*Attention, Perception, & Psychophysics*
*2010, 72 (4), 1013-1031*
*doi:10.3758/APP.72.4.1013*

# Automaticity in rule-based and information-integration categorization

Sébastien Hélie, Jennifer G. Waldschmidt, and F. Gregory Ashby
*University of California, Santa Barbara, California*

Three experiments studied the effects of category structure on the development of categorization automaticity. In Experiment 1, participants were each trained for over 10,000 trials in a simple categorization task with one of three category structures. Results showed that after the first few sessions, there were no significant behavioral differences between participants who learned rule-based versus information-integration category structures. Experiment 2 showed that switching the locations of the response keys after automaticity had developed caused a similar highly significant interference, regardless of category structure. In Experiment 3, a simultaneous dual task that engaged executive functions did not interfere with either rule-based or information-integration categorization. These novel results are consistent with a theory assuming separate processing pathways for initial rule-based and information-integration category learning but a common processing pathway after the development of automaticity.

A typical adult makes hundreds of categorization judgments every day. Almost all of these are automatic. When we sit in a chair, pick up a book, or swerve to avoid a pothole, we are making an automatic categorization judgment. Adults sometimes make categorization decisions that are not automatic. For example, a dog owner might be learning to differentiate between Briards and Bouviers. Nevertheless, for most adults, categorization decisions based on newly acquired knowledge are far less common than categorization decisions made automatically. Despite this imbalance, initial category learning has been investigated much more extensively than categorization automaticity. For example, a search of PsycINFO yields 4,655 articles in response to the keywords "category or categorization learning," but only 57 articles in response to "category or categorization automaticity"—a ratio of 82 to 1.

Despite the many studies that have examined the ability of people to learn new perceptual categories, we know of only a few that have trained participants for more than a session or two on novel categories. In all of these, participants received at most a few thousand trials of training. For example, Maddox, Ashby, and Gottlob (1998) reported the results of an experiment in which each participant received about 7,000 trials of training. Ashby, Waldron, Lee, and Berkman (2001) reported one that included 4,000 trials of practice. Nosofsky and Palmeri (1997) reported results from an experiment in which each participant received 1,800 trials of training. Even so, the last two articles examined only a single type of category structure, and the former two focused only on asymptotic performance.[1] Thus, we know of no published studies that examined performance changes across a variety of different category structures as participants transitioned from novice to automatic responding.

This article aims to fill this void in the literature. Toward this end, we report the results of three experiments in which 36 participants each received more than 10,000 trials of categorization training spread over 20 or more separate experimental sessions (for a total of approximately 480,000 trials spread over 830 experimental sessions). The 36 participants each learned one of three qualitatively different category structures.

## Overview of the Experiments

The three category structures studied in this article are illustrated in Figure 1. In all three cases, the two categories were each composed of circular sine-wave gratings that varied across trials in the width and orientation of the dark and light bars (e.g., see the bottom right of Figure 1). The dotted lines denote the category boundaries. On each trial of Experiment 1, participants were shown one disk randomly selected from one of the two categories. The participant's task was to assign this disk to Category A or B by pressing the appropriate response key. Feedback was then provided about the accuracy of the response. Every participant repeated this procedure more than 10,000 times over the course of more than 20 experimental sessions. Each participant learned only one of the three category structures.

The two category structures shown at the top of Figure 1 are examples of rule-based categorization tasks, because they can be learned via an explicit reasoning process. In rule-based tasks, the rule that maximizes accuracy (i.e., the optimal strategy) is easy to describe verbally (Ashby, Alfonso-

**S. Hélie, helie@psych.ucsb.edu**

 

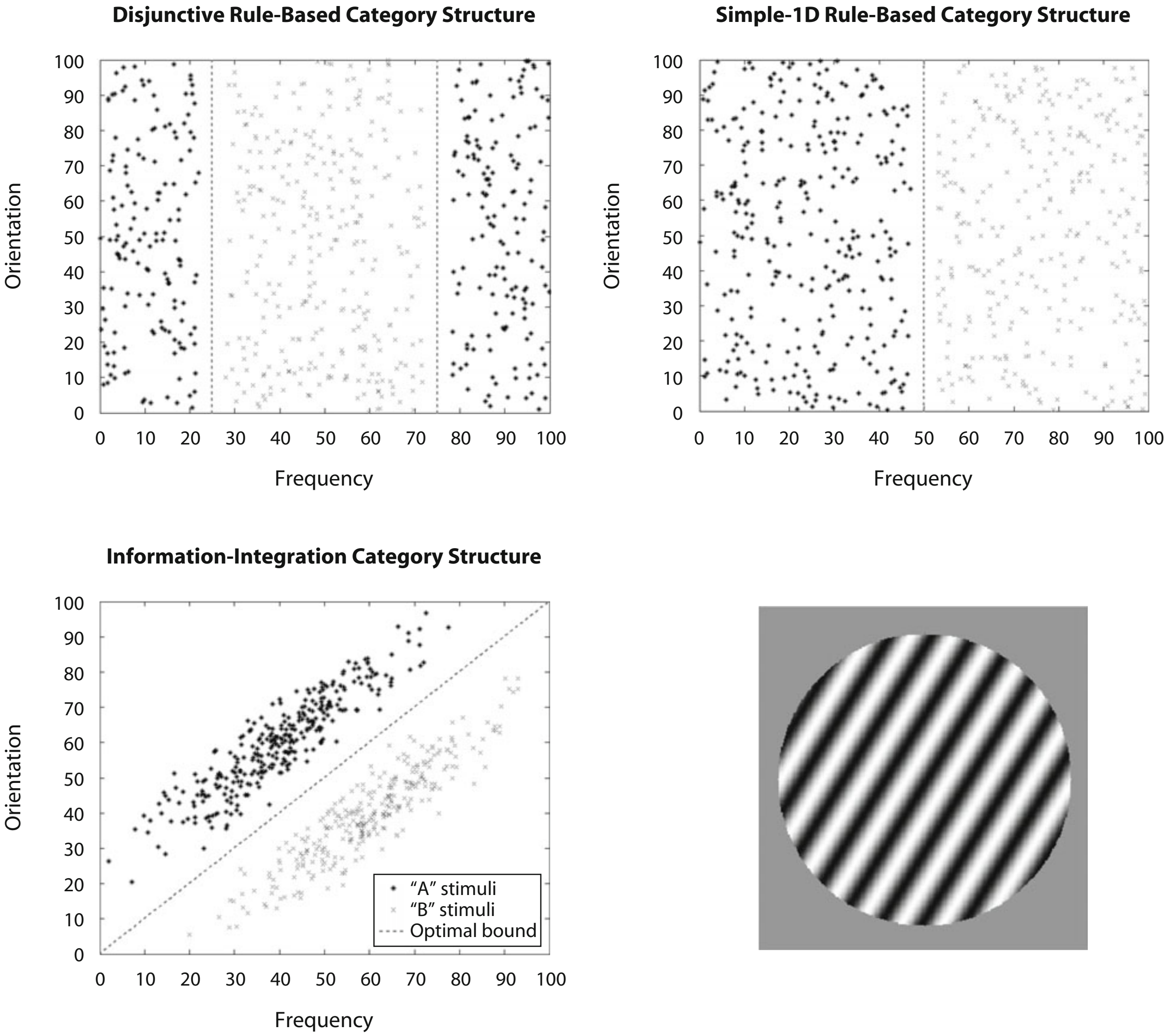


**Figure 1. Category structures used in the experiments. The top panels are rule-based conditions (left = disjunctive rule; right = simple-1D rule), and the bottom-left panel is an information-integration category structure. The optimal bounds in the top-left panel are $x_1 = 25$ and $x_1 = 75$. The optimal bound in the top-right panel is $x_1 = 50$. The optimal bound in the bottom-left panel is $x_2 = x_1$. The bottom-right panel shows an example stimulus.**

Reese, Turken, & Waldron, 1998). The top-right panel of Figure 1 shows the simplest and most widely studied rule-based task. Note that the optimal 1D rule here is "respond A if the bars are thick and B if they are thin." The top-left panel shows a more complex rule-based task in which the optimal strategy is to apply the disjunctive rule "respond A if the bars are thin or thick; otherwise respond B."

The bottom-left panel of Figure 1 shows an example of an information-integration categorization task. In information-integration tasks, accuracy is maximized only if information from two or more stimulus components (or dimensions) is integrated at some predecisional stage (Ashby & Gott, 1988). In many cases, the optimal strategy is difficult or impossible to describe verbally (Ashby et al., 1998). This is true in the bottom-left panel of Figure 1, because no simple verbal rule correctly separates the disks into the two categories. Nevertheless, many studies have shown that people reliably learn such categories, provided they receive consistent and immediate feedback after each response (for a review, see Ashby & Maddox, 2005).

## Category Learning and Memory

There is a growing consensus that human memory is mediated by multiple qualitatively distinct systems (Rolls, 2000; Squire & Schacter, 2002; Tulving, 2002). A growing body of recent evidence suggests that category learning uses many or perhaps all of the major memory systems that have been hypothesized by memory researchers (Ashby & O'Brien, 2005). This section briefly reviews some of that evidence (for more thorough reviews, see, e.g., Ashby & Maddox, 2005; Ashby & O'Brien, 2005).

**The role of working memory in rule-based categorization**. *Working memory* is the ability to maintain and manipulate limited amounts of information during brief periods of cognitive activity (Baddeley, 1986). It is heavily used in reasoning and problem solving and often associated with a wide variety of cognitive tasks. Because working memory is effective only for brief time intervals, it cannot store a lasting category representation, but it could be the primary mediating memory system in tasks where the categories are learned quickly. An obvious candidate for working memory is simple rule-based tasks.

Perhaps the best cognitive evidence that working memory is crucial for rule-based category learning comes from studies in which participants performed a dual task that required working memory and executive attention at the same time that they learned either simple one-dimensional rule-based categories or more difficult information-integration category structures that required attention to multiple stimulus dimensions (Waldron & Ashby, 2001; Zeithamova & Maddox, 2006). If the same memory system is used to learn both types of category structures, one would expect the dual task to interfere more strongly with the more difficult categorization task. The opposite result was observed. The dual task slowed learning in the one-dimensional rule-based task by as much as 350% without significantly affecting the difficult three-dimensional information-integration task. Thus, a dual task that requires working memory interferes with a simple rule-based task, but not with a more difficult information-integration task.

**The role of procedural memory in information-integration categorization**. Procedural memories are the memories of skills learned through practice (Willingham, 1998). Traditionally, these have been motor skills, such as those used when playing golf or tennis. Because procedural learning requires many repetitions, it is not likely to influence performance when the categories have a simple structure that can be discovered via logical reasoning. Instead, it seems more likely that procedural memory might mediate learning in tasks that are not easily learned via a logical reasoning process. In fact, such categories are common in everyday life. For example, the set of all X-rays displaying a tumor forms a perceptual category, but deciding whether a particular X-ray shows a tumor requires years of training, and expert radiologists are only partially successful at describing their categorization strategies.

Several studies have provided direct evidence that learning in information-integration tasks is mediated primarily by procedural memory. The quintessential paradigm for studying procedural learning is the serial reaction time (SRT) task (Nissen & Bullemer, 1987), in which participants press keys as quickly as possible in response to stimuli that appear in various locations on the screen. A large response time (RT) improvement is observed when the stimulus sequence is repeated, even when participants are unaware that a sequence exists. In addition, changing the location of the response keys interferes with SRT learning, but changing the fingers that push the keys does not (Willingham, Wells, Farrell, & Stemwedel, 2000). Thus, if procedural learning is used in information-integration tasks, switching the locations of the response keys should disrupt performance, but switching the fingers that depress the keys should not. In fact, Ashby, Ell, and Waldron (2003) reported evidence that directly supported this prediction. They also reported that neither manipulation had any effect on rule-based categorization. These results were replicated and extended in a number of subsequent studies (Maddox, Bohil, & Ing, 2004; Maddox, Glass, O'Brien, Filoteo, & Ashby, 2010; Spiering & Ashby, 2008).

**Summary**. Given that initial categorization performance is different in rule-based and information-integration tasks (because they seem to rely on different memory systems; Ashby et al., 2003; Waldron & Ashby, 2001), a natural question to ask is whether these differences persist after rule-based and information-integration judgments have been practiced long enough to become automatic. The main goal of the present experiments is to answer this question.

## Features of Automaticity

Given this goal, it is tempting to adopt some specific criteria that can be used to determine whether the categorization behaviors we are studying have become truly automatic. However, for two different but related reasons, we adopt instead a conservative position in which we operationally define automaticity simply as the result of extensive overtraining after the category structures are well learned (Moors & de Houwer, 2006; Schneider & Chein, 2003; Shiffrin & Schneider, 1977).

The first problem with identifying specific criteria to assess automaticity is that many different criteria have been proposed in the literature. A number of these have come from Schneider and Shiffrin (1977; Shiffrin & Schneider, 1977; for an updated list, see also Schneider & Chein, 2003). Perhaps the most popular of these is that a behavior should be considered automatic if it can be executed successfully while the participant is simultaneously engaged in some other secondary task (i.e., efficiency). Another widely used criterion proposed by Shiffrin and Schneider is that a behavior should be considered automatic if it becomes difficult to modify after training (i.e., behavioral inflexibility).

Even so, other authors have proposed different criteria. For instance, Logan (1988) proposed using a process-based definition of automaticity. According to Logan's (1988) instance-based theory, automatic behavior is the result of single-step memory retrieval. Hence, identifying the presence of automaticity becomes a problem of detecting the signature (features) of single-step memory retrieval in task performance. These features depend on assumptions about how memory retrieval is achieved. For instance, Logan (1988) assumed, among other things, that instances were automatically encoded and that memory retrieval was the result of a race among independent memory traces. As such, automaticity could be detected by the presence of a power law speedup of mean RTs and their standard deviations (*SD*s, with equal rates), item-specific facilitation for repeated stimuli, and the presence of separate memory traces for each stimulus presentation.

In addition, the RT distributions should be Weibull with a shape parameter constrained by the rate of the power law speedup (Logan, 1992). These criteria have little resemblance to Schneider and Shiffrin's (1977).

In the animal learning literature,[2] the most widely used automaticity criterion is that the behavior is largely independent of any ensuing reward (Dickinson, 1985). Again, this differs substantially from both Schneider and Shiffrin's (1977) and Logan's (1988) criteria. In addition, automaticity is frequently associated with unobservable features such as "unconsciousness" (for a review, see Moors & de Houwer, 2006). Lastly, it is unclear whether all these criteria need to be simultaneously present or how many need to be observed for a behavior to be labeled "automatic" (Moors & de Houwer, 2006). In summary, there is no single widely accepted criterion for assessing automaticity.

A second problem is that many of the popular behavioral criteria of automaticity were proposed before multiple memory systems were modeled and observed. For example, this is true for all of the criteria suggested by Shiffrin and Schneider (1977). To our knowledge, there have been no careful empirical investigations of whether these criteria should apply equally, regardless of the memory systems implicated. In fact, there is reason to believe that the memory systems do matter. For example, as mentioned above, several studies reported that a dual task that required working memory and executive attention (a measure of efficiency) interfered with initial rule-based category learning but not with information-integration category learning (Waldron & Ashby, 2001; Zeithamova & Maddox, 2006). Also, Ashby et al. (2003) reported that switching the position of the response buttons (a measure of behavioral inflexibility) interfered with initial information-integration performance, but not with initial rule-based performance. Therefore, blindly applying Shiffrin and Schneider's efficiency and inflexibility criteria would lead to the erroneous conclusion that information-integration categorization is automatic after the first training session. Such a conclusion would be incompatible with the intuitive notion of automaticity, because accuracy in information-integration tasks requires several thousand trials to asymptote.[3] For this reason, more work is required before the classic behavioral signatures of automaticity can be reconciled with multiple-memory-systems theories.

**Summary**. A second goal of this article is to examine whether several well-known automaticity criteria hold equally for overtrained behaviors that were initially mediated by declarative versus procedural memory systems. In particular, Experiments 2 and 3 focus on the Shiffrin and Schneider (1977) behavioral inflexibility (i.e., button switch) and efficiency (i.e., dual task) criteria (see also Crabb & Dark, 2003).

### Automaticity Versus Expertise

It is important to distinguish between automaticity and expertise. Expertise typically connotes some extra unusual training or experience not shared by most people. For example, according to Palmeri, Wong, and Gauthier (2004) "experts know more than novices. They can verbalize more properties, describe more relationships, make more inferences . . ." (p. 378). According to these definitions, a person who walks into a room and sits down in a chair without consciously making a categorization decision is showing evidence of automaticity, but such behavior by itself provides no evidence of that person's expertise in any furniture categories. Many studies have compared the categorization abilities of experts and novices (e.g., Johnson & Mervis, 1997; Medin, Lynch, Coley, & Atran, 1997), but because of the specialized training experts receive, these results tell us relatively little about normal, everyday categorization automaticity.

## EXPERIMENT 1

This experiment studies the effect of extensive practice (overtraining) on categorization performance (accuracy and RT). As mentioned earlier, many studies have reported that the initial learning and performance of information-integration and rule-based category structures are qualitatively different (for reviews, see Ashby & Maddox, 2005; Maddox & Ashby, 2004). The goal of this experiment is to overtrain participants to eliminate these performance differences as much as possible. Each participant in this experiment received feedback training on one of the three category structures shown in Figure 1 for 10,440 trials spread over 18 different experimental sessions. Three of these sessions were completed inside an fMRI scanner. This article focuses exclusively on the behavioral data from this experiment.

### Method

**Participants**. Thirty-seven healthy participants, predominantly undergraduate students at the University of California, Santa Barbara, were recruited to participate in Experiment 1. Fourteen participants were in the disjunctive-rule condition. Six of those participants completed 18 training sessions in the laboratory, and the remaining 8 completed 15 training sessions in the laboratory and 3 sessions in a 3T Siemens fMRI scanner. Twelve participants were trained in the simple-1D rule condition. Four of those participants completed 18 training sessions in the laboratory, and the remaining 8 completed 15 training sessions in the laboratory and 3 sessions in a 3T Siemens fMRI scanner. Eleven participants were trained in the information-integration condition, all of whom participated in 15 training sessions in the laboratory and 3 sessions in a 3T Siemens fMRI scanner.

Each participant was given credit or was paid between \$230 and \$350 for participation (depending on the amount of time spent in the fMRI scanner). One participant in the information-integration condition was excluded from the experiment due to an inability to learn the correct category structures by Session 5.

**Apparatus**. The stimuli were circular sine-wave gratings of constant contrast and size presented on a 21-in. monitor (1,280 × 1,024 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: frequency $= x_1/30 + 0.25$ cpd, and orientation $= 9x_2/10 + 20^\circ$. This yielded stimuli that varied in orientation from 20° to 110° and in frequency from 0.25 to 3.58 cpd. The stimuli were generated with MATLAB using Brainard's (1997) Psychophysics Toolbox, and occupied an approximate visual angle of 5°. An example is shown in the bottom-right panel of Figure 1.

For the disjunctive-rule condition (top-left panel in Figure 1), Category A stimuli were uniformly distributed in two different regions separated on the frequency dimension. Category A was defined as $x_1 < 23$ or $x_1 > 77$. Category B was defined as $27 < x_1 < 73$. These boundaries were chosen so that the areas of the two categories were the same. The optimal decision-bound model in this condition is the interval-based one-dimensional classifier (IB1D; see the Appendix).

For the simple-1D rule condition (top-right panel in Figure 1), the A category stimuli were uniformly distributed as $x_1 < 48$, and the B category stimuli were uniformly distributed as $x_1 > 52$. The optimal decision-bound model in this condition is the 1D model (see the Appendix).

For the information-integration condition (bottom-left panel in Figure 1), Category A stimuli were generated from a multivariate normal distribution with the following parameters (Ashby & Gott, 1988): $\mu_A = \{40, 60\}$; $\Sigma_A = \{185, 170; 170, 185\}$. The same sampling method was used to generate Category B stimuli: $\mu_B = \{60, 40\}$; $\Sigma_B = \Sigma_A$. The optimal decision-bound model in this condition is the *general linear classifier* (GLC; see the Appendix). Note that perfect accuracy was possible in all three conditions.

Stimulus presentation, feedback, response recording, and RT measurement were acquired and controlled using MATLAB 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). Auditory feedback was given for a correct (high-pitched tone) or incorrect (low-pitched tone) response. If a response was too late (more than 5 sec), participants saw the words "Too Slow." A participant who hit a wrong key heard a distinct beep and saw the words "Wrong Key."

**Procedure**. The experiment lasted for 18 sessions over 18 consecutive workdays. There were two types of sessions. "Regular" sessions occurred in the laboratory and were composed of 12 blocks of 50 stimuli (for a total of 600 stimuli). "Scanning" sessions could happen either in an fMRI scanner or in the laboratory (see the Participants subsection). In both cases, "scanning" sessions were composed of 6 blocks of 80 stimuli (for a total of 480 stimuli). "Scanning" sessions were Days 2, 4, and 10 in the information-integration condition and Days 1, 4, and 10 in the rule-based conditions. The remaining sessions were "regular." It should be noted that all participants had the same number of training trials. Participants who were not scanned still had simulated "scanning" sessions on Days 1, 4, and 10. In each session, half the stimuli were As and half were Bs.

Participants were told that they were taking part in a categorization experiment and that they had to assign each stimulus into either an A or a B category. The participants were allowed to take a break between blocks if they wished. A trial went as follows: A fixation point (crosshair) appeared on the screen for 1,500 msec and was followed by the stimulus, which remained on the screen until the participant made a response; correct or incorrect auditory feedback was given for 1,000 msec; "wrong key" or "too slow" feedback was given for 2,000 msec. Each participant completed a total of 10,440 trials.[4]

## Results

All the power analyses throughout this article were designed to detect a difference of 2.5% for accuracy and 20 msec for RTs, with $\alpha = .05$ and a within-subjects correlation of $\rho = .5$ (Barcikowski & Robey, 1985).[5]

**Accuracy**. The mean accuracy across sessions is shown in Figure 2A. As can be seen, all groups improved their accuracy with practice. The simple-1D rule condition was easiest at first, but accuracy in all three conditions reached roughly 95% correct after the third session. This accuracy level remained approximately constant throughout the remaining 15 sessions.

A condition (3, between subjects) × session (18, within subjects) ANOVA showed that the effect of session reached statistical significance [$F(17,561) = 19.15$, $p < .001$]. The mean accuracy was 84.9% in the first session and increased to 94.2% in the last session. The condition × session interaction was also statistically significant [$F(34,561) = 4.48$, $p < .001$]. Decomposition of the effect of condition within each level of session showed that the conditions differed during Sessions 1 [$F(2,33) = 6.37$, $p < .01$] and 3 [$F(2,33) = 4.32$, $p < .05$]. In the first session, the mean accuracies were 93.0% for the simple-1D rule group, 82.1% for the disjunctive-rule group, and 79.0% for the information-integration group. The accuracy of all three conditions was the same in all the following sessions [i.e., Sessions 4–18; all $F$s$(2,33) < 1.89$, n.s.]. The condition factor did not reach statistical significance [$F(2,33) = 0.75$, n.s.; power = 1].

**RTs**. Because the response environment and response keys were different inside the scanner, the RTs from scanning sessions were not analyzed (Sessions 2, 4, and 10 for the information-integration group, and Sessions 1, 4, and 10 for the rule-based groups). Median correct RTs in each session were individually computed for each participant.[6] The group-averaged medians of "regular" sessions are shown in Figure 2B. As can be seen, the median RTs from all three groups diminished with practice. Also, RTs seemed similar across conditions.

Because RTs from Session 2 were missing in the information-integration condition and the RTs from Session 1 were missing from the two rule-based conditions (they were "scanning" sessions), the following analysis included only "regular" sessions between 3 and 18. A condition (3, between subjects) × session (14, within subjects) ANOVA[7] showed that the median RTs diminished with practice [$F(13,429) = 6.90$, $p < .001$]. The mean median RT was 565 msec in Session 3 and decreased to 495 msec in Session 18. The main effect of condition [$F(2,33) = 1.95$, n.s.; power = 1] and the interaction between the factors [$F(26,429) = 1.13$, n.s.; power = 0.19] failed to reach statistical significance.

**Model-based analyses**. The accuracy-based analyses suggest that performance in the rule-based and information-integration conditions was similar; yet it is important to know whether each participant eventually adopted a decision strategy of the optimal type. To answer this question, we fit three different types of decision-bound models[8] (e.g., Maddox & Ashby, 1993) to the data from each individual participant in every session: rule-based, information-integration, and guessing models (see the Appendix for details). The rule-based models assumed either a single vertical or a horizontal bound, or that participants used either a conjunction or a disjunction rule. The information-integration models assumed that the decision bound was either a single line of arbitrary slope and intercept or a quadratic curve. Finally, as their name implies, the guessing models assumed that participants guessed randomly on each trial.

The percentage of participants whose data were best fit by a model that assumed a decision strategy of the optimal type is shown in Table 1. As can be seen, the responses of all the participants in the disjunctive-rule condition were best fit by a model that assumed an interval-based strategy, except for 1 participant in the first session (best fit by a model that assumed information integration). Like-

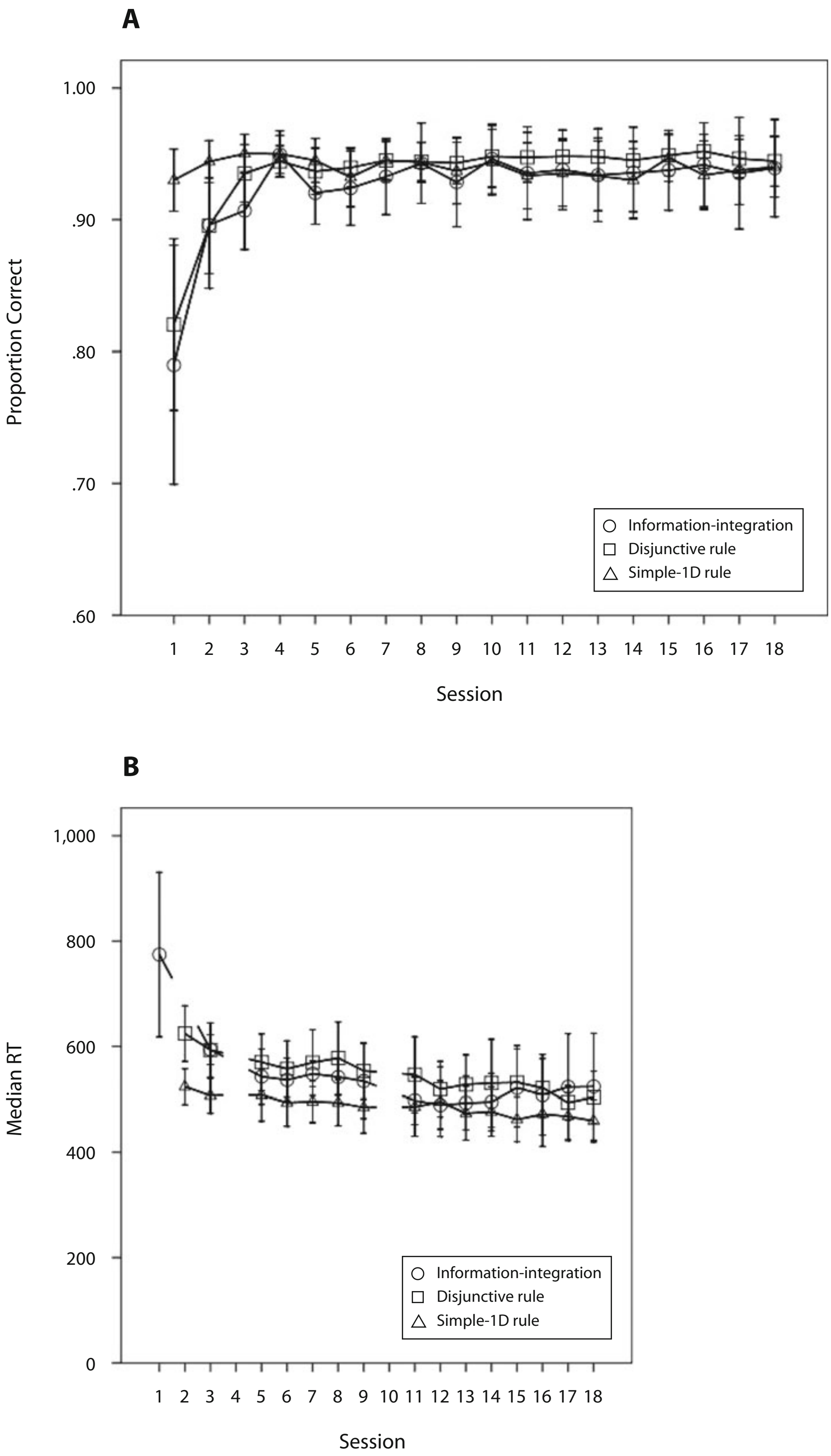


**Figure 2. (A) Proportion correct per training session. (B) Mean median correct response time (RT) per training session. The error bars are 95% confidence intervals.**

wise, the responses of 1 participant in the information-integration condition were best fit by a guessing model in the first session, but, in all other cases, the best-fitting model assumed information integration.

The responses of most participants in most sessions were best described by a model that assumed the appropriate one-dimensional rule in the simple-1D rule condition. However, there were some sessions in which the best fits were by models that assumed other strategies (mostly information integration; in one case, a conjunction rule fit best). In each one of these cases, however, the best-fitting models emulated a one-dimensional rule (i.e., the decision bound was a vertical line for the information-integration models, or, for the model that assumed a conjunction rule, the horizontal bound had a negative intercept on the orientation dimension).

One measure of learning is whether participants eventually adopt a strategy of the optimal type. Table 1 shows that participants in all three conditions successfully met this criterion. Another measure of learning is to ask whether the consistency with which participants applied this strategy improved with training. This issue can be addressed by examining estimates of the noise variance parameter from each of the best-fitting models. For every model, the noise variance will increase with greater perceptual or criterial noise, or if there is trial-by-trial variability in the participant's decision strategy. Noise reductions are therefore helpful in improving categorization performance and are important factors in the development of automaticity.

Figure 3A shows mean estimates of the noise *SD*s from the best-fitting model for each session of all three conditions. As can be seen, these *SD*s decreased sharply across the first four sessions in the information-integration and disjunctive-rule conditions. After Session 4, they fluctuated about a steady-state value.[9] In the simple-1D rule condition, they fluctuated about a (higher) steady-state value across all sessions. This was the easiest condition, and good performance was possible without having to discriminate between the stimuli lying close to the category boundary. This may explain the higher amount of estimated noise in this condition.

**Distance-to-bound analyses**. Many studies have established that stimuli close to the category boundary are more difficult to categorize than stimuli that are farther away (i.e., with longer RTs and higher error rates; e.g., Maddox et al., 1998). An interesting question is whether this distance-to-bound effect diminishes with practice. For example, stimuli near the bound might be processed less efficiently, because perceptual noise makes category membership uncertain. Because perceptual learning should reduce perceptual noise, it may also reduce distance-to-bound effects.

Figure 4 plots the group-averaged proportion correct against distance-to-bound.[10] As can be seen, the distance-to-bound effect diminished with practice for the disjunctive-rule and information-integration groups, but not for the simple-1D group. This is consistent with the higher noise estimates for the simple-1D group (see Figure 3A).

Figure 5 plots the group-averaged median correct RTs against distance-to-bound. As can be seen, participants in the information-integration condition initially had a smaller distance-to-bound effect than participants in the two rule conditions. However, the differences between the conditions tend to decrease with practice, and the curves are strongly overlapping after 14 sessions of training (bottom line in Figure 5); so, even though regular accuracy and RT analyses did not show any difference between the conditions after 3 sessions of practice, learning continues and the conditions become increasingly similar with extended practice.

**Table 1**
**Percentages of Participants in Experiment 1 Whose Data Were Best Fit by a Model That Assumed a Decision Strategy of the Optimal Type**

| Session | Simple-1D | Disjunctive Rule | Information-Integration |
|---|---|---|---|
| 1 | 92 | 93 | 90 |
| 2 | 100 | 100 | 100 |
| 3 | 100 | 100 | 100 |
| 4 | 92 | 100 | 100 |
| 5 | 92 | 100 | 100 |
| 6 | 100 | 100 | 100 |
| 7 | 92 | 100 | 100 |
| 8 | 100 | 100 | 100 |
| 9 | 100 | 100 | 100 |
| 10 | 92 | 100 | 100 |
| 11 | 83 | 100 | 100 |
| 12 | 100 | 100 | 100 |
| 13 | 83 | 100 | 100 |
| 14 | 92 | 100 | 100 |
| 15 | 92 | 100 | 100 |
| 16 | 100 | 100 | 100 |
| 17 | 83 | 100 | 100 |
| 18 | 100 | 100 | 100 |

Note—In the simple-1D condition, the performance of all participants in each session was best described by a model that assumed a vertical decision bound. Entries less than 100 denote cases in which a model that assumed an information-integration strategy or a conjunction rule fit best. In all of these cases, however, the best-fitting model emulated predictions from a 1D model (i.e., the best-fitting bound was a vertical line).

## Discussion

The results of this experiment show that participants in all three conditions reached a similar level of speed and accuracy after extensive practice. All groups were performing at the same speed after the second session of practice, and had similar accuracies after the third session of practice. In addition, the accuracy results in the first three sessions suggest that the disjunctive-rule condition was as difficult to learn as the information-integration condition; so, any difference between these two conditions cannot be attributed to task difficulty alone. Meanwhile, the simple-1D rule condition was easier than the other conditions, as suggested by the high accuracy of the participants in the first session. Even so, performance in all conditions was similar in the later sessions. Model-based analyses showed that the responses of most participants in all three conditions were best fit by optimal decision-bound models, and that perceptual/criterial noise tended to decrease with practice. Finally, the distance-to-bound effect seemed to diminish with practice. The only exception was in the simple-1D condition, where participants remained relatively poor at categorizing stimuli close to the boundary, even after 18 sessions of practice. As de-

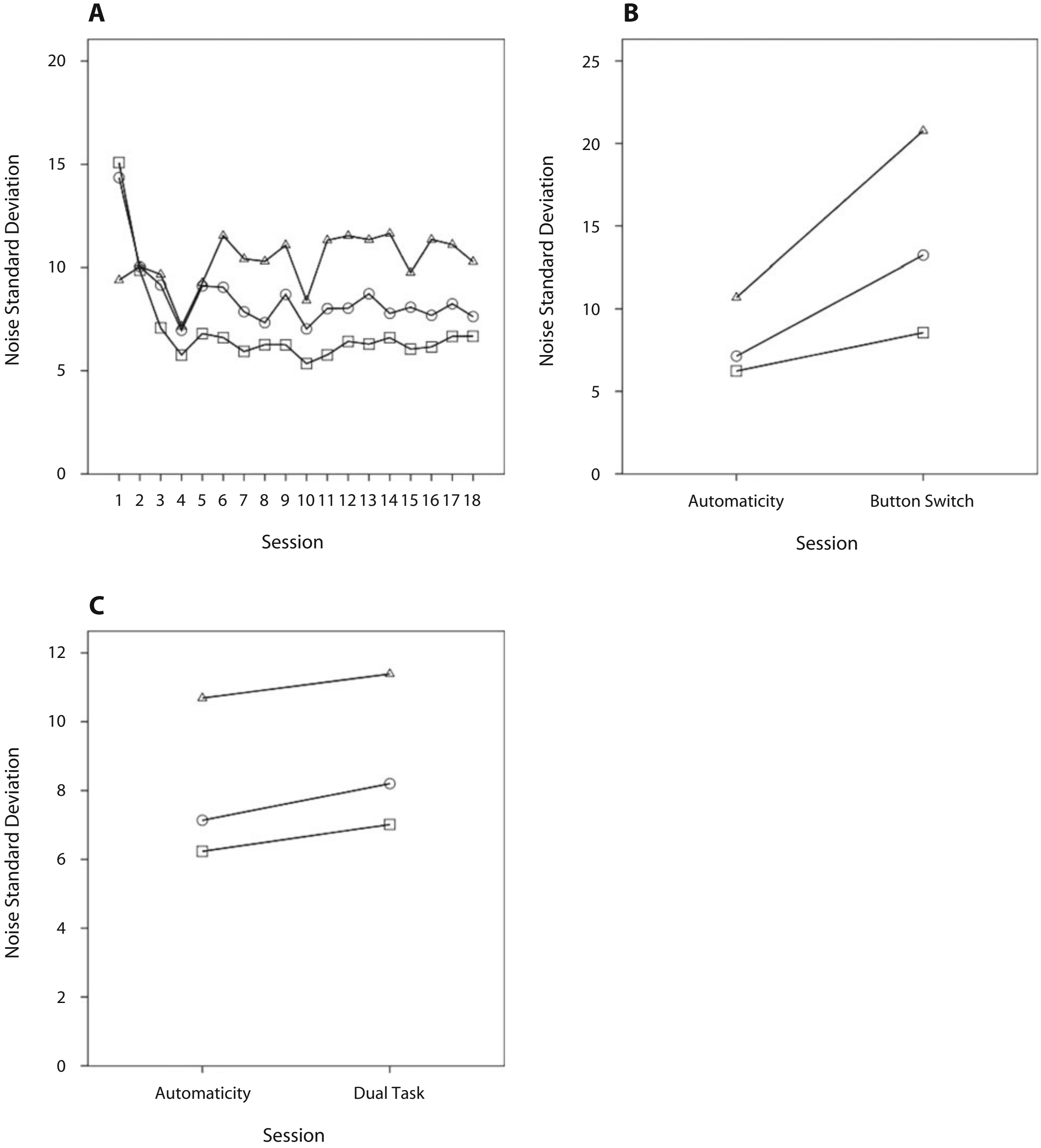


**Figure 3. Mean estimated noise parameter of the correct best-fitting models (A) during the development of automaticity (Experiment 1), (B) for automatic categorization performance and the button switch session (Experiment 2), and (C) for automatic categorization performance and the dual-task session (Experiment 3). In all panels, circles represent the information-integration condition, squares represent the disjunctive-rule condition, and triangles represent the simple-1D rule condition.**

scribed earlier, this was the only condition in which high accuracy on stimuli near the category boundary was not required for good performance.

This experiment shows that performance in information-integration and rule-based categorization becomes similar after overtraining. Specifically, the quantitative differences found in early learning were effectively eliminated. The high power values calculated in this experiment (due to the large within-subjects sample sizes) strongly suggest that a true accuracy difference of 2.5% or a true RT difference of 25 msec would have been detected by these analyses. Thus, the nonsignificant results imply that any real difference not detected must almost surely be less than these values (and of little importance

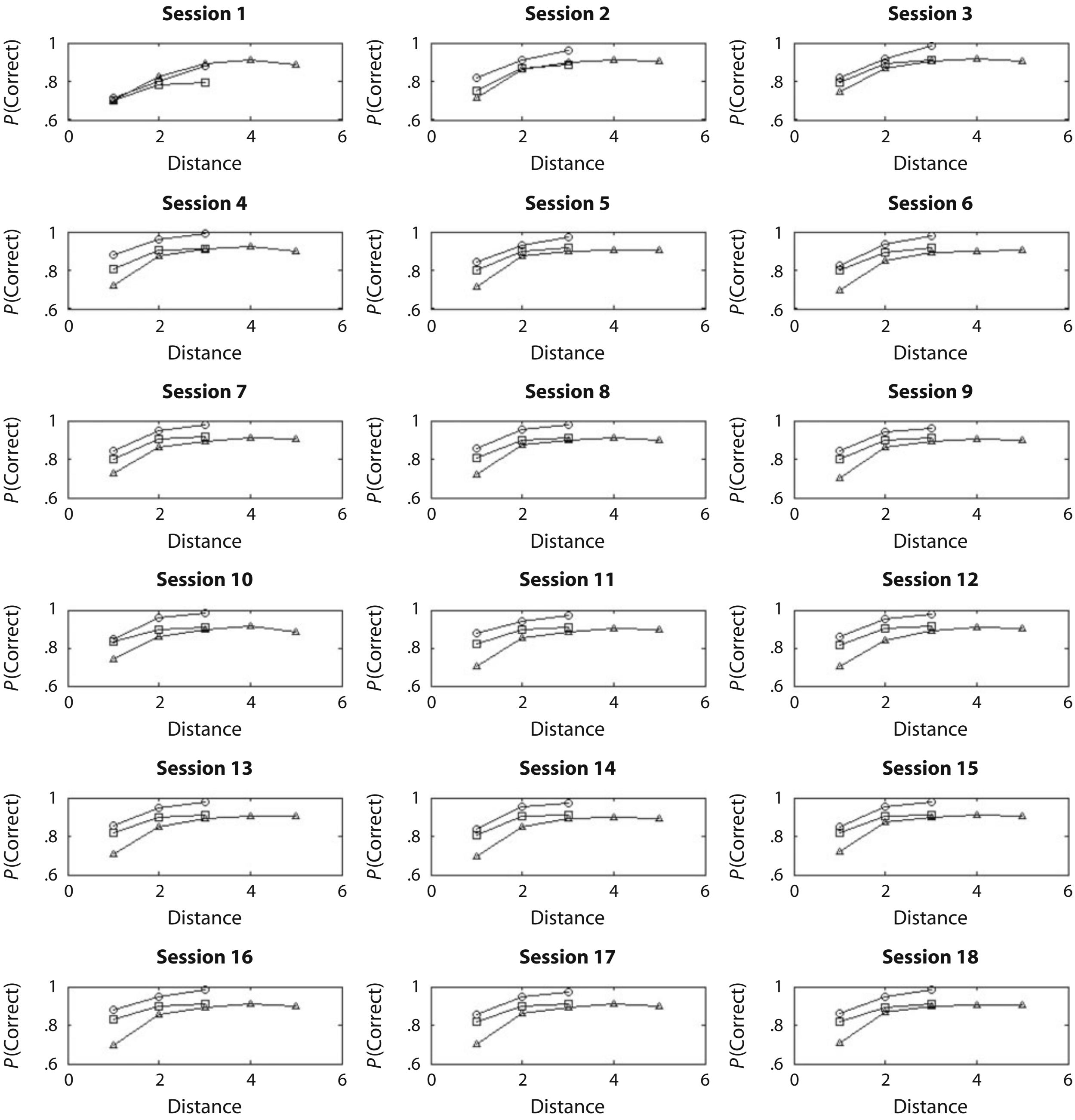


**Figure 4. Distance-to-bound effect on proportion correct. Circles represent the information-integration condition, squares represent the disjunctive-rule condition, and triangles represent the simple-1D rule condition.**

for testing our hypotheses). However, the categorization literature also points to important qualitative differences between initial rule-based and information-integration categorization (e.g., Ashby & Maddox, 2005; Maddox & Ashby, 2004). Experiments 2 and 3 focus on two of these dissociations.

## EXPERIMENT 2

Experiments 2 and 3 test whether two prominent behavioral dissociations found during initial rule-based and information-integration category learning are still present after overtraining. Both dissociations are also popular behavioral criteria for automaticity (Moors & de Houwer, 2006; Shiffrin & Schneider, 1977), so Experiments 2 and 3 also allow us to test whether these criteria are equally valid for tasks that depend on declarative (e.g., rule-based) versus procedural (e.g., information-integration) memory systems.

Several studies reported that switching the location of the response buttons interfered with the expression of information-integration category learning, but not with

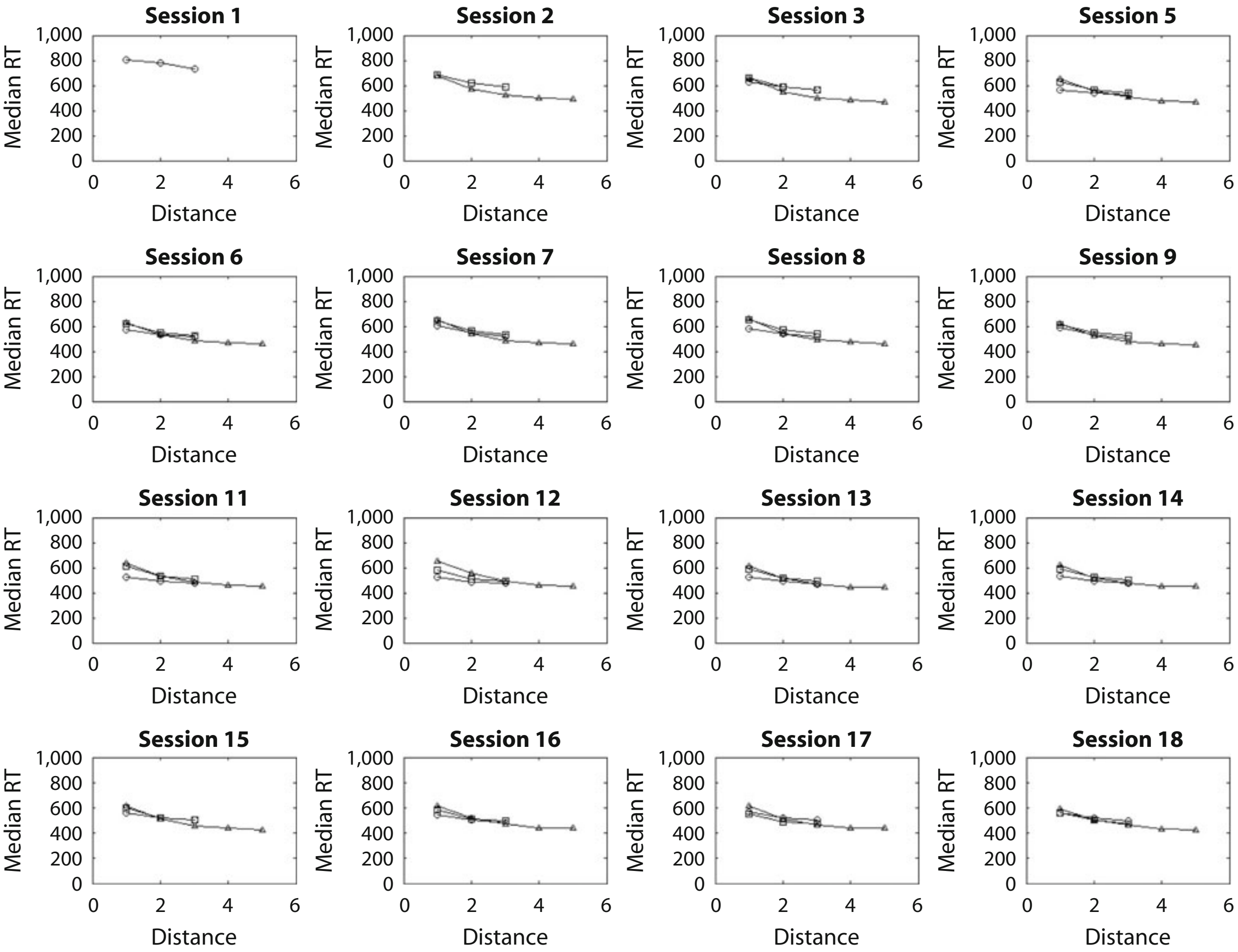


**Figure 5. Distance-to-bound effect on median correct response times (RTs). Circles represent the information-integration condition, squares represent the disjunctive-rule condition, and triangles represent the simple-1D rule condition.**

rule-based category learning (Ashby et al., 2003; Maddox et al., 2004; Maddox et al., 2010; Spiering & Ashby, 2008). Ashby et al. (2003) included only 100 trials after the buttons were switched. This was not long enough to examine recovery. In contrast, Maddox et al. (2010) included 300 trials of transfer after 300 trials of categorization. They reported significant recovery in accuracy over the course of these 300 trials. In addition, in all studies, model analyses showed that many participants in the information-integration condition switched to rule-based strategies after the response buttons had been changed. This change in strategy did not occur when the participants were trained for two sessions prior to the button switch (Ashby et al., 2003).

These results suggest that switching response buttons after overtraining should interfere with information-integration categorization. The absence of a button-switch interference in rule-based categorization, however, is consistent with the hypothesis that the learning in rule-based tasks is of abstract category labels not tied to any response. This hypothesis seems to predict no interference from a button switch, even after overtraining. On the other hand, as mentioned earlier, behavioral inflexibility (as expressed by a button-switch interference) has been proposed as a *feature of automaticity* (Shiffrin & Schneider, 1977; termed *goal independence* by Moors & de Houwer, 2006). This is problematic, since it suggests that information-integration performance is automatic during the first session. Even so, note that the automaticity literature makes the opposite prediction—namely, that interference should develop in rule-based tasks as a result of overtraining. Experiment 2 will test between these two predictions.

## Method

**Participants**. The participants in Experiment 2 were a subset of the participants from Experiment 1. There were 12 participants in the disjunctive-rule condition, 12 participants in the simple-1D rule condition, and 8 participants in the information-integration condition. There were between one and three additional sessions of practice between the end of Experiment 1 and the beginning of Experiment 2. For each participant, one of these sessions was a "scanning" session, whereas the remaining were "regular" sessions (as described in the Method section of Experiment 1). No button-switch fMRI scanning data were collected. The minimum number of practice trials before the button-switch session was 10,920, and the maximum number of practice trials was 12,120 (the mode was 11,520 trials). There was no time lag or break between Experiments 1 and 2.

**Apparatus**. The material was the same as in Experiment 1, except that the response locations were switched (i.e., the A response key occupied the location of the B response key, and vice versa).

**Procedure**. For all the participants in the simple-1D and the disjunctive-rule conditions, the button-switch session was a "regular" session consisting of 12 blocks of 50 trials (for a total of 600 trials). The participants were instructed at the beginning of the session that all stimuli and procedures were identical to those in the preceding 20+ days, except that the location of the two response keys was reversed.

In the information-integration condition, 4 participants had a "regular" button-switch session (as described above). The remaining participants started their button-switch session with four blocks (200 trials) of categorization with the practiced response-key assignment (as in their first 20+ days of practice), followed by eight blocks (400 trials) of categorization with the new (switched) response-key assignment. The procedures were the same as in Experiment 1.

## Results

In all the following analyses, results from the button-switch session were compared with results from the last three sessions of training for each participant. As in Experiment 1, RTs from the "scanning" sessions were excluded from the analyses. Also, trials from the button-switch session that occurred before the button switch (i.e., with the standard button locations) were not included in the following analyses (for 4 participants in the information-integration condition; see the Method section, above).

**Accuracy**. The mean accuracies from the last 3 days of training and from the button-switch session are shown in Figure 6A. As can be seen, accuracy was lower in the button-switch session for all groups. A condition (3, between subjects) $\times$ session (automaticity vs. button switch, within subjects) ANOVA showed a significant effect of session [$F(1,29) = 24.25$, $p < .001$], with accuracy decreasing from 94.0% correct (automaticity) to 88.8% correct (button switch). The main effect of condition [$F(2,29) = 0.93$, n.s.; power = 1] and its interaction with session [$F(2,29) = 2.17$, n.s.; power = 1] failed to reach statistical significance.

To verify that the interference on accuracy was present within each group, separate one-tailed paired $t$ tests were performed to compare training and button-switch performance. For the simple-1D group, performance in the button-switch session was significantly worse [$t(11) = 2.71$, $p < .01$], with performance decreasing from 93.4% to 89.3%. A similar result was found for participants in the disjunctive-rule group [$t(11) = 1.79$, $p < .05$], with performance decreasing from 94.6% (automatic) to 91.0% (button switch). Finally, the performance in the information-integration group diminished from 94.0% to 84.9% (for training and button-switch performances, respectively), which reached statistical significance [$t(7) = 3.69$, $p < .01$]. Hence, it can be concluded that the interference caused by introducing a button switch is genuine and affects all groups.

Another question of interest is whether the observed interference was transient or resilient. Figure 6B shows the mean accuracy for automatic performance (Block 0) and for each 50-trial block of the button-switch session. A condition (3, between subjects) $\times$ block (12, within subjects) ANOVA confirmed that accuracy did not improve during the button-switch session [$F(11,275) = 1.21$, n.s.; power = 0.89]. Furthermore, there was no difference among the conditions [$F(2,25) = 0.96$, n.s.; power = 1] and no interaction between the factors [$F(22,275) = 1.08$, n.s.; power = 0.23]. In Ashby et al. (2003) and Maddox et al. (2010), there was an improvement in accuracy after the first block of button switch (although the interference was not completely canceled).

**RTs**. The group-averaged median correct RTs are shown in Figure 6C. As can be seen, the participants in all three conditions were slower in the button-switch session. A condition (3, between subjects) $\times$ session (automaticity vs. button switch, within subjects) ANOVA was performed on median correct RTs. As in the analysis of response accuracy, the effect of session reached statistical significance [$F(1,29) = 40.94$, $p < .001$]. The correct median RTs increased from 489 msec (automaticity) to 612 msec (button switch). The condition factor [$F(2,29) = 3.21$, n.s.; power = 1] and its interaction with session [$F(2,29) = 3.00$, n.s.; power = 1] failed to reach statistical significance.[11] Recovery of RTs was not explored, as in previous studies.

**Model-based analyses**. In previous studies, when the locations of the response buttons were switched after a single session of information-integration categorization training, many participants switched from an information-integration response strategy to a rule-based strategy (Ashby et al., 2003; Maddox et al., 2010). However, this tendency was reduced in the one experiment that extended the initial (i.e., prebutton switch) training into a second session (Ashby et al., 2003). To examine this issue with the present data, we fit the same models as in Experiment 1 to the data of each participant during the button-switch session.

The percentages of optimal best-fitting models for the button-switch session are shown in Table 2. As can be seen, none of the participants in the information-integration and the disjunctive-rule conditions switched strategy during the button-switch session. However, there were 2 participants in the simple-1D rule condition whose responses were best described by a model that assumed a conjunction rule. A closer look at the parameter estimates showed that the best-fitting conjunction model emulated a 1D model (with a negative intercept on the orientation dimension). Overall, these observations are consistent with the interpretation in Ashby et al. (2003)—that is, that extensive practice stabilizes the choice of categorization strategy and makes strategy shifts less likely.

The estimated noise $SD$s during the training and button-switch sessions are shown in Figure 3B. As can be seen, these values increased during the button-switch session for all three groups. This increase in perceptual/criterial noise reflects the significant button-switch interference found in the accuracy of all three conditions (see Figure 6A). Hence, even though model fitting did not detect a shift of categorization strategy during the button-switch session, the participants seemed generally less proficient at applying their well-practiced strategy.

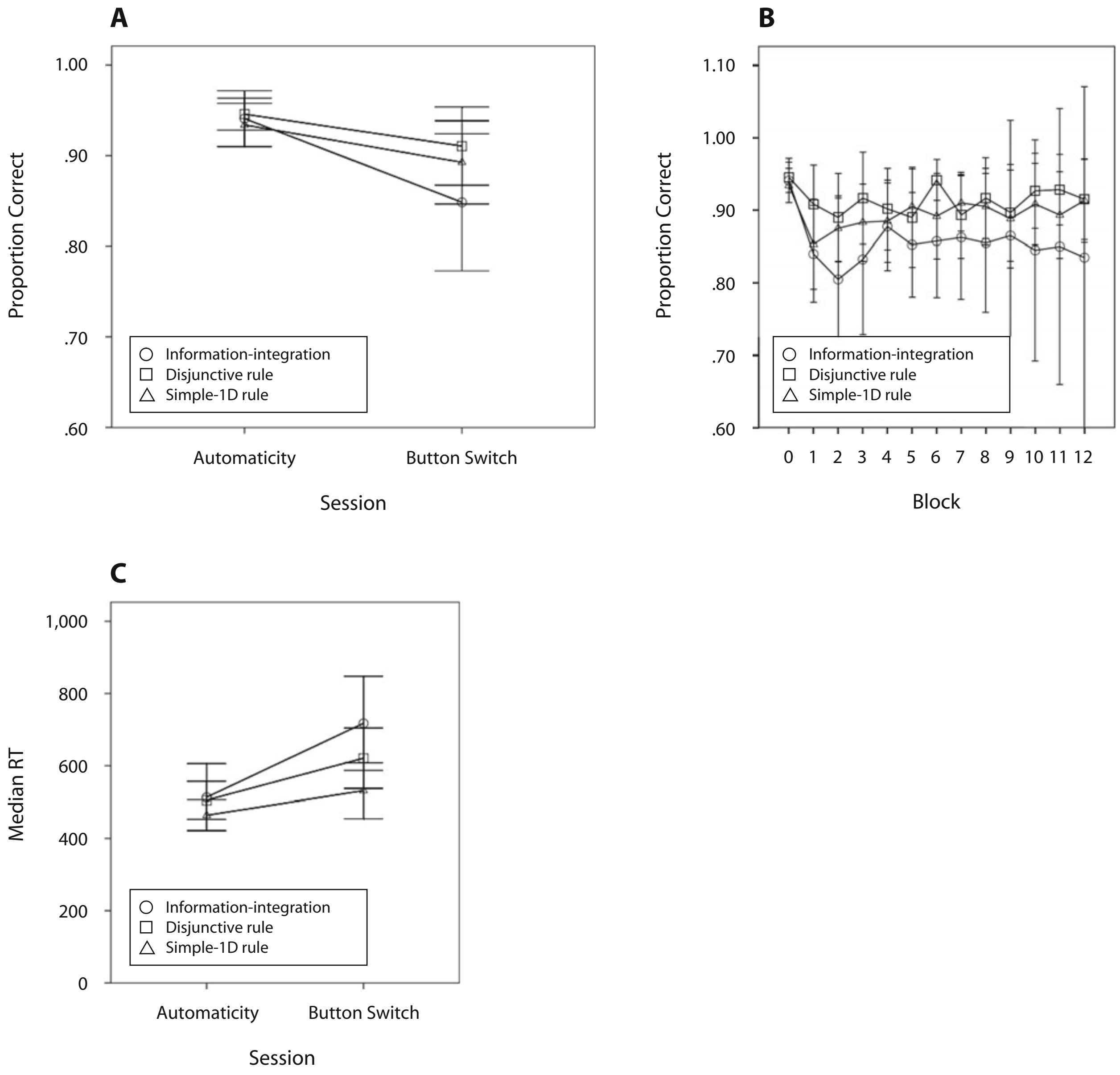


**Figure 6. (A) Proportions correct for automatic categorization performance and button switch. (B) Proportions correct per block during the button switch session. Block 0 represents automatic performance. (C) Mean median correct response times (RTs) for automatic categorization performance and button switch. In each panel, the error bars are 95% confidence intervals.**

## Discussion

This experiment introduced a button-switch session after automaticity had developed in information-integration and rule-based categorization. Past experiments have shown that this manipulation impairs accuracy after initial learning in information-integration categorization, but not in rule-based categorization (Ashby et al., 2003; Maddox et al., 2004; Maddox et al., 2010; Spiering & Ashby, 2008). Furthermore, Maddox et al. (2010) reported that this information-integration impairment was transitory; 300 trials of practice with the new button locations were enough for participants to recover significantly from their initial deficit. In contrast to those earlier results, our results showed that after extensive overtraining, switching the location of the response keys interfered with both information-integration and rule-based performance. Regardless of category structure, switching the buttons caused a significant decrease in accuracy and increase in RT.

These results suggest that, although rule-based category learning initially may be abstract, with enough training it eventually comes to include a response-specific component—just like information-integration categorization. Furthermore, in both rule-based and information-integration categorization, this response-specific component becomes such an essential part of the behavior that switching the location of the response buttons produces an impairment so great that there is no

**Table 2**
**Percentages of Participants in Experiments 2 and 3 Whose Data Were Best Fit by a Model That Assumed a Decision Strategy of the Optimal Type**

| Session | Simple-1D | Disjunctive Rule | Information-Integration |
|---|---|---|---|
| Button switch | 83 | 100 | 100 |
| Dual task | 82 | 100 | 100 |

Note—In the simple-1D condition, the performance of all participants in each session was best described by a model that assumed a vertical decision bound. Entries less than 100 denote cases where a model that assumed an information-integration strategy or a conjunction rule fit best. In all of these cases, however, the best-fitting model emulated predictions from a 1D model (i.e., the best-fitting bound was a vertical line).

significant recovery after 600 trials of practice (as shown by the absence of a block effect in the ANOVA on the button-switch trials).

However, it should be noted that performance did improve for the simple-1D group [$t(11) = 3.67$, $p < .01$]. The mean accuracy was 85.3% in the first button-switch block and increased to 91.3% in the last button-switch block. This recovery was complete, since the accuracy in the last button-switch block was similar to the accuracy in automatic performance [$t(11) = 0.78$, n.s.]. Hence, even though the effect of block was not statistically significant in the ANOVA, the possibility of button-switch interference recovery in the simple-1D group is not ruled out by the data (due to limited statistical power of the interaction term).

## EXPERIMENT 3

The results of Experiment 2 suggest that overtraining eliminates one qualitative difference between initial rule-based and information-integration categorization—namely, that information-integration learning includes a response-specific component, whereas rule-based learning does not. Experiment 3 examines another qualitative difference between initial rule-based and information-integration categorization. As mentioned earlier, several studies have shown that a simultaneous dual task that requires working memory and executive attention interferes with rule-based category learning, but not with information-integration learning (Waldron & Ashby, 2001; Zeithamova & Maddox, 2006). Experiment 3 examines how this dissociation is affected by overtraining.

The absence of dual-task interference is also among the best known behavioral criteria of automaticity (i.e., efficiency; Moors & de Houwer, 2006; Shiffrin & Schneider, 1977). Again, however, the absence of dual-task interference in initial information-integration learning suggests that this is not by itself a sufficient test of automaticity.

### Method

**Participants**. Experiment 3 consisted of a subset of participants from Experiment 1. There were 14 participants in the disjunctive-rule condition, 11 participants in the simple-1D rule condition, and 6 participants in the information-integration condition. All but 1 participant (from the information-integration condition) had one "regular" session of categorization training between the button-switch session and the dual-task session. As in Experiment 2, no data were collected in the fMRI scanner. The minimum number of practice trials before the dual-task session was 11,520, and the maximum number of practice trials was 12,720 trials (the mode was 12,120 trials), excluding the button-switch trials. There was no time lag or break between Experiments 2 and 3.

**Apparatus**. The materials in the categorization task were the same as those in Experiment 1. The same numerical Stroop dual task as in Waldron and Ashby (2001) was used. In this task, two different digits were randomly chosen on every trial (ranging from 2 to 8), and displayed on each side of the crosshair (fixation point) during the categorization experiment (6.5 cm from the fixation point). One of the digits was displayed in a bigger font and occupied 3.3º of visual angle. The size of the other digit was 1.9º of visual angle.

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 are similar to the well-known Stroop effect, because participants must inhibit the automatic response of identifying the value of the digit (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).

**Procedure**. As in the other experiments, the participants were initially shown a crosshair (fixation point) for 1,200 msec. Next, the digits from the numerical Stroop task appeared on both sides of the crosshair for 200 msec. The participant needed to memorize the numerical value and physical size of the digits. The digits disappeared, but the crosshair stayed on the display for another 100 msec.[12] The crosshair disappeared and was replaced by the categorization stimulus. The categorization stimulus stayed on the screen until a categorization response was made, and feedback was given (same as in Experiments 1 and 2).

After the feedback, the screen went blank for 500 msec. Next came a cue that was either the word "Size" or the word "Value." If the cue was "Size," the participant needed to indicate whether the number of larger size was on the right or the left of the crosshair. If the cue was "Value," the participant needed to indicate whether the number of larger value was on the right or the left of the crosshair. The cue remained on the screen until the participant responded. Feedback was given in the same way as in the categorization task, and the procedure started again for another trial.

As in Experiments 1 and 2, half the categorization stimuli were As and the remaining were Bs. In the numerical Stroop task, 510 trials were incongruent (85%), and the remaining 90 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. They were told that their data would not be used if they did not perform well enough in the numerical Stroop task.

### Results

In all analyses, "automatic performance" was defined as in Experiment 2.

**The numerical Stroop task**. The performance in the numerical Stroop task was analyzed to confirm that the participants took the secondary task seriously. The minimum accuracy was 82%, and the maximum accuracy was 99% (mean accuracy = 92.2%). The mean accuracy was 90.7% in the information-integration condition, 93.7% in the disjunctive-rule condition, and 91.2% in the simple-1D condition. Thus, every participant in every condition devoted sufficient attentional resources to the numerical Stroop task to perform at a high level. The results in the numerical Stroop task are not discussed further.

**Accuracy**. Mean categorization accuracy is shown in Figure 7A. Note that the addition of the dual task did not have much effect on accuracy in any condition. A condition (3, between subjects) × session (automaticity vs. dual task, within subjects) ANOVA was performed. The session factor [$F(1,28) = 4.02$, n.s.; power = 1] and its interaction with condition [$F(2,28) = 2.98$, n.s.; power = 1] failed to reach statistical significance. The mean accuracy during the dual-task session was 93.1% (i.e., a drop of less than 1%). The condition factor [$F(2,28) = 0.20$, n.s.; power = 1] also failed to reach statistical significance.

As in Experiment 2, separate paired $t$ tests were performed to compare training and dual-task performance. Here, two-tailed tests were used, because not all the differences had the same signs. None of the accuracy changes was statistically reliable. Accuracy in the disjunctive-rule condition decreased by 1.3% [$t(13) = 1.65$, n.s.], accuracy in the information-integration condition decreased by 2.9% [$t(5) = 1.64$, n.s.], and accuracy in the simple-1D rule condition increased by 0.7% [$t(10) = -1.02$, n.s.]. These more fine-grained analyses suggest no clear interference effects of the dual-task manipulation. Hence, it can be concluded that the dual task did not interfere with categorization performance in any condition. These results differ from those in Waldron and Ashby (2001), who found a highly significant interference in a rule-based condition.

**RTs**. Figure 7B shows the mean median correct RTs for automatic performance and during the dual-task session. A condition (3, between subjects) × session (automaticity vs. dual task, within subjects) ANOVA was performed on group-averaged correct median RTs. The effect of session reached statistical significance [$F(1,28) = 37.84$, $p < .001$]. The correct median RTs increased from 494 msec (automaticity) to 684 msec (dual task). Neither the condition factor [$F(2,28) = 0.14$, n.s.; power = 1] nor its interaction with session [$F(2,28) = 1.67$, n.s.; power = 1] reached statistical significance. RTs were not analyzed in any of the previous dual-task category-learning studies.

**Model-based analyses**. The percentages of participants whose responses were best described by a model that assumed a decision strategy of the optimal type are shown in Table 2. As can be seen, there was no propensity for participants to change their response strategy during the dual-task session. The responses of all participants in the information-integration and disjunctive-rule conditions were best fit by a model that assumed a strategy of the optimal type. For the simple-1D rule condition, the responses of 2 participants were better fit by a nonoptimal decision-bound model—one was best fit by a model that assumed a conjunction rule, and another by a model that assumed information integration. Again, a closer inspection showed that in each one of these cases, these more complex models emulated a 1D model.

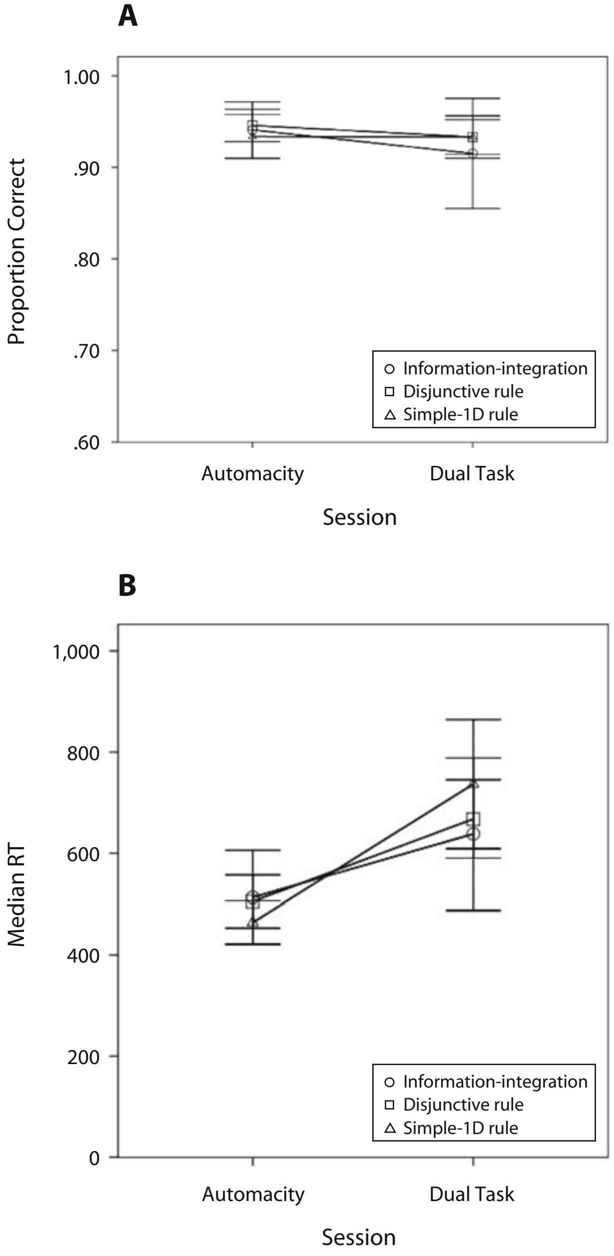


**Figure 7. (A) Proportions of correct categorizations for automatic performance and during the dual-task session. (B) Mean median correct response times (RTs) for automatic categorization performance and during the dual-task session. The error bars are 95% confidence intervals.**

The estimated noise *SD*s from the best-fitting models are shown in Figure 3C. Unlike the button-switch, the numerical Stroop task did not seem to affect the amount of perceptual/criterial noise. This further supports the hypothesis that the dual task did not interfere with categorization performance in any condition.

## Discussion

This experiment introduced a dual-task condition after participants had already completed thousands of trials of categorization training. The numerical Stroop task did not produce interference in any condition. This conclusion was supported by model-based analyses, which showed no change in response strategy and no additional perceptual/criterial noise. This result contrasts sharply with results obtained with untrained participants in rule-based tasks (Waldron & Ashby, 2001; Zeithamova & Maddox, 2006). As in Experiment 2, the behavioral dissociation of dual-task interference found between information-integration and rule-based categorization disappeared after automaticity had developed.

## GENERAL DISCUSSION

This article presents results from three experiments that explored the effects of overtraining on categorization performance. A major goal was to determine whether the qualitative differences known to occur during the early learning and performance of rule-based and information-integration category structures persist as these categorization judgments become automatic. A secondary goal was to test whether several popular automaticity criteria are equally valid for behaviors that were initially learned via declarative versus procedural memory systems. Experiment 1 showed that, consistent with previous results (e.g., Ashby & Maddox, 2005; Maddox & Ashby, 2004), participants displayed behavioral differences during the initial training sessions, depending on whether they learned rule-based or information-integration categories. Even so, after the third session, almost no significant differences could be discerned among any of these groups. More specifically, all groups showed similar accuracy levels and similar RTs (see Figure 2), and the performance of all participants in every condition was best described by a model emulating an optimal decision strategy (see Table 1). The only exception was found in distance-to-bound analyses, which suggested that participants in the simple-1D rule condition improved less than did participants in the other two conditions on stimuli close to the category boundary. This result suggests that participants learn more difficult stimuli only if absolutely necessary (Simon, 1972). Together, all these results show that quantitative behavioral differences between information-integration and rule-based categorization can be eliminated with overtraining.

Experiments 2 and 3 focused on previously reported behavioral dissociations between rule-based and information-integration categorization performance. Experiment 2 showed that switching the location of the response keys after more than 10,000 trials of practice produced interference in all conditions (on both accuracy and RT), and that there was almost no recovery from this interference over the course of 600 trials (see Figure 6). Model-based analyses suggested that this interference was not the result of a strategy shift but instead was due to an increase in the estimated perceptual/criterial noise (see Table 2 and Figure 3B). Interference stemming from switching the location of the response keys suggests a lack of control (or inflexibility) of response production, which has been used as a criterion for assessing automaticity in past research (Logan, 1988; Moors & de Houwer, 2006; Shiffrin & Schneider, 1977). Previous research showed that switching the location of the response keys early in training has no effect on rule-based categorization performance (Ashby et al., 2003; Maddox et al., 2004; Maddox et al., 2010), and that although there is an initial interference in information-integration tasks, participants show significant recovery after a few hundred trials of practice with the new response locations (Maddox et al., 2010). Experiment 2 showed that both of these effects disappear with overtraining; that is, after extensive practice, switching the response keys interferes with rule-based and information-integration performance. In both cases, there is almost no recovery from this switch, even after 600 trials of practice.

Experiment 3 included a dual-task manipulation. The results showed that the numerical Stroop task did not interfere with categorization accuracy in any of the conditions (see Figure 7A). This contrasts sharply with results from early performance, where a number of studies have shown that a dual task requiring executive attention and working memory strongly interferes with rule-based but not with information-integration categorization performance (Waldron & Ashby, 2001; Zeithamova & Maddox, 2006). The absence of dual-task interference has been used as a criterion to assess the presence of automaticity in previous research (efficiency; Moors & de Houwer, 2006; Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977).

### Theoretical Implications

The data presented in this article have important implications for the development of theories of automaticity in categorization. The main finding is that the initial quantitative and qualitative differences between rule-based and information-integration categorization that have been observed in many studies (for a review, see Ashby & Maddox, 2005) are reduced after overtraining. This is consistent with a theory assuming separate rule-based and information-integration learning systems but a common "automatic" processing mode.

Hélie and Ashby (2009) recently proposed a model of rule maintenance and application. In the new model, rule application initially relies on rule maintenance in working memory, which is eventually replaced by associative processing (i.e., a direct stimulus→response association). The Hélie and Ashby model readily predicts the absence of an initial button-switch interference in rule-based categorization (because rule application is controlled by working memory) and the presence of a button-switch interference after the development of automaticity (because of the stimulus→response associative processing). Also, because the numerical Stroop task occupies working memory, the model predicts the initial presence of dual-task interference and its disappearance after the development of automaticity. In the Hélie and Ashby model, working memory is required only for early rule application.

Ashby, Ennis, and Spiering (2007) proposed a computational model of automaticity for information-integration categorization tasks. The SPEED model suggests that information-integration categorization is a procedural process and that the stimulus→response association becomes more direct (i.e., less flexible) after extensive training. The reliance on procedural and associative processing readily predicts the presence of button-switch interference and the absence of dual-task interference (because working memory is not used). Together, the Hélie and Ashby model and the SPEED model might allow for a natural explanation of automaticity in categorization.

The convergence of performance in the three conditions can also be explained by Logan's (1988, 1992) instance theory of automaticity. According to Logan, algorithms that can achieve the task compete (race) with a single-step memory retrieval process to provide a response on each trial. Each category structure might be processed by a different algorithm, which would explain the performance differences early in training (when the response is algorithm driven). However, the responses become memory-driven after extensive practice, and memory retrieval would be similar regardless of category structure (because the stimuli were the same). Note that responding gradually becomes more stimulus specific with both Logan's memory retrieval explanation and Ashby and colleagues' associative processing explanation; hence, only learned material is automatized. This suggests that stimuli that have not been learned (e.g., the stimuli close to the boundary in the simple-1D rule condition) are not automatized. This could partially explain the recovery found in this condition in Experiment 2.

The results in this article also provide an initial opportunity to expand the use of some classical behavioral criteria of automaticity to tasks that depend on multiple memory systems. Experiments 2 and 3 focused on two features of automaticity—namely, behavioral inflexibility and efficiency (Moors & de Houwer, 2006; Shiffrin & Schneider, 1977). Our results suggest that both features hold, regardless of whether the behavior is initially learned by declarative or procedural memory systems. Even so, earlier results suggest that neither feature by itself is sufficient to distinguish automatic from controlled processes. For instance, information-integration data from Ashby et al. (2003) and Waldron and Ashby (2001) displayed behavioral inflexibility (e.g., button-switch interference) and efficiency (e.g., the absence of dual-task interference) after minimal training. These features were observed even though the participants' performance had clearly not reached asymptote and was not the result of single-step memory retrieval.

The results of Experiment 2 suggest that behavioral inflexibility could be improved as a criterion for establishing automaticity if it were augmented with resistance to recovery (e.g., so that it becomes *enduring* behavioral inflexibility). For instance, Maddox et al. (2010) trained their participants for 300 trials before the button switch and found significant recovery after 300 button-switch trials. In Experiment 2, participants were trained for more than 10,000 trials before the button switch, and most participants did not show any sign of recovery after 600 button-switch trials. Thus, with information-integration tasks, the presence of a button-switch interference does not change with overtraining, but the recovery from this interference slows considerably as participants become more and more practiced. Shiffrin and Schneider (1977) reported a similar resistance to recovery in a visual search task. After 2,100 trials of practice, a change of target mapping in the consistent mapping condition produced interference that lasted for roughly 2,500 trials. Together, all these results suggest that the duration of interference due to a button (or stimulus mapping) switch increases with the duration of preswitch training. Hence, the duration of the interference or the rate of recovery can be a good indicator of the extent of behavioral automaticity. Further research is needed to evaluate the generality of this conjecture.

Enduring behavioral inflexibility provides some insight into the nature of the interference in Ashby and colleagues' models (Ashby et al., 2007; Hélie & Ashby, 2009). Because the interference was only partial (i.e., the performance was better than that of untrained participants), the data suggest that the participants did not revert to their initial processing strategy; it therefore appears that interference results from a partially unsuccessful attempt at controlling the automatic processing. This can be done by modifying the stimulus component of the associative processing responsible for automatic behavior (e.g., making it more specific to take context into consideration). Tentatively, stronger associations (resulting from more training trials) might be harder to modify than weaker associations. More work is needed to fully understand the nature of interference following the development of automaticity.

Unfortunately, there does not appear to be a similarly straightforward way to augment the efficiency criterion. Experiment 3 found that a dual task did not interfere with any of our tasks. One problem is that several studies have reported that this same dual task interferes only minimally with initial information-integration learning (Waldron & Ashby, 2001; Zeithamova & Maddox, 2006). Thus, the absence of a dual-task interference could mean either that the behavior has become automatic or that the behavior is mediated by procedural memory. This conclusion assumes that the dual task requires working memory and executive attention. It is possible that a dual task that instead depends primarily on procedural memory might interfere with initial information-integration learning. However, this is a difficult hypothesis to test, because most classic procedural memory tasks include a motor component that would make it difficult for participants to respond to the categorization stimulus (e.g., mirror tracing). Clearly, more work is needed on this problem.

**Future Work**

This article presents a detailed study of the development of automaticity in categorization. The presentation of the results was purposefully made atheoretical, since the data themselves present a challenge for proponents of existing and future theories of automaticity. Future work should be devoted to a detailed exploration of how existing theories of automaticity can account for the data, as well as the (possible) development of new detailed theories that can simultaneously account for the learning differences and similarities in automatic performance reported herein. Also, it is likely that the results presented here are dependent on the memory systems involved in the categorization task; other tasks involving the same memory systems should, therefore, show similar effects. Finally, more work is needed to link these findings with computational modeling as well as with neurological findings.

## AUTHOR NOTE

This research was supported in part by the U.S. Army Research Office through the Institute for Collaborative Biotechnologies under Grant W911NF-07-1-0072 and by Grant R01 MH3760-2 from the National Institutes of Health. The authors thank Kyle Liban, Melissa Morrison, Alecia M. Moser, Jessica L. Roeder, and Maria Schellenberger for their help with running the participants. The authors also thank three anonymous reviewers for their useful comments on an earlier draft. Correspondence and requests for reprints should be addressed to S. Hélie, Department of Psychology, University of California, Santa Barbara, CA 93106-9660 (e-mail: helie@psych.ucsb.edu).

## REFERENCES

ASHBY, F. G. (1992). Multivariate probability distributions. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 1-34). Hillsdale, NJ: Erlbaum.

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**, 442-481.

ASHBY, F. G., ELL, S. W., & WALDRON, E. M. (2003). Procedural learning in perceptual categorization. *Memory & Cognition*, **31**, 1114-1125.

ASHBY, F. G., ENNIS, J. M., & SPIERING, B. J. (2007). A neurobiological theory of automaticity in perceptual categorization. *Psychological Review*, **114**, 632-656.

ASHBY, F. G., & GOTT, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 33-53.

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

ASHBY, F. G., & O'BRIEN, J. B. (2005). Category learning and multiple memory systems. *Trends in Cognitive Sciences*, **9**, 83-89.

ASHBY, F. G., WALDRON, E. M., LEE, W. W., & BERKMAN, A. (2001). Suboptimality in human categorization and identification. *Journal of Experimental Psychology: General*, **130**, 77-96.

BADDELEY, A. [D.] (1986). *Working memory*. New York: Oxford University Press.

BARCIKOWSKI, R. S., & ROBEY, R. R. (1985, April). *Sample size selection in single group repeated measures analysis*. Paper presented at the 69th Annual Meeting of the American Educational Research Association, Chicago.

BRAINARD, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, **10**, 433-436.

CRABB, B. T., & DARK, V. J. (2003). Perceptual implicit memory relies on intentional, load-sensitive processing at encoding. *Memory & Cognition*, **31**, 997-1008.

DICKINSON, A. (1985). Actions and habits: The development of behavioural autonomy. *Philosophical Transactions of the Royal Society B*, **308**, 67-78.

HÉLIE, S. (2006). An introduction to model selection: Tools and algorithms. *Tutorials in Quantitative Methods for Psychology*, **2**, 1-10.

HÉLIE, S. (2007). Understanding statistical power using noncentral probability distributions: Chi-squared, G-squared, and ANOVA. *Tutorials in Quantitative Methods for Psychology*, **3**, 63-69.

HÉLIE, S., & ASHBY, F. G. (2009). A neurocomputational model of automaticity and maintenance of abstract rules. *Proceedings of the International Joint Conference on Neural Networks* (pp. 1192-1198). Atlanta: IEEE Press.

JOHNSON, K. E., & MERVIS, C. B. (1997). Effects of varying levels of expertise on the basic level of categorization. *Journal of Experimental Psychology: General*, **126**, 248-277.

LOGAN, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, **95**, 492-527.

LOGAN, G. D. (1992). Shapes of reaction-time distributions and shapes of learning curves: A test of the instance theory of automaticity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 883-914.

MADDOX, W. T., & ASHBY, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception & Psychophysics*, **53**, 49-70.

MADDOX, W. T., & ASHBY, F. G. (2004). Dissociating explicit and procedural-learning based systems of perceptual category learning. *Behavioural Processes*, **66**, 309-332.

MADDOX, W. T., ASHBY, F. G., & GOTTLOB, L. R. (1998). Response time distributions in multidimensional perceptual categorization. *Perception & Psychophysics*, **60**, 620-637.

MADDOX, W. T., BOHIL, C. J., & ING, A. D. (2004). Evidence for a procedural-learning-based system in perceptual category learning. *Psychonomic Bulletin & Review*, **11**, 945-952.

MADDOX, W. T., GLASS, B. D., O'BRIEN, J. B., FILOTEO, J. V., & ASHBY, F. G. (2010). Category label and response location shifts in category learning. *Psychological Research*, **74**, 219-236.

MEDIN, D. L., LYNCH, E. B., COLEY, J. D., & ATRAN, S. (1997). Categorization and reasoning among tree experts: Do all roads lead to Rome? *Cognitive Psychology*, **32**, 49-96.

MOORS, A., & DE HOUWER, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin*, **132**, 297-326.

NISSEN, M. J., & BULLEMER, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, **19**, 1-32.

NOSOFSKY, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, **115**, 39-57.

NOSOFSKY, R. M., & PALMERI, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, **104**, 266-300.

PALMERI, T. J., WONG, A. C.-N., & GAUTHIER, I. (2004). Computational approaches to the development of perceptual expertise. *Trends in Cognitive Sciences*, **8**, 378-386.

ROLLS, E. T. (2000). Memory systems in the brain. *Annual Review of Psychology*, **51**, 599-630.

SCHNEIDER, W., & CHEIN, J. M. (2003). Controlled and automatic processing: Behavior, theory, and biological mechanism. *Cognitive Science*, **27**, 525-559.

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

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

SHIFFRIN, R. M., & SCHNEIDER, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, **84**, 127-190.

SIMON, H. A. (1972). Theories of bounded rationality. In C. B. McGuire & R. Radner (Eds.), *Decision and organization: A volume in honor of Jacob Marschak* (pp. 161-176). Amsterdam: North-Holland.

SPIERING, B. J., & ASHBY, F. G. (2008). Response processes in information-integration category learning. *Neurobiology of Learning & Memory*, **90**, 330-338.

SQUIRE, L. R., & SCHACTER, D. L. (EDS.) (2002). *Neuropsychology of memory* (3rd ed.). New York: Guilford.

TULVING, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology*, **53**, 1-25.

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**, 168-176.

WILLINGHAM, D. B. (1998). A neuropsychological theory of motor skill learning. *Psychological Review*, **105**, 558-584.

WILLINGHAM, D. B., WELLS, L. A., FARRELL, J. M., & STEMWEDEL, M. E. (2000). Implicit motor sequence learning is represented in response locations. *Memory & Cognition*, **28**, 366-375.

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

## NOTES

1. As another example, Nosofsky (1986) reported the results of an experiment in which each participant completed many thousands of trials. Even so, there were only 2 participants, each of whom completed five separate tasks (one identification, and four different categorization). In each catego-

rization task, Nosofsky's participants had about 1/3 as much training as in our Experiment 1, and Nosofsky focused only on asymptotic accuracy.

2. In the animal learning literature, automatic behaviors are often called *habits*.

3. In Ashby et al. (2003) and Waldron and Ashby (2001), each stimulus was seen only once in each session. Hence, it is very unlikely that the stimuli were memorized and that performance resulted from single-step memory retrieval after one session of training.

4. For sessions in the fMRI scanner, the feedback was visual (because of the noise in the magnet). A green check mark was used on correct trials, a red X was used on incorrect trials, and a black dot was used on trials that took too much time. (It was not possible to use an incorrect response key in the scanner.) Visual feedback in the scanner was displayed for 2,000 msec.

5. For an *F* test, the effect size is

$$\varpi^2 = \frac{\sigma_m^2}{\sigma^2},$$

where $\sigma_m^2$ is the variance of the group means and $\sigma^2$ is the error variance (Hélie, 2007). The precision of the measures was taken into consideration for power calculation. Hence,

$$\hat{\sigma}_m^2 = \frac{\sum_{i=1}^{k}\left(\bar{x}_i - \bar{x}\right)^2}{k-1}$$

and

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{k}\left(n_i - 1\right)s_i^2 / m_i}{\sum_{i=1}^{k} n_i - k},$$

where $\bar{x}_i$ is the mean of group $i$, $\bar{x}$ is the grand mean, $n_i$ is the number of participants in group $i$, $k$ is the number of groups, $s_i^2$ is the variance of group $i$, and $m_i$ is the number of data used to calculate $s_i^2$.

6. All RT analyses described in this article were also performed on mean correct RTs. The results of the ANOVAs on mean correct RTs were the same as the results of the ANOVAs on median correct RTs, except when mentioned otherwise. Only the analyses on median correct RTs are reported throughout.

7. Sessions 4 and 10 were also excluded from the ANOVA, because they were "scanning" sessions in all three conditions (and are not shown in Figure 2B).

8. It should be noted that decision-bound models are used here for descriptive purposes only; no claim is made as to whether the participants are using decision bounds or other categorization strategies.

9. Which, coincidentally, roughly corresponds to the Euclidean distance between the nearest exemplars in contrasting categories in Figure 1 (i.e., 6 units in the 100 × 100 space).

10. Distances were binned by calculating the Euclidean distance between the optimal boundary and each stimulus in the 100 × 100 space (from Figure 1). The resulting distances were divided by 10 and rounded up.

11. This is the only analysis where the mean correct RTs gave a slightly different result than the median correct RTs. For mean correct RTs, the interaction was statistically significant [$F(2,29) = 4.57$, $p < .05$]. The significant interaction suggests that the slowest automatic performance (i.e., information-integration) suffered the most interference (239 msec), and that the fastest automatic performance (i.e., simple-1D) suffered the least interference (64 msec; with the disjunctive-rule condition lying in the middle with 154 msec).

12. Overall, the fixation point/crosshair stayed on the display for 1,500 msec, as in Experiments 1 and 2.

## APPENDIX

Here, we briefly describe the decision-bound models. For more details, see Maddox and Ashby (1993).

### Rule-Based Models

Three models assume that the observers use an explicit rule-based strategy.

**The one-dimensional model (1D)**. This model assumes that the observer sets a criterion on a single perceptual dimension and then makes an explicit decision about the level of the stimulus on that dimension (Ashby & Gott, 1988). It has two free parameters: a decision criterion on the relevant perceptual dimension and the variance of internal (perceptual and criterial) noise. This strategy is optimal with the simple-1D category structure (top-right panel in Figure 1).

**The conjunction model**. Another possible rule-based strategy is that the observer uses a conjunction rule in which s/he makes separate decisions about the levels on the two dimensions and then selects a response on the basis of the outcomes of these two decisions. Conjunction models have three parameters (i.e., two decisions criteria on separate dimensions and internal noise).

**The interval-based one-dimensional classifier (IB1D)**. A rule-based strategy can also be used to create a disconnected response region in the stimulus space. The interval-based one-dimensional classifier includes two decision criteria on the same dimension and the variance of internal noise. Hence, the IB1D has three parameters. This strategy is optimal with the disjunctive-rule category structure (top-left panel in Figure 1).

### Information-Integration Models

**The general linear classifier (GLC)**. This model assumes that the decision bound between each pair of categories is linear. This produces an information-integration decision strategy, because it requires linear integration of frequency and orientation. The GLC has three parameters (slope and intercept of the linear bound and the variance of the internal noise). This strategy is optimal with the information-integration category structure used in this article (bottom-left panel in Figure 1).

**The general quadratic classifier**. A natural extension of the GLC is to assume that the observer uses a quadratic, rather than a linear, decision bound. This model also produces an information-integration strategy, but the integration of frequency and orientation is nonlinear. The general quadratic classifier has six free parameters (five describing the form of the decision bound and the variance of the internal noise).

### Guessing Models

Guessing models simply assume that the participant blindly responds A on proportion $p$ of the trials without using the stimulus information (and responds B on proportion $1 - p$ of the trials). Here, we fit two versions of the guessing model. The *pure guessing* model assumes that $p = .5$ and has no free parameters. In contrast, the

**APPENDIX (Continued)**

*biased guessing* model allows $p$ to vary between zero and one and has one free parameter. As implied by its name, the biased guessing model can be used to account for any prior response bias that a participant can have before the beginning of the experiment.

**Model Fitting**

Each of these models was fit separately to the data for every observer in each session. The model parameters were estimated using maximum likelihood (Ashby, 1992; Hélie, 2006), and the goodness-of-fit statistic was

$$\mathrm{BIC} = r \times \ln(N) - 2 \times \ln(L),$$

where $N$ is the sample size, $r$ is the number of free parameters, and $L$ is the likelihood of the model given the data (Hélie, 2006; Schwarz, 1978). The BIC statistic penalizes a model for bad fit and for extra free parameters. To find the best model among a set of competitors, one simply computes a BIC value for each model and then chooses the model with the smallest BIC.

(Manuscript received September 16, 2009;
revision accepted for publication January 15, 2010.)