



Cognitive Science 46 (2022) e13128

© 2022 The Authors. *Cognitive Science* published by Wiley Periodicals LLC on behalf of Cognitive Science Society (CSS).

ISSN: 1551-6709 online

DOI: 10.1111/cogs.13128

A Computational Model of Context-Dependent Encodings During Category Learning

Paulo F. Carvalho,^a Robert L. Goldstone^b

^a*Human–Computer Interaction Institute, Carnegie Mellon University*

^b*Department of Psychological and Brain Sciences, Cognitive Science Program, Indiana University*

Received 8 January 2021; received in revised form 28 December 2022; accepted 7 March 2022

Abstract

Although current exemplar models of category learning are flexible and can capture how different features are emphasized for different categories, they still lack the flexibility to adapt to local changes in category learning, such as the effect of different sequences of study. In this paper, we introduce a new model of category learning, the Sequential Attention Theory Model (SAT-M), in which the encoding of each presented item is influenced not only by its category assignment (global context) as in other exemplar models, but also by how its properties relate to the properties of temporally neighboring items (local context). By fitting SAT-M to data from experiments comparing category learning with different sequences of trials (interleaved vs. blocked), we demonstrate that SAT-M captures the effect of local context and predicts when interleaved or blocked training will result in better testing performance across three different studies. Comparatively, ALCOVE, SUSTAIN, and a version of SAT-M without locally adaptive encoding provided poor fits to the results. Moreover, we evaluated the direct prediction of the model that different sequences of training change what learners encode and determined that the best-fit encoding parameter values match learners' looking times during training.

Keywords: Category learning models; Sequencing; Interleaving; Attention; Encoding

Some of this work was part of a doctoral dissertation submitted by the first author to Indiana University as partial fulfillment of the degree of Doctor of Philosophy. The authors thank Rob Nosofsky, Rich Shiffrin, and Linda Smith for helpful discussion and suggestions.

The model code and fitting procedures are available at: <https://osf.io/q782h/>

Correspondence should be sent to Paulo F. Carvalho, Human–Computer Interaction Institute, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15213, USA. E-mail: pcarvalh@cs.cmu.edu

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

1. Introduction

Our categorization decisions are often based on a subset of potential features. For example, when at the supermarket trying to pick the perfectly ripe avocado for dinner, we look for a softer, darker avocado. Although avocados can differ on many properties, color seems to be particularly relevant for ripeness. Not all features are equally important, and humans are particularly good at discerning what matters from what does not. Moreover, discerning the features that matter is context dependent. For example, when looking for a book among bowls, shape is a critical feature. However, when looking for a particular book among other books that is unlikely to be the case. How do we flexibly learn what to attend to and encode?

Extant theories and models of human and artificial category learning try to answer this question by taking the attentional flexibility of human categorization into account (e.g., Bareiss, Porter, & Wier, 1990; Nosofsky, 1986). One common way to address the variable importance of different features is to assume that different properties of to-be-categorized items have different weights that modulate their overall impact on categorization decisions (Anderson, Matessa, & Lebiere, 1997; Ashby, Paul, & Maddox, 2011; Hintzman, 1984; Kruschke, 1992; Love, Medin, & Gureckis, 2004; Nosofsky, 2011; Thiessen & Pavlik, 2013).

For example, the Generalized Context Model (GCM; Nosofsky, 1986) proposes a set of selective-attention weights that represent the learned strategy of attending more to some properties than others when making decisions about stimuli represented in a multidimensional space. In GCM, the attentional weights are either free parameters fit to the data or assigned plausible values to approximate human behavior in a categorization task. Although the assumption is that these attentional weights to different dimensions are learned, GCM is a high-level representation model, not meant to model trial-by-trial changes and as such does not include an explicit process for how attention changes over time during learning.

Other models, however, have addressed this question. For example, the Attention Learning COVERing map (ALCOVE) model of categorization (Kruschke, 1992) includes a process through which attentional weights are learned during learning. On each learning trial, ALCOVE computes the discrepancy between the classification provided and the correct classification and adjusts the attention weights to reduce the classification error. Other models provide similar mechanisms for learned attentional weights: SUSTAIN (Love et al., 2004) includes an attentional tuning algorithm, whereby attentional weights are updated to provide greater salience to relevant feature dimensions and MINERVA 2 (Hintzman, 1984) includes an abstraction mechanism through which items are re-encoded emphasizing learning-relevant information (for a further specification of attentional processes in MINERVA, see iMINERVA, Thiessen & Pavlik, 2013).

2. Global and local pressures on category learning

Although taking steps toward a mechanism by which we learn what to attend to and flexibly adapt to the properties of the task and the stimuli being learned, these models still fall short

when considering learning as a process through time. Broadly speaking, even in models with a learning process, such as ALCOVE and SUSTAIN, attention is updated to change the salience of dimensions based on their overall relevance for categorization, that is, considering the global context of all of items in the same category.

Thus, current formal and informal models of categorization assign attentional weights during learning to facilitate human flexibility in categorization. However, there are two major issues with this approach. First, it implies that learners have access to the entire set of previously seen examples when making a local categorization decision and that all items and all their features are equally important, which at face value seems unlikely. Indeed, previous research has demonstrated a “recency effect” in categorization, whereby more recently studied items have a larger impact on categorization decisions than earlier ones (Jones & Sieck, 2003; Jones, Love, & Maddox, 2006). One extrapolation of these findings is that, at the extremes, some previously seen items might have a small or negligible effect on categorization decisions. Second, it implies that category learning is guided only by overall attention weights that might or might not be relevant for the task at hand. That is, when deciding which features to attend during learning, existing models optimize for overall categorization decisions: is feature x predictive of category X ? Although feature x might, overall, predict category X , in the moment-to-moment context of the other items studied (as in the book/bowl example above), that might not be the case: a feature that generally predicts X might not be relevant in the local context of a series of items that do not have that feature. Although feature X might be relevant to categorization, if that is not the case at the local context level of the recent few items seen, then it is unlikely to receive much attention.

Empirically, although some work has suggested that the attention parameters in ALCOVE (Rehder & Hoffman 2005a) and GCM (Rehder & Hoffman 2005b) are related to looking time and can account for learners’ patterns of looking and encoding, others have suggested that its power is limited. For example, Blair, Watson, Walshe, and Maj (2009) showed that different stimuli can elicit different attentional allocations, a result that, as the authors point out, is not captured by current models. Similarly, Matsuka and Corter (2008) have proposed that learners take situationally specific constraints into account when allocating their attention, something not captured by current models.

If, instead, we think of category learning as a process through time in which the learner is concerned only in solving the local task of making sense of how a current item and its temporal neighbors are grouped and in which every item has local contextual pressures on attention and encoding (from other stimuli, variation in task, etc.), we can create more flexible and informationally leaner models of learning. In this view, human learning as a process through time is the result of pressures at a small timescale. The *local context* of categorization is not only what category an item belongs to, but how that item fits with items recently categorized into one or another category. For example, a feature that overall predicts categorization (global context) might be missing from a specific exemplar or be present in an exemplar of the opposite category just presented in the previous trial (local context), in which case a different feature might be more relevant *locally*. This is a critical distinction in a world where categorization decisions are potentially seldomly based on all the information available about each category but instead only a few cases, possibly because of memory or expertise

constraints (Elio & Anderson, 1984; Jones & Sieck, 2003). While models like EBRW (Nosofsky & Palmeri, 1997) selectively sample a restricted set of previously seen examples according to their similarity to an item to be categorized, we are proposing to selectively sample examples according to their temporal proximity to the item, regardless of their similarity to the item.

Consistent with this view, there is evidence that what information is attended to and encoded depends not only on its global context (what category it belongs to), which current models of categorization focus on, but also on its local context (neighbor items), which most current formal and informal models of categorization ignore. For example, Aha and Goldstone (1992) demonstrated that humans are able to learn categories that require attributes to be weighted differently for different category *exemplars*. Across two category learning experiments, participants showed sensitivity not only to the global context (the category boundaries), but also local context, by selectively attending different dimensions in different regions of the space of stimuli. That is, human categorization is sensitive to the local context of an exemplar relative to its stimulus space. Indeed, other models and researchers have proposed that a psychological plausible model of human learning should categorize items in a context-sensitive manner (e.g., Barsalou & Medin, 1986; Tversky, 1977).

Another example of how human learning is sensitive to not only the global context of which category an item belongs to but also the local context of which other items are studied in proximity comes from a wealth of studies demonstrating sequencing effects in category learning. The sequence of events has been shown to have an impact on what we learn (Goldstone, 1996; Schyns & Rodet, 1997) and how well we learn it (Kornell & Bjork, 2008; Wahlheim, Dunlosky, & Jacoby, 2011). There is a wide array of evidence that the sequence in which information is presented influences how we perceive (Goldstone, 1996; Schyns & Rodet, 1997), remember (Jones & Sieck, 2003), represent (Clapper, 2014; Corcoran, Epstude, Damisch, & Mussweiler, 2011; Qian & Aslin, 2014), discriminate (Lipsitt, 1961; Samuels, 1969; Sandhofer & Doumas, 2008; Zotov, Jones, & Mewhort, 2011), and learn (Bloom & Shuell, 1981; Elio & Anderson, 1984; Helsdingen, van Gog, & Van Merriënboer, 2011; Li, Cohen, & Koedinger, 2013; Mack & Palmeri, 2015; McDaniel, Fadler, & Pashler, 2013; Zeithamova & Maddox, 2009) new information. These effects are both evidence that human categorization is sensitive to local variation and an ideal test situation to test the impact of local context in category learning because it is a situation where different sequences will result in the same item having different temporal neighbors for each learning event (i.e., different local but not global contexts). Moreover, because, as discussed above, existing models of categorization consider only the global context of categorization (how diagnostic each feature is for category assignment), each item will be encoded in almost the same way regardless of the local context of the items just seen. In practice, this means that the sequence of learning is unlikely to have a strong impact on learning. We start by confirming this prediction in *ALCOVE* and *SUSTAIN*.

To account for some of the findings that the sequence of learning changes what is learned, Carvalho and Goldstone (Carvalho & Goldstone, 2014b, 2015a, 2015b, 2017) have proposed the Sequential Attention Theory (SAT; Carvalho & Goldstone, 2017). According to SAT, one of the ways in which the sequence of learning influences learning is by creating

pressures on what information is attended and encoded. Specifically, Carvalho and Goldstone, proposed—and empirically demonstrated (Carvalho & Goldstone, 2014a, 2014b, 2015a)—that studying items of the same category consecutively (blocked study) makes similarities among temporally neighboring items more salient and more likely to be used for categorization, whereas studying items of different categories consecutively (interleaved study) makes their differences more salient and more likely to be used for categorization. This proposal is able to capture many findings in the literature showing, for example, that interleaved study of categories improves learning of similar categories (for which detecting differences is particularly important), whereas blocked study of categories improves learning of dissimilar categories (for which detecting similarities within each category is important; for a metaanalysis, see Brunmair & Richter, 2019).

In sum, one of the critical pressures affecting where attention will be directed during category learning is likely to be the local context of the other items experienced in proximity (temporal, physical, etc.). Local influences go beyond the global context of whether, overall, a dimension or property has been relevant for categorization in the past. The effect of both global and local influences on categorization is likely to be exerted through selective attending and encoding different properties with different experience, resulting in different properties being attended such as in the avocado example in the start of this paper. However, previous models implementing learning mechanisms that capture this phenomenon have focused only on the global context. In this paper, we present a new model that considers the local context of categorization as critical, focusing on the local temporal context.

3. A new model of category learning: SAT-M

We propose the Sequential Attention Theory Model (SAT-M), which computationally specifies and extends the SAT framework proposed by Carvalho and Goldstone (2017). SAT-M is a new exemplar model of human categorization based on GCM (Nosofsky, 1986) in which how items are encoded depends not only on global attentional weights that maximize correct categorization, but also on the immediate context during learning (i.e., the other stimuli studied in close proximity). If humans are, in fact, sensitive to local context in how they encode information, then SAT-M should provide a better characterization of learning behavior and category generalization, especially in situations where local context is particularly variable. One such situation is when one varies the sequence of exemplars during learning.

3.1. Overview of SAT-M

In SAT-M, as in other exemplar models of categorization, each experienced item is represented by its value properties. During learning, each stimulus is stored along with its category assignment and encoding weights for each of its features. Later classification of test items, both trained and novel, depends on their similarity to the aggregate properties of the previously trained items, where similarity is based on how distant these test items are from the previously studied items in the categorization space. A test item will tend to be placed into a category to the extent that it is more similar to items in that category than other categories.

Unlike other exemplar models of categorization, however, in SAT-M, the representation of studied items is influenced not only by its properties and the relative relevance of each property for categorization, but also by how those properties and the category assignment vary from prior items seen before it. Thus, unlike other exemplar models, the stored representation is biased toward features that were relevant for its categorization in the context of the other item studied in temporal proximity. This is achieved by storing for each item feature an encoding weight that modulates the impact that that feature will have to determine the item's similarity to new items at test. For example, if two similar items of the same category are seen in close temporal proximity, their similarities will be more strongly encoded and these stimuli will be seen as more similar than if they had been studied further apart in time.

Importantly, this process is not simply attentional filtering in the sense of reducing the number of dimensions for consideration. Our proposal is that by differentially encoding different properties of the stimuli from trial to trial, we are creating different attentional pressures and the history of categorization will result in different features playing a larger or smaller role during learning and future categorization. The encoding parameters in SAT-M lead to locally guided diagnosticity-based weighting of dimensions, not just paying attention to a subset of dimensions because there are too many dimensions to attend. In essence, each item is being encoded differently depending on where it was positioned in the training sequence and the item before it. The encoding of object 1 from category A when another object from category A was presented immediately before it will be quite different than encoding of that same object when preceded by an object of category B. In the next section, we discuss how the model works in more detail, its individual components, and principles.

3.2. *SAT-M principles and implementation*

As mentioned above, SAT-M builds on previous exemplar models. Specifically, SAT-M shares many of its principles and assumptions from the GCM (Nosofsky, 1986). SAT is based on (extends) GCM, and it crucially features a trial-by-trial encoding mechanism in which the encoding strength given to features depends on whether they are the same or different as previous item features, and whether the current and previous items belong to the same category. We view it as an advantage of the current model that it builds upon a model that has already been very successful (see, e.g., Hu & Nosofsky, 2021). Too rarely do cognitive scientists build upon previous work, instead proposing completely new models that may not pass benchmarks that are passed by extant models (Wills, 2013). Next, we will describe the general mechanisms of GCM and note how SAT-M deviates from it.

The principles of GCM and SAT-M can be divided into three main parts: item encoding, item similarity, and decision process. We will describe the model starting from its observable behavior, how it makes a categorization decision, and work backwards to the stimulus encoding that is used as a basis for the categorization decision. To foreshadow, categorization decisions are based on the perceived similarity of the to-be-categorized stimuli to all previously categorized stimuli and their respective category assignments. The similarity between stimuli is, in turn, inversely related to distance between their integrated distance along all of their component features. Finally, in SAT-M, unlike GCM, the distance between stimuli

is a function of each stimulus's properties as influenced by their temporal context (i.e., the properties of adjacent neighboring stimuli).

3.2.1. Categorization decision

To decide how to categorize a given item y , GCM takes into account how similar item y is to all other items encoded. Item y will tend to be categorized as belonging to the category containing the greatest number of items similar to y . The categorization probability uses a ratio rule, and is an application of Luce's choice rule (Luce, 1963; Shepard, 1965). The probability of categorizing an item y as belonging to a given category A is given by the summed similarity of that item to all the x exemplars of category A , divided by the summed similarity of y to all the k exemplars of all the categories, K :

$$p(A|y) = \frac{\beta_A [\sum_x s_{xy}]^\gamma}{\sum_K \beta_K [\sum_k s_{ky}]^\gamma}, \quad (1)$$

where S_{xy} denotes the similarity between item y and item x . $\beta_A > 0$ represents the bias toward responding with category A . A response-scaling parameter, γ . When $\gamma = 1$, the model responds more probabilistically by "probability matching," with greater values of γ , the model's behavior is more deterministic toward the category with greater summed similarity (Nosofsky & Zaki, 2002).

3.2.2. Item similarity

To determine how similar two items are, the model uses an exponentially decaying function of distance. The similarity between items x and y is given by:

$$s_{xy} = e^{-c d(x,y)^\rho}, \quad (2)$$

where $d(x,y)$ is the attention-weighted distance between the two items. This calculation includes a freely estimated sensitivity parameter, c , that defines the rate by which similarity decays with distance, that is, the gradient of the similarity function. As c increases, categorization decisions are disproportionately influenced by trained items that are very close to the test item. As c decreases, then categorization is more equitably determined by all trained items, close or far, from the test item. Finally, the shape of the function relating distance and similarity is defined by a parameter ρ . When $\rho = 1$, there is an exponential relation between similarity and distance, while when $\rho = 2$, there is a Gaussian relation between the two. This parameter is often set in advance, and values greater than 1 are often used for highly confusable stimuli (Nosofsky, 2011).

3.2.3. Distance in multidimensional space

The absolute difference between two items is determined for each dimension in the multidimensional space. This difference is weighted by an attention parameter ($\omega_i > 0$) that characterizes the global salience or relevance of dimension i , similar to previous exemplar models. The attention-weighted sum of the differences for all dimensions is the total distance between

the two stimuli. Thus, the distance between stimuli x and y is computed by:

$$d(x, y) = \left[\sum_i \omega_i \varepsilon_i^x |x_i - y_i|^r \right]^{\frac{1}{r}}, \quad (3)$$

where ω_i is the attention allocated to Dimension i and x_i and y_i are the feature values of stimuli x and y on dimension i . Notice that when ω is high, that dimension will have a greater influence in the distance calculations, whereas smaller values of ω will render any differences along that dimension less influential. This parameter can reflect learned information about which dimensions are relevant for categorization and which are not. A scaling parameter r is used to define the form of the distance metric. Although ρ defines the relation between distance and similarity in calculating items' similarities (with $\rho = 1$ implementing an exponential relation, while $\rho = 2$ implements a Gaussian relation), r defines the form of the distance metric when calculating the items' distance in psychological space. When $r = 1$, a city-block metric is used, while when $r = 2$, a Euclidean distance metric is used. While ω is often a free parameter fit to subject data, r is often defined by the type of stimuli used. When stimuli are highly discriminable and psychologically separable, a city-block metric is often used, while a Euclidean metric is used for integral dimensions that are not easily separable (Nosofsky, 2011).

3.2.4. Influence of temporal context

SAT-M incorporates variable encoding strengths for the different features that comprise a stimulus (so named to differentiate them from the attention weights ω defined above which represents the relevance of features for global category assignment, that is attention modulation based on global context). The critical parameter $\varepsilon_i^x > 0$ in Eq. 3 is the encoding strength of Feature i in Exemplar x , which depends on the sequence of study. SAT-M encodes items by comparing them to previously presented items. Differences and similarities between features from successive items and the items' category assignments define the strength of encoding of each feature ε_i , that is, the encoding weight given to that feature. The assumption here is that each object is encoded not relative to the overall task of categorizing objects into one of the categories, but rather each object is encoded relative to how it compares to the most recently categorized item. Previous work has suggested that this is likely to take place. For example, Jones and Sieck (2003) demonstrated clear recency effects in categorization, where the most recent items seemed to be considered more heavily than earlier items for categorization decisions (see also Jones et al., 2006). The current version of SAT-M makes the simplifying assumption that the previous item and its category assignment is perfectly remembered. As we will discuss later, future work should expand on this by including a memory decay function.

Thus, for each dimension, the difference between x and the preceding item y is modulated by ε_i^x and its value depends on the match between the feature value in dimension i between x and y and their category assignments. The four possibilities are formally represented by four different encoding weights ($\varepsilon^{\text{different feature, same category}}$, $\varepsilon^{\text{different feature, different category}}$, $\varepsilon^{\text{same feature, same category}}$, $\varepsilon^{\text{same feature, different category}}$). Because the encoding

weights depend on the properties of the current and previous stimuli and their category assignments, the same feature might (in fact, is likely to) receive different encoding weights from trial to trial and from participant to participant. Feature i 's encoding weight is stored along with x during learning and modulates the influence that feature i has on the overall similarity metric and thus categorization.

The encoding weight assigned to a given feature is the result of the relation of that feature with those of the immediately preceding item and works to increase or decrease the importance of that feature for a stimulus as a function of sequence of study. Note that although we separate these four encoding weights, because there are four ε_i free parameters fit to the data, it is possible that the value is the same for all four possibilities. This means that the differences in encoding between these four possibilities are not hard coded into the model, and the best fitting solution becomes an indicator for how important different features are as a function of their encoding context.

4. Applications

We start by describing the target phenomenon we will use to demonstrate the role of local context in category learning and apply SAT-M. We will then show evidence that previous models cannot account for human behavior in such situations and demonstrate that SAT-M provides a good fit to the data, suggesting that local context—in addition to global context captured by other models—has an important influence on category learning. Finally, we will probe SAT-M's assumption that attention shifts on a local context, trial-by-trial, basis by using eye-tracking data to test whether the local attentional/encoding behavior of the model matches those of learners.

4.1. Target phenomenon: The effect of sequence on category learning

In our application, we will focus on the behavioral results of one of the early studies on the impact that different sequences of study have on category acquisition: the results from Carvalho and Goldstone (2014b). As described above, this phenomenon allows us to demonstrate the impact that local context changes have on category learning and how SAT-M captures this effect, whereas models that do not include such a process of local attention changes do not closely match the results.

Carvalho and Goldstone (2014b) investigated whether different sequences of study would improve learning for different types of categories. Their main proposal was that different sequences of study lead to encoding different properties because learners' encoding of each of the items is dependent on what they have experienced before. Thus, if study involves repetition of the same category close in time, the encoding would emphasize what is similar among those items, whereas if study involves alternation between categories close in time, encoding would instead emphasize what is different among those items. This effect would be most salient when comparing learning of different types of categories. If the categories studied are highly *dissimilar*, noticing small *similarities* among items of the *same* categories—and

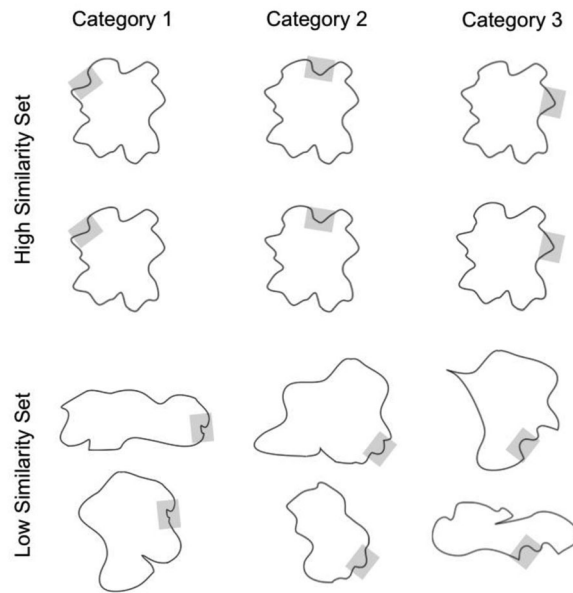


Fig. 1. Example stimuli used in Carvalho and Goldstone (2014b). Stimuli in the high similarity set (top panel) differed on only two features between any two categories and only one feature among items of the same category. Stimuli in the low similarity set (bottom panel) differed in many features between categories and among items of the same category. Gray boxes (not presented to participants) highlight the category-defining features. For details of the structure of the categories used see: <https://osf.io/s87tf/>.

by hypothesis blocked study—should improve learning outcomes. Conversely, if the different categories studied are highly *similar*, noticing the small *differences* between items of *different* categories—and by hypothesis interleaved study—should improve learning outcomes.

To test these predictions, Carvalho and Goldstone (2014b) used two different sequences of study: blocked practice where items of a category are often followed by other items of the same category, and interleaved practice where items of one category are often followed by items of one of the other categories. They also manipulated the type of category being learned: either low similarity categories where items in the same category differed from each other and from items of other categories on most of their features, or high similarity categories where items of different categories differed from items of the same and different categories on only a small number of features.

The categories used were blob shapes composed of multiple curvilinear segments (see Fig. 1 for examples). The structure of the categories is shown in Appendix A. Each dimension corresponds to a segment location in the blob, whereas each feature value corresponds to a particular curvilinear segment. High and low similarity categories varied on how many differences existed between and within categories and thus on how many feature values there were for each dimension (but not on the number of dimensions). All categories were defined by the presence of a particular curvilinear segment in a particular location of the space. In the high similarity set, exemplars shared most of their features with all the other exemplars

in the same category and in each of the other five categories. Moreover, variation within each category was exactly the same for all categories, so that a difference that could exist between two exemplars in category 1 would also exist between two exemplars of each of the other categories in the set. In the low similarity set, exemplars within each category shared only the category-relevant feature. Moreover, exemplars from different categories differed in all their features. Some of the exemplars had an overall round shape, and others an overall oblique shape (this variability was equally distributed across categories).

Participants started by studying three categories in one of the sequences and then being tested (in random order) on those followed by another study phase in the other sequence and a new test phase. Interleaved study was achieved by alternating categories 75% of the time, whereas blocked study was achieved by alternating categories only 25% of the time. Study phase was composed of four blocks of 72 trials each. During study, eight items of each category were presented three times each. The test phase included 48 trials, each a presentation of one of the 16 items of each category studied (half studied and half novel). Thus, when we refer to the sequence, we refer to how training was sequenced. Test sequence was never manipulated.

Overall, Carvalho and Goldstone (2014b) found that for high similarity categories, interleaved study improved the classification of novel items at test, whereas for low similarity categories, blocked study improved the classification of novel items at test. This is the critical pattern of results that we will model in this paper.

Of course, one way to capture Carvalho and Goldstone's proposal and results described above would be to stipulate that the attentional weights for similarities and differences are different for blocked and interleaved sequences, for example, by fitting the two sequences separately. However, such an approach brings us back to the issue stated at the start of this paper. In so doing, one would be assuming that even though the sequence is a local change in context, it would be explained solely by global context, that is, by what category the item belongs to and the pressures that exerts on attention and encoding. Instead, we propose that SAT-M can parsimoniously account for the effect of different sequences on category learning and, moreover, better characterize the flexibility of human learning.

4.2. ALCOVE and SUSTAIN and categorization performance following different sequences

Our main proposal is that attention changes that occur over the course of category learning are the result not only of global context pressures arising from category classification, but also local context pressures arising from what other items were studied in closed temporal proximity. Importantly, previous models and theories of category learning fail to capture the effect of local context, focusing only on the global context.

To confirm this assertion, we fit two models that, like SAT-M, include an attention learning process during category learning: ALCOVE and SUSTAIN. Our point with these fits is to demonstrate that without considering the importance of local temporal context, these models cannot easily account for our previously presented category learning results, as we argued is the case when comparing the effect of different sequences of learning. It is, of course, possible that if we were to fit the models separately for each sequence and added the flexibility of

Table 1

Best-fit parameter values for ALCOVE and SUSTAIN fits to Carvalho and Goldstone (2014b)

Parameter	ALCOVE				SUSTAIN			
	c	\emptyset	λ_w	λ_α	r	β	d	η
Best-fitting value	15.85	1	0.718	0.99	11.99	0.99	1.13	0.49

Note: Models were simulated using the implementation provided in the R package Catlearn.

allowing different parameter values for interleaved versus blocked training, then ALCOVE and SUSTAIN might fit the data, but, unlike our proposal, this would involve many degrees of freedom and would not offer a parsimonious mechanistic explanation.

We fit both models to the target phenomenon described above. We used the implementations of ALCOVE and SUSTAIN available in Catlearn (Wills, O’Connell, Edmunds, & Inkster, 2017), an open archive of formal psychological models implemented in R (R Core Team, 2019). We trained the models with the study sequences from each participant and compared the model performance in the test phase with that of the participant trained with that sequence. In both models, category learning feedback was presented during training but not during test. We used sum of squared errors, implemented as part of the Catlearn package, as the objective function to determine best fitting parameters. The category representation described in Appendix A was adapted—maintaining its main characteristics—to match how ALCOVE and SUSTAIN define category spaces.

ALCOVE has a total of six parameters. From these, we defined the distance metric as Euclidean distance metric ($r = 2$) and used a Gaussian similarity gradient ($q = 2$), because as noted before and exemplified in Fig. 1, both category structures were designed to have highly confusable stimuli for which a Gaussian similarity gradient is more appropriate and the dimensions are not easily separable, for which Euclidian distance is more appropriate. We consider these stimuli to be integral because previous work with similar stimuli found that with practice, people did not seem to be categorizing blobs by separately combining evidence along each spatially defined feature (Goldstone, 2000), suggesting that during categorization tasks, these types of stimuli are treated as integral. Importantly, using a city-block metric ($r = 1$) and an exponential similarity gradient ($q = 1$) did not change the overall pattern of results for this or any of the subsequent model fits. The remaining four parameters were free parameters fit to the data. These parameters include the specificity constant (c), the decision constant (\emptyset), and the associative (λ_w) and attentional (λ_α) learning rates. The best fitting values of these parameters are presented in Table 2.

SUSTAIN has a total of five parameters. From these, we defined the threshold to create a new cluster (τ) as 0 because the study includes only supervised trials. We fit the following four free parameters: attentional focus (r), cluster competition (β), decision constancy (d), and learning rate (η). The best fitting parameter values are presented in Table 1. We defined the initial attentional weights and associative strengths as the same to all dimensions. Moreover, we defined the initial receptive fields as equally tuned and the network was always started with a single cluster with zero-strength weights.

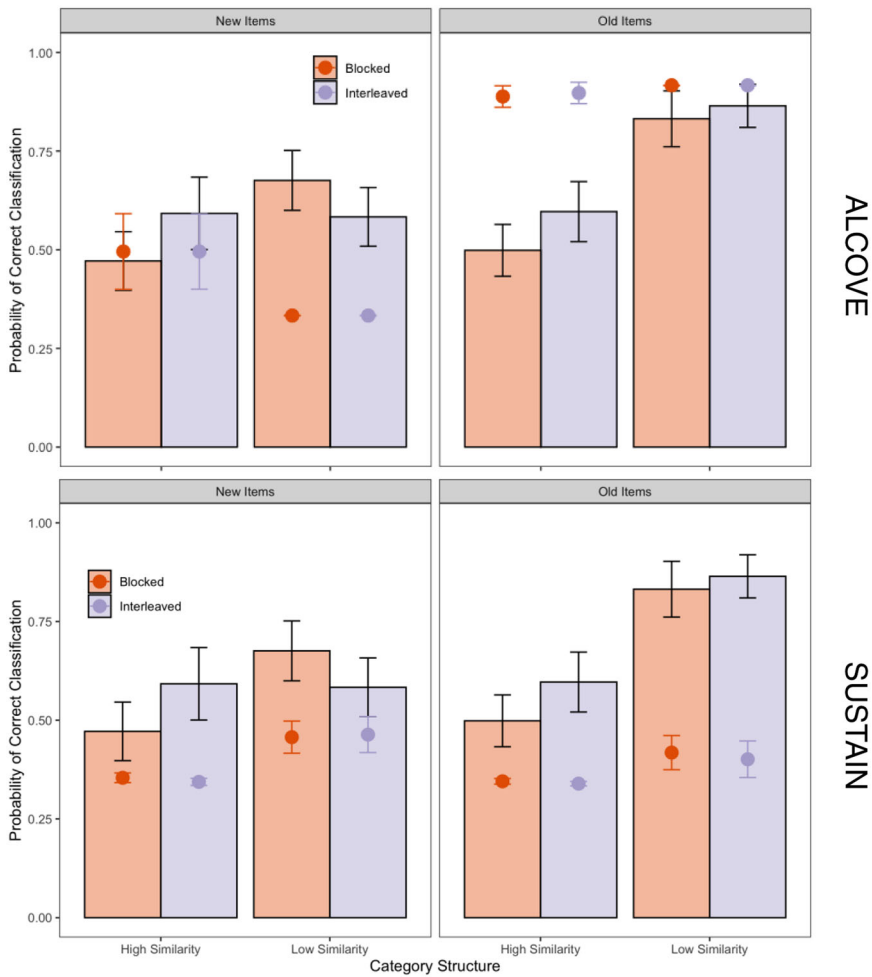


Fig. 2. Fitting results (dots) for ALCOVE (top panel) and SUSTAIN (bottom panel) over the empirical results from Carvalho and Goldstone (2014b; represented by the bars). Best-fitting parameter values are presented in Table 1.

Of note, our modeling approach here and in all subsequent model fitting was to fit the models to both sequences (and category structures, in this case) together. We use this approach because it is parsimonious. A model accounting for all conditions with the same parameter values is preferable to a model that requires additional assumption about how parameters vary across conditions. One could easily tweak parameter values and improve fit to the data without providing a parsimonious account by fitting each condition separately.

The results of our simulations with best fitting parameters are shown in Fig. 2. As can be seen in Fig. 2, both models provide an overall poor fit to the data ($r^2 = .005$; $SSE = 25.85$, and $r^2 = .001$, $SSE = 31.54$, for ALCOVE and SUSTAIN, respectively). Critically, both models also fail to show an interaction between category structure and sequence of study

for novel items presented at test, a hallmark of the effect that local context has on category learning. The results of these simulations are consistent with our proposal that part of what is happening during category learning is learning to attend to information that is or is not relevant in the context of the other items studied in close temporal proximity (local context) and not only the global context of all the items belonging to either category. Although their specific mechanisms differ, attention in both ALCOVE and SUSTAIN changes during category learning as a result of the dimensions' general relevance for categorization. While both models are process models that take as input a specific sequence of trials, their dimension learning mechanisms are designed to work over long time courses, rather than the short trial-to-trial fluctuations of encoding that is implemented in SAT-M. One could argue that the poor fits of ALCOVE and SUSTAIN are in part the result of trying to fit both the high and low similarity categories simultaneously. One reason to fit each category structure separately would be to assume that each category structure constitutes a different global context where different information needs to be attended to and ignored. However, doing so would also assume that the cognitive system is "reset" to learn different structures in different ways and that it has a special way to tell which context it is while not specifying what that way might be. Similarly, one could argue that both ALCOVE and SUSTAIN could capture the results of Carvalho and Goldstone if fit separately to each sequence and category structure. Although that might be case, those fits do not answer the question of where learned attention to different features in different contexts originates. SAT-M, we propose, provides a parsimonious solution to this question.

4.3. SAT-M and categorization performance following different sequences

In the previous section, we demonstrated that extant models that include a learning process and a mechanism for attentional changes during learning cannot account for the empirical results reported by Carvalho and Goldstone (2014b). In this section, we explore whether the mechanism of sequential encoding of similarities and differences among successive items, when paired with different sequences of stimuli, can yield these results. To that end, we fit SAT-M to the results from Carvalho and Goldstone (2014b, Experiment 1). The version of SAT-M we used had three free parameters (ϵ , γ , c), and three fixed parameters (ρ , r , ω). We fit SAT-M using maximum likelihood estimation with the four versions of the encoding weight (ϵ same feature, same category, ϵ same feature, different category, ϵ different feature, same category, ϵ different feature, different category), γ , and c as free parameters fit to the data. We set $\rho = 2$, because stimuli were relatively confusable and $r = 2$ (see above for there details on the reasoning for these decisions). Importantly, setting $\rho = 1$ and $r = 1$ did not change the overall pattern of results. We set $\omega = 1$ because the stimuli used were designed such that all dimensions were approximately equally salient, and the assignment of feature positions to roles in the category structure was counterbalanced. This counterbalancing makes it likely that there are no systematic nonlinear interactions between feature values on the same blob. To the extent that there are interactions between particular feature values, that will only handicap SAT-M from predicting the results given that it does not incorporate those interactions. Additionally, the different models being compared are all similarly handicapped. We used maximum likelihood

Table 2

Average best-fitting c and γ parameters for SAT-M fits and average best-fitting c , γ , and ε parameters SAT-M-R to results from Carvalho and Goldstone (2014b, Experiment 1)

	c	γ	ε
SAT-M	1.01 (0.46)	4.12 (0.18)	–
SAT-M-R	1.00 (0)	1.32 (0.00001)	0.9 (0.00001)

Note: Standard deviation in parenthesis.

estimation to allow for direct comparisons between SAT-M and SAT-M-R with quantification of relative evidence in favor of each.

We fit the model using sequences generated in the same way as those shown to participants. For each run, the model was trained with a sequence similar to that of a human participant and then performance in a test phase, including old and novel items, was tested. In our simulations, each stimulus was represented as an eight-dimensional vector. In the low similarity set, each dimension could assume a value 1–48 representing the many different components of that set of stimuli. For the high similarity set, each dimension could assume a value 1–4, representing the few differing components among stimuli. This representation matches how the stimuli were initially created (see Carvalho & Goldstone, 2014b, Appendix A, and materials stored in OSF: <https://osf.io/s87tf/>). The best-fitting c and γ parameters are shown in Table 2. Interestingly, for both fits, c approximates 1. This might suggest dependencies between c and γ as previously suggested (Ashby & Maddox, 1993; Myung, Pitt, & Navarro, 2007; Smith & Minda, 2002).

The results of the model are displayed as dots over the bars representing Carvalho and Goldstone's (2014b) results in the top panel of Fig. 3. As it can be seen, the model provides a good fit to the data ($r^2 = .96$, $SSE = 0.008$, $BIC = 7053.60$), suggesting that the effect of different sequences on category learning and generalization can be captured by a process of sequential comparison between successive items. For comparison, we constructed a restricted version of the model where the sequence of study does not influence the encoding weights of each stimulus property (Sequential Attention Theory Model Restricted; SAT-M-R). In SAT-M-R, there is a single encoding weight ε that is a function of the feature's category relevance and does not vary depending on sequential variation among items (note that $\omega = 1$ as in SAT-M such that all dimensions are equally relevant for categorization). Thus, in SAT-M-R, ε is allowed to vary based on the global relevance of a dimension for categorization but does not vary depending on the local context of the value of that feature in the previous item. SAT-M-R embodies the plausible alternative proposal that the effect of different sequences—and local variation during study—can be well-accounted by global weighting based only on category assignment information. As such, SAT-M-R can be construed as a version of GCM with a learning process. Best-fitting parameters are presented in Table 2.

As it can be seen in the bottom panel of Fig. 3, SAT-M-R provides a poor fit to the data ($r^2 = .44$, $SSE = 0.103$, $BIC = 7319.03$). Interestingly, although SAT-M-R displays the often-shown effect of stimulus similarity on categorization (better categorization of old items for low similarity categories, but better categorization of novel items for high similarity

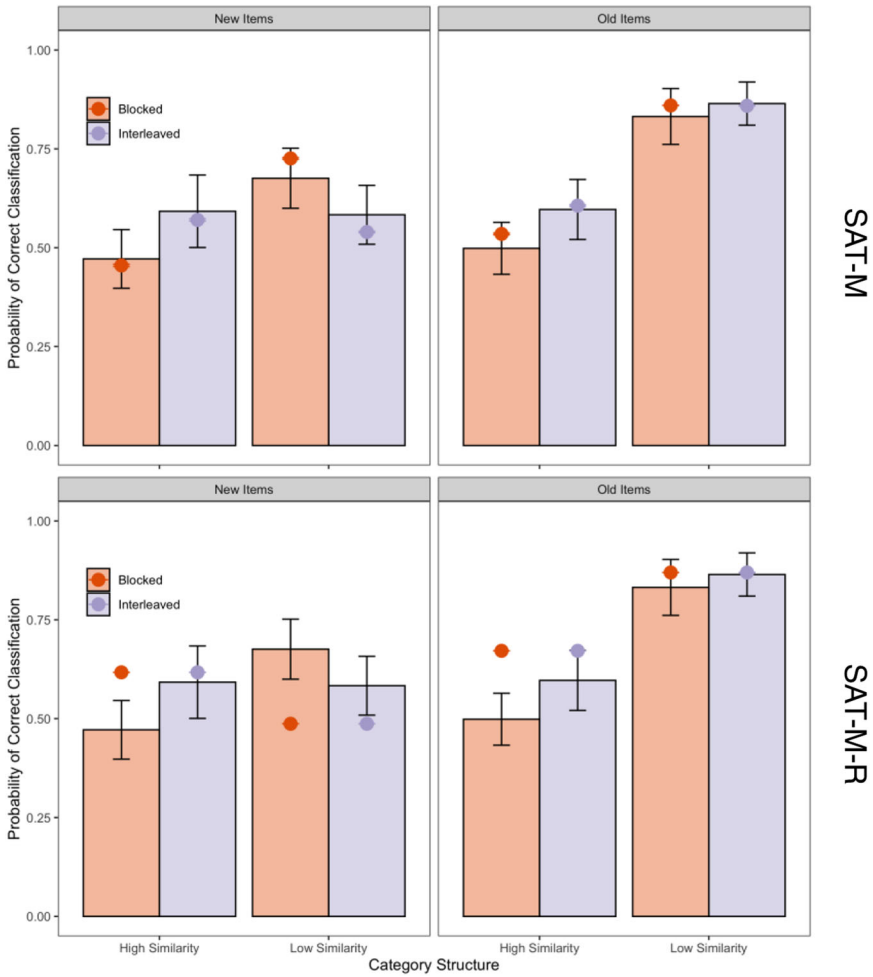


Fig. 3. Fitting results (dots) for SAT-M (top panel) and SAT-M-R (bottom panel) over the empirical results from Carvalho and Goldstone (2014b; represented by the bars). SAT-M provides a much better fit to the data than SAT-M-R.

categories, e.g., Palmeri & Nosofsky, 1995), it is not sensitive to the sequence of learning and cannot capture the interaction pattern evident in the human data. The Bayes factor comparing the two models directly suggests that SAT-M-R is much less likely to have generated the data than SAT-M ($B < 0.001$).

The ϵ parameters in SAT-M are sensitive to local context (i.e., the encoding parameters vary depending on the match in feature properties and category assignment across trials) and can be directly interpreted as the encoding strength of a feature. If encoding is sensitive to local context, then we should not only see better fit of SAT-M compared to a version in which ϵ is not allowed to vary due to local context as shown, but we should also see that the best-fit ϵ values differ depending on local context. Importantly, in SAT-M, the encoding

Table 3

Average best-fitting encoding parameters ϵ when fitting SAT-M to results from Carvalho and Goldstone (2014b, Experiment 1)

	Different category	Same category
Different feature	0.96 (0.08)	0.0015 (0.01)
Same feature	0.11 (0.03)	0.79 (0.22)

Note: Standard deviation in parenthesis.

parameters are fit to the data and, therefore, the best fitting solution could be one where the encoding parameters, like is the case in SAT-M-R, are equivalent or vary in any number of ways. Table 1 shows the best fitting parameter values.

As can be seen in Table 3, encoding strengths are stronger for differences between items of different categories and similarities among items of different categories. This result confirms that the SAT-M solution that best fits the data does so by varying encodings depending on local context (i.e., the match on category and feature properties between current and previous items). Moreover, these results are consistent with the theoretical explanation of the phenomenon offered by Carvalho and Goldstone (2014b): if interleaved study improves learning of high similarity categories because it emphasizes encoding of differences between categories, whereas blocked study improves learning of low similarity categories because it emphasizes encoding of similarities among items of the same category, then we should see that the best fitting parameters should be higher for differences among different categories and similarities among items of the same category. Because different sequences include a disproportionate number of transitions across the same category (blocked study), or different categories (interleaved study), whether differences or similarities are better encoded depends on the sequence of study.

4.4. Further tests of SAT-M involving categorization performance following different sequences

To expand on the results above, we fit SAT-M to two other studies investigating the effect that interleaved and blocked sequences have on category learning: Carpenter and Muller (2013) and Zulkipli and Burt (2013).

Carpenter and Muller (2013) demonstrated that when learning French pronunciation rules (e.g., “eau” in “bateau”), non-French speakers were better at classifying the correct pronunciations of novel words at test if they trained by practicing with words blocked by rule (i.e., studying multiple words with the “eau” pronunciation before studying words with the “ou” pronunciation), instead of interleaved by rule. We fit the results of Experiment 1 in Carpenter and Muller (2013). To do so, we created a feature space based on the characteristics of the words used in their original experiment (see Appendix C). We considered that all words had eight dimensions and aligned the features such that the letters associated with the pronunciation rule occupied the same dimensions for all words. The other dimensions represented random variation across words. For example, take the words “bateau” and “carreau,” these

Table 4

Average best-fitting parameters when fitting SAT-M to results from Carpenter and Mueller (2013; Experiment 1) and Zulkiply and Burt (2013; Experiment 2)

	c	γ	ε			
			Same cat, same feature	Same cat, different feature	Different cat, same feature	Different cat, different feature
Carpenter & Mueller, 2013; Experiment 1	0.97 (0.22)	0.99 (0.12)	0.88 (0.12)	0.37 (0.14)	0.12 (0.12)	0.86 (0.12)
Zulkiply & Burt, 2013; Experiment 2	1.01 (0.0001)	2.22 (0.0001)	1.00 (0.0001)	0.40 (0.0001)	0.24 (0.0001)	0.85 (0.0001)

Note: Standard deviation in parenthesis.

two words differ on all dimensions except the last three, representing “e,” “a,” and “u,” for the critical sounds for the rule “-eau.” The word “adosser” differs from “bateau” and “carreau” on all feature dimensions but has the same feature values for features 7 and 8 as “attraper,” representing the letters associated with the critical pronunciation rule “-er.”

In the study, participants practiced by seeing a word along with hearing the correct pronunciation rule. Participants studied four different words from each of eight rules. Four of the rules were presented interleaved and the other four were presented blocked. Following training, participants completed a classification test in which they were presented with 64 words (the 32 words seen during training plus 32 new words) and they had to identify the correct pronunciation rule by choosing one of three alternatives (correct pronunciation, incorrect rule, and incorrect pronunciation but correct rule). In our modeling, we replicated this procedure closely except the final test was a classification test requiring identifying the correct rule among all the rules studied. We used this procedure to simplify the modeling approach and remain close to the characteristics of SAT-M as a categorization model. Previous research has suggested that recognition and categorization are related (e.g., Nosofsky, 1991), and thus this simplification should not affect the main characteristics of the task. As before, we fit SAT-M using maximum likelihood estimation with the four encoding weights (ε same feature, same category, ε same feature, different category, ε different feature, same category, ε different feature, different category), γ , and c as free parameters fit to the data. We set $\rho = 2$, because we evaluate the stimuli as relatively confusable and $r = 2$ because, as encoded, the dimensions are not easily separable. Importantly, as before, changing $r = 1$ and $\rho = 1$ did not change the overall pattern of results. We set $\omega = 1$ under the assumption that in our coding of the stimuli, all dimensions are equally salient. The model was fit to the average population results from the paper, as we did not have access to individual participants’ data. The best fitting parameters are presented in Table 4.

As can be seen in Fig. 4, SAT-M provides a good fit to the data ($r^2 = .37$; SSE = 0.108; BIC = 14), capturing both the effect of novel versus old words (Panel A in Fig. 4) and the

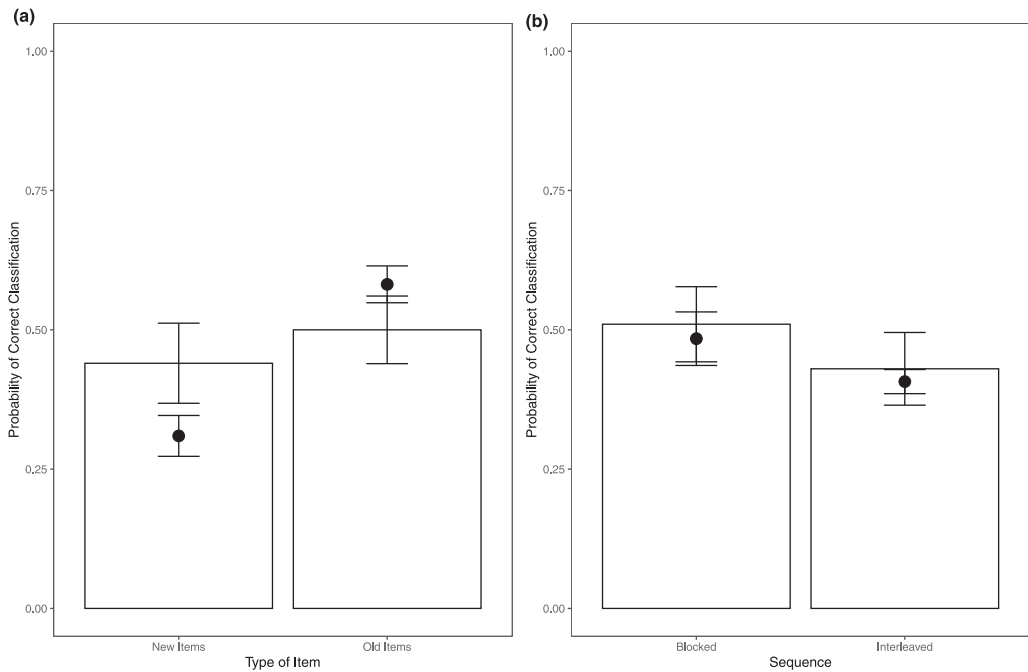


Fig. 4. Fitting results (dots) for SAT-M over the empirical results from Carpenter and Mueller (2013, Experiment 1, represented by the bars), comparing New vs. Old items (panel a) and blocked vs. interleaved study (panel b).

benefit of blocked study compared to interleaved study when classifying all words at test (Panel B in Fig. 4).

SAT-M provides a natural account for why blocked presentation of pronunciation rules leads to better learning than interleaving. Blocked presentation allows for encodings of examples that emphasize the word features that are characteristics of a particular rule. This is useful for the words used by Carpenter and Mueller (2013) because there are many nondiagnostic word features that need to be ignored, and not all of the characteristic features discriminate between rule categories.

In Experiment 2, Zulkipli and Burt (2013) compared generalization of abstract categories learned either through blocked or interleaved practice. In this study, participants learned either high similarity or low similarity categories by studying six pictures of each of 12 categories (six interleaved and six blocked). Following training, participants classified four new pictures of each category, presented in random order. The authors report an interaction between category similarity and sequence of study, such that interleaved training improved test performance for high similarity categories, whereas blocked training improved test performance for low similarity categories.

Each category was composed of pictures created by positioning different shapes on a black background. The location and properties of some of the shapes defined the category assignment (see Zulkipli & Burt, 2013 for details). Based on the description of the categories and

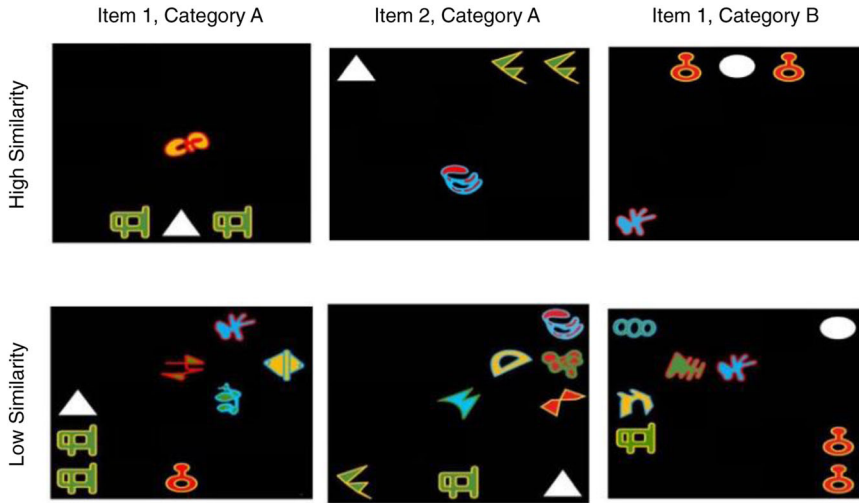


Fig. 5. Example stimuli from Zulkiply and Burt (2013). For modeling, we converted the stimuli into a feature space with nine dimensions: shape of the main object (dimension 1), color of the other category objects (dimension 2), match of shapes across category objects (dimension 3), position of the category object (dimension 4), and five distractor shape dimensions (dimensions 5–9). For details, see text and Appendix D.

the examples provided in their paper, we created a feature space that closely matched the categories used (see Appendix D). The feature space had nine dimensions: shape of the main object (dimension 1), color of the other category objects (dimension 2), match of shapes across category objects (dimension 3), position of the category object (dimension 4), and five distractor shape dimensions (dimensions 5–9). As can be seen in the examples in Fig. 5, category assignment was determined by a combination of shape of the main object (dimension 1; two values: circle or triangle), and color of the other category objects (dimension 2; six values for a combination of pairs of two out of four possible colors). For high similarity categories, the shape matched across category objects (dimension 3), whereas for low similarity categories, it could match or mismatch (dimension 3). Similarly, for high similarity categories, the position of the category object (dimension 4) was either top or bottom of the image, whereas for low similarity categories, it could be one of four positions: top, bottom, left, or right of the image. Finally, high similarity categories had only one distractor shape (dimension 5) with every item having a different shape/feature value and the same feature value for the other distractor shapes (dimensions 6–9) to indicate no shape, whereas low similarity categories had a different value in each one of the dimensions for each item (dimensions 5–9).

As before, we fit SAT-M using maximum likelihood estimation with the four encoding weights (ϵ same feature, same category, ϵ same feature, different category, ϵ different feature, same category, ϵ different feature, different category), γ , and c as free parameters fit to the data. We set $\rho = 2$, because we considered that the stimuli were relatively confusable and $r = 2$ because, as designed, the dimensions are not easily separable. Importantly, as before, changing $r = 1$ and $\rho = 1$ did not change the overall pattern of results. We set $\omega = 1$ because the stimuli were designed with an

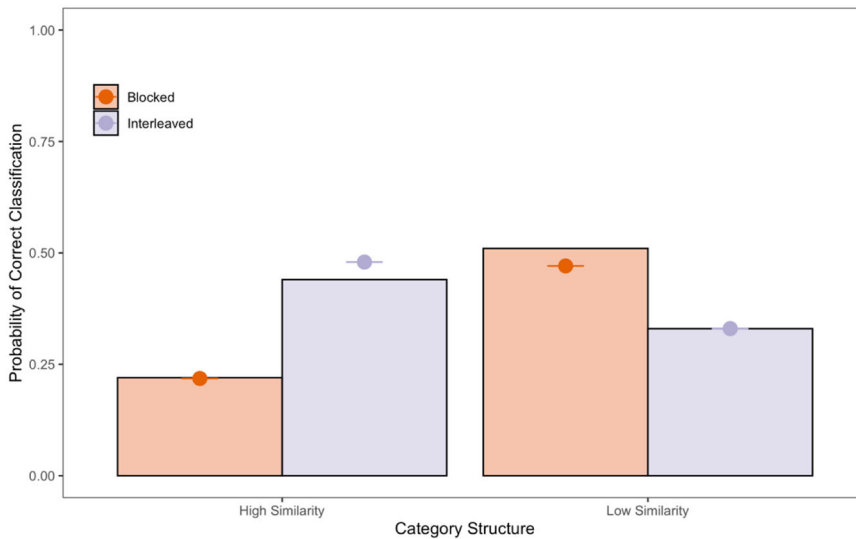


Fig. 6. Fit results (dots) for SAT-M over the empirical results from Zulkiply and Burt (2013, Experiment 2, represented by the bars).

effort to make all dimensions approximately equally salient. The model was fit to the average population results from the paper, as we did not have access to individual participants' data. The best fitting parameters are presented in Table 4.

As it can be seen in Fig. 6, SAT-M provides a good fit to the data ($r^2 = .94$, $SSE = 0.003$; $BIC = 13$), capturing the interaction between category similarity structure and sequence of study on classification of novel items at test. When the stimuli are dissimilar, it is difficult to find the features shared by members of the same category, and blocking helps to emphasize these features. When the stimuli are similar, it is difficult to find the few features that discriminate between the categories, and interleaving increases the encoding strength of these discriminating features.

The modeling of other researchers' empirical results presented here shows that the SAT-M is able to accommodate results from experiments that were not explicitly designed to test SAT. Other existing models, such as GCM, would be hard-pressed to account for these results because they do not have trial-by-trial learning mechanisms. Other models, like ALCOVE, do have trial-by-trial learning mechanisms, but they tend to slowly change attention to dimensions based on their overall relevance to a categorization task. SAT-M features rapid changes to stimulus encodings based on the immediate temporal context, in particular, the dimension-by-dimension similarities between the current and previous items. This component is particularly useful for accommodating differences between interleaved and blocked presentation. SAT also offers a different kind of explanation of observed interleaving benefits compared to contextual interference and spacing accounts. An advantage of SAT's explanation is that it can explain when interleaving does not always confer benefits over blocking.

4.5. Encoding weights and attention

The main novelty introduced by SAT-M is the sensitivity to local temporal context during category learning through differential encoding of each feature of a stimulus depending on how it compares to immediately preceding items and their category assignment. To this end, SAT-M introduces in the context of GCM the ε parameters that modulate encoding of a feature and vary depending on whether the feature and category assignment change relative to the previous item. Theoretically, we have argued that the inclusion of these parameters is sensible because learners are sensitive to local context in addition to global information about which dimensions are overall most relevant for category learning. Furthermore, we proposed that this sensitivity can be thought of as a modulation in how each stimulus experience is encoded and stored. The same item studied after a different previous item would be encoded differently.

In the earlier sections, we showed that to fit data from an experiment showing how learners are sensitive to local context, the ε parameters varied in their best-fitting values depending on whether the feature was the same or not and whether the category assignment was the same or not across two stimuli. Furthermore, we showed that SAT-M can fit a range of existing data comparing categorization and memory performance following training with different sequences. If in fact local sensitivity can be captured by differential encoding patterns as in SAT-M, then we should see that looking patterns (often taken as a measure of overt attention) should be correlated to best-fit ε values. This new insight from SAT-M can be tested empirically by comparing looking time with best-fit ε values.

Carvalho and Goldstone (2017) tested whether different sequences of study would lead learners to attend to different properties of the stimuli during different sequences of study. In their study, participants studied the same two categories of Aliens either blocked or interleaved (see left panel of Fig. 5 for examples of the stimuli used during study). The two categories studied had a structure such that multiple feature values and dimensions could predict category assignment (discriminative features), and some features, despite being common in a category, did not predict category assignment (as they were also common in the other category—characteristic features; see Appendix B and stimuli available in <https://osf.io/2n8gy/>). After study, participants completed a test phase where participants classified novel items at test. The novel items could vary on the frequent but not discriminative (characteristic) feature or on other features (see right panel of Fig. 7 for examples). Importantly, during learning, participants' eye movements were recorded using eye tracking and how long learners spent looking at each of the items' features was analyzed.

Overall, the authors found that during interleaved study, learners looked longer at properties that differed from those of the preceding item, whereas during blocked study, learners attended to both differences and similarities equally (see left panel of Fig. 7). The authors argued that the interaction is overall consistent with SAT in that following blocked study, there is no bias toward looking at differences, even though those might be more salient because people often orient toward novelty or task-relevance (Rehder & Hoffman, 2005b; Wang & Mitchell, 2011). Instead, participants attend to repeated similarities to the same extent. During interleaved study, on the other hand, participants' attention is heavily biased toward

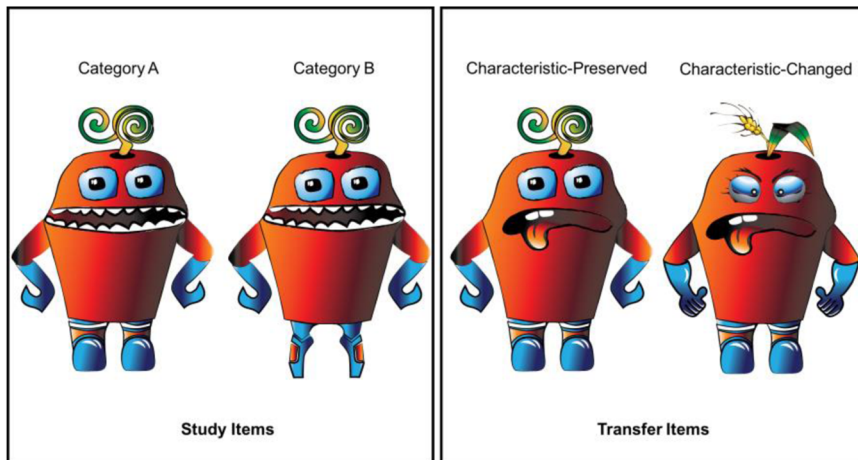


Fig. 7. Example of stimuli from one of the families used by Carvalho and Goldstone (2017). Left panel includes an example of each of the categories studied. Right panel includes an example of each of the novel items presented during the transfer task (both transfer items belong to category A; equivalent items existed for category B). For details of the structure of the categories used, see: <https://osf.io/2n8gy/>.

differences between successive items, which are likely to be properties that differentiate between the categories.

Can these different attentional patterns be captured by SAT-M? If the encoding weights in SAT-M correspond to differential looking times as proposed, then the best-fitting parameters when we fit SAT-M to the categorization data should match the looking times observed in Carvalho and Goldstone (2017). To test these predictions, we fit SAT-M to the results of the testing phase of Experiment 3 of Carvalho and Goldstone (2017). Each stimulus used in the experiment was defined as a five-dimensional vector, with each dimension taking a value between 1 and 5. After fitting the model using the same parameters as described in the previous section, we extracted the best fitting encoding weight parameters for each feature of the studied items. The average best-fitting c parameter was 1.07 and the average best-fitting γ parameter was 0.89. As described above, in SAT-M for each event, the stimulus along with encoding parameters ε for each of its features is stored. As in the previous applications, SAT-M had four encoding parameters depending on the match of the properties and category assignment among the currently encoded item and the previous one— $\varepsilon_{\text{different feature, same category}}$, $\varepsilon_{\text{different feature, different category}}$, $\varepsilon_{\text{same feature, same category}}$, $\varepsilon_{\text{same feature, different category}}$. After fitting SAT-M to the behavioral results of the transfer task, we calculated, separately for the blocked and interleaved sequences, the sum across trials and features of the encoding parameter values for features that differed across successive items and features that were the same among successive items. Conceptually, summing up the encoding parameters is similar to summing up total looking time for each feature based on sequential differences and similarities, as done by Carvalho and Goldstone (2017).

As it can be seen in the right panel of Fig. 8, for blocked study, the summed best-fitting encoding weight parameters are similar for differences and similarities across successive

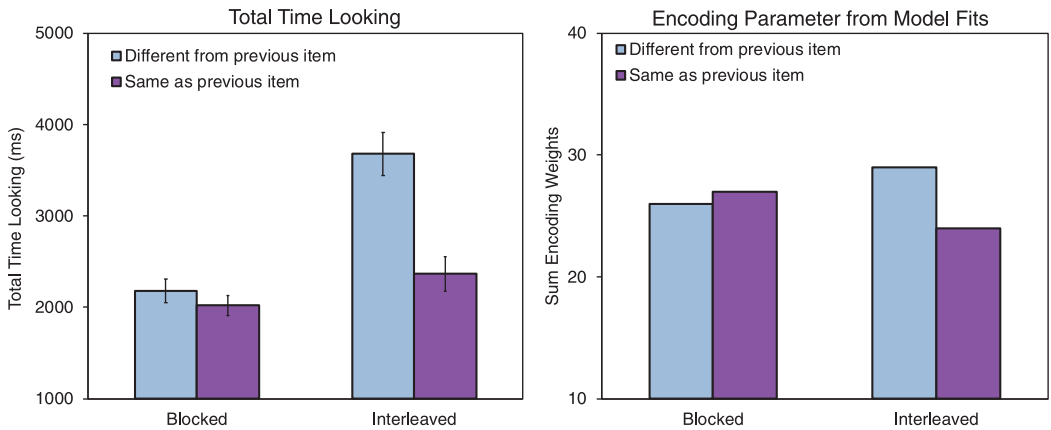


Fig. 8. Results from total looking time during study in Carvalho and Goldstone's (2017) Experiment 3 (left panel) and summed best-fit values for encoding weights (ϵ) when SAT-M is fit to the categorization results of Carvalho and Goldstone's Experiment 3 (right panel).

items. For interleaved study, on the other hand, the summed encoding weights are higher for differences than similarities across successive items. This pattern of results is similar to the empirical results found in the eye-tracking study (see left panel of Fig. 8; Carvalho & Goldstone, 2017). It is also consistent with evidence from other eye-tracking studies investigating the influence of study sequence on category learning (Zaki & Salmi, 2019).

Overall, the results from these simulations suggest that the encoding weights in the SAT-M can be directly related to encoding differences in behavioral data. Importantly, we did not fit the encoding weights to the looking data in Carvalho and Goldstone (2017). Instead, we fit the model to the categorization test results and used the resulting best-fitting encoding parameter values as a measure of the model's encoding or looking time. Finding that when fit to the test data, the model converges on encoding weights with a similar pattern to that of human looking time in the same paradigm is striking and speaks to the adequacy of the model's processes to those of human cognition.

5. Implications

The flexibility of current models of categorization lies in large part on their ability to selectively weight some features more than others during learning. This procedure—although often successful at capturing how different features are differently relevant for different categories—fails to account for the full flexibility of human categorization. Beyond considering global variables such as the relevance of a property for correct classification, humans are sensitive to the local context of categorization and the same feature can become more or less relevant depending on the other items studied in close proximity. We developed a more flexible model of categorization—the Sequential Attention Theory Model (SAT-M), based on Carvalho and Goldstone's Sequential Attention Theory of category learning (Carvalho & Goldstone, 2017).

The main assumption of our model is that how each item is encoded during category learning is not only the result of the global context of the categorization task, but also the local context of the temporally immediately preceding item. This assumption makes SAT-M not only flexible but also computationally efficient: in order to selectively encode properties that will tend to be useful for categorization, SAT-M takes only the preceding item during learning. Thus, SAT-M requires very few items that need to be stored while relying mostly on the item that is expected to be the most accessible due to its recency.

To test SAT-M, we compared its predictions with human behavior in a situation where local context strongly affects performance—the effect that difference sequences of items have on category learning. We have proposed that during category learning, people engage in a process of sequential comparison to decide what is relevant and should, therefore, be encoded and what is not relevant and can be ignored. Engaging in this process, we proposed, is the reason why the sequence of study changes what is learned—different sequences create different sequential statistics that create attentional patterns toward encoding different types of features. The model introduced in this paper instantiated this proposal by including an initial sequential encoding process during which different features are assigned encoding weights depending on their relation to the features in the item encoded immediately before and whether both items belong to the same category or not. These encoding weights represent the likelihood of encoding a particular feature during a particular presentation of the item.

Importantly, these weights do not reflect the overall relevance of the feature for categorization as do the attention parameters in models, such as GCM, ALCOVE, or SUSTAIN. A similarity shared by two successive items of the same category does not necessarily mean that it is diagnostic of categorization. Similarly, a difference between two successive items of different categories does not guarantee its diagnosticity for categorization. In this way, the sequence of study is biasing local encoding on a trial-by-trial level, as opposed to global attention to relevant versus irrelevant properties at a task level. By strongly encoding locally relevant similarities that are not globally relevant (e.g., irrelevant within-category similarities in the high similarity category set), study in a blocked sequence might, at times, deter learning. Similarly, by encoding locally relevant differences that are not globally relevant (e.g., irrelevant between-category differences in the low similarity category set), interleaved study might, at times, deter learning. The opposite is equally important—encoding locally relevant similarities or differences that are relevant for categorization might boost performance because it does not depend on having experience with the full set of items and, therefore, knowing what has been relevant so far. Locally determined encodings could be seen as a catalyst for learning when a learner has not yet collected a lot of information yet as to the global relevance of different dimensions.

Notwithstanding the significance of attending to and encoding locally relevant features, the importance of global allocation of attention to category-relevant dimensions and away from category-irrelevant ones should not be ignored. There is widespread behavioral and eye-tracking evidence that people learn to attend to certain dimensions while ignoring others, depending on their global predictive power for correct categorization (e.g., Blair et al., 2009; Chen, Meier, Blair, Watson, & Wood, 2012; Rehder & Hoffman, 2005b). In the current modeling, we have focused only on local allocation of attention and encoding for simplicity.

However, this means that because each item is encoded relative to only the previous item, dimensions that do not offer any predictive value continue to be attended. How could our model be extended to account for both local and global attention modulation? One possibility is to use the global attention parameter from GCM (ω) to model how attention is modulated by the relevance of each dimension or feature in the context of all category items. However, this approach sidesteps understanding how global context affects attentional patterns during categorization. Another possibility would be to include global attention as a learned aspect of categorization and have a process of attention accumulation for each feature starting at the beginning of encoding, based on the geometric average of the encoding weights assigned for that dimension up to that point. In essence, high global attention to a dimension would be the result of continued high local attention and would reflect an inference of “If this dimension keeps being relevant locally, it must be important globally.” With sufficient exposure, this global attention accumulator could then be used to modulate future local encoding decisions: A locally relevant difference between two categories on a seldom-predictive dimension might not be attended or encoded, while a locally irrelevant similarity between two categories on a frequently predictive dimension might be encoded. One potential advantage of this process compared to current models of categorization is that attention modulation would be a learned process at both the local (trial-by-trial) and global (task) levels.

In future extensions of this modeling work, it would be important to include both local and global attention processes, and also to fit the data to individual participants’ data by feeding the model with the sequence that a particular participant saw, using maximum-likelihood estimation of the best-fitting parameters for each participants’ data. This would allow for a more finely tuned analyses of the differences between the encoding weights for different types of local changes and how this process affects global attention. Another possibility would be to allow earlier trials (beyond $N-1$ as used here) to affect the encoding weights. Although our decision to choose only the immediately preceding item is consistent with empirical evidence showing that people use the information from the previous trial heavily when making a new categorization decision (Jones & Sieck, 2003; Stewart, Brown, & Chater, 2002), influences of preceding items could also be modeled by a function that decays more gradually (Stewart & Brown, 2004), perhaps varying with several factors, such as how easy it is to encode different items (i.e., how confusable the exemplars are). Additionally, the current version of SAT-M assumes that encoding happens post-feedback, when the correct assignment of the current item is known and can be used. Although this assumption is plausible given current evidence in supervised learning, it will be important to test extensions of the model to account for situations where feedback is not presented or delayed.

We developed SAT-M-R as a demonstration that the local context has a direct and important influence on how information is encoded and used for categorization. Of course, this key insight can potentially be instantiated in other models as well. For example, EBRW (Nosofsky & Palmeri, 1997) could be modified such that sampling of stored examples is based not on similarity but on temporal proximity. Likewise, similar encoding parameters could be added to SUSTAIN and ALCOVE. Our goal is not to demonstrate that no other model could possibly ever account for the data SAT-M was developed to account for, but instead to demonstrate that to account for those data, the impact of local encoding context needs to be considered.

The results of the modeling work presented here have direct implications for theories of how the sequence of study influences learning by providing further evidence that a sequential comparison mechanism is plausible and can account for the results presented. We showed that there is no “best sequence” shortcut to learning. Instead, each sequence of study and the local attentional patterns it creates will shape what is learned. To understand whether a sequence is optimal or not, one needs to understand what is encoded during learning and what is required during later training or transfer. A match will result in optimal learning. This proposal casts doubt over alternative proposals that focus on differentiation as the only basis for category learning and advocate for processes that maximize it (Kang & Pashler, 2012), or proposals arguing that the relative benefit of interleaved practice lies on the increased temporal spacing between items of the same category (Birnbaum, Kornell, Bjork, & Bjork, 2013; Kornell & Bjork, 2008) because it demonstrates that blocked study can result in best encoding of features repeated close in time and improved learning in situations that do not require discrimination.

Our work also has implications for theories of category learning. Our model goes beyond conceptualizing attentional changes in category learning as the result of processing statistics over the entire course of study, that is, how the properties of a given item compare to the properties of all items of all categories. This is a powerful conceptual change from taking categorization as a broad learning process to a temporally local process, where each encoding moment is influenced by what just happened. Here, we demonstrated that such a model can capture the effect of different sequences, but it might also capture other effects. Although typically described in the context of memory tasks, the spacing effect has also been shown in the context of category learning (e.g., Birnbaum et al., 2013). One way to conceptualize this effect in terms of changes in local context is to consider that encoding an item (or category of items) among many varied items increases encoding of different features because different features will be more relevant in pairwise comparisons, leading to a more robust encoding of the category and promoting later transfer.

More broadly, SAT-M and the importance of local context for attention modulation and encoding might also account for well-known effects in categorization. For example, local attention changes can provide a mechanism through which category structure modulates category learning (Gureckis & Goldstone, 2008; Livingston & Andrews, 1995). By modulating attention on a trial-by-trial basis, it is likely that category boundaries would be emphasized even if they are not overall relevant for categorization. Similarly, the effect of local context can help explain the transfer of category learning across tasks (Gauthier, James, Curby, & Tarr, 2003), and the effect of different tasks (Trippas & Pachur, 2019) on category learning. These effects can be conceptualized in terms of changes in local context: when an item is learned among a certain type of items or semantic contexts, different properties will be emphasized that matter for that local context which can promote transfer (or hurt it).

Conflict of interest

The authors have no conflict of interest to report.

Funding

Parts of this work were supported by the National Science Foundation (grant #1824257 to PFC) the Department of Education (grant # R305A1100060 to RLG), and the Portuguese Foundation for Science and Technology, co-sponsored by the European Social Fund (Graduate Training Fellowship grant SFRH/BD/78083/2011 to PFC).

Open Research Badges



This article has earned Open Data and Open Materials badges. Data and materials are available at <https://osf.io/s87tf/>, <https://osf.io/2n8gy/>, <https://osf.io/q782h/>.

References

- Aha, D. W., & Goldstone, R. L. (1992). Concept learning and flexible weighting. *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society*, 534–539.
- Anderson, J. R., Matessa, M., & Lebiere, C. (1997). ACT-R: A theory of higher level cognition and its relation to visual attention. *Human–Computer Interaction*, 124, 439–462. https://doi-org.cmu.idm.oclc.org/10.1207/s15327051hci1204_5
- Ashby, F. G., & Maddox, W. T. (1993). Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology*, 37(3), 372–400.
- Ashby, F. G., Paul, E. J., & Maddox, W. T. (2011). COVIS. In E. M. Pothos & A. J. Wills (Eds.), *Formal approaches in categorization* (pp. 65–87). Cambridge University Press. <https://doi.org/10.1017/CBO9780511921322.004>
- Bareiss, E. R., Porter, B. W., & Wier, C. C. (1990). PROTOS: An exemplar-based learning apprentice. In Y. Kodratoff & R. S. Michalski (Eds.), *Machine learning* (pp. 112–127). Morgan Kaufmann. <https://doi.org/10.1016/B978-0-08-051055-2.50009-2>
- Barsalou, L. W., & Medin, D. L. (1986). Concepts: Static definitions or context-dependent representations? *Cahiers de Psychologie*, 6(2), 187–202.
- Birnbaum, M. S., Kornell, N., Bjork, E. L., & Bjork, R. A. (2013). Why interleaving enhances inductive learning: The roles of discrimination and retrieval. *Memory & Cognition*, 41(3), 392–402.
- Blair, M. R., Watson, M. R., Walshe, R. C., & Maj, F. (2009). Extremely selective attention: Eye-tracking studies of the dynamic allocation of attention to stimulus features in categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(5), 1196–1206.
- Bloom, K. C., & Shuell, T. J. (1981). Effects of massed and distributed practice on the learning and retention of second-language vocabulary. *Journal of Educational Research*, 74(4), 245–248.
- Brunmair, M., & Richter, T. (2019). Similarity matters: A meta-analysis of interleaved learning and its moderators. *Psychological Bulletin*, 145(11), 1029–1052. <https://doi.org/10.1037/bul0000209>
- Carpenter, S. K., & Mueller, F. E. (2013). The effects of interleaving versus blocking on foreign language pronunciation learning. *Memory & Cognition*, 41(5), 671–682.
- Carvalho, P. F., & Goldstone, R. L. (2014a). Effects of interleaved and blocked study on delayed test of category learning generalization. *Frontiers in Psychology*, 5, 956. <https://doi.org/10.3389/fpsyg.2014.00936>
- Carvalho, P. F., & Goldstone, R. L. (2014b). Putting category learning in order: Category structure and temporal arrangement affect the benefit of interleaved over blocked study. *Memory & Cognition*, 42(3), 481–495. <https://doi.org/10.3758/s13421-013-0371-0>
- Carvalho, P. F., & Goldstone, R. L. (2015a). The benefits of interleaved and blocked study: Different tasks benefit from different schedules of study. *Psychonomic Bulletin & Review*, 22(1), 281–288. <https://doi.org/10.3758/s13423-014-0676-4>

- Carvalho, P. F., & Goldstone, R. L. (2015b). What you learn is more than what you see: What can sequencing effects tell us about inductive category learning? *Frontiers in Psychology*, 6, 505. <https://doi.org/10.3389/fpsyg.2015.00505>
- Carvalho, P. F., & Goldstone, R. L. (2017). The sequence of study changes what information is attended to, encoded, and remembered during category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(11), 1699–1719. <https://doi.org/10.1037/xlm0000406>
- Chen, L., Meier, K. M., Blair, M. R., Watson, M. R., & Wood, M. J. (2012). Temporal characteristics of overt attentional behavior during category learning. *Attention, Perception, & Psychophysics*, 75(2), 244–256.
- Clapper, J. P. (2014). The impact of training sequence and between-category similarity on unsupervised induction. *Quarterly Journal of Experimental Psychology*, 68(7), 1370–1390. <https://doi.org/10.1080/17470218.2014.981553>
- Corcoran, K., Epstude, K., Damisch, L., & Mussweiler, T. (2011). Fast similarities: Efficiency advantages of similarity-focused comparisons. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(5), 1280–1286. <https://doi.org/10.1037/a0023922>
- Elio, R., & Anderson, J. R. (1984). The effects of information order and learning mode on schema abstraction. *Memory & Cognition*, 12(1), 20–30.
- Gauthier, I., James, W., Curby, K. M., & Tarr, M. J. (2003). The influence of conceptual knowledge on visual discrimination. *Cognitive Neuropsychology*, 20(3), 507–523.
- Goldstone, R. L. (1996). Isolated and interrelated concepts. *Memory & Cognition*, 24(5), 608–628.
- Goldstone, R. L. (2000). Unitization during category learning. *Journal of Experimental Psychology: Human Perception and Performance*, 26(1), 86–112.
- Gureckis, T. M., & Goldstone, R. L. (2008). The effect of the internal structure of categories on perception. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 1876–1881). Cognitive Science Society.
- Helsdingen, A. S., van Gog, T., & Van Merriënboer, J. J. G. (2011). The effects of practice schedule on learning a complex judgment task. *Learning and Instruction*, 21(1), 126–136. <https://doi.org/10.1016/j.learninstruc.2009.12.001>
- Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers*, 16(2), 96–101.
- Hu M., Nosofsky R. M. (2021). Exemplar-model account of categorization and recognition when training instances never repeat. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, <https://doi.org/10.1037/xlm0001008>
- Jones, M., Love, B. C., & Maddox, W. T. (2006). Recency effects as a window to generalization: Separating decisional and perceptual sequential effects in category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(2), 316–332. <https://doi.org/10.1037/0278-7393.32.3.316>
- Jones, M., & Sieck, W. R. (2003). Learning myopia: An adaptive recency effect in category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(4), 626–640. <https://doi.org/10.1037/0278-7393.29.6.1118>
- Kang, S. H. K., & Pashler, H. (2012). Learning painting styles: Spacing is advantageous when it promotes discriminative contrast. *Applied Cognitive Psychology*, 26(1), 97–103.
- Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the “enemy of induction”? *Psychological Science*, 19(6), 585–592.
- Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, 99(1), 22–44. <https://doi.org/10.1037/0033-295X.99.1.22>
- Li, N., Cohen, W. W., & Koedinger, K. R. (2013). Problem order implications for learning. *International Journal of Artificial Intelligence in Education*, 23(1), 71–93. <https://doi.org/10.1007/s40593-013-0005-5>
- Lipsitt, L. P. (1961). Simultaneous and successive discrimination learning in children. *Child Development*, 32(2), 337–347.
- Livingston, K. R., & Andrews, J. K. (1995). On the interaction of prior knowledge and stimulus structure in category learning. *Quarterly Journal of Experimental Psychology Section A*, 48(1), 208–236.

- Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review*, 1112, 309–332.
- Luce, R. D. (1963). Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (pp. 103–189). Wiley.
- Matsuka, T., & Corter, J. E. (2008). Observed attention allocation processes in category learning. *Quarterly Journal of Experimental Psychology*, 617, 1067–1097.
- McDaniel, M. A., Fadler, C. L., & Pashler, H. (2013). Effects of spaced versus massed training in function learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 395, 1417–1432. <https://doi.org/10.1037/a0032184>
- Myung, J. I., Pitt, M. A., & Navarro, D. J. (2007). Does response scaling cause the generalized context model to mimic a prototype model? *Psychonomic Bulletin & Review*, 146, 1043–1050.
- Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, 1151, 39–61.
- Nosofsky, R. M. (1991). Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception and Performance*, 171, 3–27.
- Nosofsky, R. M. (2011). The generalized context model: An exemplar model of classification. In E. M. Pothos & A. J. Wills (Eds.), *Formal approaches in categorization* (pp. 18–39). Cambridge University Press.
- Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, 1042, 266–300.
- Nosofsky, R. M., & Zaki, S. R. (2002). Exemplar and prototype models revisited: Response strategies, selective attention, and stimulus generalization. *Journal of Experimental Psychology: Learning Memory and Cognition*, 285, 924–940. <https://doi.org/10.1037/0278-7393.28.5.924>
- Palmeri, T. J., & Mack, M. L. (2015). How experimental trial context affects perceptual categorization. *Frontiers in Psychology*, 6, 180. <https://doi.org/10.3389/fpsyg.2015.00180>
- Palmeri, T. J., & Nosofsky, R. (1995). Recognition memory for exceptions to the category rule. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 213, 548–568.
- Qian, T., & Aslin, R. N. (2014). Learning bundles of stimuli renders stimulus order as a cue, not a confound. *Proceedings of the National Academy of Sciences*, 11140, 14400–14405. <https://doi.org/10.1073/pnas.1416109111>
- R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Rehder, B., & Hoffman, A. B. (2005a). Eyetracking and selective attention in category learning. *Cognitive Psychology*, 511, 1–41.
- Rehder, B., & Hoffman, A. B. (2005b). Thirty-something categorization results explained: Selective attention, eyetracking, and models of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 315, 811–829. <https://doi.org/10.1037/0278-7393.31.5.811>
- Samuels, S. J. (1969). Effect of simultaneous versus successive discrimination training on paired-associate learning. *Journal of Educational Psychology*, 601, 46–48. <https://doi.org/10.1037/h0026671>
- Sandhofer, C. M., & Dumas, L. A. A. (2008). Order of presentation effects in learning color categories. *Journal of Cognition and Development*, 92, 194–221. <https://doi.org/10.1080/15248370802022639>
- Schyns, P. G., & Rodet, L. (1997). Categorization creates functional features. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 233, 681–696.
- Shepard, R. N. (1965). Approximation to uniform gradients of generalization by monotone transformation of scale BT — Stimulus generalization. In D. I. Mostofsky (Ed.), *Stimulus generalization* (pp. 94–110). Stanford University Press.
- Smith, J. D., & Minda, J. P. (2002). Distinguishing prototype-based and exemplar-based processes in dot-pattern category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 284, 800–811.
- Stewart, N., & Brown, G. D. A. (2004). Sequence effects in the categorization of tones varying in frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 302, 416–430.
- Stewart, N., Brown, G. D. A., & Chater, N. (2002). Sequence effects in categorization of simple perceptual stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 281, 3–11.

- Thiessen, E. D., & Pavlik, P. I. (2013). iMinerva: A mathematical model of distributional statistical learning. *Cognitive Science*, 37(2), 310–343. <https://doi.org/10.1111/cogs.12011>
- Trippas, D., & Pachur, T. (2019). Nothing compares: Unraveling learning task effects in judgment and categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45(12), 2239–2266. <http://doi.org/10.1037/xlm0000696>
- Tversky, A. (1977). Features of similarity. *Psychological Review*, 84(4), 327–352.
- Wahlheim, C. N., Dunlosky, J., & Jacoby, L. L. (2011). Spacing enhances the learning of natural concepts: An investigation of mechanisms, metacognition, and aging. *Memory & Cognition*, 39(5), 750–763.
- Wang, T., & Mitchell, C. J. (2011). Attention and relative novelty in human perceptual learning. *Journal of Experimental Psychology: Animal Behavior Processes*, 37(4), 436–445. <https://doi.org/10.1037/a0023104>
- Wills, A. J. (2013). Models of categorization. In D. Reisberg (Ed.), *Oxford handbook of cognitive psychology*, (pp. 346–357). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780195376746.013.0023>
- Wills, A. J., O’Connell, G., Edmunds, C. E. R., & Inkster, A. B. (2017). Progress in modeling through distributed collaboration. In Brian H. Ross (Ed.), *Psychology of learning and motivation* (Vol. 66, pp. 79–115). Academic Press. <https://doi.org/10.1016/bs.plm.2016.11.007>
- Zaki, S. R., & Salmi, I. L. (2019). Sequence as context in category learning: An eyetracking study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45(11), 1945–1954. <https://doi.org/10.1037/xlm0000693>
- Zeithamova, D., & Maddox, W. T. (2009). Learning mode and exemplar sequencing in unsupervised category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(3), 731–741. <https://doi.org/10.1037/a0015005>
- Zotov, V., Jones, M. N., & Mewhort, D. J. K. (2011). Contrast and assimilation in categorization and exemplar production. *Attention, Perception, & Psychophysics*, 73(2), 621–639. <https://doi.org/10.3758/s13414-010-0036-z>
- Zulkiply, N., & Burt, J. S. (2013). The exemplar interleaving effect in inductive learning: Moderation by the difficulty of category discriminations. *Memory & Cognition*, 41(1), 16–27.

Appendix A

Table 1 Representation of the category structure used in Carvalho and Goldstone (2014b)

D1	D2	D3	D4	D5	D6	D7	D8	Category	Similarity structure	Item
1	2	2	2	2	2	1	1	1	Low similarity	LS_101
1	3	4	3	3	3	2	2	1	Low similarity	LS_102
1	4	5	4	4	4	3	3	1	Low similarity	LS_103
1	5	6	5	5	5	4	4	1	Low similarity	LS_104
1	6	7	6	6	6	5	5	1	Low similarity	LS_105
1	7	8	7	7	7	6	6	1	Low similarity	LS_106
1	8	9	8	8	8	7	7	1	Low similarity	LS_107
1	9	10	9	9	9	8	8	1	Low similarity	LS_108
1	10	11	10	10	10	9	9	1	Low similarity	LS_109
1	11	12	11	11	11	10	10	1	Low similarity	LS_110

(Continues)

D1	D2	D3	D4	D5	D6	D7	D8	Category	Similarity structure	Item
1	12	13	12	12	12	11	11	1	Low similarity	LS_111
1	13	14	13	13	13	12	12	1	Low similarity	LS_112
1	14	15	14	14	14	13	13	1	Low similarity	LS_113
1	15	16	15	15	15	14	14	1	Low similarity	LS_114
1	16	17	16	16	16	15	15	1	Low similarity	LS_115
1	17	18	17	17	17	16	16	1	Low similarity	LS_116
2	1	19	18	18	18	17	17	2	Low similarity	LS_201
3	1	20	19	19	19	18	18	2	Low similarity	LS_202
4	1	21	20	20	20	19	19	2	Low similarity	LS_203
5	1	22	21	21	21	20	20	2	Low similarity	LS_204
6	1	23	22	22	22	21	21	2	Low similarity	LS_205
7	1	24	23	23	23	22	22	2	Low similarity	LS_206
8	1	25	24	24	24	23	23	2	Low similarity	LS_207
9	1	26	25	25	25	24	24	2	Low similarity	LS_208
10	1	27	26	26	26	25	25	2	Low similarity	LS_209
11	1	28	27	27	27	26	26	2	Low similarity	LS_210
12	1	29	28	28	28	27	27	2	Low similarity	LS_211
13	1	30	29	29	29	28	28	2	Low similarity	LS_212
14	1	31	30	30	30	29	29	2	Low similarity	LS_213
15	1	32	31	31	31	30	30	2	Low similarity	LS_214
16	1	33	32	32	32	31	31	2	Low similarity	LS_215
17	1	34	33	33	33	32	32	2	Low similarity	LS_216
18	18	1	34	34	34	33	33	3	Low similarity	LS_301
19	19	1	35	35	35	34	34	3	Low similarity	LS_302
20	20	1	36	36	36	35	35	3	Low similarity	LS_303
21	21	1	37	37	37	36	36	3	Low similarity	LS_304
22	22	1	38	38	38	37	37	3	Low similarity	LS_305
23	23	1	39	39	39	38	38	3	Low similarity	LS_306
24	24	1	40	40	40	39	39	3	Low similarity	LS_307
25	25	1	41	41	41	40	40	3	Low similarity	LS_308
26	26	1	42	42	42	41	41	3	Low similarity	LS_309
27	27	1	43	43	43	42	42	3	Low similarity	LS_310
28	28	1	44	44	44	43	43	3	Low similarity	LS_311
29	29	1	45	45	45	44	44	3	Low similarity	LS_312
30	30	1	46	46	46	45	45	3	Low similarity	LS_313
31	31	1	47	47	47	46	46	3	Low similarity	LS_314
32	32	1	48	48	48	47	47	3	Low similarity	LS_315
33	33	1	49	49	49	48	48	3	Low similarity	LS_316
1	2	2	2	4	3	2	2	1	High similarity	HS_101
1	2	3	2	3	3	2	2	1	High similarity	HS_102
1	2	4	3	2	4	2	2	1	High similarity	HS_103
1	2	2	3	4	4	3	3	1	High similarity	HS_104
1	3	3	4	3	2	3	3	1	High similarity	HS_105
1	3	4	4	2	2	3	3	1	High similarity	HS_106

(Continues)

D1	D2	D3	D4	D5	D6	D7	D8	Category	Similarity structure	Item
1	3	2	2	4	4	4	4	1	High similarity	HS_107
1	3	3	2	3	4	4	4	1	High similarity	HS_108
1	4	4	3	2	3	4	4	1	High similarity	HS_109
1	4	2	3	4	3	4	4	1	High similarity	HS_110
1	4	3	4	3	3	4	4	1	High similarity	HS_111
1	4	4	4	2	3	4	3	1	High similarity	HS_112
1	4	2	2	4	2	2	2	1	High similarity	HS_113
1	4	3	2	3	2	2	3	1	High similarity	HS_114
1	4	4	3	2	2	2	2	1	High similarity	HS_115
1	4	2	3	3	2	2	3	1	High similarity	HS_116
2	1	2	2	4	3	2	2	2	High similarity	HS_201
2	1	3	2	3	3	2	2	2	High similarity	HS_202
2	1	4	3	2	4	2	2	2	High similarity	HS_203
2	1	2	3	4	4	3	3	2	High similarity	HS_204
3	1	3	4	3	2	3	3	2	High similarity	HS_205
3	1	4	4	2	2	3	3	2	High similarity	HS_206
3	1	2	2	4	4	4	4	2	High similarity	HS_207
3	1	3	2	3	4	4	4	2	High similarity	HS_208
4	1	4	3	2	3	4	4	2	High similarity	HS_209
4	1	2	3	4	3	4	4	2	High similarity	HS_210
4	1	3	4	3	3	4	4	2	High similarity	HS_211
4	1	4	4	2	3	4	3	2	High similarity	HS_212
4	1	2	2	4	2	2	2	2	High similarity	HS_213
4	1	3	2	3	2	2	3	2	High similarity	HS_214
4	1	4	3	2	2	2	2	2	High similarity	HS_215
4	1	2	3	3	2	2	3	2	High similarity	HS_216
2	2	1	2	4	3	2	2	3	High similarity	HS_301
2	3	1	2	3	3	2	2	3	High similarity	HS_302
2	4	1	3	2	4	2	2	3	High similarity	HS_303
2	2	1	3	4	4	3	3	3	High similarity	HS_304
3	3	1	4	3	2	3	3	3	High similarity	HS_305
3	4	1	4	2	2	3	3	3	High similarity	HS_306
3	2	1	2	4	4	4	4	3	High similarity	HS_307
3	3	1	2	3	4	4	4	3	High similarity	HS_308
4	4	1	3	2	3	4	4	3	High similarity	HS_309
4	2	1	3	4	3	4	4	3	High similarity	HS_310
4	3	1	4	3	3	4	4	3	High similarity	HS_311
4	4	1	4	2	3	4	3	3	High similarity	HS_312
4	2	1	2	4	2	2	2	3	High similarity	HS_313
4	3	1	2	3	2	2	3	3	High similarity	HS_314
4	4	1	3	2	2	2	2	3	High similarity	HS_315
4	2	1	3	3	2	2	3	3	High similarity	HS_316

Note: Numbers represent a specific feature value on each dimension. They represent independent feature values across dimensions (i.e., a 2 on dimension 1 is unrelated to a 2 on dimension 2). Each line represents a unique item.

Appendix B

Category structure for the stimuli used in Experiment 3 of Carvalho and Goldstone (2017)

Category	Item	Dimension 1	Dimension 2	Dimension 3	Dimension 4	Dimension 5
A	1	2 (1, 0.3)	1 (0.5, 0.7)	1 (0.5, 0.7)	1 (0.5, 0.7)	4 (0.5, 0.3)
A	2	2 (1, 0.3)	1 (0.5, 0.7)	1 (0.5, 0.7)	1 (0.5, 0.7)	5 (0.5, 0.3)
A	3	2 (1, 0.3)	2 (1, 0.3)	1 (0.5, 0.7)	1 (0.5, 0.7)	3 (0.5, 0.3)
A	4	1 (0.5, 0.7)	2 (1, 0.3)	1 (0.5, 0.7)	1 (0.5, 0.7)	3 (0.5, 0.3)
A	5	1 (0.5, 0.7)	2 (1, 0.3)	2 (1, 0.3)	1 (0.5, 0.7)	5 (0.5, 0.3)
A	6	1 (0.5, 0.7)	1 (0.5, 0.7)	2 (1, 0.3)	1 (0.5, 0.7)	4 (0.5, 0.3)
A	7	1 (0.5, 0.7)	1 (0.5, 0.7)	2 (1, 0.3)	2 (1, 0.3)	4 (0.5, 0.3)
A	8	1 (0.5, 0.7)	1 (0.5, 0.7)	1 (0.5, 0.7)	2 (1, 0.3)	3 (0.5, 0.3)
A	9	1 (0.5, 0.7)	1 (0.5, 0.7)	1 (0.5, 0.7)	2 (1, 0.3)	5 (0.5, 0.3)
B	1	3 (1, 0.3)	1 (0.5, 0.7)	1 (0.5, 0.7)	1 (0.5, 0.7)	4 (0.5, 0.3)
B	2	3 (1, 0.3)	1 (0.5, 0.7)	1 (0.5, 0.7)	1 (0.5, 0.7)	5 (0.5, 0.3)
B	3	3 (1, 0.3)	3 (1, 0.3)	1 (0.5, 0.7)	1 (0.5, 0.7)	3 (0.5, 0.3)
B	4	1 (0.5, 0.7)	3 (1, 0.3)	1 (0.5, 0.7)	1 (0.5, 0.7)	3 (0.5, 0.3)
B	5	1 (0.5, 0.7)	3 (1, 0.3)	3 (1, 0.3)	1 (0.5, 0.7)	5 (0.5, 0.3)
B	6	1 (0.5, 0.7)	1 (0.5, 0.7)	3 (1, 0.3)	1 (0.5, 0.7)	4 (0.5, 0.3)
B	7	1 (0.5, 0.7)	1 (0.5, 0.7)	3 (1, 0.3)	3 (1, 0.3)	4 (0.5, 0.3)
B	8	1 (0.5, 0.7)	1 (0.5, 0.7)	1 (0.5, 0.7)	3 (1, 0.3)	3 (0.5, 0.3)
B	9	1 (0.5, 0.7)	1 (0.5, 0.7)	1 (0.5, 0.7)	3 (1, 0.3)	5 (0.5, 0.3)

Note: Numbers represent a specific feature value on each dimension. They represent independent feature values across dimensions (i.e., a 2 on dimension 1 is unrelated to a 2 on dimension 2). A value of 2 or 3 is always a discriminative feature, whereas a value of 1 is a characteristic feature. Which part (eyes, legs, arms, antenna, and mouth) corresponded to each dimension was counterbalanced across participants. Cue and category validity values for each feature value are presented in parentheses (cue validity and category validity).

Appendix C

Category structure for the stimuli used in Experiment 1 of Carpenter and Mueller (2013)

Word	Rule	D1	D2	D3	D4	D5	D6	D7	D8
bateau	eau	1	1	1	1	1	1	1	1
carreu	eau	2	2	2	2	2	1	1	1
corbeau	eau	3	3	3	3	3	1	1	1
fardeau	eau	4	4	4	4	4	1	1	1
panneau	eau	5	5	5	5	5	1	1	1
rameau	eau	6	6	6	6	6	1	1	1
tableau	eau	7	7	7	7	7	1	1	1
tonneau	eau	8	8	8	8	8	1	1	1
adosser	er	9	9	9	9	9	1	1	1
attraper	er	10	10	10	10	10	3	2	2
baver	er	11	11	11	11	11	4	2	2

(Continues)

Word	Rule	D1	D2	D3	D4	D5	D6	D7	D8
combler	er	12	12	12	12	12	5	2	2
darder	er	13	13	13	13	13	6	2	2
raconter	er	14	14	14	14	14	7	2	2
tamiser	er	15	15	15	15	15	8	2	2
valser	er	16	16	16	16	16	9	2	2
allimette	e	17	17	17	17	17	10	6	3
cervelle	e	18	18	18	18	18	11	7	3
carrolle	e	19	19	19	19	19	12	8	3
cuvette	e	20	20	20	20	20	13	9	3
emplette	e	21	21	21	21	21	14	10	3
frappe	e	22	22	22	22	22	15	11	3
malade	e	23	23	23	23	23	16	12	3
serpe	e	24	24	24	24	24	17	13	3
bouton	ou	25	25	25	25	25	18	3	4
genou	ou	26	26	26	26	26	19	3	4
goudron	ou	27	27	27	27	27	20	3	4
gourou	ou	28	28	28	28	28	21	3	4
mouton	ou	29	29	29	29	29	22	3	4
tabou	ou	30	30	30	30	30	23	3	4
verrou	ou	31	31	31	31	31	24	3	4
voulu	ou	32	32	32	32	32	25	3	4
archipel	ch	33	33	33	33	33	26	4	5
capuchon	ch	34	34	34	34	34	27	4	5
chacal	ch	35	35	35	35	35	28	4	5
chardon	ch	36	36	36	36	36	29	4	5
charnel	ch	37	37	37	37	37	30	4	5
chiffon	ch	38	38	38	38	38	31	4	5
cochon	ch	39	39	39	39	39	32	4	5
fichu	ch	40	40	40	40	40	33	4	5
brumeux	eux	41	41	41	41	41	2	5	6
neveux	eux	42	42	42	42	42	2	5	6
osseux	eux	43	43	43	43	43	2	5	6
paresseux	eux	44	44	44	44	44	2	5	6
sablonneux	eux	45	45	45	45	45	2	5	6
somptueux	eux	46	46	46	46	46	2	5	6
vaniteux	eux	47	47	47	47	47	2	5	6
venimeux	eux	48	48	48	48	48	2	5	6
admis	s	49	49	49	49	49	34	14	7
brebis	s	50	50	50	50	50	35	15	7
coloris	s	51	51	51	51	51	36	16	7

(Continues)

Word	Rule	D1	D2	D3	D4	D5	D6	D7	D8
compris	s	52	52	52	52	52	37	17	7
lambris	s	53	53	53	53	53	38	18	7
tandis	s	54	54	54	54	54	39	19	7
verglas	s	55	55	55	55	55	40	20	7
vernis	s	56	56	56	56	56	41	21	7
brevet	t	57	57	57	57	57	42	22	8
carnet	t	58	58	58	58	58	43	23	8
pavot	t	59	59	59	59	59	44	24	8
rabot	t	60	60	60	60	60	45	25	8
sifflet	t	61	61	61	61	61	46	26	8
soldat	t	62	62	62	62	62	47	27	8
sommet	t	63	63	63	63	63	48	28	8
tricot	t	64	64	64	64	64	49	29	8

Note: Numbers represent a specific feature value on each dimension. They represent independent feature values across dimensions (i.e., a 2 on dimension 1 is unrelated to a 2 on dimension 2). Each line represents a unique item.

Appendix D

Category structure for the stimuli used in Experiment 2 of Zulkipli and Burt (2013)

High similarity set:

F1	F2	F3	F4	F5	F6	F7	F8	F9	Category
1	1	1	1	2	1	1	1	1	1
1	1	1	2	3	1	1	1	1	1
1	1	1	1	4	1	1	1	1	1
1	1	1	2	5	1	1	1	1	1
1	1	1	1	6	1	1	1	1	1
1	1	1	2	7	1	1	1	1	1
1	1	1	1	8	1	1	1	1	1
1	1	1	2	9	1	1	1	1	1
1	1	1	1	10	1	1	1	1	1
1	1	1	2	11	1	1	1	1	1
1	2	1	1	2	1	1	1	1	2
1	2	1	2	3	1	1	1	1	2
1	2	1	1	4	1	1	1	1	2

(Continues)

F1	F2	F3	F4	F5	F6	F7	F8	F9	Category
1	2	1	2	5	1	1	1	1	2
1	2	1	1	6	1	1	1	1	2
1	2	1	2	7	1	1	1	1	2
1	2	1	1	8	1	1	1	1	2
1	2	1	2	9	1	1	1	1	2
1	2	1	1	10	1	1	1	1	2
1	2	1	2	11	1	1	1	1	2
1	3	1	1	2	1	1	1	1	3
1	3	1	2	3	1	1	1	1	3
1	3	1	1	4	1	1	1	1	3
1	3	1	2	5	1	1	1	1	3
1	3	1	1	6	1	1	1	1	3
1	3	1	2	7	1	1	1	1	3
1	3	1	1	8	1	1	1	1	3
1	3	1	2	9	1	1	1	1	3
1	3	1	1	10	1	1	1	1	3
1	3	1	2	11	1	1	1	1	3
1	4	1	1	2	1	1	1	1	4
1	4	1	2	3	1	1	1	1	4
1	4	1	1	4	1	1	1	1	4
1	4	1	2	5	1	1	1	1	4
1	4	1	1	6	1	1	1	1	4
1	4	1	2	7	1	1	1	1	4
1	4	1	1	8	1	1	1	1	4
1	4	1	2	9	1	1	1	1	4
1	4	1	1	10	1	1	1	1	4
1	4	1	2	11	1	1	1	1	4
1	5	1	1	2	1	1	1	1	5
1	5	1	2	3	1	1	1	1	5
1	5	1	1	4	1	1	1	1	5
1	5	1	2	5	1	1	1	1	5
1	5	1	1	6	1	1	1	1	5
1	5	1	2	7	1	1	1	1	5
1	5	1	1	8	1	1	1	1	5
1	5	1	2	9	1	1	1	1	5
1	5	1	1	10	1	1	1	1	5
1	5	1	2	11	1	1	1	1	5
1	6	1	1	2	1	1	1	1	6
1	6	1	2	3	1	1	1	1	6
1	6	1	1	4	1	1	1	1	6
1	6	1	2	5	1	1	1	1	6
1	6	1	1	6	1	1	1	1	6

(Continues)

F1	F2	F3	F4	F5	F6	F7	F8	F9	Category
1	6	1	2	7	1	1	1	1	6
1	6	1	1	8	1	1	1	1	6
1	6	1	2	9	1	1	1	1	6
1	6	1	1	10	1	1	1	1	6
1	6	1	2	11	1	1	1	1	6
2	1	1	1	2	1	1	1	1	7
2	1	1	2	3	1	1	1	1	7
2	1	1	1	4	1	1	1	1	7
2	1	1	2	5	1	1	1	1	7
2	1	1	1	6	1	1	1	1	7
2	1	1	2	7	1	1	1	1	7
2	1	1	1	8	1	1	1	1	7
2	1	1	2	9	1	1	1	1	7
2	1	1	1	10	1	1	1	1	7
2	1	1	2	11	1	1	1	1	7
2	2	1	1	2	1	1	1	1	8
2	2	1	2	3	1	1	1	1	8
2	2	1	1	4	1	1	1	1	8
2	2	1	2	5	1	1	1	1	8
2	2	1	1	6	1	1	1	1	8
2	2	1	2	7	1	1	1	1	8
2	2	1	1	8	1	1	1	1	8
2	2	1	2	9	1	1	1	1	8
2	2	1	1	10	1	1	1	1	8
2	2	1	2	11	1	1	1	1	8
2	3	1	1	2	1	1	1	1	9
2	3	1	2	3	1	1	1	1	9
2	3	1	1	4	1	1	1	1	9
2	3	1	2	5	1	1	1	1	9
2	3	1	1	6	1	1	1	1	9
2	3	1	2	7	1	1	1	1	9
2	3	1	1	8	1	1	1	1	9
2	3	1	2	9	1	1	1	1	9
2	3	1	1	10	1	1	1	1	9
2	3	1	2	11	1	1	1	1	9
2	4	1	1	2	1	1	1	1	10
2	4	1	2	3	1	1	1	1	10
2	4	1	1	4	1	1	1	1	10
2	4	1	2	5	1	1	1	1	10
2	4	1	1	6	1	1	1	1	10
2	4	1	2	7	1	1	1	1	10
2	4	1	1	8	1	1	1	1	10
2	4	1	2	9	1	1	1	1	10
2	4	1	1	10	1	1	1	1	10
2	4	1	2	11	1	1	1	1	10

(Continues)

F1	F2	F3	F4	F5	F6	F7	F8	F9	Category
2	5	1	1	2	1	1	1	1	11
2	5	1	2	3	1	1	1	1	11
2	5	1	1	4	1	1	1	1	11
2	5	1	2	5	1	1	1	1	11
2	5	1	1	6	1	1	1	1	11
2	5	1	2	7	1	1	1	1	11
2	5	1	1	8	1	1	1	1	11
2	5	1	2	9	1	1	1	1	11
2	5	1	1	10	1	1	1	1	11
2	5	1	2	11	1	1	1	1	11
2	6	1	1	2	1	1	1	1	12
2	6	1	2	3	1	1	1	1	12
2	6	1	1	4	1	1	1	1	12
2	6	1	2	5	1	1	1	1	12
2	6	1	1	6	1	1	1	1	12
2	6	1	2	7	1	1	1	1	12
2	6	1	1	8	1	1	1	1	12
2	6	1	2	9	1	1	1	1	12
2	6	1	1	10	1	1	1	1	12
2	6	1	2	11	1	1	1	1	12

Low similarity set:

F1	F2	F3	F4	F5	F6	F7	F8	F9	category
1	1	1	1	2	2	2	2	2	1
1	1	2	2	3	3	3	3	3	1
1	1	1	3	4	4	4	4	4	1
1	1	2	4	5	5	5	5	5	1
1	1	1	1	6	6	6	6	6	1
1	1	2	2	7	7	7	7	7	1
1	1	1	3	8	8	8	8	8	1
1	1	2	4	9	9	9	9	9	1
1	1	1	1	10	10	10	10	10	1
1	1	2	2	11	11	11	11	11	1

(Continues)

F1	F2	F3	F4	F5	F6	F7	F8	F9	category
1	2	1	1	2	2	2	2	2	2
1	2	2	2	3	3	3	3	3	2
1	2	1	3	4	4	4	4	4	2
1	2	2	4	5	5	5	5	5	2
1	2	1	1	6	6	6	6	6	2
1	2	2	2	7	7	7	7	7	2
1	2	1	3	8	8	8	8	8	2
1	2	2	4	9	9	9	9	9	2
1	2	1	1	10	10	10	10	10	2
1	2	2	2	11	11	11	11	11	2
1	3	1	1	2	2	2	2	2	3
1	3	2	2	3	3	3	3	3	3
1	3	1	3	4	4	4	4	4	3
1	3	2	4	5	5	5	5	5	3
1	3	1	1	6	6	6	6	6	3
1	3	2	2	7	7	7	7	7	3
1	3	1	3	8	8	8	8	8	3
1	3	2	4	9	9	9	9	9	3
1	3	1	1	10	10	10	10	10	3
1	3	2	2	11	11	11	11	11	3
1	4	1	1	2	2	2	2	2	4
1	4	2	2	3	3	3	3	3	4
1	4	1	3	4	4	4	4	4	4
1	4	2	4	5	5	5	5	5	4
1	4	1	1	6	6	6	6	6	4
1	4	2	2	7	7	7	7	7	4
1	4	1	3	8	8	8	8	8	4
1	4	2	4	9	9	9	9	9	4
1	4	1	1	10	10	10	10	10	4
1	4	2	2	11	11	11	11	11	4
1	5	1	1	2	2	2	2	2	5
1	5	2	2	3	3	3	3	3	5
1	5	1	3	4	4	4	4	4	5
1	5	2	4	5	5	5	5	5	5
1	5	1	1	6	6	6	6	6	5
1	5	2	2	7	7	7	7	7	5
1	5	1	3	8	8	8	8	8	5
1	5	2	4	9	9	9	9	9	5
1	5	1	1	10	10	10	10	10	5
1	5	2	2	11	11	11	11	11	5
1	6	1	1	2	2	2	2	2	6
1	6	2	2	3	3	3	3	3	6
1	6	1	3	4	4	4	4	4	6

(Continues)

F1	F2	F3	F4	F5	F6	F7	F8	F9	category
1	6	2	4	5	5	5	5	5	6
1	6	1	1	6	6	6	6	6	6
1	6	2	2	7	7	7	7	7	6
1	6	1	3	8	8	8	8	8	6
1	6	2	4	9	9	9	9	9	6
1	6	1	1	10	10	10	10	10	6
1	6	2	2	11	11	11	11	11	6
2	1	1	1	2	2	2	2	2	7
2	1	2	2	3	3	3	3	3	7
2	1	1	3	4	4	4	4	4	7
2	1	2	4	5	5	5	5	5	7
2	1	1	1	6	6	6	6	6	7
2	1	2	2	7	7	7	7	7	7
2	1	1	3	8	8	8	8	8	7
2	1	2	4	9	9	9	9	9	7
2	1	1	1	10	10	10	10	10	7
2	1	2	2	11	11	11	11	11	7
2	2	1	1	2	2	2	2	2	8
2	2	2	2	3	3	3	3	3	8
2	2	1	3	4	4	4	4	4	8
2	2	2	4	5	5	5	5	5	8
2	2	1	1	6	6	6	6	6	8
2	2	2	2	7	7	7	7	7	8
2	2	1	3	8	8	8	8	8	8
2	2	2	4	9	9	9	9	9	8
2	2	1	1	10	10	10	10	10	8
2	2	2	2	11	11	11	11	11	8
2	3	1	1	2	2	2	2	2	9
2	3	2	2	3	3	3	3	3	9
2	3	1	3	4	4	4	4	4	9
2	3	2	4	5	5	5	5	5	9
2	3	1	1	6	6	6	6	6	9
2	3	2	2	7	7	7	7	7	9
2	3	1	3	8	8	8	8	8	9
2	3	2	4	9	9	9	9	9	9
2	3	1	1	10	10	10	10	10	9
2	3	2	2	11	11	11	11	11	9
2	4	1	1	2	2	2	2	2	10
2	4	2	2	3	3	3	3	3	10
2	4	1	3	4	4	4	4	4	10
2	4	2	4	5	5	5	5	5	10
2	4	1	1	6	6	6	6	6	10
2	4	2	2	7	7	7	7	7	10
2	4	1	3	8	8	8	8	8	10
2	4	2	4	9	9	9	9	9	10
2	4	1	1	10	10	10	10	10	10

(Continues)

F1	F2	F3	F4	F5	F6	F7	F8	F9	category
2	4	2	2	11	11	11	11	11	10
2	5	1	1	2	2	2	2	2	11
2	5	2	2	3	3	3	3	3	11
2	5	1	3	4	4	4	4	4	11
2	5	2	4	5	5	5	5	5	11
2	5	1	1	6	6	6	6	6	11
2	5	2	2	7	7	7	7	7	11
2	5	1	3	8	8	8	8	8	11
2	5	2	4	9	9	9	9	9	11
2	5	1	1	10	10	10	10	10	11
2	5	2	2	11	11	11	11	11	11
2	6	1	1	2	2	2	2	2	12
2	6	2	2	3	3	3	3	3	12
2	6	1	3	4	4	4	4	4	12
2	6	2	4	5	5	5	5	5	12
2	6	1	1	6	6	6	6	6	12
2	6	2	2	7	7	7	7	7	12
2	6	1	3	8	8	8	8	8	12
2	6	2	4	9	9	9	9	9	12
2	6	1	1	10	10	10	10	10	12
2	6	2	2	11	11	11	11	11	12