The impact of category structure and training methodology on learning and generalizing within-category representations

Shawn W. Ell[1] · David B. Smith[2] · Gabriela Peralta[2] · Sébastien Hélie[3]

Published online: 5 June 2017

Abstract When interacting with categories, representations focused on within-category relationships are often learned, but the conditions promoting within-category representations and their generalizability are unclear. We report the results of three experiments investigating the impact of category structure and training methodology on the learning and generalization of within-category representations (i.e., correlational structure). Participants were trained on either rule-based or information-integration structures using classification (Is the stimulus a member of Category A or Category B?), concept (e.g., Is the stimulus a member of Category A, Yes or No?), or inference (infer the missing component of the stimulus from a given category) and then tested on either an inference task (Experiments 1 and 2) or a classification task (Experiment 3). For the information-integration structure, within-category representations were consistently learned, could be generalized to novel stimuli, and could be generalized to support inference at test. For the rule-based structure, extended inference training resulted in generalization to novel stimuli (Experiment 2) and inference training resulted in generalization to classification (Experiment 3). These data help to clarify the conditions under which within-category representations can be learned. Moreover, these results make an important contribution in highlighting the impact of category structure and training methodology on the generalization of categorical knowledge.

Keywords Knowledge representation · Training methodology · Generalization · Category learning

The ability to learn categorical representations is foundational for cognition. Categories enable the navigation of familiar situations with increasing efficiency and can also be generalized to facilitate function in novel situations. Not surprisingly, much research has been dedicated to understanding categorical representations and how they are learned. This research has been fertile ground for a vigorous and healthy debate regarding the nature of category representations. Throughout this debate, most research groups advocating for one theory or another have tended to focus on a single paradigm, suggesting that some theoretical disagreements may be driven by methodological differences. This article investigates the impact of two methodological variants on the learning of category representations focusing on within-category similarities. Namely, how does variability in the structure of the categories, and the training methodology that dictates how participants interact with the to-be-learned information, impact within-category representations? Furthermore, once learned, what are some of the limits on the generalization of within-category representations?

Category representations

Category representations are largely dependent upon the goal of the task (Goldstone, 1996; Hoffman & Rehder, 2010; Markman & Ross, 2003; Minda & Ross, 2004; Yamauchi & Markman, 1998). For instance, in the typical category learning

✉ Shawn W. Ell
shawn.ell@umit.maine.edu

[1] Department of Psychology, Graduate School of Biomedical Sciences and Engineering, University of Maine, 5742 Little Hall, Room 301, Orono, ME 04469-5742, USA

[2] Department of Psychology, University of Maine, Orono, ME, USA

[3] Department of Psychological Sciences, Purdue University, West Lafayette, IN, USA

experiment, participants are presented with stimuli (each drawn from one of a number of contrasting categories) and instructed to make a decision about the category membership of each stimulus. Such classification instructions have often been argued to lead to the development of a representation that focuses on between-category differences (e.g., learn what dimensions are relevant for classification, along with decision criteria or category boundaries; Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Erickson & Kruschke, 1998; Maddox & Ashby, 1993; Nosofsky, Palmeri, & McKinley, 1994; Smith & Minda, 2002). In a slightly different paradigm, participants are presented with a subset of the stimulus features as well as a category label and instructed to infer the missing feature. Such inference instructions lead to the development of a category representation that focuses on within-category similarities (e.g., the correlational structure of the stimulus dimensions; Chin-Parker & Ross, 2002; Markman & Ross, 2003). Thus, the goal of classifying the stimuli into one of a number of contrasting categories may lead to a between-category representation, whereas the goal of inferring missing information for stimuli from a known category may lead to a within-category representation.

Task goal is clearly an important factor, but it is not the only factor in producing within-category representations. For instance, observational training (Carvalho & Goldstone, 2015; Levering & Kurtz, 2015), training emphasizing the comparison of members from the same category (Hammer, Diesendruck, Weinshall, & Hochstein, 2009), and blocked training (Carvalho & Goldstone, 2014; Goldstone, 1996) can promote within-category representations. Another factor that is investigated in the present article involves a seemingly minor tweak of the typical classification instructions to emphasize concept learning called the yes/no task (i.e., participants learn categories by classifying stimuli as a member/nonmember of a target category; Maddox, Bohil, & Ing, 2004; Posner & Keele, 1968; Reber, 1998; Smith & Minda, 2002; Zeithamova, Maddox, & Schnyer, 2008). Both classification and concept training are active tasks and have the goal of classification on a trial-by-trial basis. Concept training, however, has been argued to shift the emphasis from between-category differences to within-category similarities (Casale & Ashby, 2008; Hélie, Shamloo, & Ell, 2017).

The very structure of the categories themselves can influence category representations (Ashby et al., 1998; Carvalho & Goldstone, 2014). Consider, for example, the distinction between rule-based (RB) and information-integration (II) category structures (Ashby & Ell, 2001). RB structures can be learned using logical rules. Although logical rules can be based on either within- or between-category representations (e.g., large or larger than), the subset of logical rules learned with RB structures tends to depend upon between-category representations (Casale, Roeder, & Ashby, 2012; Ell & Ashby, 2012; Ell, Ing, & Maddox, 2009; Hélie et al., 2017). In contrast, II structures are those in which information from multiple dimensions needs to be integrated prior to making a categorization response. Unlike RB structures, II structures generally promote within-category representations (Ashby & Waldron, 1999; Hélie et al., 2017; Thomas, 1998). Again, even when classification is the goal, RB structures would be expected to promote between-category representations, whereas II structures would be expected to promote within-category representations. Neurocomputational models that have been applied to RB and II structures implicitly echo this between- versus within-category distinction (Ashby et al., 1998; Ashby & Crossley, 2011).

Utility of within- and between-category representations

Categorical representations in and of themselves have little value. Rather, it is the efficiencies afforded by categories that are a better measure of their cognitive utility (Hoffman & Rehder, 2010; Markman & Ross, 2003). Category representations can facilitate interactions with category members (e.g., Rosch & Mervis, 1975). Arguably more important is that category representations can also facilitate interactions with novel stimuli. Indeed, the field has a well-established tradition of probing the extent to which learned category representations can support the classification of novel stimuli (e.g., Smith & Minda, 1998). Clearly this is an important function of category representations.

Importantly, we argue that the generalizability of category representations depends upon the nature of the representation itself (Carvalho & Goldstone, 2014; Hoffman & Rehder, 2010; Levering & Kurtz, 2015). For instance, between-category representations may be better suited to generalize to novel stimuli that are beyond the range of the previously encountered stimuli (e.g., because the representation is not tied to the stimuli themselves but rather between-category differences; Casale et al., 2012; Hoffman & Rehder, 2010; Maddox, Filoteo, Lauritzen, Connally, & Hejl, 2005). Similarly, within-category representations may be better suited for generalization that would benefit from knowledge of within-category regularities, such as prototypicality or the covariation of stimulus dimensions (Chin-Parker & Ross, 2002, 2004; Yamauchi & Markman, 1998).

The ability to generalize between-category representations, however, may be task dependent. Although knowledge of between-category differences would facilitate classification of novel stimuli, such knowledge is inextricably tied to the goal of classification. When successful generalization depends upon the ability to reconfigure knowledge acquired during training to solve a new decision-making problem, within-category representations would seem to have far greater utility than between-category representations. Indeed, within-category representations can support generalization to novel tasks (Chin-Parker & Ross, 2002). Within-category representations are also better able than between-category representations to support the reconfiguration of categorical knowledge (Hélie et al., 2017; Hoffman & Rehder, 2010).

Formal models of categorization have been successful in accounting for generalization of learned representations to support the classification of novel stimuli. Attempts to test the ability of formal models to account for generalization to novel tasks, however, are not as common (see Maddox & Bogdanov, 2000; Nosofsky & Zaki, 1998; Smith & Minda, 2001, for notable examples). Thus, the approach taken in the current study—investigating generalization to novel stimuli and tasks—will provide an important test bed for the development and testing of formal models.

The current study

Although participants often demonstrate an initial bias toward between-category representations (Ashby, Queller, & Berretty, 1999; Ell & Ashby, 2006; Medin, Wattenmaker, & Hampson, 1987; Smith, Beran, Crossley, Boomer, & Ashby, 2010), within-category representations may be a common outcome of interacting with categories (e.g., Anderson & Fincham, 1996; Hélie et al., 2017; Hoffman & Rehder, 2010; Thomas, 1998). Previous work suggests that numerous methodological factors can promote within-category representations, but there is variability in how within-category representations were measured, if measured at all. For example, some studies used a two-alternative, forced-choice procedure (e.g., Hoffman & Rehder, 2010) while others asked for typicality ratings (e.g., Levering & Kurtz, 2015).

Studies using inference training consistently demonstrate the development of within-category representations but have not given much attention to the impact of category structure. For example, when trained by inference, participants learn the correlational structure of the categories despite such information possibly being irrelevant to category membership (e.g., Chin-Parker & Ross, 2002). Motivated by this work, we employ knowledge of correlational structure as our primary dependent measure of within-category representation and extend this work by considering variability in training methodology and category structure.

Using a transfer task that required the reconfiguration of within-category representations, Hélie and colleagues (2017) showed that learning an II structure resulted in successful transfer with both concept and classification training. In contrast, learning a RB structure resulted in successful transfer with concept training, but not classification training. Although these data are consistent with the claim that within-category representations may be a more common outcome of categorization, this claim would be bolstered by using a more traditional measure of within-category representations (i.e., knowledge of the correlational structure).

A second goal of the current study is to investigate the extent to which within-category representations can be generalized to support performance with novel stimuli and/or novel tasks. Within-category representations developed with inference training appear to be quite versatile and can support generalization to novel tasks (Chin-Parker & Ross, 2002). Although the within-category representations developed with concept and classification training can support knowledge reconfiguration (Hélie et al., 2017), it is unclear if these within-category representations can also support generalization to novel stimuli and tasks. For example, some researchers have demonstrated that within-category correlations can be learned during a classification task (Anderson & Fincham, 1996; Thomas, 1998), whereas others have argued that such demonstrations are a byproduct of simplistic stimuli, overtraining, and/or classification tasks that incorporate additional inference-like training (e.g., Chin-Parker & Ross, 2002).

The current study tests these hypotheses using classification, concept, and inference training methodologies to learn RB and II structures. For classification training, participants were instructed to distinguish between members of contrasting categories (e.g., Is the image a member of Category A or Category B?—hereafter referred to as A/B training). For concept training, participants were instructed to distinguish between category members and nonmembers (e.g., Is the image a member of Category A?—hereafter referred to as YES/NO training; Hélie et al., 2017; Maddox, Bohil, et al., 2004). For inference training, participants were instructed to produce the missing stimulus feature given the category label and another stimulus feature (hereafter referred to as INF training; Chin-Parker & Ross, 2002; Thomas, 1998; Zotov, Jones, & Mewhort, 2011).

In Experiment 1, participants learned RB or II structures that incorporated a correlation between the stimulus dimensions using either A/B, YES/NO, or INF training. Knowledge of the correlation between the stimulus dimensions was subsequently tested using inference. The test phase included stimuli that were consistent with the training categories (allowing for the assessment of within-category representations developed during training) and novel stimuli (allowing for the assessment of generalization of within-category representations beyond the trained stimuli). Importantly, the design also enabled an analysis of the extent to which knowledge could be generalized across methodologies (e.g., from classification to inference).

Following Hélie et al. (2017), we hypothesized that within-category representations would be learned in all but the RB-A/B condition, and that within-category representations could be generalized to support inference across stimuli and methodologies. Experiments 2 and 3 were designed to replicate and extend Experiment 1. Experiment 2 investigated the impact of extended training on the ability to generalize within-category representations. Experiment 3 aimed to investigate if the generalization results were specific to using an inference procedure at test by testing participants on A/B classification rather than inference.

To anticipate, the results of Experiments 1 and 2 demonstrate that the II structure consistently resulted in within-category representations that could be generalized to novel stimuli and across methodologies (Experiments 1 and 2). The RB structure, however, resulted in within-category representations only when paired with INF training (Experiments 1 and 2). The within-category representations acquired in the RB-INF condition could be generalized to novel stimuli and across methodologies, but only when provided with extended INF training (Experiment 2). Furthermore, generalization across methodologies was asymmetric as the within-category representations acquired with INF training could be generalized to support A/B classification, but only with the RB structure (Experiment 3).

Experiment 1

The goals of Experiment 1 were twofold. First, Experiment 1 investigated the extent to which training methodology and category structure promotes the learning of within-category representations. Second, Experiment 1 investigated the ability of within-category representations to support knowledge generalization across stimuli and tasks. Specifically, participants were trained on either RB or II category structures using classification training (A/B), concept training (YES/NO), or inference training (INF). The stimulus dimensions were correlated within each category, thereby allowing the use of knowledge of the correlational structure of the categories as a probe for within-category representations. All participants were subsequently tested using an inference procedure that included exemplars from the training categories as well as novel exemplars. Knowledge of the within-category correlations for training exemplars indexed learning of the within-category representations whereas knowledge of the within-category correlations for transfer exemplars indexed generalization to novel stimuli. Successful test performance for participants in the A/B and YES/NO conditions provided a measure of generalization across methodologies. It was predicted that all but the RB-A/B condition would evidence within-category representations and that these representations would be able to be generalized across stimuli and methodology.

Method

Participants and design

In all experiments, a target sample size of approximately 30 participants in each experimental condition was determined a priori (based upon previous experience with similar experiments). Participants (193 total) were recruited from the University of Maine community and received partial course credit for participation. Participants were randomly assigned to one of six experimental conditions in the 2 category structure (RB vs. II) × 3 training methodology (A/B, YES/NO, INF) design. A total of seven participants were excluded from analysis: two participants due to a software error (RB-INF: 1; II-INF: 1), three participants did not complete the task within the hour-long experimental session (II-AB: 1; II- YES/NO: 1; II-INF: 1), and two participants were statistical outliers (i.e., more than three standard deviations from the mean on both average training accuracy and accuracy during the final training block; RB-YES/NO: 2). The resulting sample sizes by condition were RB-A/B: 32; RB-YES/NO: 29; RB-INF: 32; II-A/B: 32; II- YES/NO: 30; II-INF: 31. All participants reported normal (20/20) or corrected-to-normal vision. Each participant completed one session of approximately 60 minutes duration.

Stimuli and apparatus

The stimuli in all experiments comprised circles and lines that varied continuously in diameter and orientation, respectively (see Fig. 1). These dimensions were selected in an effort to facilitate the ability of participants to complete the inference task. The training categories were generated using a variation of the randomization technique introduced by Ashby and Gott (1988), in which the stimuli were generated by sampling from bivariate normal distributions defined in a Diameter × Angle (from horizontal) space in arbitrary units. For the II structure, the category means were $\mu_A = [485, -20]$ and $\mu_B = [415, 40]$. For the RB structure, the category means were $\mu_A = [635, -20]$ and $\mu_B = [265, 40]$. The covariance matrix $\Sigma = \begin{bmatrix} 3125 & 2875 \\ 2875 & 3175 \end{bmatrix}$ (i.e., a correlation of 1 between diameter and angle) was the same for all tasks and categories. Recall that the primary dependent measure of within-category knowledge was the extent to which participants learned the diameter-angle correlation. As a consequence, it was necessary to have a nonzero covariance within each category and to increase the category separation in the RB task in order to allow for a unidimensional rule on diameter to produce optimal accuracy.

On each trial, a random sample (x, y) was drawn from the Category A or B distribution, and these values were used to construct a stimulus with circle of $\frac{x}{2}$ pixels in diameter and line of $\frac{180y}{800}$ degrees (counterclockwise from horizontal) with length of 200 pixels. The line was always connected at the circle's highest point. The scaling factors were selected in an effort to equate the perceived salience of the stimulus dimensions. Eighty stimuli (40 from each category) were generated for

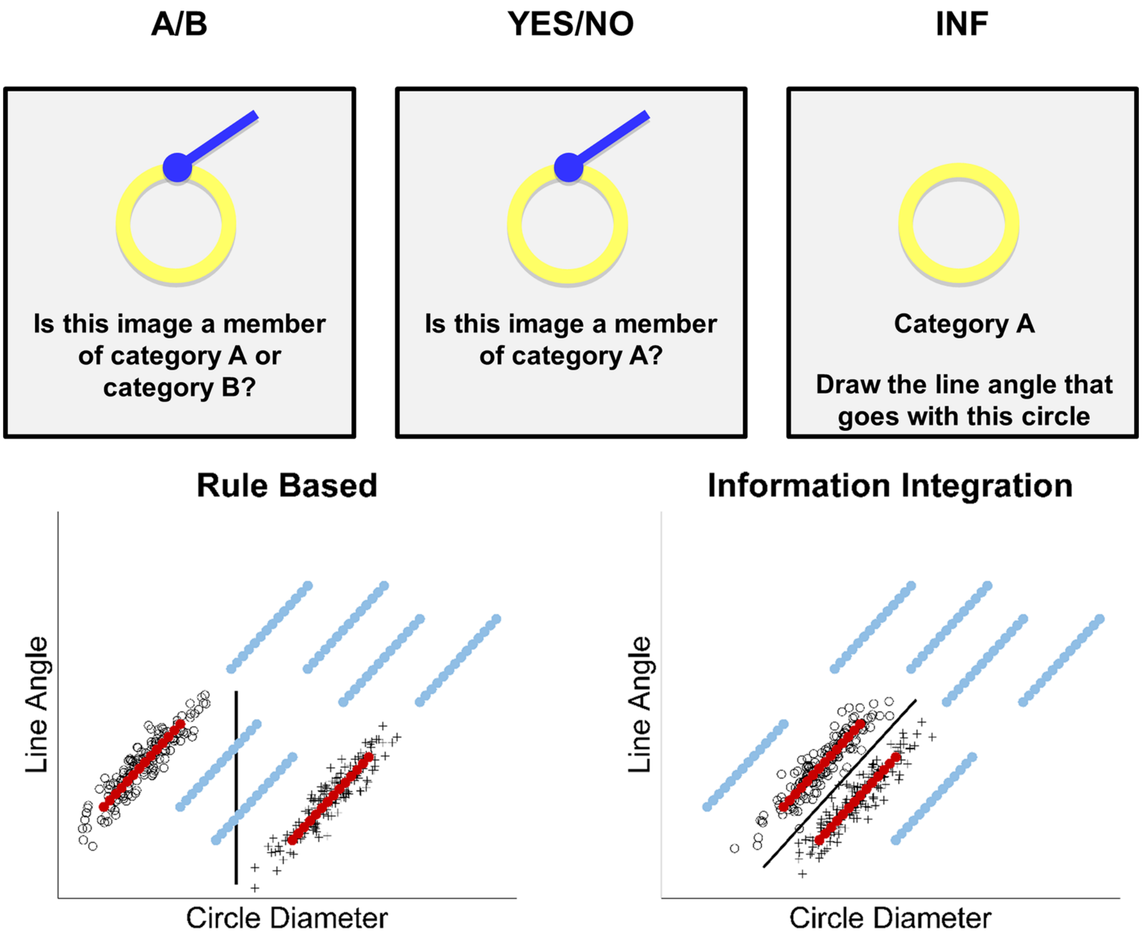

Fig. 1 (*Top*) Example displays for three training methodologies. (*Bottom*) RB and II category structures. Category A and B exemplars used during the training phase are plotted as *black crosses* and *circles*, respectively. Stimuli used during the test phase are plotted as red/dark (training set) and blue/light (transfer set) circles. (Color figure online)

each of the four blocks of trials. All stimuli were generated off-line, and a linear transformation was applied to ensure that the sample statistics matched the population parameters. The experiment was run using the Psychophysics Toolbox (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997) in the MATLAB computing environment. Each stimulus was displayed on a 20-inch LCD with 1600 × 1200 pixel resolution at a viewing distance of 20 inches in a dimly lit room.

Two sets of test phase stimuli (112 total) were selected to assess the learning and generalization of the within-category correlations. The training set was selected to approximate the training categories and was used to assess learning of the within-category correlations (red circles in Fig. 1). The transfer set was selected to broadly sample the untrained region of the stimulus space while maintaining the within-category correlation from the training categories and was used to assess generalization of the within-category correlations (blue circles in Fig. 1). The coordinates of the test phase stimuli are presented in Appendix 1.

Consistent with previous work, participants were expected to learn unidimensional rules in the RB task (Ell & Ashby, 2006). Given the large category separation, however, there are many alternative strategies that would also yield perfect performance (e.g., the optimal strategy for the II task). Thus, probe stimuli were included to differentiate between unidimensional and integration strategies (e.g., the solid lines in Fig. 1). A subset (14) of the test stimuli that lie between Category A and B were included as probe stimuli during the final block of training (resulting in a total of 94 trials during the final block). In an effort to increase the similarity between the RB and II conditions, these same probe stimuli were also included during the final block with the II structure. Because the probe stimuli do not aid in the identifiability of the decision strategy used with the II structure, the probe stimuli were excluded from the analysis of the II training data. No feedback was provided for probe trials. The coordinates of the probe stimuli are presented in Appendix 1.

Procedure

Each participant was run individually. At the beginning of the training phase, participants were told that stimuli would comprise a circle with a line connected at the top and that the stimuli would be presented individually but would vary across trials in circle diameter and line angle. In the A/B condition, participants were instructed that their goal was to learn, by trial and error, to distinguish between members of Category A and B. On each trial, a stimulus was presented, and participants were prompted "Is this image a member of Category A or Category B?" and responded by pressing the button labeled "A" or "B" on the keyboard. In the YES/NO condition, participants were instructed that their goal was to learn, by trial and error, if each image is a member of a particular category or not. On each trial, a stimulus was presented and participants were prompted with either "Is this image a member of Category A?" or "Is this image a member of Category B?" (with equal probability) and responded by pressing the button labeled "Yes" or "No" on the keyboard. In the INF condition, participants were instructed that their goal was to learn, by trial and error, to draw the missing stimulus component. Example stimulus displays are shown in Fig. 1. On each trial, a partial stimulus (i.e., line or circle along with the category label) was presented and participants were prompted to draw the missing component—that is, "Draw the circle that goes with this line angle" or "Draw the line angle that goes with this circle" (with equal probability). Participants initially responded by using the mouse to select the location of either the bottom of the circle (indicating the diameter of the circle relative to the dot at the beginning of the line) or the end of the line (indicating the orientation of the line relative to horizontal). The circle or line was drawn by the computer based upon the participant's selection with a line beginning at the dot at the top of the circle (at a constant length of 200 pixels). After the line was drawn, participants were able to adjust the diameter or angle using the arrow keys, pressing the space bar when satisfied. Any selected stimulus values outside the allowable range were reset to the nearest allowable value (allowable range: diameter 10 to 600 pixels, angle: -50 to 110 degrees).

Stimulus presentation was response terminated with an upper limit of 60 s. After responding, feedback was provided. In the A/B and YES/NO conditions, the screen was blanked and the word "CORRECT" (in green, accompanied by a 500 Hz tone) or "WRONG" (in red, accompanied by a 200 Hz tone) was displayed. In the INF condition, the correct circle or line was overlaid upon the participant's response. In all conditions, feedback duration was 2 s and the screen was then blanked for 1 s prior to the appearance of the next stimulus.

In addition to trial-by-trial feedback, summary feedback was given at the end of each 80-trial block, indicating percentage correct for that block (A/B and YES/NO, participants were informed that higher numbers are better) or the root mean square error between the drawn and correct stimulus components (INF, participants were informed that lower numbers are better). The presentation order of the stimuli was randomized within each block, separately for each participant. Prior to starting the training phase, participants completed several practice trials to familiarize themselves with the task using stimuli randomly sampled (with equal probability) from the training categories.

During the test phase, all participants performed the inference task (one block of 112 trials). Instruction was provided for all conditions and participants completed several practice trials prior to beginning the test phase using stimuli randomly sampled (with equal probability) from all test phase stimuli. No feedback was provided during the test phase.

Results

Training phase

The dependent measure varied across training methods, thus the training phase data from the A/B, YES/NO, and INF conditions were analyzed separately. Performance generally improved across blocks for all training methodologies (Fig. 2). A 2 category structure × 4 block mixed ANOVA conducted on the data from the A/B condition revealed significant main effects, structure: $F(1, 62) = 282.14, p < .05, \eta_p^2 = .82$; block: $F(2.62, 162.41) = 27.37, p < .05, \eta_p^2 = .31$, and a significant interaction, $F(2.62, 162.41) = 4.16, p < .05, \eta_p^2 = .06$.[1] To decompose the interaction, a series of pairwise comparisons were conducted within each structure. For the II structure, accuracy increased across the first three blocks (*p*s < .05), but not from Block 3 to Block 4 (*p* = .80). For the RB structure, there was no significant block-over-block increase in accuracy (*p*s > .10), but there was a more general increase with Block 4 accuracy being higher than Block 1 (*p* < .05). These results suggest that with A/B training, there was more consistent improvement across blocks in the II structure, but caution is warranted given a possible ceiling effect in the RB structure.

A 2 category structure × 4 block mixed ANOVA conducted on the data from the YES/NO condition revealed significant main effects, structure: $F(1, 57) = 261.62, p < .05, \eta_p^2 = .82$; block: $F(2.14, 122.06) = 46.97, p < .05, \eta_p^2 = .45$, but the interaction was not significant, $F(2.14, 122.06) = 2.85, p = .06, \eta_p^2 = .05$. Pairwise comparisons indicated that the main effect of block was driven by an increase in accuracy across the first three blocks (*p*'s < .05), but not from Block 3 to Block 4 (*p* = .63). With YES/NO training, participants in both structures learned, but accuracy was higher in the RB structure.

To analyze the data from the INF condition, the correlation between the presented and produced dimensions was computed separately for each category, then averaged across categories (Fig. 2, right panel). A 2 category structure × 4 block

[1] A Huynh-Feldt correction for violation of the sphericity assumption has been applied to all mixed ANOVAs (when appropriate). All post-hoc comparisons have been Šidák corrected unless otherwise noted.

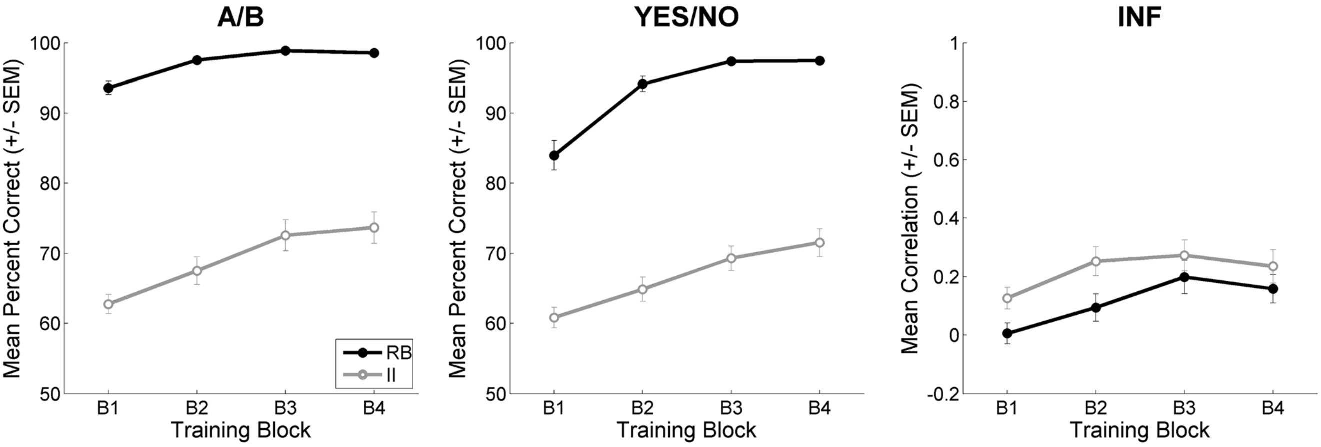

Fig. 2 Training performance in the A/B, YES/NO, and INF conditions of Experiment 1

mixed ANOVA indicated a significant effect of block, F(2.89, 176.22) = 16.26, p < .05, η_p^2 = .21, with consistent improvement across Blocks 1–3 [*p*'s < .05. No other effects were statistically significant, category structure: $F(1, 61) = 3.06$, $p = .08$, $\eta_p^2 = .05$; Category Structure × Block: $F(2.89, 176.22) = 1.23$, $p = .3$, $\eta_p^2 = .02$. In sum, participants in the two inference training conditions evidenced learning of the within-category correlation although this learning was modest, only being statistically greater than zero in Blocks 2–4 (*p*'s < .05) and asymptoting near a correlation of .2.

Categorization performance in the RB task was expected to be mediated by unidimensional decision strategies (Ell & Ashby, 2006), but given the large separation between the RB categories, a number of qualitatively different decision strategies could have produced high accuracy. In order to confirm that participants were using unidimensional strategies in the RB task, a number of decision-bound models (Ashby, 1992a; Maddox & Ashby, 1993) were fit to the individual participant data from the A/B and Yes/No conditions. Three different types of models were evaluated, each based on a different assumption concerning the participant's strategy. Rule-based models assume that the participant sets decision criteria on one (or both) stimulus dimensions (e.g., unidimensional model: If the circle is large, respond A; otherwise respond B). Information-integration models assume that the participant integrates the stimulus information from both dimensions prior to making a categorization decision. Finally, random responder models assume that the participant guessed. Each of these models were fit separately to the data from the final block, for each participant, using a standard maximum likelihood procedure for parameter estimation (Ashby, 1992b; Wickens, 1982) and the Bayes information criterion for goodness of fit (Schwarz, 1978; see Appendix 2 for a more detailed description of the models and fitting procedure).

As expected, most participants in the RB task were best fit by a unidimensional model assuming participants attended selectively to diameter (A/B: 91%, YES/NO: 86%). Similarly, most participants in the II task were best fit by information-integration models (A/B: 63%, YES/NO: 67%). The results of the model-based analysis indicate that the majority of participants used task appropriate strategies at the end of training.

Test phase

Correlations between the presented and produced dimensions were computed separately for each cluster of test stimuli in Fig. 1. Preliminary analyses were conducted on the correlations to determine if the data could be safely aggregated across clusters. For test phase data from the two training clusters, a 2 cluster × 2 category structure × 3 training methodology mixed ANOVA did not reveal any significant effects of cluster (main effect and interactions: all $F < 1, p \geq .38, \eta_p^2 \leq .01$). Similarly, for test phase data from the six transfer clusters, a 6 cluster × 2 category structure × 3 training methodology mixed ANOVA did not reveal any significant effects of cluster (main effect and interactions: all $F \leq 1.6, p \geq .1, \eta_p^2 \leq .018$). Thus, the subsequent analyses average across clusters within the two sets of test stimuli (i.e., training and transfer).

Inspection of the correlations during the test phase (see Fig. 3) suggests more consistent learning and generalization of the correlational structure of the training categories for the II structure. A series of one-sample *t* tests (see Table 1) were consistent with this observation. For the RB structure, neither the correlations for the training stimuli nor the transfer stimuli were significantly greater than zero. In contrast, for the II structure, almost all of the correlations were significantly greater than zero, with the correlation for training items in the YES/NO condition not surviving the correction for multiple comparisons. Consistent with the previous analysis, a 2 stimulus set (training, transfer) × 2 category structure × 3 training methodology mixed ANOVA comparing the magnitude of the

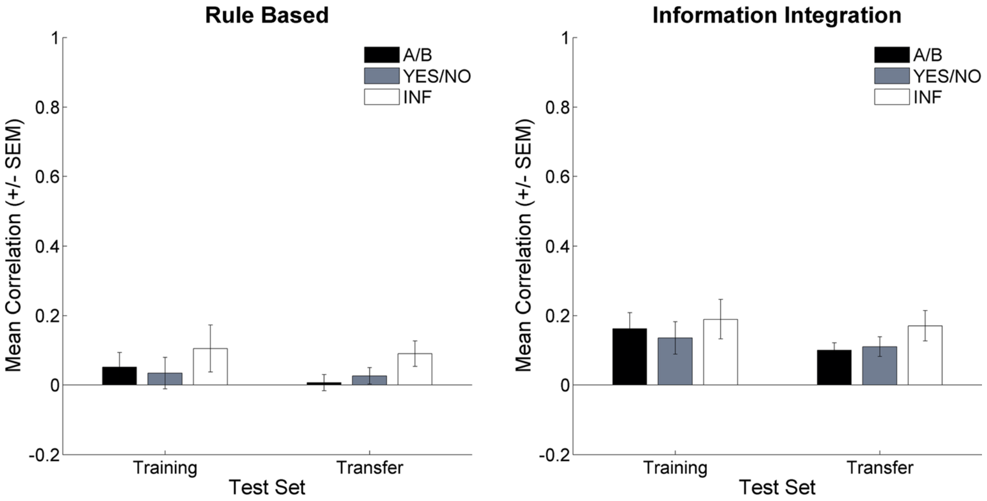

Fig. 3 Performance on the inference task used during the test phase of Experiment 1

correlation across conditions indicated only a significant main effect of category structure, $F(1, 180) = 9.98$, $p < .05$, $\eta_p^2 = .05$. None of the other effects were statistically significant (all $F \leq 2.43$, $p \geq .12$, $\eta_p^2 \leq .02$). In sum, these data suggest learning and generalization of the within-category correlations, but only for the II category structure.

Summary

The goal of Experiment 1 was to investigate the impact of category structure and training methodology on the ability to learn and generalize within-category representations (i.e., correlational structure of the categories). Structure and methodology were predicted to interact such that within-category representations would be learned in all but the RB-A/B condition. The results, however, did not support these predictions. First, although participants demonstrated some evidence of learning within-category representations during the training phase of the INF condition, this information was only significantly maintained for the II structure. That being said, there may have been learning in the RB-INF condition that did not survive the statistical correction for multiple comparisons given the small-to-moderate effect sizes for the training and transfer stimuli during the test phase. Second, YES/NO training did not generally result in the learning of within-category representations. Instead, the results suggest that the II structure consistently resulted in the learning of within-category representations, regardless of training methodology. Moreover, the within-category representations could be generalized to a novel task (i.e., from categorization to inference) and to novel stimuli.

Table 1 Knowledge of the within-category correlational structure during the test phase of Experiment 1

		Training stimuli			Transfer stimuli		
Rule based	*df*	*t*	*p*	*d*	*t*	*p*	*d*
A/B	31	1.23	.11	.22	.3	.38	.05
YES/NO	28	.76	.23	.14	1.1	.14	.2
INF	31	1.55	.07	.27	2.45	.01	.43
Information integration							
A/B	31	3.61*	.0005	.64	4.62*	.00003	.82
YES/NO	29	2.91	.007	.53	3.91*	.0002	.71
INF	30	3.34*	.001	.6	3.93*	.0002	.71

Note. One-sample *t* tests comparing observed correlations to no learning (i.e., correlation = 0), *df* are constant across training and transfer. *Statistically significant at Šidák corrected $\alpha = .004$

Experiment 2

The results of Experiment 1 suggest that learning and generalization of within-category representations may be limited to II category structures. The inference task, however, was fairly challenging. Thus, it may be that there would be more robust evidence of within-category knowledge with extended training on the inference task. In addition, providing extended training may also provide more of an opportunity for participants given categorization training to learn the within-category representations. The goal of Experiment 2 was to investigate the impact of extended training on the ability to learn and generalize within-category representations. The design of Experiment 2 was identical to Experiment 1 with two exceptions. First, the amount of training was doubled (across two training sessions). Second, given the similarity of the results in the A/B and YES/NO conditions of Experiment 1, only A/B training was included in Experiment 2.

Method

Participants and design

Participants (143 total) were recruited from the University of Maine community and received partial course credit for participation. Participants were randomly assigned to one of four experimental conditions in the 2 category structure (RB vs. II) × 2 training methodology (A/B, INF) design. The experiment was conducted across two 60-min sessions on consecutive days. Twenty-one participants failed to return on Day 2. Although the attrition rate was high, it was similar across the four conditions. The sample sizes (and number of participants that did not return for Day 2) by condition were RB-A/B: 33 (3), RB-INF: 29 (6), II-A/B: 30 (6), II-INF: 30 (6). All participants reported normal (20/20) or corrected-to-normal vision.

Procedure

The stimuli and procedure were identical to Experiment 1 with two exceptions. First, the YES/NO condition was not included. Second, the experiment was conducted across two, consecutive days. On Day 1, participants completed four blocks of A/B or INF training. Day 2 was identical to Experiment 1 with participants completing four additional blocks of training followed by the test phase.

Results

Training phase

The results from the training phase were similar to Experiment 1 (Fig. 4). In the A/B condition, learning was evident in both category structures, but accuracy was generally higher and increased more quickly in the RB structure. A 2 category structure × 8 block mixed ANOVA was consistent with these observations with all effects being significant, structure: $F(1, 61) = 199.36, p < .05, \eta_p^2 = .77$; block: $F(3.96, 241.56) = 24.41$, $p < .05$, $\eta_p^2 = .29$; Structure × Block: $F(3.96, 241.56) = 5.52$, $p < .05$, $\eta_p^2 = .08$]. As with the A/B condition of Experiment 1, the interaction was driven by a more consistent increase in accuracy across blocks in the II structure than in the RB structure, likely due to a ceiling effect in the RB structure. The decision-bound models described in Experiment 1 were fit to the final training block in the A/B conditions and indicated that most participants in the RB task were best fit by a unidimensional model on diameter (88%) and most participants in the II task were best fit by an information-integration model (73%).

In the INF condition, a 2 category structure × 8 block mixed ANOVA indicated a significant effect of block, $F(4, 227.9) = 6.41$, $p < .05$, $\eta_p^2 = .1$, reflecting improved performance with training (e.g., Block 8 > Block 1, $p < .05$). No other effects were statistically significant, structure: $F(1, 57) = .42, p = .52, \eta_p^2 = .007$; Block × Structure: $F(4, 227.9) = 2.02$, $p = .09$, $\eta_p^2 = .03$. In sum, as in Experiment 1, there was evidence of learning in all conditions, albeit to varying degrees.

Test phase

Inspection of the correlations during the test phase (Fig. 5) suggests learning and generalization of the correlational structure of the training categories for the II structure and for inference training with the RB structure. A series of one-sample t tests (see Table 2) were consistent with this observation. For the RB structure, the correlations for the training and transfer stimuli were significantly greater than zero in only the INF condition. In contrast, for the II structure, all correlations were significantly greater than zero, suggesting learning and generalization of the correlational structure. A 2 stimulus set (training, transfer) × 2 category structure × 3 training methodology ANOVA comparing the magnitude of the correlation across conditions indicated that learning was greater for the trained region of the stimulus space [main effect of stimulus set:

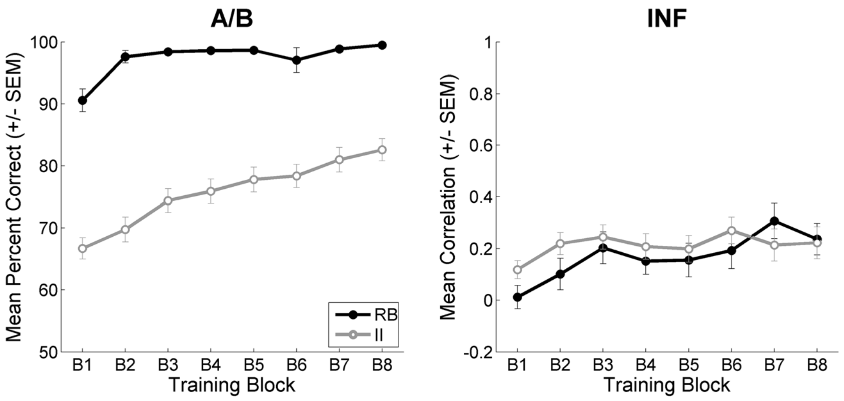

Fig. 4 Training performance across two, four-block sessions in the A/B and INF conditions of Experiment 2

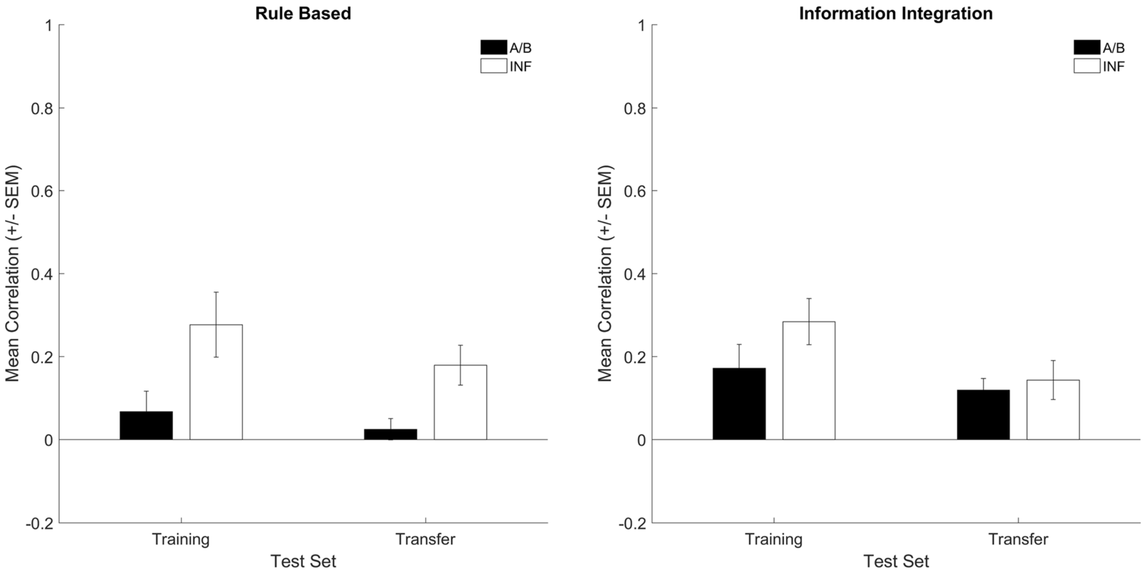

Fig. 5 Performance on the inference task used during the test phase of Experiment 2

$F(1, 118) = 12.24$, $p < .05$, $\eta_p^2 = .09$, and that INF training was superior to A/B training, main effect of methodology: $F(1, 118) = 7.95$, $p < .05$, $\eta_p^2 = .06$. None of the other effects were statistically significant (all $F \leq 2.23$, $p \geq .14$, $\eta_p^2 \leq .02$). In sum, with extended training participants in the RB condition were able to learn and generalize the within-category correlation, but these results were specific to inference.

Summary

The goal of Experiment 2 was to investigate if extended training might facilitate the learning and generalization of within-category information. For the II structure, the results mirrored those from Experiment 1 with evidence of learning and generalization regardless of training methodology. For the RB task, participants were able to learn and generalize with extended training, but these effects were specific to inference training. In sum, II structures and inference training seem to be best suited for the learning and generalization of within-category information.

Table 2 Knowledge of the within-category correlational structure during the test phase of Experiment 2

		Training stimuli			Transfer stimuli		
Rule based	*df*	*t*	*p*	*d*	*t*	*p*	*d*
A/B	32	1.39	.09	.24	.98	.17	.17
INF	29	3.54*	.001	.66	3.78*	.001	.7
Information integration							
A/B	29	3.04*	.005	.56	4.31*	.0002	.79
INF	29	5.08*	.0002	.93	3.05*	.005	.56

Note. One-sample *t* tests comparing observed correlations to no learning (i.e., correlation = 0), *df* are constant across training and transfer. *Statistically significant at Šidák corrected $\alpha = .006$.

Experiment 3

Experiments 1 and 2 focused on the ability to generalize within-category representations to an inference task. In principle, within-category representations should also be able to support categorization. The primary goal of Experiment 3 was to investigate the extent to which INF training would support generalization to A/B classification with the same stimuli. A second, more exploratory question concerned the extent to which between- and within-category representations could be generalized to support the classification of novel stimuli. RB-A/B training would be expected to result in between-category representations that could be generalized to support the classification of novel stimuli (Casale et al., 2012). More specifically, a unidimensional boundary on circle diameter could be applied to novel stimuli during the test phase (e.g., an extension of the vertical boundary plotted with the RB structure in Fig. 1). Casale et al. found, however, that the within-category representations resulting from II-A/B training had limited generalizability to novel stimuli.

Method

Participants and design

Participants (170 total) were recruited from the University of Maine community and received partial course credit for participation. Participants were randomly assigned to one of four experimental conditions in the 2 category structure (RB vs. II) × 2 training methodology (A/B, INF) design. A total of five participants were excluded from analysis: one participant due to software error (RB-A/B) and four participants were statistical outliers (i.e., more than three standard deviations from the mean on both average training accuracy and accuracy during the final training block; RB-INF: 1; II-INF: 3). The resulting sample sizes by condition were RB-A/B: 44; RB-INF: 40; II-A/B: 43; II-INF: 38. All participants reported normal (20/20) or corrected-to-normal vision. Each participant completed one session of approximately 60 minutes duration.

Procedure

The stimuli and procedure were identical to Experiment 1 with two exceptions. First, the YES/NO condition was not included. Second, all participants were tested on classification (i.e., A/B) without corrective feedback.

Results

Training phase

In the A/B conditions, learning was evident in both category structures and accuracy was higher in the RB structure (Fig. 6, left). A 2 category structure × 4 block mixed ANOVA was consistent with these observations, structure: $F(1, 85) = 485.14, p < .05, \eta_p^2 = .85$; block: $F(2.71, 230.03) = 26.31, p < .05, \eta_p^2 = .24$; Structure × Block: $F(2.71, 230.03) = 2.49, p = .07, \eta_p^2 = .03$. As was the case in the previous experiments, most participants in the RB task were best fit by a unidimensional model during the final training block (86%). However, information-integration strategies were not as prevalent in the II task (37%) with the majority of participants using rule-based strategies (47%) or guessing (16%).

The correlations in the INF conditions followed a similar pattern as the accuracy rates in the A/B conditions (Fig. 6, right). A 2 category structure × 4 block mixed ANOVA conducted on the correlations between the presented and inferred dimension indicated that all effects were significant, structure: $F(1, 76) = 13.81, p < .05, \eta_p^2 = .15$; block: $F(2.58,\ 196.2) = 11.88, p < .05, \eta_p^2 = .14$; Structure × Block: $F(2.58, 196.2) = .3.15, p < .05, \eta_p^2 = .04$. The Structure × Block interaction was driven by a superior performance in the II structure during Blocks 1–3 (*p*s < .05), but equivalent performance by the end of training (*p* = .24). In sum, as in Experiments 1 and 2, there was evidence of learning in all conditions, albeit to varying degrees.

Test phase

Recall that unlike Experiments 1 and 2, Experiment 3 participants performed A/B classification during the test phase. The two training clusters had an objectively correct response and, therefore, were analyzed by computing categorization accuracy (Fig. 7). For participants in the INF conditions, accuracy on the training clusters provided a measure of generalization from inference to classification. A series of one-sample *t* tests (Šidák corrected $\alpha = .0127$) indicated that accuracy in all conditions was significantly greater than chance, rule-based: A/B - $t(43) = 22.96, p < .001, d = 7$; INF - $t(39) = 6.6, p < .001, d = 2.11$; information-integration: A/B - $t(42) = 6.72, p < .001, d = 2.07$; INF - $t(37) = 2.54, p = .007, d = .84$. A 2 category structure × 2 training methodology ANOVA comparing accuracy across conditions indicated that test phase accuracy was generally greater for the RB structure, main effect of structure: $F(1, 161) = 92.24, p < .05, \eta_p^2 = .36$, and when participants continued to categorize at test, main effect of methodology: $F(1, 161) = 37.72, p < .05, \eta_p^2 = .19$]. The Category Structure × Training Methodology interaction was also significant, $F(1, 118) = 5.24, p < .05, \eta_p^2 = .03$, with higher A/B accuracy for both the RB (*p* < .001) and II (*p* = .008) structures. The interaction likely reflects a greater A/B advantage for the RB structure (*d* = 8.57) than the

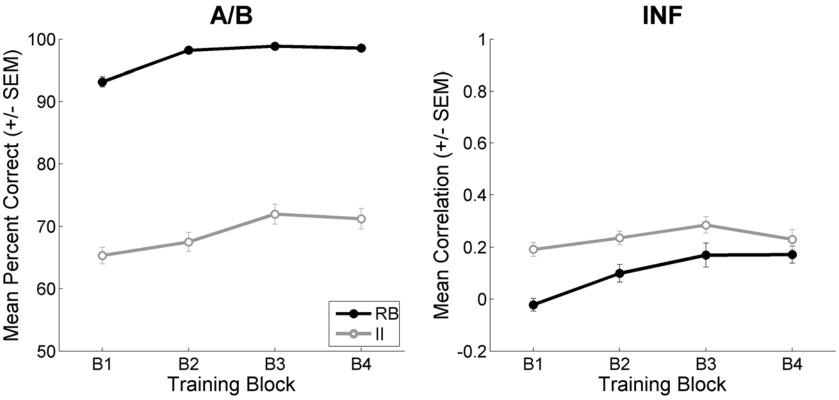

Fig. 6 Training performance in the A/B and INF conditions of Experiment 3

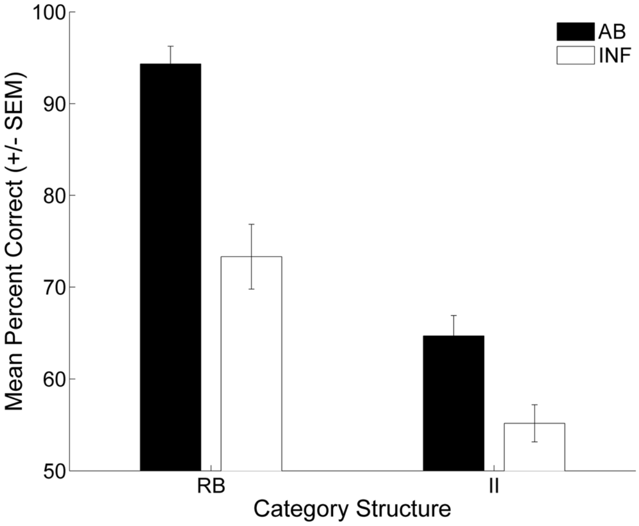

Fig. 7 Test phase accuracy for the training clusters

II structure (d = 2.5). In sum, these data suggest that participants could generalize knowledge from inference to categorization, but performance was inferior to participants that did not have to generalize across methodologies.

No feedback was provided at test, thus the test transfer items did not have an objectively correct response. Nevertheless, the analysis of how the transfer items were classified permitted an exploration of generalization profiles to novel regions of the stimulus space. Previous work would suggest that participants in the RB-A/B condition would apply the between-category representation acquired during training (i.e., a unidimensional rule on diameter) to the transfer items. In contrast, participants in the II-A/B condition would be expected to be limited in their ability to generalize their decision strategy to novel stimuli. To investigate this question, the decision-bound models described in Experiment 1 were fit to each participant's test phase data. As expected, unidimensional strategies on diameter dominated in the RB-A/B condition (see Table 3). In contrast, strategy use was variable in the other conditions, and there was little evidence of the consistent application of a task-appropriate strategy during the test phase.

Table 3 Percentage of participants best fit by each model type during the test phase

	Condition			
Model Type	RB-A/B	RB-INF	II-A/B	II-INF
UD-Diameter	93.2	32.5	16.3	10.5
Other RB	2.3	17.5	51.1	31.6
II	4.5	25.0	25.6	26.3
RR	0	25.0	7.0	31.6

Note. UD = unidimensional; RB = rule based; II = information integration; RR = random responder

Summary

The primary goal of Experiment 3 was to investigate if the within-category representations learned in the INF condition would support classification with the same stimuli. Participants in the INF conditions were able to generalize across tasks when classifying the training items. Their performance, however, was inferior to those participants that classified during training and test. The exploratory analysis of generalization profiles to novel stimuli replicated previous work demonstrating that the between-category representations learned with RB-A/B training could be generalized to novel stimuli.

General discussion

This manuscript reports the results of three experiments designed to investigate the impact of category structure and training methodology on the ability to learn and generalize within-category representations (i.e., within-category correlations). In Experiment 1, with an II structure, participants were able to learn within-category representations regardless of whether participants used classification, concept, or inference training. Participants were also able to generalize within-category representations to novel tasks and novel stimuli. Experiment 2 revealed that with extended training, within-category representations learned with a RB structure could be generalized to novel stimuli. Experiment 3 showed that within-category representations learned in a RB structure could also be generalized to a novel task. These results complement the growing body of work highlighting the impact of category structure and training methodology on category representations (Carvalho & Goldstone, 2015; Hammer et al., 2009; Levering & Kurtz, 2015). These results also build upon previous work by investigating the relationship between these factors and the generalization of categorical knowledge (Carvalho & Goldstone, 2014; Chin-Parker & Ross, 2002; Hoffman & Rehder, 2010).

Learning of within-category representations

Previous research has shown that II structures and concept training result in a bias towards learning within-category representations (Hélie et al., 2017). Moreover, inference training results in knowledge of within-category correlations that can be generalized to novel tasks (Chin-Parker & Ross, 2002; Markman & Ross, 2003). Based on these data, it was predicted that within-category representations would be learned in all but the RB-A/B condition, a combination of category structure and training methodology that leads to a between-category representation (Casale et al., 2012; Ell & Ashby, 2012; Ell et al., 2009; Hélie et al., 2017). With the exception of the RB-YES/NO condition, these predictions were supported. Within-category correlations could be learned with an II structure (see also Thomas, 1998) and with extended training in the RB-INF condition.

Extended training in the RB-INF condition, and II training (Experiment 2) resulted in similar knowledge of the within-category correlations. The process of producing a missing stimulus value (i.e., the production task), much like producing a missing category label, is a direct measure of knowledge. It could be, however, that a direct measure of within-category information was not an ideal match for measuring the category representation that is learned with II structures (e.g., Roediger, Marsh, & Lee, 2002). Thus, it is possible that the observed within-category correlations are an underestimate of within-category representations in the II structure.

The present data do not support previous claims that within-category correlations cannot be learned during the course of categorization (Markman & Ross, 2003). Instead, and consistent with previous work (Anderson & Fincham, 1996; Thomas, 1998), within-category correlations could be learned during categorization. Such learning, however, is constrained by the structure of the categories. Furthermore, because the tasks and stimulus dimensions were the same for both RB and II structures, the present data argue against the hypothesis that within-category correlations learned during categorization are a byproduct of simplistic stimuli, overtraining, and/or classification tasks that incorporate additional inference-like training (e.g., Chin-Parker & Ross, 2002).

Hélie et al. (2017) showed that, even with an RB structure, concept training (i.e., YES/NO) results in a bias towards within-category representations. In Hélie et al., participants learned two RB category structures (simultaneously) along a single diagnostic stimulus dimension (Category A vs. Category B and Category C vs. Category D). Participants were subsequently tested on a novel categorization problem using the same categories (i.e., Category B vs. Category C). Participants were successfully able to generalize the knowledge when receiving concept training, but not when receiving classification training, suggesting that concept training promoted a representation based on the categories themselves rather than between-category differences (see Hoffman & Rehder, 2010, for a related finding). Consistent with this interpretation, a computational model assuming within-category representations fit to the concept training data was more successful in predicting test phase performance than a model assuming between-category representations. Thus, it may be the case that concept training promotes a minimal within-category representation that is sufficient to support classification on a novel RB categorization problem (e.g., the range of values on the stimulus dimensions) but not so rich so as to include knowledge that was not required during training (e.g., the correlational structure of the categories).

Indeed, it has often been argued that participants learn what is necessary to perform the task at hand (Markman & Ross, 2003; Pothos & Chater, 2002; Yamauchi & Markman, 1998). This would suggest that the learning of within-category correlations depends upon their relevance to the training methodology. With inference training, knowledge of the within-category correlations would facilitate performance regardless of the category structure. In the categorization conditions, however, the relevance of within-category correlations depends upon the category structure. With the RB structure, successful performance during training did not depend upon learning the relationship between diameter and angle. In this task, the vast majority of participants selectively attended to diameter during training, using unidimensional decision rules. It is thus possible that participants in the YES/NO task learned within-category representations that did not include correlations (i.e., they could be limited to containing the range of the stimulus dimensions). With the II structure, however, successful performance depends upon knowledge of the relationship between diameter and angle. Thus, the II structure may promote learning within-category representations that contain correlation information. The present data suggest that within-category representations may only include information that is necessary for successful performance. As a result, within-category representations learned with inference training, or an II structure, would contain correlation information, but within-category representations learned with a RB structure might not.

Generalization of within-category representations

In addition to investigating the factors that promote within-category representations, another goal of the present work was to investigate the extent to which within-category representations can be generalized to support performance with novel stimuli and/or on novel tasks. Successful generalization performance on the inference task used during the test phase of Experiments 1 and 2 depended upon a within-category representation that contained correlation information. For the II structure, generalization to novel stimuli and tasks was supported by within-category representations. Participants in the II-INF conditions were able to generalize knowledge of the within-category correlations to the transfer stimuli and participants in the II-A/B and II-YES/NO conditions were able to generalize knowledge of the within-category correlations to the inference task used during the test phase. In contrast, with the RB structure, generalization was limited to novel stimuli with participants in the RB-INF condition of Experiment 2 being able to generalize knowledge of the within-category correlations to the transfer stimuli. The lack of successful generalization in the RB-YES/NO condition may reflect a mismatch between the nature of the within-category information (i.e., range of the stimulus dimensions) and the knowledge necessary at test (i.e., within-category correlations), instead of reflecting the absence of a within-category representation.

In the categorization conditions, training accuracy was consistently higher in the RB structure than the II structure, but the opposite pattern was observed on the inference task during the test phase. Moreover, in Experiment 3, inference training performance was higher for the II structure than the RB structure, but the pattern of test phase performance on the categorization task was reversed. Thus, it may be the case that increased difficulty during training somehow benefits the ability to learn and generalize within-category representations—a common observation in the learning and memory literature (e.g., Schmidt & Bjork, 1992). However, if task difficulty were the primary determinant of learning and generalization of within-category representations, then we should have also observed a difference in transfer performance between the RB-YES/NO and RB-A/B conditions of Experiment 1. Despite average training accuracy across blocks being lower in the RB-YES/NO (M = 93.2, SD = 3.9) condition than the RB-A/B condition (M = 97.1, SD = 1.7), $t(59) = 3.9$, $p < .05$, d = 1.28, within-category representations were not learned in either condition. That being said, we cannot rule out the possibility that with extended training, or a generalization task dependent upon aspects of within-category representations other than within-category correlation, generalization in the RB-YES/NO condition would have been stronger than generalization in the RB-A/B condition.

Inference training was also able to support generalization to classification—an effect that was more pronounced for the RB structure (Experiment 3). Participants in the RB-INF condition demonstrated learning of the within-category correlations that was similar to the performance observed in Experiments 1 and 2. It is possible, however, that the seemingly successful generalization from inference to classification may actually be a consequence of a preference to use unidimensional rules in the absence of feedback in category structures where logical rules would be successful (Ashby et al., 1999; Ell & Ashby, 2012; Ell et al., 2012; Medin et al., 1987; Milton & Wills, 2004; Pothos & Chater, 2005; Pothos & Close, 2008). Although the training clusters were interleaved with the transfer clusters during the test phase, many unidimensional rules would have resulted in high test phase accuracy when focusing on the widely separated training clusters. Moreover, if inference training generally supported classification, then one might have also expected successful generalization in the II-INF condition. In light of these points, the present data provide limited evidence of the ability to generalize within-category representations learned by inference to support classification.

The RB structure had a greater range on diameter than the II structure. This was done to equate the accuracy of a unidimensional rule on diameter in the RB structure with the accuracy of information-integration models in the II structure while also equating the within-category correlations between category structures. One consequence of this design choice is that in the RB conditions, a higher proportion of the transfer stimuli fell within the range of stimuli used during training. Thus, test-phase differences as a function of category structure could be at least partly attributable to interference (e.g., producing a line angle consistent with the training set rather than the transfer set). Several considerations, however, would make this possibility seem unlikely. For instance, this account would predict that test phase performance for the transfer stimuli would have varied by cluster, but this was not the case. In addition, in Experiment 2, inference training with either category structure produced similar levels of transfer performance. Thus, the difference in the range of the training stimuli between category structures does not seem to be problematic for interpretation of the test phase data.

Conclusions

These data suggest that both category structure and training methodology are important factors in shaping the way categorical knowledge is represented and ultimately used. These results extend previous research on inference training to demonstrate the generality of this training methodology as a means for learning within-category representations that can support generalization. In addition, regardless of training methodology, II structures promote the learning of within-category representations that contain correlation information. Although between-category representations may be efficient for classification, such representations would seem to have limited utility for generalization. Intuitively, learning within-category representations appears to be more useful in most cases. For example, learning within-category representations about cats and dogs is more useful than exclusively learning between-category representations between cats and dogs. Knowing what a cat is will facilitate distinguishing a cat from another animal, whereas only learning what is different between cats and dogs may not help. Indeed, the present data are consistent with these speculations and make the contribution of clarifying the conditions that promote within-category representations. Moreover, these data will be useful in developing a new generation of computational models better equipped to deal with knowledge generalization.

Acknowledgments This work was supported by the National Science Foundation under Grants #1349677-BCS and #1349737-BCS to S.H. and S.W.E., respectively. The authors would like to thank Alex Groat, Mikael Heikkinen, Steve Hutchinson, Erin Perry, Nicole Rubin, and Lauren Szymula for their help with data collection.

Appendix 1

Table 4 Test stimuli coordinates (arbitrary units)

II training/RB transfer		II training/RB transfer		RB training/II transfer		RB training/II transfer	
Diameter	Angle	Diameter	Angle	Diameter	Angle	Diameter	Angle
410.0	**-95.0**	**340.0**	**-35.0**	560.0	-95.0	190.0	-35.0
421.5	-83.5	351.5	-23.5	571.5	-83.5	201.5	-23.5
433.1	**-71.9**	**363.1**	**-11.9**	583.1	-71.9	213.1	-11.9
444.6	-60.4	374.6	-0.4	594.6	-60.4	224.6	-0.4
456.2	**-48.8**	**386.2**	**11.2**	606.2	-48.8	236.2	11.2
467.7	-37.3	397.7	22.7	617.7	-37.3	247.7	22.7
479.2	**-25.8**	**409.2**	**34.2**	629.2	-25.8	259.2	34.2
490.8	-14.2	420.8	45.8	640.8	-14.2	270.8	45.8
502.3	**-2.7**	**432.3**	**57.3**	652.3	-2.7	282.3	57.3
513.8	8.8	443.8	68.8	663.8	8.8	293.8	68.8
525.4	**20.4**	**455.4**	**80.4**	675.4	20.4	305.4	80.4
536.9	31.9	466.9	91.9	686.9	31.9	316.9	91.9
548.5	**43.5**	**478.5**	**103.5**	698.5	43.5	328.5	103.5
560.0	55.0	490.0	115.0	710.0	55.0	340.0	115.0
II and RB transfer		II and RB transfer		II and RB transfer		II and RB transfer	
Diameter	Angle	Diameter	Angle	Diameter	Angle	Diameter	Angle
440.0	215.0	590.0	215.0	660.0	155.0	810.0	155.0
451.5	226.5	601.5	226.5	671.5	166.5	821.5	166.5
463.1	238.1	613.1	238.1	683.1	178.1	833.1	178.1
474.6	249.6	624.6	249.6	694.6	189.6	844.6	189.6
486.2	261.2	636.2	261.2	706.2	201.2	856.2	201.2
497.7	272.7	647.7	272.7	717.7	212.7	867.7	212.7
509.2	284.2	659.2	284.2	729.2	224.2	879.2	224.2
520.8	295.8	670.8	295.8	740.8	235.8	890.8	235.8
532.3	307.3	682.3	307.3	752.3	247.3	902.3	247.3
543.8	318.8	693.8	318.8	763.8	258.8	913.8	258.8
555.4	330.4	705.4	330.4	775.4	270.4	925.4	270.4
566.9	341.9	716.9	341.9	786.9	281.9	936.9	281.9
578.5	353.5	728.5	353.5	798.5	293.5	948.5	293.5
590.0	365.0	740.0	365.0	810.0	305.0	960.0	305.0

Note. Bold coordinates were used as probe stimuli during the final block of training

Appendix 2

Model-based analyses

To get a more detailed description of how participants categorized the stimuli, a number of different decision bound models (Ashby, 1992a; Maddox & Ashby, 1993) were fit separately to the data for each participant. Decision bound models are derived from general recognition theory (Ashby & Townsend, 1986), a multivariate generalization of signal detection theory (Green & Swets, 1966). It is assumed that, on each trial, the percept can be represented as a point in a multidimensional psychological space and that each participant constructs a decision bound to partition the perceptual space into response regions. The participant determines which region the percept is in and then makes the corresponding response. While this decision strategy is deterministic, decision bound models predict probabilistic responding because of trial-by-trial perceptual and criterial noise (Ashby & Lee, 1993).

This appendix briefly describes the decision bound models. For more details, see Ashby (1992a) or Maddox and Ashby (1993). The classification of these models as either *rule-based* or *information-integration* models is designed to reflect current theories of how these strategies are learned (e.g., Ashby et al., 1998) and has received considerable empirical support (see Ashby & Maddox, 2005; Maddox & Ashby, 2004, for reviews).

Rule-based models

Unidimensional classifier (UC). This model assumes that the stimulus space is partitioned into two regions by setting a criterion on one of the stimulus dimensions. Two versions of the UC were fit to the data. One version assumes that participants attended selectively to diameter and the other version assumes participants attended selectively to angle. The UC has two free parameters, one corresponds to the decision criterion on the attended dimension and the other corresponds to the variance of internal (perceptual and criterial) noise (σ^2). A special case of the UC, the optimal unidimensional classifier, assumes that participants use the unidimensional decision bound that maximizes accuracy. This special case has one free parameter (σ^2).

Conjunctive classifier (CC) An alternative rule-based strategy is a conjunction rule involving separate decisions about the stimulus value on the two dimensions with the response assignment based on the outcome of these two decisions (Ashby & Gott, 1988). The CC assumes that the participant partitions the stimulus space into four regions. Based on an initial inspection of the data, two versions of the CC were fit to these data. One version assumes that individuals assigned a stimulus to Category B if it was low on diameter and high on angle; otherwise, the stimulus would be assigned to Category A. The other version assumes that individuals assigned a stimulus to Category A if it was high on diameter and low on angle; otherwise, the stimulus would be assigned to Category B. The CC has three free parameters: the decision criteria on the two dimensions and a common value of σ^2 for the two dimensions.

Information-integration models

The linear classifier (LC). This model assumes that a linear decision bound partitions the stimulus space into two regions. The LC differs from the CC in that the LC does not assume decisional selective attention (Ashby & Townsend, 1986). This produces an information-integration decision strategy because it requires linear integration of the perceived values on the stimulus dimensions. The LC has three parameters, slope and intercept of the linear bound and σ^2.

The minimum distance classifier (MDC). This model assumes that there are a number of units representing a low-resolution map of the stimulus space (Ashby & Waldron, 1999; Ashby, Waldron, Lee, & Berkman, 2001; Maddox, Filoteo, Hejl, & Ing, 2004). On each trial, the participant determines which unit is closest to the perceived stimulus and produces the associated response. The version of the MDC tested here assumes two units because the category structures were generated from two multivariate normal distributions. Because the location of one of the units can be fixed, and because a uniform expansion or contraction of the space will not affect the location of the minimum-distance decision bounds, the MDC has four free parameters (three determining the location of the units and σ^2).

Random responder models

Equal response frequency (ERF) This model assumes that participants randomly assign stimuli to the two response frequencies in a manner that preserves the category base rates (i.e., 50% of the stimuli in each category). This model has no free parameters.

Biased response frequency (BRF). This model assumes that participants randomly assign stimuli to the two response frequencies in a manner that matches the participant’s categorization response frequencies. This model has one free parameter, the proportion of stimuli in Category A. Although the ERF and BRF are assumed to be consistent with guessing, these models would also likely provide the best account of participants that frequently shift to very different strategies.

Model fitting

The model parameters were estimated using maximum likelihood (Ashby, 1992b; Wickens, 1982) and the goodness-of-fit statistic was

$$BIC = rlnN{-}2lnL$$

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

References

Anderson, J. R., & Fincham, J. M. (1996). Categorization and sensitivity to correlation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 259–277.

Ashby, F. G. (1992a). Multidimensional models of categorization. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 449–483). Hillsdale, NJ: Erlbaum.

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

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

Ashby, F. G., & Crossley, M. J. (2011). A Computational model of how cholinergic interneurons protect striatal-dependent learning. *Journal of Cognitive Neuroscience, 23,* 1549–1566. doi:10.1162/jocn.2010.21523

Ashby, F. G., & Ell, S. W. (2001). The neurobiology of human category learning. *Trends in Cognitive Science, 5,* 204–210.

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., & Lee, W. W. (1993). Perceptual variability as a fundamental axiom of perceptual science. In S. C. Masin (Ed.), *Foundations of percpetual theory* (pp. 369–399). Amsterdam: Elsevier.

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology, 56,* 149–178. doi:10.1146/annurev.psych.56.091103.070217

Ashby, F. G., Queller, S., & Berretty, P. M. (1999). On the dominance of unidimensional rules in unsupervised categorization. *Perception & Psychophysics, 61,* 1178–1199.

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review, 93,* 154–179.

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

Ashby, F. G., Waldron, E. M., Lee, W. W., & Berkman, A. (2001). Suboptimality in human categorization and identification. *Journal of Experimental Psychology: General, 130,* 77–96.

Brainard, D. H. (1997). Psychophysics software for use with MATLAB. *Spatial Vision, 10,* 433–436.

Carvalho, P. F., & Goldstone, R. L. (2014). Putting category learning in order: Category structure and temporal arrangement affect the benefit of interleaved over blocked study. *Memory & Cognition, 42,* 481–495. doi:10.3758/s13421-013-0371-0

Carvalho, P. F., & Goldstone, R. L. (2015). The benefits of interleaved and blocked study: Different tasks benefit from different schedules of study. *Psychonomic Bulletin & Review, 22,* 281–288. doi:10.3758/s13423-014-0676-4

Casale, M. B., & Ashby, F. G. (2008). A role for the perceptual representation memory system in category learning. *Perception & Psychophysics, 70,* 983–999.

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

Chin-Parker, S., & Ross, B. H. (2002). The effect of category learning on sensitivity to within-category correlations. *Memory & Cognition, 30,* 353–362.

Chin-Parker, S., & Ross, B. H. (2004). Diagnosticity and prototypicality in category learning: A comparison of inference learning and classification learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 30,* 216–226. doi:10.1037/0278-7393.30.1.216

Ell, S. W., & Ashby, F. G. (2006). The effects of category overlap on information-integration and rule-based category learning. *Perception and Psychophysics, 68,* 1013–1026.

Ell, S. W., & Ashby, F. G. (2012). The impact of category separation on unsupervised categorization. *Attention, Perception, & Psychophysics, 74,* 466–475.

Ell, S. W., Ashby, F. G., & Hutchinson, S. (2012). Unsupervised category learning with integral-dimension stimuli. *Quarterly Journal of Experimental Psychology, 65,* 1537–1562.

Ell, S. W., Ing, A. D., & Maddox, W. T. (2009). Criterial noise effects on rule-based category learning: The impact of delayed feedback. *Attention, Perception, & Psychophysics, 71,* 1263–1275.

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

Goldstone, R. L. (1996). Isolated and interrelated concepts. *Memory & Cognition, 24,* 608–628.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Hammer, R., Diesendruck, G., Weinshall, D., & Hochstein, S. (2009). The development of category learning strategies: what makes the difference? *Cognition, 112,* 105–119. doi:10.1016/j.cognition.2009.03.012

Hélie, S., Shamloo, F., & Ell, S. W. (2017). *The effect of training methodology on kowledge representation in perceptual categorization.* Manuscript submitted for publication.

Hoffman, A. B., & Rehder, B. (2010). The costs of supervised classification: The effect of learning task on conceptual flexibility. *Journal of Experimental Psychology: General, 139,* 319–340. doi:10.1037/a0019042

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3? *Perception, 36* (ECVP Abstract Supplement).

Levering, K. R., & Kurtz, K. J. (2015). Observation versus classification in supervised category learning. *Memory & Cognition, 43,* 266–282. doi:10.3758/s13421-014-0458-2

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

Maddox, W. T., & Ashby, F. G. (2004). Dissociating explicit and procedural-learning based systems of perceptual category learning. *Behavioral Processes, 66,* 309–332.

Maddox, W. T., & Bogdanov, S. V. (2000). On the relation between decision rules and perceptual representation in multidimensional perceptual categorization. *Perception & Psychophysics, 62*(5), 984–997.

Maddox, W. T., Bohil, C. J., & Ing, A. D. (2004). Evidence for a procedural learning-based system in category learning. *Psychonomic Bulletin & Review, 11,* 945–952.

Maddox, W. T., Filoteo, J. V., Hejl, K. D., & Ing, A. D. (2004). Category number impacts rule-based but not information-integration category learning: Further evidence for dissociable category learning systems. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 30,* 227–235.

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, & Cognition, 31,* 654–669. doi:10.1037/0278-7393.31.4.654

Markman, A. B., & Ross, B. (2003). Category use and category learning. *Psychological Bulletin, 129,* 529–613.

Medin, D. L., Wattenmaker, W. D., & Hampson, S. E. (1987). Family resemblance, conceptual cohesiveness, and category construction. *Cognitive Psychology, 19,* 242–279.

Milton, F., & Wills, A. J. (2004). The influence of stimulus properties on category construction. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 30,* 407–415. doi:10.1037/0278-7393.30.2.407

Minda, J. P., & Ross, B. H. (2004). Learning categories by making predictions: An investigation of indirect category learning. *Memory & Cognition, 32,* 1355–1368.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review, 101,* 53–79.

Nosofsky, R. M., & Zaki, S. R. (1998). Dissociations between categorization and recognition in amnesic and normal individuals: An exemplar-based interpretation. *Psychological Science, 9,* 247–255.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437–442.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77,* 353–363.

Pothos, E. M., & Chater, N. (2002). A simplicity principle in unsupervised human categorization. *Cognitive Science, 26,* 303–343.

Pothos, E. M., & Chater, N. (2005). Unsupervised categorization and category learning. *Quarterly Journal of Experimental Psychology: A, 58,* 733–752.

Pothos, E. M., & Close, J. (2008). One or two dimensions in spontaneous classification: A simplicity approach. *Cognition, 107,* 581–602. doi:10.1016/j.cognition.2007.11.007

Reber, P. J., Stark, C. E. L., & Squire, L. R. (1998). Cortical areas supporting category learning identified using functional MRI. *Proceedings of the National Academy of Sciences of the United States of America, 95,* 747–750.

Roediger, H. L., Marsh, E. J., & Lee, S. C. (2002). Kinds of memory. In H. Pashler & D. L. Medin (Eds.), *Stevens' handbook of experimental psychology* (Memory and cognitive processes 3rd ed., Vol. 2, pp. 1–41). New York: John Wiley & Sons.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of natural categories. *Cognitive Psychology, 7,* 573–605.

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.

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

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,* 54–65. doi:10.1037/a0015892

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

Smith, J. D., & Minda, J. P. (2001). Journey to the center of the category: The dissociation in amnesia between categorization and recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 27,* 984–1002.

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, & Cognition, 28,* 800–811.

Thomas, R. D. (1998). Learning correlations in categorization tasks using large, ill-defined categories. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 24,* 119–143.

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

Yamauchi, T., & Markman, A. B. (1998). Category learning by inference and classification. *Journal of Memory and Language, 39,* 124–148.

Zeithamova, D., Maddox, W. T., & Schnyer, D. M. (2008). Dissociable prototype learning systems: Evidence from brain imaging and behavior. *Journal of Neuroscience, 28,* 13194–13201. doi:10.1523/JNEUROSCI.2915-08.2008

Zotov, V., Jones, M. N., & Mewhort, D. J. (2011). Contrast and assimilation in categorization and exemplar production. *Attention, Perception, & Psychophysics, 73,* 621–639. doi:10.3758/s13414-010-0036-z