

Linear Separability, Irrelevant Variability, and Categorization Difficulty

Luke A. Rosedahl & F. Gregory Ashby
University of California
Santa Barbara

Abstract

In rule-based (RB) category-learning tasks, the optimal strategy is a simple explicit rule, whereas in information-integration (II) tasks, the optimal strategy is impossible to describe verbally. This study investigates the effects of two different category properties on learning difficulty in category learning tasks – namely, linear separability and variability on stimulus dimensions that are irrelevant to the categorization decision. Previous research had reported that linearly separable II categories are easier to learn than nonlinearly separable categories, but Experiment 1, which compared performance on linearly and nonlinearly separable categories that were equated as closely as possible on all other factors that might affect difficulty, found that linear separability had no effect on learning. Experiments 1 and 2 together also established a novel dissociation between RB and II category learning: increasing variability on irrelevant stimulus dimensions impaired II learning but not RB learning. These results are all predicted by the best available measures of difficulty in RB and II tasks.

Keywords: Categorization; Classification; Linear Separability; Category-Learning Difficulty; Information Integration

Introduction

Categorization is fundamental to much of our daily lives. From the moment we wake in the morning to the moment we close our eyes at night, we are continually perceiving objects, categorizing them, and acting upon those categorization judgments. Given the importance of categorization, it is natural that much research has focused on what makes a categorization task difficult. Since the classic study of Shepard et al. (1961), the field has examined a number of factors that might affect difficulty, including the nature of

the category structures (Alfonso-Reese et al., 2002), the motivational state of the learner (Ell et al., 2011), and various methods of training and testing (e.g., Rosedahl et al., 2018; Spiering & Ashby, 2008a). Here we examine the role of two factors that could potentially impact performance: whether the contrasting categories are linearly or nonlinearly separable, and the effect of increasing variability on stimulus dimensions that are irrelevant to the categorization decision.

Linear Separability

One of the oldest debates about category-learning difficulty concerns the role of linear separability. Two categories are *linearly separable* (LS) if a decision strategy based on a linear combination of stimulus dimension values produces optimal performance. If optimal performance requires a nonlinear weighting of dimensional values then the categories are said to be *nonlinearly separable* (NLS). Another way to determine linear separability is by examining the optimal decision bound – that is, the set of dimensional values that separate stimuli associated with different optimal responses. The optimal decision bound is always linear with LS categories and nonlinear with NLS categories.

There is striking disagreement in the literature about whether or not NLS categories are more difficult to learn than LS categories. Some studies have reported that NLS categories are easier to learn than LS categories (Levering et al., 2019; Wattenmaker et al., 1986; Wattenmaker, 1995), some studies have reported that LS and NLS categories are equally difficult (Medin & Schwanenflugel, 1981; Smith et al., 1997)¹, and some studies have reported that NLS categories are more difficult to learn than LS categories (Blair & Homa, 2001; Maddox & Filoteo, 2001; Wattenmaker et al., 1986; Wattenmaker, 1995).

One hypothesis that explains the variation in results is that linear separability affects category learning difficulty for some strategies, but not for others. This hypothesis predicts that the discrepancies in the literature are because different studies used tasks that encouraged different categorization strategies. Note that this hypothesis is consistent with the now overwhelming evidence that humans can apply qualitatively different types of strategies during category learning (e.g., Ashby & Valentin, 2017; Davis et al., 2012; Nomura & Reber, 2008; Patalano et al., 2001; Reber et al., 2003).

Several other findings support this hypothesis. First, by altering the context of the stimulus information and the participant instructions, Wattenmaker et al. (1986) were able to make LS categories either easier or more difficult to learn than NLS categories. Second, Ashby et al. (2019) recently showed that there is no single difficulty measure that simultaneously accounts for difficulty in both rule-based (RB) and information-integration (II) tasks. Instead, difficulty in these two types of categorization tasks depends on qualitatively different features. In other words, features of an RB task that make it difficult are different from the features that make an II task difficult. If difficulty in different types of categorization tasks depends on different features, then it seems possible that linear separability might affect the difficulty of some tasks, but not others.

For this hypothesis to be true, there should be some consistent task differences between studies that reported LS categories are easier to learn than NLS categories, and

¹On the other hand, it should be noted that the Smith et al. (1997) participants performed poorly for the exception stimuli that made the categories nonlinearly separable, which suggests they may have been using a linear strategy.

studies that reported the reverse ordering. Looking at the previous findings, such task differences are readily apparent. Virtually all studies that used stimuli that vary on binary-valued dimensions reported either that NLS categories are easier to learn than LS categories (Levering et al., 2019) or that there was no difference in difficulty (Medin & Schwanenflugel, 1981; Smith et al., 1997). In contrast, the studies reporting that NLS categories are more difficult than LS categories used stimuli that varied on continuous-valued dimensions (Blair & Homa, 2001; Maddox & Filoteo, 2001).

Feldman hypothesized that in the case of stimuli that vary on binary-valued dimensions, learning difficulty increases with the Boolean complexity of the optimal classification rule (Feldman, 2000, 2004). If stimuli that vary on binary-valued dimensions are divided into two or more categories, then the exemplars of any category always can be identified by a verbal rule that may include some number of "ands" and "ors." For example, suppose we create large and small squares that are either red or green, and category A includes the single large red square. Then the optimal categorization rule is: Respond A if the square is large *and* red; otherwise respond B. This rule can be described using a simple Boolean expression,² and Feldman (2000) defined its Boolean complexity as the length of the shortest logically equivalent Boolean expression (two in this example; i.e., large red). He also showed that Boolean complexity gave a good account of difficulty differences across 41 different category structures that all had optimal rules that could be described verbally. Linear separability is irrelevant to Boolean complexity³, so this hypothesis predicts that linear separability should not predict learning difficulty when the categories are constructed from stimuli that vary on binary-valued dimensions or in any task in which the optimal strategy can be described by a simple Boolean expression (e.g., as in RB tasks).

This prediction of the Boolean complexity hypothesis is supported by a re-analysis of previous studies that examined the effects of linear separability on category-learning difficulty and that used stimuli that varied on binary-valued dimensions. First, Levering et al. (2019) reported that NLS categories are easier to learn than LS categories, but the NLS categories in that study had a lower Boolean complexity than the LS categories (i.e., 5 versus 8). Second, several other studies that used stimuli varying on binary-valued dimensions found no effect of linear separability (Medin & Schwanenflugel, 1981; Smith et al., 1997). Third, consider the well-known categories originally studied by Shepard et al. (1961). Structures III, IV, and V all have a Boolean complexity of 6, but types III and V are NLS whereas type IV is LS. In support of the Boolean complexity hypothesis, Nosofsky et al. (1994) reported similar error rates for all three types, and that a rank ordering of the types by error rate placed the NLS categories (type IV) in the middle (error rates: type III = 6.1%, type IV = 6.5%, type V = 7.5%).

In contrast, the effect of linear separability on category-learning difficulty when the stimuli vary on continuous-valued dimensions is much less clear. Only a few such

²Denote perceived values on the size dimension by X and values on the color dimension by Y . Let X_C denote a size value intermediate between small and large, and Y_C denote a color value that is intermediate between red and green. Then the condition that a stimulus belongs to category A if it is large and red is equivalent to the Boolean expression: $X > X_C$ *and* $Y > Y_C$.

³If the linear bound is orthogonal to one of the stimulus dimensions, then only one dimension is relevant, Boolean complexity is low, and learning is fast. In contrast, in the case of all other LS categories, Boolean complexity is considerably higher, and learning is slower. Therefore, knowledge that categories are LS does not allow for accurate prediction of whether learning will be fast or slow.

studies have examined this issue, and they have generally agreed that NLS categories are more difficult to learn than LS categories (Blair & Homa, 2001; Maddox & Filoteo, 2001). However, there are several reasons why more research is needed. First, the previous studies were designed for other purposes and made no attempt to control the LS and NLS categories on other features that might affect difficulty (e.g., optimal accuracy was different in the Maddox & Filoteo, 2001, LS and NLS conditions), or they included four categories, which complicates the definition of LS (Blair & Homa, 2001).

Second, there is no theoretical reason to expect that linear separability, by itself, should affect category-learning difficulty – at least, not in II tasks, in which accuracy is maximized only if information from two or more incommensurable stimulus dimensions is integrated perceptually at a pre-decisional stage, and category membership depends on perceptual similarity (Ashby & Gott, 1988). The optimal strategy in II tasks has no Boolean algebraic analogue, and therefore note that the Boolean complexity hypothesis makes no predictions in II tasks. Instead, the best predictor of difficulty in II tasks is the Striatal Difficulty Measure⁴ (SDM) (Rosedahl & Ashby, 2019). Rosedahl and Ashby (2019) showed that the SDM accounted for 87% of the variance in final-block accuracy across a wide range of II category-learning data sets, and that this measure provided consistently better predictions than 12 alternative measures. The data sets came from four previously published studies that each included multiple conditions that varied in difficulty. The studies were highly diverse and included experiments with a variety of stimulus types, and both LS and NLS categories. The SDM uses only the ratio of within-category and between-category similarity and therefore assigns no weight to whether the categories are LS or NLS, so it predicts that linear separability, by itself, is irrelevant to II category-learning difficulty.

Variability Along Irrelevant Stimulus Dimensions

Variability within the stimulus set along irrelevant dimensions has been shown to benefit learning in a wide range of tasks, including learning to read (Apfelbaum et al., 2013; Apfelbaum & McMurray, 2011), second-language learning (Lively et al., 1993; Per-rachione et al., 2011), and the learning of motor skills, such as throwing a ball (Kerr & Booth, 1978) or landing an airplane (Huet et al., 2011). In contrast, several categorization studies have reported evidence that variability along irrelevant dimensions impairs learning. For example, category-learning deficits of patients with Parkinson’s disease increase with stimulus variability along irrelevant dimensions (Filoteo et al., 2007) and unsupervised category-learning performance is reduced with increased variability along irrelevant stimulus dimensions (Zeithamova & Maddox, 2009). However, these studies focused exclusively on RB categories, so little is known about how irrelevant variability might affect learning in II tasks.

As mentioned previously, no single measure simultaneously accounts for difficulty in both RB and II tasks (Ashby et al., 2019). Currently, the best predictor of difficulty in RB tasks is Boolean complexity and the best predictor in II tasks is the SDM. Strikingly, these two measures make different predictions about how variability along irrelevant stimulus

⁴It is called the Striatal Difficulty Measure because it was derived from the procedural-learning component of the COVIS model of category learning, which assumes that the key site of learning in II tasks is within the striatum (Ashby et al., 1998; Cantwell et al., 2015).

dimensions should affect difficulty. Boolean complexity predicts that difficulty depends only on the complexity of the optimal strategy and therefore increasing variability on irrelevant dimensions should have no effect on difficulty (i.e., because the optimal strategy does not change). In contrast, the SDM predicts that increasing variability on irrelevant dimensions increases difficulty because it decreases within-category similarity.

Research Goals

Together, these findings motivate the goals of the present article: to determine the role that linear separability plays in II category learning and the possibly different roles that variability along irrelevant stimulus dimensions might play in RB versus II category learning. Toward this end, in Experiment 1 we compared learning in three II conditions that were carefully constructed so that performance differences allow strong inferences to be made about the role of linear separability and irrelevant variability. The results strongly suggested that linear separability plays no significant role in II difficulty, whereas increasing stimulus variability in a direction irrelevant to the categorization decision (i.e., parallel to the category decision bound) increased difficulty. In Experiment 2, we then tested whether increased irrelevant variability affects difficulty in RB tasks. Results showed no effect of variability along the irrelevant dimension on task difficulty, which further supports the hypothesis that different task features determine difficulty in RB and II tasks.

Experiment 1

An ideal test of whether linear separability affects categorization difficulty would compare performance on LS and NLS categories that were equated on other factors that might affect difficulty. If successful, any performance differences would then have to be due to the presence or absence of linear separability. Our approach to this problem was to construct LS and NLS II categories that were equated as closely as possible on difficulty according to the SDM. The SDM was derived from the procedural-learning model of COVIS (Ashby et al., 1998; Cantwell et al., 2015) and predicts difficulty using a ratio of between- and within-category similarity using a similarity function that models the tuning curves of neurons in the visual cortex (Rosedahl & Ashby, 2019). It therefore places no weight on whether the categories are LS or NLS.

We followed this procedure to create the II categories shown in Figure 1. We began with the Short LS condition. Next, we created the NLS condition to be as similar as possible to the Short LS condition, except for the shape of the optimal decision bound. For example, the width of the empty region between the categories is the same for the Short LS and NLS categories. According to the SDM, the Short LS and NLS categories are approximately equal in difficulty (NLS: SDM = .574; Short LS: SDM = .571). As a further test of the role that linear separability and irrelevant variability play in II difficulty, we also constructed the Long LS condition by stretching the Short LS categories in the direction parallel to the category boundary. This stretching does not affect the linear separability of the categories, the optimal decision strategy, or optimal accuracy, but because increasing irrelevant variability decreases within-category similarity, the SDM predicts that the Long LS categories should be more difficult to learn than the Short LS and NLS conditions (Long LS: SDM = .585).

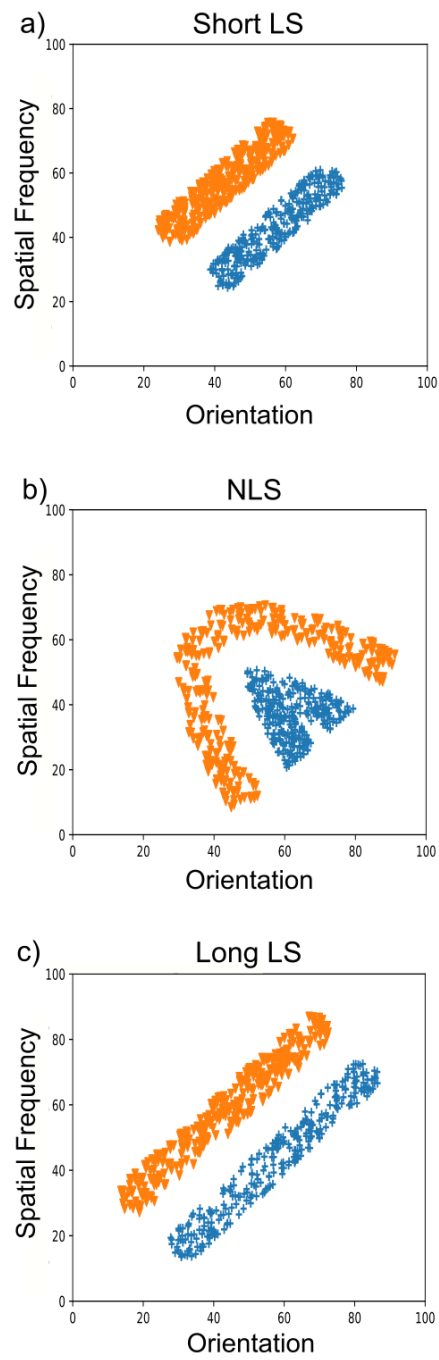


Figure 1

II category structures used in Experiment 1.

Methods

Participants

One hundred and twenty students at the University of California, Santa Barbara (UCSB) participated in this one-hour study in exchange for course credit. The participants were split equally among the three conditions (one for each category structure shown in Figure 1), and each participant learned only one set of categories. All relevant ethical regulations were followed and the study protocol was approved by the Human Subjects Committee at UCSB. Informed consent was obtained from all participants, and every participant was allowed to quit the experiment at any time for any reason without penalty.

Stimuli and Categories

The stimuli were circular sine-wave gratings 3° degrees visual angle in size presented at the center of the screen. All stimuli had the same shape, size and contrast. The stimuli varied only in bar width (as measured by cycles per degree of visual angle or cpd) and bar orientation (measured in degrees counterclockwise rotation from horizontal).

The category structures were all identical in category separation but varied in bound linearity and stimulus density. The linearly separable categories were created by drawing stimuli from uniform cigar shaped distributions centered around the point (50,50) and separated by the ideal bound $y=x$. The distributions were defined by the central line and width of each distribution (see Figure 1 A).

Each Short LS category was uniformly distributed over a cigar-shaped region with length 50 units and width 10 units, and the categories were separated by an empty region that was 10 units wide. Each Long LS category had the same width and separation but had a length of 80 units. The NLS categories were created by drawing stimuli from a uniform distribution and discarding all samples that had an Euclidean distance of less than 10 units or more than 20 units from the parabola $y = -.1(x-50)^2 + 35$ for x in the range 31.3 to 68.7, resulting in a boundary length of 90 units. This increased boundary length was necessary to match the NLS and Short LS categories in SDM-predicted difficulty. The NLS categories were then rotated 45° counterclockwise (see Figure 1b). Stimulus values were converted from 0-100 space to cpd/orientation space using the transforms $CPD = x/30 + .25$ and $\theta = .011\pi(.045y + 1)$ radians counterclockwise from horizontal.

Procedure

At the start of the experiment, all participants were told that they would be shown striped disks and that their task was to categorize these stripes as either an A or B by pressing the 'd' or 'k' keys, respectively. Participants were given no information about the nature of the category structure they were trying to learn.

The disks were shown for 2 seconds and if participants took longer than 5 seconds to respond they were told to respond faster and the trial was not counted. After each response feedback was provided in the form of a green correct or red incorrect label displayed in the center of the screen for 1 second. The participants were trained for 600 trials split into twelve blocks of 50 trials and participants were allowed to take short breaks (less than 2 minutes) between blocks.

Results

Computing difficulty using the SDM

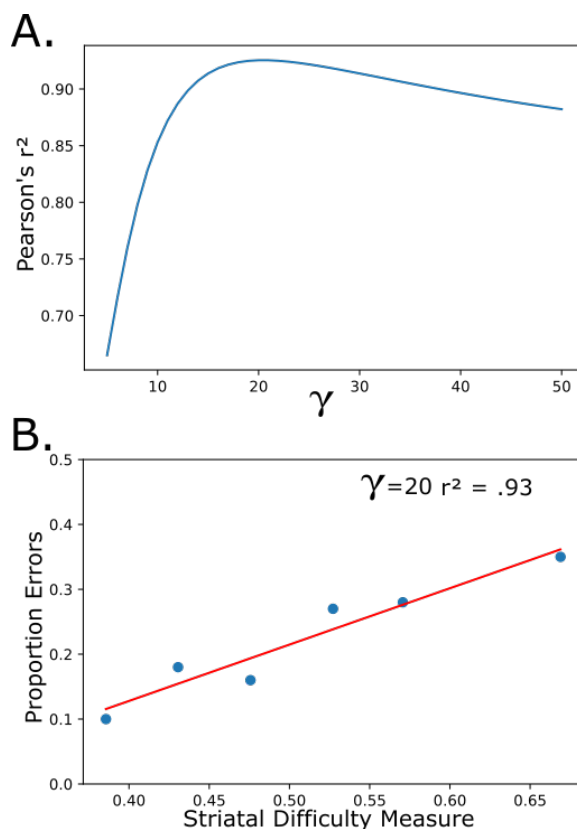
The SDM includes one free parameter, γ , which increases with visual noise and the width of the tuning functions of neurons in visual cortex. Because tuning and visual noise for attended stimuli averaged across participants should remain relatively constant, this direct tie to neurobiology allows us to estimate γ from previous studies, thereby leaving the SDM with no free parameters in the current applications. To this end, we performed a meta-analysis of category-learning studies that used the same stimuli as in the current experiments (i.e., Gabor disks). Specifically, we estimated the value of γ that allowed the SDM to best account for accuracy differences across all studies in this database.

To identify relevant studies, we performed two Google Scholar searches – one on the phrase "category learning" and the other on "category learning Gabor." Both searches added the requirement that selected articles include either the word "implicit" or the phrase "information integration." The first 100 results for each of these searches (so 200 total articles) were examined and included for analysis only if they: 1) used Gabor stimuli with immediate feedback; 2) included at least 10 participants per condition; 3) included at least 500 trials per participant; and 4) reported the category distribution parameters as well as any transformations necessary to convert the units of measurement to cycles per degree of visual angle in the case of bar width and degrees of counterclockwise rotation from horizontal in the case of orientation. This search identified six previous studies: Freedberg et al. (2017), Maddox et al. (2006), Paul et al. (2011), Smith et al. (2010), Spiering and Ashby (2008b), and Vandist et al. (2009). For each of these studies, 600 stimuli were randomly sampled from each category used by the authors. These stimuli were then converted to cycles per degree of visual angle and degrees of counterclockwise rotation from horizontal using the information provided in the articles and were then converted to the 0-100 space shown in Figure 1 using the inverse of the transformations used to create the stimuli in this article. This process ensured all stimuli were in the same space and that the dimensions were approximately equated in terms of just noticeable differences (JNDs).

An SDM analysis was then run for all these category structures following a similar procedure to Rosedahl and Ashby (2019). Briefly, the SDM was computed for all integer values of γ in the interval [5,50] for each category structure. For each value of γ , the SDM was computed separately on 50 different sets of 300 random stimulus samples (150 per category) and then averaged across all sets to determine the SDM value for that value of γ . The best-fitting value of γ was chosen as the value that maximized the correlation between SDM and the proportion of errors during the last 100 trials across all six studies.

Figure 2A shows the r^2 value for all tested values of γ . As can be seen, r^2 peaks at $\gamma = 20$, where it accounts for 93% of the variance in asymptotic accuracy across the 6 studies. Note, however, that any value of γ between about 14 and 37 provides an excellent account of asymptotic accuracy (i.e., $r^2 \geq .90$ throughout this entire range). Figure 2B shows the SDM predictions for each study for $\gamma = 20$.

Finally, using this best-fitting value of $\gamma = 20$, we computed the SDM value of difficulty for each of the three conditions from Experiment 1. The SDM predicts that the difficulty of the NLS condition is approximately equal to (or marginally greater than) the Short LS condition (NLS: SDM = .574; Short LS: SDM = .571), whereas the Long LS

**Figure 2**

A) Proportion of variance accounted for in final-block accuracy across the 6 studies identified in our meta-analysis for different values of the SDM parameter γ . B) Proportion of errors in the 6 studies during the last 100 trials versus predicted difficulty according to the best-fitting version of the SDM. Also shown is the best-fitting regression line, along with the corresponding r^2 and the value of γ used to compute SDM.

categories are substantially more difficult than the Short LS and NLS conditions (SDM = .585).

Performance

Figure 3A shows the learning curves of all participants in each condition and Figure 3B shows asymptotic performance (i.e., average accuracy during the last 100 trials). Note that as predicted by the SDM, performance on the NLS and Short LS categories is almost identical, whereas performance on the Long LS categories is substantially worse. For example, during the last 100 trials, performance was significantly worse in the Long LS condition than in either the Short LS [$t(78) = 3.12, p = .003, BF_{Alt} = 12.55$] or NLS [$t(78) = 3.28, p = .002, BF_{Alt} = 19.2$] conditions, whereas Short LS and NLS accuracies were not significantly different [$t(78) = .18, p = .85, BF_{Null} = 5.9$]. In these comparisons, BF_{Alt} is the JZS Bayes factor favoring the alternative hypothesis that performance is different over the null hypothesis of no difference, whereas BF_{Null} is the JZS

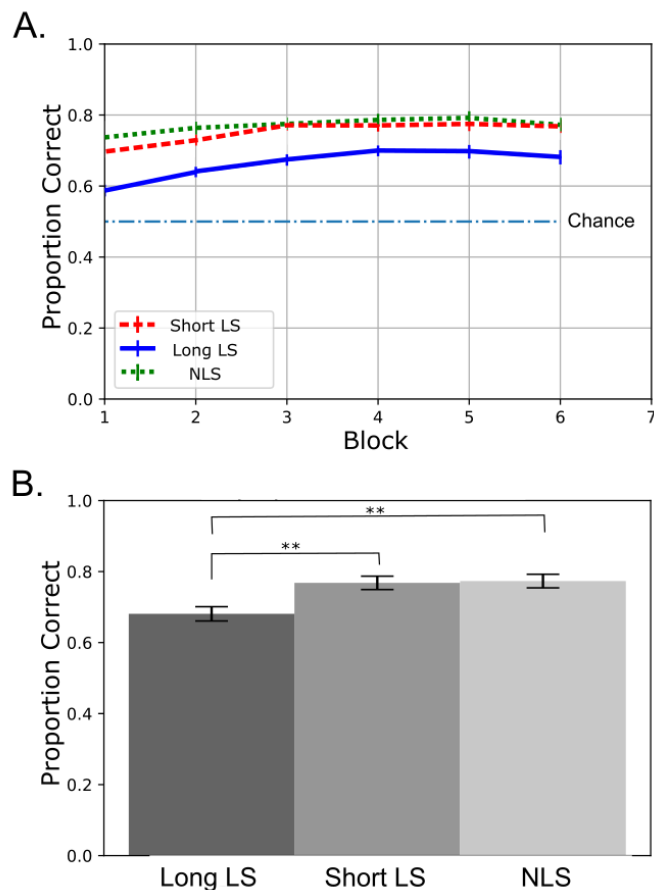


Figure 3

Learning curves (panel A) and asymptotic accuracy (panel B) for all participants in Experiment 1.

Bayes factor favoring the null over the alternative hypothesis (calculated following the method outlined in Rouder et al., 2009 as implemented in their online calculator located at <http://pcl.missouri.edu/bayesfactor>, with the scale factor on effect size set to $r = 1$). Therefore, for example, it is 19.2 times more probable that performance in the Long LS and NLS conditions was different than that there was no difference in performance.

As described earlier, different difficulty measures are required for explicit strategies versus the procedural strategies required to perform well in II tasks (Ashby et al., 2019). Therefore, it is possible that including participants who guessed or used explicit strategies might mask a true effect of linear separability. To correct for this, we assessed the decision strategy of each participant using decision bound modeling (Maddox & Ashby, 1993). Specifically, we determined whether the last 200 responses of each participant were best described by a guessing strategy, the use of some simple explicit rule, or a procedural strategy similar to the optimal rule (for details, see Ashby & Valentin, 2018). Next, we eliminated all participants whose responding was best described by a guessing or explicit rule strategy,

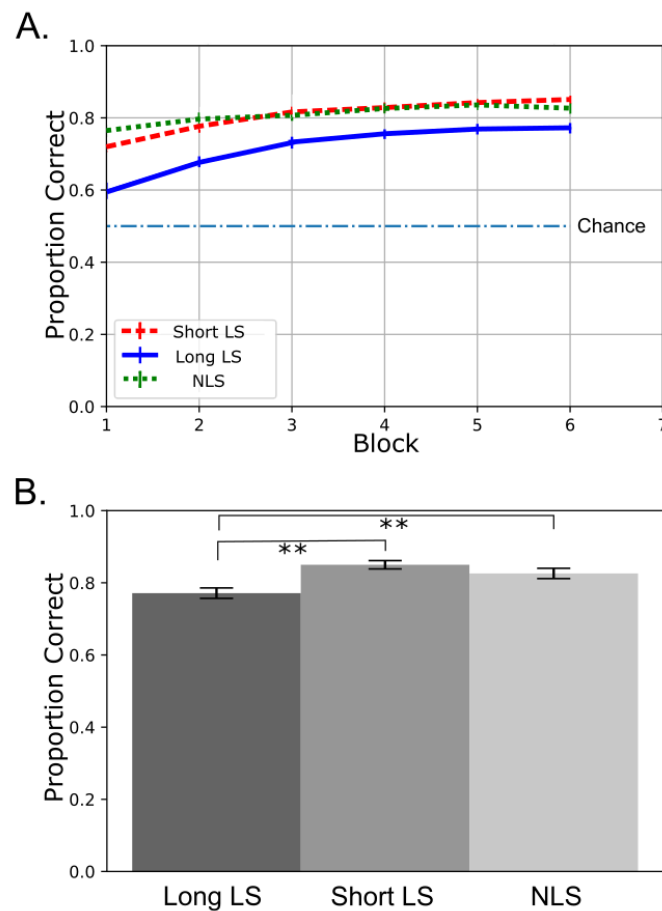


Figure 4

Learning curves (panel A) and asymptotic accuracy (panel B) for all Experiment 1 participants whose responses were best accounted for by a model that assumed a procedural decision strategy.

and only retained data from participants whose responding was best described by a procedural strategy. This left 23 participants in the Long LS condition, 23 in Short LS condition, and 29 in the NLS condition.

Figure 4 shows the learning curves and asymptotic accuracy of these procedural-strategy participants. Note that they are qualitatively identical to Figure 3. The mean accuracies during the last 100 trials were 85% correct for Short LS, 82% correct for NLS, and 77% correct for Long LS. As before, accuracy in the Short LS and NLS conditions were both significantly greater than in the Long LS condition [Short LS vs Long LS: $t(44) = 4.18, p = .0001, BF_{ALT} = 181$; NLS vs Long LS: $t(50) = 2.64, p = .01, BF_{ALT} = 4.1$], and the difference between the Short LS and NLS conditions was not significant [$t(50) = 1.29, p = .2, BF_{Null} = 2.27$].

The SDM, which assigns no role to linear separability, therefore correctly predicted

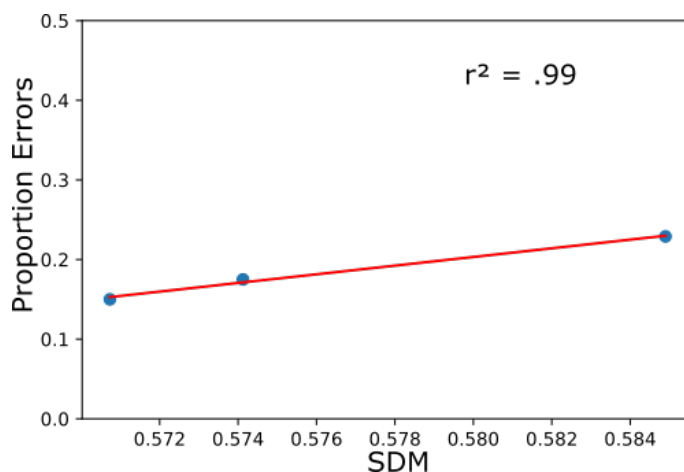


Figure 5

Mean error rate during the last 100 trials in each of the three conditions for all participants who used a procedural strategy plotted against the predicted difficulty of each condition according to the SDM. Also shown is the best-fitting regression line and the proportion of variance accounted for by the SDM.

that the Long LS condition was most difficult and that the Short LS and NLS conditions were approximately equally difficult. Figure 5 examines its quantitative predictions, by plotting the mean error rate during the last 100 trials against the predicted SDM value. Note that the SDM accounts for 99% of the variance in the error rates across these three conditions.

Next, we analyzed the learning curves using a series of Generalized Linear Mixed Models (GLMMs). A traditional linear mixed-effects model analysis would consider this a four-factor design with separate factors for trial, linear separability, predicted difficulty, and participant, with the former three factors fixed and participants random. However, the trial-by-trial data are Bernoulli distributed, and so the assumptions of the linear mixed-effects model are violated. Therefore, we fit a set of GLMMs with logit link functions to the trial-by-trial responses of all participants using MATLAB’s `fitglm` function with maximum likelihood estimates computed via the Laplace method.

The models and their performance are described in Table 1. The null model included a fixed intercept that represented average baseline performance (i.e., β_0 in Table 1), a random intercept for participant that represented individual deviation from the group baseline performance ($P_{0p}, p \in (1, 75)$), a fixed effect for trial (learning slope: $T_i, i \in (1, 600)$), and a random trial \times participant interaction that modeled variation in learning rates across participants ($T_i \times P_{1p}$).

There were four alternative models: model LS included an additional fixed effect of linear separability (i.e., either 0 or 1, $LS_j, j \in (0, 1)$), model LS Int included a fixed effect of linear separability and an interaction between linear separability and trial number ($LS_j \times T_i$), model SDM included an additional fixed effect for SDM difficulty (0 or 1, $SDM_k, k \in (0, 1)$), and model SDM Int included a fixed effect for difficulty plus an interaction between difficulty and trial number ($SDM_k \times T_i$). All Bayes factors in Table 1 estimate the odds that the

Table 1*Generalized Linear Mixed Models for Experiment 1*

Model	Terms	-Log L	BIC	BF
Null	$\beta_0 + P_{0p} + T_i + (T_i \times P_{1p})$	22459	44971	1.0
LS	$\beta_0 + P_{0p} + T_i + (T_i \times P_{1p}) + LS_j$	22453	44970	1.6
LS Int	$\beta_0 + P_{0p} + T_i + (T_i \times P_{1p}) + LS_j + (LS_j \times T_i)$	22450	44974	.22
SDM	$\beta_0 + P_{0p} + T_i + (T_i \times P_{1p}) + SDM_k$	22442	44949	59874
SDM Int	$\beta_0 + P_{0p} + T_i + (T_i \times P_{1p}) + SDM_k + (SDM_k \times T_i)$	22441	44958	665

Note. β_0 = fixed effect of group average baseline (intercept). $P_{0p}, p \in (1, 75)$ = random intercept for participant-specific baseline differences. $T_i, i \in (1, 600)$ = trial. $LS_j, j \in (0, 1)$ = linear separability. $SDM_k, k \in (0, 1)$ = difficulty predicted by the SDM. BF = Bayes factor

alternative model is correct, assuming that either the alternative or null models are correct (estimated using the BIC scores; Raftery, 1995; Wagenmakers, 2007).

The best fitting GLMM was SDM, which included a fixed effect of difficulty as predicted by the SDM (low or high) in addition to the effects included in the null model. SDM had a BIC score of 44949⁵ and the Bayes factor for SDM vs. Null was 59,874 (far above the BF cutoff of 150 for very strong evidence; Raftery, 1995). Additionally, the SDM model fit substantially better than the next best model (SDM Int) with a Bayes Factor of 90 suggesting the data were 90 times more likely to occur under the model with just the SDM vs. the SDM and interaction. Critically, both SDM models fit substantially better than the linear separability models, and the best of the LS models (model LS) barely outperforms the Null model (according to Kass and Raftery (1995) its Bayes factor of 1.6 is "not worth more than a bare mention" (p. 777)).

Discussion

Experiment 1 found no evidence that linear separability has any effect on II category-learning difficulty – either positive or negative. First, participants performed approximately equally well in the Short LS and NLS conditions. This is critical because the categories in these two conditions differed sharply on linear separability but were nearly equated on difficulty according to the SDM and were exactly equated on optimal accuracy and on the width of the empty region that separated the categories. Second, and most importantly, all of our results were accurately predicted by the SDM, which assigns no weight to linear separability when estimating category-learning difficulty. Specifically, the SDM correctly predicted the order of conditions by difficulty and it accounted for 99% of the variance in asymptotic accuracy of participants who used a procedural strategy. Therefore, we found no evidence that linear separability had any effect on categorization difficulty.

Although Experiment 1 failed to find any effect of linear separability on II difficulty, our results did show a clear effect of another factor: within-category variability along an irrelevant stimulus dimension. Increasing the variability of stimuli along the direction parallel

⁵The values are so large because each model was fit to 72000 data values (120 participants \times 600 trials each). Furthermore, the models predict the probability of a correct response on each trial, and each data point is a 0 (error) or 1 (correct), so even a correct model will have a misprediction on every trial.

to the category bound dramatically increased difficulty. Specifically, the asymptotic error rate in the Long LS condition was 40% higher than in the Short LS condition (i.e., 32% versus 23%). Although this effect is predicted by the SDM, to our knowledge, it has not been previously demonstrated. Experiment 2 explores whether this is a general property of category learning, or unique to the procedural learning that is thought to dominate in II tasks.

Experiment 2

Experiment 1 showed that increasing within-category variability in a direction parallel to the optimal decision boundary increased II category-learning difficulty.⁶ However, overwhelming evidence suggests that the categories used in Experiment 1 recruit procedural learning, which raises the question of whether this finding is true for all category types, even those in RB tasks that are learned explicitly. The Boolean complexity hypothesis, as well as models that assume RB learning is mediated by a rule-discovery process, such as COVIS (Ashby et al., 1998), predict that variability along an irrelevant dimension should not affect difficulty. Experiment 2 examines this issue by comparing performance in the four RB conditions described in Figure 6. Note that these four conditions cross two levels of two different factors – variability on the irrelevant dimension and category separation.

Participants

Eighty students at the University of California, Santa Barbara participated in this one hour study in exchange for course credit. The participants were split equally between the four category structures. All relevant ethical regulations were followed and the study protocol was approved by the Human Participants Committee at UCSB. Informed consent was obtained from all participants, and every participant was allowed to quit the experiment at any time for any reason and still receive credit.

Stimuli and Categories

The stimuli and procedures were the same as in Experiment 1. The only difference was the category structures.

Each category was created by random sampling from a uniform distribution defined over a cigar-shaped region of 100×100 stimulus space. The Long Far and Short Far categories were created by rotating the Long Linear and Short Linear structures from Experiment 1 by 45° counterclockwise. The Near structures had the same length and width as their respective Far structures but had a within-category separation of 0 (such that the categories were touching but not overlapping).

Results

The learning curves for the four category structures are plotted in Figure 7A. Note that accuracy depended strongly on category separation. In particular, performance was

⁶The optimal category bound in the Short LS and Long LS conditions was a diagonal line through the origin with slope 1 (i.e., see Figure 1). Therefore, the only variability that is relevant to categorization accuracy is variability orthogonal to this bound. Collapsing the two-dimensional category distributions onto this one relevant dimension produces identical results in the Short LS and Long LS conditions.

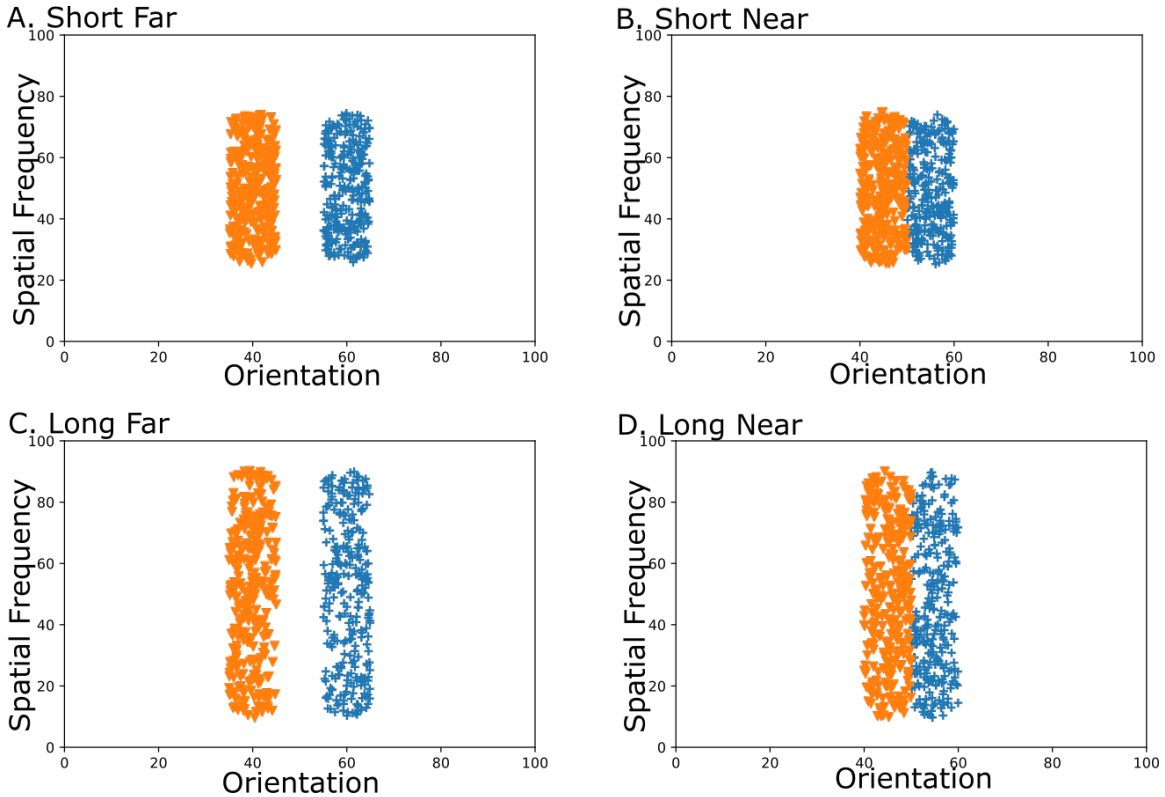


Figure 6

RB category structures used in Experiment 2.

substantially higher in the two Far conditions than in either Near condition. In contrast, unlike Experiment 1, there appears to be little or no effect of variability along the irrelevant dimension. Specifically, the Short Far and Long Far learning curves are almost superimposed, as are the Short Near and Long Near curves.

As in Experiment 1, we tested these conclusions more rigorously by comparing performance of a series of nested GLMMs. These models are described in Table 2. Note that the Null model was the same as in Experiment 1. The alternative GLMMs included models with the addition of just the fixed effects of length ($L_j, j \in (0, 1)$) or separation ($S_k, k \in (0, 1)$), the addition of the fixed effects of length or separation along with an interaction, and a full model with both additional fixed effects and the relevant two-way and three-way interactions (i.e., model LS Int; see Table 2).

Model S (which included a fixed effect for separation) fit the best, with a Bayes factor on the order of 10^{32} vs. the null model and a Bayes factor on the order of 10^{16} vs. the second best fitting model (model S Int; which included the same main effect of separation plus a trial \times separation interaction). Importantly, the separation models outperformed the models that included an effect of length. Additionally, the null model was strongly favored over the model that included a main effect of length (i.e., model L); the Bayes factor for these two models suggests that the null model is 250 times more likely to be correct than

Table 2

Generalized Linear Mixed Models for Experiment 2

Model	Terms	-Log L	BIC	BF
Null	$\beta_0 + P_{0p} + T_i + (T_i \times P_{1p})$	15859	31772	1.0
L	$\beta_0 + P_{0p} + T_i + (T_i \times P_{1p}) + L_j$	15859	31783	.004
L Int	$\beta_0 + P_{0p} + T_i + (T_i \times P_{1p}) + L_j + (T_i \times L_j)$	15858	31792	0
S	$\beta_0 + P_{0p} + T_i + (T_i \times P_{1p}) + S_k$	15812	31624	1.38e32
S Int	$\beta_0 + P_{0p} + T_i + (T_i \times P_{1p}) + S_k + (T_i \times S_k)$	15811	31697	1.93e16
LS	$\beta_0 + P_{0p} + T_i + (T_i \times P_{1p}) + L_j + S_k$	15812	31700	4.3e15
LS Int	$\beta_0 + P_{0p} + T_i + (T_i \times P_{1p}) + L_j + S_k + (L_j \times S_k)$ $+ (L_j \times T_i) + (S_k \times T_i) + (L_j \times S_k \times T_i)$	15807	31733	2.94e8

Note. β_0 = fixed effect of group average baseline (intercept). $P_{0p}, p \in (1, 80)$ = random intercept for participant-specific baseline differences. $T_i, i \in (1, 600)$ = trial. $L_j, j \in (0, 1)$ = length. $S_k, k \in (0, 1)$ = separation. BF = Bayes factor

model L.

Figure 7B shows the accuracies during the final block (after participants had reached asymptotic performance). Overall, performance was good, reaching an average accuracy of 94% correct in the Short Far and Long Far conditions and 82% correct in the Short Near and Long Near conditions. Post-hoc t-tests found no difference between the Short and Long structures in either separation condition [Near: $t(38) = .21, p = .84, BF_{Null} = 4.17$. Far: $t(38) = .18, p = .86, BF_{ALT} = 4.17$]. Performance was significantly worse in the Low separation conditions than in the High separation conditions [Short: $t(38) = 5.99, p < .0001, BF_{ALT} = 24903$. Long: $t(38) = 7.76, p < .0001, BF_{ALT} = 4170662$].

While our analysis found no effect of irrelevant variability on performance, it is possible that the irrelevant variability influenced initial strategy use. If so, we would expect a performance difference during the early stages of learning. To test this, we compared performance during the first 50 trials in the Long and Short conditions. Long and Short accuracies were not significantly different in either the Near or Far conditions (Near: $t(38) = .18, p = .86, BF_{Null} = 4.24$. Far: $t(38) = 1.93, p = .06, BF_{ALT} = 1.13$). The effect was close to significant in the Far conditions, but the Bayes factor was only 1.13, which suggests that this difference is "not worth more than a bare mention" (Kass & Raftery, 1995).

Discussion

Our results found no effect of irrelevant-dimension variability on RB learning. Neither the GLMM analysis nor the post-hoc t-tests found any significant difference between the low and high variability conditions for either category separation. These results provide strong evidence that the increased difficulty we observed in Experiment 1 when within-category variability increased along an irrelevant stimulus dimension is not a universal property of category learning, and is perhaps unique to II category learning.

One potential critique could be that there is an effect of irrelevant-dimension variability on RB learning, but this effect was masked by the high levels of accuracy in our

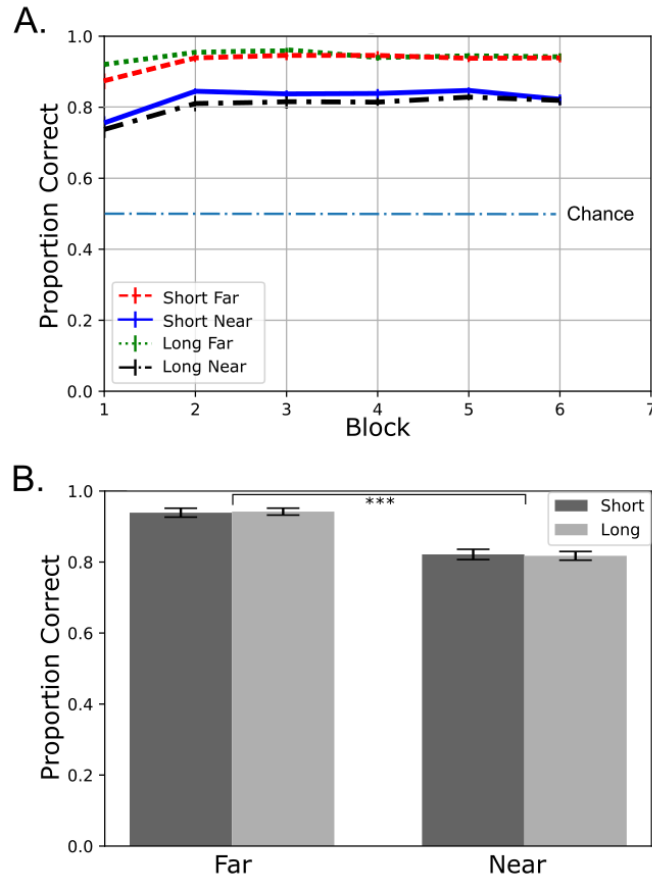


Figure 7

Learning curves (panel A) and asymptotic performance (panel B) for all participants in Experiment 2.

experiment. Note however, that participants in the Near separation conditions had average accuracies of around 82% correct, which is less than the Experiment 1 Short LS participants who used a procedural strategy (i.e., their mean accuracy was 85%). Therefore, it seems unlikely that the absence of an effect of irrelevant-dimension variability in Experiment 2 was because of high accuracy. Additionally, performance in the first 50 trials (where high irrelevant variability might be expected to influence strategy selection) was not different in the low and high irrelevant variability conditions.

In addition to showing that variability along an irrelevant dimension does not affect RB learning difficulty, the Experiment 2 results provide further evidence that no single measure of difficulty can predict performance in both RB and II categories (Ashby et al., 2019). Any measure that predicts difficulty should increase with variability on irrelevant dimensions would successfully predict better performance in the Experiment 1 Short LS condition than in the Long LS condition, but would fail to predict equal performance in the Short and Long Experiment 2 conditions. And of course, any measure predicting no

effect of irrelevant-dimension variability would make the opposite error. A measure that somehow assigned weight to irrelevant-dimension variability except when the optimal bound is horizontal or vertical might succeed, but this would be equivalent to applying qualitatively different measures to RB and II tasks.

Experiments 1 and 2 together also establish another novel dissociation between RB and II learning – II learning is affected by irrelevant-dimension variability, whereas RB learning is not. This dissociation adds to a long list of qualitative differences between human learning and performance in these two tasks. Currently, somewhere around 30 different RB versus II dissociations have been identified (for a review of most of these, see Ashby & Valentin, 2017). Collectively, these results provide overwhelming evidence that humans learn RB and II categories in qualitatively different ways.

General Discussion

Linear Separability

The role of linear separability in categorization difficulty has long been uncertain. Some studies have reported LS categories are easier to learn than NLS categories, some studies have reported the opposite ordering, and some studies have reported no difference in difficulty. In the Introduction, we reviewed evidence that LS categories are not inherently easier to learn than NLS categories when the stimuli vary on binary-valued dimensions. In these cases, the optimal strategy can be described using Boolean algebra, and Feldman (2000) showed that learning difficulty is well predicted by the Boolean complexity of the optimal strategy. In RB tasks, the optimal strategy can also be described using Boolean algebra, even when the stimuli vary on continuous-valued dimensions. For this reason, we predict that RB learning difficulty is also well predicted by the Boolean complexity of the optimal strategy. Boolean complexity assigns no weight to linear separability, so this hypothesis correctly predicts that LS categories are not inherently easier to learn than NLS categories when the stimuli vary on binary-valued dimensions, and we predict a similar non-effect of linear separability in RB tasks.

In II tasks, however, the optimal strategy has no Boolean description, and we know of only one previous study that compared the difficulty of LS and NLS categories in an II task (Maddox & Filoteo, 2001). This study concluded that NLS categories are more difficult to learn than LS categories. Even so, the study was designed for other purposes and made no attempt to control their LS and NLS categories on other features that might affect difficulty. Experiment 1 included LS and NLS II categories that were carefully constructed to be equally difficult according to the best current measure of difficulty in II tasks (i.e., the SDM). For example, optimal accuracy, and the width of the empty region between the Short LS and NLS categories of Experiment 1 were identical. Performance in these two conditions was essentially equivalent, which strongly suggests that linear separability, by itself, has no effect on difficulty in II tasks.

Experiment 1 produced no empirical evidence that linear separability affects II category learning and the best available measure of II category-learning difficulty (i.e., the SDM) predicts no effect of linear separability. But could there be some other, undiscovered measure of II difficulty that outperforms the SDM and predicts an effect of linear separability? Many years ago, Ashby and Waldron (1999) reported evidence that II category learning is

nonparametric – that is, that participants make no assumptions about the parametric form of the optimal decision bound (Ashby & Alfonso-Reese, 1995). Thus, all nonparametric classifiers assign no role to linear separability because in this class of models, the decision bound has no psychological meaning. Therefore, whether it has the parametric form of a line or a quadratic curve is irrelevant. The SDM was derived from a nonparametric model of II category learning, which is why it assigns no weight to linear separability. But neither do any other nonparametric measures. For example, exemplar models of categorization (Medin & Schaffer, 1978; Nosofsky, 1986) are also nonparametric and they assign no role to linear separability. In fact, Ashby and Rosedahl (2017) showed that the exemplar model (i.e., the generalized context model) is a special case of the striatal pattern classifier that motivated the SDM. Therefore, difficulty predictions of exemplar models are essentially equivalent to the SDM. For these reasons, evidence that II category learning is nonparametric is also evidence that linear separability does not affect II category-learning difficulty.

Our results, taken together with existing work, suggest that linear separability plays little or no role in human category learning in tasks where the optimal strategy can be described using Boolean algebra or in II tasks. Much previous research suggests that the former tasks are learned via an explicit, rule-discovery process, whereas II categories are learned procedurally (e.g., Ashby et al., 1998; Ashby & Valentin, 2017). However, it is important to note that none of these results rule out the possibility that linear separability plays a role in other tasks, especially tasks that might be learned using a strategy that does not include rule discovery or procedural learning. One interesting possibility is the prototype-distortion task (Posner & Keele, 1968).

In prototype-distortion category-learning tasks, the category exemplars are created by randomly distorting a single category prototype. The most widely known example uses a constellation of dots (often 7 or 9) as the category prototype, and the other category members are created by randomly perturbing the spatial location of each dot. Sometimes the dots are connected by line segments to create polygon-like images. A popular version begins with a single central Category A and participants are presented with stimuli that are either exemplars from Category A or random patterns that do not belong to Category A. The participant’s task is to respond “Yes” or “No” depending on whether the presented stimulus was or was not a member of Category A. Critically, there have been multiple proposals that prototype-distortion learning is mediated by the perceptual-representation memory system, rather than by rule or procedural learning (Casale & Ashby, 2008; Reber & Squire, 1999). As a result, we believe that our results, along with previous results from studies that used stimuli that vary on binary-valued dimensions, should not be generalized to prototype-distortion tasks.

In fact, Blair and Homa (2001) reported that LS categories are easier to learn than NLS categories in prototype-distortion tasks. Although more research is needed to confirm this finding, this result also has theoretical support. Specifically, prototype models of categorization assume LS categories, and performance in prototype-distortion tasks is well described by prototype models of categorization (e.g., Homa et al., 1981; Smith & Minda, 2002). Furthermore, Wattenmaker et al. (1986) were able to reduce the difficulty of LS categories by changing the stimuli in a way that made the category prototypes more prominent. The stimuli were descriptions of four behaviors emitted by the same person, and the participant’s task was to assign each hypothetical person to one of two categories. When

the four behaviors described four different personality traits then NLS categories were easier to learn. However, when the four behaviors all described the same trait (e.g., honesty), LS categories were easier. In the former case, the categories had no clear prototype, whereas they did in the latter case.

The perceptual-representation memory system is thought to improve processing of a stimulus merely as a result of repeating its presentation – a phenomenon known as repetition priming (e.g., Wiggs & Martin, 1998). Repetition priming occurs even if the two stimuli are different, so long as they are perceptually similar (e.g., Biederman & Cooper, 1992; Cooper et al., 1992; Seamon et al., 1997). Prototype distortion tasks typically use stimuli that vary on many continuous-valued dimensions. For example, the dot patterns used in the original prototype distortion experiments of Posner and Keele (1968), and in many subsequent studies, vary on 18 continuous-valued dimensions. This is critical because random distortions of the prototype are likely to produce more exemplars highly similar to the prototype when the stimuli vary on many perceptual dimensions (Casale & Ashby, 2008).⁷ Therefore, the high-dimensional stimuli typically used in prototype-distortion tasks are likely to induce more repetition priming than the low-dimensional stimuli commonly used in other types of category-learning tasks. So an interesting goal of future research should be to investigate whether repetition priming and prototype-distortion learning are facilitated by linear separability.

Variability Along Irrelevant Stimulus Dimensions

In addition to linear separability, we also examined the effect on learning of stimulus variability on dimensions that are irrelevant to the categorization decision. Such variability has been found to facilitate learning in a variety of different tasks, including second-language learning (Lively et al., 1993; Perrachione et al., 2011), motor-skill learning (Huet et al., 2011; Kerr & Booth, 1978), and teaching children to read (Apfelbaum et al., 2013; Apfelbaum & McMurray, 2011). Previous categorization research has suggested that irrelevant variability increases categorization difficulty (Filoteo et al., 2007; Zeithamova & Maddox, 2009). However, these studies focused exclusively on RB tasks and they either omitted feedback or used neuropsychological patient groups with known category-learning deficits. Therefore, the general role that irrelevant variability plays in category learning is poorly understood.

The difficulty measures used here – namely, the SDM for II tasks and Boolean complexity for RB tasks – make opposite predictions about the effects of irrelevant variability. The SDM predicts that increases in variability along irrelevant stimulus dimensions should increase difficulty. This is because the SDM is based on a model (i.e., the procedural-learning model of COVIS; Cantwell et al., 2015) that assumes II learning is mediated at cortical-striatal synapses. In the II tasks used here, every stimulus is novel. The SDM

⁷See Ashby (2019, pp. 355-360) for a thorough description of this phenomenon. Briefly however, with stimuli that vary on only one dimension, random distortions of a prototype necessarily produce stimuli with a lower or higher value than the prototype on the stimulus dimension. As a result, a few distortions will be similar to the prototype and many will be dissimilar. In fact, in one dimension, the prototype can have only two nearest neighbors. All other exemplars must be more dissimilar to the prototype than these nearest neighbors. In two dimensions, however, the prototype can have five nearest neighbors, in 8 dimensions it can have 240 nearest neighbors, and with stimuli that vary on 24 dimensions, the prototype can have 196,560 nearest neighbors (Odlyzko & Sloane, 1979).

assumes that learning occurs because a novel stimulus is similar to some previously trained stimulus. Similar stimuli excite visual cortical neurons with overlapping tuning curves, so a novel stimulus will excite previously trained striatal neurons, which makes the correct response more likely. Increasing variability along an irrelevant stimulus dimension reduces this effect and therefore impairs learning. In contrast, Boolean complexity predicts that irrelevant variability should have no effect of category-learning difficulty. This is because changing the variability on an irrelevant stimulus dimension does not change the optimal decision strategy.

Experiment 1 verified the predictions of the SDM that increasing variability on an irrelevant stimulus dimension should increase the difficulty of II category learning (i.e., Long LS performance was significantly worse than Short LS performance).⁸ Furthermore, Experiment 2 verified the predictions of Boolean complexity for RB tasks. Specifically, although performance depended strongly on category separation, it did not depend on irrelevant variability (i.e., we found no difference in performance between the Long and Short conditions for either level of separation).

Some previous studies that found stimulus variability affects learning have attributed these effects to the hypothesis that high levels of variability on a stimulus dimension attract attention to that dimension (Apfelbaum & McMurray, 2011). For example, Pashler and Mozer (2013) hypothesized that:

Selective attention to a dimension may be based on the sheer amount of variability present along this dimension. According to this account, it is not the distance between examples on the relevant dimensions that helps but rather is just the magnitude of the variation on that dimension. This might be expected to occur if, for example, learners search for the relevant dimension by starting with dimensions showing the most salient variation. This account makes an interesting prediction that can be tested in future research: increasing variability within an irrelevant dimension should produce a very marked interference with learning. (p. 1171)

The results of Experiment 1 support this hypothesis, but the results of Experiment 2 do not. Even so, we believe there are several reasons that our results should not be interpreted as evidence against this interesting and reasonable hypothesis. First, our experiments were not designed to test this hypothesis. For example, the variability we included along irrelevant dimensions may not have been great enough, relative to the variability along relevant dimensions, for these hypothesized attentional effects to affect performance. Second, our use of trial-by-trial feedback may have masked attentional effects because participants who attended to the irrelevant dimension would have received immediate error feedback informing them that the irrelevant dimension was not diagnostic of category membership. Future

⁸Note that this is opposite to what a decision bound model seems to predict. If participants were estimating the slope and intercept of a linear decision bound, then estimation should be better in the Long LS condition than in the Short LS condition – not worse. This is because the Long LS condition provides more data points for estimation than the Short LS condition – that is, the rise and run of the optimal decision bound are both greater in the Long LS condition. Even so, note that decision bound models are parametric since they assume participants are estimating parameters of a specific parametric function (e.g., a line), and, as noted earlier, Ashby and Waldron (1999) reported evidence that II learning is nonparametric.

research is needed to disentangle the possible attentional effects of changing stimulus variability from the more purely difficulty-based effects.

Conclusions

In addition to clarifying the roles of linear separability and irrelevant variability on categorization difficulty, this article makes several other contributions. First, Experiments 1 and 2 document another novel and striking dissociation between RB and II category learning. II learning is impaired when irrelevant variability is increased, whereas RB learning is unaffected. This result adds to the growing list of 30+ empirical dissociations that have been documented between learning and performance in RB and II tasks (Ashby & Valentin, 2017). Collectively, these dissociations add to the already substantial evidence that RB and II tasks recruit qualitatively different decision strategies, and therefore require qualitatively different measures to predict learning difficulty.

Second, our results further validate the efficacy of the SDM. The SDM has one free parameter, γ , which represents a physical quality of the brain that combines the tuning of visual neurons and visual noise. Because of this physical interpretation, Rosedahl and Ashby (2019) predicted that γ should be constant across studies that used the same stimuli and amount of training. Our results support this prediction. We estimated the appropriate value of γ for the stimuli used in Experiment 1 by fitting the SDM to final block accuracies from 6 previous studies that used the same stimuli. The SDM accounted for 93% of the variance in accuracy across these studies, even though they were run in different labs and spanned more than a decade of time. Using this value of γ , the SDM accounted for 95% of the variance in asymptotic accuracy across the three conditions of Experiment 1, and 99% after excluding participants who did not use a procedural strategy. Furthermore, the SDM correctly predicted the nearly identical performance we observed in the Short LS and NLS conditions, and that the Long LS condition would be the most difficult of the three.

Acknowledgments

This research was supported by the National Science Foundation Graduate Research Fellowship under Grant No. 1650114 and by NIMH grant 2R01MH063760. The funders had no role in the conceptualization, design, data collection, analysis, decision to publish, or preparation of the manuscript. Thanks to A. Campbell, E. Kim, R. Serota, and T. Timsit for their assistance with data collection.

References

- Alfonso-Reese, L. A., Ashby, F. G., & Brainard, D. H. (2002). What makes a categorization task difficult? *Perception & Psychophysics*, *64*(4), 570–583.
- Apfelbaum, K. S., Hazeltine, E., & McMurray, B. (2013). Statistical learning in reading: Variability in irrelevant letters helps children learn phonics skills. *Developmental Psychology*, *49*(7), 1348.
- Apfelbaum, K. S., & McMurray, B. (2011). Using variability to guide dimensional weighting: Associative mechanisms in early word learning. *Cognitive Science*, *35*(6), 1105–1138.
- Ashby, F. G. (2019). *Statistical analysis of fMRI data, Second Edition*. Cambridge, MA: MIT Press.

- Ashby, F. G., & Alfonso-Reese, L. A. (1995). Categorization as probability density estimation. *Journal of Mathematical Psychology*, *39*(2), 216–233.
- Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *105*(3), 442–481.
- Ashby, F. G., & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 33–53.
- Ashby, F. G., & Rosedahl, L. A. (2017). A neural interpretation of exemplar theory. *Psychological Review*, *124*(4), 472–482.
- Ashby, F. G., Smith, J. D., & Rosedahl, L. A. (2019). Dissociations between rule-based and information-integration categorization are not caused by differences in task difficulty. *Memory & Cognition*, *48*, 541–552.
- Ashby, F. G., & Valentin, V. V. (2017). Multiple systems of perceptual category learning: Theory and cognitive tests. *Handbook of categorization in cognitive science (second edition)* (pp. 157–188). Elsevier.
- Ashby, F. G., & Valentin, V. V. (2018). The categorization experiment: Experimental design and data analysis. In E. J. Wagenmakers & J. T. Wixted (Eds.), *Stevens' handbook of experimental psychology and cognitive neuroscience, Fourth Edition, Volume Five: Methodology* (pp. 307–347). New York: Wiley.
- Ashby, F. G., & Waldron, E. M. (1999). On the nature of implicit categorization. *Psychonomic Bulletin & Review*, *6*(3), 363–378.
- Biederman, I., & Cooper, E. E. (1992). Size invariance in visual object priming. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(1), 121–133.
- Blair, M., & Homa, D. (2001). Expanding the search for a linear separability constraint on category learning. *Memory & Cognition*, *29*(8), 1153–1164.
- Cantwell, G., Crossley, M. J., & Ashby, F. G. (2015). Multiple stages of learning in perceptual categorization: Evidence and neurocomputational theory. *Psychonomic Bulletin & Review*, *22*, 1598–1613.
- Casale, M. B., & Ashby, F. G. (2008). A role for the perceptual representation memory system in category learning. *Attention, Perception, & Psychophysics*, *70*(6), 983–999.
- Cooper, L. A., Schacter, D. L., Ballesteros, S., & Moore, C. (1992). Priming and recognition of transformed three-dimensional objects: Effects of size and reflection. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(1), 43–57.
- Davis, T., Love, B. C., & Preston, A. R. (2012). Learning the exception to the rule: Model-based fMRI reveals specialized representations for surprising category members. *Cerebral Cortex*, *22*(2), 260–273.
- Ell, S. W., Cosley, B., & McCoy, S. K. (2011). When bad stress goes good: Increased threat reactivity predicts improved category learning performance. *Psychonomic bulletin & review*, *18*(1), 96–102.
- Feldman, J. (2000). Minimization of Boolean complexity in human concept learning. *Nature*, *407*(6804), 630.
- Feldman, J. (2004). How surprising is a simple pattern? Quantifying Eureka! *Cognition*, *93*(3), 199–224.

- Filoteo, J. V., Maddox, W. T., Ing, A. D., & Song, D. D. (2007). Characterizing rule-based category learning deficits in patients with parkinson's disease. *Neuropsychologia*, *45*(2), 305–320.
- Freedberg, M., Glass, B., Filoteo, J. V., Hazeltine, E., & Maddox, W. T. (2017). Comparing the effects of positive and negative feedback in information-integration category learning. *Memory & Cognition*, *45*(1), 12–25.
- Homa, D., Sterling, S., & Trepel, L. (1981). Limitations of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Human Learning and Memory*, *7*(6), 418–439.
- Huet, M., Jacobs, D. M., Camachon, C., Missenard, O., Gray, R., & Montagne, G. (2011). The education of attention as explanation of variability of practice effects: Learning the final approach phase in a flight simulator. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(6), 1841.
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795.
- Kerr, R., & Booth, B. (1978). Specific and varied practice of motor skill. *Perceptual and Motor Skills*, *46*(2), 395–401.
- Levering, K. R., Conaway, N., & Kurtz, K. J. (2019). Revisiting the linear separability constraint: New implications for theories of human category learning. *Memory & Cognition*, 1–13.
- Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America*, *94*(3), 1242–1255.
- Maddox, W. T., & Ashby, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception & Psychophysics*, *53*(1), 49–70.
- 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*(6), 710–727.
- Maddox, W. T., Ing, A. D., & Lauritzen, J. S. (2006). Stimulus modality interacts with category structure in perceptual category learning. *Perception & Psychophysics*, *68*(7), 1176–1190.
- Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*(3), 207–238.
- Medin, D. L., & Schwanenflugel, P. J. (1981). Linear separability in classification learning. *Journal of Experimental Psychology: Human Learning and Memory*, *7*(5), 355–368.
- Nomura, E., & Reber, P. J. (2008). A review of medial temporal lobe and caudate contributions to visual category learning. *Neuroscience & Biobehavioral Reviews*, *32*(2), 279–291.
- Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, *115*(1), 39–57.
- Nosofsky, R. M., Gluck, M. A., Palmeri, T. J., McKinley, S. C., & Glauthier, P. (1994). Comparing models of rule-based classification learning: A replication and extension of Shepard, Hovland, and Jenkins (1961). *Memory & Cognition*, *22*(3), 352–369.

- Odlyzko, A. M., & Sloane, N. J. (1979). New bounds on the number of unit spheres that can touch a unit sphere in n dimensions. *Journal of Combinatorial Theory, Series A*, *26*(2), 210–214.
- Pashler, H., & Mozer, M. C. (2013). When does fading enhance perceptual category learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(4), 1162.
- Patalano, A. L., Smith, E. E., Jonides, J., & Koeppe, R. A. (2001). PET evidence for multiple strategies of categorization. *Cognitive, Affective, & Behavioral Neuroscience*, *1*(4), 360–370.
- Paul, E. J., Boomer, J., Smith, J. D., & Ashby, F. G. (2011). Information–integration category learning and the human uncertainty response. *Memory & Cognition*, *39*(3), 536–554.
- Perrachione, T. K., Lee, J., Ha, L. Y., & Wong, P. C. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America*, *130*(1), 461–472.
- Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, *77*(3p1), 353–363.
- Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, *25*, 111–163.
- Reber, P. J., Gitelman, D. R., Parrish, T. B., & Mesulam, M. M. (2003). Dissociating explicit and implicit category knowledge with fMRI. *Journal of Cognitive Neuroscience*, *15*(4), 574–583.
- Reber, P. J., & Squire, L. R. (1999). Intact learning of artificial grammars and intact category learning by patients with Parkinson’s disease. *Behavioral Neuroscience*, *113*(2), 235–242.
- Rosedahl, L. A., & Ashby, F. G. (2019). A difficulty predictor for perceptual category learning. *Journal of Vision*, *19*(6), 20–20.
- Rosedahl, L. A., Eckstein, M. P., & Ashby, F. G. (2018). Retinal-specific category learning. *Nature Human Behaviour*, *2*(7), 500.
- Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237.
- Seamon, J. G., Ganor-Stern, D., Crowley, M. J., Wilson, S. M., Weber, W. J., O’Rourke, C. M., & Mahoney, J. K. (1997). A mere exposure effect for transformed three-dimensional objects: Effects of reflection, size, or color changes on affect and recognition. *Memory & Cognition*, *25*(3), 367–374.
- Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs: General and Applied*, *75*(13), 1–42.
- Smith, J. D., Beran, M. J., Crossley, M. J., Boomer, J., & Ashby, F. G. (2010). Implicit and explicit category learning by macaques (*Macaca mulatta*) and humans (*Homo sapiens*). *Journal of Experimental Psychology: Animal Behavior Processes*, *36*(1), 54–65.

- Smith, J. D., & Minda, J. P. (2002). Distinguishing prototype-based and exemplar-based processes in dot-pattern category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 800–811.
- Smith, J. D., Murray Jr, M. J., & Minda, J. P. (1997). Straight talk about linear separability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(3), 659–680.
- Spiering, B. J., & Ashby, F. G. (2008a). Initial training with difficult items facilitates information integration, but not rule-based category learning. *Psychological Science*, *19*(11), 1169–1177.
- Spiering, B. J., & Ashby, F. G. (2008b). Response processes in information–integration category learning. *Neurobiology of Learning and Memory*, *90*(2), 330–338.
- Vandist, K., De Schryver, M., & Rosseel, Y. (2009). Semisupervised category learning: The impact of feedback in learning the information-integration task. *Attention, Perception, & Psychophysics*, *71*(2), 328–341.
- Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, *14*(5), 779–804.
- Wattenmaker, W. D., Dewey, G. I., Murphy, T. D., & Medin, D. L. (1986). Linear separability and concept learning: Context, relational properties, and concept naturalness. *Cognitive Psychology*, *18*(2), 158–194.
- Wattenmaker, W. D. (1995). Knowledge structures and linear separability: Integrating information in object and social categorization. *Cognitive Psychology*, *28*(3), 274–328.
- Wiggs, C. L., & Martin, A. (1998). Properties and mechanisms of perceptual priming. *Current Opinion in Neurobiology*, *8*(2), 227–233.
- Zeithamova, D., & Maddox, W. T. (2009). Learning mode and exemplar sequencing in unsupervised category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(3), 731–741.