributed (in the 100 × 100 space) in the interval [50,100] on the spatial frequency dimension and [0, 100] on the orientation dimension.

3.1.3. Procedure

The trial-by-trial procedures were identical to Experiment 1, except participants were informed that they would be participating in two different kinds of sessions, primary and secondary. They were instructed that the optimal strategy would be different on the secondary days, but they were given no instructions about the nature of the categories or about the type of strategies they should employ. At the beginning of each session, participants were informed about whether they would practice primary or secondary categories on that day.

3.2. Results

Fig. 10 shows the accuracy results for each 300-trial block and Fig. 11 shows the means of the median RTs. Data from the first two days are omitted because at this point in the experiment – that is, before the first secondary session – there were no incongruent stimuli. Note that accuracy is considerably higher for congruent stimuli in every session and RT is lower. A comparison back to Fig. 8 shows that these results are strikingly different from those of Roeder and Ashby (2016).

To test these conclusions statistically, we used the same GLMM analyses as in Experiment 1. We ran these analyses separately for all the data combined, the data only from primary sessions, and the data only from secondary sessions. The results were similar in all cases, but the results from the primary sessions are most important because the

[5] Note that the transformation to spatial frequency was nonlinear in Experiment 1 and linear in Experiment 2. This is because the Experiment 1 IDT transfer stimuli differed from the training stimuli in spatial frequency, so the range of perceived bar widths was much greater in Experiment 1 than in Experiment 2. In fact, the range was great enough that we felt it important to account for the nonlinear relationship between spatial frequency and perceived bar width.

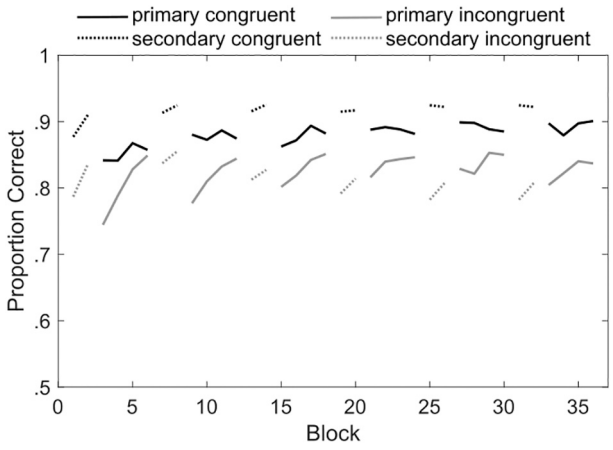

Fig. 10. Proportion correct in Experiment 2 shown separately for congruent and incongruent stimuli on primary and secondary days.

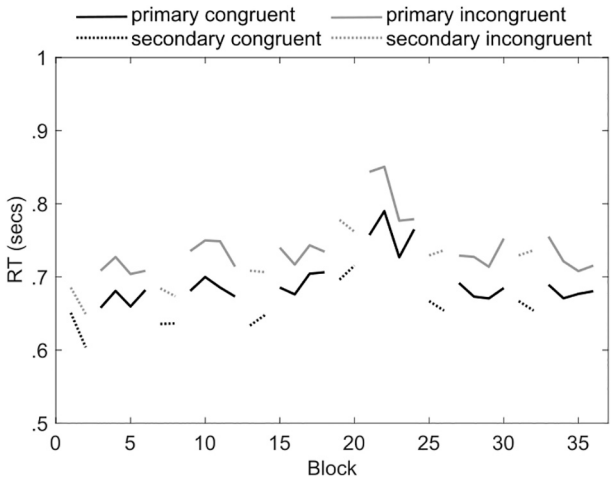

Fig. 11. Means (across participants) of the median RTs in Experiment 2 shown separately for congruent and incongruent stimuli on primary and secondary days.

number of primary sessions (i.e., 14) was chosen to ensure that responding had become automatic by the end of training (according to results of Hélie et al., 2010). As a result, this section focuses on the results from the primary categories only.

The accuracy analyses are shown in Table 4 and the RT analyses are shown in Table 5. In both cases, we tested models that included a main effect of session, a main effect of congruence (congruent stimuli versus incongruent stimuli), and an interaction. As described in the Methods, due to COVID restrictions, all participants performed the experiment at home on their personal computers. As a result, there were more frequent extreme RT outliers than in typical laboratory experiments. Therefore, as a conservative approach, we excluded from the RT analyses all RTs longer than 5 s. Fig. 11 shows that the median RTs were all well below 1 s, so any RT $>$ 5 s was almost surely due to some irrelevant distraction.

Table 4
GLMM results for the accuracy data from the Experiment 2 primary sessions.

Model	Terms	Log L	BIC	BF
Null	β_0	93,366	186,745	1
Congruence	$\beta_0 + C$	92,676	185,376	2.3e297
Session	$\beta_0 + S$	93,086	186,321	2.0e92
CongSess	$\beta_0 + C + S$	92,394	184,947	2.8e390
Full	$\beta_0 + C + S + (C \times S)$	92,334	184,963	9.6e386

Table 5
GLMM results for the RTs from the Experiment 2 primary sessions.

Model	Terms	Log L	BIC	BF
Null	β_0	1,720,526	3,441,076	1
Congruence	$\beta_0 + C$	1,720,494	3,441,024	2.5e11
Session	$\beta_0 + S$	1,720,171	3,440,503	5.0e124
CongSess	$\beta_0 + C + S$	1,720,139	3,440,450	1.4e136
Full	$\beta_0 + C + S + (C \times S)$	1,720,131	3,440,569	2.2e110

Table 6
Decision bound modeling results for Experiment 2.

Model	Number of sessions	Percentage
Primary Sessions		
Disjunctive Classifier	431	99.3
1D: Bar Width	3	0.7
Procedural Strategy	0	0
Guessing	0	0
Secondary Sessions		
Disjunctive Classifier	60	32.3
1D: Bar Width	122	65.6
Procedural Strategy	4	2.2
Guessing	0	0

This criterion excluded 1.2% of the RTs from the primary sessions (2638 out of 223,200 total RTs).

Table 4 shows that for the accuracy analysis, the best-fitting model (CongSess) included both main effects but no interaction. The Bayes factors (BF) suggest that the evidence for both main effects is extreme, as is the evidence that there is no interaction (Lee & Wagenmakers, 2014). An examination of Fig. 8 suggests that the main effect of congruency is driven by the higher accuracy for congruent stimuli that was evident in every experimental block. Note that this same difference also occurred with the secondary categories, where it was even more extreme. In fact, the main effect of congruency was highly significant even when we analyzed data from all sessions together and when we analyzed data from the secondary sessions only.

Table 5 shows that for the RTs, the best-fitting model again included both main effects but no interaction. And as with the accuracy analysis, the Bayes factors (BF) suggest that the evidence for both main effects is extreme, as is the evidence that there is no interaction. Fig. 9 shows that the main effect of congruency is driven by the faster RTs for congruent stimuli that was evident in every experimental block, and that this same effect was seen with both primary and secondary categories.

The accuracy and RT results support the predictions of the revised Kovacs et al. (2021) model only if participants were using the disjunction rule shown in Fig. 9 on primary days. For example, our labeling of stimuli as congruent or incongruent assumed this strategy. High accuracy and low RT is possible with multiple strategies, so a strategy analysis is needed to supplement our GLMM analyses of accuracy and RT. For this reason, we fit decision-bound models to the responses of each individual participant from each of their 20 experimental sessions. The models, which are described in the Appendix, were the same as the models used in Experiment 1, except the rule-based models also included a model that assumed participants used a disjunction rule.

Each of the 31 participants completed 14 sessions with the primary categories and 6 sessions with the secondary categories. Therefore, we fit all the models to 434 sets of primary session data (31 $\times$ 14) and 186 sets of secondary session data (31 $\times$ 6). The results are summarized in Table 6. The disjunctive classifier assumed a disjunction rule of the type that is optimal on primary days, the "1D: bar width" model assumed a one-dimensional rule of the type that is optimal on secondary days, the procedural strategy model assumed that perceptual information from both dimensions was (pre-decisionally) integrated, and the guessing models assumed random guessing (see the Appendix for details). Note that on the critical primary days, the participants used a disjunction rule

of the optimal type during almost every session. This result greatly increases confidence in our interpretation of the GLMM results.

Several points are worth noting about the results from the secondary sessions. First, participants almost always used a rule-based strategy (i.e., on 97.8% of the sessions). Second, participants used a rule of the optimal type (i.e., a one-dimensional rule on bar width) on most of the sessions (i.e., about two-thirds). Third, the disjunctive classifier that was optimal on primary days provided the best fit on about one-third of the sessions. This is not too surprising since participants had twice as much practice with the disjunction rule, and by the end of training they had automatized this rule. Note though, from Fig. 9, that if the disjunction rule was used on every trial during secondary sessions, accuracy would be only 50%, whereas Fig. 8 shows that accuracy on secondary sessions averaged about 85% correct. A closer examination of the secondary sessions for which the disjunctive classifier provided the best fit indicated that in almost every case, only a few responses were incompatible with the optimal one-dimensional rule. These few responses allowed the disjunctive classifier to fit better, even though the great majority of responses were compatible with a one-dimensional rule.[6] Therefore, we believe that our results suggest that virtually all participants used a one-dimensional rule of the optimal type on all but a few trials on each secondary day. However, about a third of the secondary sessions included a few trials in which participants inadvertently applied the more well-practiced disjunction rule.

3.3. Discussion

Although the design of Experiment 2 was nearly identical to the design used by Roeder and Ashby (2016), the results of the two experiments were strikingly different. Whereas Roeder and Ashby (2016) found no difference on primary days in either accuracy or RT for congruent versus incongruent stimuli, we found that responding was more accurate and faster for congruent than for incongruent stimuli. A comparison of Figs. 8 and 9 shows that the two experiments used identical primary categories, and in both experiments the secondary categories required a simple one-dimensional decision rule. The only difference was that in the Roeder and Ashby (2016) experiment, the relevant dimension on secondary days was irrelevant on primary days, whereas in our Experiment 2, the same stimulus dimesion was relevant on all days.

Our results are inconsistent with the conclusions of Roeder and Ashby (2016) that participants automatize an abstract rule in RB tasks. In both experiments, the rule on primary and secondary days was different, so if participants had automatized a rule, the two experiments should have yielded identical results. On the other hand, the results of both experiments are predicted by the revised version of the Kovacs et al. (2021) model in which the projections from visual cortex to PFC and PMC are restricted to visual representations of the relevant stimulus dimension only (see Fig. 3). In the Roeder and Ashby (2016) experiment, the relevant dimension changed from primary to secondary days, and as a result the model predicts that the visual input to PMC was fundamentally different on primary and secondary days. In other words, the model predicts that the effective stimuli were completely different on primary and secondary days, and as a result, the network mediating automaticity did not recognize any stimuli as being incongruent. In contrast, in our Experiment 2, because the same stimulus dimension was relevant on primary and secondary days, the model predicts that the visual projections into PMC were the same on every day, and therefore performance was worse on incongruent stimuli because of the interference that was caused by practicing competing motor responses on primary and secondary days.

[6] The maximum-likelihood-based goodness-of-fit statistic that we used (i.e., BIC) assigns an extreme penalty to any response that is incompatible with the assumed decision rule (e.g., to any B response in the presumed A response region), and this penalty gets much worse the further the discrepant response is from the decision boundary.

4. General discussion

This article describes the results of two extensive experiments that included a combined total of 633,000 categorization trials. The experiments investigated the nature of what is automatized after lengthy practice with a rule-guided behavior by testing novel predictions of a recent neurocomputational model (Kovacs et al., 2021). The results of both experiments suggest that an abstract rule, if interpreted as a verbal-based strategy, was not automatized during training, but rather the automatization linked a set of stimuli with similar values on one visual dimension to a common motor response.

It is important to note, however, that our results do not suggest that participants no longer had easy access to an abstract rule after automaticity developed. In fact, the Kovacs et al. (2021) model predicts that access to the abstract rule is always available via projections from visual cortex to PFC (see Fig. 3). However, the model predicts that after automaticity has developed, the behavior is not initiated by this indirect path to PMC, but rather by a faster, direct projection from visual cortex, and that it is only this direct projection that links stimuli with similar values on one visual dimension to a common motor response. Support for this prediction comes from reports that, after automaticity has developed, rule-sensitive neurons in the PMC of monkeys fire before rule-sensitive neurons in PFC (Wallis & Miller, 2003).

Our results clarify a number of puzzling results in the literature. First, categorization tasks, like the ones used here, in which the optimal bound is a vertical or horizontal line (and in which the stimulus dimensions are perceptually separable) are known as rule-based (RB) tasks in the literature. These are often compared to information-integration (II) tasks that are identical, except the categories are rotated 45° in stimulus space (so the separating decision bound is now diagonal). One curious, and previously unexplained result is that capuchin and macaque monkeys both learn these one-dimensional RB categories more quickly and to a higher asymptotic accuracy than the rotated II categories (Smith et al., 2010; Smith et al., 2015; Smith, Crossley, et al., 2012). Humans show an even more pronounced RB advantage than macaques, whereas pigeons and rats learn rotated RB and II category structures at exactly the same rate (Ashby et al., 2020; Broschard et al., 2019; Smith et al., 2011; Smith et al., 2012). Furthermore, the RB advantage shown by humans (and monkeys) is not because of an inherent difference in task difficulty, but rather because humans learn the two tasks in qualitatively different ways (Ashby et al., 2020).

One leading account of human category learning, called COVIS, proposes that humans learn RB categories by experimenting with simple, explicit rules and that in II tasks they instead rely on procedural learning (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Ashby & Waldron, 1999). The COVIS acronym stands for COmpetition between Verbal and Implicit Systems because the original proposal was that the learning of rules depends on verbal strategies. However, the superior performance of macaques in RB versus II tasks is strong evidence that verbalization is not a necessary condition for the RB advantage. So why are monkeys better at RB tasks than in rotated II tasks?

The present results offer an answer to this question. Monkeys are better at one-dimensional RB tasks than in rotated II tasks because they can allocate executive attention selectively to the single relevant stimulus dimension in the RB task, and this ability is not language dependent. In fact, the evidence is good that PFC plays a key role in this type of top-down selective attention (e.g., Desimone & Duncan, 1995). Macaque monkeys have a well-developed PFC, and so it is not surprising that there is much neural evidence for feature-based selective attention in monkeys (e.g., Fuster, 1990; Maunsell & Treue, 2006). Therefore, our results suggest that the most fundamental difference between rotated RB and II tasks may not so much be that language facilitates RB learning, but rather that selective visual attention does, whereas this attentional

ability provides no benefit in II tasks.

Second, our results offer an alternative interpretation of the many reports of rule-sensitive neurons in PMC (Muhammad et al., 2006; Vallentin et al., 2012; Wallis & Miller, 2003). These studies reported single-unit recordings from neurons in PMC that fired when a monkey applied one of two categorization rules. Furthermore, these neurons did not fire when the alternative rule was applied, and the neural responses were the same regardless of which stimulus was shown and what cue was used as a signal to the animal about which rule to apply. Neurons with similar firing properties have frequently been found in PFC (Asaad, Rainer, & Miller, 2000; Hoshi, Shima, & Tanji, 2000; White & Wise, 1999), but finding such neurons in PMC is somewhat surprising, given that the primary function of PMC has long been thought to be the selection of motor actions. Our results suggest that rule-sensitive neurons in PMC might not be implementing a categorization rule as it is commonly interpreted, but rather linking a set of stimuli with similar values on one visual dimension to a common motor response.

Third, our results suggest that the automatization of rule-guided behaviors and procedural skills might not be fundamentally different. Ashby, Ennis, and Spiering (2007) proposed that the automatic execution of procedural skills is mediated entirely within cortex and that the development of automaticity is associated with a gradual transfer of control from the basal ganglia circuits that mediate initial procedural learning to cortical-cortical projections from the relevant sensory areas directly to units in areas of PMC that initiate the behavior. According to this account, a critical function of the basal ganglia is to train purely cortical representations of automatic procedural behaviors (Hélie, Ell, & Ashby, 2015). The Kovacs et al. (2021) model proposes a similar account of the automatization of rule-guided behaviors, except for two key differences. First, in the case of rule-guided behaviors, the PFC trains the automatic cortical representations, rather than the basal ganglia. And second, the PMC targets are rule-sensitive units, rather than units associated with a specific motor goal. Despite these differences, both models assume that the development of automaticity is a gradual transfer of control from neural networks that mediate initial learning to direct projections between sensory association areas of cortex and PMC. The current results reduce the differences between these two theories because they suggest that the PMC targets in the two models are not fundamentally different. For both procedural and rule-guided behaviors, the PMC targets link sensory representations to motor behaviors. Our results suggest that the only real difference might be in the nature of the visual representations – gestalts in the case of procedural skills and single stimulus dimensions in the case of rule-guided behaviors.

Finally, at a more speculative level, our results might also be used to reflect on possible developmental origins of rule use. If rules are only abstract sets of verbal instructions, then their learning must necessarily be language dependent. If so, then procedural learning that is mediated by basal ganglia circuits can play at most a minor role in their acquisition. However, our results elevate the role that selective attention might play in this process and, together with the capuchin and macaque results (Smith et al., 2010; Smith et al., 2015; Smith, Crossley, et al., 2012), suggest that rule automatization might not necessarily even require language. Furthermore, the fact that our results reinforce neuroscience theories of automaticity that propose similar accounts for behaviors that are initially rule-guided versus mediated by procedural learning, suggests that rules might develop from an initial period of procedural learning. Together, all of these considerations suggest an intriguing hypothesis that might be worth developing and testing. First, initial rule use begins with a period of procedural learning that is facilitated by dopamine-mediated reinforcement learning in the basal ganglia (as described e.g., by Ashby & Crossley, 2010 and Cantwell, Crossley, & Ashby, 2015). Second, this process simultaneously trains cortical-cortical projections from the visual areas that respond to the stimulus to the relevant PMC targets (as proposed by Ashby et al., 2007). Finally, PFC selective-attention circuits directed at these visual representations begin to filter out irrelevant stimulus information (e.g., Feldman, 2021), leading to an end result in which the PMC targets receive input only about the relevant stimulus dimension.

In summary, our results suggest that the common interpretation that rule-guided behavior is mediated by a verbal-based strategy that implements a set of explicit instructions, is valid, at most, only for a period of initial learning. After rule-guided behaviors are practiced long enough to become automatic, they appear to no longer be mediated by anything resembling a rule, but instead to be triggered directly by the visual stimulus. Similar proposals have been made for automatic behaviors that are initially acquired via procedural learning, so our results suggest that behaviors that are acquired via rule or procedural learning, although initially depending on very different neural networks, may be mediated in almost identical ways after they become automatized. The only real difference appears to be that in the case of rule-guided behaviors, top-down selective attention whittles away irrelevant visual information, in the sense that the automatic behavior is triggered by visual representations that depend only on relevant stimulus information.

CRediT authorship contribution statement

Paul Kovacs: Conceptualization, Methodology, Software, Formal analysis, Investigation. **F. Gregory Ashby:** Conceptualization, Methodology, Resources, Visualization.

Declaration of Competing Interest

None.

Appendix

This appendix provides a brief overview of the decision bound modeling (DBM) used to investigate the strategies participants used in both experiments. For more details, including exact equations that describe each model, see Ashby and Valentin (2018), or Maddox and Ashby (1993).

In DBM, a series of models are fit to each participant's response data. To monitor the learning process, all models were fit to the 600 trials from each successive training session separately for each participant. In Experiment 1, the strategies participants used during each transfer block were examined by fitting all models separately to each of the three 200-trial blocks of the transfer session. Experiment 1 included a total of 29 participants, who each completed 14 training sessions, and 3 blocks of 200 trials during the final transfer session. So in total, all models were fit to 493 different data sets [i.e., $29 \times (14 + 3)$]. Experiment 2 included 31 participants who each completed 14 sessions with the primary categories and 6 sessions with the secondary categories. In this case, we fit all the models to 434 sets of primary session data (31×14) and 186 sets of secondary session data (31×6). For each of these data sets, we compared the performance of three qualitatively different types of models: models that assumed the use of an explicit rule, models that assumed a procedural strategy, and models that assumed participants guessed on every trial.

A.1. Models that assume an explicit rule

There were two types of models in this class. The one-dimensional (1D) model assumes that the participant sets a criterion on a single stimulus dimension and uses that criterion to separate the categories. The 1D model has two free parameters: the decision criterion and the variance of perceptual and criterial noise. There were two versions of this model – one that assumed selective attention to orientation, and one that assumed selective attention to bar width.

The disjunctive classifier assumes that the participant sets two criteria on a single dimension and uses those criteria to partition the attended dimension into three intervals – small, medium, and large – and then to give one response to stimuli that fall in the small or large regions and the contrasting response to stimuli falling in the medium region. The disjunctive classifier has three free parameters: two decision criteria and the variance of perceptual and criterial noise. There were four versions of this model, depending on which dimension was attended, and whether an A or B response was given to the middle response region. This model was only used in the analysis of the Experiment 2 results.

A.2. Models that assume a procedural strategy

One model assumed a procedural strategy – namely, the general linear classifier (GLC). The GLC assumes the participant separates the categories using a linear decision bound. When the decision bound is neither vertical nor horizontal, it mimics a procedural strategy in which information from the two dimensions is integrated pre-decisionally (i.e., in a linear fashion). The GLC has three free parameters: the slope and intercept of the decision bound, and the noise variance.

A.3. Models that assume guessing

Two models assumed the participant guessed on every trial. One model assumed A and B responses each were emitted with probability 0.5, and one model assumed that an A response was given with probability p and a B response was given with probability $1 - p$, where p is a free parameter. The former model has zero free parameters and the latter model has one (i.e., p). The former model is useful for identifying participants who try, but fail to learn, and the latter model is useful for identifying participants who ignore the stimulus and simply press the same response key on every trial (in which case, the best-fitting parameter value is either $p = 0$ or $p = 1$).

A.4. Model comparison

All model parameters were estimated using the method of maximum likelihood. The Bayesian Information Criterion (BIC) was used to determine which model best fit the data:

$$BIC = rln(N) - 2ln(L), \tag{1}$$

where N = sample size (i.e., number of trials in the sample), r = the number of free parameters (e.g., 3 for the GLC), and L is the model likelihood. Note that BIC penalizes models for both a bad fit and for the number of free parameters. A lower BIC is better, so the best-fitting model for each data set is the one with the lowest BIC.

Appendix B. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2022.105168.

References

Asaad, W. F., Rainer, G., & Miller, E. K. (2000). Task-specific neural activity in the primate prefrontal cortex. *Journal of Neurophysiology, 84*(1), 451–459.

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. (2010). Interactions between declarative and procedural-learning categorization systems. *Neurobiology of Learning and Memory, 94*(1), 1–12.

Ashby, F. G., Ennis, J. M., & Spiering, B. J. (2007). A neurobiological theory of automaticity in perceptual categorization. *Psychological Review, 114*(3), 632–656.

Ashby, F. G., Smith, J. D., & Rosedahl, L. A. (2020). Dissociations between rule-based and information-integration categorization are not caused by differences in task difficulty. *Memory & Cognition, 48*, 541–552.

Ashby, F. G., & Valentin, V. V. (2018). The categorization experiment: Experimental design and data analysis. In E. J. Wagenmakers, & J. T. Wixted (Eds.), *Stevens' handbook of experimental psychology and cognitive neuroscience, fourth edition, volume five: Methodology* (pp. 307–347). Wiley.

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

Broschard, M. B., Kim, J., Love, B. C., Wasserman, E. A., & Freeman, J. H. (2019). Selective attention in rat visual category learning. *Learning & Memory, 26*(3), 84–92.

Cantwell, G., Crossley, M. J., & Ashby, F. G. (2015). Multiple stages of learning in perceptual categorization: Evidence and neurocomputational theory. *Psychonomic Bulletin & Review, 22*, 1598–1613.

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

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18*(1), 193–222.

Feldman, J. (2021). Mutual information and categorical perception. *Psychological Science, 32*(8), 1298–1310.

Fuster, J. M. (1990). Inferotemporal units in selective visual attention and short-term memory. *Journal of Neurophysiology, 64*(3), 681–697.

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., 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.

James, W. (1914). *Habit*. New York: H. Holt.

Kovacs, P., Hélie, S., Tran, A. N., & Ashby, F. G. (2021). A neurocomputational theory of how rule-guided behaviors become automatic. *Psychological Review, 128*(3), 488–508.

Kumle, L., Võ, M. L.-H., & Draschkow, D. (2021). Estimating power in (generalized) linear mixed models: An open introduction and tutorial in R. *Behavior Research Methods, 53*(6), 2528–2543.

Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.

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

Maunsell, J. H., & Treue, S. (2006). Feature-based attention in visual cortex. *Trends in Neurosciences, 29*(6), 317–322.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience, 24*(1), 167–202.

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.
Nosofsky, R. M. (1991). Typicality in logically defined categories: Exemplar-similarity versus rule instantiation. *Memory & Cognition, 19*(2), 131–150.
Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods, 162*(1–2), 8–13.
Qadri, M. A., Ashby, F. G., Smith, J. D., & Cook, R. G. (2019). Testing analogical rule transfer in pigeons (*Columba livia*). *Cognition, 183*, 256–268.
Roeder, J. L., & Ashby, F. G. (2016). What is automatized during perceptual categorization? *Cognition, 154*, 22–33.
Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. detection, search, and attention. *Psychological Review, 84*(1), 1–66.
Sherrington, C. S. (1906). Observations on the scratch-reflex in the spinal dog. *The Journal of Physiology, 34*(1–2), 1–50.
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., Ashby, F. G., Berg, M. E., Murphy, M. S., Spiering, B., Cook, R. G., & Grace, R. C. (2011). Pigeons' categorization may be exclusively nonanalytic. *Psychonomic Bulletin & Review, 18*(2), 414–421.
Smith, J. D., Beran, M. J., Crossley, M. J., Boomer, J. T., & Ashby, F. G. (2010). Implicit and explicit category learning by macaques (*macaca mulatta*) and humans (*homo sapiens*). *Journal of Experimental Psychology: Animal Behavior Processes, 36*, 54–65.
Smith, J. D., Berg, M. E., Cook, R. G., Murphy, M. S., Crossley, M. J., Boomer, J., et al. (2012). Implicit and explicit categorization: A tale of four species. *Neuroscience & Biobehavioral Reviews, 36*(10), 2355–2369.
Smith, J. D., Crossley, M. J., Boomer, J., Church, B. A., Beran, M. J., & Ashby, F. G. (2012). Implicit and explicit category learning by capuchin monkeys (*Cebus apella*). *Journal of Comparative Psychology, 126*(3), 294–304.
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, 41*, 322–335.
Treutwein, B., Rentschler, I., & Caelli, T. (1989). Perceptual spatial frequency—orientation surface: Psychophysics and line element theory. *Biological Cybernetics, 60*(4), 285–295.
Vallentin, D., Bongard, S., & Nieder, A. (2012). Numerical rule coding in the prefrontal, premotor, and posterior parietal cortices of macaques. *Journal of Neuroscience, 32*(19), 6621–6630.
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.
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.
Watanabe, K., & Funahashi, S. (2014). Neural mechanisms of dual-task interference and cognitive capacity limitation in the prefrontal cortex. *Nature Neuroscience, 17*(4), 601–611.
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.