Title
Understanding and Optimizing the Inductive Learning of Categories and Concepts

Permalink
https://escholarship.org/uc/item/8396m4q0

Author
Birnbaum, Monica Sachiko

Publication Date
2013

Peer reviewed|Thesis/dissertation
Understanding and Optimizing
the Inductive Learning of Categories and Concepts

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy
in Psychology

by

Monica Sachiko Birnbaum

2013
ABSTRACT OF THE DISSERTATION

Understanding and Optimizing the Inductive Learning of Categories and Concepts

by

Monica Sachiko Birnbaum

Doctor of Philosophy in Psychology

University of California, Los Angeles, 2013

Professor Elizabeth Ligon Bjork, Co-Chair

Professor Robert A. Bjork, Co-Chair

Inductive learning refers to learning categories and concepts from exemplars of those categories or concepts. Such learning, which is a fundamental component of human cognition, enables us to classify never-before-encountered examples as instances of a given category or concept. How, though, should the exemplars of different categories, such as pictures of different species of butterflies, be presented in order to enhance inductive learning? The prevailing view has been that exemplars of a given category should be presented close together in time to highlight the commonalities that define that category, but Kornell and Bjork (2008)—to their surprise and against participants’ intuitions—found that interleaving, not blocking, the exemplars of separate to-be-learned categories enhanced inductive learning. These findings, together with results from subsequent studies, suggest that the opportunity interleaving provides to contrast exemplars of different categories, such as an Admiral butterfly versus an Elfin butterfly, may be key to optimizing inductive learning.

The eight experiments reported in this dissertation, which involved having participants learn species of butterflies, families of birds, the handwriting styles of different individuals, and
the styles of different artists, were designed to clarify the roles of contrast processing (encoding differences between exemplars of different categories) and commonality processes (encoding commonalities across the exemplars within a category).

Overall, the results obtained suggest that contrast processing and commonality processing are both critical, but that contrast processes occur automatically, whereas noticing commonalities is a deliberate and conscious process in category learning. Importantly, interleaving exemplars of different categories appears not only to facilitate contrast processing, but also, under some circumstances, commonality processing. Finally, it appears that learners—even when they know about the benefits of interleaving and the importance of contrast processing for achieving category learning—cannot effectively engage in contrast processing as a self-initiated study strategy without the aid of an interleaved presentation schedule.

The present findings help to provide a more complete picture of the processes involved in feature extraction and category generalization. From a practical standpoint, the results also have implications for enhancing category learning in educational settings, where such learning is prevalent and critical to mastery and achievement.
The dissertation of Monica Sachiko Birnbaum is approved.

Douglas Bell

Barbara Knowlton

Elizabeth Ligon Bjork, Committee Co-Chair

Robert A. Bjork, Committee Co-Chair

University of California, Los Angeles

2013
TABLE OF CONTENTS

List of Figures ............................................................................................................................... vii
Acknowledgments ........................................................................................................................... x
Vita ................................................................................................................................................ xii
CHAPTER 1 – General Introduction ............................................................................................. 1
  Interleaved Study in Inductive Learning ..................................................................................... 2
  Discrimination Hypothesis: Noting Features that Distinguish between Categories ............... 3
  Outline of the Present Dissertation ........................................................................................ 12
CHAPTER 2 – The Role of Spaced Retrieval in Inductive Learning ................................................. 15
  Experiment 1a: Effects of temporal contrast and spaced retrieval on inductive learning .... 18
  Experiment 1b: A closer look at the role of temporal contrast and spaced retrieval in inductive learning .................................................................................................................. 23
CHAPTER 3 – Effects of Interleaved Study on Recall Learning and Inductive Learning ............... 29
  Experiment 2: Does recall learning drive the interleaving effect?: Using category label pretraining ........................................................... 30
CHAPTER 4 – Effective Study Schedules for Incidental and Non-directed Category Learning ........ 37
  Experiment 3: The effect of interleaving on incidental learning ............................................. 39
  Experiment 4: Manipulating directed and non-directed learning within subjects ................. 45
CHAPTER 5 – Transfer in Inductive Learning ............................................................................. 56
  Transfer of the Benefits of Interleaving to New Inductive Learning ..................................... 57
  Experiment 5: Applying the “interleaving strategy” to new learning .................................. 58
  The Testing Effect, the Generation Effect, and Feedback Training ....................................... 65
  Transfer of the Benefits of the Testing Experience to New Inductive Learning ................. 68
Experiment 6a: The effect of feedback training on new inductive learning ........................................... 69

Experiment 6b: Transfer of the “testing strategy” to new inductive learning ................................. 76

CHAPTER 6 – General Discussion ........................................................................................................ 85

Summary of methods and findings ...................................................................................................... 85

Possible explanation for the present findings ..................................................................................... 88

Concluding remarks .............................................................................................................................. 90

Appendix .............................................................................................................................................. 92

References .......................................................................................................................................... 94
LIST OF FIGURES

Figure 1.1. Study conditions for Birnbaum, Kornell, Bjork, and Bjork (2013) Experiment 2. TQ stands for a 10-s trivia question used to space apart the exemplar presentations ............... 7

Figure 1.2. Study conditions for Birnbaum, Kornell, Bjork, and Bjork (2013) Experiment 1. TQ stands for a 10-s trivia question used to space apart the exemplar presentations ............... 9

Figure 2.1. Study conditions for Experiment 1a ................................................................. 20

Figure 2.2. Classification test performance for Experiment 1a ........................................... 21

Figure 2.3. Study conditions for Experiment 1b. Letter denotes butterfly species, subscript value represents exemplar number .................................................................................... 26

Figure 2.4. Classification test performance for Experiment 1b ........................................... 27

Figure 3.1. Classification test performance for Experiment 2. Left panel: classification test performance for control condition. Right panel: classification test performance for pretraining condition ................................................................................................. 33

Figure 3.2. Judgments of learning for Experiment 2. Left panel: judgments of learning and actual learning effectiveness for control condition. Right panel: judgments of learning and actual learning effectiveness for pretraining condition ........................................................................................................... 35

Figure 4.1. Training phase trial for the incidental-learning condition .................................. 41

Figure 4.2. Classification test performance for Experiment 3 ............................................... 42

Figure 4.3. Examples of the paintings used in Experiment 4. Those in the top row represent the Baroque era and are paintings, from left to right, by Rubens, Poussin, and Rembrandt, respectively. Those in the bottom row represent the Impressionism era and are paintings by Renoir, Cassatt, and Sisley, respectively ......................................................................................................................... 47
Figure 4.4. Classification performance for eras as a function of presentation conditions in Experiment 4 ................................................................. 50

Figure 4.5. Classification performance for artists as a function of presentation conditions in Experiment 4 ................................................................. 51

Figure 5.1. Classification test performance for Part 1 and Part 2 as a function of blocked and interleaved study schedules in Experiment 5 ........................................................................ 62

Figure 5.2. Examples of the handwriting samples used in Experiment 6a by two authors, with those in the top row all from one author and those in the bottom row all from a different author................................................................................................. 70

Figure 5.3. Classification performance following feedback and observational training collapsed across Part 1 and Part 2 in Experiment 6a ........................................................................ 72

Figure 5.4. Averaged classification performance across Part 1 and Part 2 for the four training conditions in Experiment 6a, where F signifies feedback training and O signifies observational training, and the left and right letters signify the type of training used in Part 1 and Part 2, respectively, during the study phase .................................................... 73

Figure 5.5. Performance during feedback training in Experiment 6a. Left panel shows performance during the Part-1 study phase for the Feedback-Feedback and Feedback-Observational conditions combined. Middle panel shows performance during the Part-2 study phase of the Feedback-Feedback condition. Right panel shows performance during the Part-2 study phase of the Observational-Feedback condition .............................................. 74

Figure 5.6. Classification test performance Part 1 and Part 2 for the Feedback-Feedback, Feedback-Observational, Observational-Feedback, and Observational-Observational conditions of Experiment 6a ........................................................................ 75
Figure 5.7. Experiment 6b Part-1 classification test performance for feedback training and observational training conditions, averaged across control and information conditions . 78

Figure 5.8. Classification performance for Experiment 6b. Left panel shows classification test performance for control condition. Right panel shows classification test performance for the information condition ................................................................. 79

Figure 5.9. Experiment 6b classification accuracy during feedback training for Part 1, for Part 2 of control condition, and for Part 2 of information condition ............................................. 80

Figure 5.10. Control condition participants’ judgments regarding best study method and participant’s actual performance for Experiment 6b ................................................................. 80
ACKNOWLEDGMENTS

Thank you first and foremost to Dr. Elizabeth Ligon Bjork and Dr. Robert A. Bjork. You gave me the opportunity to learn, grow, and interact with brilliant people in a one of a kind academic setting. It has been an honor to observe the thought processes of and work with two experts in the field, nay experts in life. Thank you for inviting me to be a part of something amazing. Your insight and help have meant so much. Elizabeth, thank you so very much for your help in editing this dissertation.

I would like to thank Dr. Barbara Knowlton for being so cool, funny, and brilliant. Dr. Douglas Bell, I greatly appreciated your support and helpful feedback. Thank you Dr. Nate Kornell; you gave me great direction and are an inspiration. Dr. Jesse Rissman and Dr. Alan Castel, thank you for great ideas to pursue and interesting interpretations of data. Dr. Steve Smith, you are a wonderful professor and I am lucky to know you. Dr. Benjamin Storm, thank you for being a great mentor.

Saskia Giebl, thank you for your help with the development and design of Experiment 4. Mikey Garcia, thank you for writing the Collector program, which made conducting studies exponentially faster. Thank you current and former Bjork Lab friends: Genna Angello, Michael Friedman, Doe Buchli, Saskia Giebl, Toshiya Miyatsu, Jeri Little, John Nestojko, Mikey Garcia, Adam Blake, Carole Yue, Veronica Yan, and Courtney Clark, for fun conversations and keeping it real. The laboratory environment has been so constructive and I have really appreciated everyone’s thoughtful comments and help. Thank you, Maverick C. and Alex J. for your assistance with research.

Thank you to my family for being so supportive and encouraging. Okasan, your wisdom amazes me and always makes me see the world from a different perspective. I love you. Thank
you to my brother Alex for always thinking so highly of me. I am so very proud of you. Dad, thank you for believing in me. Thank you, Aunt Elise, for encouraging me to pursue education. And thank you Nicholas, my love, for being an amazing partner.
<table>
<thead>
<tr>
<th>Year Range</th>
<th>Position/Program</th>
<th>Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006 – 2008</td>
<td>Laboratory Assistant</td>
<td>University of California, Los Angeles</td>
</tr>
<tr>
<td>2007</td>
<td>Bachelor of Arts, Psychology</td>
<td>University of California, Los Angeles</td>
</tr>
<tr>
<td>2008</td>
<td>Graduate StudentResearcher</td>
<td>University of California, Los Angeles</td>
</tr>
<tr>
<td>2009 – 2011</td>
<td>Teaching Assistant/Teaching Associate</td>
<td>University of California, Los Angeles</td>
</tr>
<tr>
<td>2010</td>
<td>Master of Arts, Psychology</td>
<td>University of California, Los Angeles</td>
</tr>
<tr>
<td>2012</td>
<td>Candidate of Philosophy, Psychology</td>
<td>University of California, Los Angeles</td>
</tr>
<tr>
<td>2012</td>
<td>Research Intern</td>
<td>Institute for Creative Technologies</td>
</tr>
<tr>
<td></td>
<td></td>
<td>University of Southern California</td>
</tr>
<tr>
<td>2012 – 2013</td>
<td>Teaching Fellow</td>
<td>University of California, Los Angeles</td>
</tr>
</tbody>
</table>
PUBLICATIONS AND PRESENTATIONS


CHAPTER 1 – GENERAL INTRODUCTION

Essentially from birth, humans need to engage in the process of learning by example. That is, to make sense of the world in which we live, we must engage in the fundamental cognitive process of abstracting concepts and categories from our encounters with specific exemplars from those different categories. This process of abstracting concepts and categories from exemplars is referred to as induction, or inductive learning. Inductive learning allows us to extract commonalities among objects belonging to the same category (Chin-Parker & Ross, 2002; Queller, Schell, & Mason, 2006), detect differences between objects belonging to different categories (Goldstone, 1996), identify new items as being a member of a newly learned category (Kornell & Bjork, 2008), and construct entirely new categories (Zeithmova & Maddox, 2009). Inductive learning is thus a fundamental component of human cognition that begins at birth and continues to be critical in daily life as we are faced with an array of categorization tasks. The ability to induce categories and accurately categorize unfamiliar objects is also an important goal of formal education and training in many fields. Medical students, for example, must learn to diagnose disorders in patients who may present a novel combination of symptoms and of varying severity. For all of these reasons, the goal of gaining a fuller understanding of how inductive learning can be optimized is important to pursue.

The overarching goals of my dissertation research were, first, to investigate which training methods appear to optimize inductive learning, then to uncover why those methods produce more effective learning, and, finally, to explore ways to make these strategies more accessible to people. I followed two themes in the conduction of my research. The first theme focused on gaining a better understanding of the cognitive processes underlying efficient inductive learning. To this end, I explored different mechanisms of category feature processing
to determine why certain study conditions are more beneficial to inductive learning than others with the goal of being able to recommend study methods that maximize the efficiency of category learning. The focus of the second research theme was to determine whether people—having experienced effective inductive learning strategies—might then be able to adopt or successfully employ them during new inductive learning situations. An ultimate goal for this research theme was to uncover the circumstances that led to effective transfer of learning strategies, which could have important implications for how to improve category learning in a variety of contexts, including formal education and training contexts where the need to engage in successful and efficient inductive learning is both prevalent and critical for mastery and achievement.

**Interleaved Study in Inductive Learning**

Since as early as the 1800s, researchers have investigated how study events should be sequenced in order to maximize learning (Ebbinghaus, 1885/1964). In such research, temporally distributing study trials has been repeatedly shown to benefit learning, and this observed benefit of spacing study trials versus massing them is referred to as the *spacing effect* (for a review, see Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006). The spacing effect is a robust phenomenon that has been replicated with different materials, such as foreign language vocabulary (e.g., Bahrick, Bahrick, Bahrick, & Bahrick, 1993) and text passages (e.g., Dempster, 1986) and across various conditions, including ones highly applicable to educational settings (e.g., Kornell, 2009).

Recently, however, Kornell and Bjork (2008; Experiment 1a) explored whether there might be one type of learning situation for which massing of study trials would be more beneficial than spacing them: namely, the learning of categories or inductive learning. Thus, in their study, the learning task for participants was not to memorize the items presented for study
in preparation for a later recall or recognition test, but rather to study the items in preparation for a later classification test. Specifically, participants studied different paintings by several different artists with the goal of trying to learn the artistic style of each painter. At test, new paintings by each artist were shown along with the names of all the artists that had been studied, and the participants’ task was to classify each as being painted by one of the studied artists. For half of the artists, their paintings had been presented in a spaced schedule, such that the paintings of one artist were intermixed with paintings by other artists, also referred to as *interleaved* study. For the other half of the artists, all of the paintings by one artist were presented in succession, also referred to as *blocked* study. Kornell and Bjork assumed that blocked study would assist in the difficult task of inducing a given painter’s artistic style more than would interleaved study because noticing the commonalities across the different paintings of a given artist would be facilitated by studying them one after the other or in a blocked manner.

What Kornell and Bjork (2008) found, however, was the opposite. Inductive learning was yet another case in which spacing (more specifically interleaving) proved more beneficial to learning than did massing (or blocking). Additional investigators have also demonstrated a benefit of interleaved study in inductive category and skill learning. Rohrer and Taylor (2007), for example, have found that intermixing the practice of solving different types of math problems led to better performance at a later test than did blocked practice. Zulkiply, McLean, Burt, and Bath (2012) have found that interleaved exposure to case studies of different psychological disorders led to a better ability to classify new case studies of those disorders correctly as compared to when they had been studied in a blocked fashion. And Hatala, Brooks, and Norman (2003) have demonstrated that medical students learned to diagnose different disorders from electrocardiographs better after interleaved practice as compared to blocked practice. Thus, the
benefits of interleaved practice for effective induction have been replicated numerous times and across a range of different types of skill or category learning.

**Discrimination Hypothesis: Noting Features that Distinguish between Categories**

Among the possible interpretations for the benefit of interleaving over blocking, Kornell and Bjork (2008) posited that when category exemplars were interleaved, exemplars of different categories were presented in succession, allowing participants to make contrasts between categories and to identify features that were unique to each category. According to this discrimination hypothesis, interleaving allows for the critical process of category contrast in inductive learning. An obstacle in attributing such benefits to interleaving, however, is that interleaving confounds the effects of placing exemplars of different categories next to each other in time (temporal juxtaposition) and the effects of separating exemplars of the same category in time (temporal spacing). If spacing alone can account for the benefits observed in category learning, then interleaving is simply a time-efficient way to schedule the spaced presentations of exemplars from the same category. But if spacing alone does not provide the same benefits as interleaving, the between-category juxtapositions of exemplars from different categories inherent to interleaving may be critical to category learning.

Taylor and Rohrer (2010) recognized this issue and designed an experiment that equated temporal spacing in the interleaved and blocked conditions. In their experiment, college students learned to calculate the volume for four obscure shapes. The study phase consisted of eight practice problems for each of the four shapes, which were either presented in a blocked schedule or interleaved schedule. In the interleaved condition, the presentations of practice problems for a particular shape were spaced out by 30 s worth of practice problems for other shapes. To match this temporal spacing in the blocked condition, 30 s of a filler task was inserted between each
practice problem for the same type of shape. Finally, to equate for total time on task in the two presentation conditions, participants in the interleaved condition engaged in the filler tasks that were used in the blocked condition after completing all of the practice problems. Even after controlling for spacing between practice problems for the same shape and total time on task, Taylor and Rohrer found that the interleaved study condition led to better performance at test and to fewer errors of solution discrimination.

Kang and Pashler (2012) were also interested in whether Kornell and Bjork’s (2008) findings of better inductive learning with an interleaved presentation schedule was simply due to the temporal spacing inherent to the interleaved presentation schedule or whether the juxtaposition of exemplars from different categories during interleaving was a critical contributor to the observed benefits for interleaving. To explore this issue further, they (Experiment 1) presented participants with 24 paintings by three artists in different scheduling conditions followed by a test in which, similar to the Kornell and Bjork procedure, new paintings by the studied artists were presented along with a list of the three studied painters’ names, and participant’s task was to classify each as being painted by one of the studied artists. Of most relevance to the present issue, participants either saw the paintings in one of the following three presentation schedules: (a) Massed (all paintings by the same artist were presented in a blocked manner), (b) Interleaved (presentations of the paintings by the same artist were mixed in with the paintings of the other artists), or (c) Temporal Spaced (paintings by the same artist were blocked but separated by the same temporal spacing (filled in this case by presentation of a cartoon drawing that did not have to be remembered) as those in the interleaved condition). The idea of their temporal-spaced condition was to introduce temporal spacing without temporal juxtaposition of exemplars from different categories. Thus, if temporal spacing of paintings per
se—rather than the temporal juxtaposition of paintings by different artists—was the critical condition for producing the benefits of interleaving for inductive learning observed by Kornell and Bjork, there should be no difference in performance between the temporal-spaced and the interleaved conditions, which, in turn, should both produce better performance than would the massed condition. What Kang and Pashler found was that the interleaved condition led to significantly better induction performance than the massed and temporal spaced conditions. Thus, they concluded that temporal juxtaposition of exemplars of different categories played a role in the benefits of interleaving for inductive learning.

Birnbaum, Kornell, Bjork, and Bjork (2013) have also assessed the separate effects of temporal juxtaposition and temporal spacing on inductive learning in a series of studies. In one of their studies (Experiment 2), participants studied four exemplars each for 16 different species of butterflies during an initial training phase in which the exemplars for half of the species were presented in a blocked fashion while the exemplars for the remaining half were presented intermixed with one another. Additionally, half of the participants viewed all of the exemplars contiguously—that is, with no temporal spacing between exemplar presentations—whereas, for the other half of the participants, the presentations of the exemplars were spaced apart from each other by a 10-s interval during which a trivia question was presented. The crossing of these two variables (blocked vs. interleaved and spaced vs. contiguous) created four presentation conditions: blocked-contiguous, blocked-spaced, interleaved-contiguous, and interleaved-space. Thus, by examining all four of these presentation conditions, which are illustrated in Figure 1.1, their design allowed for a more complete analysis of the potential separate effects of interleaving and spacing than the previous designs of Taylor and Rohrer (2010) and Kang and Pashler (2012), and the pattern of findings observed was as follows. First, performance on the inductive learning
task—correctly classifying the species of a new butterfly—was better when the exemplars had been studied in an interleaved rather than a blocked manner in the contiguous conditions, thus replicating the findings from Kornell and Bjork (2008). When trivia questions were used to space out study trials however, the benefit of interleaving went away, suggesting that being able to contrast exemplars of different categories is advantageous for inductive learning. That is, the pattern of these results demonstrates that spacing alone cannot account for the benefit of interleaving as the benefit of interleaving disappeared when temporal juxtaposition of exemplars of different categories was taken away.

Blocked-contiguous

 Blocked-spaced

 Interleaved-contiguous

 Interleaved-spaced
In another study conducted by Birnbaum et al. (2013, Experiment 1), the effect of temporal juxtaposition was further assessed using exemplars of bird species (created by Jacoby, Wahlheim, & Coane, 2010) rather than exemplars of butterfly species. Additionally, in this study, the effect of time on task was also examined. Four exemplars of eight different bird species—although always interleaved with exemplars of other species—were presented in three different conditions: contiguous, grouped trivia, and alternating trivia, which are illustrated in Figure 1.2. In the contiguous condition, exemplars of the eight different species were presented intermixed with one another in the standard way. In the grouped trivia condition, one exemplar from each species was presented in a random order in a first series of exemplars, followed by eight 10-s trivia questions, which was then followed by another series of one exemplar from each species randomly ordered, then another block of 10-s trivia question, and so on until all four exemplars from each of the eight species had been presented. In the alternating trivia condition, exemplars were presented in an interleaved order and a 10-s trivia question appeared between presentations of each exemplar.
Alternating Trivia

**Figure 1.2.** Study conditions for Birnbaum, Kornell, Bjork, and Bjork (2013) Experiment 1. TQ stands for a 10-s trivia question used to space apart the exemplar presentations.

In the grouped trivia and alternating trivia conditions, temporal juxtaposition was manipulated, while the length of the study phase and the average temporal spacing between exemplars of the same category was kept the same. Thus, the comparison of performance between these two conditions allowed Birnbaum et al. (2013) to examine whether category contrasts provided a special benefit to inductive learning. They found that the contiguous and grouped trivia conditions led to significantly better performance than the alternating trivia condition, providing additional evidence that being able to make contrast between categories during the presentation schedule is critical to inductive learning.

In a recent review article of research investigating the effects of different types of presentation schedules on inductive learning, Rohrer (2012) argues for the importance of discrimination learning in perceptual category learning and mathematics learning. In his review, Rohrer points out that category learning, not only involves the learning of a category’s features, but also the learning of the boundaries that separate similar categories. Thus, in mathematics learning, interleaved practice helps students learn to distinguish between different kinds of problems. Rohrer argues that when practice problems are blocked, the learner must go through the steps to solve each problem. But when practice problems are interleaved, the learner must first identify the kind of problem it is, select the appropriate solution, and then implement the solution. He concludes that interleaved practice feels more difficult to learners and that is
because it is. The additional steps that the learner must engage in during interleaved practice, however, are what then lead to the learner’s better performance on a later test.

In research discussed thus far, the superiority of an interleaved study schedule over a blocked study schedule have mostly been framed in terms of the advantages of interleaved study schedule as opposed to the relative disadvantage of blocked study for inductive learning. Interleaving allows for critical between-category discriminations to be made (e.g., Taylor & Rohrer, 2010; Kang & Pashler, 2012; Wahlheim, Dunlosky, & Jacoby, 2011; Birnbaum et al., 2013) and for skill learning, interleaving allows for task switching and orientation to be practiced (Rohrer, 2012; Goode & Magill, 1986). I posit that the differences in inductive learning performance due to interleaved versus blocked study conditions can also be attributed to the disadvantages of blocked study.

In blocked practice, learners do not have the opportunity to engage in cognitive processes that are critical for experiencing category boundaries. When examples of the same category are presented in succession, the learner may experience a false sense of understanding of the category. Learners may identify features that are common within a category and mistakenly believe that those features are diagnostic of the category even though they also occur in other categories as well. For example, a person may mistakenly believe that having two doors is a feature distinctive to sports cars. Not until the person sees non-sports cars with two doors, will he or she realize that the two-door characteristic is not diagnostic of category membership. Interleaved practice allows the learner to sample the range of variability of features across categories so that learners do not mistakenly latch on to features that are not critical in telling categories apart.
Additionally, when exemplars of a category are presented in succession, learners are at risk of committing confirmation bias. Confirmation bias occurs when one seeks evidence that supports what he or she already believes. Confirmation bias is illustrated by the Wason card selection task (Wason, 1966). In this task, the participant is shown four cards. The participant is told that each card has a number on one side and a letter on the other side. The participant is also told that all of the cards must follow a rule: If a card has a vowel on one side of the card, then the other side must have an even number. The four cards are presented so that only one side of each is showing. The symbols are a 3, 4, A, and K. If one wanted to see if each of the cards followed the rule, which two cards should one turn over?

Given this task, most people mistakenly choose the card showing a 4 and the card showing an A. People want to see examples that confirm the rule. That is, they are hoping to turn over the card with the 4 and find a vowel on the other side, and turn over the card with the vowel and find an even number on the other side. While it is correct that the card with the vowel should be turned over to assure that the other side shows an even number, turning over the card with a 4 is not helpful. The participant should instead look at the other side of the card showing a 3 to be sure that the other side does not contain a vowel.

Confirmation bias can occur during inductive learning as well. Blocked study has been argued to facilitate the extraction of within-category correlations (Goldstone, 1996). That is, during blocked study, a participant may notice features that are consistent across exemplars of the same category. If, however, a learner incorrectly believes that a given feature is diagnostic of and unique to a category, seeing successive exemplars of that category that share that feature only serve to support the false belief. That is, viewing such a series of exemplars will only continue to support the hypothesis that a feature is not only common among the exemplars of the
category, but also diagnostic of the category. Blocked presentation can thus lead learners to engage in inefficient processes that result in erroneous conclusions. Noticing within-category correlations can thus become a liability and may steer the learner away from learning features critical for telling categories apart.

**Outline of the Present Dissertation**

The experiments reported in the present dissertation were designed to reveal a greater understanding of the processes used to contrast or notice differences between to-be-learned categories and those used to compare or notice commonalities among the exemplars within the same category that occur during inductive learning and how these processes can be optimized. Additionally, in this undertaking, the factor of awareness and automaticity of contrast and commonality processing were examined. That is, to understand more fully how contrast and commonality processing occurs during inductive learning, I examined whether these processes require effort and awareness. Thus, intentionality and the ability to actively employ different study strategies during inductive learning were also examined.

In the studies previously discussed (i.e., Taylor & Rohrer, 2010; Kang & Pashler, 2012), the role of spacing out exemplars of the same category and the role of temporally juxtaposing exemplars of different categories were examined. Results from these studies revealed that spacing out exemplars did not lead to the same benefits to inductive learning as did interleaved study, and it was concluded that being able to make contrasts between categories is crucial for efficient inductive learning. The benefits of temporal spacing in conjunction with temporal juxtaposition, however, were not examined in these studies. Thus, Experiments 1a and 1b of the present dissertation were designed to examine the role of varying the temporal spacing between
exemplars of the same category on inductive learning while keeping intact the temporal juxtaposition of exemplars of different categories.

Interleaved practice has time and again been shown to be better than blocked practice for category learning. The typical category-learning paradigm consists of a study phase during which participants view exemplar/label pairs, followed by a classification test phase in which learners are required to classify new exemplars as members of the studied categories. The classification test, however, examines two abilities: (a) the ability to include the test item as correctly belonging to a studied group or category and (b) the ability to recall the group or category names. It is critical to ensure that what is believed to be the benefit of interleaved study for inductive learning is not simply an artifact of the benefit of spaced study for recall learning of category labels. The present Experiment 2 was designed to separate the effects of spacing on recall learning and interleaving on inductive learning in order support the argument that interleaving has unique and potent benefits for category induction.

To understand the cognitive processes involved in category learning, the effects of intentionality and study schedule (blocked versus interleaved) on inductive learning were also examined in the present dissertation. In Experiments 3 and 4, participants’ awareness of the goal to learn category information was manipulated between-subjects and within-subjects, respectively. Examining how interleaving affects unintentional and non-directed category learning helped reveal how contrast processing occurs. Experiment 5 was designed to explore whether participants could deliberately engage in contrast processing in the absence of an interleaved study schedule. In Experiment 5, participants first experienced the beneficial effects of interleaving and were then encouraged to make contrasts between categories in a separate inductive learning opportunity in which exemplars within categories were presented in a blocked
rather than an interleaved fashion. Thus, Experiment 5 investigated whether learners could actively engage in and transfer the contrast processing strategy to new inductive learning. Together, Experiments 3, 4, and 5 help reveal the automaticity of contrast processes and how they occur during inductive learning.

Experiments 6a and 6b built upon the question of the transfer of study strategies. Instead of examining the possibility of transferring the benefits of interleaving, however, the possibility of transferring benefits of feedback training to new learning were examined. Feedback training, like interleaved study, has been shown to be beneficial to inductive learning (e.g., Jacoby, Wahlheim, & Coane, 2010). If the ability to engage in between-category contrast processing cannot be intentionally and successfully applied to new learning opportunities in the absence of an interleaved presentation schedule, perhaps requiring learners to respond with a category label on each study trial is a tactic that people can effectively transfer to new category learning situations. A primary goal of the present dissertation was that the sum of the results from the studies outlined above would create a more cohesive picture of the cognitive mechanisms that underlie inductive category learning.
CHAPTER 2 – THE ROLE OF SPACED RETRIEVAL ON INDUCTIVE LEARNING

As discussed in Chapter 1, between-category juxtapositions are critical to the
differentiation of categories. Birnbaum, Kornell, Bjork, and Bjork (2013, Experiment 2) found
that interleaved study was beneficial to learning, but that when interleaved exemplars were
spaced out with trivia questions, the critical contrast processes were interrupted and classification
performance suffered. Their finding that presenting exemplars of different categories
consecutively is better than spacing them apart, illustrates the point that making between-
category discriminations is important. Their finding does not, however, indicate that the
increased temporal spacing between exemplars of the same category per se is detrimental to
category learning. Provided that the temporal juxtaposition of between-category exemplars
remains intact, and the benefits of between-category contrasts are not interrupted, increasing
temporal spacing of exemplars from the same category may provide benefits to inductive
learning. Indeed, the combined effect of facilitating between-category contrast processing and
maximizing temporal spacing between exemplars of the same category could be particularly
beneficial. Finding evidence for an additive benefit would reveal that interleaved study is
beneficial for at least two reasons: It supports between-category contrast and spaces out
exemplars within categories. But why would spacing of exemplars of the same category be
beneficial? I discuss the roles of retrieval, forgetting, and generalization in temporal spaced
exemplar presentation as evidence that helps to predict a benefit of increased temporal spacing
for inductive learning.

When learners engage in category exemplar study, they may be thinking back to, or
retrieving from memory, other exemplars from the same category in an attempt to create a
cohesive category representation. According to the study-phase retrieval account of the spacing
effect, studying an item that is similar to a previously presented item will prompt its retrieval (Thios & D’Agostino, 1976). Additionally, according to Bjork (1975), retrieval of information is more beneficial to learning if it is made more challenging. Spacing out the exemplars of a given category would make retrieval of previously seen exemplars of the same category more challenging. The term desirable difficulty, coined by Robert Bjork (1994), refers to practices that make learning seem more difficult during acquisition, but ultimately enhance learning. Temporal spacing of exemplars of the same category may thus prove to be a desirable difficulty and enhance category learning. The retrieval effort hypothesis (Pyc & Rawson, 2009) similarly states that more effortful retrievals leads to greater long-term learning. Both the study-phase retrieval account and the retrieval effort hypothesis can be applied to inductive learning. Studying an exemplar may lead to the retrieval of previously studied exemplars from the same category. Provided that the retrieval is successful, it may be beneficial to make this retrieval as challenging as possible by increasing the temporal spacing between exemplars of the same category.

Furthermore, greater temporal spacing between exemplars of the same category presumably leads to greater forgetting for the details of the previously studied exemplars. This forgetting, however, may help learners to extract a generalized representation of a category. Successful category learning does not rely on memory for every detail of each encountered exemplar. In fact, remembering every detail would be very inefficient. The features that should be recalled and used later to determine category membership are features that are correlated within a category and are diagnostic of category membership. Some amount of forgetting could thus play an advantageous role in critical feature extraction.

If the previously viewed exemplar of a given category is difficult to retrieve, the presentation a subsequent exemplar of that category could cue the retrieval of the previously
viewed exemplar in as much as they have overlapping features. Furthermore, because the overlapping features act as the cues for the retrieval of previously seen exemplars of the same category, those cues can become strengthened. For example, when a child is learning the concept of a dog, he or she may think back to other exemplars referred to as dogs. One dog may have had a green collar and another dog may have had a brown spot on its left ear. These features, however, are not important to the concept of dog and do not occur in most examples of dogs. Thus such unique features are not likely to be strengthened upon further exposures to examples of dogs. On the other hand, the overlapping features (e.g., four legs, snout, and wagging tail), which are highly correlated across exemplars, act as reminding cues and are strengthened in memory. Overlapping features are thus highlighted and irrelevant features are deemphasized. Therefore, spacing-induced forgetting leads to the retrieval of relevant information (Vlach, Ankowski, & Sandhoff, 2011). Gist information is thereby strengthened and consequently leads to better inductive learning.

Ross, Perkins, and Tenpenny (1990) also discuss the importance of discovering commonalities within categories, and describe category learning as being similar to analogical reasoning. Just as the similar structural characteristics between scenarios are extracted during analogical reasoning, correlated features between exemplars are extracted during category learning; and, importantly, both lead to the derivation of more general knowledge. Ross et al. argue that the retrievals of exemplars of the same category lead to explicit comparisons between exemplars within a category and to the storage of characteristics that are common to that category. They further argue that the non-automatic process of finding commonalities relies on the successful reminding of previous instances and successful detection of common features—very similar to Bjork and Bjork’s (2011) assertion that desirable difficulties are only desirable to
the extent that the difficulty they pose can be overcome. That is, while retrievals should be challenging, retrievals must be successful in order for learners to be able to discover common characteristics and reap benefits to inductive learning.

I have discussed several reasons to predict that temporal spacing of exemplars of the same category will lead to enhanced category learning. Whereas researchers that have investigated the benefits of interleaving have mostly concluded that interleaving is good because it enhances between-category contrast processing, temporal spacing may also provide a benefit to category learning. Although the work of Taylor and Rohrer (2010) and Kang and Pashler (2012) showed that interleaved presentation led to better inductive learning than presentations spaced out by filler tasks—thus demonstrating the importance of temporal juxtaposition—these results did not prove that temporal spacing was detrimental to or provides no benefits for inductive learning. It is still possible, provided the exemplars are interleaved and temporal juxtaposition of exemplars of different categories remains intact, more temporal spacing between exemplars of the same category (as opposed to less temporal spacing), would be beneficial to inductive learning. I assessed this possibility in Experiments 1a and 1b of the present dissertation, which were designed to combine the benefits of temporal juxtapositions of exemplars of different categories while enhancing the effectiveness of discovering within-category commonalities.

**Experiment 1a: Effects of temporal contrast and spaced retrieval on inductive learning.**

In order to investigate whether increasing temporal spacing between exemplars within categories during study can play a beneficial role in inductive learning, Experiment 1a (also presented in Birnbaum, Kornell, Bjork, and Bjork (2013) as Experiment 3) was devised. The
The purpose of Experiment 1a was to examine the effects of varying temporal spacing on category learning while keeping temporal juxtapositions intact.

Participants. Participants were 53 people recruited through Amazon Mechanical Turk, a website that pays individuals to partake in small tasks.

Materials. The materials were 80 photographs depicting butterflies of 16 different species. The participants studied four exemplars per butterfly species and were tested on a fifth exemplar. The butterfly names presented to participants for each species were based on the actual name of each species, but some names were shortened, made into one word, or changed entirely if the real name described physical characteristics of the butterfly. The names employed were Admiral, American, Baltimore, Cooper, Tiger, Streak, Harvester, Mark, Lady, Elfin, Pipevine, Sprite, Tipper, Satyr, Viceroy, and Nymph.

Design and Procedure. The participants studied four exemplars from each of 16 species. Each butterfly image was presented for 4 s, with the species’ name appearing below the image. Participants were instructed to attend to the presented images and corresponding species’ names, and were told that there would be a categorization test following the study phase. After participants viewed all study images, participants were tested on their ability to classify never-before-seen images of the butterfly species that were studied. One new image for each of the 16 butterfly species was presented, one at a time, along with the names of the 16 species below the image. Participants clicked on one of the 16 options for each test item. Test trials were not timed, and no corrective feedback was given.

There were two between-subjects conditions: the small-spacing condition and large-spacing condition, whose corresponding study schedules are shown in Figure 2.1. Both conditions presented the exemplars in an interleaved order. The two conditions differed by the
number of exemplars that were presented between the presentations of exemplars of the same category. In the small-spacing condition (n = 26), participants studied images of a given species separated by an average of three other exemplars of different butterfly species. In the large-spacing condition (n = 27), exemplars of a given species were separated on average by 15 exemplars of other butterfly species.

**Small spacing**

![Small spacing image]

**Large spacing**

![Large spacing image]

*Figure 2.1. Study conditions for Experiment 1a.*

In the small-spacing condition, four exemplars (one from each of four species) were presented in a random order, then a different four exemplars (one from each of the same four species) were presented in a newly randomized order, and so on, until all four exemplars from each of those four species had been presented. Then, another set of four species would be presented in a similar manner, and so on until all exemplars of all 16 species had been presented. In the large-spacing condition, exemplars (one from each of all 16 species) were presented, then different exemplars (one from each of the 16 species) were presented in a different order, and so on, until all exemplars had been presented. In the large-spacing condition, rather than randomizing all 16 exemplars within one grouping, the 16 species were divided into groupings of four, and order was randomized within each grouping of four. For example, if an exemplar of a Lady, Admiral, Cooper, and American butterfly appeared in that order within the first
presentation of 16 exemplars, exemplars of the same species would be grouped together and randomized among one another within each subsequent grouping of 16 exemplars.

**Results.** As indicated in Figure 2.2, the participants in the large-spacing condition performed significantly better (M = .42, SD = .198) on the induction test than did those in the small-spacing condition (M = .29, SD = .125), $t(51) = 2.74, p = .008, d = 0.752$.

![Figure 2.2. Classification test performance for Experiment 1a.](image)

**Discussion.** The small-spacing and large-spacing conditions were equated in terms of the exemplars studied, the tasks in which the learners engaged, and the total time spent on task. Additionally, both the small-spacing and large-spacing conditions allowed for approximately the same number of between-category exemplar juxtapositions to occur. That is, each category’s exemplars were juxtaposed with exemplars of roughly the same number of other categories. These controls between the two conditions allowed for an examination of the effect of temporal spacing.

The results of Experiment 1a indicate that larger spacing led to better inductive learning. An interleaved study schedule creates temporal juxtapositions of exemplars of different
categories and also creates spacing between exemplars of the same category. Results from Experiment 1a revealed that each of these features—inherent to interleaved study—may play a role in why interleaving is beneficial to inductive learning. Additionally, the pattern of results observed in the present experiment provides support for accounts discussed earlier, including the study-phase retrieval account, retrieval effort hypothesis, and the category generalization explanation.

According to Ross, Perkins, and Tenpenny (1990), the act of retrieving exemplars and seeking commonalities among exemplars of a given category is an explicit process. Presumably, when learners are aware that there will be a later test on category classification, learners actively attempt to retrieve previous exemplars of the same category in order to create a cohesive category grouping. This consciously exerted effort to retrieve studied material strengthens memory and enhances category learning. When spacing between exemplars of the same category is increased, retrieval becomes more difficult and overcoming this difficulty appears to hold advantages to learning. Birnbaum et al. (2013, Experiment 2) demonstrated that spacing out the presentations of blocked exemplars led to better category learning than did consecutively presentations of blocked exemplars. The blocked-spaced condition, relative to the blocked-contiguous condition, had the advantage of increased difficulty of retrieval of previous exemplars. The large-spacing condition in the present experiment combined the benefits of difficult retrievals of exemplars of the same category and the benefits of between-category juxtapositions. Results revealed that when the benefits of between-category juxtapositions are present, increased spacing between exemplars of the same category, thereby making retrieval processes more effective, is optimal for category learning.
Experiment 1b:

A closer look at the role of temporal contrast and spaced retrieval in inductive learning.

One potential limitation of Experiment 1a is that increased temporal spacing (the manipulated variable of interest) was confounded with the number of opportunities for between-category contrasts. In the small spacing condition, four exemplars, one from each of four species, were shuffled among one another and temporal contrasts among these four categories could be made. For example, exemplars of a Lady, Admiral, Cooper, and American butterfly were presented consecutively, allowing for contrasts to be made between these four butterfly species. The transition between groupings of four exemplars, however, led to additional opportunities for between-category contrasts. For example, after all Lady, Admiral, Cooper, and American butterflies were presented, exemplars from the Elfin, Nymph, Baltimore, and Mark categories were presented. The last exemplar of the initially presented four species was then temporally juxtaposed with the first exemplar of the following four species (e.g., an American would be followed by an Elfin). In the small-spacing condition, these contrasts occurred when the first set of four species was exhaustively studied and presentations of the second set of four species began, when the second set was complete and the third set began, and when the third set was complete and the fourth set began. Thus, there were three additional between-category comparisons.

In the large-spacing condition, however, there were 15 opportunities for contrasts that occurred between groupings of four exemplars. This difference in additional contrast opportunities (three versus 15) could have created a substantial benefit to inductive learning and driven the effect observed in Experiment 1a. In order to control for the differences in between-category contrasts, the additional contrast opportunities were eliminated by the insertion of a simple
math addition question between every four exemplars in both study conditions of the present
Experiment 1b.

Two additional changes from the design of Experiment 1a were made in Experiment 1b. In most
inductive learning experiments, the final test is a classification task that requires the learner to
recognize or recall the appropriate category label. Given this type of test, some or all of the
enhanced performance observed in the interleaved practice conditions may be due to the
beneficial effects on participants’ memory for the category labels. In order to explore this
possibility, a pretraining phase for labels was introduced in Experiment 1b. During the pre-
training phase, participants studied all 16 species names and were tested on them with a free-
recall test; then, this study-test cycle was repeated once more. After the category labels were
learned, participants proceeded on to normal category learning task using the butterfly images.

The last change made to Experiment 1b was to use a “medium-spacing” condition in which
ever exemplars of the same category were spaced out on average by seven exemplars, rather
than the small-spacing condition of Experiment 1a, in which exemplars of the same category
were spaced out on average by three exemplars. This change was introduced to address a
possible criticism of comparing the small-spacing and large-spacing conditions in Experiment
1a: namely, that the presentation of exemplars in the small-spacing condition was more
predictable than that in the large-spacing condition. Groupings of four species are cycled through
four times. Thus, the cyclic nature of the task could potentially have created a false sense of
mastery that the large-spacing condition did not. The small-spaced condition was thus altered to
isolate the differences of temporal spacing more accurately and to minimize differences in item
predictability and potential attention attenuation.

Method
Participants. Participants were recruited from Amazon Mechanical Turk, a website that employs individuals to partake in small tasks in exchange for money. A total of 61 participants from the United States (26 females and 35 males) participated.

Materials. Materials were identical to those used in Experiment 1a, with the addition of 15 slides that each presented a simple math addition problem.

Design and Procedure. Participants first partook in a pre-training phase, during which they studied a list of all the 16 butterfly species names used in the experiment, with the names appearing on the computer screen for 30 s. Following presentation of the list, a textbook appeared on the screen for 45 s, and participants were to type as many of the names as they could recall into the textbook. This 30-s study and 45-s free-recall cycle was then repeated once more. There were no performance criteria for the free-recall test; all participants, regardless of free-recall performance, continued on to the study phase.

During the category-learning study phase, participants viewed a total of four exemplars for each of 16 butterfly species and 15 simple math addition questions (e.g., 4 + 5). Each butterfly image was presented for 4 s and each math question was presented for 4 s. Presentation order (medium spacing-spaced and large spacing-spaced) was manipulated between-subjects. In both conditions, illustrated below in Figure 2.3, four butterfly images were presented consecutively, followed by one math question, and so on, until all materials for the study phase had been presented. In the medium spacing-spaced condition (n = 30), each exemplar from a given species was spaced out on average by seven exemplars of other species as well as by two math questions. In the large spacing-spaced condition (n = 31), each exemplar from a given species was spaced out on average by 15 exemplars of other species as well as by four math
questions. Each grouping of four exemplars was block randomized. Items in Figure 2.3 do not show the blocked randomization that was employed in the design.

Medium spacing – spaced

\[
\begin{array}{cccccccc}
A_1 & B_1 & C_1 & D_1 & E_1 & F_1 & G_1 & H_1 \\
A_2 & B_2 & C_2 & D_2 & E_2 & F_2 & G_2 & H_2 \\
\end{array}
\]

Large spacing – spaced

\[
\begin{array}{cccccccc}
A_1 & B_1 & C_1 & D_1 & E_1 & F_1 & G_1 & H_1 \\
I_1 & J_1 & K_1 & L_1 & M_1 & N_1 & O_1 & P_1 \\
\end{array}
\]

\[
3+4\quad 5+2\quad 3+4\quad 5+2 \\
...\quad ...\quad ...\quad ...
\]

Figure 2.3. Study conditions for Experiment 1b. Letter denotes butterfly species, subscript value represents exemplar number.

Following the study phase, participants were given a classification test. One never-before-seen exemplar from each of the 16 butterfly species was presented one at a time, and each test image was presented with the 16 species names appearing below it. Participants clicked on what the deemed to be the appropriate species name. No feedback was given, and test trials advanced when the participant entered a response.

Results. As indicated in Figure 2.4, the participants in the large spacing-spaced condition performed significantly better (M = .39, SD = .143) than did those in the medium spacing-spaced condition (M = .29, SD = .195), t(59) = 2.27, p = .027, d = 0.582.
Discussion. Even when controlling for temporal juxtaposition, the larger temporal spacing between within-category exemplars led to better category learning. This finding provides further evidence that interleaving is beneficial to category learning for at least two reasons: (a) exemplars of different categories are temporally juxtaposed with one another allowing for category boundaries to be learned, and (b) exemplars of the same category are spaced out, creating a challenging yet gainful retrieval of previously viewed exemplars of the same category.

As discussed earlier, desirable difficulties are only desirable when the difficulty can be overcome, suggesting that there is likely an upper limit to the amount of temporal spacing that can be introduced before retrievals are not successful. In the present Experiments 1a and 1b, temporal spacing as large as 15 intervening exemplars of other categories did not impair learning. Suppose, however, that the task was to learn 30 categories by studying exemplars from each category. In this situation, an upper limit to the benefits of temporal spacing might well be revealed: that is, fully interleaving category exemplars in groupings of 30 could prove
suboptimal for learning because retrievals of prior exemplars from the same category might become too difficult to be successful.

The results of Experiments 1a and 1b demonstrate that the temporal spacing of within-category exemplars plays a role in the benefits of interleaved study. Two factors—not studied in the present studies—could, however, affect the amount of temporal spacing that is optimal for learning: presentation time of exemplars and the nature of the to-be-learned categories. Increasing the presentation time of each exemplar may alter the optimal number of intervening exemplars for inductive learning. By increasing presentation time, the retrieval strength for each exemplar may be strengthened. On the other hand, increasing presentation time of exemplars also increases the temporal spacing between within-category exemplars by a factor of presentation time, thus making retrievals more difficult. Further investigation is needed to clarify the effect of presentation time on how best to interleave to-be-learned category exemplars.

Additionally, the nature of the study materials will likely affect the upper limit to the benefits of temporal spacing. Categories that have low between-category similarities, high within-category similarities, and are otherwise easier to learn because of pre-experimental familiarity will benefit from larger spacing between within-category exemplars. In conclusion, although results from Experiments 1a and 1b revealed that greater spacing was beneficial to inductive learning, whether there is an absolute upper limit that applies to all category-learning scenarios has not been adequately explored or established. Researchers and educators should conduct specified investigations of a variety of learning materials, taking into account the number of to-be-learned categories and the nature of the categories, to determine the optimal temporal spacing for inductive learning of those materials.
CHAPTER 3 – EFFECTS OF INTERLEAVINED STUDY ON RECALL LEARNING AND INDUCTIVE LEARNING

In many studies examining category learning, classification tests are used to measure learning. An accurate classification response results from two processes: accurate category representation and the ability to associate that category to its appropriate category label. Recall learning, such as the learning that occurs as a result of repeated presentations of category labels, benefits from the spaced presentations of the to-be-learned information (for a review on the Spacing Effect see Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006). Temporal spacing of repeated presentations of category labels is inherent to an interleaved schedule. It thus becomes unclear whether the benefits of interleaving for category exemplar classification are due to enhanced category learning or enhanced label learning.

Kornell and Bjork (2008) recognized this potential problem and devised a test that would measure category knowledge independent of category label knowledge. Instead of presenting participants at test with a new painting and asking participants which of the 12 studied artists had painted it, participants were presented new paintings by studied artists and paintings that were highly similar but painted by different artists. Participants were asked whether each painting was painted by a studied artist or a different artist. Thus, this test allowed Kornell and Bjork to examine participants’ category knowledge independent of their ability to associate the category to the appropriate category label, and they found that interleaved study led to better performance than did blocked study.

Another way to examine the effects of interleaving on inductive learning while controlling for effects of spacing on recall learning (e.g., learning the names of the artists or butterfly species) is to eliminate the need for recall learning. In many category-learning studies,
the study phase in which the exemplars are presented is also the first time participants are exposed to the labels of the categories they are to learn. Thus, participants have to learn category features and category labels simultaneously during the study phase, making the effects of interleaving on inductive learning versus the effects of spacing on recall learning difficult to separate. By introducing a pretraining phase, during which participants learn the category labels, the exemplar study phase will require less category label learning and focus more on exemplar feature learning. Final test performance therefore will reveal the effects of interleaved study schedule on category learning and less the effects of spacing on category label learning. Whereas Kornell and Bjork’s solution was to isolate and assess only category-grouping knowledge at test, pretraining allows researchers to examine the effects of interleaving on category learning while diminishing any contributions of interleaving to category-label learning.

**Experiment 2:**

Does recall learning drive the interleaving effect?: Using category label pretraining.

Pretraining is becoming more frequently used in category learning experiments, but a straightforward comparison between a procedure using pretraining and a procedure without pretraining has not yet been conducted. To make such a direct comparison was the primary goal of the present Experiment 2. One condition was a replication of Kornell and Bjork’s (2008) procedure, and the other condition was identical to the replication but with the inclusion of a pretraining phase. If the results are similar between the two conditions, we can be confident that interleaved study provides a unique benefit for category learning and its benefit is not simply an artifact of the spacing benefit for recall learning of category labels.

**Participants.** Participants were recruited through Amazon Mechanical Turk, and 34 females and 28 males (mean age = 33.58) participated.
**Materials.** Materials were identical to those used by Kornell and Bjork (2008), which were images of ten paintings by each of 12 artists (Braque, Cross, Hawkins, Juras, Lewis, Mei, Mylrea, Pessani, Schlorff, Seurat, Stratulat, and Wexler). All paintings depicted landscapes or skyscapes.

**Design and Procedure.** Study schedule (interleaved vs. blocked) was manipulated within-subjects and pretraining was manipulated between-subjects (control or no pertaining vs. pretraining). The control condition was a replication of Kornell and Bjork’s (2008) Experiment 1a design. Specifically, participants viewed six exemplars for each of the 12 artists. Six of the artists’ exemplars were presented in a blocked order and the other six artists’ exemplars were presented in an interleaved order. The study phase was ordered as BIIBBIIBBIIB, with B standing for a series of six paintings of the same artists presented consecutively and I standing for a series of six paintings, one by each of the six painters assigned to the interleaved schedule condition. Each painting was presented along with the artist name printed below the painting and appeared for 5 s each.

In the pretraining condition, participants were presented with the names of all 12 artists on the screen for 30 s and asked to study them. Then they were asked to free recall as many of the 12 names as possible, typing them into a textbox that appeared on the computer screen for 45 s. Then, this study-test cycle was repeated once more. Following this pretraining exercise, participants proceeded to the study phase, which was identical to the study phase for the control condition.

After all paintings were presented during the study phase, the ability of participants to classify new paintings by the 12 studied artists was assessed. Four new exemplars for each of the 12 artists were tested in four test cycles, with each cycle containing one new exemplar from each
studied artist. After each response, corrective feedback was given to the participant. There were thus four test cycles with feedback. After the test phase was completed, participants were asked which study schedule they believed led to better classification performance at test, with their choices being: blocked, interleaved, or same.

**Results.** Performance observed for the control condition closely matched the results from Kornell and Bjork’s (2008) Experiment 1a. For the control condition, interleaved study ($M = .55$, $SD = .27$) led to significantly better classification performance in the first test block than did blocked study ($M = .37$, $SD = .22$), $t(33) = 3.5$, $p = .001$. These results are very similar to Kornell and Bjork’s results. Average performance for their first test block was $M = .61$ ($SD = .24$) for categories that were studied in an interleaved schedule and $M = .35$ ($SD = .24$) for categories studied in a blocked schedule.

Performance across test blocks for the control condition, shown in the left panel of Figure 3.1, was also similar to those observed by Kornell and Bjork. Because participants received corrective feedback on each test item, performance improved across test blocks, $F(3,31) = 3.5$, $p = .03$, with most of this effect being driven by the marked improvement across test blocks for blocked categories, $F(3,31) = 7.72$, $p = .001$. The interaction of study schedule and test block was also significant for the control condition, $F(3,31) = 3.86$, $p = .02$. Kornell and Bjork also observed an interaction between study schedule and test block, such that learning across test blocks for the blocked artist categories improved whereas performance for interleaved artist categories remained steady across test blocks.

For the pretraining condition, the results of which are shown in the right panel of Figure 3.1, performance from the first test block alone did not reveal a significant benefit of interleaving. There was however an overall benefit (collapsed across all four test cycles) of
interleaving, $F(1,25) = 10, p = .004$. There was a significant main effect of test block such that performance improved across test blocks, $F(3,23) = 5.2, p = .007$. Improvement across test blocks was significant in both the blocked condition $F(3,23) = 3.12, p = .046$, and the interleaved condition $F(3,23) = 3.05, p = .049$. Because for each test block, interleaving was better than blocking, test block and study schedule did not interact.

To analyze the data further, a 3-way analysis of variance (control vs. pretraining, study schedule, and test block) was also performed. When collapsing across the pretraining and control conditions, a significant effect of test block emerged, $F(3,57) = 8, p < .001$, as well as a significant benefit of interleaving over blocking $F(1,59) = 9.28, p = .003$. The interactions in the control condition and pretraining condition were different. Whereas for the control condition, only the blocked condition showed improvement across test blocks; for the pretraining condition, both the blocked and interleaved study conditions improved across test blocks, with this difference thus producing a significant interaction among the three independent variables (control vs. pretraining, study schedule, and test block), $F(3,57) = 2.76, p = .05$. 

![Graph showing Proportion Correct Classification vs. Test Block for Control and Pretraining conditions.](image)
Figure 3.1. Classification test performance for Experiment 2. Left panel: classification test performance for control condition. Right panel: classification test performance for pretraining condition.

Similar to the judgements of learning made by participants in Kornell and Bjork (2008), most participants in the present study judged blocked study to be more effective for their learning than interleaved study. Figure 3.2 shows a comparison of participants’ judgments regarding in what condition they learned better versus their actual performance for the first test block, with the left panel showing the results for the control condition and the right panel showing the results for the pretraining condition. In the control condition, 56% of the participants correctly classified more interleaved categories than blocked categories and 18% correctly classified more blocked categories than interleaved categories. Only 9% of the participants, however, judged that interleaved study had been better for their learning, while 56% of participants said that blocked study had been better. In the pretraining condition, 44% of the participants correctly classified more interleaved categories than blocked categories and 6% correctly classified more blocked categories than interleaved categories, but 20% of the participants judged interleaved study to be better and 72% believed that blocked study to be better.
Discussion. The control condition of the present experiment replicated the interleaving benefit for inductive learning and the judgements of learning observed in Kornell and Bjork (2008). Performance across test blocks for the pretraining condition, however, differed slightly from that of the control condition and also from that observed by Kornell and Bjork, with performance in the interleaving condition being superior to that in the blocking condition across test trials. Importantly, however, in both the control and pretraining conditions, interleaved study led to better category learning than blocked learning.

The continued improvement in performance observed in the pretraining condition across test blocks could be due to the effect of feedback on category labels with high retrieval strength. That is, because the category labels are well-learned, they are presumed to have high storage strength. If during a test trial, the correct label is not identified (in some cases, indicating low retrieval strength of category label), the corrective feedback would act as a restudy opportunity and result in great benefits to storage strength and retrieval strength. Therefore, participants in
the pretraining condition may have gained more from feedback training because feedback
provided greater gains to storage and retrieval strength of the names.

Despite the difference observed across test blocks, introducing pretraining did not
ultimately result in a lack of a benefit for interleaved study for category learning. These findings
thus suggest that the benefits of an interleaved study schedule is not driven by the benefits of
spaced study for recall learning and that an interleaved study schedule is specifically beneficial
to inductive category learning.
CHAPTER 4 – EFFECTIVE STUDY SCHEDULES FOR INCIDENTAL AND NON-DIRECTED CATEGORY LEARNING

People frequently engage in category learning in formal educational settings. Botany students, for example, learn to classify plants into specific species. Math students learn to select the appropriate algorithm to solve a word problem. And medical students learn to diagnose patients from the symptoms that they present. In these situations, students are aware their learning goal is to understand category groupings, and students anticipate future situations in which that category knowledge will be tested. This type of category learning is called intentional category learning.

Many of the categories we learn throughout our lifetimes, however, are likely acquired under incidental conditions, without explicit direction or instruction. For example, a person may be aware that diners typically serve burgers and shakes but not pizza and beer. The category “diner” was likely acquired through experiences with diners and other types of restaurants that are not diners, such as pizza parlors and delicatessens. The process of learning the category “diner” occurred without the explicit goal to do so and is thus referred to as incidental category learning.

How category exemplar features are processed may depend on the intentionality of the learner during exposures to exemplars. According to Brooks (1978), categories learned under incidental conditions tend to have greater within-category similarity than between-category similarity. In other words, exemplars of a given category must share several common features, albeit not all of these are strictly present in each exemplar, in order for incidental learners to notice grouping patterns. Furthermore, children do not learn to categorize their surroundings based on explicit lessons from adults; rather, much of their learning occurs incidentally. Children
tend to process functionally meaningful groupings in the world at what Rosch and Mervis (1975) refer to as the “basic level,” where categories have maximized within-category similarity and minimized between-category similarity.

The cognitive processes recruited during incidental and intentional category learning may also differ from one another. Brooks (1978) states that incidental learners tend to compute overall similarity of exemplars of a given category, whereas intentional learners look for rules that help define a category. Intentional learners consciously look for commonalities that cluster exemplars of a category together, which drives rule-seeking and rule-testing strategies. Incidental learners, on the other hand, do not partake in these strategies, as there is no awareness of the learning goal. Incidental learners lack the same conscious motivation to find rules that determine groupings. Instead, incidental learners are sensitive to sets of overlapping characteristics among exemplars. Brooks argues that analytical learning mechanisms, which occur during intentional learning, assist in the learning of categories that can be diagnosed by a rule (i.e., rule-based categories) and that non-analytical learning mechanisms, which are at work during incidental learning, serve better during learning of categories that are grouped by correlated features (i.e., family resemblance-based categories).

As evidenced by the way people effectively make sense of and interact with the world, people are highly proficient at learning categories without explicit instructions to do so. Interleaved study of exemplars is believed to be beneficial because of the contrast opportunities aided by temporal juxtaposition. But must a learner be aware of the learning goal in order to make such beneficial between-category contrasts? And can contrast processes be implemented intentionally during analytical and/or non-analytical learning? The purpose of the following two
experiments was to explore whether interleaving, which is beneficial to intentional learning, would also provide benefits to incidental category learning.

**Experiment 3: The effect of interleaving on incidental learning.**

Previous research has demonstrated that interleaved presentation of category exemplars is better for intentional inductive learning than is blocked presentation (Kornell & Bjork, 2008). In the present Experiment 3, I investigated whether interleaved study is also superior, compared to blocked study, for incidental inductive learning. Prior to the study phase, participants were either (a) made aware of the goal to induce categories during study and that there would be a subsequent classification test (i.e., the *intentional learning* condition), or (b) they were not told that the goal was to learn the categories nor that there would be a classification test at the end (i.e., the *incidental learning* condition). Incidental learners were asked to make a judgment on each exemplar viewed during the study phase. This judgment task was employed to encourage participants to attend to and process each image while keeping participants from becoming aware of the categorization task or goal.

A task that keeps participants’ attention on study materials in the absence of an explicit learning goal is called an orienting task (Postman & Adams, 1956). Eagle and Mulliken (1974) found that orienting tasks, specifically those that ask participants to make affective ratings such as pleasantness ratings and likeability ratings, lead to better recall learning as compared to intentional learning without an orienting task. In the present Experiment 3, participants took part in an affective-ratings orienting task in the incidental learning condition. Specifically, participants rated how talented they felt the artist of each painting was. Eagle and Mulliken’s findings suggest that affective-ratings orientation tasks not only effectively get participants to attend to the task, they also get participants to process information more deeply. Thus, in the
present experiment, the orienting task used in the incidental learning condition ensured that participants in both the incidental and intentional learning conditions would process the study materials (exemplar image and category label), allowing an examination of the role of intentionality on between-category discrimination processing while keeping attention constant across both learning conditions.

**Participants.** Participants were 54 females and 52 males (mean age = 32.4), recruited through Amazon Mechanical Turk.

**Materials.** A portion of the materials created by Kornell and Bjork (2008) (ten paintings each by six artists) were used. The artists used were Braque, Juras, Hawkins, Lewis, Mei, and Pessani. Only six artists’ paintings were used for this study because it may not be possible for incidental learners to induce a large number of similar categories.

**Design and Procedure.** Intentionality (intentional learning vs. incidental learning) was manipulated between-subjects and study schedule (blocked vs. interleaved) was manipulated within-subjects. The presentation schedule was BIBIBI where each B stands for six paintings by the same artist presented consecutively (i.e., blocked), and each I stands for six paintings, two by each of three artists) presented in an interleaved manner. The presentations of the six paintings within each B and I were randomized. Each painting with the name of the artist printed below it appeared on a computer screen for 8 s.

The intentional learners (n = 48) were instructed to learn each category such that when a new example of the category was presented, the exemplar could be correctly classified. Participants received the following instruction: “Associate the artist with their painting style. You will be tested with new paintings on your ability to identify who painted it.”
The incidental learners (n = 46) were not told to learn the painters’ styles. Instead, during each trial, they were told to evaluate each painting, an example of which is shown in Figure 4.1, and answer the following question for each painting: “For each trial, please rate from 1 to 7 how talented this artist is based on that particular painting.” This orienting task was meant to keep participants’ attention on each painting and to associate the painting to the artist’s name. The question incorporated the evaluation of the artist and that particular artwork. Use of this particular instruction was to make the incidental-learning task as similar as possible to the intentional-learning task without explicitly asking participants to attend to the artist name and associate it with the painting. For each painting, participants clicked on a scale beneath the painting that spanned from 1 to 7. Although an additional action was required for each trial (clicking on a value from 1 to 7) compared to the intentional learning condition, each trial in both learning conditions lasted for the same amount of time, but durations were increased to 8 s (rather than 5 s as in Kornell & Bjork, 2008) to allow for sufficient processing.
Figure 4.1. Training phase trial for the incidental-learning condition.

After the study phase, participants were tested on four new paintings by each of the six studied artists. Each image was presented with the six artist names below, and participants clicked on their answer on the computer screen. Following the test phase, participants were asked which study schedule they believed helped them perform better on the final test: blocked or interleaved study.

**Results.** Performance on the final classification task for both learning conditions and scheduling conditions is shown in Figure 4.2. As indicated there, intentional learners ($M = .76$, $SD = .2$) performed better on the final test compared to incidental learners ($M = .65$, $SD = .22$), $F(1,92) = 6.9, p = .01$. Collapsed across both learning conditions, interleaved study ($M = .79$, $SD = .21$) led to significantly better inductive learning than did blocked study ($M = .63$, $SD = .28$), $F(1,92) = 40.8, p < .001$. Furthermore, planned comparisons revealed a significant benefit of interleaving in both the incidental-learning condition ($t[47] = 4.09, p < .001$) and the intentional-learning condition ($t[45] = 4.88, p < .001$).

![Figure 4.2](image.png)

*Figure 4.2. Classification test performance for Experiment 3.*
Metacognitive judgments were very similar across the two learning conditions. In the intentional-learning condition, 13 participants believed interleaved study was better, whereas 33 participants believed blocked study was better, and two participants believed that both study schedules were equally effective. In the incidental-learning condition, 13 participants believed interleaved study was better, whereas 33 participants believed blocked study was better.

**Discussion.** Results revealed that intentional learners performed better than incidental learners. According to Kelmer Nelson (1984), incidental learners make categorization judgments based on overall group similarity, whereas intentional learners actively search for rules that define category membership. The categories studied in Experiment 3 were painting styles of different artists, which are more resemblance-based than rule-based, yet incidental learners did not perform better than the intentional learners at classifying the paintings. Although it may seem that Kelmer Nelson’s (1984) position would predict that incidental learners should perform better than intentional learners on the learning of resemblance-based categories, it is important to clarify that Kelmer Nelson’s position suggests that different category features are attended to depending on the intention of the learner. That is, incidental learners will attend more to groupings of features that are highly correlated within categories and will thus do better at learning categorizing family-resemblance based categories than rule-based categories. Intentional learners, on the other hand, will approach the study phase with the intent to find a rule that defines each category and will thus learn rule-based categories better than family-resemblance based categories. Although in the present Experiment 3, incidental and intentional learners are both learning family-resemblance based categories, the question of interest is not which group of learners performs better overall, but rather how each group of learners responds to interleaved and blocked study schedules.
In Experiment 3, the interleaving effect was replicated for the intentional learning condition. Interestingly, however, interleaved study was found to be better than blocked study for incidental learning as well—a finding that has not previously been investigated or demonstrated. Interleaving is thought to benefit intentional category learning because interleaving allows exemplars of different categories to be contrasted. This same reasoning can be used to explain why interleaved practice is better for incidental learning.

Intentional category learners are aware of the importance of comparing exemplars of the same category and extracting commonalities. This awareness can be seen to underlie peoples’ preference to block exemplars as opposed to interleaving them during study (Tauber, Dunlosky, Rawson, Wahlheim, & Jacoby, 2013), presumably because blocking makes noticing within-category commonalities more fluent. And it is true that finding within-category similarities is important for category learning. As demonstrated in Experiments 1a and 1b, increasing spacing between exemplars of the same category improves category learning, suggesting that manipulations that enhance within-category commonalities are important to learning.

Intentional learners, however, do not appear to be aware of the benefits of temporal juxtaposition for allowing contrast processes to occur. When exemplars are interleaved, exemplars of different categories can be contrasted, providing valuable category boundary information. Although intentional learners know that the goal of study is to learn categories, they often do not report interleaving to be better for learning than blocking when metacognitive measures are taken (e.g., Kornell & Bjork, 2008; Birnbaum, et al., 2013). That is, even though intentional learners are aware of the task goal, they seem to be unaware of the benefits of contrast processing. Such lack of awareness on the part of intentional learners can thus help to explain why interleaved study was observed to benefit incidental learning as well as intentional
learning in the present Experiment 3. If even intentional learners are not aware of the benefit of interleaving, the mechanisms that occur during interleaving may be similar in both learning conditions. That is, the between-category discrimination processes that are facilitated by interleaved presentation may occur automatically and without directed effort.

Inductive learning requires processing of differences between categories and of similarities within categories. Ross, Perkins, and Tenpenny (1990) argue that the commonality processes that occur during category learning are explicit (as discussed earlier in Chapter 2: The Role of Spaced Retrieval in Inductive Learning); that is, learners intentionally think back to other examples of the same category and make mindful comparisons. Results from the present Experiment 3, however, suggest that learners do not explicitly or intentionally implement contrast processes. The argument that contrast processes occur outside of awareness and are automatic is in line with previous findings that participants do not believe that interleaving is better than blocking for category learning. Furthermore, that interleaved study is more beneficial than blocked study for both intentional and incidental category learning, as demonstrated in the present Experiment 3, provides further evidence for a dissociation of awareness versus automaticity for commonality and contrast processes respectively.

**Experiment 4: Manipulating directed and non-directed learning within subjects.**

Unintentional category learning occurs in a variety of settings. When a child learns the concept of “dog,” for example, he or she—without explicit instruction—builds a representation based on exemplars of dogs. In other settings, unintentional and non-directed category learning may occur in conjunction with intentional category learning. In the process of learning different eras of art, for example, an art history student may pick up on which artists were prevalent in each era. The information specific to the different artists would be learned by non-directed
learning in this case, because no attention had been directed toward that type of information about the artists. Noticing that each era is comprised of several key artists is essentially the process of extracting subcategories. Subcategories can sometimes assist with learning broader categories because they help to organize information within the different broader categories. If each artist’s work is very distinct and the era is comprised of these artists, it may be beneficial to learn each artist within a category and use that knowledge to then connect to era knowledge.

This question was addressed in the present Experiment 4 by examining the effects of different study schedules for directed learning and the simultaneously occurring non-directed learning of subcategories within the to-be-learned broader categories. Several questions were of interest: When the broader categories of eras must be learned, can subcategory information about key artists within that category be utilized in doing so? And what study schedules will optimize the learning of both the broader categories (directed) and the subcategories (non-directed)?

Unlike the incidental learning condition in Experiment 3, which employed an orienting task to direct attention to the processing of individual paintings and the name of the artist, no instruction was given in the present experiment to orient participants’ attention to the processing of specific artist information.

Participants. Participants were recruited from Amazon Mechanical Turk, a website that employs individuals to partake in small tasks in exchange for money. A total of 76 participants from the United States (30 females and 46 males, mean age = 31.14) participated.

Materials. A total of 72 images of paintings, some of which are illustrated in Figure 4.3, were used. The paintings were categorized into artwork from four eras: Impressionism, Romanticism, Renaissance, and Baroque. Each era (the broad categories) was exemplified by the work three artists, and six paintings by each artist were used. The artists from the Impressionist
group were Cassatt, Renoir, and Sisley; those from the Romanticism group were Constable, Delacroix, and Turner; those from the Renaissance group were Angelico, Botticelli, and Gozzoli; and, those from the Baroque group were Poussin, Rembrandt, and Rubens. Of the six paintings by each of the 12 artists, five were used during the study phase and the sixth painting was used at test.

![Examples of the paintings used in Experiment 4.](image)

*Figure 4.3. Examples of the paintings used in Experiment 4. Those in the top row represent the Baroque era and are paintings, from left to right, by Rubens, Poussin, and Rembrandt, respectively. Those in the bottom row represent the Impressionism era and are paintings by Renoir, Cassatt, and Sisley, respectively.*

**Design and Procedure.** Participants were told that they would be shown various paintings from four eras of art and that their task was to learn the different eras such that at test, when presented with a never-before-seen painting, they would be able to identify in what era it had been painted. Sixty paintings were presented to participants during the study phase at a rate of 5 s each. During each image presentation, the era and the artist name were also shown below the image. The image appeared in the center of the screen with the word “by” and the artist’s
name (e.g. “By Renoir”) appearing directly below the image and the era label appearing just below the artist’s name. Although both artist and era information were presented, the explicit goal was to learn the eras associated with each painting. Learning the artistic style representative of each era was thus the directed and intentional goal, and whether participants would also achieve the non-directed and unintentional learning of each painter’s style, given that the name of the artist as well as the era were presented with each image, was one of the experimental questions.

Paintings were presented in one of four study schedules: separate-blocked, separate-interleaved, mixed-blocked, and mixed-interleaved. In the separate-blocked and separate-interleaved conditions, all paintings from a given era were presented consecutively, then all paintings from another era were presented consecutively, and so on until the paintings from all four eras had been presented. In the separate-blocked condition, all paintings of an era were presented consecutively, and the paintings of the three artists within that era were presented in a blocked manner. For example, all paintings by Rubens (Baroque era) would be presented in a row, followed by all paintings by Poussin (Baroque era), followed by all paintings by Rembrandt (Baroque era), and so on for the other artists of the other eras of art. In the separate-interleaved condition, all paintings of an era were presented consecutively, but the paintings from the three artists representing that era were interleaved among one another. For example, paintings by Rubens, Poussin, and Rembrandt (all Baroque artists) would be interleaved among one another and presented consecutively. This interleaved series would then be followed by the interleaved presentation of paintings from another era, and so on, until all the paintings from each of the four eras had been presented.
In the mixed-blocked condition, paintings of a given artists were blocked while the eras were interleaved. For example, all exemplars from Rubens (Baroque era) would be presented consecutively, followed by all exemplars by Renoir (Impressionism era), and so on. In the mixed-interleaved condition, paintings and eras were interleaved in a blocked-randomized order. Thus, for example, the first block would contain one painting from each of the 12 artists with their presentation order randomly determined; then the second block would contain one painting from each of the 12 artists appearing in a different random order, and so forth until all 60 of the paintings had been presented.

After the study phase, participants were shown one new painting by each artist and asked to identify the era from which the painting came by clicking on one of the four era labels that appeared underneath the test image. Then the same painting was re-presented on the screen, with the names of the 12 artists appearing underneath the image, and participants were asked to identify the artist by clicking on one of these 12 names. Thus, participants were asked first to identify the era and then the name of the artist for each of the 12 test items.

**Results.** Two measures of performance were obtained in the present study: performance on classification of eras and performance on classification of a painting’s artist. Performance on era classification as a function of the four presentation conditions is illustrated in Figure 4.4. As indicated there, a significant main effect of study schedule was observed, $F(3,72) = 5.35, p = .002$. Furthermore, post-hoc comparisons revealed that the mixed-interleaved condition ($M = .73, SD = .14$) led to significantly better performance than that observed in the mixed-blocked condition ($M = .57, SD = .16$) and in the separate-blocked ($M = .6, SD = .19$) conditions, $p = .002$ and $p = .013$, respectively. Additionally, the separate-interleaved ($M = .73, SD = .16$) condition led to significantly better performance than that observed in the mixed-blocked ($M = .57, SD = .16$) condition.
.16) and separate-blocked ($M = .6, SD = .19$) conditions, $p = .003$ and $p = .017$, respectively. The separate-interleaved ($M = .73, SD = .16$) and mixed-interleaved ($M = .73, SD = .14$) conditions did not significantly differ, $p = .983$.

Figure 4.4. Classification performance for eras as a function of presentation conditions in Experiment 4.

The pattern of results for artist classification, shown in Figure 4.5, was quite different from that of era classification. Firstly, performance for era classification was much better ($M = .66, SD = .17$) than that for artist classification ($M = .17, SD = .12$), $t(75) = 22.24, p < .001$. Additionally, the outcome of a 2 x 4 ANOVA with test type (era vs. artist) and presentation schedule (mixed-blocked, mixed-interleaved, separate-blocked, and separate-interleaved) as the independent variables revealed a significant interaction between type of classification and presentation schedule, $F(3,72) = 6.57, p = .001$, indicating that the schedules that were optimal for directed learning were not the best schedules for non-directed learning. Post-hoc t-tests revealed that the separate-blocked presentation schedule seemed to be the best for artist classification, with the separate-blocked schedule ($M = .22, SD = .14$) being marginally better than both the mixed-blocked ($M = .15, SD = .13$) and separate-interleaved ($M = .15, SD = .09$).
conditions, \( p = .065 \) and \( p = .074 \) respectively. The separate-blocked condition (\( M = .22, SD = .14 \)) was not significantly better than the mixed-interleaved condition (\( M = .17, SD = .11 \), \( p = .134 \).

![Proportion Correct Classification](chart.png)

**Figure 4.5.** Classification performance for artists as a function of presentation conditions in Experiment 4.

The artists that were accurately classified in the separate-blocked condition were more closely examined. In the separate-blocked schedule, paintings of each artist were blocked and artists of a given era were presented together in a block. There were therefore three blocks of artists for each era. A chi-square goodness of fit test examined from which block the correctly classified artists came. Of the 96 correct classifications of artists, 23 were the first artist presented for an era, 27 were the second artist, and 46 were the third artist. Results indicate that most correct classifications of artists occurred for those artists whose paintings were presented in the third block of paintings within a given era (\( \chi^2 = 9.438, p = .009 \)).

**Discussion.** For the explicit goal of learning art eras, interleaving the exemplars by artist was most beneficial. Based on previous research, the exemplars of the to-be-learned categories
should be interleaved in order to enhance inductive learning (e.g., Kornell & Bjork, 2008; Kang & Pashler, 2012). That is, it is the interleaving of eras that should have led to improved inductive learning of the eras. The only study condition that interleaved presentations by era was the mixed-interleaved schedule. Although all paintings from a given era were not presented consecutively in the mixed-blocked condition, paintings did alternate on each trial by era. In the mixed-blocked condition, the exemplars of each artist were blocked, which meant that five paintings from the same era were presented consecutively. Given that only the mixed-interleaved condition interleaved by era, the mixed-interleaved condition should have led to the best inductive learning of the eras. Yet, the results revealed that, not only did the mixed-interleaved condition yield good performance, the separate-interleaved condition led to equally good performance.

What was good about the separate-interleaved study schedule for era induction? If the study schedule of era presentation was the only schedule feature that should have had an effect on the classification of eras, then the separate-interleaved and separate-blocked conditions should have led to similar performance. The manner in which the artists within each era were presented, however, had an effect on the inductive learning of eras.

A possible explanation for the differences in the separate-blocked and separate-interleaved conditions is that the nature of the materials had an effect on the processes involved in efficient learning. The era categories have high within-category variability: they are made up of subcategories of artists. And, although artists from the same era are more similar to one another than are artists across different eras, the fact that eras are made up of subcategories means that there is more variability among paintings within each era. According to Goldstone (1996), blocking helps learners notice within-category correlations and is especially helpful for
learning categories with large within-category variability. In the case of the present experiment, because there was large within-category variability within each era, blocking by era helped learners span the variability of artists within each era and experience the various characteristics of paintings within each era.

The era categories have large within-category variability and benefitted from separated presentation of exemplars of the same era. But why was interleaving the artists within each separated era beneficial to inductive learning? Again, separated presentation of eras allows the participant to sample the variability of artists within an era; furthermore, an interleaved presentation of artists within the era allows learners to sample the variability of artists within the category most efficiently. As discussed in the General Introduction, blocked presentation may lead to confirmation bias: Learners may assume they have successfully learned the features that make up a category. When the paintings of a given artists were presented in a row within a separated era, learners may have begun creating a category representation of that era driven by the painting style of the initially presented artist. By interleaving the artists, learners were given the opportunity to realize how different each artist within an era is and notice commonalities among those artists. By interleaving the artists within each era, learners were able to create a cohesive representation of the artists within each era.

For the unintentional and non-directed process of inducing artist style, a different pattern of results emerged. The separate-blocked schedule provided the largest benefit for classifying artists. In the case of information that is peripheral to the explicit task, such as the names of the artists in the present study, conditions that make names of the artists as salient as possible and as easy to attend to as possible are likely to lead to the best induction of artist information. In the separate-blocked condition, as compared to the other three conditions, each consecutive painting
was as predictable as possible. Not only were all exemplars of a given era presented consecutively, all exemplars of a given artist were presented consecutively as well. In the other study conditions, artist names alternated on each trial, most likely making the indirect learning of artist names more challenging. The successive study trials in the separate-blocked condition changed minimally and may have thus led to a sense of fluency of processing and ease of learning.

The increased fluency of painting-era associations may have led to a decrease in attention paid to era information, with this reduction in attention paid to era information possibly freeing up attention to be directed toward artist information instead. This possibility may explain the better classification observed for the third artist presented within an era. By the third artist within an era, attention to era information may have gradually waned, resulting in greater attention paid to artist information. The less attention paid to learning the era information meant that more attention would be paid to processing artist information. Thus the poor performance for era and enhanced performance for artist in the separate-blocked condition may reflect a trade off in attention.

Lastly, it is important to note that the poor accuracy for artist classification may not be a result of a lack of inductive learning. Participants could have noticed the artist subcategories and induced the painting style of each artist, but performed poorly on a classification test owing to an inability to associate an artist style with the artist name. Participants were never told that they would be tested on artists, so it would make sense for associative learning of the artist’s name and painting style to be poor. Although the separate-blocked schedule condition led to best performance for artist classification, it is not certain whether this condition led to enhanced learning of artist styles or enhanced learning of the artist name-painting style association.
In conclusion, although results from Experiment 3 demonstrated that interleaved study schedule is beneficial for both intentional and incidental category learning, when both incidental and intentional category learning occurred together, greater attention was required to attend to both categories and subcategories. In cases where learners must engage in intentional and non-directed learning in the same setting, study conditions that help direct attention to secondary information will likely lead to most effective non-directed inductive learning.
CHAPTER 5 – TRANSFER IN INDUCTIVE LEARNING

Transfer of learning is the application of learning in one context to the learning and/or performance in another similar context. Lee (2012) found that interleaved study proved to be beneficial not only for direct tests of studied material, but also for transfer tasks in which learned information is applied to a related task. Lee suggests that interleaved practice creates varying conditions of processing that lead to abstraction and ultimately effective transfer of learned information.

Lee (2012) discusses three transfer tasks that have benefited from previous interleaved study. One such task was the making of badminton serves with the left hand after practicing a variety of serves with the right hand (Goode & Magill, 1986); another was the learning of more complex motor sequences following interleaved practice on three similar, but simpler, motor sequences (Shea & Morgan, 1979); and, finally, the third transfer task was that used by Kornell and Bjork (2008) in which participants classified new paintings as those of one artist from a set of artists based on prior study of other exemplars of their different painting styles.

Across all three transfer tasks that Lee (2012) describes, each transfer task required an application of some learned information or skill to a new task that was in some way related, but different from that learned in the initial study trials. What differed across the transfer tasks was not only whether the task was a motor or cognitive task, but also the degree to which the transfer task differed from the initially learned task. The term ‘transfer task’ thus refers to a broad set of possible tasks, and leads us to question how different a transfer task can be from the initially learned task yet still realize a benefit from the initially learned task. For inductive category learning, Lee describes the categorization test commonly used in inductive learning experiments as the application of previously learned exemplar knowledge to the classification of new
exemplars from studied categories. Perhaps further transfer can occur in inductive learning beyond the exemplar classification of studied categories and, possibly, even to the learning of entirely new categories. Testing a variety of transfer tasks can reveal the kinds of information and strategies that participants may be learning during category exemplar study under different study schedules.

**Transfer of the Benefits of Interleaving to New Inductive Learning**

Experiences with learning can be seen as having two facets: learning and learning to learn. An undergraduate student, for example, accumulates knowledge from coursework throughout the years. Beyond the specific course material being learned, however, he or she is presumably also becoming a better student over the years. The student becomes more efficient at taking notes in class, studying for exams, and taking exams. The student therefore not only learns the actual information taught in classes, but also develops strategies to accomplish academic goals.

Transfer tests can probe what kinds of information and strategies learners are extracting from the initial learning. That is, performance on transfer tasks reveal the kinds of strategies learned during initial training that can be effectively transferred to a new context. The greater goal of research on transfer in inductive learning is to discover learning strategies that can be effectively transferred to near transfer tasks, such as classification of new exemplars of the studied categories, as well as far transfer tasks, and to educate students on which study strategies can be implemented in order to benefit new learning opportunities.

Hiew (1977) examined the effects of interleaved and blocked presentations on rule-based category learning, specifically participants’ ability to figure out the defining rules for category membership. After the study phase, participants first reported what they believed were the
defining rules for each category. Participants then performed a transfer task, which was to
discover a category-defining rule for a new category. Hiew found that the interleaved condition
led to both better rule-discovery and better performance on the transfer task. Hiew reasoned that
the interleaving of exemplars created greater contiguity of instances demonstrating the
overarching nature of the rules for category membership, which led to better acquisition of a
rule-finding strategy. Noticing the nature of all of the rules functioned as a strategy that could
then be applied to discovering other rules. In contrast to the transfer task of applying category
knowledge to the correct classification of never-before-seen exemplars of the learned categories
(as done by Kornell & Bjork, 2008), Hiew’s transfer task required the application of a strategy to
new opportunities for learning.

Just as Hiew’s (1977) study demonstrated that enhanced performance on a transfer task
can reveal the type of skills that were acquired during initial study, performance on a variety of
transfer tasks can inform us about what information participants are gaining from the study phase
and how that information is used in future learning. Experiments 5, 6a, and 6b of the present
dissertation were designed to assess study strategies that benefit inductive learning, whether
learners can become better aware of those strategies, and, even if they do, whether learners have
the control to implement such strategies during learning in new contexts.

**Experiment 5: Applying the “interleaving strategy” to new learning.**

In Experiment 5, participants were evaluated on their ability to learn more efficiently
after they had experienced the benefits of interleaving in an initial learning task. Much like the
way a student may discover that quizzing one’s self is effective for learning and, then, uses that
strategy to study for other exams, learners that experience the benefits of contrasting exemplars
of different categories may be able to apply that process to subsequent category learning.
Participants. A total of 24 females and 18 males (mean age = 35.4), who were recruited via Amazon Mechanical Turk, participated in Experiment 5.

Materials. The materials were 80 photographs, five for each of 16 species of butterflies. The species names employed were Admiral, American, Baltimore, Cooper, Tiger, Streak, Harvester, Mark, Lady, Elfin, Pipevine, Sprite, Tipper, Satyr, Viceroy, and Nymph. Eight of the butterfly species were used in Part 1 of the experiment, and the remaining eight butterfly species were used in Part 2.

Design and Procedure. The study consisted of two parts, with each part consisting of two phases: a study phase and a test phase. In Part 1, participants studied four exemplars from each of eight butterfly species. Half of the species’ exemplars were presented in a blocked schedule and the other half of the species’ exemplars were presented in an interleaved schedule. The schedule was as follows for the Part 1 study phase: BIIBBIIB, with each B standing for a blocked series of four exemplars for a given butterfly species and each I standing for an interleaved series of four exemplars (one exemplar from each of the four interleaved species). Each butterfly image plus its species label appeared on the screen for 5 s. After the study phase, participants were tested on their ability to classify new exemplars of the studied butterfly species. One new exemplar for each of the eight species was tested. Each test image appeared on the computer screen with the names of the eight studied species appearing below the image. For each test trial, the participant clicked on one of the species labels. Test trials were not timed and no corrective feedback was given.

After the Part 1 test phase was complete, participants read the following: “Thank you for partaking in the test. You may have noticed that during the study phase, some species' examples were all presented in a row and other species' examples were intermixed with other species. Your
test results indicate that you did better for species that were presented to you in an intermixed fashion. This is in line with how most other people perform. Intermixed study is better for learning categories because it gives the observer the opportunity to compare and contrast the categories. Observing all examples in a row makes this process more difficult, albeit not impossible. Part 2 is identical to Part 1, with the exception that you will be learning a different set of 8 butterfly species. Examples of half of the species will be presented consecutively and examples of the other half will be intermixed. Please try to apply the same "compare and contrast" strategy to the species examples that are presented in a row in order to optimize your performance. Are you ready to begin?"

The instructions told participants, regardless of how they had actually performed on the test for Part 1, that they had performed better on the classification task for the species studied in an interleaved manner. Participants were also told how interleaving benefits category learning and why it is more effective than blocking. Finally, participants were explicitly told to try to apply the same “compare and contrast strategy” to new learning. After these instructions were given, the study phase of Part 2 began.

The study phase of Part 2 was similar to that of Part 1, with the exception that a different set of eight butterfly species were studied. Additionally, the study order was different. Instead of BIIBBIIB, the study order used was IIBBIIBB. This change was made in order to expose participants to an interleaved series of exemplars first in the study phase. Exposing participants to exemplars of different categories from the beginning should increase participants’ ability to begin contrasting between categories. If the study phase began again with a blocked series of exemplars, those exemplars would not be readily comparable to any other species’ exemplars because no other categories would have been viewed at that point. By the time the first blocked
species appeared, participants would have viewed exemplars of four other species and would be better able to contrast exemplars of different species.

After the Part-2 study phase, participants engaged in the Part-2 test phase. The test phase was identical to that of Part 1, except that the eight species learned in Part 2 were tested. After the test was complete, participants were asked, “Did you feel that you were able to compare and contrast across all categories regardless of how the presentation was ordered?” Participants typed in their response in a textbox on the computer screen.

Results. Correct classification performance for Part 1 and Part 2 and as a function of interleaved versus blocked study schedules is shown in Figure 5.1. As indicated there, overall test performance (collapsed across blocked and interleaved study schedules) neither significantly improved nor declined from Part 1 \((M = .57, SD = .21)\) to Part 2 \((M = .52, SD = .22)\), \(F(1,40) = 1.76, p = .193\). A significant main effect of study schedule was obtained, with the benefit of interleaved study being marginally significant for Part 1, \(t(40) = 1.337, p = .189\), and for Part 2, \(t(40) = 1.622, p = .113\). Collapsed across Part 1 and Part 2, interleaved study \((M = .58, SD = .22)\) led to significantly better performance than blocked study \((M = .5, SD = .18)\), \(F(1,40) = 5.34, p = .026\).
Figure 5.1. Classification test performance for Part 1 and Part 2 as a function of blocked and interleaved study schedules in Experiment 5.

When asked whether they had been able to use the compare and contrast strategy to learn in Part 2, 25 participants stated that they had been able to do so, 11 participants stated that they were unable to do so, and the remaining six participants gave unclear responses that did not address or answer the question posed, and thus these latter responses were excluded from further analysis. A nonparametric binomial test revealed that there were significantly more people who stated that they had been able to employ the contrast strategy during Part-2 study than people who said they had not been able to do so ($z = 2.33, p = 0.04$).

**Discussion.** Participants were given the opportunity in Part 1 to experience both blocked and interleaved presentation schedules during the study phase and to experience classification performance for blocked versus interleaved species during the test phase. Furthermore, following Part 1, participants were told that they had performed better on the classification of interleaved
species, received instruction for why interleaving is better, and were urged to use the contrast strategy that interleaving facilitates during Part-2 study.

If participants were successfully able to make contrasts between exemplars of different categories, even when exemplars within categories were presented in a blocked order, performance for both the blocked and interleaved conditions in Part 2 should have been similar to one another. Despite the experimental design and instructions to participants, however, performance in Part 2 did not reflect a transfer of the contrast strategy. Rather, the performance observed in Part 2 indicated that participants were unable to make contrasts between exemplars of different species when those species’ exemplars were blocked. Thus, the overall pattern of results suggests that either interleaving is not beneficial for the supposed reason that it allows for between-category contrasts or that participants are not able actively to employ contrast processes beyond what is made possible by an interleaved presentation scheduling. The benefits of interleaving may thus occur only as a result of preset interleaved scheduling and not by an explicit strategy to contrast exemplars. (See appendix for additional experiments conducted that revealed a similar lack of transfer of interleaving benefits.)

Thus, ostensibly, the present pattern of results would seem not to agree with those of Hiew’s (1990), which showed that the benefit of interleaving could be transferred to new learning. They are not, however, necessarily at odds with Hiew’s findings in that his results were based on categorization learning of rule-based concepts. In order to develop rules, explicit problem solving must be carried out. Hiew’s task relied heavily on reminding and keeping track of characteristics in order to formulate rules. When it came time to apply the same strategy to a new learning opportunity, participants were able to use the explicit techniques for rule finding that were practiced in the initial learning session.
The categories used in the present study, however, were not based on rules and could not be solved via explicit reasoning. Learning natural categories like butterflies is instead an “information-integration” task, a task that requires responses based on an entire collection of perceptual features (Ashby & Gott, 1988). According to the competition between verbal and implicit system account of category learning (COVIS), categories that can be defined by a verbalizable rule are mediated by an explicit system and categories that are information-integration based are mediated by an implicit system (Ashby, Alfonso-Reese, Turken, & Waldron, 1998). Knowlton and Squire (1993) found that amnesic patients with damage to the limbic or diencephalic structures of the brain were still able to acquire category knowledge. Additionally, Knowlton, Mangels, and Squire (1996) demonstrated a compelling double dissociation between the limbic-diencephalic regions damaged in amnesics and the neostriatum damaged in Parkinson's patients. The amnesics could acquire category information (in a probabilistic classification task) but were unable to recall declarative memories, and the Parkinson’s patients were able to recall declarative memories but were unable to acquire category information.

Just as probabilistic category learning and explicit memory tasks are mediated by different brain regions, Ashby and Ell (2001) argue that information-integration and rule-based category learning are mediated by different brain structures. According to neuroimaging data, information-integration category learning may be mediated by the tail of the caudate nucleus, while rule-based category learning may be mediated by the prefrontal cortex, anterior cingulate cortex, and head of the caudate nucleus.

The nature of the studied categories may thus explain the differences between the results obtained in the present Experiment 5 and those obtained by Hiew (1990). The categories studied
in Experiment 5 were more information-integration based whereas Hiew’s categories were rule-based, which could thus account for the difference in how category information was processed. I argue earlier that the benefits of interleaving occur outside of awareness—an observation consistent with the results of the present Experiment 3 demonstrating that contrast processes occur even when category learning is incidental. In Experiment 5, participants were explicitly told about the benefits of contrasting between categories and were told to apply this contrast strategy to the learning of the new set of categories. Even given this information and instructions, however, participants were largely unsuccessful at implementing the contrast strategy in their new learning, suggesting that contrast processes occur automatically, and efforts to engage contrast processes explicitly are not effective.

As indicated by the present pattern of results, the best way to ensure the occurrence of contrasts processing in the learning of natural categories, such as butterflies, is to learn under conditions that facilitate the automatic occurrence of such processing, such as appears to happen in interleaving study. That is, if you need to induce natural categories, although it may be nice to know why interleaving is beneficial, it is better for your learning if you study the category exemplars in an interleaved manner.

**The Testing Effect, the Generation Effect, and Feedback Training**

The *testing effect* refers to the finding that testing oneself on information leads to better learning and later recall of that information than does simply restudying the information (Roediger & Karpicke, 2006). The testing effect has also been demonstrated in skill learning (Kromann, Jensen, & Ringsted, 2009). Medical students who participated in a 4-hour in-hospital resuscitation course performed better after a lesson that reserved 30 minutes for case study scenario-testing compared to students who continued with study during the last 30 minutes. A
similar phenomenon called the *generation effect* is the phenomenon that stimuli that are generated as prompted by a cue are better recalled than stimuli that are simply observed (Jacoby, 1978; Slameka & Graf, 1978).

Karpicke and Zaromb (2010) state that the main difference between testing and generation is the intentionality of retrieval. With testing, the production of a response is a directed retrieval of a previously studied item. With generation, the produced answer is an incidental retrieval – meaning that whatever answer comes to mind and fits the cue is an appropriate generation response. Karpicke and Zaromb found that while both testing and generation lead to better recall than reading an item, testing leads to better later recall than generation. Despite this difference, testing and generation are similar in that effortful retrieval processes (be it intentional or incidental) enhance encoding and also provide study conditions that are more similar to the test format.

Testing and generation can also be incorporated into inductive learning with the use of *feedback training*. In a trial of feedback training, an exemplar is presented with no category label and the participant attempts to classify the exemplar correctly. Then, the generation (attempt) is followed by presentation of the correct category label. In contrast, in a trial of *observational training*, which is what is typically used in experiments of inductive learning, presentations of the exemplars are accompanied by the correct category labels. Studies by Jacoby, Wahlheim, and Coane (2010) and Ashby, Maddox, and Bohil (2002) demonstrate that feedback training is more beneficial than observational training for information-integration concept learning.

Jacoby, Wahlheim, and Coane (2010) had participants learn eight bird families. Exemplars of half of the bird families were studied via feedback training and the other half were studied via observational training. Participants studied five exemplars for each bird family. For
both feedback training and observational training conditions, all exemplars along with label name were first presented one at a time in a randomized order (thus, these trials were essentially observational trials). In the observational training condition, each exemplar was presented again three additional times in a block-randomized order (SSSS). For the feedback training condition, exemplars were presented without category label followed by feedback, three times each, in a block-randomized order (ST\textsuperscript{S}T\textsuperscript{S}T\textsuperscript{S}). Because exemplars were repeated and participants were told to select the category label that appeared with the exemplar, the feedback training was similar to a test with feedback (rather than a generation). Results demonstrated that bird families that were studied via feedback training were better learned than were families studied via observational training.

Ashby, Maddox, and Bohil (2002) also demonstrated a benefit of feedback training for the learning of information-integration categories. Their feedback training procedure differed from that of Jacoby, Wahlheim, and Coane’s (2010) in that their trial items were never repeated. Additionally, feedback training began on the very first trial. That is, participants never viewed observational trials. The feedback training was thus more similar to a generation task (rather than a test trial). Ashby et al. were also interested in how feedback training affects both rule-based category learning and information-integration category learning, and they found that—whereas the learning of rule-based categories did not benefit from feedback training—the learning of information-integration categories did benefit from feedback training. Based on this pattern of results, Ashby et al. argued that feedback training is a form of dopamine-mediated reward-based learning because when a learner produces an answer and it turns out to be correct, there is a spike in dopamine release. Furthermore, because procedural learning is mediated by this neural system (Ashby et al., 2002) and the learning of information-integration categories is believed to be
mediated by the same system (Ashby & Ell, 2001), information-integration category learning benefits more from feedback training than does rule-based category learning.

**Transfer of the Benefits of the Testing Experience to New Inductive Learning**

A strategy can only benefit later learning situations if that strategy can be actively applied by the learner. Testing one’s self, as opposed to making between-category contrasts, may be a strategy that participants can actively employ during new learning. One could, for example, presumably “cover up” the correct answer and attempt to retrieve the correct response before allowing one’s self to look at the correct answer. The testing strategy can be self-directed by learners and can likely be transferred to new learning opportunities.

DeWinstanley and Bjork (2004) demonstrated that participants can become aware of the generation effect and then transfer the generation strategy successfully to new learning. In deWinstanley and Bjork’s Experiment 1a, participants studied a text passage that contained critical read items and critical generate items, and then they were given a fill-in-the-blank test on the critical read and generate items. Thus, after the first study phase and test, participants had experienced the effects of both read and generate strategies on their test performance. Participants then received a second text passage to study, one again containing both critical read and critical generate items, and then and took a test on the materials from the second passage. Interestingly, a generation effect, which had appeared in the first test, did not occur in the second test—presumably because participants were able to employ the “generate” strategy to both read and generate items when studying the second passage. DeWinstanley and Bjork also found evidence that learners became aware of the generation effect. In deWinstanley and Bjork’s Experiment 1b, participants were asked after the first passage and test, “what did you notice about your performance on the previous memory test?” Seventeen out of 31 participants reported
noticing that generating target items led to better recall. Thus, a majority of their participants became aware of the connection between improved test performance and generating items. Furthermore, as indicated by the lack of a generation advantage on the second test, participants were able to implement the generation strategy when learning a new passage.

The purpose of Experiments 6a and 6b of the present dissertation was to examine the effect of feedback training on inductive learning, to determine whether participants could recognize the superiority of the feedback training/testing strategy, and whether the benefits of feedback training would transfer to new learning. In Experiment 6a, training style (feedback vs. observational) was manipulated between-subjects, as was done in deWinstanley and Bjork’s Experiment 2, for which no significant transfer of the generation strategy was observed. The purpose for this manipulation was to conduct a similar experiment so that the findings could be compared across studies. In Experiment 6b, training style was manipulated within-subjects, as was done in deWinstanley and Bjork’s Experiments 1a and 1b, so that the present results could be compared to those studies in which participants had been able to transfer the generation strategy to new learning.

**Experiment 6a: The effect of feedback training on new inductive learning.**

Experiment 6a was designed to investigate the effects of feedback training versus observational training on subsequent inductive learning.

**Participants.** Participants were 76 individuals, all from the United States of America, recruited through Amazon Mechanical Turk. Age and gender were not recorded.

**Materials.** Materials were nine handwriting samples, illustrated in Figure 5.2, written by each of six right-handed undergraduate females between the ages of 18 and 19, henceforth referred to as the authors of the handwriting samples, and the same nine phrases were written by
each author. The phrases were two to three words long and were of neutral content, such as “At the top,” and “Beside the point.” Rather than using author names as the labels for the different categories, the numbers “1,” “2,” “3,” “4,” “5,” and “6” were used to designate each author. Reasoning for this labeling is further explained in the Design and Procedure of Experiment 6a.

Figure 5.2. Examples of the handwriting samples used in Experiment 6a by two authors, with those in the top row all from one author and those in the bottom row all from a different author.

**Design and Procedure.** The study consisted of two parts, and each part consisted of a study phase and a test phase. Handwriting samples of three of the authors were used in Part 1 and the handwriting samples of the remaining three authors were used in Part 2. Type of training was manipulated between subjects for both Part 1 and Part 2. That is, participants studied all exemplars of Part 1 by either observational training or feedback training and all exemplars of Part 2 by either observational training or feedback training. There were thus four between subjects conditions: Observational – Observational (n = 18), Feedback – Feedback (n = 20), Observation – Feedback (n = 19), and Feedback – Observational (n = 19), with the first training style signifying the training style used for Part 1 and the second training style signifying the
training style used for Part 2. Conditions henceforth are referred to as O-O, F-F, O-F, and F-O, respectively.

During the study phase of Part 1, six handwriting samples written by each of three authors were used. Presentation order was similar to that used by Jacoby, Wahlheim, and Coane (2010). In the observational training condition, six samples by each of three authors were presented four times each in a block-randomized order, for a total of 72 presentations. The author label appeared below the image of the handwriting sample, and each presentation was 5 s. In the feedback training condition, each exemplar and label pair was presented once for 5 s. Then each exemplar appeared again 3 additional times, in a block-randomized order as feedback training trials. In a feedback training trial, the handwriting sample, without the label, appeared on the screen for 3 s. During this time, the participant was to type in their response in a text box on the screen below the handwriting sample. After 3 s, the handwriting sample remained on the screen and the correct answer appeared below for 2 s.

For Part 1, regardless of which three categories were randomly selected to be used during the study phase, the categories were labeled “1,” “2,” and “3.” Simple number labels as opposed to real names were used in order to make the response task during feedback training physically easy. All three keys are next to each other on the keyboard and could be entered quickly without interrupting or diverting visual attention away from the screen during the study phase.

After the study phase, three new handwriting samples for each of the three studied categories were tested. For each test trial, participants typed in whether the sample was written by “1,” “2,” or “3.” No feedback was given for test trials, and test trials were not timed.

Participants were then told that the same study and test phases would occur with three new categories of handwriting. Participants did not receive any special instructions about the
benefits of feedback training. The study phase and test phase were the same as Part 1 with the exception that three new categories were used and new labels ("4," "5," and "6") were used.

**Results.** Classification performance, collapsed across Part 1 and Part 2, is shown in Figure 5.3 for the feedback training and observational training conditions. As indicated there, feedback training ($M = .69, SD = .19$) led to significantly better overall classification performance than did observational training ($M = .62, SD = .21$), $t(150) = 2.03, p = .044$, $d = .33$.

![Figure 5.3](image.png)

*Figure 5.3.* Classification performance following feedback and observational training collapsed across Part 1 and Part 2 in Experiment 6a.

The correct classification performance, collapsed across Part 1 and Part 2, obtained for each of the combined training procedures is shown in Figure 5.4. Collapsing across Part-1 and Part-2 test performance, there was a main effect of condition, $F(3, 72) = 2.971, p = .037$. Additional analyses revealed that performance in the F-F training condition ($M = .73, SD = .18$) was significantly better than that in the O-F training condition ($M = .6, SD = .14$), $p = .015$, and also that in the O-O training conditions ($M = .59, SD = .18$), $p = .014$. Participants in conditions whose Part-1 study was feedback training (i.e., F-F and F-O) performed significantly better.
overall ($M = .7, SD = .16$) than did the participants in conditions whose Part-1 study was observational training (i.e., O-F and O-O; $M = .6, SD = .15$), $t(74) = 2.78, p = .007, d = .65$.

![Figure 5.4](image-url)

**Figure 5.4.** Averaged classification performance across Part 1 and Part 2 for the four training conditions in Experiment 6a, where F signifies feedback training and O signifies observational training, and the left and right letters signify the type of training used in Part 1 and Part 2, respectively, during the study phase.

During feedback training, responses to the test/feedback trials were recorded and scored. Classification performance improved across the three test/feedback cycles, as indicated in Figure 5.5 in which performance across the three cycles for different training conditions is shown. Part-1 feedback training performance was averaged across the F-F and F-O conditions, and there was a significant main effect of performance by test/feedback cycle, $F(2,76) = 12.7, p < .01$, with performance significantly increasing from the first ($M = .65, SD = .17$) to the second cycle ($M = .76, SD = .16$), $t(38) = 3.9, p < .01$.

Performance also significantly improved across test/feedback training cycles during the study phase of Part 2 in the F-F condition [$F(2,38) = 10.14, p < .01$] and in the O-F condition [$F(2,36) = 5.02, p = .012$].
Figure 5.5. Performance during feedback training in Experiment 6a. Left panel shows performance during the Part-1 study phase for the Feedback-Feedback and Feedback-Observational conditions combined. Middle panel shows performance during the Part-2 study phase of the Feedback-Feedback condition. Right panel shows performance during the Part-2 study phase of the Observational-Feedback condition.

Using a 2 x 4 analysis of variance, performance on the Part-1 test and the Part-2 test for all four conditions, shown in Figure 5.6, was examined. Although an interaction between test part and conditions was not significant, significant differences between groups were observed. The difference in performance on the Part-2 test for the F-F condition ($M = .74, SD = .19$) and that for the O-O condition ($M = .56, SD = .22$) was significant, $p = .004$. Part-2 test performance significantly benefited from having studied in Part 1 via feedback training (collapsed across Part-2 test performance of F-F and F-O; $M = .7, SD = .19$) as opposed to observational training (collapsed across Part-2 test performance of O-F and O-O; $M = .58, SD = .19$), $t(74) = 2.89, p = .005, d = .67$. Specifically, Part-2 test performance for the F-O condition ($M = .66, SD = .19$) was marginally better than was Part-2 test performance for the O-O condition ($M = .56, SD = .22$), $p = .12$. And, Part-2 test performance for the F-F condition ($M = .74, SD = .19$) was significantly better than Part-2 test performance for the O-F condition ($M = .59, SD = .166$), $p = .015$. 74
Figure 5.6. Classification test performance Part 1 and Part 2 for the Feedback-Feedback, Feedback-Observational, Observational-Feedback, and Observational-Observational conditions of Experiment 6a.

**Discussion.** Participants that learned initially with feedback training tended to do better overall than participants that learned initially with observational training. Although initial feedback training did not lead to improvements in subsequent learning contexts, it seemed to keep participants’ performance at a comparably high level, regardless of whether subsequent learning was done with observational training or feedback training. Participants who learned with observational training in Part 1 performed worse on the Part-1 test than did participants who leaned with feedback training and, furthermore, their performance from subsequent learning opportunities did not improve from that point. Even though the training method was identical between the F-O and O-O conditions during Part 2, having studied with feedback training first (as in F-O) was marginally better than having studied with observational training first (as in O-O). This 10% difference in performance suggests that participants are gaining something from experiencing the feedback training condition first.
Furthermore, even though Part-2 training was identical for the F-F and O-F conditions, Part-2 test performance for F-F was significantly better than Part-2 test performance for O-F, indicating that learning via feedback training in Part 1 led to better performance at Part-2 test. The results thus indicate that initially studying exemplars via feedback training leads to better subsequent learning in a transfer task, regardless of whether the subsequent learning occurs via observational or feedback training.

**Experiment 6b: Transfer of the “testing strategy” to new inductive learning.**

Experiment 6a revealed benefits of feedback training on both a classification test for the studied categories and on new inductive learning. Experiment 6a, however, was not designed to probe participants on whether a testing strategy was actively applied during observational learning. To determine more fully the conditions under which participants became aware of the benefits of testing, feedback training and observational training are manipulated within-subjects in Experiment 6b, as in deWinstanley and Bjork’s (2001) Experiment 1a.

**Participants.** Thirty-four participants from the United States, all fluent in English (20 females and 14 males, mean age = 33.68) recruited through Amazon Mechanical Turk participated in Experiment 6a.

**Materials.** Materials were identical to those used in Experiment 6a, with the exception that handwriting samples by eight authors, rather than six, were used. Thus a total of 72 different handwriting samples were used.

**Design and Procedure.** The procedure used in Experiment 6b was very similar to that of Experiment 6a, with the exception that four authors’ handwriting was studied for each of Part 1 and Part 2. Additionally, training style was manipulated within-subjects. Rather than four
between-category conditions, all participants experienced both feedback and observation training in both Part 1 and Part 2.

Furthermore, for a half of the participants (n = 17), no information after the Part-1 test about the effectiveness of feedback training was given (control condition), whereas the other half of the participants (n = 17) received information after the Part-1 test (information condition). Specifically, participants in the information condition were told that they had performed better on exemplar classification for categories studied with feedback training than for categories studied via observational training. These participants received this instruction regardless of how they had performed on the Part-1 test. The explanation was as follows: “Thank you for completing phase 1. According to your performance in the previous test, you were better able to identify handwriting by people whose samples were presented as test/feedback trials. Your results are in line with others who perform well on a variety of memory tests. Learners who tend to generate an answer before allowing themselves to see the correct answer learn more effectively and retain information better.” This information was intended to convince participants that feedback training is more effective than observational training. The explanation for why feedback training is effective and how self-testing is possible during observational training was provided with the intent of influencing participants to apply the study strategy to new learning.

In Part 2, participants in both conditions studied four categories, two via feedback training and two via observational training, followed by a classification test for all four categories. Participants in the control condition were then asked which study strategy they felt was more effective: feedback training or observational training or neither/same. Participants in the information condition were not asked this question because they had already been told that feedback training is more effective. All participants were then asked: “In the second phase of the
study, did you 'test yourself' on the handwriting samples whether or not it was a test/feedback trial? In other words, even if the label was presented with the handwriting sample, did you try to produce the answer before looking at the answer?’”

**Results.** Test performance in Part 1, averaged across the information and control conditions, is shown in Figure 5.7. Although not significant, feedback training ($M = .53, SD = .25$) was numerically more beneficial in Part 1 than was observational training ($M = .46, SD = .28$), $t(33) = 1.26, p = .22$.

![Figure 5.7](image)

*Figure 5.7. Experiment 6b Part-1 classification test performance for feedback training and observational training conditions, averaged across control and information conditions.*

Classification performance for the control and information conditions on feedback training and observational training trials on the Part-1 and Part-2 tests are shown in Figure 5.8. No significant main effects or interactions were observed across the eight experimental conditions.
In Part 1, feedback training was numerically better than observational training for both control and information conditions, but the differences were not significant. Between the Part 2 tests and across control and information conditions, no significant differences or interactions emerged. For the control condition, the two variables (Part 1 vs. Part 2 and feedback vs. observational) did not interact significantly ($p = .12$).

Response accuracy across the test/feedback cycles with feedback training is shown in Figure 5.9. Feedback training performance generally improved across test blocks. Feedback training performance was collapsed across the control and information condition because there were no differences between the two conditions on the Part-1 test (before instruction was given to participants). Response accuracy went up across test/feedback cycles in Part 1 of the control condition, $F(2,76) = 16.88, p < .001$. The same was the case for Part-2 feedback training performance for the control condition [$F(2,16) = 6.746, p = .008$] and the information condition [$F(2,19) = 6.057, p = .009$].
Figure 5.9. Experiment 6b classification accuracy during feedback training for Part 1, for Part 2 of control condition, and for Part 2 of information condition.

The judgments of participants in the control condition, all of whom were asked at the end of the Part-2 test which type of training condition they believed to be better for learning category membership, are illustrated in Figure 5.10 along with their actual performance. One out of 17 participants reported thinking that observational training was best, 11 said that feedback training was best, and five said that each study method provided the same benefit to learning.
The participants in both the control condition and the information condition were asked whether they felt they found themselves testing themselves on the observational learning trials during the study phase of Part 2. Fourteen of the 17 participants in the control condition said that they did, in fact, test themselves on observational trials as well as feedback trials. Note that the participants in the control condition were never instructed to do so and were never told that generating answers and testing oneself can be beneficial to learning. In the information condition, all 17 participants stated that they tested themselves on observational trials as well as feedback trials.

**Discussion.** The proposed questions and potential conclusions to be drawn from the results of Experiment 6b relied heavily upon the previous finding that feedback training was superior to observational training for category learning. This superiority, however, did not arise in the study phase of Part 1, thus limiting the conclusions that could be drawn. A possible explanation for why the benefit of feedback training in Part 1 was not significantly better than observational training is that participants could have begun applying the testing strategy to all study trials during Part-1 study. Perhaps the procedure of producing an answer became habitual across trials. Because trials alternated between observing and recalling answer/receiving feedback, participants may have found it fluent to repeat the same generation task on every trial. Alternatively, participants may have figured that testing one’s self provided benefits to learning and began employing the tactic more or less throughout the study phase of Part 1. In a future study, asking participants whether they found themselves producing an answer on observational trials as well as feedback trials during Part-1 study would help elucidate this issue.
Another possibility for the lack of a benefit for feedback training over observational training in Part 1 was that alternating between giving a response and observing an exemplar on a time-controlled presentation may have been too confusing for participants to manage. This explanation does not seem plausible, however, because Jacoby et al. (2010) manipulated feedback/observational training within-subjects and the benefits of feedback were apparent. A more likely reason for the lack of a feedback training benefit in the present experiment was the number of categories used. The materials were a major limitation because while Jacoby et al. (2010) used eight categories during the study phase (four for observational training and four feedback training), eight categories were used in the present experiment across the two study phases (four during Part 1 and four during Part 2). Two of the four used in each part were assigned to the feedback training condition and the other two were assigned to the observational training condition.

Participants in the present experiment likely realized that exemplars of two of the categories were consistently presented with the category label and that the other two categories consistently required responses. Furthermore, given that the present experimental design used only two categories per condition meant that during a feedback trial, the correct answer would have been one of only two possible responses. Generating a response, when there were only two response options, may have provided little to no benefit beyond that of an observational trial.

Not only were the feedback training trials not challenging in the present experiment, they also appeared not to have engaged the participant and promote learning. Consistent with this possibility, the feedback training performance for Part 1 of Experiment 6b ($M = .52, SD = .23$) was poorer than the feedback training performance for Part 1 of Experiment 6a ($M = .72, SD = .17$), even though there were only two possible responses for Experiment 6b, as opposed to three
possible responses for Experiment 6a. This difference indicates that the task was not engaging or challenging enough to activate the same retrieval processes that account for the benefits of the testing effect. Participants were essentially guessing throughout training blocks and learning very little. In future studies that require two study-test parts, a set of materials that has a sufficient number of categories must be used for a proper exploration of the effects of feedback training.

In contrast to the metacognitive judgments regarding blocking and interleaving, participants seem to be aware of the benefits of testing even when they never received information about the benefits of testing. It is unclear, however, whether this awareness is due to the procedure of the experiment or to pre-experimental theories of learning. According to Roediger and Karpicke (2006) and Kornell and Son (2009), learners are generally unaware that testing one’s self is more beneficial to memory than is restudying material. It seems unusual that even though performance did not benefit significantly from feedback training, and learners are typically not aware of the benefits of testing according to previous research, participants in the present study reported beliefs that feedback training was better than observational training. The judgements that feedback training was better could be driven by the fact that performance across feedback training cycles improved. Participants could have been sensitive to this improvement and used that improvement to inform their belief that feedback training is better than observational training.

As demonstrated in Experiment 6a, participants benefited from experiencing feedback training relative to experiencing observational training for inductive learning. Due to the limitations of Experiment 6b, however, a follow-up study with an appropriate set of materials should be pursued. Given that participants would experience both types of training and their
relative effects on category learning, results of such a follow-up could potentially demonstrate an even more robust transfer of the benefits of feedback training.

In conclusion, although contrast processes seem to occur outside of awareness and cannot be explicitly applied to new learning situations, testing one’s self seems to be a strategy that learners can actively employ during new learning opportunities. Although further investigation of the benefits of feedback training are needed, the present results suggest that feedback training benefits inductive learning and is also more advantageous than observational training for subsequent category learning.
GENERAL DISCUSSION

Without our ability to organize our surroundings into concepts and categories, we would essentially be faced with a barrage of meaningless and chaotic stimuli. The study of how people induce concepts and categories is thus an essential component of understanding human cognition, and one important goal for the study of category induction is to find ways in which to improve how we organize our world. In nearly all formal educational settings, category learning is essential. For example, in art and architecture education, students must learn to recognize different artistic eras and movements and even create new pieces of work that are reminiscent of those styles. In medical school, students must learn to identify cells, organs, and other biological units, not just in one patient, but across many different patients. And even in mathematics and engineering, recognizing patterns within problem spaces is critical to finding a solution (Holyoak & Thagard, 1995). In all of these situations and more, inductive learning is crucial for successful and creative performance. By finding ways to make category learning more efficient through cognitive psychology research, educational practices can be improved in all of fields of study.

Summary of methods and findings

The research conducted for the present dissertation examined two critical processes that are believed to occur during category learning: discriminating or finding differences between categories and finding commonalities within categories. Previously reported findings and present findings suggest that interleaved study of exemplars from different categories is beneficial because it enhances both of these processes. Firstly, interleaved study places exemplars of different categories next to one another in time, thus allowing learners to draw category boundaries and extract critical between-category discriminations (Birnbaum et al., 2013). Secondly, interleaved study spaces out the presentations of exemplars of the same category, thus
helping learners to generalize features within a category. In Experiments 1a and 1b of the present dissertation, study schedules of experimental conditions provided the benefits of contiguous interleaved presentation, while spacing between exemplars of the same categories was varied across conditions. If the only critical feature of interleaved study was that it placed exemplars of different categories next to one another to allow for contrast processing, then performance across the conditions should not have varied. If the temporal spacing of exemplars from the same category added a benefit to learning, then conditions that increased temporal spacing should have led to better performance. Results from Experiment 1a and 1b indicated that temporal spacing that is inherent to an interleaved study schedule does indeed provide an additional benefit to inductive learning.

Experiment 2 was designed to check for an important possible confounding factor that may be present in most category induction studies: the effect of spacing on recall learning. The benefits of interleaving for category learning have been demonstrated numerous times. Extensive research has also demonstrated the benefits of spacing for recall learning. One way to isolate the effects of interleaving on category inductive learning from recall learning is to reduce or eliminate recall learning during the study phase. In Experiment 2, participants partook in a pretraining phase for category labels prior to exemplar study. The pretraining phase minimized recall learning of category labels during study phase. For both the control condition and pretraining condition, the benefits of interleaving were observed. Thus, Experiment 2 revealed that there is a distinct benefit to category induction from an interleaved schedule. That is, the effect of better classification performance following interleaved study is not simply driven by enhanced learning of category labels owing to the spacing inherent to interleaved study. Thus,
the design and results of Experiment 2 help to add another piece of evidence in support of the important role of between-category contrast processing for category learning.

Experiments 3 and 4 were designed to examine further the mechanisms of between-category contrast processes and whether contrast processes occur even when participants were not made aware of the goal to learn category information. In Experiment 3, intentionality of the learning goal was manipulated between-subjects: Half of the participants were told to learn categories for a later classification test while the other half of the participants were asked to rate each exemplar image. Results revealed that the benefits of interleaving were apparent even when category learning occurred unintentionally, suggesting that the mechanisms enhanced by interleaving were active during incidental learning as well as intentional learning. I discussed the possibility that contrast processes occur largely without conscious effort—a possibility that would also help to account for the finding that people generally judge blocking to be better than interleaving for their learning as well as the finding that interleaving benefits unintentional category learning.

Another piece of evidence that corroborates the argument that contrast processes are carried out automatically is that learners find it very difficult to enact contrast processes when exemplar presentation is not interleaved. Experiment 5 was designed to test the transfer of interleaving benefits. That is, after learning about why interleaving helps category learning, would participants be able to employ the contrast strategy to new learning even in conditions that do not facilitate contrast processes (i.e., blocked study)? Results revealed that people were not able to apply the contrast strategy successfully when study schedule was blocked. It would seem that the most effective way to facilitate contrast processes is to study exemplars in an interleaved schedule.
Although engaging in contrasts processing may be difficult or essentially impossible for learners to accomplish on their own volition in the absence of being presented with exemplars from different categories in an interleaved manner, some strategies can be actively and successfully applied to new learning. The act of testing one’s self on category membership during learning enhances category learning. Testing one’s self is not something that is done automatically. In fact, many students make a conscious effort to do so when studying for exams. In the case of inductive learning, results from Experiment 6a demonstrated that learners could realize the benefits of testing and subsequently use this strategy to improve their learning of entirely new categories.

Possible explanation for the present findings

As a whole, the experiments presented in the present dissertation help reveal why interleaving is beneficial for inductive learning and provide further reasons for why educators should schedule exemplar study in an interleaved fashion. In cases where there are numerous categories to be learned, exemplars should be mixed up and exemplars of the same category should not appear again until all other category exemplars are studied once, maximizing spacing between exemplars within categories and maximizing temporal juxtapositions of between-category exemplars. The benefits of interleaving occur outside of awareness, suggesting that contrast processes are automatic rather than effortful. Additionally, attempting consciously to make category contrasts may not be sufficient to enhance learning. Studying in a blocked schedule, even after knowing why interleaving is beneficial for inductive learning, is not as good as studying in an interleaved schedule. Finally, whenever possible, learners should test themselves throughout exemplar study.
Much of these results fit into a theoretical framework that argues that contrast processes occur largely outside of awareness whereas as commonality processes occur consciously. Ross, Perkins, and Tenpenny (1990) argue that the reminding that occurs during exemplar study leads to explicit comparisons between exemplars within a category. Results of Experiment 1a and 1b reveal that increased spacing likely leads to more difficult retrievals, but that this desirable difficulty leads to enhanced memory for features that are common across exemplars of a given category. Participants often report blocked study as seeming to have been best for inductive learning. This judgment is based on the accurate understanding that commonality processes are critical for category learning, but is also based on the inaccurate belief that making this process as easy as possible is most beneficial.

Participants are rarely aware of the benefits of contrasting between categories. This lack of awareness fits in with the framework that contrast processes occur automatically and outside of awareness. People are constantly learning categories without explicit instructions to do so. Experiment 3 revealed that interleaving, which is beneficial to intentional category learning, is also beneficial to incidental category learning. This finding suggests that contrast processes occur even when learners do not intend to induce categories or think that they will have to perform in a later classification test situation.

Not only do contrast processes seem to occur without awareness and effort, instructing participants to implement contrast processes appears to be ineffective. In Experiment 5, participants were directed to notice differences between categories during study. Even with this instruction, however, participants were largely unsuccessful at transferring the contrast strategy to new learning when the study schedule was blocked. A concerted effort to pick out differences that would effectively define category boundaries did not hinder category learning when study
schedules were interleaved, but did not help when study schedules were blocked. The results from Experiment 5 thus fit well with the framework that contrast processes occur automatically. Contrast processes function best when exemplars are interleaved, rather than when the learner attempts to make effortful comparisons.

Neuropsychology research reveals that different regions of the brain mediate different types of learning. Ashby and Ell (2001) assert that information-integration and rule-based category learning are mediated by different brain structures. Although the category learning that takes place in the studies reported in the present dissertation use information-integration based categories, different brain regions may mediate different processes that occur during information-integration category learning. For example, the prefrontal cortex that mediates rule-based category learning may also be active in within-category exemplar comparison. During study presentation, participants examine exemplars for commonalities within categories, similar to how participants make comparisons to find a common rule in rule-based category learning. Although no neuropsychological evidence has been reported in the present dissertation, it would be interesting to examine the brain structures that are active during category learning via blocked and interleaved study.

Concluding remarks

Inductive category learning is pervasive in countless settings encountered throughout development and life. As children, people learn categories of objects to make sense of, speak about, and interact with the world. And people continue to organize and categorize new encounters throughout life. Most categorical knowledge is acquired unintentionally, while a great deal of category learning in social and formal educational settings is intentional. Inductive learning, whether incidental or intentional, benefits from interleaved study, suggesting that
contrast processes are critical for both types of learning. The present dissertation uncovers characteristics of the mechanisms that underlie inductive category learning in general, while its practical implications apply mostly to intentional category learning for which study manipulations can be implemented. Firstly, not only should study be interleaved, temporal spacing should also be maximized. And even if learners are made aware of the benefits of interleaving, an interleaved study schedule should be used to help facilitate those benefits. Finally, testing one’s self can help during exemplar study, even in conditions where category labels are presented intact. Educators and students are encouraged to use these study techniques to enhance category learning of instructional materials.
Appendix

Three additional studies also revealed a lack of transfer of the benefits of interleaving. The first study was designed to examine whether the benefits of learning categories in an interleaved order could transfer to the learning of new categories. This study had two parts, each consisting of a study phase and a test phase. Participants first studied eight butterfly species in a blocked or interleaved order and were then tested on the eight butterfly species. Test performance revealed a significant benefit of interleaving. Participants then studied a different set of eight butterfly species in a blocked or interleaved order and were then tested on those eight butterfly species. Results revealed that having studied the butterflies in an interleaved schedule during the first learning phase provided no benefit to learning in the second phase. That is, the benefits of interleaved study did not transfer to new learning of a similar but different set of categories.

The remaining two studies were designed to investigate whether interleaved study could not only improve classification of studied categories, but also improve the ability to differentiate similar never-before-studied categories. In the first of these studies, participants studied handwriting samples of six authors either in a blocked schedule or interleaved schedule. At test, participants classified new handwriting samples written by the six studied authors and by three unstudied authors. If participants believed a handwriting sample to have been written by someone other than the original six authors, they were to classify it as written by someone else. Results revealed that participants who studied the six authors in an interleaved schedule were no better than participants who studied them in a blocked schedule at distinguishing examples of the handwriting of the three unstudied authors from examples of the handwriting of the studied authors.
A very similar study in which participants studied four authors and were tested on the four studied authors and one unstudied author revealed similar results. No benefit of interleaved study for correct classification of the unstudied author was observed.
References


Pomerantz (Eds.), Psychology and the real world: Essays illustrating fundamental contributions to society (pp. 56-64). New York: Worth Publishers.


