Stimulus range and discontinuity effects on information-integration category learning and generalization

W. Todd Maddox · J. Vincent Filoteo

Abstract Two studies are reported that support the hypothesis that categories that require a multiple-unit, as opposed to a single-unit, representation lead to worse initial acquisition but better generalization. On the basis of the constraints imposed by the procedural-based learning system thought to mediate information-integration categorization, we argue that the need to train multiple units during initial category acquisition slows the procedural-based category learning process and adversely affects learning performance. However, we speculate that better generalization occurs because of the increased likelihood that a novel stimulus will activate at least one of the multiple units needed to represent the category. Relations to other findings in the literature and the implications of this work for training and clinical assessment are discussed.

Keywords Category learning · Information-integration · Procedural-based learning · Category variance · Category coherence · Retention · Generalization · Striatal pattern classifier

W. T. Maddox (✉)
Department of Psychology and Institute for Neuroscience, Center for Perceptual Systems, University of Texas at Austin,
1 University Station A8000,
Austin, TX 78712, USA
e-mail: maddox@psy.utexas.edu

J. V. Filoteo (✉)
Psychology Service 116B, VA San Diego Healthcare System and University of California, San Diego,
3350 La Jolla Village Drive,
San Diego, CA 92161, USA
e-mail: vfiloteo@ucsd.edu

Understanding the experimental factors and psychological processes that facilitate acquisition and generalization has been the focus of psychological research since its inception. This focus has implications for education and training at all levels, from formal classroom training (e.g., reading, writing, and arithmetic), to skill learning (e.g., driving, airline screening, medical diagnosis, or radiology), to rehabilitation and interventions (e.g., improving memory and attention or reducing drug relapse).

One cognitive skill for which acquisition and generalization processes are critical is information-integration perceptual classification (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Ashby & Maddox, in press; Maddox & Ashby, 2004). These are classification problems for which there is no verbal analogue to the optimal rule and for which learning is gradual and incremental. Some of the most important occupations in our society—such as airport screening, radiology, or medical diagnosis, to name a few—involve classification, and we spend an enormous amount of time and money devoted to training individuals to perform these jobs. In a typical classification learning task, participants are presented with a stimulus and are asked to classify it into a single category. Once a response is made, participants are provided with corrective feedback. A common assumption in classification research (as well as other research areas) is that training regimens that lead to good initial acquisition when feedback is presented will also lead to good generalization to novel items within and outside the range of training items, even when corrective feedback is removed (for a review, see Schmidt & Bjork, 1992; Smith, Redford, Washburn, & Taglialatela, 2005).

Early learning theorists, however, recognized that this assumption is often false, and they noted that experimental

factors that improve initial acquisition can either lead to good *or* poor generalization once feedback is removed (Estes, 1955; Hull, 1943; Skinner, 1938; Tolman, 1932). Perhaps counterintuitively, there is evidence in some domains that experimental factors that lead to *worse* initial acquisition actually lead to *better* generalization (Schmidt & Bjork, 1992). This pattern has been observed in the motor and verbal learning domains (Balota, Duchek, & Logan, 2007; Bjork, 1994; Bjork & Linn, 2006; Karpicke & Roediger, 2007, 2008; Landauer & Bjork, 1978; Roediger & Karpicke, 2006). For example, in motor learning, spaced practice of motor movements leads to worse initial acquisition but better generalization, whereas massed practice of the same movements leads to better initial acquisition but worse generalization (see, e.g., Shea & Morgan, 1979).

Given the important role that classification plays in many real-world skills, and given the fact that good acquisition training does not necessarily imply good generalization, it is critical to evaluate the efficacy of any training procedure by incorporating a transfer phase that includes novel items from within and outside the range of training items and during which feedback is not provided (Schmidt & Bjork, 1992). Although acquisition training usually involves presentation of a fixed set of items, the true test of generalization lies in one's ability to classify not only items that are similar to the training items (i.e., those from within the range of trained items), but also items that are dissimilar from them (i.e., items from outside that range) (Erickson & Kruschke, 1998, 2002; Smith et al., 2005). We take this approach in the present study.

In addition, we use a model-based approach to determine how different training regimens affect the types of processes people use to perform that task and how these affect the nature of generalization. To anticipate, the model-based analysis turns out to be critical in the interpretation of the data.

Overview of the present study

The overriding aim of the present work is to examine the effects of category range and category discontinuity on both acquisition of and generalization from a broadly sampled set of stimuli. *Category range* is defined as the breadth of stimulus values along the stimulus dimensions (often referred to as *category variance* in the literature; Cohen, Nosofsky, & Zaki, 2001; Hahn, Bailey, & Elvin, 2005; Rips & Collins, 1993). *Category discontinuity* results when each category is composed of distinct subclusters of stimuli that are separated by unsampled portions of the stimulus space.

Category range and discontinuity effects have been examined in the literature, but often the two factors are confounded, making it difficult to determine their independent impacts. For example, Maddox, Filoteo, and Lauritzen

(2007; for related work, see Kornell & Bjork, 2008; Maddox, Filoteo, Lauritzen, Connally, & Hejl, 2005) examined the effects of continuous versus discontinuous category training on information-integration acquisition and generalization. Scatterplots of the exemplars from the (small-range) continuous and discontinuous training conditions for the information-integration categories are displayed in Fig. 1 (along with the transfer items). Maddox et al. (2007) found that, for information-integration categories, acquisition was adversely affected by discontinuous category training, but that no-feedback transfer performance was better in the discontinuous training condition than in the continuous training condition. Unfortunately, discontinuity was confounded with range, making it impossible to determine whether the increased range or the category discontinuity led to the observed acquisition and generalization performance difference.

Real-world categories differ in their ranges and levels of continuity. For example, members of the category "hand guns" (which an airline screener must learn) are highly similar, and thus have a relatively small category range and are fairly cohesive (i.e., highly continuous). On the other hand, the category "weapon" is highly variable and contains items such as knives, bombs, guns, and so on, which are highly discontinuous. The differences in continuity and range between these two categories could have implications for acquisition and generalization under various training conditions. Thus, it is important to disentangle category range and discontinuity to understand the effects these factors would independently have on acquisition and generalization in the real world.

The present study provides an unconfounded test of the effects of category range and category discontinuity on information-integration acquisition and generalization, across two experiments. Each experiment includes a small-range continuous, a large-range continuous, and a discontinuous acquisition training condition. Comparison of performance across the small- and large-range continuous conditions provides a test of the effects of increased category range on acquisition and generalization while holding discontinuity constant, whereas comparison of performance across the large-range continuous and discontinuous conditions provides a test of the effects of category discontinuity on acquisition and generalization while holding category range constant.

A number of factors that are often left uncontrolled are held constant across our experimental conditions. These include the number of acquisition training trials, the nature of the optimal decision bound, and optimal accuracy. A no-feedback transfer phase is also included that tests performance for items from within the trained portion of the stimulus space and generalization to items outside the trained portion of the stimulus space.
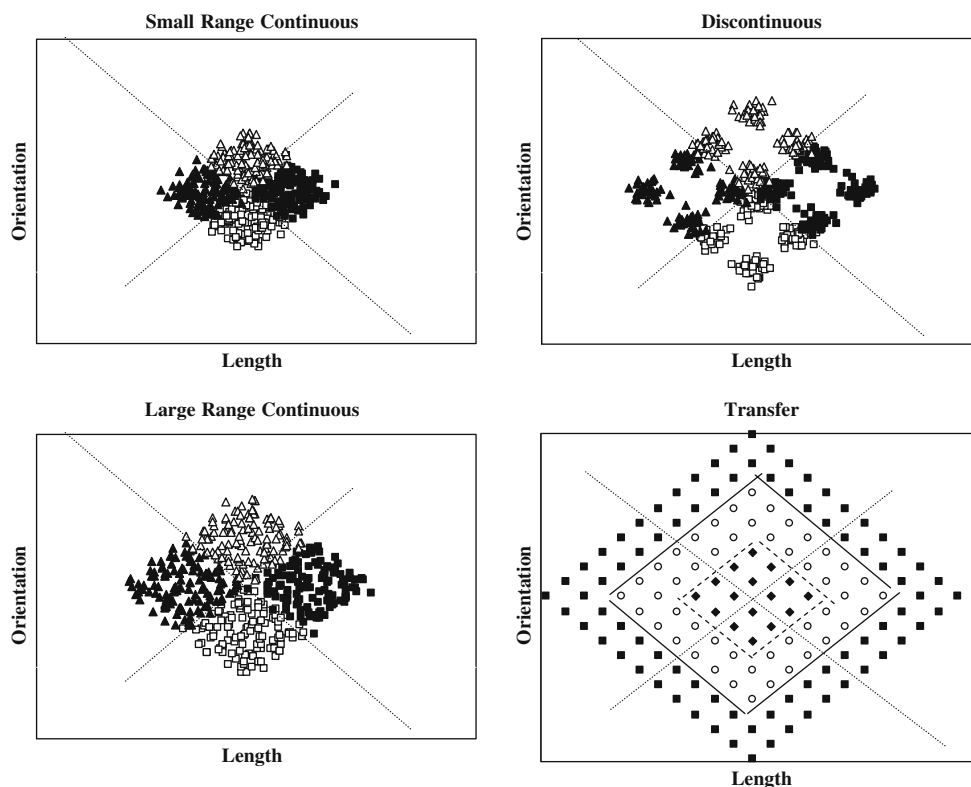
Fig. 1 Categorization conditions used for Experiment 1. The *x*-axis denotes the line length in pixels, and the *y*-axis denotes the line orientation in degrees. Filled triangles denote stimuli from Category A, open triangles denote stimuli from Category B, open squares denote stimuli from Category C, and filled squares denote stimuli from Category D. The solid lines that form an "x" in the small-range continuous, discontinuous, and large-range continuous conditions denote the optimal decision bounds. Small-range continuous condition: In the transfer stimulus plot, all items that lie within the small, broken-line parallelogram (*filled diamonds*) denote novel transfer items from within the range of training items, whereas all items outside this parallelogram (*open and filled squares*) denote novel transfer items from outside the range of training items. Large-range continuous condition and discontinuous condition: All items that lie within the larger, solid-line parallelogram (*filled diamonds and open squares*) denote novel transfer items from within the range of the training items, whereas all items outside this parallelogram (*filled squares*) denote novel transfer items from outside the range of training items

As the results (presented below) suggest, one of these experiments supports the hypothesis that category discontinuity, and not category range, leads to poor initial acquisition but better generalization, whereas the other supports the hypothesis that category range, and not discontinuity, leads to poor initial acquisition but better generalization. Importantly, were we to focus our interpretation only on these empirical data, we would be left with a contradictory set of findings. However, by applying computational models, we offer a unified explanation of these findings that is consistent with the known processing characteristics of the procedural-based learning system that is thought to mediate information-integration classification acquisition (Ashby et al., 1998; Ashby & Ennis, 2006; Cincotta & Seger, 2007; Nomura et al., 2007). To do so, we apply a procedural-based learning model, called the *striatal pattern classifier* (SPC; Ashby & Waldron, 1999), to the data. To anticipate, the model-based analyses suggest that neither increased category range nor category discontinuity

accounts for the results. Rather, the more direct mediator of performance appears to be whether a single-unit or multiple-units representation best represents each category. We now briefly outline the model-based approach.

Model-based approach

The goal of the model-based approach is two-fold. First, we use the models to determine whether and when participants are using the task-appropriate process—that is, a process consistent with the known characteristics of the procedural-based learning system—or an alternative process. We focus our analyses on those individuals who used the appropriate process, since they are the ones who will be most telling in regard to the questions we ask. Second, we use the models to determine when participants use a multiple-unit representation and whether these situations correspond to cases in which a multiple-unit representation is predicted. Although this

approach will be important for future research, we have not used this report to further develop the SPC as a formal model of procedural-based classification learning.

The model-based approach involves applying three models separately to the data from each participant (the details are provided in the Appendix). The first of these is the SPC, which is a computational model whose processing is consistent with what is known about the neurobiology of the procedural-based category learning system thought to underlie information-integration classification performance (Ashby et al., 1998; Ashby & Ennis, 2006; Ashby & Waldron, 1999; Nomura et al., 2007; Seger & Cincotta, 2005). The second is rule-based and instantiates hypothesis-testing strategies such as the application of unidimensional or conjunctive rules. These are verbalizable strategies that are suboptimal in the present studies, but are often utilized by participants. The third is a random-responder model that assumes that the participant guesses on each trial. The model parameters are estimated using maximum likelihood procedures (Ashby, 1992; Wickens, 1982). When the models are nested, G^2 (likelihood ratio) tests will be applied below in order to determine the best model. When models are not nested, the goodness-of-fit statistic will be the Akaike information crieterion:

$$AIC = 2r - 2\ln L,$$

where r is the number of free parameters and L is the likelihood of the model given the data (Akaike, 1974; Takane & Shibayama, 1992). The AIC statistic penalizes a model for extra free parameters in such a way that, the smaller the AIC, the closer a model is to the "true model," regardless of the number of free parameters (for a discussion of the complexities of model comparisons, see Myung, 2000; Pitt, Myung, & Zhang, 2002).

Because the focus of this research is on information-integration learning and generalization, we here describe the SPC in more detail. The SPC assumes that stimuli are represented perceptually in higher-level visual areas, such as the inferotemporal cortex. Because of the massive many-to-one (approximately 10,000-to-1) convergence of afferents from the cortex to the striatum (Ashby & Ennis, 2006; Wilson, 1995), a low-resolution map of perceptual space is represented among the *striatal units*. During acquisition training, the striatal units become associated with one of the category labels so that, after acquisition training is complete, a category response label is associated with each of a number of different regions of perceptual space. In effect, the striatum learns to associate a response with clumps of cells in the visual cortex. When all of the stimuli coming from a category are perceptually similar and form a coherent (or continuous) group, the category can be represented by a small number of units. However, when the stimuli coming

from a category are perceptually dissimilar and form a less coherent, or even discontinuous, group, a multiple-unit representation will be needed.

It is important to be clear that the SPC is a computational model inspired by what is known about the neurobiology of the striatum. Because of this fact, the striatal "units" are hypothetical and could be interpreted within the language of other computational models (e.g., as "prototypes" in a multiple-prototype model like SUSTAIN; Love, Medin, & Gureckis, 2004). In addition, we do not model learning with the SPC, in the sense that we do not update association weights between units and category labels. Learning models have been proposed (e.g., Ashby, Paul, & Maddox, 2011), but because our focus is on asymptotic acquisition and generalization (see below), computational versions of the model are adequate to capture behavior at the end of learning.

Acquisition and generalization predictions

If we make the reasonable assumption that category learning is computationally and biologically more efficient when each category can be represented by fewer rather than more units, it then follows that the procedural-based learning system should be more efficient when stimuli coming from the same category form a coherent (or continuous) group, because fewer units might then be required. Alternatively, it should be less efficient when stimuli coming from the same category are perceptually dissimilar and form a less coherent (or discontinuous) group, because more units might be required to represent the categories. In line with this prediction, previous research has suggested that initial acquisition is slower when more units are needed to represent each category, such as when the decision bound is nonlinear (Ashby & Maddox, 1990, 1992; Ashby & Waldron, 1999; Maddox et al., 2007). In fact, Maddox et al. (2007) found that more units were needed to account for final-block acquisition in their discontinuous condition than in their small-range continuous condition.

Although initial acquisition might be slower when a multiple-unit representation is required, it is reasonable to hypothesize that a more distributed multiple-unit representation might lead to better generalization. This follows because novel stimuli presented during generalization would be more likely to activate one of the many units associated with a multiple-unit representation and would be less likely to activate the one unit associated with a single-unit representation. Importantly, this prediction holds in cluster models such as SUSTAIN (Love et al., 2004). Taken together, this reasoning implies that the more striatal cells are involved in representing a category, the more difficult

category acquisition will be, but the more likely it will be that a novel stimulus (e.g., stimuli presented during transfer) will be associated with that category. This leads to two predictions. First, multiple-unit SPC models should provide better model fits to data collected in generalization conditions for which a multiple-unit representation was needed during acquisition. Second, generalization accuracy rates should be higher in these conditions, and should be especially high when a multiple-unit SPC model provides the best account of the data.

We turn now to the experiments. For each experiment, we first generate performance predictions based on the category structure and the models. We then use the models to help organize the results by breaking participants into groups based upon the best-fitting model: SPC, rule-based, or random-responder.

Experiment 1

Experiment 1 used stimuli that were lines varying in length and orientation across trials. Scatterplots of the training stimuli in the small-range continuous, large-range continuous, and discontinuous conditions are displayed in Fig. 1, along with a scatterplot of the transfer stimuli. The transfer block included test items from within and outside the trained portion of the stimulus space. Each point in Fig. 1 denotes a unique stimulus, with each symbol denoting stimuli from different categories.

Maddox et al. (2007) found that the small-range continuous condition required a single-unit representation for each category, whereas the discontinuous condition require a multiple-unit (specifically, four-unit) representation. They reasoned that because the stimuli in each of the small-range continuous categories were tightly packed around a single, central prototype, a single-unit representation would suffice. They also reasoned that because the stimuli in each of the discontinuous categories were from discontinuous clusters of stimuli, a multiple-unit representation would be needed. In the large-range continuous condition of Experiment 1, the stimuli were not tightly packed around a central prototype, but they were continuously sampled and, more importantly, spread evenly around a central prototype. This made a single-unit representation (or, at the very least, one strong central unit surrounded by other weaker units) likely in this condition. Thus, for Experiment 1, a single-unit representation was predicted in the two continuous conditions, and a multiple-unit representation was predicted in the discontinuous condition. This implies that initial acquisition should be superior in the two continuous conditions but that generalization should be superior in the discontinuous condition.

Method

Participants A total of 90 participants (30 per condition) completed the study and received course credit for their participation. All participants had normal or corrected-to-normal vision. Each participant served in only one condition, and all met a learning criterion of 55% in the final acquisition training block.

Stimuli and stimulus generation The stimuli are displayed in Fig. 1, along with the optimal decision bounds. The category distribution parameters are outlined in Table 1, and optimal accuracy was 95%. In the small-range continuous condition and the discontinuous condition, each category was composed of four "subclusters" (16 total), with 30 stimuli being sampled randomly from each, for a total of 480 stimuli. In the large-range continuous condition, each category was composed of nine "subclusters" (36 total), with 13 stimuli being sampled randomly from each and 3 additional stimuli being sampled randomly from each category, for a total of 480 stimuli. The random samples were linearly transformed so that the sample mean vector and sample variance–covariance matrix equaled the population mean vector and variance–covariance matrix for each subcluster. Each random sample (x, y) was converted to a stimulus by deriving the length (in pixels) as $l = x$ and the orientation (in degrees counterclockwise from horizontal) as $o = 18y/50$. These scaling factors were chosen to roughly equate the salience of each dimension. The resulting 480 stimuli were randomized and divided into five 96-trial blocks separately for each participant. These were presented during category acquisition training. A total of 144 stimuli (36 from each of the four response regions) were used during the no-feedback transfer phase (see Fig. 1).

Procedure Each participant was run individually in a dimly lit testing room with an approximate viewing distance of 35 cm. The participants were informed that there were four equally likely categories. They were also informed that perfect performance was impossible, but that high levels of accuracy could be achieved. They were instructed to learn about the categories, to be as accurate as possible, and not to worry about their speed of responding. On each trial, the stimulus appeared and remained on the screen until the participant generated a response by pressing one of two keys. The correct category label was then presented on the screen for 1 s, along with the word "wrong," if the response was incorrect, or "right," if the response was correct. Once feedback was given, the next trial was initiated. The procedure for the transfer trials was identical, except that feedback was omitted.

Table 1 Category distribution parameters from Experiment 1

μ	x	μy
Small-Range Continuous Condition		
A	240	54
A	216	78
A	264	78
A	240	102
B	192	102
B	168	126
B	216	126
B	192	150
C	288	102
C	264	126
C	312	126
C	288	150
D	240	150
D	216	174
D	264	174
D	240	198
Large-Range Continuous Condition		
A	240	6
A	216	30
A	192	54
A	264	30
A	240	54
A	216	78
A	288	54
A	264	78
A	240	102
B	168	78
B	144	102
B	120	126
B	192	102
B	168	126
B	144	150
B	216	126
B	192	150
B	168	174
C	312	78
C	288	102
C	264	126
C	336	102
C	312	126
C	288	150
C	360	126
C	336	150
C	312	174
D	240	150
D	216	174
D	192	198
D	264	174

Table 1 (continued)

μ	x	μy
D	240	198
D	216	222
D	288	198
D	264	222
D	240	246
Discontinuous Condition		
A	240	6
A	192	54
A	288	54
A	240	102
B	168	78
B	120	126
B	216	126
B	168	174
C	312	78
C	264	126
C	360	126
C	312	174
D	240	150
D	192	198
D	288	198
D	240	246

The standard deviation along the x- and y-dimensions was 10 for all subclusters, and the covariance was 0

Results

The Results section is organized as follows. First, we apply the models to the final acquisition block and to the transfer block to determine whether each participant was using a procedural-based, rule-based, or random process. Because of concerns with modeling aggregate data, each participant's data were fit separately (e.g., Ashby, Maddox, & Lee, 1994; Estes, 1956; Maddox, 1999; Maddox & Ashby, 1998; Smith & Minda, 1998). Four versions of the striatal pattern classifier (SPC) were fit to the data (SPC-1, SPC-2, SPC-4, and the optimal model). The SPC-1 assumed one unit per category, the SPC-2 assumed two units per category, and the SPC-4 assumed four units per category. Models with more units were not examined since, at most, a category contained four clusters of stimuli. The optimal model assumed that the optimal decision bounds were applied. Each of the SPC models assumed that, on each trial, the participant determined which unit was closest to the perceptual effect and gave the associated response— with the only difference among the models being the difference in the number of units. If one of these four

models provided the best account of the data, the participant was classified as an "SPC user." A number of conjunctive and unidimensional hypothesis-testing models, as well as the random-responder model, were also applied to the data (see the Appendix). If one of the hypothesis-testing models provided the best account of the data, the participant was classified as a "rule-based user." If the random-responder model provided the best account of the data, the participant was classified as a "random responder."

Second, because our main focus is on participants who used procedural-based learning strategies during the final acquisition block *and* during the transfer block, we display the learning curves and transfer performance for participants who were classified as SPC users during both the final acquisition block *and* the transfer block. For these same participants, we also examine transfer performance in greater detail by examining transfer performance for items sampled from within the trained portion of the space separately from items sampled from outside the trained portion of the space. For completeness, we also display the learning curves and transfer performance for those participants who did not use a procedural-based learning strategy in the final acquisition block *or* the transfer block. Finally, we examine the nature of strategy shifts across the final acquisition and transfer blocks and the performance under various strategy-shift conditions. The focus of this analysis was to compare and contrast performance for single- versus multiple-unit SPC users.

Learning curves and transfer performance for participants best fit by the SPC in the final acquisition and transfer blocks Figure 2a displays the average proportions correct for the small-range continuous, large-range continuous, and discontinuous conditions for each of the five acquisition training blocks and for the transfer block, *only* for those participants who were SPC users (i.e., those best fit by the SPC-1, SPC-2, SPC-4, or optimal model) in the final acquisition block and the transfer block. To reiterate, it was important to focus on these individuals because they are using the task-appropriate process. This analysis included 57%, 73%, and 93% of the participants from the small-range continuous, large-range continuous, and discontinuous conditions, respectively. A 3 Condition (small-range continuous vs. large-range continuous vs. discontinuous) × 5 Acquisition Block ANOVA was conducted. There was a significant effect of condition [$F(2, 64) = 16.99$, $p < .001$, $\eta^2 = .347$] that suggested worse acquisition in the discontinuous condition relative to the two continuous conditions (ps $< .001$ for both comparisons), with the latter two conditions showing no significant performance differences. There was a significant effect of block [$F(4, 256) = 29.29$, $p < .001$, $\eta^2 = .314$] suggesting that learning
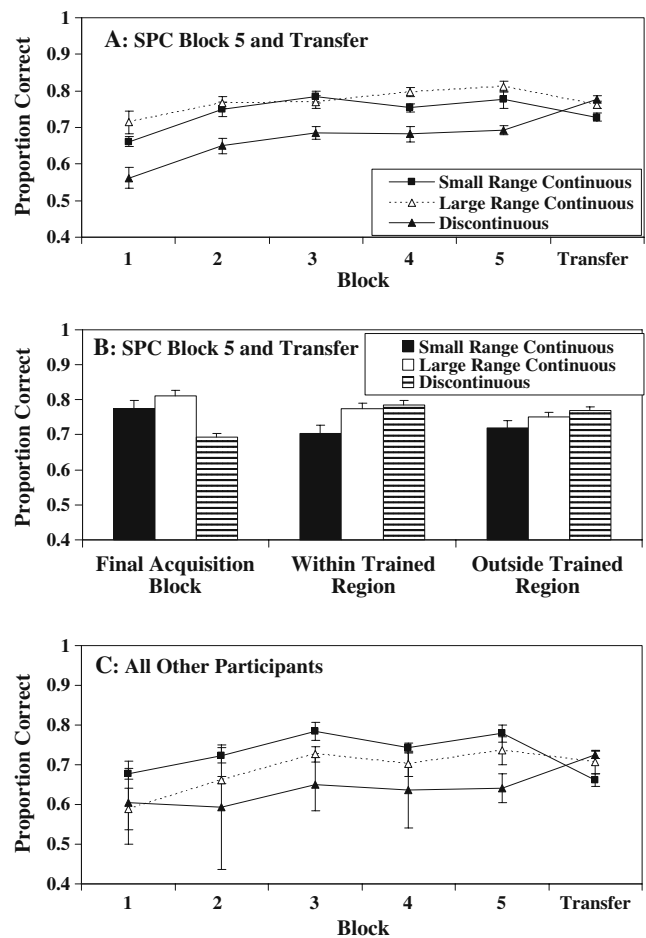

Fig. 2 **a** Proportions correct (averaged across participants) from the acquisition training and transfer phases of Experiment 1 for participants best fit by any of the SPC models in the final acquisition block *and* the transfer block. **b** Absolute proportions correct for the final acquisition block and for no-feedback generalization transfer items from within and outside the trained region of the space, for the same participants shown in panel A. **c** Proportions correct (averaged across participants) from the acquisition training and transfer phases of Experiment 1 for all participants not included in panel A. Standard error bars are included

occurred, and the interaction was nonsignificant ($F < 1$). Thus, category range had no effect on initial acquisition, as suggested by a comparison of performance in the small- and large-range continuous conditions, whereas category discontinuity had a large attenuating effect on initial acquisition, as suggested by a comparison of performance in the large-range continuous and discontinuous conditions.

We also examined the change in performance from the final acquisition block to the transfer block. The performance drop was nonsignificant in the small-range continuous condition but was significant in the large-range continuous condition ($p < .01$). In the discontinuous condition, there was a significant performance *increase* ($p < .01$).

Figure 2b displays the transfer performance for items from within and outside the trained region (along with

performance in the final acquisition block) for the participants displayed in Fig. 2a. The effect of condition was significant for the transfer items from within the trained region of the space [$F(2, 64) = 6.39$, $p < .01$, $\eta^2 = .165$] and was characterized by significantly worse performance in the small-range continuous condition (.70) than in the large-range continuous (.78) and discontinuous (.79) conditions (both ps $< .01$), with no performance difference emerging for the latter two conditions (n.s.). The effect of condition was nearly significant for the transfer items from outside the trained region of the space [$F(2, 64) = 2.81$, $p = .068$, $\eta^2 = .081$] and was characterized by a significant performance difference between the small-range continuous (.72) and discontinuous (.77) conditions ($p < .05$).

For completeness, Fig. 2c displays the learning curves and transfer block performance for the remaining 43%, 27%, and 7% of the participants from the small-range continuous, large-range continuous, and discontinuous conditions, respectively. These are the individuals whose final acquisition block *or* transfer block data were best fit by a rule-based model or by the random-responder model. Given the small sample size, ANOVAs were not conducted.

Taken together, these data suggest that for SPC users (those best fit by one of the SPC models in the final acquisition and transfer blocks), acquisition is worse but transfer is better in the discontinuous condition than in the two continuous conditions. This supports our initial claim that discontinuous categories should be more difficult to acquire but should lead to better transfer. What these data do not tell is whether this performance pattern is due to the increased use of multiple-unit representations in the discontinuous condition. To answer this important question, we turn now to a more detailed analysis that examines performance separately for single-unit SPC users and multiple-unit SPC users. As outlined in the introduction, we predict that a multiple-unit representation will be more likely to provide a better account of the data in the discontinuous than in the two continuous conditions.

Single- versus multiple-unit SPC analyses The percentage of participants in each condition whose final acquisition block and transfer block of data were best fit by specific model pairings, along with the proportions correct achieved by those participants, is presented in Table 2. Five model pairings are examined. First, we examine performance for participants whose final acquisition block of data was best fit by the single-unit SPC (sSPC). These participants were divided into those whose transfer block was also fit by an SPC model (SPC-1, SPC-2, SPC-4, or optimal; henceforth, sSPC–SPC) or by some other model (i.e., one of the hypothesis-testing models or the random-responding model; SPC–other). Next, we examine performance for participants

whose final acquisition block of data were best fit by a multiple-unit SPC (mSPC; SPC-2, SPC-4, or optimal model in transfer block). These participants were divided into those whose transfer block was also fit by an SPC (mSPC–SPC) or by some other model (mSPC–other). Finally, we examined performance for all participants whose final acquisition block of data was best fit by a rule-based model or the random-responder model.

Several comments are in order. First, the proportion of mSPC–SPC participants was largest in the discontinuous condition (47%) and was smaller in the small-range (13%) and large-range (20%) continuous conditions. This supports our hypothesis that discontinuous training, relative to continuous training, is more likely to lead to a multiple-unit representation. Second, final-acquisition-block accuracy was higher for mSPC–SPC participants than for any other participant group, and this was especially salient in the discontinuous condition. Finally, transfer block accuracy was higher for mSPC–SPC participants than for any other participant group, and this also was especially salient in the discontinuous condition.

Discussion

These data suggest that discontinuous category training leads to worse initial acquisition but better performance from the final acquisition block to the transfer block, whereas continuous-category training (even when the range of values is equated with that from the discontinuous categories) leads to better initial acquisition but a decrease in performance from the final acquisition block to the transfer block. The model-based analyses suggest that discontinuous, but not continuous, categories are more likely to lead to a multiple-unit representation, and that this might explain the finding of worse initial acquisition but better transfer in the discontinuous conditions. In Experiment 2, we tested this hypothesis against the alternative hypothesis that category discontinuity alone drives the effect. To achieve this aim, we compared discontinuous-condition performance against performance in a continuous condition for which a multiple-unit representation was likely.

Experiment 2

Experiment 1 examined category range and discontinuity effects in a four-category information-integration task and found support for the hypothesis that discontinuous-category training leads to worse initial acquisition but better generalization when category range is held constant. To our knowledge, this is the first study to rigorously test these two hypotheses.

Table 2 Model results from Experiments 1 and 2 for the final acquisition block and the no-feedback transfer block (see the text for details)

Condition	Final Acquisition–Transfer Block Model	Percentage of Participants	Proportion Correct	
			Final Acquisition	Generalization
	Experiment 1			
Small-Range Continuous	sSPC–SPC	43	.77	.73
	SPC–Other	27	.76	.64
	mSPC–SPC	13	.80	.74
	mSPC–Other	10	.80	.67
	Rule-Based or Random	7	.86	.74
Large-Range Continuous	sSPC–SPC	53	.79	.75
	SPC–Other	10	.74	.66
	mSPC–SPC	20	.86	.81
	mSPC–Other	NA	NA	NA
	Rule-Based or Random	17	.73	.74
Discontinuous	sSPC–SPC	47	.67	.76
	SPC–Other	3	.68	.74
	mSPC–SPC	47	.72	.79
	mSPC–Other	NA	NA	NA
	Rule-Based or Random	3	.60	.72
	Experiment 2			
Small-Range Continuous	sSPC–SPC	4	.84	.90
	SPC–Other	NA	NA	NA
	mSPC–SPC	32	.81	.86
	mSPC–Other	14	.78	.79
	Rule-Based or Random	50	.75	.78
Large-Range Continuous	sSPC–SPC	NA	NA	NA
	SPC–Other	NA	NA	NA
	mSPC–SPC	88	.72	.84
	mSPC–Other	4	.65	.67
	Rule-Based or Random	8	.61	.76
Discontinuous	sSPC–SPC	NA	NA	NA
	SPC–Other	NA	NA	NA
	mSPC–SPC	76	.75	.84
	mSPC–Other	8	.75	.82
	Rule-Based or Random	16	.66	.80

NA = not applicable, because no participant's data were best fit by this model

Notice from Fig. 1 that the exemplars from each category in the discontinuous condition are sampled from four subclusters of stimuli. Many computational models of categorization would predict that under these conditions some sort of multiple-unit (e.g., SPC; Ashby & Waldron, 1999), multiple-prototype (e.g., rational model; Anderson, 1991), or multiple-cluster (e.g., SUSTAIN; Love et al., 2004) representation would be required for learning. In other words, each of these subclusters of stimuli would be represented by a separate unit, prototype, or cluster, with each of these being assigned to a specific category. When presented with a new item, the distance between the item and each unit would be calculated and the item assigned to the category associated with the nearest unit.[1] Notice also that the exemplars from each category in the small-range and large-range continuous condition are sampled continuously from the space and are spread evenly around a central prototype. This makes a single-unit representation (or at the very least one strong central unit surrounded by other weaker units) likely in this condition. These observations,

---

[1] It is important to note that we are not necessarily arguing that such representations are formed or used in the manner put forth by classic "prototype" theory or other theories that have suggested a central representation (e.g., SUSTAIN or SPC). How such representations are computationally implemented awaits further study.

along with the finding that acquisition was worse but generalization was better in the discontinuous condition, support the hypothesis that the need for a multiple-unit representation impedes initial learning but results in better generalization.

In Experiment 2, we tested the generalizability of these findings to a two-category case (Maddox et al., 2005) for which multiple units would likely be required to represent the categories, even though no discontinuity existed. Scatterplots of the training stimuli in the small-range continuous, large-range continuous, and discontinuous conditions are displayed in Fig. 3, along with a scatterplot of the transfer stimuli. The transfer block included novel items from within and outside the range of training items, to evaluate generalization. First, notice that each category of training items in the discontinuous condition is composed of two distinct and dissimilar clusters of items. Thus, a two-unit representation per category would likely be required. Next, notice that each category of training items in the large-range

condition, although not composed of discontinuous clusters, is composed of items that are "spread" out parallel along the decision bound. Using prototype terminology, there is no single prototype or central tendency that adequately describes the training stimuli. In this case, it is likely that a multiple-unit representation would also be required. Finally, notice that each category of training items in the small-range condition is tightly packed, likely yielding a single-unit representation. If the large-range continuous condition does, in fact, require a multiple-unit representation, then the performance pattern in that condition should be similar to that observed in the discontinuous condition, which should yield poorer initial acquisition but better generalization relative to the small-range continuous condition.

Method

Participants A total of 90 participants (30 per condition) completed the study and received course credit for their
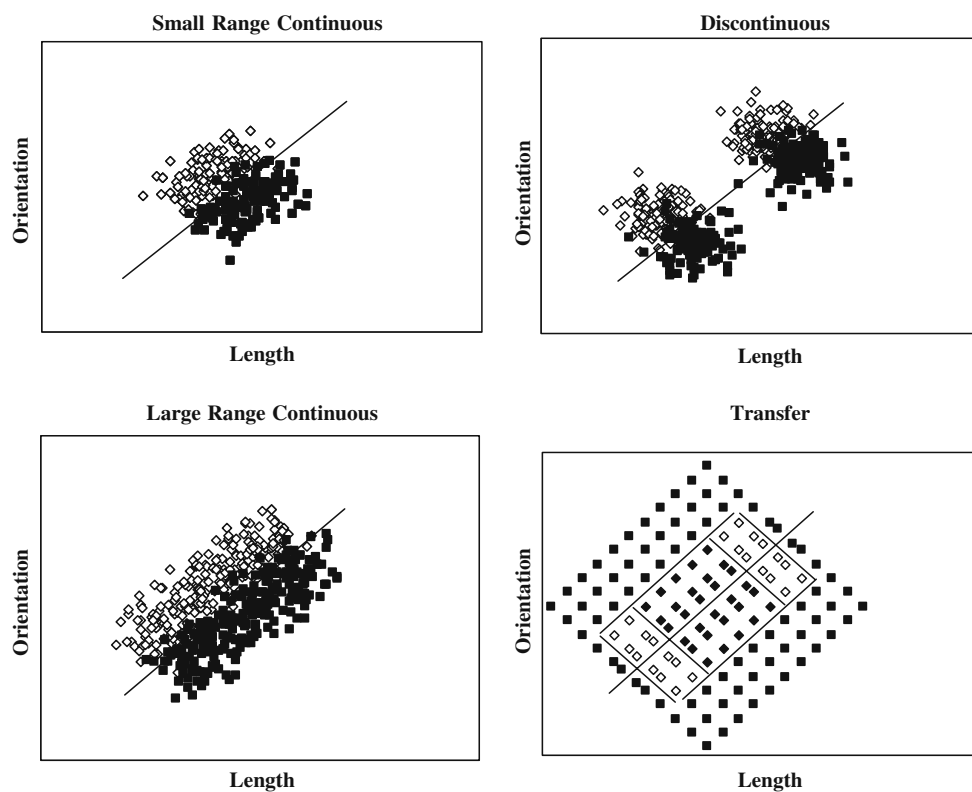

Fig. 3 Categorization conditions used for Experiment 2. The *x*-axis denotes the line length in pixels, and the *y*-axis denotes the line orientation in degrees. Open diamonds denote stimuli from Category A, and filled squares denote stimuli from Category B. The solid lines in the small-range continuous , large-range continuous , and discontinuous conditions denote the optimal decision bounds. Small-range continuous condition: In the transfer stimulus plot, all items (denoted by *filled diamonds*) that lie within the small solid-line parallelogram denote novel transfer items from within the range of the

training items, whereas all items outside this parallelogram denote novel transfer items from outside the range of training items (*open diamonds and filled squares*). Large-range continuous condition and discontinuous condition: All items (denoted by *filled and unfilled diamonds*) that lie within the larger solid-line parallelogram denote novel transfer items from within the range of training items, whereas all items outside this parallelogram (*filled squares*) denote novel transfer items from outside the range of training items

participation. All participants had normal or corrected-to-normal vision, and each participant served in one condition. To ensure that only participants who showed some initial learning during the acquisition phase of the experiment were included in the analyses, a learning criterion of 55% correct during the final acquisition block was applied. All but 12 of the participants met the performance criterion (small range, $n = 28$; large range, $n = 25$; discontinuous, $n = 25$).

Stimuli and stimulus generation The stimuli are displayed in Fig. 3, along with the optimal decision bounds. The category distribution parameters are outlined in Table 3, and optimal accuracy was 90%. In the small-range continuous and discontinuous conditions, each of the two training categories was composed of two "subclusters" (four total) with 120 stimuli being sampled randomly from each, for a total of 480 stimuli. In the large-range continuous condition, each of the two categories was composed of four "subclusters" (eight total) with 60 stimuli being sampled randomly from each, for a total of 480 stimuli. All other aspects of the acquisition training stimuli were identical to Experiment 1. A total of 132 stimuli (66 from the "A" response region and 66 from the "B" response region) were used during the transfer phase and were randomized separately for each participant (see Fig. 4).

Table 3 Category distribution parameters from Experiment 2

	μx	μy
Small-Range Continuous Condition		
A	122	150
A	150	178
B	150	122
B	178	150
Large-Range Continuous Condition		
A	93	122
A	122	150
A	150	178
A	178	206
B	122	93
B	150	122
B	178	150
B	206	178
Discontinuous Condition		
A	93	122
A	178	206
B	122	93
B	206	178

The standard deviation along both the *x*- and *y*-dimensions was 15 for all subclusters, and the covariance was 0
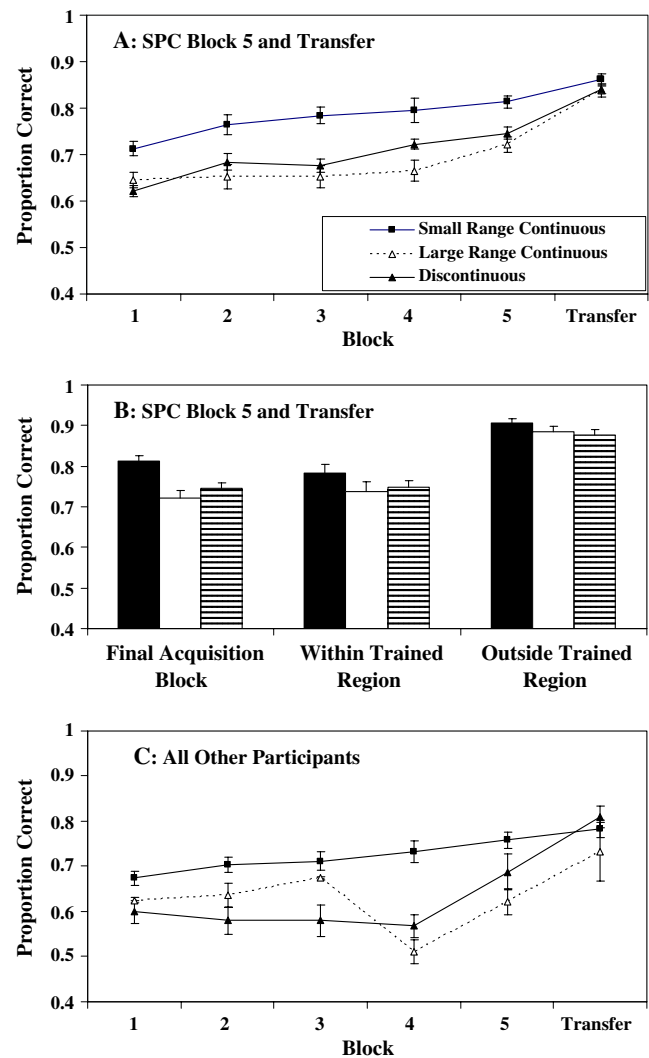

Fig. 4 a Proportions correct (averaged across participants) from the acquisition training and transfer phases of Experiment 2 for participants best fit by the SPC in the final acquisition block *and* the transfer block. **b** Absolute proportions correct for the final acquisition block and for the no-feedback generalization transfer items from within and outside the trained region of the space, for the same participants shown in panel A. **c** Proportions correct (averaged across participants) from the acquisition training and transfer phases of Experiment 2 for all participants not included in panel A. Standard error bars are included

Procedure The procedure was identical to that of Experiment 1, except that two response buttons were used instead of four.

Results

We followed the same data-analytic approach in Experiment 2 that we used in Experiment 1. First, we fitted the models to the final acquisition block and transfer block of data and characterized each participant as an SPC user, rule-based

user, or a random responder. Next, we plotted learning curves, overall transfer performance, and transfer performance broken down by trained versus untrained region for *only* those participants who were classified as SPC users in the final acquisition and transfer blocks. For completeness, we included learning curves for the remaining participants as well. Finally, we examined the nature of strategy shifts across the final acquisition and transfer blocks and performance under various strategy-shift conditions. The focus of this analysis was to compare and contrast performance for single- versus multiple-unit SPC users. The only caveat was that we did not fit the SPC-4 model because, at most, a single category was composed of just two subclusters of stimuli.

Learning curves and transfer performance for participants best fit by the SPC in the final acquisition and transfer blocks The average proportions correct for the small-range continuous, large-range continuous, and discontinuous conditions for each of the five acquisition training blocks and the transfer block for participants classified as SPC users in the final acquisition block *and* the transfer block are displayed in Fig. 4a. This included 36%, 88%, and 76% of the participants from the small-range continuous, large-range continuous, and discontinuous conditions, respectively. A 3 Condition (small-range continuous vs. large-range continuous vs. discontinuous) $\times$ 5 Acquisition Block ANOVA was conducted. There was a significant effect of condition [$F(2, 48) = 9.77$, $p < .001$, $\eta^2 = .289$] that suggested better acquisition in the small-range continuous condition relative to the large-range continuous and discontinuous conditions (ps $< .001$ for both comparisons), with the latter two conditions showing no significant performance differences. There was a significant effect of block [$F(4, 192) = 19.34$, $p < .001$, $\eta^2 = .2874$], suggesting that learning occurred, and no interaction [$F(8, 192) = 1.75$, n.s.). Thus, category range had an effect on initial acquisition, as suggested by a comparison of performance in the small-range continuous condition with the large-range continuous and discontinuous conditions, whereas category discontinuity did not have an effect, as suggested by a comparison of performance in the large-range continuous and discontinuous conditions.

We also examined the change in performance from the final acquisition block to the transfer block. There was a performance increase in all three conditions (all ps $< .001$). Even so, the increase was larger in the large-range continuous and discontinuous conditions than in the small-range continuous condition (both ps $< .05$).

Figure 4b displays the transfer performance for items from within and outside the trained region (along with performance in the final acquisition block) for the same participants. The effect of condition was nonsignificant for the transfer items from within the trained region of the space ($F < 1$) and for the transfer items from outside the trained region of the space ($F < 1$).

For completeness, we also plotted the learning curves and transfer block performance for the remaining 64%, 12%, and 24% of the participants from the small-range continuous, large-range continuous, and discontinuous conditions, respectively. These data are plotted in Fig. 4c. Given the small sample size, ANOVAs were not conducted.

These data suggest that for SPC users (those best fit by one of the SPC models in the final acquisition and transfer blocks), acquisition is worse but transfer is better in the large-range continuous and discontinuous conditions relative to the small-range continuous condition. This supports our initial claim that the two larger-variance conditions (i.e., the large-range continuous and discontinuous conditions) should be more difficult to acquire but should lead to better transfer. What these data do not tell is whether this performance pattern is due to the increased use of multiple-unit representations in the large-range continuous and discontinuous conditions. To answer this important question, we turn to a more detailed analysis that examines performance separately for single-unit SPC users and multiple-unit SPC users. As outlined in the introduction, we predict that a multiple-unit representation will be more likely to provide a better account of the data in the large-range continuous and discontinuous conditions than in the small-range continuous condition.

Single- versus multiple-unit SPC analyses The percentages of participants in each condition whose final acquisition block and transfer block of data were best fit by the five model pairings outlined in Experiment 1 are presented in Table 2. Two comments are in order. First, and as predicted, the proportions of mSPC–SPC participants were largest in the discontinuous (76%) and large-range continuous (88%) conditions, with a much smaller proportion in the small-range continuous condition (32%). Second, final acquisition and transfer accuracies were higher for these participants than for other groups of participants.

Discussion

These data suggest that a large category range, regardless of discontinuity, leads to worse initial acquisition but an increase in performance from the final acquisition block to the transfer block, whereas a small category range leads to better initial acquisition but a smaller increase in performance from the final acquisition block to the transfer block. Based solely on accuracy, this finding appears in conflict with that from Experiment 1, but if one hypothesizes that the two large-range conditions (large-range continuous and discontinuous) required a multiple-unit

representation, which is supported by the model-based analyses, then these data converge nicely with those from Experiment 1, in that both sets of results suggest that the pattern of acquisition and transfer observed is not necessarily related to the breadth or discontinuity of the category structures, but to whether multiple units will provide a better representation of the category structures.

General discussion

Previous research had suggested that discontinuous-category training leads to worse initial acquisition, an increase in performance from acquisition to generalization blocks, and better overall generalization (Maddox et al., 2007). However, in that previous study, discontinuity was confounded with an increase in the range of training items. In this article, two experiments are reported that pitted a discontinuity-and-range explanation of the results against the hypothesis that categories that require multiple-unit representations, as opposed to single-unit representations, lead to worse initial acquisition but better generalization. Taken together, the data support the multiple-unit hypothesis.

In the remainder of this discussion, we will address a number of relevant issues.

SPC versus other non-neuroscience-based models

The multiple-unit hypothesis converges nicely with predictions from a number of computational models that share many properties with the SPC, including the grid model (Ashby & Maddox, 1989), the covering version of Kruschke's (1992) ALCOVE model, Anderson's (1991) rational model, and Love et al.'s (2004) SUSTAIN model. Consider Love et al.'s SUSTAIN model as just one example. Although the exact behavior of the model is parameter dependent, across a wide range of parameter settings, SUSTAIN would predict a multiple-unit (called "clusters" in SUSTAIN) representation in the discontinuous condition from Experiment 1, and in the large-range continuous and discontinuous conditions from Experiment 2. In addition, the model would predict a single-unit representation in the small- and large-range continuous conditions from Experiment 1 and the small-range continuous condition from Experiment 2. The model would also most likely predict the immediate generalization advantages, mainly because the representation is spread out, and thus the similarity between a trained unit and the transfer items would be higher. Thus, the present findings are congruent with predictions from a popular computational model (SUSTAIN), as well as with the neurobiologically inspired SPC model.

Other generalization effects in classification

Identifying training conditions that enhance learning and generalization is a fundamental problem facing learning theorists. The present study suggests that training on discontinuous clusters of stimuli can enhance generalization. Other studies have examined this topic, and we briefly review two that are directly relevant. The first is a study by Spiering and Ashby (2008), who examined the effects of different training sequences on information-integration category acquisition using categories similar to those from Experiment 2 above. They compared a condition in which participants began by classifying easy stimuli (far from the decision bound), then classified stimuli of intermediate difficulty (intermediate distance to the decision bound), then classified difficult stimuli (near the decision bound) (easy-to-hard condition) with a condition in which participants began by classifying difficult stimuli, then classified intermediate-difficulty stimuli, then classified easy stimuli (hard-to-easy condition). This training was followed by a transfer block that required classification of all of the training items with feedback. With information-integration categories, they found that transfer performance was superior when difficult items were trained first as opposed to last. Interestingly, no effects emerged for rule-based categories. Future work should determine whether this effect holds when feedback is removed during transfer, as a true test of the permanence of the learning, and whether the effect holds across a broader sampling of stimuli from both within and outside the trained portion of the stimulus space.

In a related study, Kornell and Bjork (2008) examined the effects of spaced versus massed observational training on the learning of artist categories. Participants studied multiple paintings by different artists, with paintings from a given artist being presented sequentially along with the artist's name (massed training), or randomized with other artists' paintings (spaced training), with each painting being accompanied by the appropriate artist's name. A subsequent transfer test with new paintings from the same artists was administered. Participants viewed each new painting and were asked to give a categorization judgment followed by corrective feedback. Transfer was better for spaced than for massed training. Although initial acquisition performance could not be assessed in this study because initial acquisition training was observational (i.e., the category label was presented along with the training stimulus), and thus no response was required, the transfer advantage for spaced observational training is interesting and is likely related in some sense to the multiple-unit training used in the present study. As with the Spiering and Ashby (2008) result, future work should determine whether this effect holds when feedback is removed during transfer, as a true test of the permanence of the learning. Thus, it appears that

generalization advantages might emerge for other types of acquisition training and might not be constrained only to cases in which a multiple-unit representation is required. In fact, we deem this likely.

Training and clinical implications

The implications of this work for training, neurorehabilitation, and clinical assessment should not be overlooked. First, the procedural-based system is critically involved in acquisition and generalization for complex categorization problems—such as the interpretation of medical imaging, the reading of sophisticated instrumentation, the diagnosis of complex illnesses, the identification of threats to security, and so on—yet little is known about the effects of different training regimens. The present study takes a first step toward addressing these issues and suggests that categorization training that builds a multiple-unit representation facilitates generalization. As just one example, these data suggest that when training radiologists to identify benign versus malignant tumors, it would be advantageous to select training samples that cluster into disparate subgroups, with x-rays within each subgroup being highly similar but dissimilar from x-rays in the other subgroups. Second, measures of procedural-based categorization are largely absent from clinical assessment batteries, whereas one popular rule-based categorization task (the Wisconsin card sorting task; WCST) has been used extensively. This lack of procedural-based categorization measures exists despite the plethora of evidence that striatal functioning is impacted in Parkinson's disease, Huntington's disease, and normal aging (Ashby, Noble, Filoteo, Waldron, & Ell, 2003; Filoteo & Maddox, 1999, 2004; Filoteo, Maddox, & Davis, 2001a, 2001b; Filoteo, Maddox, Ing, Zizak, & Song, 2005; Filoteo, Maddox, Salmon, & Song, 2005; Filoteo, Maddox, Simmons, et al., 2005; Maddox & Filoteo, 2001, 2005, 2007; Maddox, Filoteo, Delis, & Salmon, 1996; Maddox, Filoteo, & Huntington, 1998). In fact, in a recent study (Filoteo, Maddox, Salmon, & Song, 2007), we showed that performance in nondemented Parkinson's disease patients on an information-integration task that required a multiple-unit representation was highly predictive of future cognitive decline, whereas the number of perseverative errors on the WCST (a rule-based task) was not predictive. In contrast, their performance on a task that required fewer SPC units was not predictive of cognitive decline. Finally, some neuropsychological disorders (e.g., Alzheimer's disease) do not impact the procedural-based system, which opens the possibility of rehabilitation approaches that emphasize the intact procedural-based learning system (Schacter, Rich, & Stampp, 1985). Thus, a deeper understanding of procedural-based acquisition and generalization could facilitate the success of neurorehabilitation by identifying the subprocesses that can

replace damaged learning processes in various patient groups (Krakauer, 2006) and by identifying optimal conditions under which procedural-based learning processes should be implemented. Taken together, these findings suggest that tasks that tap the procedural-based system might provide more useful clinical assessment tools than do those tasks that are currently used.

Conclusions

Two experiments provided strong support for the hypothesis that the need for a multiple-unit, as opposed to a single-unit, category representation leads to worse initial acquisition, a performance increase from acquisition to generalization, and better no-feedback generalization. We argue that some category structures are more conducive to the need for a multiple-unit representation, and that under these conditions initial category acquisition is slowed, but transfer is enhanced. In line with other models, such as SUSTAIN, we speculate that acquisition is slower because it is more taxing on the system to train multiple units; however, during transfer, a multiple-unit representation increases the likelihood that a novel stimulus will activate at least one of the multiple units needed to represent the category, enhancing transfer performance.

Appendix

Procedural-based category learning models

The SPC-1, SPC-2, SPC-4, and optimal models were applied to the data from each participant separately. The SPC-1 model assumed that there was one striatal "unit" in the length–orientation space for each category, yielding a total of four units in Experiment 1 and two in Experiment 2. The SPC-2 assumed that there were two striatal "units" in the length–orientation space for each category, yielding a total of eight units in Experiment 1 and four in Experiment 2. The SPC-4 assumed that there were four striatal "units" in the length–orientation space for each category, yielding a total of sixteen units in Experiment 1 and eight in Experiment 2. Because the location of one of the units could be fixed, and since a uniform expansion or contraction of the space would not affect the location of the resulting response region partitions, the SPC-1 contained 6 free parameters in Experiment 1—5 that determined the locations of the units and 1 that represented the noise associated with the placement of the striatal units—and 3 free parameters in Experiment 2. The noise parameter estimated the variability associated with the participant's responding, with large variability estimates being associated with less deterministic responding and

small variability estimates being associated with more deterministic responding. The SPC-2 and SPC-4 models were the same as the SPC-1 model, except that the SPC-2 model assumed that each category had two striatal category units (14 parameters in Experiment 1, and 6 in Experiment 2), and the SPC-4 model assumed that each category had four striatal category units (30 parameters in Experiment 1, and 14 in Experiment 2). The optimal model assumed optimal placement of the decision bounds and contained only the single noise parameter.

Hypothesis-testing models

Three conjunctive hypothesis-testing models were applied to the data from Experiment 1. The *conjunctive(1) model* assumed that the participant made one decision about the length of the line (short or long), a separate decision about the orientation of the line (shallow or steep), and then integrated this information postdecisionally (i.e., after deciding whether the line was short or long and after deciding whether the orientation was shallow or steep). The model assumed that the participant used the following decision rule: Respond "A" if the line length is short and the orientation is shallow; respond "B" if the line length is short and the orientation is steep; respond "C" if the line length is long and the orientation is shallow; and respond "D" if the line length is long and the orientation is steep. This model had three free parameters—two decision criterion parameters and a noise parameter that estimated the variability associated with the participant's trial-by-trial memory and application of the decision criteria. The *conjunctive(2) model* instantiated an "extreme values" type of decision rule. This model assumed that the participant set two criteria along the length dimension that partitioned that dimension into three regions. The model assumed that the participant set a criterion along the orientation dimension that was invoked only when the perceived length fell into the intermediate length region. The model assumed that the participant used the following rule: Respond "B" if the length is short, and respond "C" if the length is long; if length is intermediate, respond "A" if the orientation is shallow, and respond "D" if the orientation is steep. The *conjunctive(3) model* was similar. This model assumed that the participant set two criteria along the orientation dimension that partitioned that dimension into three regions. The model assumed that the participant set a criterion along the length dimension that was invoked only when the perceived orientation fell into the intermediate orientation region. The model assumed that the participant used the following rule: Respond "A" if the orientation is shallow, and respond "D" if the orientation is steep; if the orientation is intermediate, respond "B" if the line is short, and respond "C" if the line is long. The conjunctive(2) and

conjunctive(3) models contained four parameters—three criteria and the noise parameter.

Two conjunctive and two unidimensional hypothesis-testing models were applied to the data from Experiment 2. Both conjunctive models assumed that the participant made one decision about the length of the line (short or long), a separate decision about the orientation of the line (shallow or steep), and then integrated this information postdecisionally (i.e., after deciding whether the line was short or long and after deciding whether the orientation was shallow or steep). One version assumed that the participant responded "A" to all short, steep-angle lines and "B" to all others, and the second version assumed that the participant responded "B" to all long, shallow-angle lines and "A" to all others. Both models had three free parameters. The unidimensional length model assumed that the participant set a criterion along length and ignored orientation, generating one response to short lines and the other to long lines. The unidimensional orientation model assumed that the participant set a criterion along orientation and ignored length, generating one response to shallow-angle lines and the other to steep-angle lines. Both models had two free parameters.

The random-responder model assumed fixed probabilities of responding "A," "B," "C," and "D." In Experiment 1, the model had three free parameters to denote the predicted probabilities of responding "A," "B," and "C," with the probability of responding "D" equal to 1 minus the sum of the other three categories. In Experiment 2, the model had one free parameter, to denote the predicted probability of responding "A."

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*, 716–723.

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review, 98*, 409–429.

Ashby, F. G. (1992). *Multivariate probability distributions*. Hillsdale: 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., & Ennis, J. M. (2006). The role of the basal ganglia in category learning. In B. H. Ross (Ed.), *The psychology of learning and motivation (Vol. 46)* (pp. 1–36). San Diego, CA: Academic Press.

Ashby, F. G., & Maddox, W. T. (1989). *Toward a theory of natural categorization.* Paper presented at the annual meeting of the Psychonomic Society, Atlanta, GA.

Ashby, F. G., & Maddox, W. T. (1990). Integrating information from separable psychological dimensions. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 598–612.

Ashby, F. G., & Maddox, W. T. (1992). Complex decision rules in categorization: Contrasting novice and experienced performance. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 50–71.

Ashby, F. G., & Maddox, W. T. (in press). Human category learning 2.0. *Annals of the New York Academy of Sciences.* doi:10.1111/j.1749-6632.2010.05874.x

Ashby, F. G., Maddox, W. T., & Lee, W. W. (1994). On the dangers of averaging across subjects when using multidimensional scaling or the similarity-choice model. *Psychological Science, 5*, 144–151.

Ashby, F. G., Noble, S., Filoteo, J. V., Waldron, E. M., & Ell, S. W. (2003). Category learning deficits in Parkinson's disease. *Neuropsychology, 17*, 115–124.

Ashby, F. G., Paul, E. J., & Maddox, W. T. (2011). COVIS. In E. M. Pothos & A. J. Wills (Eds.), *Formal approaches in categorization* (pp. 65–87). Cambridge: Cambridge University Press.

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

Balota, D. A., Duchek, J. M., & Logan, J. M. (2007). *Is expanded retrieval practice a superior form of spaced retrieval? A critical review of the extant literature.* New York: Psychology Press.

Bjork, R. A. (1994). *Memory and metamemory considerations in the training of human beings.* Cambridge, MA: MIT Press.

Bjork, R. A., & Linn, M. C. (2006). The science of learning and the learning of science. *APS Observer, 19*, 1–2.

Cincotta, C. M., & Seger, C. A. (2007). Dissociation between striatal regions while learning to categorize via feedback and via observation. *Journal of Cognitive Neuroscience, 19*, 249–265.

Cohen, A. L., Nosofsky, R. M., & Zaki, S. R. (2001). Category variability, exemplar similarity, and perceptual classification. *Memory & Cognition, 29*, 1165–1175.

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 127*, 107–140.

Erickson, M. A., & Kruschke, J. K. (2002). Rule-based extrapolation in perceptual categorization. *Psychonomic Bulletin & Review, 9*, 160–168.

Estes, W. K. (1955). Statistical theory of distributional phenomena in learning. *Psychological Review, 62*, 369–377.

Estes, W. K. (1956). The problem of inference from curves based on group data. *Psychological Bulletin, 53*, 134–140.

Filoteo, J. V., & Maddox, W. T. (1999). Quantitative modeling of visual attention processes in patients with Parkinson's disease: Effects of stimulus integrality on selective attention and dimensional integration. *Neuropsychology, 13*, 206–222.

Filoteo, J. V., & Maddox, W. T. (2004). A quantitative model-based approach to examining aging effects on information-integration category learning. *Psychology and Aging, 19*, 171–182.

Filoteo, J. V., Maddox, W. T., & Davis, J. D. (2001a). A possible role of the striatum in linear and nonlinear category learning: Evidence from patients with Huntington's disease. *Behavioral Neuroscience, 115*, 786–798.

Filoteo, J. V., Maddox, W. T., & Davis, J. D. (2001b). Quantitative modeling of category learning in amnesic patients. *Journal of the International Neuropsychological Society, 7*, 1–19.

Filoteo, J. V., Maddox, W. T., Ing, A. D., Zizak, V., & Song, D. D. (2005). The impact of irrelevant dimensional variation on rule-based category learning in patients with Parkinson's disease. *Journal of the International Neuropsychological Society, 11*, 503–513.

Filoteo, J. V., Maddox, W. T., Salmon, D. P., & Song, D. D. (2005). Information-integration category learning in patients with striatal dysfunction. *Neuropsychology, 19*, 212–222.

Filoteo, J. V., Maddox, W. T., Salmon, D., & Song, D. D. (2007). Implicit category learning performance predicts rate of cognitive decline in nondemented patients with Parkinson's disease. *Neuropsychology, 21*, 183–192.

Filoteo, J. V., Maddox, W. T., Simmons, A. N., Ing, A. D., Cagigas, X. E., Matthews, S., et al. (2005). Cortical and subcortical brain regions involved in rule-based category learning. *NeuroReport, 16*, 111–115.

Hahn, U., Bailey, T. M., & Elvin, L. B. (2005). Effects of category diversity on learning, memory, and generalization. *Memory & Cognition, 33*, 289–302.

Hull, C. (1943). *Principles of behavior.* New York: Appleton-Century-Crofts.

Karpicke, J. D., & Roediger, H. L., III. (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 33*, 704–719.

Karpicke, J. D., & Roediger, H. L., III. (2008). The critical importance of retrieval for learning. *Science, 319*, 966–968.

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? *Psychological Science, 19*, 585–592.

Krakauer, J. W. (2006). Motor learning: Its relevance to stroke recovery and neurorehabilitation. *Current Opinion in Neurology, 19*, 84–90.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99*, 22–44.

Landauer, T. K., & Bjork, R. A. (1978). *Optimum rehearsal patterns and name learning.* London: Academic Press.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review, 111*, 309–332.

Maddox, W. T. (1999). On the dangers of averaging across observers when comparing decision bound models and generalized context models of categorization. *Perception & Psychophysics, 61*, 354–375.

Maddox, W. T., & Ashby, F. G. (1998). Selective attention and the formation of linear decision boundaries: Comment on McKinley and Nosofsky (1996). *Journal of Experimental Psychology: Human Perception and Performance, 24*, 301–339.

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., & Filoteo, J. V. (2001). Striatal contributions to category learning: Quantitative modeling of simple linear and complex nonlinear rule learning in patients with Parkinson's disease. *Journal of the International Neuropsychological Society, 7*, 710–727.

Maddox, W. T., & Filoteo, J. V. (2005). The neuropsychology of perceptual category learning. In H. Cohen & C. Lefebvre (Eds.), *Handbook of categorization in cognitive science* (pp. 573–599). Amsterdam: Elsevier.

Maddox, W. T., & Filoteo, J. V. (2007). Modeling visual attention and category learning in amnesiacs, striatal-damaged patients and normal aging. In R. W. J. Neufeld (Ed.), *Advances in clinical cognitive science: Formal modeling of processes and symptoms* (pp. 113–146). Washington DC: American Psychological Association.

Maddox, W. T., Filoteo, J. V., Delis, D. C., & Salmon, D. P. (1996). Visual selective attention deficits in patients with Parkinson's disease: A quantitative model-based approach. *Neuropsychology, 10*, 197–218.

Maddox, W. T., Filoteo, J. V., & Huntington, J. R. (1998). Effects of stimulus integrality on visual attention in older and younger adults: A quantitative model-based analysis. *Psychology and Aging, 13*, 472–485.

Maddox, W. T., Filoteo, J. V., & Lauritzen, J. S. (2007). Within-category discontinuity interacts with verbal rule complexity in

perceptual category learning. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 33*, 197–218.

Maddox, W. T., Filoteo, J. V., Lauritzen, J. S., Connally, E., & Hejl, K. D. (2005). Discontinuous categories affect information-integration but not rule-based category learning. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 31*, 654–669.

Myung, I. J. (2000). The importance of complexity in model selection. *Journal of Mathematical Psychology, 44*, 190–204.

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

Pitt, M. A., Myung, I. J., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review, 109*, 472–491.

Rips, L. J., & Collins, A. (1993). Categories and resemblance. *Journal of Experimental Psychology: General, 122*, 468–486.

Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*, 249–255.

Schacter, D. L., Rich, S. A., & Stampp, M. S. (1985). Remediation of memory disorders: Experimental evaluation of the spaced-retrieval technique. *Journal of Clinical and Experimental Neuropsychology, 7*, 79–96.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*, 207–217.

Seger, C. A., & Cincotta, C. M. (2005). The roles of the caudate nucleus in human classification learning. *The Journal of Neuroscience, 25*, 2941–2951.

Shea, J. B., & Morgan, R. L. (1979). Contextual interference effects on the acquisition, retention, and transfer of a motor skill. *Journal of Experimental Psychology: Human Learning and Memory, 5*, 179–187.

Skinner, B. F. (1938). *The behavior of organisms*. New York: Appleton-Century-Crofts.

Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 24*, 1411–1436.

Smith, J. D., Redford, J. S., Washburn, D. A., & Taglialatela, L. A. (2005). Specific-token effects in screening tasks: Possible implications for aviation security. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 31*, 1171–1185.

Spiering, B. J., & Ashby, F. G. (2008). Initial training with difficult items facilitates information integration, but not rule-based category learning. *Psychological Science, 19*, 1169–1177.

Takane, Y., & Shibayama, T. (1992). *Structure in stimulus identification data*. Hillsdale, NJ: Erlbaum.

Tolman, E. C. (1932). *Purposive behavior of animals and men*. New York: Century.

Wickens, T. D. (1982). *Models for behavior: Stochastic processes in psychology*. San Francisco: W. H. Freeman.

Wilson, C. J. (1995). *The contribution of cortical neurons to the firing pattern of striatal spiny neurons*. Cambridge, MA: MIT Press.