

# The impact of training methodology and category structure on the formation of new categories from existing knowledge

Sébastien Hélie

Department of Psychological Sciences, Purdue University

Farzin Shamloo

Department of Psychological Sciences, Purdue University

Shawn W. Ell

Department of Psychology, University of Maine

Categorization decisions are made thousands of times every day, and a typical adult knows tens of thousands of categories. It is thus relatively rare that adults learn new categories without somehow reorganizing pre-existing knowledge. Yet, most perceptual categorization research has investigated the ability to learn new categories without considering their relation to existing knowledge. In this article, we test the ability of young adults to merge already known categories into new categories as a function of training methodology and category structures using two experiments. Experiment 1 tests participants' ability to merge rule-based or information-integration categories that are either contiguous, semi-contiguous, or non-contiguous in perceptual space using a classification paradigm. Experiment 2 is similar to Experiment 1 but uses a YES/NO learning paradigm instead. The results of both experiments suggest a strong effect of the contiguity of the merged categories in perceptual space that depends on the type of category representation that is learned. The type of category representation that is learned, in turn, depends on a complex interaction of the category structures and training task. We conclude by discussing the relevance of these results for categorization outside the laboratory.

**Keywords:** knowledge representation, training methodology, compositionality, perceptual category learning

Categorization is a ubiquitous cognitive process and categorization decisions are made thousands of times every day (Hélie, Waldschmidt, & Ashby, 2010). As a result, a typical adult knows tens of thousands of categories. It is thus relatively rare that adults learn new categories without relying on prior knowledge. For example, if one already knows the category “red objects” and the category “large objects”, then learning the category “large red objects” should be fairly straightforward and would not need re-learning about the color red or large sizes. Instead, the already known categories can be merged together to form the new category. This property has been referred to as *compositionality* (Fodor & Pylyshyn, 1988). Although this issue has been investigated in the context of natural categories (Aerts, Gabora, & Sozzo, 2013; Cohen & Murphy, 1984; Prinz, 2012; E. E. Smith, Osherson, Rips, & Keane, 1988; Voorspoels, Storms, & Vanpaemel, 2012; Wisniewski, 1997; Ling Wu & Barsalou, 2009; Zadeh, 1982), most perceptual category learning tasks have investigated the ability to learn artificial categories without relying on prior knowledge (e.g., F. G. Ashby & Maddox, 2010; Erickson & Kruschke, 1998; Medin & Schaffer, 1978; Nosofsky, 1986; Posner & Keele, 1968; Shepard, Hovland, & Jenkins, 1961; J. D. Smith & Minda, 2002). In this article, we test the ability of young adults to merge already

known categories into new categories as a function of training methodology and category structures using two experiments. The results show that categories that are contiguous in perceptual space are easier to merge, and that the magnitude of the merging cost may depend on the type of category representation that is learned. The type of category representation that is learned, in turn, depends on both the category structures of the learned categories and the training task (Ell, Smith, Peralta, & Hélie, 2017; Hélie, Shamloo, & Ell, 2017).

## Category representation and generalization

Category representations (i.e., the way in which information is stored and used: Markman, 2002) are the building blocks of decision-making from the most routine to the most novel contexts (Hélie & Ashby, 2012). Generally speaking, category representations can be broadly classified as within-category representations or between-category representations (Levering & Kurtz, 2015; Markman & Ross, 2003). Specifically, within-category representations contain information about the categories themselves. For example, a within-category representation of *humans* could contain information about what is common among category members (e.g., one head), the correlation between the features (e.g., as height increases, so does arm length), or the range of feature

values (e.g., adult height typically varies between 5–6.5 feet). In contrast, between-category representations would contain information about the distinguishing features between two categories. For example, a between-category representation contrasting *humans* and *dogs* might contain information about the relevant features for separating humans and dogs (e.g., number of legs) and criteria on these feature-values (e.g., less than three legs is generally a human; more than three legs is generally a dog). Hélie, Ell, and colleagues (Ell et al., 2017; Hélie et al., 2017) argued that within-category representations may be more useful when inferring missing attributes (e.g., one may infer that the author's spouse has two legs without being told) whereas between-category representations may be useful for extrapolation outside of the learning space (e.g., an animal with one leg is even less likely to be a dog than a human). In the category learning literature, prototype models (Reed, 1972; J. D. Smith & Minda, 2002) and exemplar models (Nosofsky, 1986) assume that a within-category representation is the basis for response selection, whereas criterion-setting models (Erev, 1998; Treisman & Williams, 1984) assume a between-category representation.

Given that category representations may contain different types of information, it is reasonable to expect that within- and between-category representations may assemble differently when merged to create new categories. For example, between-category representations can be represented as decision criteria in perceptual space (e.g., a threshold on height separating tall peoples from short peoples), so compositionality could consist of linking multiple decision criteria together using logical connectors (e.g., AND, OR). In the earlier example of large red objects, these would be objects that fall above the large threshold on size AND above some threshold in the hue spectrum. The decisions could be made independently on each dimension using a conjunction rule (F. G. Ashby, Alfonso-Reese, Turken, & Waldron, 1998). Accordingly, applying multiple rules consecutively could be taxing for working memory resources (Miles & Minda, 2011), and one would expect that difficulty would increase with the number of decision criteria that need to be applied to merge the categories (Erickson, 2008). In other words, the effect would be similar to increasing the boolean complexity of the categories (Feldman, 2003; Shepard et al., 1961).

In contrast, within-category representations can be represented by the generative model (e.g., density distribution) that is most likely to have generated the category members (Hélie et al., 2017). For example, imagine four categories formed based on peoples' height (e.g., short, average, tall, very tall). Let's further imagine that each category was generated using a normal distribution, so an ideal observer learning within-category information would learn 4 generative models of the categories (i.e., 4 separate normal distributions

– one for each category) and use the sample means and variances to estimate the distribution parameters. The top row of Figure 1 shows an example resulting model where the  $x$ -axis would represent the height of an individual and the  $y$ -axis would represent the probability.

After learning the 4 categories, imagine that the ideal observer is asked to merge the learned categories into 2 new categories. There are many ways in which generative models can be merged. If the newly formed categories are contiguous, one could aggregate the learned generative distributions by calculating a grand mean and variance to include all the stimuli from the merged categories, or learn a mixture model of the training categories. Mixture models are statistical models where a distribution is formed using a weighed sum of basis functions (in this case normal distributions) with different parameter values (Bishop, 2006). These two possibilities are shown in the middle row of Figure 1. These two merged models are fairly straightforward and not much complexity is added since there are no stimuli with a different category label between the merged categories. However, if the categories become non-contiguous, then the best that can be done by the ideal observer is to create mixture models that are semi-contiguous (Figure 1, bottom-left) or non-contiguous (Figure 1, bottom-right). These mixture models are discontinuous, which could drastically increase the difficulty of the categorization task.

## Generating predictions

To provide an intuition for the difficulty of merging the generative models, we simulated the merging models included in Figure 1. The simulation went as follows: (1) A set of 96 training stimuli was generated from 4 Gaussian distributions. The resulting maximum likelihood Gaussian models are shown in the top row of Figure 1. (2) Next, the training models identified in (1) were merged and tested using 96 new stimuli generated from the original training distributions. For the contiguous condition, both models in the middle row of Figure 1 were tested. For the semi-contiguous and non-contiguous conditions, the models in the bottom row of Figure 1 were tested. (3) In all cases, the test stimuli were presented one at a time and the probability of the stimulus under each merged category was calculated based on the distribution models. (4) A response was selected stochastically. Each condition was simulated 10,000 times.

When the learning densities (top row of Figure 1) are merged with optimal mixture weights (i.e., the weights linearly combining the merged learning densities), all conditions produce a test accuracy > 99%. However, human participants do not have the opportunity to learn the mixture weights: They are only told that the categories are merged. To produce more realistic predictions, we added mean-centered Gaussian noise to the optimal mixture param-

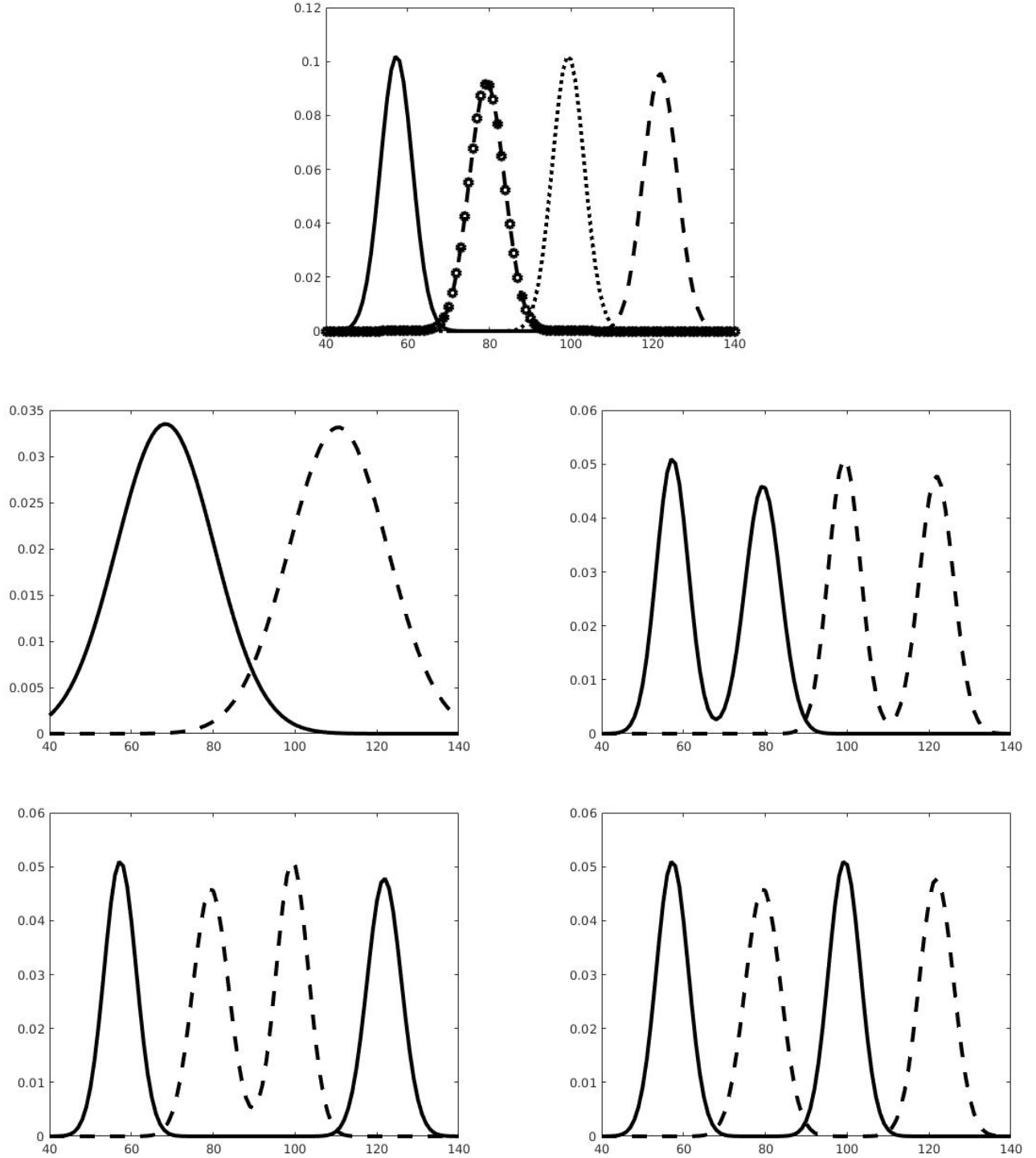


Figure 1. Top row: Generative models learned by an ideal observer. Middle row: Possible merging models for contiguous categories. Bottom row: Possible merging models for semi-contiguous and non-contiguous categories.

eter estimates ( $SD = 1$ ).<sup>1</sup> With the noisy weight estimates, the contiguous mixture model (Figure 1, middle-right) pro-

<sup>1</sup>The noise distribution was cut so that the mixture parameters could not be negative and summed to 1.

duced a test accuracy of 94.2%, the semi-contiguous merged model (Figure 1, bottom-left) produced a test accuracy of 80.2%, and the non-contiguous merged model (Figure 1, bottom-right) produced a test accuracy of 73.4%. Finally, because the integrated merged model (Figure 1, middle-left) did not require estimating mixture weights, it produced a test accuracy of 96.7%.

These simulation results thus allow for the following predictions. If participants are merging categories using the integrated model, then there should be nearly perfect transfer from training to test. However, all the mixture models should produce a transfer cost. The transfer cost for the mixture model is small in the contiguous condition (about 5%), but is substantial when category contiguity is broken. For the semi-contiguous condition, the transfer cost is about 20%; for the non-contiguous condition, the transfer cost was about 26%. Hence, breaking contiguity of the merged categories produces a large transfer cost, with a small difference between semi- and non-contiguous conditions. This suggests an *all or none* type of transfer cost when breaking contiguity with within-category representations.

## Hypotheses

The present experiments test the effects of category contiguity in perceptual space on merging difficulty as a function of category structures and training methodology. Hélie et al. (2017) and Ell et al. (2017) showed that category structures and training methodology interact in determining the type of information that is learned in perceptual categorization. In a typical classification (A/B) experiment, participants are shown a stimulus and asked to assign the stimulus to one of a number of contrasting categories by pressing a response button corresponding to the category. For example, a participant might be asked to press the left button if the animal is a *human* and the right button if the animal is a *dog*. Hélie and colleagues showed that more than half of the participants trained in A/B with rule-based (RB) categories learned between-category information. In contrast, over 70% of the participants trained in A/B with information-integration (II) categories learned within-category information. We thus predict that, with A/B training, the difficulty of merging already learned RB categories will increase with the number of decision bounds that need to be joined with logical connectors. In contrast, the difficulty of merging already learned II categories with A/B categorization will be increased abruptly by the mere presence of a discontinuity.

Experiment 1 tests for these hypotheses by training participants in a 4-category A/B task using either RB or II categories. After training, participants transferred to a condition where they had to merge the 4 training categories into 2 new categories. The new categories could be formed using contiguous (C), semi-contiguous (SC), or non-contiguous (NC) training categories. To anticipate, the results show that,

as predicted, each additional decision bound increased the transfer cost with RB categories. Hence, the C condition was easiest (requiring 1 bound), followed by the SC condition (requiring 2 bounds), and the NC condition was the most difficult (requiring 3 bounds). In contrast, breaking contiguity increased the transfer cost in an all-or-none fashion with II categories. Specifically, the transfer cost for the SC and NC conditions was higher than for the C condition (as with the RB categories), but there was no evidence of a transfer cost difference between the SC and NC conditions (unlike with the RB categories).

Another task popular in perceptual categorization is the YES/NO task. In a typical YES/NO experiment, participants are shown a stimulus with a category label and asked to accept or reject the association by pressing a different response button for *yes* and *no*. For example, a participant might be shown an animal with the label *human* and be asked to press the right button if the animal is a member of the category *human* (i.e., “yes”) or the left button if the animal is not a member of the category *human* (i.e., “no”). Hélie et al. (2017) showed that most participants trained with YES/NO learned within-category information for both RB and II category structures. We thus predict that, with YES/NO training, the difficulty of merging already learned categories will depend on the contiguity of the categories in perceptual space (similar to merging II categories with A/B training). However, with YES/NO training, this prediction should hold for both RB and II categories.

Experiment 2 tests for this hypothesis by reproducing Experiment 1 with the only difference being that participants were trained with YES/NO categorization (same stimuli, categories, and transfer conditions). To anticipate, the results show that, as predicted, the transfer cost was higher for the NC condition than for the C condition, and did not depend on the category structures. However, unlike in Experiment 1, there was no evidence suggesting a difference in transfer cost between the SC condition and either the C or NC condition.

## Experiment 1

Experiment 1 tested the effects of category structures and category contiguity on the compositionality of categories learned using an A/B paradigm. Participants learned four II or RB categories using trial-and-error and then transferred to a two-category task where two new categories were created by merging learned categories. The merged categories at test could be contiguous, semi-contiguous, or non-contiguous.

## Method

**Participants.** One hundred eighty-eight Purdue University undergraduate students were recruited to participate in this experiment. There were two category structures (RB and II) and three testing conditions (C, SC, and NC). Participants were randomly assigned to one of the six combinations

of category structure  $\times$  testing conditions: RB/C ( $n = 32$ ), RB/SC ( $n = 31$ ), RB/NC ( $n = 30$ ), II/C ( $n = 33$ ), II/SC ( $n = 32$ ), and II/NC ( $n = 30$ ). Each participant was given credit for participation as partial fulfillment of a course requirement.

**Material.** The stimuli were lines of various lengths and orientations presented on a 21-inch monitor ( $1,920 \times 1,080$  resolution). Each stimulus was defined in a 2D space by a set of points (length, orientation) where length was calculated in pixels and orientation (counterclockwise rotation from horizontal) was calculated in degrees. The stimuli were generated with the Matlab Psychophysics toolbox (Brainard, 1997) and occupied an approximate visual angle of 5 degrees. Figure 2a shows an example stimulus.

Four categories (arbitrarily labeled “A”, “B”, “C” and “D”) were generated using the randomization technique of F. G. Ashby and Gott (1988). Each category was generated using a bivariate normal distribution. The parameters to generate the RB category structures were as follows (Figure 2b):  $\mu_A = (110, 67)$ ,  $\Sigma_A = \begin{pmatrix} 50 & 0 \\ 0 & 350 \end{pmatrix}$ ;  $\mu_B = (150, 67)$ ,  $\Sigma_B = \Sigma_A$ ;  $\mu_C = (190, 67)$ ,  $\Sigma_C = \Sigma_A$ ;  $\mu_D = (230, 67)$ ,  $\Sigma_D = \Sigma_A$ . To generate the II category structures, we used the following parameters (Figure 2c):  $\mu_A = (122, 88)$ ,  $\Sigma_A = \begin{pmatrix} 646 & 313 \\ 313 & 179 \end{pmatrix}$ ;  $\mu_B = (159, 77)$ ,  $\Sigma_B = \Sigma_A$ ;  $\mu_C = (182, 61)$ ,  $\Sigma_C = \Sigma_A$ ;  $\mu_D = (210, 44)$ ,  $\Sigma_D = \Sigma_A$ . RB categories can be separated using a rule on line length while ignoring the line orientation: the shortest lines are from category “A”, medium-short lines are from category “B”, medium-long lines are from category “C”, and the longest lines are from category “D”. No such verbalizable rule exist for II categories. Perfect accuracy was possible in all conditions.

Twenty-four stimuli were generated from each category for a total of 96 stimuli. The resulting stimuli were re-shuffled at the beginning of each block and each stimulus was presented once in each block. In each trial, a single stimulus was presented in the center of the screen with a question in the center-top of the screen asking a specific categorization question: “X or Y?”, where X and Y stand for one of the category labels used in the experiment. During the training blocks, the category labels were A, B, C, or D. For example, substituting A for X and B for Y would produce the questions “A or B?”. The question indicated the possible choices for the categorization trial. By creating all the possible combinations there were six possible questions. Each question appeared 16 times in each training block. Correct responses for each question were also equally split (e.g., the correct response to half of “A or B?” was “A” and other half “B” and so on). Positive feedback was indicated by the word “Correct” in green font, negative feedback was indicated by the word “Incorrect” in red font, and late responses (i.e., more than 5 seconds) were followed by the words “Too slow!” in black font.

During the test block, two new non-overlapping cate-

gories were formed by merging together two training categories. The new categories were arbitrarily labeled “1” and “2”. In the C condition,  $1 = \{A, B\}$  and  $2 = \{C, D\}$ . In the SC condition,  $1 = \{A, D\}$  and  $2 = \{B, C\}$ . In the NC condition,  $1 = \{A, C\}$  and  $2 = \{B, D\}$ . The test categories are shown in Figure 3 for the RB conditions and in Figure 4 for the II conditions. Trials during the test block were identical to those in the training blocks except that the categorization question was always “1 or 2?”.

Participants responded using a standard keyboard. Key “d” always corresponded to category “A” and key “x” was the category that merged with “A” in the test phase. The keys “k” and “m” were used for the other two categories. Therefore, the key locations depended on the testing condition. The reason for this was to have the response buttons of the categories that were merged together at test be on the same side during training in order to exclude any possible motor effect when comparing different testing conditions. Keys “e” and “i” corresponded to test categories “1” and “2” respectively for all testing conditions. The keyboard configurations for all conditions are shown in Figure 5.

**Procedure.** Each experimental session was composed of five training blocks and one test block. Participants were told that they would be doing a categorization task for six blocks, and that the stimuli were lines varying in length and orientation. They were also told that there are four categories “A”, “B”, “C” and “D” and that on each trial they would see a stimulus and be asked to choose between the two categories mentioned in a question on top of the screen. They were told that the first five blocks would be training blocks in which they receive feedback while the last block would be a test block where they would not receive feedback. Participants were told that they would see instructions on the screen about the test phase after finishing the last training block. The test instructions varied based on the testing condition, but they were all similar and told participants which categories would be merged in the test block. For example, the instructions for the semi-contiguous conditions was: “Categories A and D will form a new category, ‘1’. Categories B and C will form a new category, ‘2’.”

A training trial went as follows: (1) a fixation cross was presented in the center of the screen for 1500 ms; (2) The crosshair disappeared and was replaced by the line stimulus and the question. The stimulus and question stayed on screen until the participant pressed a key corresponding to one of the two categories in the question. (3) After a key was pressed, feedback was presented for 750 ms. Test trials were identical to training trials except that no feedback was presented.

## Results

A binomial distribution was used to identify and exclude participants who performed randomly during the last training block (i.e., non-learners). The rationale was that participants

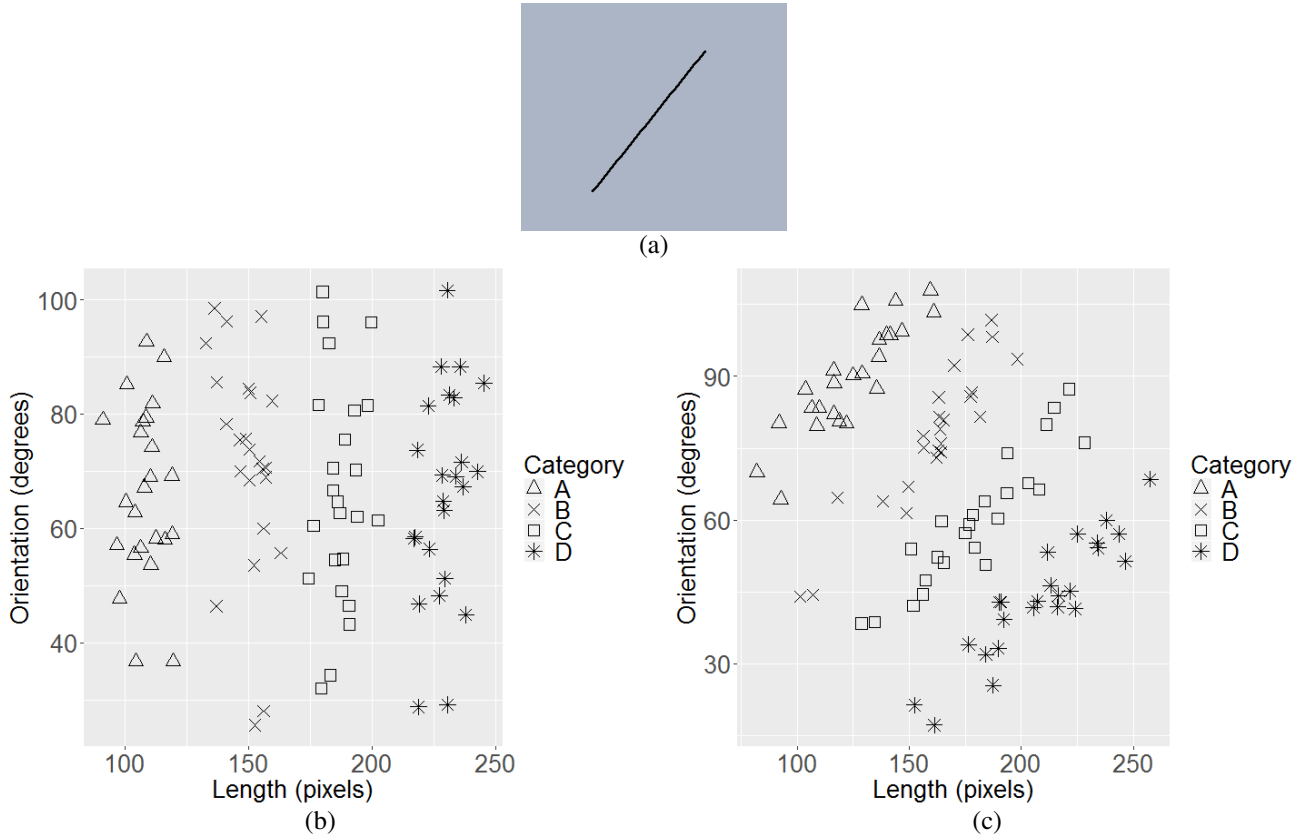


Figure 2. Stimuli used in the experiments. (a) An example stimulus. (b) RB category structures used in Experiments 1 and 2. (c) II category structures used in Experiments 1 and 2. Symbols in panels (b) and (c) denote different categories.

who did not learn the training categories should not be able to merge the training categories (which is the main goal of the experiment). Specifically, we excluded participants whose accuracy in Block 5 was not above chance ( $p < .05$ ) according to a binomial distribution ( $p = 0.5$ ,  $n = 96$ ).<sup>2</sup> This corresponded to a 59% accuracy threshold. Using this threshold, 31 participants were excluded (16.5% of the sample), and 157 participants remained in the analysis, with at least 25 participants left in each condition (see Figure 6 for exact counts per condition).

**Learning phase.** Figure 6 shows the mean accuracy for each block for each testing condition. The left panel (a) shows the RB categories while the right panel (b) shows the II categories. In both panels, the first five blocks were training and the last block was the testing block. A  $2$  (RB, II)  $\times$   $3$  (C, SC, NC)  $\times$   $5$  (Block) mixed effect ANOVA was performed on the training data. As expected, the main effect of Block was statistically significant ( $F(4, 628) = 144.11$ ,  $p < .001$ ,  $\eta^2 = 0.18$ ), showing that participants were able to learn the task. The effect of Category was also significant ( $F(1, 157) = 70.51$ ,  $p < .001$ ,  $\eta^2 = 0.19$ ), showing that participants were more accurate with RB categories than II categories. However, these main effects need to be interpreted

with care since the Category  $\times$  Block interaction also reached statistical significance ( $F(4, 628) = 7.29$ ,  $p < .001$ ,  $\eta^2 = 0.01$ ). The interaction was decomposed by computing the effect of Block within each level of Category. The results show that the effect of Block reached statistical significance for both RB ( $F(4, 328) = 91.70$ ,  $p < .001$ ,  $\eta^2 = 0.27$ ) and II ( $F(4, 300) = 54.09$ ,  $p < .001$ ,  $\eta^2 = 0.18$ ) categories, confirming that participants were able to learn both category structures. The interaction was thus likely caused by a larger increase in accuracy with the RB categories than with the II categories. Mean accuracy in Block 1 with RB categories was 70.9%, which improved to 88.2% in Block 5. For II categories, mean accuracy in Block 1 was 64.8%, which improved to 76.2% in Block 5. All other main effects and interactions failed to reach statistical significance (all  $F < 1.54$ ,  $n.s.$ ).

**Testing phase.** The main goal of this experiment was to test whether participants could merge learned categories together to form new categories. The transfer cost was calculated as the difference in accuracy between Blocks 5 and 6 and is shown in Figure 7. Again, the left panel (a) shows

<sup>2</sup>On any given trial, participants chose one of two response buttons so chance performance was 0.50, and each block had 96 trials.

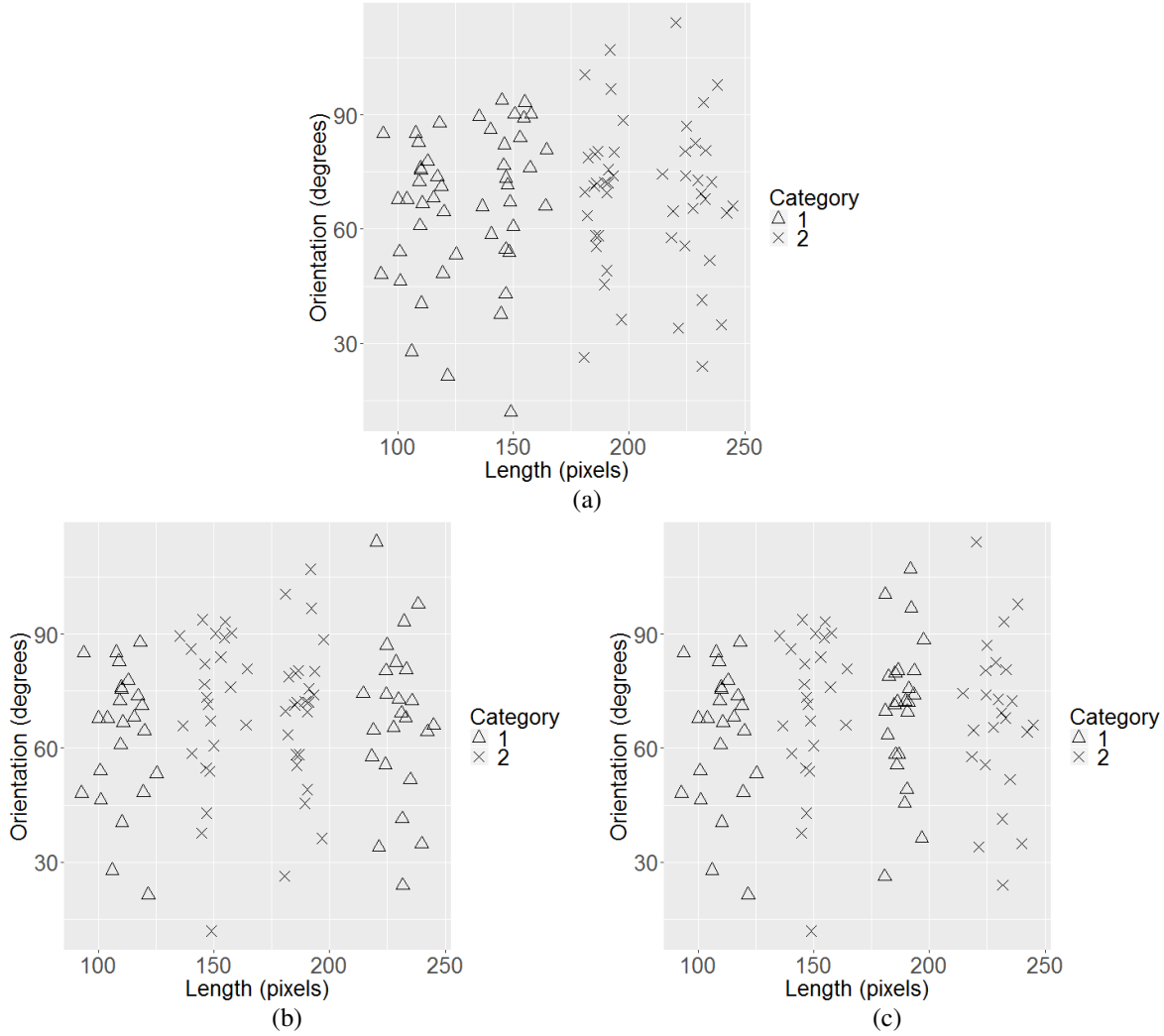
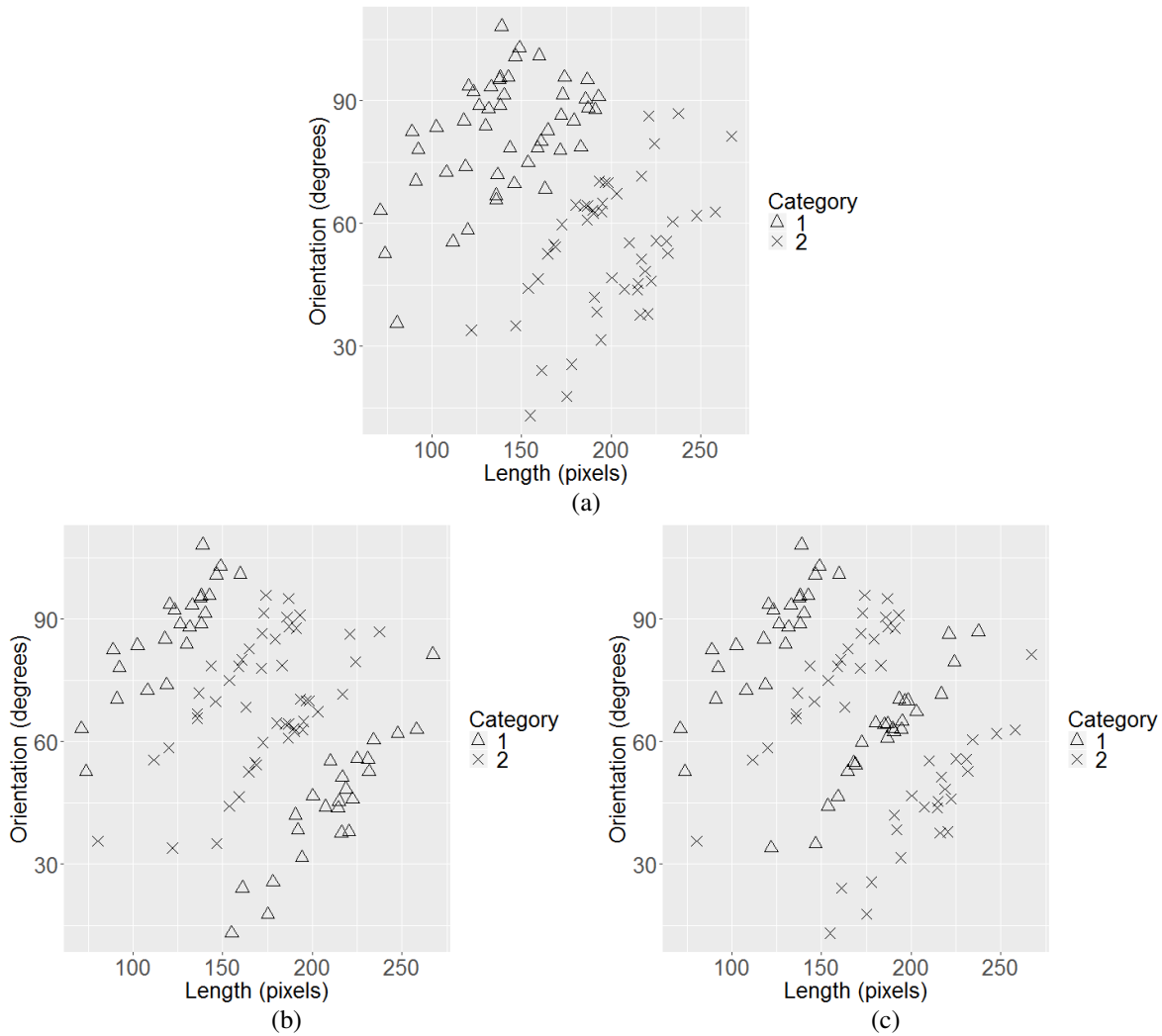


Figure 3. Test categories in the RB conditions. (a) RB/C: “A” and “B” formed category “1” and “C” and “D” formed category “2”. (b) RB/SC: “A” and “D” formed category “1” and “B” and “C” formed category “2”. (c) RB/NC: “A” and “C” formed category “1” and “B” and “D” formed category “2”.

the RB categories whereas the right panel (b) shows the II categories. A  $2 \text{ (RB, II)} \times 3 \text{ (C, SC, NC)}$  ANOVA was performed on the transfer cost. Both the effects of testing condition ( $F(2, 151) = 111.97, p < .001, \eta^2 = 0.57$ ) and category ( $F(1, 151) = 4.39, p < .05, \eta^2 = 0.01$ ) reached statistical significance. However, the main effects need to be interpreted in the context of a statistically significant interaction ( $F(2, 151) = 6.29, p < .01, \eta^2 = 0.03$ ). We proceeded by decomposing the effect of testing condition within each level of Category. For RB categories, the effect of testing condition was statistically significant ( $F(2, 79) = 33.76, p < .001, \eta^2 = 0.46$ ). Bonferroni-corrected pairwise comparisons show that all pairwise differences were statistically significant ( $p < .001$ ). The mean transfer costs were: C = 00.0%; SC = 11.4%; and NC = 21.6%. For II categories, the effect of testing condition also reached statistical signifi-

cance ( $F(2, 72) = 106.00, p < .001, \eta^2 = 0.75$ ). Again, Bonferroni-corrected pairwise comparisons show that the C condition differs from both the SC and NC conditions ( $p < .001$ ). However, unlike for RB categories, there was no statistical difference between the SC and NC conditions. The mean transfer costs were: C = -8.4%; SC = 14.3%; and NC = 18.4%.

Next, a  $t$ -test was performed to assess whether the transfer cost was statistically different from zero in each testing condition of each category. For RB categories, the transfer cost was not statistically significant in the RB/C condition ( $t(24) = 0.09, n.s.$ ), but reached statistical significance for both the RB/SC and RB/NC conditions (both  $t > 6.83, p < .001$ ). For II categories, all transfer costs were statistically different from zero (all  $|t| > 6.00, p < .001$ ), but note that this difference is negative for the C condition, showing a fa-



*Figure 4.* Test categories in the II conditions. (a) II/C: “A” and “B” formed category “1” and “C” and “D” formed category “2”. (b) II/SC: “A” and “D” formed category “1” and “B” and “C” formed category “2”. (c) II/NC: “A” and “C” formed category “1” and “B” and “D” formed category “2”.

cilitation effect instead of a cost. In contrast, the transfer costs were negative for the SC and NC conditions, showing a true cost of merging categories (similar to RB categories). Hence, breaking contiguity had a transfer cost for both RB and II categories, but the cost was progressive for RB categories and all-or-none for II categories.

## Discussion

The results of Experiment 1 show no evidence of a transfer cost for C conditions with either RB or II category structures. One surprising result is that there was facilitation when merging II categories. It is possible that participants averaged the distributions of the merged categories and used a single integrated distribution for the “1” category and another single integrated distribution for the “2” category (instead of form-

ing mixture models – see middle-left of Figure 1). No increase in accuracy was observed in the simulations of this model because of a ceiling effect in training accuracy, but if training accuracy is reduced by biasing the estimated means of the training generative models the integrated model does produce a higher test accuracy. It is thus possible that participants in the contiguous II condition used this response strategy at test. Note that this “single integrated distribution” strategy is only possible with within-category information, so it was unlikely to be used with RB categories, which could explain why no facilitation was observed in the RB/C condition. This result was unexpected and the experiment was not designed to test for this possibly. Still, clearly, there was no transfer cost for both RB and II categories.

In contrast, a transfer cost was present for all other conditions. Critically, the SC condition was differently affected



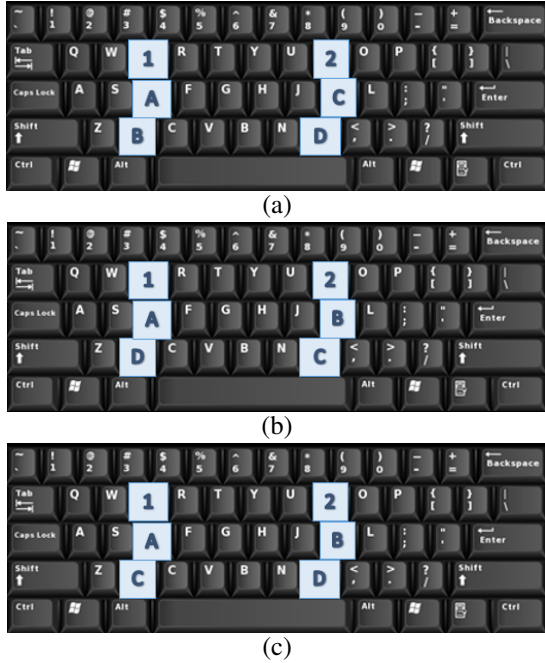


Figure 5. Category labels on the keyboard of Experiment 1 for (a) RB/C and II/C, (b) RB/SC and II/SC and (c) RB/NC and II/NC.

by the category structures. Specifically, the SC condition differed from both the C and the NC conditions with RB category structures, with a transfer cost falling somewhere between these two conditions. This result is in line with the hypothesis that participants learn between-category information in A/B with RB categories (Hélie et al., 2017), so the transfer cost increase with the number of decision bounds that needs to be assembled. In contrast, there was no evidence of a different transfer cost between the SC and NC conditions with II category structures. This suggests that, when trained with an A/B paradigm, category contiguity may be an all-or-none phenomenon with II category structures because participants are learning a within-category representation and are forming mixture models of the generating distributions (at least when the merged categories are not fully contiguous). Experiment 2 tested whether these effects were also present with YES/NO training.

### Experiment 2

Experiment 2 tested the effects of category structures and category contiguity on the compositionality of categories learned using a YES/NO task. Experiment 1 showed that transfer cost increased gradually with the required number of decision bounds with RB categories (consistent with between-category representations) but that this increase was all-or-none with II categories (consistent with within-category representations). However, Hélie et al. (2017)

showed that, unlike A/B categorization, YES/NO categorization leads to learning within-category information with both RB and II categories. Hence, Experiment 2 tests whether breaking the contiguity of the training categories at test would increase the transfer cost in an all-or-none fashion. As in Experiment 1, participants learned four II or RB categories using trial-and-error and then transferred to a two-category task where two new categories were created by merging learned categories. The merged categories at test could be contiguous, semi-contiguous, or non-contiguous. The only difference between Experiments 1 and 2 is that Experiment 2 used YES/NO training instead of A/B training.

### Method

**Participants.** One hundred eighty-one participants were recruited from the Purdue University undergraduate population to participate in this experiment. As in Experiment 1, there were two category structures (RB and II) and three testing conditions (C, SC, and NC). Participants were randomly assigned to one of the six combinations of category structure and testing conditions: RB/C ( $n = 32$ ), RB/SC ( $n = 29$ ), RB/NC ( $n = 30$ ), II/C ( $n = 31$ ), II/SC ( $n = 30$ ), and II/NC ( $n = 29$ ). None of the participant participated in Experiment 1. Each participant was given credit for participation as partial fulfillment of a course requirement.

**Material.** The stimuli and category structures were the same as those used in Experiment 1 (see Figures 2–4). The only differences were the question included at the top of the screen and the response keys. In the YES/NO task, the question asked about category membership: “Is this a X?”, where X is replaced by one of the categories (A, B, C, D). For example, by substituting B for X, the question would be “Is this a B?”. The participant responded YES by pressing the ‘d’ key or NO by pressing the ‘k’ key (sticker labeled). During training, one quarter of the trials asked about category “A”, another quarter asked about category “B”, etc. For each question, the correct response was YES in half the trials and NO in the other half. At test, half the trials asked about category “1” and the other half asked about category “2”. Again, the correct response to each test question was YES on half the trials and NO on the other half. The same YES and NO response keys were used at test and at training. The key configuration for Experiment 2 is shown in Figure 8.

**Procedure.** The procedure was the same as Experiment 1.

### Results

The same procedure as in Experiment 1 was used to exclude participants who failed at learning the task. Fifty participants were excluded (27.6% of the sample), and 131 participants remained in the analysis. The exact number of participants left in each condition is shown in Figure 9.

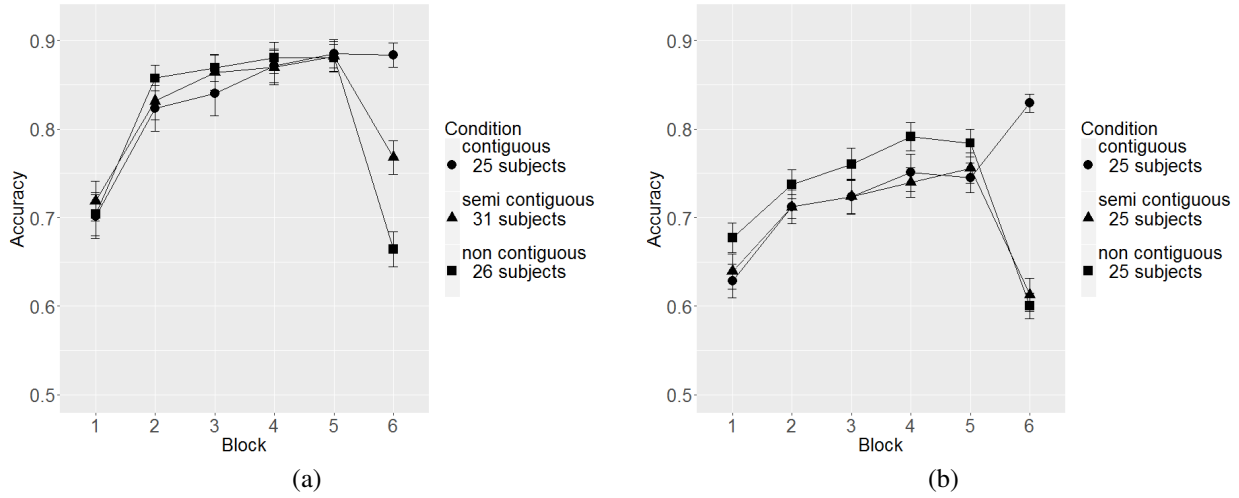


Figure 6. Mean accuracy per block in Experiment 1. (a) RB categories; (b) II categories. In both panels, Blocks 1–5 are the training phase and Block 6 is the test phase. Error bars are between-subject standard error of the mean.

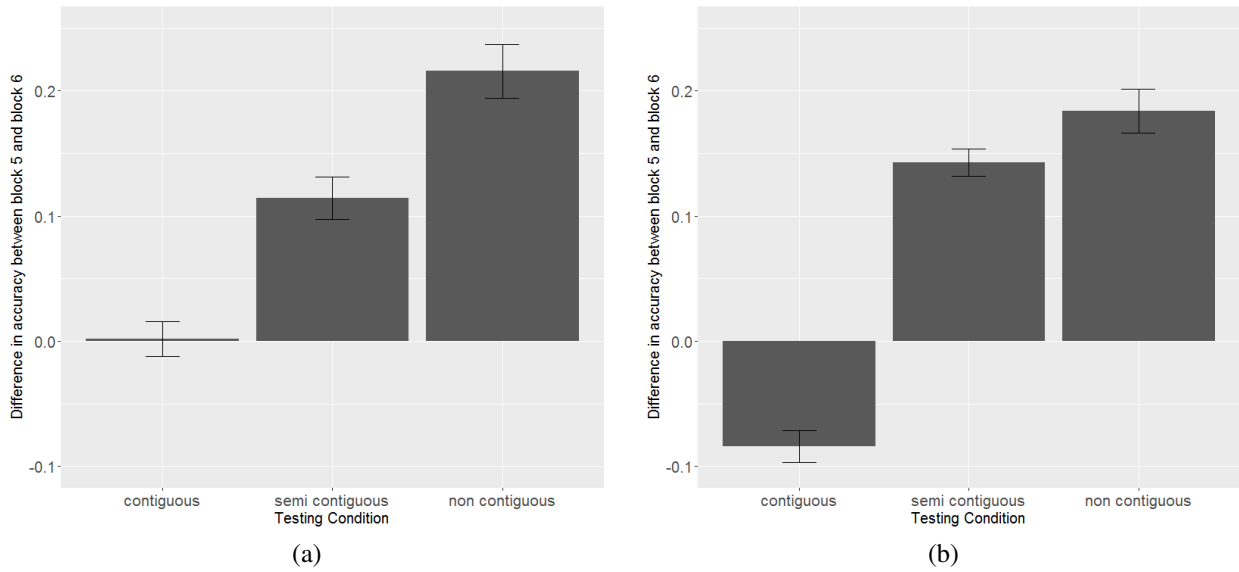


Figure 7. Accuracy differences between Blocks 5 and 6 (transfer cost) in each testing conditions in Experiment 1. (a) RB categories; (b) II categories. Error bars are between-subject standard error of the mean.



Figure 8. Category labels on keyboard in Experiment 2.

**Learning phase.** Figure 9 shows the mean accuracy for each block in each testing condition. The left panel (a) shows the RB categories while the right panel (b) shows the II categories. As in Experiment 1, the first five blocks

were training and the last block was the testing block. A  $2 \text{ (RB, II)} \times 3 \text{ (C, SC, NC)} \times 5 \text{ (Block)}$  mixed effect ANOVA was performed on the training data. As in Experiment 1, the main effect of Block was statistically significant ( $F(4, 524) = 104.12, p < .001, \eta^2 = 0.19$ ), showing that participants were able to learn the task. The effect of Category was also significant ( $F(1, 131) = 27.22, p < .001, \eta^2 = 0.01$ ), showing that participants were again more accurate with RB categories than II categories. However, these main effects need to be interpreted with care since the Category  $\times$  Block interaction again reached statistical significance ( $F(4, 524) = 3.84, p < .01, \eta^2 = 0.01$ ). The interaction was decomposed by computing the effect of Block

within each level of Category. The results show that, similar to Experiment 1, Block had a statistically significant effect for both RB ( $F(4, 292) = 74.98, p < .001, \eta^2 = 0.30$ ) and II ( $F(4, 232) = 36.42, p < .001, \eta^2 = 0.16$ ) categories. The interaction was likely caused by a larger improvement in accuracy for RB, which began with a mean accuracy of 67.3% (Block 1) and ended up with a mean accuracy of 86.4% (Block 5). For II categories, Block 1 accuracy was 63.4% and increased to 78.2% in Block 5. As in Experiment 1, none of the other main effects and interactions reached statistical significance (all  $F < 1.49, n.s.$ ).

**Testing phase.** The transfer cost for each testing condition is shown in Figure 10. The left panel (a) shows the RB categories while the right panel (b) shows the II categories. A  $2 \text{ (RB, II)} \times 3 \text{ (C, SC, NC)}$  ANOVA was performed on the transfer cost. As in Experiment 1, the effect of testing condition reached statistical significance ( $F(2, 125) = 4.98, p < .01, \eta^2 = 0.07$ ), showing that transfer cost was affected by the contiguity of the merged categories. However, unlike in Experiment 1, the effect of categories ( $F(1, 125) = 0.54, n.s., \eta^2 = 0.00$ ) and the interaction between the factors ( $F(2, 125) = 1.54, n.s., \eta^2 = 0.02$ ) both failed to reach statistical significance. Bonferroni-corrected pairwise comparisons show a statistically significant difference between the C and the NC conditions ( $p < .01$ ). All other pairwise comparisons failed to reach statistical significance. The mean transfer costs were: C = 8.1%; SC = 13.1%; and NC = 18.7%. The transfer cost was statistically significant for all testing conditions (all  $t > 3.54, p < .001$ ).

## Discussion

The results in Experiment 2 suggest that contiguous categories learned with the YES/NO task are more difficult to merge into new categories than contiguous categories learned with the A/B task. For the YES/NO task, even the C condition showed a significant transfer cost. This suggests that the “single integrated distribution” strategy was not used in this case. However, as predicted, there was no interaction between contiguity and categories. There was thus no evidence of a differential effect of contiguity on RB and II categories. A more detailed comparison of the two experiments is described in the following section.

### Comparing Experiments 1 and 2

While Experiments 1 and 2 were separate experiments and any direct comparison needs to be interpreted with care, comparing the experiment results may still be informative to better understand the effect of training task on transfer cost. In this section, separate Training paradigm (A/B vs. YES/NO)  $\times$  Testing condition (C, SC, NC) ANOVAs were computed for the RB and II transfer costs.

For RB categories, the effect of testing condition was statistically significant ( $F(2, 149) = 21.60, p < .001, \eta^2 =$

0.22), with a mean transfer cost of C = 4.45%, SC = 10.44%, and NC = 20.78%. This effect is not surprising given that the oneway ANOVA on transfer cost in both experiments showed a statistically significant effect. In contrast, the effect of training paradigm did not reach statistical significance ( $F(1, 149) = 0.51, n.s., \eta^2 = 0.00$ ), and the interaction between the factors was trending but was also not statistically significant ( $F(2, 149) = 2.87, p < .10, \eta^2 = 0.03$ ). Given the exploratory nature of these analyses, we decomposed the trending interaction to compute the effect of training paradigm in each level of testing condition. The results show a statistically significant effect of training paradigm in the C testing condition ( $F(1, 50) = 4.42, p < .05, \eta^2 = 0.08$ ). For the C testing condition, the mean transfer cost was 00.0% for A/B and 8.4% for YES/NO. This suggests that participants may have used a single decision criterion in the A/B task (i.e., the criterion learned at training between the B and C categories) but used a mixture model in the YES/NO task (which produced a transfer cost). In contrast, the effect of training paradigm did not have a statistically significant effect in the SC or NC conditions (both  $F < 0.57, n.s.$ ).

For II categories, both the effect of testing condition ( $F(2, 127) = 42.05, p < .001, \eta^2 = 0.36$ ) and the effect of training paradigm ( $F(1, 127) = 8.34, p < .01, \eta^2 = 0.04$ ) reached statistical significance. However, these main effects need to be interpreted in the context of a statistically significant interaction ( $F(2, 127) = 7.00, p < .01, \eta^2 = 0.06$ ). As with RB categories, decomposing the interaction to compute the effect of training paradigm within each level of testing condition shows a statistically significant effect of training paradigm in the C condition ( $F(1, 39) = 15.31, p < .001, \eta^2 = 0.28$ ). For the C testing condition, the mean transfer cost was -8.4% for A/B and 7.6% for YES/NO. This difference appears to be linked to the possibility of using the “single integrated distribution” strategy in the A/B condition but not in the YES/NO condition. Again, the effect of training paradigm did not have a statistically significant in the SC or NC conditions (both  $F < 0.88, n.s.$ ). These analyses further support the previous interpretation that training methodology (i.e., A/B vs. YES/NO) mostly affected transfer costs in the C conditions.

## General discussion

This article presents the results of two experiments exploring the effect of training methodology and category structure on the ability to merge already known categories. We hypothesized that the ability to merge categories (i.e., compositionality) would depend on the type of knowledge contained in the categorical representation (Ell et al., 2017; Hélie et al., 2017). Specifically, between-category representations can be connected using logical operators and therefore each additional discontinuity in the merged category would require additional connectors, which would tax working memory and

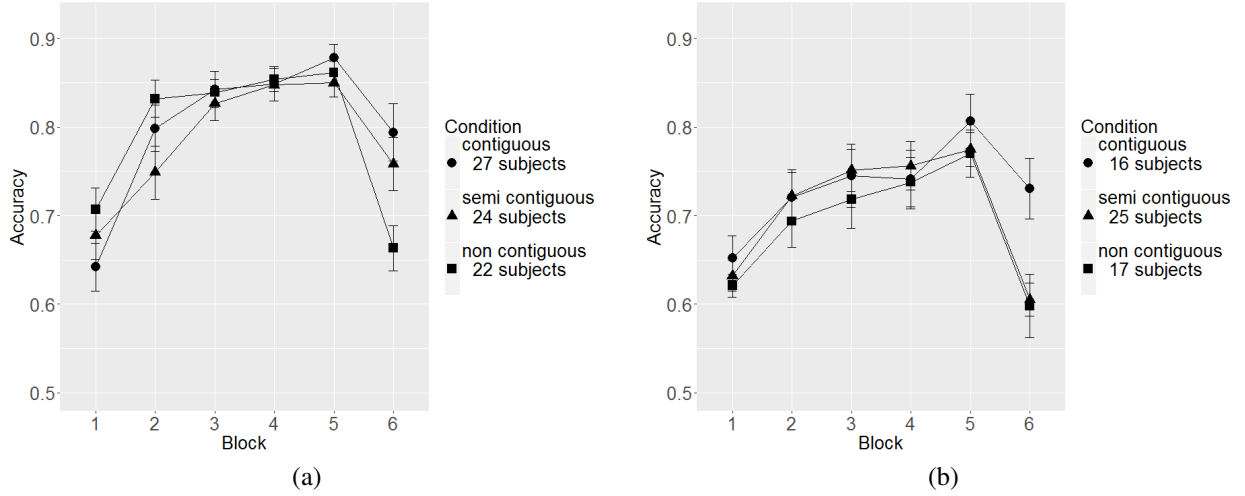


Figure 9. Mean accuracy per block in Experiment 2. (a) RB categories; (b) II categories. In both panels, Blocks 1–5 are the training phase and Block 6 is the test phase. Error bars are between–subject standard error of the mean.

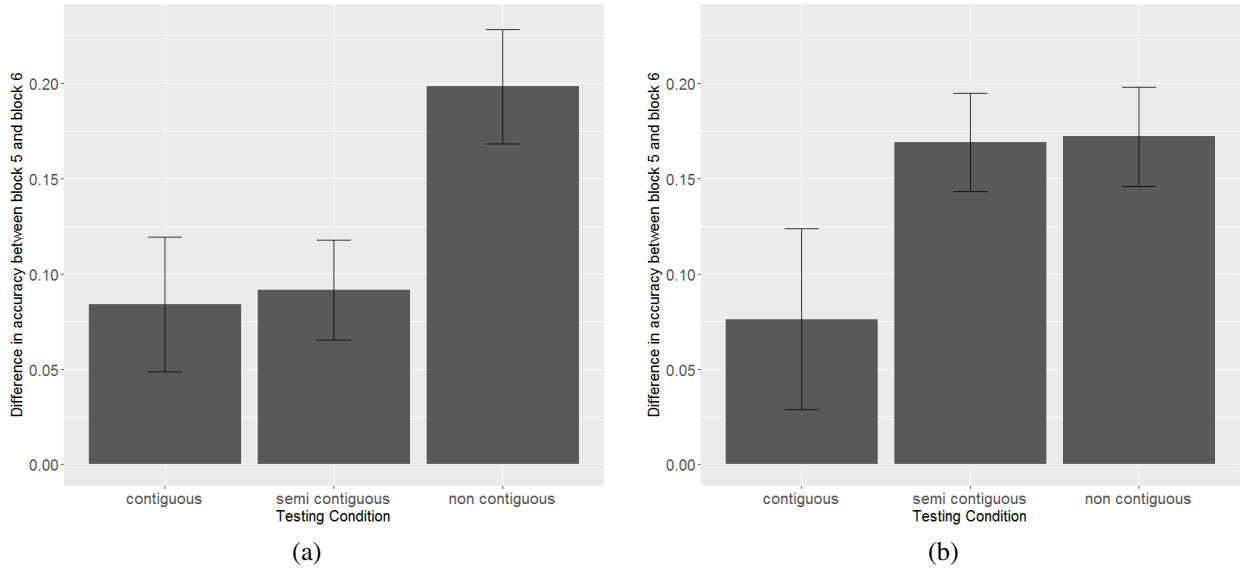


Figure 10. Accuracy differences between Blocks 5 and 6 (transfer cost) in each testing conditions in Experiment 2. (a) RB categories; (b) II categories. Error bars are between–subject standard error of the mean.

gradually increase transfer cost. In contrast, within–category representations could be grouped using generative mixture models, and simulation results suggest that breaking the continuity of the merged categories would significantly increase the transfer cost, but additional breaks in continuity would only marginally increase transfer cost. Hélie et al. (2017) showed that learning RB categories with A/B training would produce a between–category representation, and we thus expected that the contiguity of the merged categories would have a progressive effect on transfer cost. However, learning II categories with A/B training, or learning both RB and II categories with YES/NO training, would produce within–category representations, and thus a break in the contiguity

of the merged category would produce an abrupt increase in transfer cost.

The results of the experiments were largely consistent with the above hypotheses. First, Experiment 1 showed that with A/B training, each added level of discontinuity in the merged categories increased the transfer cost with RB categories. With II categories, the SC and NC conditions produced a larger transfer cost than the contiguous condition, but we found no evidence of a difference in transfer cost between the SC and NC conditions. Second, Experiment 2 showed that with YES/NO training, the NC condition produced a larger transfer cost than the contiguous condition. However, there was no evidence that transfer cost in the SC condition

differed from the other two, and there was also no evidence of an effect of category structure on transfer cost. Together, these results are consistent with earlier results showing that A/B training produces different category representations depending on the category structures whereas YES/NO training produces the same type of category representation for RB and II structures (Ell et al., 2017; Hélie et al., 2017).

Another interesting result follows from the comparison of the experiments. The effect of training condition was mostly relevant for the contiguous condition. Specifically, contiguous categories were easier to merge with A/B training than with YES/NO training. For both tested category structures, contiguous categories could be merged without a significant transfer cost with A/B training, whereas a transfer cost was always present with YES/NO training. This main effect of task could be linked to the response modality of the tasks. For participants trained with A/B, the training task used 4 response buttons whereas the test used only 2 response buttons. This made some aspect of the test easier, which could compensate for the added difficulty of merging categories. In contrast, the number of response buttons did not change between training and test for the YES/NO task, so overall the test phase may be more difficult. Another possible explanation is that the A/B and YES/NO tasks likely rely on different cognitive mechanisms and brain circuits. For example, the A/B task has been shown to rely on a circuit centered around the ventrolateral prefrontal cortex for RB categories (Hélie, Roeder, & Ashby, 2010) and a different circuit centered around the sensorimotor striatum for II categories (Waldschmidt & Ashby, 2011). While much less is known about the brain circuit supporting YES/NO learning, there is evidence that it differs from the brain circuit used for A/B training (Zeithamova, Maddox, & Schnyer, 2008), which could explain the difference in RB representations (Hélie et al., 2017; Ell et al., 2017) and merging difficulty for contiguous categories. However, this result was unexpected and the present experiments were not design to directly test these possibilities. Future research should attempt to directly test these two competing hypotheses.

### Implications for research on category learning

Beyond the effects of category structure and training methodology, the experiments included in this article clearly show that, in perceptual categorization, participants can merge already known categories when needed. Given the ubiquity of categorization in everyday life, and the large number of known categories, it appears very unlikely that adults engage in learning new categories without restructuring existing categorical knowledge. Thus, the process of merging learned category representations is likely closer to real-life category learning. Indeed, much empirical and theoretical work has been devoted to the investigation of how natural categories are combined (Cohen & Murphy,

1984; Prinz, 2012; E. E. Smith et al., 1988; Wisniewski, 1997). This work has focused on the merging of within-category representations of well-learned, natural categories (e.g., merging the categories *red* and *apple* to form the category *red apple*). The present research extends this work to perceptual categorization and emphasizes how different factors during learning may affect the combination of categories and suggests that multiple kinds of category representations demonstrate characteristics of compositionality.

The present methodological approach may prove useful in investigating the neural substrates of combining category representations. Much is known about the neural substrates mediating learning in rule-based and information-integration tasks (e.g., F. G. Ashby & Ell, 2001; Hélie et al., 2010; Seger & Miller, 2010; Waldschmidt & Ashby, 2011). Very little research, however, has investigated the neural substrates of merging category representations. At least with between-category representations, knowledge can be reorganized into higher-order, hierarchical representations. For instance, merging a set of stimulus-response rules into a common, superordinate category recruits regions of prefrontal cortex rostral to the subregions of prefrontal cortex that learned the original stimulus-response rules (Badre, Kayser, & D'Esposito, 2010). Although a similar neural substrate may support the merging of discontinuous between-category representations, it is unclear if such a rostro-caudal distinction would be the neural substrate of knowledge reorganization with within-category representations. Thus, the current approach may provide a useful method for comparing and contrasting knowledge reorganization supported by between-category differences versus knowledge reorganization supported by within-category similarities.

### Limitations and future work

One important limitation of the present experiments is that merging cost was only measured for one transfer block. It is possible that re-organization of existing takes longer than the duration of the included test block and re-organisation was still ongoing. As a result, the differences observed between the different tasks and conditions may be transient and a longer test condition could reduce these differences. Future research should focus on increasing the length of test and observe its effect on transfer cost.

Another limitation of the experiments is that they explore the general effects of between- and within-category representation on compositionality, but there is likely more than one type of within-category representation (same with between-category representation). For example, both prototype models (J. D. Smith & Minda, 2002) and exemplar models (Nosofsky, 1986) are treated as producing within-category representations in the current framework because they are generative models (F. Ashby & Maddox, 1993). However, the decision rule applied to the representation

is different, which could result in different merging performance. The present experiments were not designed to address these subtler differences, and future experiments are needed to test how homogeneous (or not) the different within-category representations are and how they merge.

Finally, an emphasis on how category learning influences compositionality would increase the ecological validity of categorization research and be more useful for understanding categorization outside of the laboratory. For instance, research on the merging of category representations could inform best practices for training challenging conceptual relations in physics and math education (e.g., Heckler, 2011). The work included in this article provides an initial step, but much more work is needed to identify the factors that facilitate the re-organization of categorical knowledge in useful new ways that allow for generalization and transfer.

### Compliance with ethical standards

The authors declare that they have no conflict of interest. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants included in the study.

### Acknowledgments

This research was supported in part by NSF grants #1349677-BCS and #1349737 to SH and SWE (respectively).

### References

- Aerts, D., Gabora, L., & Sozzo, S. (2013). Concepts and Their Dynamics: A Quantum-Theoretic Modeling of Human Thought. *Topics in Cognitive Sciences*, 5, 737–772.
- Ashby, F., & Maddox, W. T. (1993). Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology*, 37, 372–400.
- 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.
- 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., & Maddox, W. T. (2010). Human category learning 2.0. *Annals of the New York Academy of Sciences*, 1224, 147–161.
- Badre, D., Kayser, A. S., & D'Esposito, M. (2010). Frontal Cortex and the Discovery of Abstract Action Rules. *Neuron*, 66, 315–326.
- Bishop, C. (2006). *Pattern Recognition and Machine Learning*. Singapore: Springer.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10, 433–436.
- Cohen, B., & Murphy, G. L. (1984). Models of concepts. *Cognitive Science*, 8, 27–58.
- Ell, S. W., Smith, D. B., Peralta, G., & Hélie, S. (2017). The impact of category structure and training methodology on learning and generalizing within-category representations. *Attention, Perception, & Psychophysics*, 79, 1777–1794.
- Erev, I. (1998). Signal detection by human observers: A cutoff reinforcement learning model of categorization and decisions under uncertainty. *Psychological Review*, 105, 280–298.
- Erickson, M. A. (2008, jun). Executive attention and task switching in category learning: Evidence for stimulus-dependent representation. *Memory & Cognition*, 36(4), 749–761. doi: 10.3758/MC.36.4.749
- Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General*, 127(2), 107.

- Feldman, J. (2003). A catalog of Boolean concepts. *Journal of Mathematical Psychology*, 47, 75–89.
- Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28, 3–71.
- Heckler, A. F. (2011). The Ubiquitous Patterns of Incorrect Answers to Science Questions: The Role of Automatic, Bottom-up Processes. *The Psychology of Learning and Motivation*, 55, 227–267.
- Hélie, S., & Ashby, F. G. (2012). Learning and transfer of category knowledge in an indirect categorization task. *Psychological Research*, 76, 292–303.
- Hélie, S., Roeder, J. L., & Ashby, F. G. (2010, oct). Evidence for cortical automaticity in rule-based categorization. *Journal of Neuroscience*, 30(42), 14225–14234. doi: 10.1523/JNEUROSCI.2393-10.2010
- Hélie, S., Shamloo, F., & Ell, S. W. (2017). The effect of training methodology on knowledge representation in categorization. *PLOS ONE*, 12, e0183904.
- Hélie, S., Waldschmidt, J. G., & Ashby, F. G. (2010). Automaticity in rule-based and information-integration categorization. *Attention, Perception, & Psychophysics*, 72(4), 1013–1031.
- Levering, K. R., & Kurtz, K. J. (2015). Observation versus classification in supervised category learning. *Memory & Cognition*, 43(2), 266–282. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/25190494> doi: 10.3758/s13421-014-0458-2
- ling Wu, L., & Barsalou, L. W. (2009). Perceptual simulation in conceptual combination: Evidence from property generation. *Acta Psychologica*, 132(2), 173–189. (Spatial working memory and imagery: From eye movements to grounded cognition) doi: <https://doi.org/10.1016/j.actpsy.2009.02.002>
- Markman, A. B. (2002). Stimulus categorization. In D. L. Pashler H. & Medin (Ed.), *Stevens' handbook of experimental psychology* (3rd ed., Vol. 2: Memory and cognitive processes, pp. 165–208). New York: John Wiley & Sons.
- Markman, A. B., & Ross, B. (2003). Category use and category learning. *Psychological Bulletin*, 129, 529–613.
- Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, 85(3), 207–238.
- Miles, S. J., & Minda, J. P. (2011). The effects of concurrent verbal and visual tasks on category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(3), 588–607. Retrieved from <http://doi.apa.org/getdoi.cfm?doi=10.1037/a0022309> doi: 10.1037/a0022309
- Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, 115, 39–57.
- Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, 77, 353–363.
- Prinz, J. J. (2012). Regaining composure: A defense of prototype compositionality. In M. Werning, W. Hinzen, & E. Machery (Eds.), *The oxford handbook of compositionality* (pp. 437–453). New York: Oxford University Press.
- Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology*, 3, 382–407.
- Seger, C. A., & Miller, E. K. (2010). Category Learning in the Brain. *Annual Review of Neuroscience*, 33, 203–219.
- Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs: General and Applied*, 75, 1–42.
- Smith, E. E., Osherson, D. N., Rips, L. J., & Keane, M. (1988). Combining prototypes: A selective modification model. *Cognitive Science*, 12, 485–527.
- 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.
- Treisman, M., & Williams, T. C. (1984). A theory of criterion setting with an application to sequential dependencies. *Psychological Review*, 91, 68–111.
- Voorspoels, W., Storms, G., & Vanpaemel, W. (2012). An exemplar approach to conceptual combination. *Psychologica Belgica*, 52, 435–458.
- Waldschmidt, J. G., & Ashby, F. G. (2011). Cortical and striatal contributions to automaticity in information-integration categorization. *Neuroimage*, 56(3), 1791–1802.
- Wisniewski, E. J. (1997). When concepts combine. *Psychonomic Bulletin & Review*, 4, 167–183.
- Zadeh, L. A. (1982). A note on prototype theory and fuzzy sets. *Cognition*, 12, 291–297.
- Zeithamova, D., Maddox, W. T., & Schnyer, D. D. M. (2008, dec). Dissociable prototype learning systems: evidence from brain imaging and behavior. *Journal of Neuroscience*, 28(49), 13194–13201. doi: 10.1523/JNEUROSCI.2915-08.2008