1-1-2015

Interleaved Effects in Inductive Category Learning: The Role of Memory Retention

Alex MacKendrick
University of South Florida, amackend@mail.usf.edu

Interleaved Effects in Inductive Category Learning: The Role of Memory Retention

by

Alex MacKendrick

A dissertation submitted in partial fulfillment
of the requirement for the degree of
Doctor of Philosophy
in Curriculum and Instruction with an emphasis in
Interdisciplinary Education
College of Education
University of South Florida

Co-Major Professor: Stephen Blessing, Ph.D.
Co-Major Professor: Robert Dedrick, Ph.D.
Emanuel Donchin, Ph.D.
Sarah Kiefer, Ph.D.
Tony Xing Tan, Ed.D.

Date of Approval:
March 27, 2015

Keywords: interleaving; memory retention; discrimination; spacing; category learning

Acknowledgments

Foremost, I would like to express my sincerest gratitude to my advisor, Dr. Stephen Blessing. I had the good fortune to work with Steve these past few years and he has been an invaluable friend and mentor. I also thank my co-advisor, Dr. Robert Dedrick for his gracious support and advisement. I am grateful to Dr. Tim Lee for his unwavering encouragement and to Dr. Emanuel Donchin for his indispensable collaboration. Finally, I thank Drs. Sarah Kiefer and Tony Tan for their continuous support.

Table of Contents

List of Tables

List of Figures

Abstract

Interleaved effects are widely documented. Research demonstrates that interleaved presentation orders, as opposed to blocked orders typically benefit inductive category learning. What drives interleaved effects is less straightforward. Interleaved presentations provide both the opportunity to compare and contrast between different types of category exemplars, which are temporally juxtaposed, and the opportunity to space study of the same type of category exemplars, which are temporally separated within the presentation span. Accordingly, interleaved effects might be driven by enhanced discrimination, enhanced memory retention, or both in some measure. Though recent studies have largely endorsed enhanced discrimination as the critical mechanism driving interleaved effects, there is no strong evidence to controvert the contribution of enhanced memory retention for interleaved effects. I further examined the role of memory retention by manipulating both presentation order and category structure. Across two experiments I found that memory retention may drive interleaved effects in categorization tasks.

Introduction[1]

One of the major issues in cognitive psychology is the study of how humans learn. Over the past century cognitive psychologists have greatly advanced our understanding of the conditions that promote learning. Ideally, this research would be effectively used to support classroom instruction via the *bench-to-trench* process of transferring knowledge derived from the scientific bench to the trenches where practitioners may apply that scientifically derived knowledge (see Proctor 2003; Thase, 2006; Whitehurst, 2003). However the transition of these findings to classrooms and curriculums has been slow and frequently lost in translation – a *bench-to-trench gap*. Recently however, a cohort of researchers (e.g.Dunlosky et al., 2013; Pashler et al., 2007; Roediger & Karpicke, 2006) have turned their attention towards bridging this bench-to-trench gap with a swell of research that seeks to bridge the mechanics of learning with the mechanics of instruction.

For instance, how to present information to maximize learning and memory is central to cognition and fundamental for most educational contexts. Accordingly, one area that has received attention recently is the investigation of how presentation orders may be used to enhance inductive category learning. In this context, learning by induction entails learning by example. Research has shown that interleaved presentation orders, which alternate the presentation of category examples, or exemplars, from a set of to-be-learned categories (e.g., ABC - BCA - CAB), typically produce better inductive category learning when compared to blocked presentation orders, which group the presentation of category exemplars (e.g., AAA -

[1] The text of this dissertation is predominantly a reprint of the material submitted to *Applied Cognitive Psychology* (under review). The co-author listed in this article supervised the research which forms the basis for the dissertation.

BBB - CCC). For instance, Kornell and Bjork (2008) demonstrated that interleaving was superior for generalization of similar style paintings by different artists when compared to blocking. This finding was later replicated using the same type of artist identification task (e.g., Kang & Pashler, 2012; Zulkiply & Burt 2013, Exp. 1), natural stimuli (e.g., birds and butterflies, Birnbaum, Kornell, Bjork, & Bjork, 2013; birds, Wahlheim, Dunlosky, & Jacoby, 2011), math problems (e.g., geometric prisms, Taylor & Rohrer, 2010), and artificial stimuli (e.g., blobs, Carvalho & Goldstone, 2014; abstract pictures, Zulkiply & Burt 2013, Exp. 2).

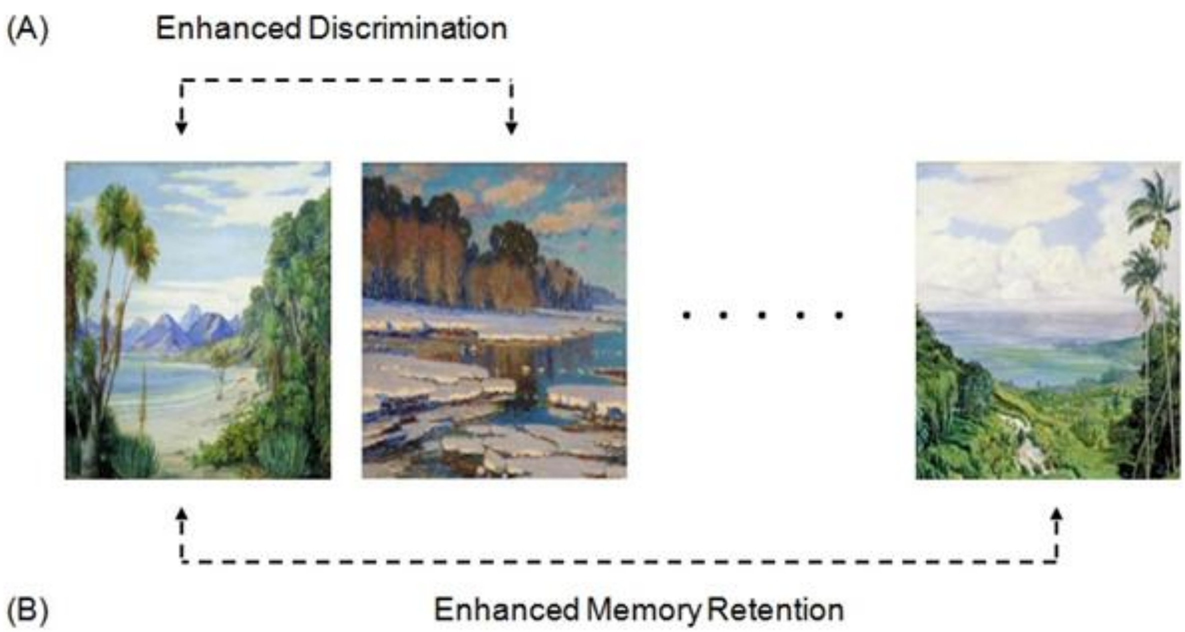


Figure 1. Illustrates the separate mechanisms that may drive interleaved effects. The temporal juxtaposition of different category exemplars afforded by interleaved presentations may facilitate learning of distinctive attributes between categories (A). Alternatively, the temporal separation of the same category exemplars afforded by interleaved presentations may facilitate memory for relevant, similar attributes within a category (B).

Such interleaved effects are widely documented, but a natural confound exists within these studies. Interleaved presentations provide both the opportunity to compare and contrast between different types of category exemplars that are temporally juxtaposed (which may enhance discrimination; Figure 1A), and the opportunity to space study of the same type of category exemplars that are temporally separated within the presentation span (which may enhance memory retention; Fig 1B). Accordingly, interleaved effects might be driven by enhanced discrimination, enhanced memory retention, or both in some measure. This paper reports new findings on the relation between interleaved effects and memory retention that derive from an experimental paradigm that manipulates both presentation order and category structure.

Enhanced Discrimination

Recent studies have largely endorsed enhanced discrimination as the critical mechanism driving interleaved effects (e.g., Birnbaum et al., 2013; Carvalho & Goldstone, 2014; Kang & Pashler, 2012; Taylor & Rohrer, 2010; Zulkipley & Burt, 2013; and for a review see Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013). This line of reasoning posits that interleaved presentations facilitate processing of differences between categorical exemplars that are temporally juxtaposed. Carlson and Yarue (1990) describe how this might work to the extent that an item presented on trial *n-1* might still be present in working memory on trial *n*. Consequently, the architecture of interleaved presentations may promote diagnostic comparisons between temporally juxtaposed items. In this way, discrimination learning may be enhanced (see *discriminative-contrast hypothesis*, Birnbaum et al., 2013). For instance, if similar landscape paintings by different artists (or categories) are interleaved (see Figure 1A), the temporal juxtaposition of two exemplars from different categories might make the differences unique to each category (e.g., tropical vs. wintry) stand out relative to what is common to both (e.g., natural landscape). One caveat bears mentioning, if enhanced discrimination (or discriminative-contrast) is to account for interleaved effects, it is reliant upon the similarity between the set of to-be-learned categories (see Kang & Pashler, 2012; Mitchell, Nash & Hall, 2008). In other words, if interleaved effects are driven by study conditions that facilitate learning to detect differences separating the categories, then the set of to-be-learned categories must possess similar attributes common to all.

Enhanced Memory Retention

The architecture of interleaved presentations additionally introduces temporal separation as a natural consequence of interpolating other category exemplars between presentations of the same type of category exemplar. Many researchers have demonstrated that when a temporal separation is placed between at least two presentations of a learning event, (i.e., spaced practice) performance on a delayed retention test is superior relative to presentations that are repeated successively in a short period of time (i.e., massed or blocked practice; for reviews see Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Dempster, 1989; Dempster & Farris, 1990; Janiszewski, Noel, & Sawyer, 2003). This is referred to as the spacing effect. Thus the inherent spaced practice afforded by interleaved presentations may confer a memory trace advantage, or enhanced memory for relevant, similar attributes shared by category members (Figure 1B).

For instance, studies by Vlach and colleagues found that spaced presentations of category exemplars promote generalization at a delayed test (e.g., Vlach, Ankowski, & Sandhofer, 2012; Vlach, Sandhofer, & Kornell, 2008). In these studies learners were presented with a series of novel objects during learning that were either spaced or massed (i.e., blocked) together. All of the category exemplars contained a relevant attribute (e.g., shape of object: *wrench shape*) elemental to a corresponding category (e.g., category label: *wug*). At test, learners were asked to generalize novel exemplars that contained these attributes. Four objects were presented simultaneously and learners had to identify the target object (e.g., Hand me the *wug*). The spaced study conditions reliably outperformed the blocked conditions.

This task paradigm parallels the type of paired-associate memory tasks found in the spacing literature. In a typical paired-associate spacing effect paradigm, pairs of words or other stimuli are associated across repeated study trials that are temporally spaced, generally by studying other paired associates during the interval(s) between study trials. In the preceding instance, learners learned to associate a relevant category attribute with a given category label (e.g., *wrench shape* -*wug*). Consequently spacing effects may have resulted in a memory trace advantage, or enhanced retention for the “paired-associates” (see Dempster, 1987; Hintzman & Rogers, 1973; and for a review see Cepeda, et al., 2006).

One class of theories posits that the benefit of spacing arises from the effects of retrieving elements from the first presentation at the time of the second presentation (e.g., Thios & D’Agostino, 1976; and for a review see Delaney, Verkoeijen, & Spirgel, 2010). Accordingly, study-phase retrieval mechanisms might account for the observed spacing effects (also see the *forgetting-as-abstraction account*, a theoretical extension of the study-phase retrieval account, to explain spacing effects in categorization; e.g., Vlach et al., 2008, and for a review see Vlach, 2014). As indicated by study-phase retrieval theories, when a temporal separation is placed between initial and subsequent presentations of a learning event (i.e., spaced or interleaved practice), forgetting, or diminished retrieval capacity, is induced. In contrast, when the presentations are presented successively (i.e., massed or blocked practice), forgetting is inhibited (see also Cuddy & Jacoby, 1982; Lee & Magill, 1983; Lee & Simon 2004; Lee & Weeks, 1987). In other words, increased levels of forgetting compel effortful retrievals at a subsequent presentation. This in turn gives rise to more potent encoding, or more durable memory structures of the learning that results (e.g., Bjork & Allen, 1970; Bjork & Bjork 1992; Lockhart, Craik, & Jacoby, 1976; Pavlik & Anderson, 2005; Vlach et al., 2008; Storm, Bjork, & Bjork, 2008; Whitten & Bjork, 1977). The outcome is a stronger memory trace, or enhanced memory retention. In contrast, when retrieval demands are minimal, such as with blocked practice, the consequence is a much less durable memory structure of the learning, or a weaker memory trace.

The notion that forgetting facilitates learning is also in keeping with research that demonstrates that increasing processing difficulty improves memory retention (e.g., Jacoby, Craik, & Begg, 1979; Slamecka & Graf 1978; and for reviews see, McDaniel & Einstein, 2005; Roediger & Karpicke, 2006) and with studies related to *desirable difficulty*, which propose that conditions that compel effortful processing during learning enhance memory retention and generalization (e.g., Bjork & Bjork, 2011; Christina & Bjork 1991; Schmidt & Bjork 1992; Whitten & Bjork, 1977). Importantly, by this line of reasoning, the advantage accorded to spaced/interleaved presentations is not reliant upon the similarity between the set of to-be-learned categories. That is the to-be-learned categories may share, but *do not* need to share, common attributes for a memory benefit to be conferred.

Past Studies Examining Spacing Effects in Interleaved Presentations

Some studies have examined latent spacing effects in interleaved presentations by holding spacing constant (e.g., Kang & Pashler, 2012; Taylor & Rohrer, 2010). In these studies learners were presented with exemplars of similar style painting categories (Kang and Pashler) or similar math problem categories (Taylor and Rohrer). The effects of spacing were examined by comparing an interleaved study condition with a study condition that inserted a temporal separation between blocked trials of category exemplars, referred to as the blocked-spaced condition. In the first instance, learners viewed unrelated cartoon drawings during the intervals between trials in the blocked-spaced condition, and in the second instance, brief puzzles were inserted between trial presentations. The lengths of the intervals between trial presentations of category exemplars in the blocked-spaced conditions were equated with corresponding interval lengths in the interleaved conditions. At test, the interleaved conditions significantly outperformed the blocked-spaced conditions, suggesting that temporal spacing alone was not sufficient to account for the benefit of interleaved effects.

However, these studies may contain a confounding by not accounting for rates of forgetting in the intervals separating trial presentations. Studies evaluating distractor task difficulty have demonstrated that rates of forgetting may be proportionally affected by the degree of task difficulty interpolated between the initial and subsequent presentations. For instance, Bjork and Allen (1970) manipulated both the difficulty of a distractor task and temporal length separating repeated study trials in which subjects were asked to remember trigrams consisting of three common four-letter nouns. Findings indicated that a difficult distractor task was more disruptive of retention (i.e., induced greater forgetting) than an easy distractor task. More important was the failure to find an interaction between distractor difficulty level and interval length. In other words, a short interval that contained a relatively difficult task produced a similar rate of forgetting when compared to a long interval that contained a relatively easy task (cf. Bui, Maddox, & Balota, 2013; Carlson & Yaure, 1990, Exp.3; Nakajima & Sato, 1989; Proctor, 1980; White, 2012).

Therefore, it is possible that the intervals between trials in the blocked-spaced conditions did not generate the same rate of forgetting to compel effortful retrievals, relative to their interleaved counterparts. For instance, Kang and Pashler inserted irrelevant cartoon drawings between painting exemplars for an interval of 10.5 seconds in the blocked-spaced condition, and learners were instructed to disregard these drawings. This is akin to an *easy* distractor task during intervals between trials. In contrast, in the interleaved condition, the interval between presentations of the same category exemplars required learners to study exemplars from other painting categories. This is akin to a *difficult* distractor task during intervals between trials. Consequently, even though interval lengths were equated in both conditions, the degree of task difficulty during the intervals between trials was not. For that reason, the potential rate of forgetting and retrieval difficulty experienced by learners in the interleaved conditions, relative to blocked-spaced conditions, may have resulted in the formation of stronger memory traces.

In sum, although past studies examining spacing effects native to interleaved presentation orders did not find a benefit of temporal spacing *per se* (but see Birnbaum et al., 2013), no strong evidence has been presented that challenges the contribution of enhanced memory retention for interleaved effects. The role of memory mechanisms underlying interleaved effects has not been explicitly investigated, nor explicitly ruled out. The benefit of enhanced memory retention as a driver of interleaved effects remains an open question.

Category Structure

A natural way of determining what things belong together is to cluster items which are similar. Indeed, similarity plays a central role in many theories of categorization. Prototype theories posit that category membership of a novel stimulus depends on its similarity to a "category standard" (Posner & Keele, 1968; Rosch, 1973; Smith & Minda, 1998). Likewise, exemplar theories posit that category membership of a novel stimulus is determined by its similarity to memory representations of previously encountered stimuli (Hintzman 1986; Medin & Schaffer, 1978; Nosofsky, 1986). By both accounts, category assignment is a function of the greatest sum of similarities.

It follows that category learning may be susceptible to the "objective structure" of to-be-learned categories (Hammer, Diesendruck, Weinshall, & Hochstein, 2009; Rosch, Mervis, Gray, Johnson & Boyes-Braem, 1976). That is the ease in which one acquires/retains category knowledge may be subject to the structural composition of a category's respective between- and within- category similarities. Goldstone (1996) proposed that as similarity *within* a category increases, category learning will be made easier, while conversely, as similarity *between* categories increases, category learning will be made more difficult (and vice versa; Table 1), albeit the respective impact of the between and within components may not necessarily be equally weighted (see Hammer et al., 2009; Rosch et al., 1976).

Table 1
Relative Impact of the Between and Within Structural Components on Category Learning.

	Learning Difficulty
Between Category Similarity	
High	Increases
Low	Decreases
Within Category Similarity	
High	Decreases
Low	Increases

Based on the possible combinations of between and within structural components (Table 1), three types of viable category structures may be denoted. First, high between- and high within category similarity (HBHW) categories have a high degree of similarity between the set of to-be-learned categories (e.g., "Elm", "Oak") and a high degree of similarity among exemplars within each category (e.g., “Live Oak”, “White Oak”). HBHW categories may increase the difficulty of learning to associate a given category with its relevant attributes because learners must filter out salient between-category similarities in order to induce relevant category attributes. As a side point, I do not consider here structures with high between- and low within-category similarity (HBLW). If categories are highly similar between categories they are by definition similar within category. Accordingly, HBLW category structures are not a viable consideration, at least for naturalistic and a majority of artifactual categories.

Second, low between- and low within-category similarity (LBLW) categories have a low degree of similarity between the set of to-be-learned categories (e.g., “plants”, “animals”) and a low degree of similarity among exemplars within each category as well (e.g., “bear”, “worm”). LBLW categories may increase the difficulty of learning to associate categories with their relevant attributes because learners must navigate somewhat striking within-category exemplar differences in order to induce relevant category attributes. Finally, low between- and high within category similarity (LBHW) categories have a low degree of similarity between the set of to-be-learned categories (e.g., “Roses”, “Horses”) and a high degree of similarity among exemplars within each category (e.g., “Thoroughbred”, “Quarter Horse”). As a result, LBHW category structures do not increase the difficulty of learning to associate categories with their relevant category attributes. LBHW categories only require that learners remember the relevant attributes associated with each category.

Category Structure and Presentation Orders

A second area that requires clarification involves the interaction between presentation orders and category structure. It has been proposed that interleaved presentation orders will benefit HBHW categories and blocked presentation orders will benefit LBLW categories (e.g., Carvalho & Goldstone, 2014; Zulkipley & Burt, 2013; see also Goldstone, 1996). The basic logic is that temporally juxtaposing different category exemplars (i.e., interleaving) may facilitate processing of between-category differences, or enhance discrimination, in HBHW categories. Put differently, interleaving should assist with learning the distinctive attributes between categories, when respective non-category members are similar. Conversely, temporally juxtaposing the same category exemplars, (i.e., blocking) may facilitate the processing of within-category similarities in LBLW categories. Accordingly, blocking should assist with learning the relevant attributes within a given category, when respective category members are dissimilar.

Although this proposal has parsimonious appeal, the interactions might not always be observed. For instance, Wahlheim et al. (2011) examined the effects of spaced study with categories of similar bird families (i.e., HBHW categories). Two spaced conditions were employed. The first presented different category exemplars separately (i.e., standard interleaved practice) while the other presented the different category exemplars in pairs (i.e., simultaneous interleaved practice). Results demonstrated that only the simultaneous interleaved practice condition yielded superior performance relative to a blocked condition for tests of novel and studied exemplars.

Why didn't the standard interleaved condition also facilitate detection of differences between category exemplars? One explanation is that presenting exemplars separately may have placed a higher demand on working memory, and as a result, diagnostic comparisons between category exemplars might have been prohibited. Whereas, presenting exemplars simultaneously would have reduced working memory demands, and diagnostic comparisons between category exemplars would have been permitted. Given that other studies have obtained interleaved effects with different HBHW materials using standard interleaved practice conditions (e.g., Birnbaum et al., 2013; Carvalho & Goldstone, 2013; Kang & Pashler, 2012; Kornell & Bjork, 2008; Zulkipley & Burt, 2013), it is possible that the materials used (i.e., increased task difficulty) may also be a factor. Importantly, while this explanation is in keeping with the discrimination account, it is also in keeping with the memory retention account. HBHW categories possess a unique structural property, inasmuch as "learning the distinctive attributes between categories" represents the inverse of "learning the unique attributes within a category." Consequently increased memory demands in a standard interleaved practice condition may amplify the difficulty of *learning to associate* category relevant attributes with a respective category label. As a result, spacing effects otherwise native to standard interleaved practice may be precluded.

With respect to LBLW categories, too few studies (e.g., Carvalho & Goldstone, 2014; Zulkiply & Burt; 2013) have looked at the interaction between LBLW categories and presentation orders to generalize these findings. In other words, different materials or procedures might produce different outcomes. For instance, it is unclear if the previously observed interactions will hold when memory loads are comparatively higher. It is possible that higher memory loads may impair the advantage of blocking LBLW categories, by amplifying the difficulty of *remembering* the relevant attributes associated with the respective categories (see Phillips, Shiffrin, & Atkinson, 1967).

Finally, the hypothesized interactions offer no basis from which to form predictions for LBHW categories. This is a critical oversight. These types of categories are commonly learned, such as pieces of art (e.g., Monets vs. Mondrians) or math problems (e.g., slope vs. inequalities). More to the point, learning to associate a given LBHW category with its relevant attributes only requires that learners remember the relevant attributes associated with each category. In other words, the low between-category similarity holds constant the benefit of learning to detect differences that separate one category from another. Accordingly, LBHW categories may be used to examine the potential contribution of memory retention. If interleaving is shown to benefit LBHW categories, this would represent direct evidence that other processes (*viz*. enhanced memory retention) may also engender interleaved effects in categorization tasks.

Overview of the present research

I compared the benefits of blocked and interleaved presentations using an inductive learning paradigm for artist identifications, similar to the types used in past work (e.g., Kang & Pashler, 2012; Kornell & Bjork, 2008). Unlike previous studies, I examined the interaction of presentation order and category structure using the full spectrum of viable structures. This afforded a more diagnostic paradigm to evaluate the underlying mechanisms driving interleaved effects. In Experiment 1, I compared the benefits of blocked and interleaved presentation orders for each of the three category structures (i.e., LBHW, HBHW and LBLW). I was interested in examining if memory retention drives interleaved effects in LBHW categories, and if the proposed interactions between presentation order and HBHW and LBLW categories would hold at higher memory loads. In Experiment 2, I extended my investigation of memory retention as a driver of interleaved effects with an experiment designed to examine if memory retention is explicitly relevant for driving interleaved effects with HBHW and LBLW categories.

Experiment 1

In this experiment I tested the prediction that enhanced memory retention alone can drive interleaved effects. I do this by comparing interleaved and blocked presentations of LBHW painting categories, which hold constant the benefit of learning to detect differences that separate one category from another. With these types of categories learners must only remember the relevant attributes associated with each artist category, similar to paired-associate memory tasks (also see Vlach et al., 2008, 2012). If it is shown that interleaved study benefits LBHW categories relative to blocked study, this would represent direct evidence that enhanced memory retention may also engender interleaved effects with categorization tasks.

I additionally test the tenability of the proposed interactions of HBHW and LBLW categories with presentation orders. Specifically, I test if these predictions hold when the memory load is comparatively high. To date, the largest number of to-be-learned HBHW categories and LBLW categories to respectively demonstrate either interleaved or blocked effects, is 12 (e.g., Zulkiply & Burt, 2013; cf. twenty categories in the present study). If increased memory demand amplifies the difficulty of learning to associate the relevant attributes with respective HBHW categories, I predict that this group should not demonstrate interleaved effects. Conversely, increased memory demands may amplify the difficulty of remembering the relevant attributes associated with respective LBLW categories in the blocked condition. Accordingly, I predict that this group will not demonstrate blocked effects.

Method

Participants and Design

One hundred and twenty participants (88 female, 32 male; age range = 18-29, mean age = 19.38) were recruited and from the University of South Florida's Department of Psychology participant pool and participated in exchange for partial course credit. This was judged to be a sufficient sample size to achieve power = .80, based on a medium effect. No problems with participation presented.

The design of the experiment was a 2 (presentation order: blocked or interleaved) x 3 (category structure: LBHW, HBHW, or LBLW) between-subjects design.

Materials

The materials consisted of 7 paintings by each of 20 different artists for three types of category structures: LBHW, HBHW, and LBLW. I selected artists who were likely to be unknown to an average college student. Nevertheless, data from any participant who scored higher than chance performance (5%) on the administered pre-test (see Procedures section) would have been excluded from the final analysis. No individual participant in any group scored higher than 5% on a pre-test. Additionally, painting selection for each group was based on adherence to the respective between and within structural components criteria for the stated category structures as described in the introduction.

LBHW Category Group. LBHW categories are distinctive between categories and similar within category. Painting selection for this group was thus based on the criteria: low similarity *between* artist categories in terms of subject matter and high similarity of exemplars *within* each category. Accordingly, the selected LBHW artist categories each depicted distinctive subject matter. This signified each category's relevant attributes (e.g., Graffiti, Native Americans, Fruit and Vegetables, Landscapes, Religious Icons, Flower Arrangements, 18th Century Ladies, Diner Tabletops, Human Figure Outlines, Couples, Rocky Seashores, Color Splashes, Tigers, Man with a Black Hat, Lines and Rectangles, Ships, Bottles and Dishes, Mountain Peaks, Street-Side Buildings, and Large Circles).

Figure 2. Examples of artist categories used in the Low Between High Within (LBHW) group. Painting exemplars from different artist categories are depicted in each respective column and painting exemplars from the same artist category are depicted in each respective row.

In this way, between-category similarity was kept low, as each artist category was readily identifiable from all other categories. Accordingly, induction of artist-attribute associates would have been relatively easy. For instance, the artist Yoshihara painted large circles while the artist Fantin painted flower arrangements. I ensured that the within category similarity was kept high by selecting sets of exemplars that were comparatively homogenous (see Figure 2). The 20 artists that formed the respective categories were: Jean-Michael Basquiat, Karl Bodmer, Fernando Botero, Henri-Edmond Cross, Giovani Cimabue, Henri Fantin-Latour, Thomas Gainsborough, Ralph Goings, Keith Haring, Robert Harris, Childe Hassam, Paul Jenkins, Antonio Ligabue, Rene Magritte, John McLaughlin, Anton Melbye, Giorgio Morandi, Nicholas Roerich, Maurice Utrillo, and JiroYoshihara

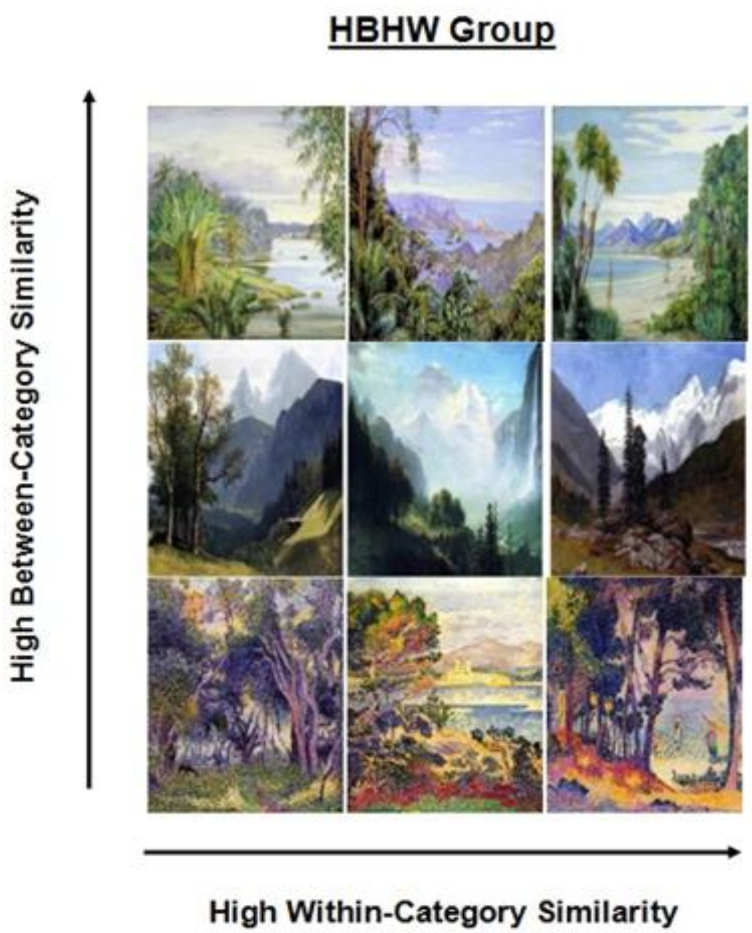


Figure 3. Examples of artist categories used in the High Between High Within (HBHW) group. Painting exemplars from different artist categories are depicted in each respective column and painting exemplars from the same artist category are depicted in each respective row.

HBHW Category Group. HBHW categories are similar between categories and similar within category. Painting selection for this group was thus based on the criteria: high similarity *between* artist categories in terms of subject matter and painting style, and high similarity of exemplars *within* each category. Within category similarity was kept high by selecting sets of exemplars that were comparatively homogenous (see Figure 3), and between-category similarity was kept high by selecting artist painting categories which depicted natural landscapes and possessed comparatively similar painting styles. It follows that induction of the artist-attributed associates in this group would have been relatively challenging. The 20 artists that formed the respective categories were: Ivan Aivazovsky, Albert Bierstad, Albert Bloch, Konstantin Bogaevesky, Georges Braque, Francis F. M. Cook, Henri Cross, Eyvind Earle, Frederick Gore, Janos Mattis-Teutsch, Alfred Munnings, Istvan Nagy, Marianne North, Vilhelms Purvitis, Pierre-Auguste Renoir, Henri Rousseau, Ivan Shishkin Frederick Short, Sidney H. Sime, Kyffin Williams.

LBLW Category Group. LBLW categories are dissimilar between categories and dissimilar within category. Painting selection for this group was thus based on the criteria: low similarity *between* artist categories in terms of subject matter, and low similarity of exemplars *within* each category. One way to delineate LBLW categories, which are by definition dissimilar by most dimensions, is to assign a unifying attribute across exemplars for a given category. For instance, the LBLW "blob" stimuli used by Carvalho and Goldstone (2013) delineated each blob category based on single attribute, a curvilinear segment, notably present in each blob category. Similarly, the LBLW categories I use here, were delineated by a single attribute, or object, notably present in each of the respective artist categories (e.g., piano, fire, moon, umbrella, glasses, apple, bed, horse, cross, snow, boat, clock, door, egg, rainbow, fish, hand, pipe, mask, ear). For instance, each painting exemplar representing the "snow" category contained snow, but this attribute was not necessarily the central focus of the painting, but rather one component within a broader composition. Accordingly exemplars within each category were perceptibly distinctive in terms of subject matter and/or composition beyond the relevant attribute that delineated each category. Importantly, no other category displayed this attribute. For instance, the "hand" category was the only category to display human hands. Other categories contained human forms in some of their exemplars (e.g., some of the piano exemplars had a person playing the piano; see Figure 4) but none of these exemplars displayed, or had "hands" visible.

Unlike the other two category structure groups, the exemplars representing a given artist category in the LBLW group, were assembled using paintings from different artists. For instance, exemplars from the LBLW category defined by the attribute "snow," were assembled using one painting by each of the following artists: Patrick Caulfield, Caspar Friedrich, William Kiddier, Vasily Polenov, Nicholas Roerich, Michael Sowa, and Peter Upton. I therefore assigned "artist" surnames to correspond to a given LBLW category during the learning phase. I generated a list of 20 surnames by randomly selecting from a list of 880 notable physicists (e.g., Abbott, Barbosa, Basov, Born, Cormack, Dirac, Fresnel, Gates, Hirn, Ising, Jacobi, Kobayashi, Landau, Mach, Millikan, Orlov, Pontecorvo, Seiberg, Umov, Wang). The 20 painting categories (e.g., apple, bed, boat etc.) were then randomly assigned one of the twenty "artist" names (e.g., Dirac, Fresnel, Jacobi, etc.).

One potential risk of using "sham artist" names is that participants might be thrown off by having to learn new artist names for artworks for which they had prior knowledge. However I expected that few if any of the participants' possessed relevant knowledge of the paintings used in the LBLW group. Indeed base rate knowledge of fine art paintings, as inferred from the pre-test scores in the LBHW and HBHW groups, supports this prospect. If however some participants in these experiments did possess any relevant knowledge of the paintings used in the LBLW group, the use of random assignment ensured that this would not be a confounding variable.

Figure 4. Examples of artist categories used in the Low Between Low Within (LBLW) group. Painting exemplars from different artist categories are depicted in each respective column and painting exemplars from the same artist category are depicted in each respective row

Procedure

I manipulated presentation order during the learning phase for each of the three category structure groups. In the blocked conditions, the painting exemplars were presented successively by artist category (e.g., $A_1A_2A_3$ A_4 -- B_1B_2 B_3 B_4 -- $C_1C_2C_3C_4$ --). In the interleaved conditions the presentation order of the paintings alternated category exemplars with the only constraint that no two exemplars from the same artist category be presented consecutively (e.g., $A_1B_1C_1$ -- $B_2C_2A_2$

-- $C_3A_3B_3$ -- $A_4C_4B_4$ --). Other than the presentation order and category structure, the conditions did not differ.

I randomly assigned participants to one of the six study conditions with 20 participants per condition. Participants were tested individually at computer work stations, within a multi-station lab with a maximum capacity of 6 persons per session. The experiment was conducted on laptop computers with 15 inch screens, set to a resolution of 1366 x 768 pixels. A computer program administered the respective learning and test phases. The experiment was created using E-Prime (Psychology Software Tools, 2012), a programming platform used for computerized behavioral experiments.

Participants in each category group saw one painting from each artist category during a pre-test that assessed their prior knowledge of the to-be-studied painting categories, four paintings from each category during the learning phase, and two more during the test phase. Data from participants with a pre-test score greater than 5% (i.e., chance) would have been excluded from the final analysis, but no individual participant in any group scored higher than 5% on the pre-test, (*Means* = 0.00, *Standard Error of the Means* = 0.00).

Each painting occupied 427 x 517 pixels, in the center of the computer screen. The paintings were set against a black background, with the respective artist's surname written below each painting during the learning phase. When necessary, I cropped or blurred the paintings to remove identifying characteristics such as names or signatures. Two artists' surnames were simplified. In the LBHW group "Fantin-Latour" was adjusted to read "Fantin," and in the HBHW group "Mattis-Teutsch was simplified to read "Mattis."

During the pre-test, participants viewed one painting individually by each of the 20 artists and asked to select, from a list of all the artists' surnames, which artist created the respective painting. They were also instructed to select the option "*I don't know,"* rather than guess the answer. The ordering of the paintings in each condition was randomized for each participant.

The learning phase followed directly. During the learning phase, participants in each condition saw 80 different paintings one a time for 4 seconds, with the artist's surname printed below each painting. The ordering of the paintings in each condition was randomized for each participant. The entire learning phase for all conditions lasted approximately 5.33 minutes. After the last painting was presented, participants were asked to complete a 5 minute distractor task, consisting of 15 rebus puzzles (20 seconds allotted per puzzle).

A self-paced generalization test was then administered. Participants saw 40 paintings (2 novel exemplars per artist category), one at a time, on the right side of the computer screen, along with an alphabetized list of the 20 artists' surnames on the left side of the computer screen. The order of the paintings was randomly determined for each participant. Participants used the computer's keyboard to select the artist who they thought painted the painting. As with the pre-test, they were instructed to select the option "*I don't know,"* rather than guess an answer. Participants had unlimited time to respond and the program provided no feedback. Upon completion, the experiment concluded. Participants were debriefed and thanked for their participation.

Statistical Analyses

Statistical analyses were conducted using IBM SPSS Statistics for Windows (IBM Corp, 2013). All data were tested for *heteroscedasticity* and normality using Levene's test and the Shapiro-Wilk's test, respectively. Bootstrapped Confidence Intervals were based on 1000 bootstrap samples. Test score reliability ranged from acceptable to excellent (see Table 2).

Table 2
Internal Consistency (Cronbach's alpha) of the Final Generalization Test used in Exp.1

Condition	**Cronbach's alpha**
Interleaved	
LBHW	.93
HBHW	.78
LBLW	.63
Blocked	
LBHW	.91
HBHW	.89
LBLW	.85

Results and Discussion

Figure 5 shows mean performance on the final generalization test as a function of study condition. Performance was analyzed using a 2 (presentation order: blocked or interleaved) x 3 (category structure: LBHW, HBHW, or LBLW) analysis of variance (ANOVA). The interaction was significant, F (2, 114) = 11.33, $p < .001$, $\eta_p^2 = .17$, indicating that the effects of presentation order differed across the different types of category structure.

Planned *t*-tests revealed significant test performance differences between the interleaved ($M = .69$, $SD = .19$) and blocked conditions ($M = .37$, $SD = .19$) for the LBHW group, t (38) = 5.28, $p < .001$. The effect sizes were large, $d = 1.68$[2], Bootstrapped 95% CI [.20, .44] denoted a fair amount of precision. These results show that interleaving was the superior condition for learning LBHW categories. With LBHW categories, processing of between-category differences

[2] The formula used to calculate Cohen's d was $(M_1 - M_2) / SD_{\text{pooled}}$ (Cohen, 1988).

or within-category similarities was not necessary. The main challenge was to remember the relevant attributes associated with each artist, such as the artist Fantin painted flower arrangements while the artist Hassam painted seashores (also see Vlach et al., 2008). I refer to these associations as *artist-attribute associates.* Accordingly, these results confirm the prediction that spaced practice afforded by the interleaved presentation conferred a significant memory advantage for the artist-attribute associates. More generally, this indicates that enhanced memory retention alone may drive interleaved effects.

Test performance differences between the interleaved (*M =.08, SD = .07*) and blocked (*M =.06, SD = .09*) conditions for the LBLW group were not significant, *t* (38) = 0.68, *p* = .50. The effect size was small, *d* = 0.25, Bootstrapped 95% CI [-.03, .06]. Interleaving was no better or worse than blocking for novel generalizations. Both conditions performed near chance levels on the final test. This means that neither presentation order was useful for learning the LBLW categories. This finding is inconsistent with previous reports which found that temporally juxtaposing the same category exemplars (i.e., blocked practice) facilitated learning LBLW categories (e.g., Carvalho & Goldstone, 2014; Zulkiply & Burt, 2013). Although there are some differences between the present and prior research (e.g., prior studies used artificial stimuli and more study trials), one explanation for the divergence is that the increased memory demand amplified the difficulty of remembering the relevant attributes associated with respective categories in the blocked condition, which may have precluded blocked effects. However, this is only speculative as the observed floor effects in this group indicate an insufficient range of measurement. It is possible that the materials employed in the present research (as opposed to the types of artificial stimuli used in past studies) may have made the task of inducing the relevant artist-attribute associates (e.g., *Abbott's paintings always had a piano*) too difficult, even for the blocked condition. If so, this would have precluded the observation of blocked, and for that matter, interleaved effects.

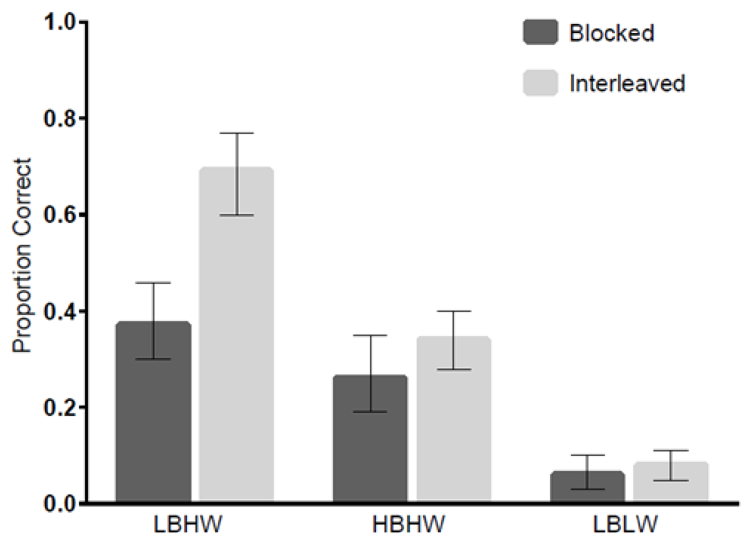


Figure 5. Proportions correct on the final test in Experiment 1. Error bars represent Bootstrapped 95% confidence intervals.

Finally as predicted, test performance differences between the interleaved (*M =.34, SD = .13*) and blocked (*M =.26, SD = .18*) conditions for the HBHW group were not found to be significant, t (38) = 1.59, p = .121. The effect size was moderate, d = 0.50[3], Bootstrapped 95% CI [-.02, .17]. This means that the temporal juxtaposition of different category exemplars (i.e., interleaved practice) did not facilitate learning significantly better than the temporal juxtaposition of the same category exemplars (i.e., blocked condition). Again, this result is inconsistent with past studies which found that temporally juxtaposing different category exemplars (i.e., interleaved practice) facilitated learning HBHW categories (e.g., Birnbaum et al., 2013; Carvalho & Goldstone, 2014; Kang & Pashler, 2010; Kornell & Bjork, 2008; Zulkiply & Burt, 2013; but see Wahlheim et al., 2011). One explanation for the divergence is that the

[3] This nominally moderate effect size is not too meaningful given the overall pattern of these results and in light of past research (e.g., Kang & Pashler , 2012, Exp. 1 & 2 and Kornell & Bjork, 2008; Exp. 1b).

increased memory load may represent a boundary condition that limits the efficacy of interleaved presentations. Put differently, there may be an upper limit to the number of categories that can be interleaved before performance suffers. Indeed a comparison of past and present research appears to suggest an inverse relationship between the efficacy of interleaved practice and the number of to-be-learned categories (see Table 3).

Table 3
Mean Test Accuracies for Novel Generalizations of High Between High Within Artist Categories in Past and Present Research[1].

	Number of Categories	**Number of Exemplars per Category**	**Interleaved Effects**	**Mean Test Accuracy[2]**	
				Interleaved	**Blocked**
Kang & Pashler, 2012 (Exp. 1)	3	24	Yes	.68	.60
Zulkipley & Burt, 2013 (Exp. 1)	12	6	Yes	.48	.32
Present Research (Exp. 1)	20	4	No	.34	.26

[1] *Each used different sets of HBHW paintings.*
[2] *Final generalization test of novel painting exemplars for standard Interleaved and Blocked conditions.*

However, this relationship may be misleading as other variables (e.g., task difficulty and/or number of study trials) may bear on final test performance. For instance, the LBHW group in the present research, admittedly an easier category structure to learn relative to HBHW category structure, demonstrated interleaved effects under the same high memory load. Accordingly an alternative explanation is that the increased memory load may have simply made an already difficult task (i.e., induction of HBHW artist-attribute associations) more difficult. Put differently, the increased memory load may have amplified the difficulty of learning to associate relevant category attributes with their corresponding artist categories in the HBHW interleaved condition. Correspondingly, any memory advantage from spaced practice, otherwise accorded by the interleaved presentation, would have been impaired.

In sum, the pattern of results obtained in Experiment 1 suggests that memory retention plays a key role in realizing interleaved effects in categorization tasks. This is not surprising as both exemplar models (e.g., Nosofsky, 1988) and prototype models (e.g., Smith & Minda, 1998) hold that generality of acquired knowledge is strongly related to memory processes. More important was the finding that enhanced memory retention alone may drive interleaved effects. This signifies that discrimination mechanisms are not the only mechanism to drive interleaved effects (also see Birnbaum et al., Exp. 3, and Vlach & Kalish, 2014, for converging evidence).

Experiment 2

The results of Experiment 1 revealed that enhanced memory retention alone may drive interleaved effects with LBHW categories. Experiment 2 examined if this finding extends to the HBHW and LBLW groups. As discussed, LBHW categories do not increase the difficulty of learning to associate categories with relevant category attributes. (i.e., induction of artist-attribute associates). In contrast, both HBHW and LBLW categories interfere with induction of artist-attribute associates by requiring learners to either respectively process challenging between-category differences or within-category similarities. This makes inducing artist-attribute associates more difficult in both of these groups. This may impair the memory advantage accorded by interleaved practice.

In Experiment 2, I diminish the respective interference in both groups with the addition of category attribute "cues," one or two word descriptions that explicitly delineate relevant category attributes, during the respective learning phases. I refer to these as *cued* groups. The cued groups parallel the functioning of the LBHW group in Experiment 1, inasmuch as the main challenge shifts away from processing between-category differences or within-category similarities, to remembering the respective category attributes associated with each artist. Blocked and interleaved presentations of HBHW-Cued and LBLW-Cued categories thus allow an objective evaluation of memory retention in these groups. I predicted that if enhanced memory retention drives interleaved effects, then diminishing the interference for inducing artist-attribute associates should produce superior performance on the final generalization test for the interleaved conditions, irrespective of category structure.

Method

Participants and Design

Eighty participants (57 female, 23 male; age range = 18-27, mean age = 19.42) were recruited from the University of South Florida's Department of Psychology participant pool and participated in exchange for partial course credit. This was judged to be a sufficient sample size to achieve power = .80, based on a medium effect. No problems with participation presented.

The design of the experiment was a 2 (presentation order: blocked or interleaved) x 2 (Category Structure: HBHW-Cued or LBLW-Cued) between-subjects design.

Materials

The materials consisted of the same set of paintings used in Experiment 1 for the HBHW and LBLW conditions. As with Experiment 1, no individual participant scored higher than 5% on the pre-test in any condition, (*Means =0.00, Standard Error of Means =0.00*)

Procedure

The experiment followed the same procedure as Experiment 1 with the following exceptions: 1) testing was limited to the HBHW and LBLW groups; and 2) during the learning phase, a one or two word cue that delineated the relevant category attributes of respective painting categories was placed below each painting, next to the artist's surname (see Figure 6).

In the LBLW groups, cue selection was based on the object attributes used to delineate each respective category (e.g., "snow", "hand"). None of the other LBLW categories displayed this attribute. Cues were thus one dimensional (i.e., presence of a unique object, such as snow) and easily verbalized.

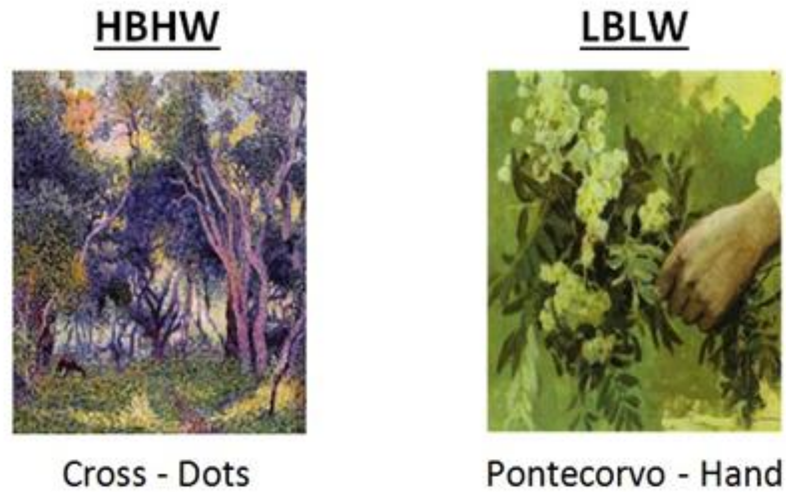


Figure 6. Illustrates artist name and respective category attribute cue for the HBHW-Cued and LBLW-Cued exemplars used in Experiment 2.

In the HBHW group, all categories depicted natural landscapes and the respective exemplars contained a number of central features that overlapped across many of the artists (e.g., pastures or woodlands, mountains, clouds, trees, foliage, bodies of water). This made selecting unique cues more challenging. In order to maintain consistency with the LBLW group, HBHW cues needed to be one dimensional (i.e., cues could not contain conjunctions such as “lagoons and palm trees”), easily verbalized, and the cues needed to be comparatively unique to a respective category. This was accomplished by creating three types of cues which could be used: (a) the presence of a unique object or feature (e.g., “sunsets”, “winter”), or (b) the presence of a unique attribute for an overlapping object or feature (e.g., “fuzzy trees”, “round bushes”), or (c) the presence of a unique attribute for a respective artist’s painting technique (e.g., “dots”, “thick paint”).

All other procedures including the test phase were exactly the same as in Experiment 1. Lastly, test score reliability was on the whole excellent (see Table 4).

Table 4

Internal Consistency (Cronbach's alpha) of the Final Generalization Test used in Exp. 2.

Condition	Cronbach's alpha
Interleaved	
HBHW Cued	.93
LBLW Cued	.90
Blocked	
HBHW Cued	.89
LBLW Cued	.91

Results and Discussion

Figure 7 shows mean performance on the final generalization test as a function of study condition. Performance was analyzed using a 2 (presentation order: blocked or interleaved) x 2 (category structure: HBHW-Cued or LBLW- Cued) analysis of variance (ANOVA). As predicted, there was a significant main effect of presentation order, F (1, 76) = 20.72, MSE = .05, p < .001, η_p^2= .21, which showed that participants were better at generalizing novel paintings in the interleaved condition (M = .49, SD = .24) compared to those in the blocked condition (M = .27, SD = .20). The effect of category structure was not significant F (1, 76) = .56, p = .46, η_p^2= .01. Performance in the HBHW-Cued group (M = .40, SD = .24) did not differ significantly from the LBLW-Cued group (M = .36, SD = .25). More important, the interaction between the two variables was not significant, F (1, 76) = .08, p = .78, η_p^2 < .01. This means that the benefit of interleaving was not dependent on category structure. Planned *t*-tests confirmed significant test performance differences between the interleaved (M =*.50*, SD = *.24*) and blocked conditions (M

=*.29, SD = .20*) for the HBHW-Cued group, *t* (38) = 3.02, *p* = .005, 95% CI [.07, .35], and significant test performance differences between the interleaved (*M =.48, SD = .24*) and blocked conditions (*M = .24, SD = .20*) for the LBLW-Cued group, *t* (38) = 3.42, *p* = .002, Bootstrapped 95% CI [.09, .38]. The respective effect sizes were large: *d* = 0.95 and *d* = 1.09.

The above analyses show that when category attribute cues were provided, interleaving was superior to blocking for novel generalizations, irrespective of category structure. Because cues diminished the respective requirements for processing between-category differences or within-category similarities, the main challenge was to remember the respective attributes associated with each artist. This means that the interleaved presentations enhanced memory retention of artist-attribute associates better than their blocked counterparts.

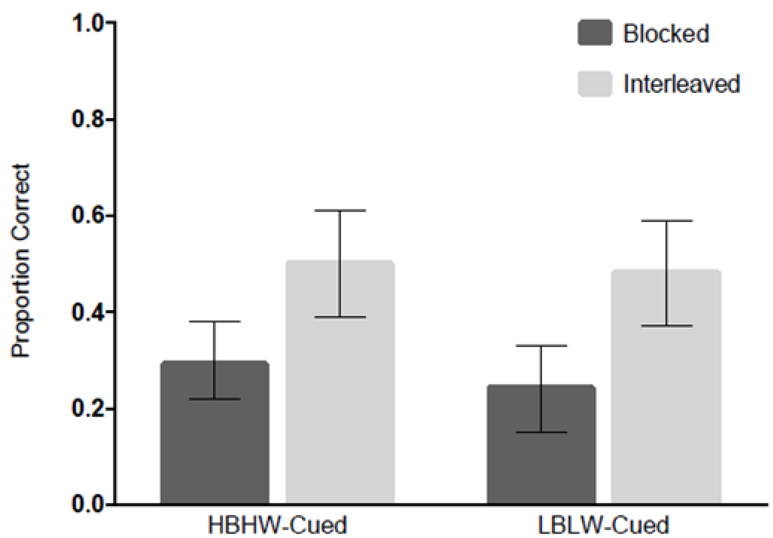


Figure 7. Proportions correct on the final test in Experiment 2. Error bars represent 95% confidence intervals.

One could argue that the cues improved discrimination processing in the HBHW-Cued interleaved condition, by highlighting between-category featural differences. In that case, discrimination mechanisms might have engendered the observed interleaved effects with this group. However, it is unclear how discrimination processing would have engendered interleaved effects in the LBLW-Cued group. According to the current state of theory, the attribute cues should have improved the processing of within-category similarities (or relevant category attributes) in the LBLW-Cued blocked condition, thereby engendering blocked effects. Put differently, the blocked LBLW-Cued condition should have shown superior performance relative to the interleaved LBLW-Cued condition. Yet this was not the case, the LBLW-Cued interleaved condition significantly outperformed the LBLW-Cued blocked condition. Taken together, discrimination mechanisms cannot account for the pattern of results across both groups, whereas memory retention mechanisms can. While these results do not necessarily provide direct evidence against the discrimination account, they do suggest that memory mechanisms are critical for engendering interleaved effects.

General Discussion

Recent studies have largely favored discrimination mechanisms as the critical driver of interleaved effects however findings from the present research suggest an alternative hypothesis. Across two experiments, memory mechanisms were shown to be the critical driver. In the framework I have adopted, the inherent spaced practice afforded by interleaved presentations confers a memory trace advantage, or enhanced memory retention for relevant attributes shared by category members. I likened the underlying machinery to paired-associate memory tasks typically found in the spacing literature, which as a large body of research prescribes, would be expected to show robust spacing effects (e.g., Cepeda et al., 2006). Moreover, the notion that memory plays a key role in realizing interleaved effects with categorization tasks is consistent with both exemplar (category membership of a novel exemplar is determined by its similarity to memory representations of previously encountered category exemplars; e.g., Nosofsky, 1988) and prototype models (category membership of a novel exemplar depends on its similarity to a "category standard" derived from previously encountered category exemplars; e.g., Smith & Minda, 1998), which hold that generality of acquired knowledge is strongly related to memory processes.

In Experiment 1, results from the LBHW group provide direct evidence that memory processes engender interleaved effects in categorization tasks. With these types of categories, the benefit of learning to detect differences that separate one category from another is held constant; learners only need to remember the relevant attributes associated with each category. The observed interleaved effects with the LBHW artist categories, thus suggest that spaced practice afforded by the interleaved presentation, conferred a significant memory advantage for the artist-attribute associates. To be precise, induction of the artist-attribute associates would have ensued early in the learning phase (e.g., *Fantin - Flowers*) and subsequent presentations served as spaced or massed practice, depending on the respective condition assignment (i.e., interleaved or blocked). Spacing effects engendered a memory trace advantage, or enhanced memory retention for the artist-attribute associates in the interleaved presentation.

How does this framework accord when other types of category structures are employed? LBLW and HBHW categories interfere with induction of the artist-attribute associates. This may impair memory mechanisms. Past studies that examined spacing of non-exact repetitions found that when participants failed to recognize that a subsequent presentation was a repetition, spacing effects were precluded (e.g., Appleton-Knapp, Bjork & Wickens, 2005; Dellarosa & Bourne, 1985; Glover & Corkill, 1987). It follows that if participants failed to induce artist-attribute associates in these groups, spacing effects native to the interleaved condition would be precluded.

While this of line reasoning may be evident for the LBLW group, there were compelling empirical arguments to expect interleaving to facilitate novel generalizations of HBHW categories, based on discrimination mechanisms alone (e.g., Birnbaum et al., 2013; Kang & Pashler, 2012). Yet in the present research interleaving was not significantly better than blocking for novel generalizations in the HBHW group (Exp. 1). While this finding seems inconsistent with prior studies, the framework I have adopted may account for the divergence. In the present study the number of to-be-learned categories was substantially higher than in prior studies (i.e., 20 categories vs. 3 to 12 categories). This higher level of memory load may have amplified the interference for inducing artist-attribute associates. If so, spacing effects native to the interleaved condition would be precluded.

More specifically, although the memory mechanisms are powered by spacing effects native to interleaved presentations, if learners fail to induce the relevant artist-attribute associates, spaced study will not commence. Accordingly, multiple factors (e.g., the number of categories, number of study trials, task difficulty and learner ability) may serve to impair or facilitate memory mechanisms. In Experiment 1, the task difficulty of learning HBHW categories coupled with a large number of to-be-learned categories likely interfered with inductions of the respective artist-attribute associates. This interference impaired memory retention. Conversely, it may also be possible to facilitate memory retention when task difficulty is high by reducing levels of interference, such as employing fewer to-be-learned categories or increasing the number of study trials. Prior studies which found a benefit for interleaving HBHW categories generally employed fewer categories and increased study trials. For instance, Kang and Pashler (2012) demonstrated interleaved effects for highly similar artist categories when the number of to-be-learned categories was quite small (3 categories; see Table 2), and the number of study trials was quite large (24 trials per artist).

In keeping with this logic, Experiment 2 tested the prediction that diminished interference would facilitate memory retention in both the HBHW and LBLW groups, neither of which realized interleaved effects in Experiment 1. In this case, interference was diminished by reducing task difficulty during the learning phase with cues that delineated category relevant attributes. Results demonstrated that when interference was diminished via attribute cueing, interleaving was uniformly superior to blocking for the generalization of novel paintings in both groups. Critically, the type of category structure (HBHW or LBLW) did not bear on this outcome. Memory retention was mutually restored. From a theoretical standpoint the research reported here therefore suggests that memory mechanisms are critical for interleaved effects, and more importantly, memory retention alone may drive interleaved effects.

Practical Implications and Future Directions

As discussed previously, I likened the underlying machinery of interleaving to a paired-associate learning task typically found in the spacing literature inasmuch as participants learned to associate relevant painting attributes with an artist's name (e.g., *Fantin* - *flowers*) and spacing effects associated with the interleaved presentations engendered a memory advantage for these associations. These findings may have implications for learning other types of ecologically relevant categories. For instance, when a student encounters mathematics problems (e.g., 10 + 3 vs. 10 x 3) the operators "+" and "x" are symbols used to denote the respective mathematical operations addition and multiplication. In order for the student to correctly execute the respective solutions, s/he must remember that the attribute "+" is associated with the operation addition and the attribute "x" is associated with multiplication. Correspondingly, if word phrases are used to denote mathematical operations (e.g., If baker A sold 10 cupcakes and baker B sold 3 times as many cupcakes as baker A, how many cupcakes did baker B sell?) the student would need to remember that the attribute, or word, "times" is associated with multiplication. In each case, efficient recall of the respective attribute-operations association is critical.

What if the attribute(s)-category associations have more room to vary (e.g., selection of an appropriate statistical test)? In this case the appropriate selection varies based on a set of inputs, or set of *attributes*, which help to determine the appropriate statistical test to select. For instance if the set of attributes includes (determine significance of mean group difference, has 1 continuous DV, has 1 dichotomous IV, and has 0 covariates) then the statistical test generally associated with this set of attributes is *t*-test. In contrast, if we assume the same set of attributes but change the number and type of IV attributes to “2 categorical IVs” then the statistical test generally associated with this set of attributes is factorial ANOVA. As a result, even though the attribute-category associations are more complex insofar as the associations encompass a *set of attributes,* efficient recall of the respective associations is nevertheless critical for the selection of an appropriate statistical test.

Accordingly, the foremost practical implication from these experiments is that interleaving as a learning strategy may be used to facilitate memory retention. There is no reason to believe that this implication will not extend beyond the type of visual categorization task I used, to broader types of learning applications. This of course has wide-ranging implications for pedagogy, as the value of education depends in large part not only on what information is learned, but on whether that information once learned will be retained. Since the late 1800s research on memory and learning has demonstrated time and again that spaced study enhances memory retention. However, no system has been developed, which practically and economically incorporates spacing into classrooms en masse, presumably because the logistical costs for spaced study are high. Interleaving may prove to be a viable and economic alternative. Continued investigations of latent spacing effects underlying interleaved presentations represent an important avenue for future research. Equally, replications with diverse subject matter domains will be necessary in order to generalize the present findings to broader learning contexts.

Secondly, findings from Experiment 2 have specific implications for pedagogy as it pertains to the efficiency of learning. Son, Smith and Goldstone (2008) demonstrated that appropriate generalization could be achieved with only one learning instance by directly teaching the relevant abstraction. Experiment 2's paradigm similarly provided explicit instruction of relevant category attributes via cueing, which facilitated interleaved effects. Taken together, these findings potentially form new insight for optimizing the adaptive application of knowledge. Namely, the development of training modules which both directly teach the relevant information, and promote memory retention via iterative presentation orders such as interleaving. Future investigations will need to bear this out. Researchers in future studies should also continue to examine the role of memory retention, particularly as it relates to proposals for training flexibility in thought, as the sum and substance of the findings reported here suggest that memory mechanisms may be critical for these processes.

That said, some limitations of the present study should be acknowledged. First my sample was drawn exclusively from a population of undergraduate students from a state university. It is possible that findings from this restricted sample may not generalize to other populations. However past studies have shown interleaving to be effective in older populations (e.g., visual categorization tasks; Kornell, Castel, Eich, & Bjork, 2010; Wahlheim, Dunlosky, & Jacoby, 2011) and in younger populations (e.g., math problems; Taylor & Rohrer, 2010). A second limitation is that I did not include a follow up test, so it is not possible to delineate to what extent the observed advantages were maintained over time. Nevertheless, some early evidence (e.g., visual categorization and textual tasks; Zulkiply & Burt, 2013) indicates that single session interleaving has the potential to improve long-term retention.

Concluding Comment

To my knowledge, this study is the first to demonstrate the extent to which memory retention drives interleaved effects. While previous research has found resultant retention benefits for interleaved presentation orders, this study advances this field by disaggregating the benefits of memory mechanisms from discrimination mechanisms. Additionally the present research is novel in that it demonstrates that interleaved presentation orders create stronger underlying memory structures from which to make future generalizations. This research adds to the understanding of how presentation orders may be used to enhance learning.

References

Appleton-Knapp, S. L., Bjork, R. A., & Wickens, T. D. (2005). Examining the spacing effect in advertising: Encoding variability, retrieval processes, and their interaction. *Journal of Consumer Research, 32*, 266-276.

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*, 392-402.

Bjork R.A. & Allen, T.W. (1970). The spacing effect: Consolidation or differential encoding? *Journal of Verbal Learning and Verbal Behavior, 9,* 567-572.

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56-64). New York: Worth Publishers.

Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. Healy, S. Kosslyn, & R. Shiffrin (Eds.), *From learning processes to cognitive processes: Essays in honor of William K. Estes* (Vol. 2, pp. 35-67). Hillsdale, NJ: Erlbaum.

Bui, D. C., Maddox, G. B., & Balota, D. A. (2013). The roles of working memory and intervening task difficulty in determining the benefits of repetition. *Psychonomic Bulletin & Review, 20,* 341-347.

Carlson, R. A., & Yaure, R. G. (1990). Practice schedules and the use of component skills in problem solving. *Journal of Experimental Psychology: Learning, Memory and Cognition, 16*, 484-496.

Carlson, R. A., & Yaure, R. G. (1990). Practice schedules and the use of component skills in problem solving. *Journal of Experimental Psychology: Learning, Memory and Cognition, 16, 484-496.*

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

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall: A review and quantitative synthesis. *Psychological Bulletin, 132*, 354-380.

Christina, R. W., & Bjork, R. A. (1991). Optimizing long-term retention and transfer. In D. Druckman & R. A. Bjork (Eds.), *In the mind's eye: Enhancing human performance* (pp. 23-56). Washington, DC: National Academy Press.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Cuddy, L. J., & Jacoby, L.L. (1982). When forgetting helps memory: An analysis of repetition effects. *Journal of Verbal Learning and Verbal Behavior*, *21*, 451-467.

Delaney, P. F., Verkoeijen, P. P., & Spirgel, A. (2010). Spacing and testing effects: A deeply critical, lengthy, and at times discursive review of the literature. *Psychology of learning and motivation*, *53*, 63-147.

Dellarosa, D., & Bourne, L. E. (1985). Surface form and the spacing effect. *Memory & Cognition*, *13*, 529-537.

Dempster, F. N. (1987). Effects of variable encoding and spaced presentations on vocabulary learning. *Journal of Educational Psychology, 79*, 162.

Dempster, F. N., & Farris, R. (1990). The spacing effect: Research and practice. *Journal of Research and Development in Education, 23,* 97-101.

Dempster, F. N. (1989). Spacing effects and their implications for theory and practice. *Educational Psychology Review, 1,* 309-330.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, *14*, 4-58.

Glover, J. A., & Corkill, A. J. (1987). Influence of paraphrased repetitions on the spacing effect. *Journal of Educational Psychology*, *79*, 198.

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

Hammer, R., Diesentruck, G., Weinshall, D., & Hochstein, S. (2009). The development of category learning strategies: What makes the difference? *Cognition, 112*, 105-199.

Hintzman, D. L. (1986). "Schema Abstraction" in a multiple-trace memory model. *Psychological Review*, 93, 411-428.

Hintzman, D. L., & Rogers, M. K. (1973). Spacing effects in picture memory. *Memory & Cognition*, *1*(4), 430-434.

IBM Corp. (2013). IBM SPSS Statistics for Windows, Version 22.0. [Computer software]. Armonk, NY: IBM Corp.

Jacoby, L. L., Craik, F. I. M., & Begg, I. (1979). Effects of decision difficulty on recognition and recall. *Journal of Verbal Learning and Verbal Behavior, 1979, 18,* 585-600.

Janiszewski, C., Noel, H., & Sawyer, A.G. (2003). A meta-analysis of the spacing effect in verbal learning: Implications for research on advertising retention and consumer memory. *Journal of Consumer Research, 30,* 138-149.

Kang, S. H., & Pashler, H. (2012). Learning painting styles: Spacing is advantageous when it promotes discriminative contrast. *Applied Cognitive Psychology, 26*, 97-103.

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

Kornell, N., Castel, A. D., Eich, T. S., & Bjork, R. A. (2010). Spacing as the friend of both memory and induction in young and older adults. *Psychology and Aging*, 25, 498-503.

Lee, T. D., & Weeks, D. J. (1987). The beneficial influence of forgetting on shorterm retention of movement information. *Human Movement Science*, *6*, 233-245.

Lee, T. D., & Magill, R. A. (1983). The locus of contextual interference in motor-skill acquisition*, Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 730-746.

Lee, T. D., & Simon, D. A. (2004). Contextual interference. In A.M. Williams & N.J. Hodges (Eds.)*, Skill acquisition in sport: Research, theory and practice.* London: Routledge.

Lockhart, R. S., Craik, F. I. M., & Jacoby, L. (1976). Depth of processing, recognition, and recall. In J. Brown (Ed.), *Recall and recognition*. New York: Wiley.

MacKendrick, A. F., & Blessing, S. B. (under review). Interleaved effects in inductive category learning: The role of memory retention. *Applied Cognitive Psychology*.

McDaniel, M. A., & Einstein, G. O. (2005). Material appropriate difficulty: A framework for determining when difficulty is desirable for improving learning. In A.F. Healy (Ed.), *Experimental cognitive psychology and its applications* (pp. 73-85). Washington DC: American Psychological Association.

Medin, D. L., & Schaffer, M. M. (1978). A context theory of classification learning. *Psychological Review*, 85, 207-238.

Mitchell, C., Nash, S., & Hall, G. (2008). The intermixed-blocked effect in human perceptual learning is not the consequence of trial spacing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 237-242.

Nakajima, Y., & Sato, K. (1989). Distractor difficulty and the long-term recency effect. *The American Journal of Psychology*, *102*, 511-521.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115,* 39-57.

Pashler, H., Bain, P. M., Bottge, B. A., Graesser, A., Koedinger, K., McDaniel, M., & Metcalfe, J. (2007). Organizing Instruction and Study to Improve Student Learning. IES Practice Guide. NCER 2007-2004. *National Center for Education Research*.

Pavlik, P. I., & Anderson, J. R. (2005). Practice and Forgetting Effects on Vocabulary Memory: An Activation-Based Model of the Spacing Effect. *Cognitive Science*, *29*, 559-586.

Phillips, J. L., Shiffrin, R. M., & Atkinson, R. C. (1967). Effects of list length on short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *6*, 303-311.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of experimental psychology*, *77*, 353-363.

Proctor, E. K. (2003). Developing knowledge for practice: Working through" trench-bench" partnerships. *Social Work Research*, *27,* 67.

Proctor, R.W. (1980). The influence of intervening tasks on the spacing effect for frequency judgments. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 254-266.

Psychology Software Tools. (2012). E-Prime (Version 2.0.10.242) [Computer software]. Available from http://www.pstnet.com.

Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1,* 181-200.

Rosch, E. H. (1973). On the internal structure of perceptual and semantic categories. In T.E. Moore (Ed.). *Cognitive development and the acquisition of language* (pp. 111-144). New York: Academic Press.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive psychology*, *8*, 382-439.

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

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning & Memory, 4*, 592-604.

Son, J. Y., Smith, L. B., & Goldstone, R. L. (2008). Simplicity and generalization: Short-cutting abstraction in children’s object categorizations. *Cognition*, *108*, 626-638.

Smith, J. D., & Minda, J. P. (1998). Prototypes in the Mist: The Early Epochs of Category Learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24,* 1411-1436.

Sokal, R. R., & Rohlf, F. J. (1995). Biometry: The principals and practice of statistics in biological research. *WH Freeman and Company, New York*.

Storm, B. C., Bjork, E. L., & Bjork, R. A. (2008). Accelerated relearning after retrieval-induced forgetting: The benefit of being forgotten. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 230-236.

Taylor, K., & Rohrer, D. (2010). The effects of interleaved practice. *Applied Cognitive Psychology, 24*, 837-848.

Thase, M. E. (2006). A tale of two paradigms. *Clinical Psychology: Science and Practice*, *13*, 94-98.

Thios, S. J., & D'Agostino, P. R. (1976). Effects of repetition as a function of study-phase retrieval. *Journal of Verbal Learning and Verbal Behavior*, *15*, 529-536.

Vlach, H. A., Ankowski, A. A., & Sandhofer, C. M. (2012). At the same time or apart in time? The role of presentation timing and retrieval dynamics in generalization. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *38*, 236-254.

Vlach, H. A., & Kalish, C. W. (2014). Temporal dynamics of categorization: Forgetting as the basis of abstraction and generalization. *Frontiers in Psychology*, *5*. doi: 10.3389/fpsyg.2014.01021.

Vlach, H. A., Sandhofer, C. M., & Kornell, N. (2008). The spacing effect in children's memory and category induction. *Cognition, 109*, 163-167.

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*, 750-763.

White, K. G. (2012). Dissociation of short-term forgetting from the passage of time. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 255

Whitten, W. B., & Bjork, R. A. (1977). Learning from tests: Effects of spacing. *Journal of Verbal Learning and Verbal Behavior*, *16*, 465-478.

Whitehurst, G. J. (2003). *The Institute of Education Sciences: New Wine, New Bottles*. Speech given at the annual meeting of the American Educational Research Association.

Zulkiply, N., & Burt, J. S. (2013). The exemplar interleaving effect in inductive learning: Moderation by the difficulty of category discriminations. *Memory & Cognition*, *41*, 16-27.

Zulkiply, N., & Burt, J. S. (2013). Inductive learning: Does interleaving exemplars affect long-term retention?. *Malaysian Journal of Learning and Instruction*, *10*, 133-155.

Appendix A:
Final Test Score Frequency Graphs for Experiments 1 & 2

Table 5
Final Test Score Frequency Graph Experiment 1.

	LBHW	**HBHW**	**LBLW**
Blocked			
0-25%	6 (1)	12 (0)	19 (6)
26-50	9	6	1
51-75	5	1	0
76-100	0	1	0
Interleaved			
0-25%	2 (0)	8 (0)	20 (4)
26-50	1	11	0
51-75	9	1	0
76-100	8	0	0

Note: the number of final tests with a score of zero is in parentheses.

Table 6
Final Test Score Frequency Graph Experiment 2.

	HBHW-Cued	**LBLW-Cued**
Blocked		
0-25%	11 (1)	14 (1)
26-50	6	3
51-75	2	2
76-100	1	1
Interleaved		
0-25%	3 (0)	3 (0)
26-50	7	10
51-75	6	3
76-100	4	4

Note: the number of final tests with a score of zero is in parentheses.

Appendix B:
IRB Approval Letter

USF UNIVERSITY OF SOUTH FLORIDA

RESEARCH INTEGRITY AND COMPLIANCE
Institutional Review Boards, FWA No. 00001669
12901 Bruce B. Downs Blvd., MDC035 • Tampa, FL 33612-4799
(813) 974-5638 • FAX (813) 974-7091

6/24/2013

Alex MacKendrick
Psychology
4202 E. Fowler Ave., PCD 4118G
Tampa, FL 33620

RE:	**Exempt Certification**
IRB#:	Pro00012323
Title:	Enhanced Learning (1)

Study Approval Period: 6/22/2013 to 6/22/2018

Approved Items:
Protocol Document:
12323_Protocol_V1_6.13.13.pdf

Consent Script:
12323_ResearchInfo_V1_6.19.13.docx

Dear Ms. MacKendrick:

On 6/22/2013, the Institutional Review Board (IRB) determined that your research meets USF requirements and Federal Exemption criteria as outlined in the federal regulations at 45CFR46.101(b):

(2) Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures or observation of public behavior, unless:
(i) information obtained is recorded in such a manner that human subjects can be identified, directly or through identifiers linked to the subjects; and (ii) any disclosure of the human subjects' responses outside the research could reasonably place the subjects at risk of criminal or civil liability or be damaging to the subjects' financial standing, employability, or reputation.

As the principal investigator for this study, it is your responsibility to ensure that this research is conducted as outlined in your application and consistent with the ethical principles outlined in the Belmont Report and with USF IRB policies and procedures. Please note that changes to this protocol may disqualify it from exempt status. Please note that you are responsible for notifying the IRB prior to implementing any changes to the currently approved protocol.

Your study qualifies for a waiver of the requirements for the documentation of informed consent as outlined in the federal regulations at 45CFR46.117(c) which states that an IRB may waive the requirement for the investigator to obtain a signed consent form for some or all subjects.

The Institutional Review Board will maintain your exemption application for a period of five years from the date of this letter or for three years after a Final Progress Report is received, whichever is longer. If you wish to continue this protocol beyond five years, you will need to submit a new application at least 60 days prior to the end of your exemption approval period. Should you complete this study prior to the end of the five-year period, you must submit a request to close the study.

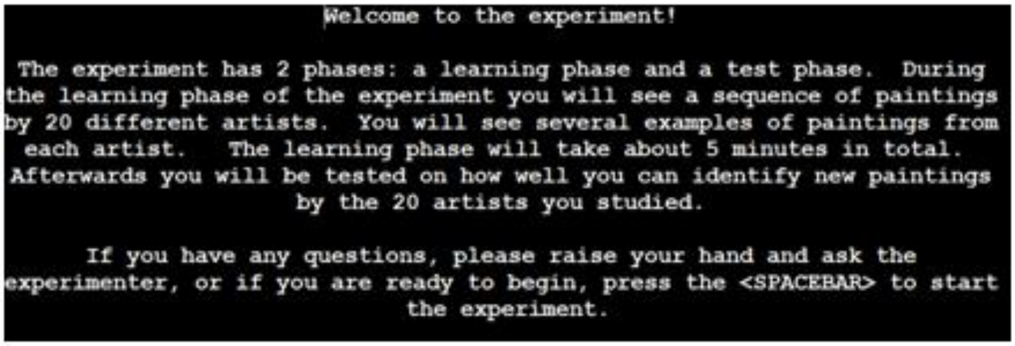
We appreciate your dedication to the ethical conduct of human subject research at the University of South Florida and your continued commitment to human research protections. If you have any questions regarding this matter, please call 813-974-5638.

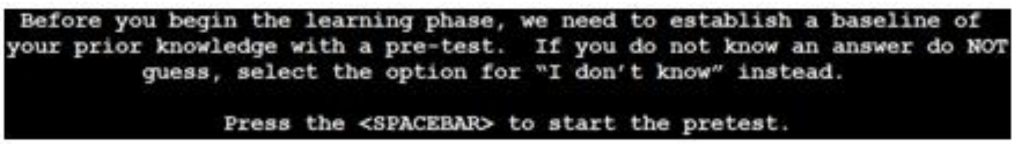
Sincerely,

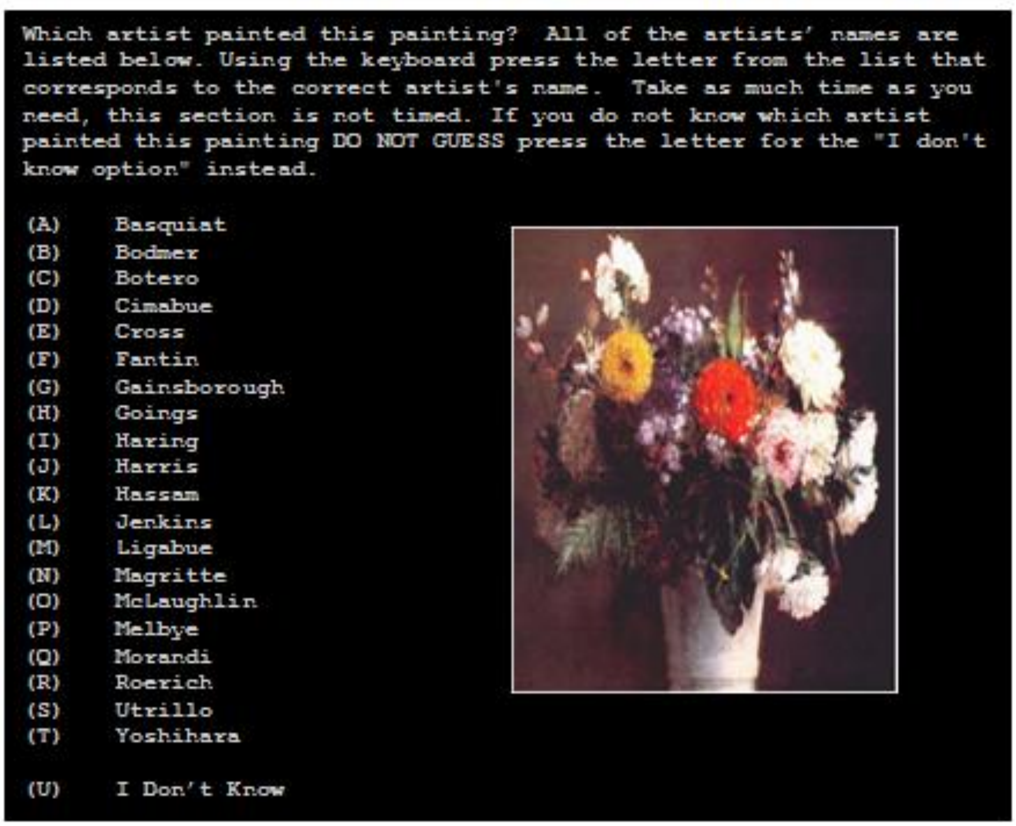
John A Schinka, Ph.D.

John Schinka, Ph.D., Chairperson
USF Institutional Review Board

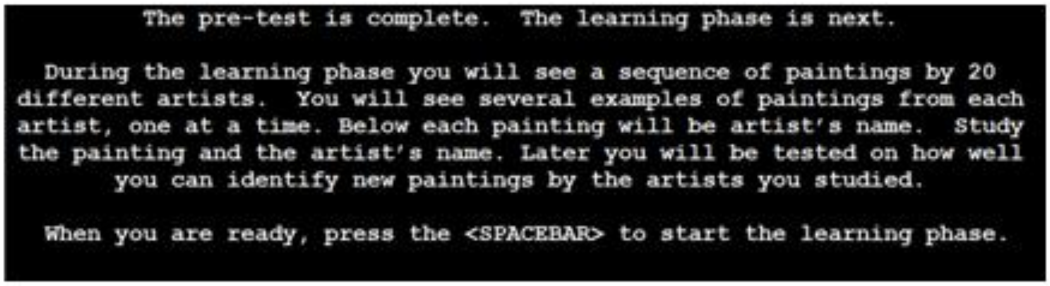
Appendix C:
Experiment Screen Shots

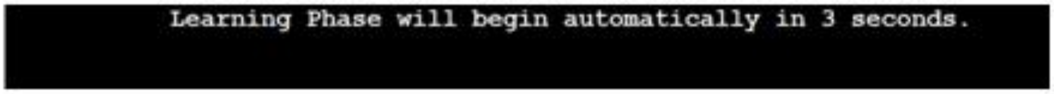






Pre-Test: Self-Paced





Learning Phase: 4 sec per exemplar / Total time ≈ 5.33 mins

Learning Phase has ended

Before the test phase begins, you will be asked to complete a set of puzzles.

Rebus puzzles consist of pictures representing syllables, words, and common phrases.

For instance: OHOLENE
represents the phrase "Hole in One" o-HOLE-ne.

Try your best to provide a solution for each rebus puzzle. This is a timed test.

When you are ready to begin press the <SPACEBAR> key.

Rebus Puzzle Instructions

Using the keyboard, you will type your solution to the rebus puzzle. (When you have a solution just start typing, your answer will appear on the top portion of the screen. If you finish before the clock runs out, wait for the puzzle to change).
You will have 20 seconds to complete each puzzle. If you do not know the answer, type the letter "U".

You will be alerted when there are 5 seconds remaining on the clock (if you have already typed your solution, ignore the warning and wait until the next puzzle appears).

These instructions will be provided again with each puzzle.

Press < SPACE BAR> when you are ready to begin

using the keyboard, type your solution to the rebus puzzle. (type your answer and wait for the puzzle to change). you have 20 seconds to complete the puzzle. if you do not know the answer, type the letter "u". you will be alerted when there are 5 seconds remaining on the clock (if you have already typed your solution ignore the warning).

Distractor Task: total time 5 mins

The TEST phase is next. Press <SPACEBAR> to begin.

Which artist painted this painting? All of the artists' names are listed below. Using the keyboard press the letter from the list that corresponds to the correct artist's name. Take as much time as you need, this section of the test is not timed. If you do not know which artist painted this painting DO NOT GUESS press the letter for the "I don't know option" instead.

(A) Basquiat
(B) Bodmer
(C) Botero
(D) Cimabue
(E) Cross
(F) Fantin
(G) Gainsborough
(H) Goings
(I) Haring
(J) Harris
(K) Hassam
(L) Jenkins
(M) Ligabue
(N) Magritte
(O) McLaughlin
(P) Melbye
(Q) Morandi
(R) Roerich
(S) Utrillo
(T) Yoshihara

(U) I Don't Know

Final Test: Self-Paced

The experiment has concluded. Thank you for participating.

Appendix D:
Descriptive Statistics

Table 7

Skewness and Kurtosis Statistics for Experiments 1 and 2.

	Skewness	Kurtosis
Exp. 1		
LBHW (I)	-1.5	1.6
LBHW (B)	-.28	-.60
HBHW (I)	1.4	1.7
HBHW (B)	.73	.62
LBLW (I)	2.1	4.4
LBLW (B)	.57	-.37
Exp. 2		
HBHW – Cued (I)	-.51	-.64
HBHW – Cued (B)	.71	.65
LBLW – Cued (I)	.59	.09
LBLW – Cued (B)	1.0	-.11