Learning Concepts and Categories from Examples: How Learners' Beliefs Match and Mismatch the Empirical Evidence

Permalink
https://escholarship.org/uc/item/91q7z7z4

Author
Yan, Veronica

Publication Date
2014

Peer reviewed|Thesis/dissertation
Learning Concepts and Categories from Examples: How Learners’ Beliefs Match and Mismatch the Empirical Evidence

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

by

Veronica Yan

2014
ABSTRACT OF THE DISSERTATION

Learning Concepts and Categories from Examples:
How Learners’ Beliefs Match and Mismatch the Empirical Evidence

by

Veronica Yan

Doctor of Philosophy in Psychology
University of California, Los Angeles, 2014

Professor Robert A. Bjork, Chair

The compelling intuition in learning to abstract categories and concepts from examples is that such inductive learning is enhanced when the examples are blocked by category, such as solving addition problems first before moving on to subtraction problems, or learning painters’ styles by comparing multiple paintings by one artist before moving on to the paintings of another artist. Recent research, however, suggest that interleaving examples of separate to-be-learned categories enhances such inductive learning. Learners typically fail to appreciate the benefit of interleaving (as compared to blocking study by category), both before study experience (Tauber, Dunlosky, Rawson, Wahlheim, & Jacoby, 2013) and even after completing a test in which their performance demonstrates better learning after interleaved study (e.g., Kornell & Bjork, 2008; Kornell, Castel, Eich, & Bjork, 2010). Across ten experiments, I investigated learners’ metacognitive beliefs as to what optimizes inductive learning and how those beliefs, when in error, can be modified. In Chapter 2, I report on my efforts, using theory- and experience-based de-biasing procedures (Koriat & Bjork, 2006), to dislodge learners’ beliefs in the superiority of
blocking. Learners’ intuitions favoring blocking proved remarkably difficult to dislodge, despite my taking multiple steps to direct participants’ attention to the link between initial study schedule and subsequent test success, and providing explanations of the interleaving benefit. In Chapter 3, I report on my efforts to extend the cognitive literature by investigating whether hybrid schedules of study—that is, schedules that combine elements of blocking and interleaving—are effective. While learners remained insensitive to the benefits of purely interleaved study, they demonstrated some appreciation of hybrid schedules that were about as effective as a purely interleaved schedule. Finally, in Chapter 4, I report on my efforts to examine learners’ a priori beliefs by analyzing self-scheduled study orders in light of the cognitive results found in Chapter 3, which revealed that learners can demonstrate some metacognitive sophistication with respect to category learning, when given the right opportunity to do so.
The dissertation of Veronica Yan is approved.

Elizabeth Ligon Bjork
Alan D. Castel
James W. Stigler
Aimee Drolet Rossi

Robert A. Bjork, Committee Chair

University of California, Los Angeles

2014
# TABLE OF CONTENTS

List of Figures vii
List of Tables ix
Acknowledgments x
Vita xi

## CHAPTER 1

**Introduction and Overview** 1
Interleaving Enhances Learning Concepts and Categories from Exemplars 1
What do Learners Believe Enhances Inductive Learning? 2
Overview of the Dissertation 5

## CHAPTER 2

**On the Difficulty of Mending Metacognitive Illusions: Trying, and Mostly Failing, to Overcome Learners’ Belief that Blocking Exemplars Enhances Inductive Learning** 7

Experiment 1: Experience- and Theory-based Debiasing 10
Experiment 2: Directing Attention to Test Performance and Initial Study Schedule 13
Experiment 3: Color-Coded Schedules 19
Experiment 4: Everything but the Kitchen Sink 22

## CHAPTER 3

**Are there Ways to Combine Blocking and Interleaving that are More Effective for Learning than Is Pure Interleaving?** 30

Experiment 1: Hybrid Schedules with 6 Exemplars per Category 34
Experiment 2: Hybrid Schedules with 12 Exemplars per Category 40
CHAPTER 4

Do Learners Blindly Block Study? Self-Scheduling and the Role of Prior Exposure to Interleaving

Experiment 1: Self-Schedule of 6 Exemplars per Category
Experiment 2: Self-Schedule of 20 Exemplars per Category
Experiment 3: Self-Schedule of Courses and Modules
Experiment 4: Self-Schedule of New or Repeated Study After Prior Exposure

CHAPTER 5

Concluding Remarks

Summary of Results
Is Interleaving Always Better Than Blocking?
When and Why does Interleaving Enhance Inductive Learning?
Future Directions in Self-Regulated Category Learning

Appendices

Appendix A: Judgments of Learning during Blocked and Interleaved Study
Appendix B: Post-test Description of the Five Schedules in Chapter 3, Experiment 1
Appendix C: Instructions for Chapter 4, Experiment 1
Appendix D: Instructions and Experiment Packet for Chapter 4, Experiment 2
Appendix E: Example of Study Rationales in Chapter 4, Experiment 2
Appendix F: Instructions and Experiment Packet for Chapter 4, Experiment 3
Appendix G: Instructions and Experiment Packet for Chapter 4, Experiment 4
Appendix H: Additional Results for Chapter 4, Experiment 4

References
LIST OF FIGURES

Figure 2.1. Participants’ actual performance and judgments of what led to better learning as a function of information condition in Experiment 1. 13

Figure 2.2. Proportion of paintings whose artist was correctly identified as a function of study schedule across test blocks in Experiment 2. Error bars represent one standard error of the mean. 15

Figure 2.3. The proportion of study schedules correctly identified as a function of correct artist identification for each presentation schedule in Experiment 2. Error bars represent one standard error of the mean. 16

Figure 2.4. Participants’ actual performance and judgments of what led to better learning as a function of information condition in Experiment 2. 18

Figure 2.5. Participants’ actual performance and judgments of what led to better learning as a function of information condition in Experiment 3. 22

Figure 2.6. An example of the two frames used to denote blocked and interleaved schedules of study in Experiment 4. 25

Figure 2.7. Participants’ actual performance and judgments of what led to better learning as a function of information condition in Experiment 4. 27

Figure 3.1. A visual representation of the five schedules in Experiment 1. Paintings were presented sequentially (in the diagram, this order is to be read left to right, top to bottom). Each letter represents different image by a given artist; the specific order of the artists, however, was block randomized—each ‘block’ being marked by a black border. 50
Figure 3.2. Performance on the classification test across test blocks, by study schedule, in Experiment 1. Error bars represent standard error of the mean. 37

Figure 3.3. Proportion of participants within each experimental condition (schedule) that judged each schedule to be the most effective in Experiment 1. 39

Figure 3.4. Average classification performance under each schedule in Experiment 2. Error bars represent standard error of the mean. 41

Figure 3.5. Proportion of participants within each experimental condition (schedule) that judged each schedule to be the most effective in Experiment 2. 43

Figure 3.6. Classification performance for each artist when studied in each of the five schedules in Experiment 1 (top) and Experiment 2 (bottom), with the artists from each study organized from most difficult to least difficult to learn. 46

Figure 4.1. The maximum run length (i.e., maximum number of paintings by the same artist in a row) in the presentation sequences constructed by the participants in Experiment 2. 62

Figure 4.2. Proportion of participants—by condition—whose schedules fell into each of the categories (left), and who selected each sample schedule (right). 68

Figure 4.3. Proportion of participants within each initial schedule x second phase study set condition that predominantly blocked or predominantly interleaved their phase 2 study in Experiment 4. 74

Figure 4.4. Experiment 4’s phase 2 classification performance (adjusted means, after covarying out phase 1 performance), by initial study schedule, phase 2 study set and phase 2 schedule. 78
LIST OF TABLES

Table 3.1  Comparison of the five schedule conditions across Experiments 1 and 2

Table 3.2  Mean rank sum of each schedule, when examined by artist for Experiments 1 and 2 separately and combined

Table 4.1  The number of participants who scheduled each order and judged each order to be the most effective for learning artist styles, in Experiment 1.

Table 4.2  The number of participants who scheduled each order and judged each order to be the most effective for learning artist styles in Experiment 2

Table 4.3  Broad and specific strategy classifications in Experiment 4, examples of schedules, and the number of participants whose schedules fell into each category

Table 5.1  Summary of the features and results of existing studies comparing interleaving and blocking
ACKNOWLEDGMENTS

This research was supported by Grant No. 29192G from the James S. McDonnell Foundation.

Thanks first and foremost to Bob and Elizabeth Bjork who not only gave me the opportunity to pursue research in psychology, but also provided incredible mentorship, showing me how to direct the inquisitive mind, and allowing me the freedom to do so. Thanks also to Alan Castel, Jim Stigler, and Danny Oppenheimer who welcomed me into their lab meetings, and to all CogFog members, who opened me up to an array of perspectives, an arsenal of diverse knowledge, and gave insightful feedback on my own research.

The “older siblings” of the Bjork lab family—John Nestojko, Colin Clark, and Jeri Little, and more recently, Kou Murayama—probably do not realize it, but I learned so much watching them discuss, critique, and lead the lab. And special thanks to all my other fellow lab members and friends who have fueled my interests through our many research-related discussions and helped me retain my sanity through our many non-research related discussions: Adam Blake, Alison Tsai, Doe Buchli, Carole Yue, Courtney Clark, Erin Sparck, KP Thai, Mikey Garcia, Mike Friedman, Monica Birnbaum, Saskia Giebl, Sharon Noh, Tara Patterson, Toshi Miyatsu, and Victor Sungkhasetteee.

Finally, to my family: My parents who instilled in me a love for learning and pushed me toward graduate school, my brother who has always looked out for me despite being five years younger, and to Shan who has buoyed me up when I’ve been stressed and pulled me out into the sunlight and fresh air after sitting indoors for far too long.
VITA

2009 B.A. Hons, Experimental Psychology
University of Cambridge
Cambridge, United Kingdom

2010-2014 Teaching Assistant
Department of Psychology
University of California, Los Angeles

2011-2013 Graduate Student Research Assistant
Department of Psychology
University of California, Los Angeles

2011 M.A., Psychology
University of California, Los Angeles

PUBLICATIONS


PRESENTATIONS


Yan, V. X. (2013, February). Efficient learning is not always easy learning: The peculiarities of optimizing memory. *Keynote Address, Pearson Mastering Leadership Conference, Austin, TX.*


Yan, V. X. (2012, September). What do learners know about their own learning, and what can we do to improve metacognition? *Keynote Address, Pearson Non-Majors Biology Leadership Conference, Seattle, WA.*

CHAPTER 1

Introduction and Overview

The ability to learn concepts and categories from exemplars is a fundamental component of the human experience, and accordingly, it has been a topic of interest in many subfields of psychology under the guise of a variety of different names, including generalization, transfer, category induction, abstraction, and perceptual learning. Part of the reason for the different labels is that concepts and categories cover so many different types of knowledge and stimuli—from perceptual categories (e.g., learning to recognize artists’ painting styles) to more cognitively-based concepts (e.g., laws of physics), from categories that can easily be defined by a set of rules to categories that cannot (e.g., What is a “game”?; Wittgenstein, 1953/2001).

Concept and category learning is important from infancy through formal schooling, and beyond. In fact, concept and category learning comprises the lion’s share of formal education (as opposed to strict memorization). For example, a medical student who can learn to spot fractures in X-rays and apply this learning to finding fractures in the X-rays of new patients is going to be a more useful doctor than the student who can simply memorize the location of fractures in the set of X-rays presented in class; the architect who can apply his knowledge of physics to designing buildings is more employable than the one who can only regurgitate Newton’s laws.

Interleaving Enhances Learning Concepts and Categories from Exemplars

Within the large corpus of research, one particularly counterintuitive and theoretically interesting finding is that the learning of multiple related categories is benefited by interleaving study of these different categories, rather than by blocking study by category. Kornell and Bjork
(2008) were the first to demonstrate this finding using a paradigm in which participants were tasked with learning artists’ painting styles so as to be able to later identify new paintings by the studied artists. This basic interleaving benefit has now been replicated in other labs using artist stimuli (e.g., Kang & Pashler, 2012; Verkoeijen & Bouwmeester, 2014), in older adults (Kornell, Castel, Eich, & Bjork, 2010; Wahlheim, Dunlosky, & Jacoby, 2011), and in learning a wide range of naturalistic and realistic materials—butterflies (Birnbaum, Kornell, Bjork, & Bjork, 2013), birds (Wahlheim et al., 2011), aurally and visually presented text materials (Zulkiply, McLean, Bath, & Burt, 2012), math formulae (Taylor & Rohrer, 2010; Rohrer & Taylor, 2007), and women’s voices (Yan, Vetter, & Bjork, 2012).

The finding that interleaving is more effective for category learning than is blocking is counterintuitive—Kornell and Bjork themselves had predicted the opposite effect, and the majority of learners report that blocking was more effective even after having experienced better test performance on the interleaved artists (Kornell & Bjork, 2008; Kornell et al., 2010). Given the importance of category learning (both in and out of formal education) and considering that the majority of learning is self-regulated, it is important that learners are able to understand how to self-regulate inductive learning effectively and efficiently.

What Do Learners Believe Enhances Inductive Learning?

One of the simplest methods of examining what learners believe is effective for inductive learning is to ask them. Such aggregate judgments are often elicited at the end of an experiment after participants have experienced the category-learning task and one or more schedules. This approach is what Kornell and Bjork (2008) and Kornell et al. (2010) used: After the test phase, participants were asked which schedule—blocked or interleaved—was more effective for
learning artists’ painting styles, or whether they were equally effective. The majority of participants, ranging from 56%-78% in these studies, reported that blocking was more effective than interleaving, and only a minority (ranging from 4%-21%) reported that interleaving was better than blocking (the remaining responded that the schedules were equally effective). Similarly 80% and 58% of participants in the two experiments reported in Zulkiply et al. (2013) claimed that blocking was more effective after learning visually and aurally presented text passages, respectively.

Why do so many learners believe that blocking is more effective than interleaving? There appear to be multiple influences: Blocking may lead to greater feelings of fluency during study; learners may hold an a priori theory that blocking enhances learning by drawing attention to commonalities between exemplars within a category; and learners have experienced blocking throughout their formal education.

To test the hypothesis that blocking increases the feeling of fluency during study, we can ask learners to make online judgments of learning (JOLs) during the course of studying. In a study reported in Appendix A of this dissertation, participants studied 6 paintings by each of 12 artists. The paintings of half of these artists were studied blocked and the paintings by the remaining artists were studied interleaved. Additionally, immediately after the study of the third and sixth paintings by half of the blocked and half of the interleaved artists, participants were asked to make a judgment of category learning (“How likely are you to correctly recognize a new painting by [name of the just-studied artist] on the final test?”; response rated on a scale of 0-100). After study of the third exemplar of a given category, participants gave reliably higher JOLs to the blocked artists ($M = 62.28, SD = 18.62$) than to the interleaved artists ($M = 52.09, SD = 21.38$), $t(108) = 5.36, p < .001, d = .51$. This difference in JOLs for the blocked and
interleaved artists was still reliable for the sixth paintings of each artist, \((M = 67.52, SD = 20.99)\) for the interleaved vs. \(M = 71.20, SD = 17.24\) for the blocked) although much reduced, \(t(108) = 2.10, p < .05, d = .19\) (and the interaction between schedule and painting position was significant, \(F(1,108) = 11.63, p < .01, \eta^2_p = .10\)).

Wahlheim et al. (2011) also collected judgments of category learning (likelihood of correctly classifying new exemplars of the studied categories, bird families) immediately after study and found that participants gave higher judgments to the bird families that were studied interleaved (\(M = .46\)) than to those families that were studied blocked (\(M = .39\)), although the differences were, again, slight. While on the surface, the judgments of category learning in Wahlheim et al. may appear to contradict the aggregate judgments, they differ in that the category learning judgment prompt makes no reference to study schedule. Given Wahlheim et al.’s findings, however, and my findings that any differences in JOLs of the blocked and interleaved categories are small by the end of study, fluency during study cannot be the only reason behind learners’ insistence that blocking is more effective than interleaving for learning.

Evidence also exists that learners bring \textit{a priori} theories of optimal scheduling into the lab with them. Instead of asking for judgments of learning during study or aggregate judgments at the end of experiments, we can also examine learners’ beliefs by examining their study-regulated study choices. Tauber, Dunlosky, Rawson, Wahlheim, and Jacoby (2013) asked participants to choose on a trial-by-trial basis from what category they wanted to study a next exemplar. Tauber et al. categorized the vast majority (ranging from 78-100% of the participants in their four experiments) “blockers”—in other words, most participants were more likely to block their learning than they were to interleave. Particularly in Tauber et al.’s first three studies (in which participants self-scheduled after an initial interleaved familiarization with all the
exemplars), participants tended to study one category exhaustively before moving onto the next. Tauber et al. (2013) therefore concluded that participants held and drew upon an *a priori* belief that category learning is enhanced by blocking the exemplars of a given category.

**Overview of the Dissertation**

The current set of studies was designed to examine more closely learners’ metacognitive beliefs about the optimal scheduling of exemplars in category induction. One speculation as to why learners give higher judgments of category learning (e.g., Wahlheim et al., 2011) to interleaved categories than to blocked categories, but are unable, when directly asked, to express that interleaving is better than blocking, is that learners do not make the connection between their test performance and the initial study schedules. The four experiments in Chapter 2 were designed, therefore, to disambiguate the link between study and test in the hopes that learners would subsequently use this information when making their judgments of schedule efficacy. This experience-based de-biasing was used in conjunction with theory-based de-biasing—explaining to learners that interleaving enhances discrimination and, therefore, enhances overall category induction.

The two experiments in Chapter 3 tested whether combinations of blocked and interleaved schedules would in fact lead to better overall induction than only interleaving study and probed whether participants, when given these additional hybrid schedule options, would still prefer blocked study or shift to one of the hybrid schedules.

Finally, the experiments in Chapter 4 adopted Tauber et al.’s (2013) basic approach in asking participants to self-schedule their own learning. In contrast to the Tauber et al. experiments and in light of the experimental findings in Chapter 3, analysis of learners’ self-
schedules used a more fine-tuned approach, allowing for the coding of various types of hybrid
schedules, examining the situations in which participants are more likely to deviate from a purely
blocked schedule, and probing their understanding of the relative merits of blocked and
interleaved study.
CHAPTER 2
On the Difficulty of Mending Metacognitive Illusions: Trying, and Mostly Failing, to
Overcome Learners’ Belief that Blocking Exemplars Enhances Inductive Learning

Given the increasing frequency with which learning is happening outside of formal classrooms and in unsupervised environments, knowing how to manage one’s own learning is increasingly important. Multiple findings demonstrate, however, that most of us are prone to beliefs about how we learn that lead us to manage our learning in non-optimal ways. Learners, for example, fail to appreciate (a) the pedagogical benefits of testing (e.g., Hartwig & Dunlosky, 2012; Karpicke, Butler, & Roediger, 2009; Kornell & Bjork, 2007); (b) that producing errors can sometimes enhance learning (e.g., Huelser & Metcalfe, 2012); and (c) that spacing, rather than massing, repeated practice or study sessions tends to enhance long-term retention (e.g., Cohen, Yan, Halamish, & Bjork, 2013; Simon & Bjork, 2001; Zechmeister & Shaughnessy, 1980).

Such findings raise an interesting question: What kinds of experiences and/or information might have the power to dislodge learners’ preferences for non-optimal conditions of learning? To pursue that issue, we focused on a particularly dramatic instance in which learners’ judgments and preferences are at odds with their performance: the inductive learning of categories and concepts via exposure to interleaved or blocked exemplars. Learners have a strong belief that blocking exemplars of to-be-learned categories enhances inductive learning, even after their own learning has profited from interleaving.

Kornell and Bjork (2008) and Kornell, Castel, Eich, and Bjork (2010), for example, using a task in which participants learned the styles of artists from examples of their paintings, found that interleaving, not blocking, enhanced inductive learning—as measured by participants’
subsequent ability to identify the studied artist responsible for never-before-seen paintings. The great majority of the participants, however, said that blocking, not interleaving, helped them learn better—even after their own final-test performance had exhibited benefits of interleaving. Kornell et al. (2010), for example, found that only 19% of younger adults and 4% of older adults judged interleaving to be better after taking the final test, and Tauber, Dunlosky, Rawson, Wahlheim, and Jacoby (2013) found, by allowing participants to select the order in which they studied exemplars of bird families, that participants came to the task already believing exemplars should be blocked. Similar benefits of interleaving exemplars have been found for the learning of various types of categories (Birnbaum, Kornell, Bjork, & Bjork, 2013; Kang & Pashler, 2012; Kornell et al., 2010; Rohrer & Taylor, 2007; Vlach, Sandhofer, & Kornell, 2008; Wahlheim, Dunlosky, & Jacoby, 2011; Zulkiply, McLean, Burt, & Bath, 2012).

Why Do Learners Judge that Blocking Is Superior?

One factor that seems a likely contributor to learners believing that blocking enhances induction is that blocking is what teachers and trainers typically do; that is, blocking is everywhere in the real world. Another factor is that learners may intuit that inducing concepts from exemplars relies on being able to see the commonalities across exemplars that define a category and that blocking will make those commonalities more apparent.

In addition, such pre-existing beliefs will tend to be supported by certain objective and subjective experiences. When learning motor and procedural skills, for example, blocking tends to produce better performance during acquisition (see, e.g., Simon & Bjork, 2001, and Rohrer & Taylor, 2007, where participants learned keystroke skills and volume formulae, respectively). Subjective experiences during an induction task may also be influential: Viewing blocked
exemplars of a given category can lead to a sense of fluency, reinforcing the feeling that blocking enhances inductive learning. Prior studies (e.g., Begg, Duft, Lalonde, Melnick, & Sanvito, 1989; Hertzog, Dunlosky, Robinson, & Kidder, 2003; Koriat & Ma’ayan, 2005) have shown that a higher subjective ease of processing can lead to higher judgments of learning (i.e., use of the heuristic: easily learned means easily remembered; Koriat, 2008), and it may simply feel easier to study exemplars in a blocked schedule.

What Does It Take to Dislodge Learners’ Belief in Blocking?

In four experiments, we examined the effectiveness of what Koriat and Bjork (2006) labeled “theory-based” and “experience-based” methods of “mending metacognitive illusions.” Using Kornell and Bjork’s paintings task, our theory-based manipulation consisted of providing participants—before asking them whether blocking or interleaving had helped them to learn better—with different levels of information regarding the benefits of interleaving, such as why interleaving has those benefits and why blocking may seem better. Our experience-based manipulations consisted of attempts to draw participants’ attention to their actual test performance vis-a-vis presentation schedules by asking participants during the test how a given artist’s paintings had been scheduled during study (Experiment 2) and by presenting paintings by blocked and interleaved artists against differently colored backgrounds or frames (Experiments 3 and 4).

General Method

Our basic task replicated Kornell and Bjork’s (2008) Experiment 1A. During study, a total of 72 landscape paintings (6 by each of 12 artists) were presented with the artist’s name
shown below each painting, one at a time, and at a 3-s rate. The six paintings by a given artist were either blocked or interleaved with those of other artists, with assignment of artist to condition randomized across participants. Order of presentation followed the scheme: B I I B B I I B B I I B, where every “B” refers to a set of six paintings by the same artist and every “I” refers to a set of six paintings by six different artists.

Participants were instructed that their task was to learn the artists’ styles in preparation for a final test on which they would be shown new paintings and have to identify the studied artist responsible for that painting (selecting the artist’s name from a response panel). A 45-s interval during which participants played Tetris intervened between the study and test phase.

Experiment 1

Method

Participants, materials, and procedure. Participants were 124 US residents recruited from Amazon Mechanical Turk. Each participant received $0.50, and the experimental procedure was created using Collector, an open-sourced, PHP-based experimental platform (https://github.com/gikeymarcia/Collector). Before the final test, which included one new painting by each studied artist, participants were made aware of the two presentation schedules and told that, after selecting the artist they thought responsible for a given new painting, they would receive feedback regarding the correct artist. After the final test, but before collecting their metacognitive judgments (“Which learning schedule led to the most effective learning?”), participants were randomly assigned to one of four information conditions: no information ($n = 29$), percentage ($n = 31$), percentage + fluency discounting ($n = 33$), and everything ($n = 31$).
The no-information condition replicated the procedure of Kornell and Bjork (2008). The additional information received by participants in the other three conditions is shown below.

Percentage: “In fact, previous research has shown that, contrary to many people's intuitions, an overwhelming 90% of individuals actually learn better when the paintings of an artist are presented intermixed with paintings by other artists. Seeing all the paintings by one artist consecutively is not as good for learning as mixing all the artists up.”

Percentage + fluency discounting: “In fact, previous research has shown that—even though it might feel easier to focus on one artist at a time—an overwhelming 90% of individuals actually learn better when the paintings by an artist are presented intermixed with paintings by other artists. Seeing all the paintings by one artist consecutively is not as good for learning as mixing all the artists up.”

Everything: Same information given to participants in the percentage + fluency discounting conditions plus: “This comes as a surprise to many people, as there is a strong sense—while you are studying the paintings—that it seems easier to pick out the defining characteristics of an artist if you study all of his/her paintings consecutively. However, the benefit of interleaving the artists is that your attention was drawn to the DIFFERENCES between the artists, and this is really what is crucial for the task of learning artists’ styles.”

Finally, all participants were again reminded of the two learning schedules and asked, “Which schedule do you think led to better learning of the artists’ styles?” and selected their response from the following options: Blocking was better, interleaving was better, or that the two schedules were equally effective.
Results and discussion

Consistent with the results of previous studies, overall correct classification of new paintings (i.e., identifying which studied artist had been responsible for it) was significantly better for those artists whose paintings were studied in an interleaved manner ($M = .52, SD = .30$) versus a blocked manner ($M = .42, SD = .24$), $t(124) = 4.12, p < .001, d = .49$.

Figure 2.1 displays participants’ judgments about the effectiveness of the two study schedules (represented by the height of the bars) as a function of information condition, together with the actual relative efficacy of the two schedules. Broadly speaking, as the information becomes more detailed, fewer participants judged blocking as more effective and more judged interleaving as more effective. What is most impressive, however, is how relatively ineffective all of the information conditions were in modifying participants’ belief that blocking facilitated their learning. In the no-information condition, only 6 of 29 participants judged interleaving to be superior. Being told that interleaving was superior for 90% of learners increased the proportion (13/31) of participants willing to judge interleaving as better, but the majority (18/31) still judged blocking to be as good or better than interleaving. Telling participants that interleaving produces better performance for 90% of learners, while also giving them a reason (fluency) why they might be fooled into thinking otherwise, had—if anything—a smaller effect: Only 10 participants in that condition judged interleaving to be better, whereas 23 judged blocking as good or better than interleaving.

Even in the very heavy-handed “everything” condition, where participants received the full set of reasons why interleaving produces better inductive learning plus the caution that people are prone to feel that the opposite is true, only a slight majority of participants (16/31) responded that interleaving was better. Basically, participants, overall, were very willing to
believe that any benefits of interleaving for other participants did not offset the benefits of blocking for their own learning.

Figure 2.1. Participants’ actual performance and judgments of what led to better learning as a function of information condition in Experiment 1.

Experiment 2

Given the results of Experiment 1, we speculated that one factor in participants’ holding fast to their faith in blocking is that they did not link their performance on the final test to the blocking-versus-interleaving manipulation. We attempted, therefore, to draw participants’ attention to this study manipulation–test performance connection in Experiment 2.
Method

Participants and procedure. Participants were 80 US residents recruited from Mechanical Turk, and each received $0.50 for participating. Participants were randomly assigned to the same information conditions as in Experiment 1: no-information \((n = 19)\), percentage \((n = 19)\), percentage + fluency discounting \((n = 21)\), and everything \((n = 21)\). The same procedure was used except that (a) the final test presented four 12-item randomized blocks of new paintings—rather than just a single block of one new painting by each artist—with the goal of allowing participants to accumulate more test-phase experience; and (b) after receiving feedback on the artist responsible for each new painting, participants were asked, “Was this artist presented massed (consecutive) or spaced (intermixed)?” to which they responded by selecting one of two buttons.

Results and discussion

Classification performance. The results of the classification test are presented in Figure 2. They replicate Kornell and Bjork’s (2008) results very closely: Interleaving produced better performance than blocking, and performance improved across test blocks (particularly for the blocked artists) owing to feedback. A 2 (schedule) x 4 (test block) repeated measures ANOVA revealed a significant main effect of schedule, \(F(1,76) = 14.80, MSE = .07, p < .001, \eta_p^2 = .16\), with interleaving \((M = .56, SD = .21)\) producing better inductive learning than blocking \((M = .48, SD = .23)\). A significant main effect of test block was also obtained, \(F(3,237) = 4.27, MSE = .03, p < .01, \eta_p^2 = .05\), with performance improving across test blocks. Finally, there was a marginal interaction, \(F(3,237) = 2.20, MSE = .03, p = .09, \eta_p^2 = .03\), reflecting the increasing performance in the blocked condition across test blocks (from .44 to .54), and the essential lack thereof in the interleaved condition (from .57 to .58). Interestingly, the improvement for the
blocked artists may reflect an interleaving benefit: Given feedback was provided on each test trial, blocked artists were receiving, for the first time, what amounted to interleaved practice.

![Figure 2.2](image_url) Figure 2.2. Proportion of paintings whose artist was correctly identified as a function of study schedule across test blocks in Experiment 2. Error bars represent one standard error of the mean.

**Memory for study schedule.** Overall, participants correctly recalled the study schedule 64\% (SD = 15\%) of the time. Participants’ accuracy in identifying the study schedule of each artist, conditional upon correct identification of the artist responsible for the painting, is presented in Figure 2.3. A 2x2 within-subjects ANOVA confirmed that schedule memory for interleaved artists ($M = .79$, $SD = .25$) was significantly higher than for blocked artists ($M = .50$, $SD = .29$), $F(1,76) = 41.29, p < .001, \eta_p^2 = .35$, and schedule memory for correctly identified artists ($M = .68$, $SD = .34$) was significantly higher than for incorrectly identified artists ($M = .60$, $SD = .30$), $F(1,76) = 21.62, p < .001, \eta_p^2 = .22$. These main effects, however, were qualified by a significant interaction, $F(1,76) = 47.26, p < .001, \eta_p^2 = .38$. When an artist’s paintings had
been interleaved during study, participants were actually more likely to report that the artists’ paintings had been interleaved when they incorrectly identified the painting as that of a different artist \( (M = .82, SD = .22) \) than when they correctly identified the artist \( (M = .76, SD = .29) \), \( t(77) = 2.90, p < .01, d = .33 \). When an artist’s paintings had been blocked during study, however, participants were more likely to judge “blocked” when they correctly identified the artist \( (M = .61, SD = .29) \) than when they misidentified the artist \( (M = .38, SD = .29) \), \( t(79) = 6.90, p < .001, d = .77 \).

In other words, when participants were incorrect in identifying the artist of a particular painting, they were more likely to judge that the artist’s paintings had been interleaved. Thus, participants’ judgments appear to be influenced by the heuristic, “If I was wrong, I must have studied that artist’s paintings interleaved”—a heuristic that seems likely to be mediated by a belief that fluency during study is an indicator of better learning.

Figure 2.3. The proportion of study schedules correctly identified as a function of correct artist identification for each presentation schedule in Experiment 2. Error bars represent one standard error of the mean.
**Metacognitive judgments.** Participants’ judgments concerning which schedule led to better learning for each information condition are shown in Figure 2.4. Compared to Experiment 1, and despite our attempts to draw participants’ attention to their own performance, an even greater majority judged blocking to be superior to interleaving! Notably, not a single participant in the no-information condition reported that interleaving was better (vs. 21% in Experiment 1).

Again, more participants indicated that interleaving was better for learning as more information was provided regarding why that might be the case. When told that 90% of participants learn better given an interleaved schedule, only 4 of 19 participants responded that interleaving was better for them. When also given the information that the feeling of fluency can be misleading, only 5 of 21 participants judged that interleaving was better. Only in the everything condition did more participants choose interleaving than blocking (9 vs. 8), but the majority (57%) of even the participants in the everything condition still judged blocking to be as good or better than interleaving.
Figure 2.4. Participants’ actual performance and judgments of what led to better learning as a function of information condition in Experiment 2.

Clearly, our attempt to draw learners’ attention to the relationship between initial study schedule and final test performance did not dislodge their belief in the benefits of blocking. Overall, participants were very willing to believe that they fell in the 10% of learners who benefit from blocking, not interleaving: Excluding those in the no-information condition, who were not told that 90% of people learn better interleaved, and those in the everything condition who were told in detail why blocking may appear superior, but is not, 78% of the participants placed themselves in the 10% of individuals who learn as well or better from a blocked schedule. Even in the everything condition, which included a complete explanation for why interleaving is better, only 43% of participants judged interleaving to be better for learning.
Perhaps, however, we observed this pattern of results because our attempt to enhance participants’ awareness of the link between test performance and study schedule was not effective. Participants identified the correct schedule only about two-thirds of the time, and, furthermore, were more likely to state that an artist had been studied interleaved when they incorrectly, not correctly, identified the artist responsible for a given test painting. We reasoned, therefore, that our manipulations might have elicited participants’ *a priori* theories about the superiority of blocking instead of linking test performance to prior study schedules. Another possible factor is that our use of four tests-with-feedback blocks, which allowed successful classification rates for blocked and interleaved artists to became similar by the final test block, made it less clear to participants which schedule had actually led to better test performance.

**Experiment 3**

In Experiment 3, we attempted to draw participants’ attention to the link between test-performance and study-schedules more forcefully and explicitly by eliminating any ambiguity about the study schedule for a given artist’s paintings and by employing only a single test block. Specifically, we made the two schedules very salient from the start by (a) informing participants that they would study half the artists blocked and half of the artists interleaved, and (b) consistently placing the blocked and interleaved artists’ paintings against different colored backgrounds.

**Method**

**Participants and procedure.** Eighty US residents recruited via Amazon Mechanical Turk received $0.50 for participating. Again, participants were randomly assigned to four
information conditions: no-information \((n = 20)\), percentage \((n = 21)\), percentage + fluency
discounting \((n = 21)\), and everything \((n = 18)\).

The procedure was the same as in Experiments 1 and 2, but included three features
designed to increase participants’ awareness of the link between study schedule and final-test
performance. First, participants were told prior to study that the paintings of half the artists
would be presented blocked and the other half presented interleaved. Second, paintings were
presented against either a black or white background to designate whether the responsible artist
was in the blocked or interleaved condition, with assignment of backgrounds to schedules
counterbalanced across subjects and with each participant informed of the background-schedule
mapping prior to both study and test phases. Third, after the final test, but before being asked to
judge the relative effectiveness of blocked and interleaved schedules, participants were asked to
recall the background–schedule mapping.

**Results and discussion**

**Classification test.** Again, we replicated the interleaving benefit. Paintings were
classified significantly better if the artist’s paintings had been interleaved \((M = .52, SD = .28)\)
than if they had been blocked \((M = .34, SD = .23)\), \(t(79) = 5.69, p < .05, d = .81\).

**Metacognitive judgments.** Participants’ judgments regarding which schedule led to
better learning are displayed in Figure 2.5. Again, the number of participants judging blocking
to be better decreased with their being told both that interleaving is better for 90% of individuals
and why people can be fooled into thinking that blocking is better.

The no-information condition revealed that—despite background-coding of the study
schedules—only 3 of 20 participants believed that interleaving was better for learning, whereas
13 participants believed that blocking was better for learning and another 4 participants thought
blocking was equal to interleaving. As more theory-based de-biasing information was introduced, however, more participants chose interleaving: In the percentage condition, the number of participants choosing interleaving rose to 8 of 21, and in the percentage + fluency discounting condition, a small majority (11/21) chose interleaving. Finally, in the everything condition, 8 of 18 participants judged interleaving to be better, whereas 6 and 4 participants, respectively, judged blocking to be better or equal to interleaving.

Overall, the percentages of participants who believed they were in the minority 10% of learners (who do not learn better with interleaving) was 62%, 48% and 56% for the percentage, percentage + fluency discounting, and everything information conditions, respectively.

**Memory for the background coding.** Overall—despite our emphasis on the background-schedule mapping—only 63% of participants recalled the mapping scheme correctly, with 22% responding “don’t know,” and 15% reporting it backwards. That more than a third of the participants did not learn (or remember) the background assignment was unexpected, but perhaps that result, too, contains a message—namely, that the inductive learning task is so compelling to participants, especially using artists’ paintings, that peripheral details, even if emphasized, are not well encoded.
Figure 2.5. Participants’ actual performance and judgments of what led to better learning as a function of information condition in Experiment 3.

Experiment 4

In Experiment 4, we attempted to make the two schedules more salient by (a) forewarning participants of the two schedules and (b) by surrounding the paintings being shown via the two schedules within two very different frames.

Method

Participants and design. One hundred and twenty US residents, recruited via Amazon Mechanical Turk, each receiving $1.00 for participating, were randomly assigned into one of the five experimental conditions: no-information \((n = 28)\), percentage \((n = 31)\), percentage + fluency discounting \((n = 31)\), and everything \((n = 30)\). For each participant, paintings by the blocked and
interleaved artists were presented within two very different frames, a gold frame or a black frame, as shown in Figure 2.6, and the mapping of frame to schedule was counterbalanced across participants.

**Materials and procedure.** The procedure for Experiment 4 expanded on the procedure of Experiment 3 in ways designed to increase participants’ awareness of the mapping of artists to the blocked and interleaved schedules.

**Pre-training phase.** Participants were first forewarned that the paintings by half of the artists would be presented consecutively (i.e., blocked) while the paintings by the other half of the artists would be all mixed together (i.e., interleaved). Additionally, they were told that the blocked and interleaved artists would be presented with different frames, and they were then shown two short demonstrations—using paintings by famous artists who were not among the studied artists—designed to ensure they knew what blocked and interleaved study meant. Three paintings by Van Gogh were shown for three seconds each, with the appropriate frame and artist name written beneath each image (the “blocking” demonstration), and then four paintings by four different artists (Claude Monet, Leonardo Da Vinci, Salvador Dali, and Vincent Van Gogh) were shown, again with the appropriate frame and appropriate name printed beneath each image (the “interleaving” demonstration). Before and after each demonstration, they were reminded of what schedule each series of paintings represented.

Participants were then informed that the different frames would be used to help them keep track, during the study phase, of whether a blocked or interleaved artist was being studied, and that—in addition to being tested on whether they had learned the artists’ painting styles—they would also be asked to identify how an artist’s painting had been studied (blocked or interleaved). Next, they were tested on whether they could recall the frame-schedule mapping
(selecting from one of two multiple choice options), and given immediate feedback of the correct mapping.

Finally, prior to the study phase, participants were trained on the 12 names of the artist that they would be studying. They were shown the six names of the interleaved artists and given 30 seconds to study them, before engaging in a free recall test of the names, with feedback. This same process was repeated with the six names of the blocked artists (the order of studying and testing blocked and interleaved names was counterbalanced between participants). Participants then engaged in one more test-with-feedback cycle on each of the two sets of artists’ names.

**Study and test phases.** The study and test phases were administered as in Experiments 1, 2, and 3, but with following exceptions: Paintings by blocked and interleaved artists were shown on different frames (see Figure 2.6) during the study phase and during the test phase; participants were reminded of the frame-schedule mapping right before the final test; and feedback during the test phase was a little more elaborate (instead of just being presented with the name of the correct artist, participants were shown, “Correct/Incorrect! [Name] was the answer”).

**Post-test information and questions.** Participants were reminded of the two study schedules and were then shown one of the information prompts (no-information, percentage, percentage + fluency discounting, or everything). They were then asked to recall which frame mapped onto which study schedule (“blocked = gold frame; interleaved = black frame”, or “blocked = black frame; interleaved = gold frame”, or “I don’t remember”), only after which they were asked to judge which schedule led to better learning.

Finally, participants were shown the 12 artist names and asked to select the six names that were studied interleaved, and then the six names that were studied blocked (order counterbalanced).
Results and discussion

Classification performance. Again, we replicated the basic interleaving benefit: Paintings by artists that were studied interleaved ($M = .54, SD = .30$) were correctly classified at a significantly higher rate than paintings by artists that were studied blocked ($M = .39, SD = .25$), $t(119) = 5.29, p < .001, d = .97$.

Memory for frame colors. We checked whether participants could recall the frame-schedule mapping twice: Once before they studied any paintings, and a second time after the final test. At both time points, participants were highly accurate in recalling the correct frame-schedule mapping—95% ($SD = 22\%$) and 97% ($SD = 18\%$), respectively. The manipulation to make presentation schedule salient, therefore, was successful.

Classification of artists’ names. Participants were reasonably good at selecting the names of the blocked and interleaved artists at the end of the study: On average, participants correctly selected 4.39 ($SD = 1.98$) names and incorrectly selected only 1.5 ($SD = 1.99$) names.

Figure 2.6. An example of the two frames used to denote blocked and interleaved schedules of study in Experiment 4.
Metacognitive judgments. Participants’ judgments concerning which schedule led to better learning for each information condition are shown in Figure 2.7. Despite forewarning participants of the two schedules, pre-training on the names of the blocked and interleaved artists, presenting the frames around the images throughout the test phase, explicitly telling people whether they were correct or not, and the fact that almost every single participant could tell us (after the test), the mapping between frame and schedule, there still remains the large tendency for participants to believe that blocked study leads to more effective learning than does interleaving study. This pattern is most striking in the “no-information” condition—despite all the measures taken to draw learners’ attention to the schedules, 19 out of 28 (68%) still claimed that blocking was better than interleaving, and only 3 (11%) appreciated the benefits of interleaving.

Among the other three conditions in which extra information was provided, there still was no condition in which a convincing majority of participants believed that interleaving was better than blocking: Only 48%, 48% and 37% of the percentage, percentage + fluency discounting, and everything conditions, respectively, reported that interleaving was more effective than blocking.
General Discussion

In teaching individuals that interleaving is effective for learning—and by extension, how better to self-regulate their own learning—a number of obstacles must be overcome. Experiment 1 demonstrated that simply allowing people to experience the interleaving benefit and then informing them of the theory behind the interleaving benefit was not sufficient to alter their preference for blocking in any significant way.

Experiment 2 revealed that participants appear to use a heuristic that makes it difficult, during the test, to draw their attention to how artists had been studied. Specifically, participants were more likely to judge artists they failed to identify as being ones studied in an interleaved manner, suggesting at least a partial (many of the correctly classified interleaved artists were
correctly judged as having been interleaved) reliance on the heuristic, “If I’m wrong, it must have been difficult and, therefore, it must have been interleaved.” Even when such a heuristic is not activated during the test, however, and study schedules are made clear, as in Experiments 3 and 4, participants still failed in any significant way to learn from experience that interleaving is more effective than blocking. Finally, across all four experiments, even the most heavy-handed theory-based de-biasing was fairly ineffectual in convincing participants that interleaving was the more beneficial schedule for learning, with only 52%, 43%, 44%, and 37% of participants in the everything conditions, respectively, judging interleaving as better for learning.

Overall, our metacognitive results seem quite remarkable, especially given the heavy demand characteristics present in the information conditions: Participants were asked which schedule led to better learning for them immediately after being told which schedule ought to lead to better learning and, furthermore, and that it does so for 90% of learners. Even were a strong theory-de-biasing method to be successful at convincing learners that interleaving is better for a particular set of stimuli, how to convince individuals to use interleaving as a general study strategy in their everyday learning of various types of study materials presents a clear challenge.

Why is a belief in the effectiveness of blocking so firmly entrenched? The present results suggest two, not necessarily mutually exclusive, sources of this belief. First, years of formal education, in which concepts are almost universally taught and studied in a blocked fashion, has likely established a strong belief in the benefits of blocking that is difficult to dislodge in a single experimental session. Additionally, the apparent, if misleading, ease of learning felt when studying exemplars in a blocked manner—and the lack thereof when studying exemplars in an interleaved manner—is probably very difficult to discount and serves to support individuals’ prior belief in the benefits of blocking.
Concluding Comments

The present research adds to the growing body of literature illustrating that learners are prone to certain misconceptions about how to learn. It also demonstrates, in a striking way, that overcoming such misconceptions—however crucial to achieving our potential as learners—is remarkably difficult.
CHAPTER 3

The experiments reported in Chapter 2 illustrate how deeply entrenched the belief that blocking is more effective for learning than is interleaving. In spite of combinations of experience-based and theory-based debiasing and the many measures taken in Experiment 4 to ensure that learners were able to successfully link their test performance to the initial study schedule (falling just short of simply telling participants that they learned better when studying artists in an interleaved manner), the majority of participants still refused to accept that interleaving was better than blocking.

Blocking and interleaving, as defined in the literature and in the Chapter 2 experiments, merely represent two ends of a spectrum along which lie many other possible schedules of study. Given the theoretical reasons for benefits of both blocking (highlights within-category similarities) and interleaving (highlights between-category differences), the experiments reported in Chapter 3 explore the efficacy of hybrid schedules that combine the two, comparing them against pure blocked and interleaved schedules. Two interesting questions follow from this line of study: Are hybrid schedules more effective than pure interleaving, and given these additional scheduling options, do people still prefer pure blocking?

The experiments reported in Chapter 2 demonstrated that people are unlikely to choose pure interleaving over pure blocking, but if hybrid schedules are just as, or more, effective than pure interleaving, then might they represent schedules that learners would be willing to adopt in their own practices?
Are there Ways to Combine Blocking and Interleaving that are More Effective for Learning than Is Pure Interleaving?

The ability to induce categories and concepts from exposure to multiple exemplars and to apply this learning to categorizing new exemplars is fundamental, because being able to categorize the world around us allows us to understand, perceive, and interact with our environments. Biologists, for example, have categorized living organisms into multiple levels (kingdom, phylum, class, order, family, genus, and species), children learn to distinguish between the friendlier furry dogs that are kept as pets and the wilder coyotes that roam the streets, physicists can see deep relational structures underlying problems that may look very different on the surface, and art authenticators learn to recognize artists’ styles so as to distinguish authentic works from forgeries.

How can we optimize category learning? One very intuitive method—and one that is commonly used in textbooks and classrooms—is to focus on learning one concept at a time. In most typical math classrooms, for example, instructors teach their students how to solve one type of problem, and they have them practice several questions using a given formula before starting on a different type of problem. Similarly, in learning to categorize different bird families or artist paintings, learners express a strong preference for studying exemplars blocked by category rather than interleaving the different categories (e.g., Kornell & Bjork, 2008; Tauber, Dunlosky, Rawson, Wahlheim, & Jacoby, 2013). So compelling is this intuition that an eminent cognitive psychologist, Ernst Rothkopf, was quoted as saying, “Spacing is the friend of recall, but the enemy of induction” (1977, personal communication, as reported in Kornell & Bjork, 2008). Intuitively, blocked study should better allow learners to notice the commonalities that tie a category together, and there is a large literature that supports the importance of comparison
processes for learning, transfer, and understanding deeper relational structures (Gentner & Namy, 1999; Gentner, Loewenstein, Thompson, & Forbus, 2009; Gick & Holyoak, 1983; Oakes & Ribar, 2005; Rittle-Johnson & Star, 2011).

The literature on the role of comparisons in category learning, however, has typically focused on making comparisons between exemplars within the same category. Benefits, however, can also be found from making comparisons (or contrasts) across different to-be-learned categories: For instance, Day, Goldstone, and Hills (2010) showed that middle school students who compared examples from two confusable types of problems (negative and positive feedback loops) were better able to distinguish between them than those who compared two examples of the same problem type. Similarly, Taylor and Rohrer (2010) demonstrated that when students intermixed practice of four different types of math problems (calculating faces, edges, corners and angles) scored significantly higher than those who practiced the problems blocked by type, and they also made significantly fewer errors in discriminating between the formulae.

Kornell and Bjork (2008) were the first to examine the relative efficacy of sequentially blocking or interleaving the exemplars of to-be-learned perceptual categories. Participants studied six paintings by each of 12 different artists either blocked by artist (i.e., all the paintings by a given artist were studied before moving onto the next artist’s paintings) or interleaved (i.e., paintings from different artists were studied mixed together). Despite Rothkopf’s intuitively compelling argument that spacing study of exemplars is the “enemy of induction,” Kornell and Bjork found that interleaving led to better classification of new paintings by the studied artists. In other words, participants were less able to discriminate between artists’ styles (as well as distinguishing the studied styles from close lures) after studying them in a way that would
promote within-category comparisons. This basic finding has now been replicated many times (e.g., Birnbaum, Kornell, Bjork, & Bjork, 2013; Kang & Pashler, 2012; Kornell, Castel, Eich, & Bjork, 2010; Wahlheim, Dunlosky, & Jacoby, 2011; Yan, Vetter, & Bjork, 2012). Interestingly, another robust finding from these studies is that learners—even after experiencing better classification performance on a final test for those artists that were studied interleaved—consistently and overwhelming report that blocking exemplars by category is better for learning than is interleaving the exemplars of different categories.

Interleaving is generally thought to be effective for learning because it enhances discrimination and contrasts between the exemplars of different categories. The artists used by Kornell and Bjork (2008) may, therefore, represent a set of stimuli in which discriminating between artists is relatively more useful than searching for the shared features that tie a given artists’ works together. Birnbaum et al. (2013) showed that introducing spacing (filled with interpolated trivia questions) between the interleaved presentation of paintings by different artists interrupted contrast processes and did not benefit inductive learning, and Kang and Pashler (2012) demonstrated that simultaneously presenting paintings by different artists was both (a) more effective than either sequentially blocking paintings by the same artist and simultaneously presenting multiple paintings by the same artist and (b) as effective for learning artist styles as was sequentially interleaving paintings by different artists.

Interleaving is not always better than blocking, however: Carvalho and Goldstone (2013) have argued that the relative efficacy of interleaving and blocking depend on the category structure. When categories are highly similar and, therefore, discrimination between categories is relatively more important, interleaving outperforms blocking. On the other hand, when the
within-category similarity is low, and it is therefore, relatively difficult to spot the commonalities within the exemplars of a given category, blocking outperforms interleaving.

But why should we choose between making within-category comparisons and between-category discriminations? The reality is that both comparisons and contrasts are generally useful for inductive learning and, thus, combining elements of blocked and interleaved study could prove to be optimal beyond using just one schedule or the other. In two experiments, using either six or 12 exemplars per category, we explored the effectiveness of hybrid schedules, including testing whether blocked study should come before interleaved study, or vice versa. Additionally, we report more detailed metacognitive analyses—beyond deciding between blocked and interleaved study, are learners able to appreciate the relative benefits of blocked and interleaved study, and thus when given the option, might they prefer hybrid schedules of learning?

**Experiment 1**

**Method**

**Participants and design.** One hundred and fifty-six undergraduate students (126 females, one undisclosed) from UCLA participated for partial course credit. Participants studied six paintings by each of twelve different artists via one of five schedules: blocked ($n = 35$), interleaved ($n = 31$), mini-blocks ($n = 31$), blocked-to-interleaved ($n = 29$), or interleaved-to-blocked ($n = 30$). They were then tested on their ability to identify the artists that painted each of 48 new paintings (four by each artist).

**Materials and procedure.** The experiment was programmed using Collector, an online, open-source, PHP-based platform (downloadable from www.gikeymarcia.com). Participants were informed that they would be presented with the paintings of 12 different artists (6 paintings
per artist for a total of 72 paintings) and that their task was to learn the artists’ painting styles. On the final test, they would be shown new paintings by these artists and asked to identify, from a list of names, the artist responsible for each painting. Participants were shown each painting for three seconds at a time; paintings were resized to be as close to 500 x 400 pixels as possible; paintings were centered on the computer screen; and the artist name was printed below each image.

In the blocked condition, the six paintings by a given artist were presented consecutively (e.g., Ax6, Bx6, Cx6…Lx6; six in a row). In the interleaved condition, the paintings were presented in six randomized blocks of twelve paintings, with each block consisting of one painting per artist (e.g., ABCDEFGHIJKL x 6; notated as 1-1-1-1-1-1). In the mini-blocks condition, the six paintings by a given artist were viewed in two cycles (although, from the participant’s point of view, study consisted only of one long sequence): three paintings by each artist were viewed consecutively in the first cycle, and then the remaining three paintings by each artist were presented consecutively in the second cycle, with the order of the artists randomized in each cycle (e.g., AAABBBCC… AAABBBCC…; notated as 3-3). In the blocked-to-interleaved condition, paintings by the artists were presented in three cycles: In the first cycle, three paintings by each artist were presented consecutively, then two paintings by each artist were presented consecutively, and finally, one painting by each artist was presented (e.g., AAABBBCC… AABBC… ABC…; notated as 3-2-1). In the interleaved-to-blocked condition, the order of the three cycles was reversed (e.g., ABC… AABBC… AAABBBCC…; notated as 1-2-3). A visual representation of the five conditions is shown in Figure 3.1 and Table 3.1 illustrates a comparison of the specific design of these conditions across
the two experiments. In all conditions, the specific order of presentation of a given artist’s paintings, and their assignment to study and test phase, were randomized for each participant.

After all 72 paintings had been studied, participants engaged in a game of Tetris for 45 seconds and then moved on to the test phase. In the test phase, participants were presented with 48 new paintings sequentially (four paintings per artist, one presented in each of four randomized blocks of twelve paintings) and asked to select the artist they believed responsible for each painting from a list of names. They were provided with the correct answer after each selection (i.e., presented the painting again with the correct name written beneath it) and were allowed to complete the test at their own pace.

Finally, participants were read descriptions of the five different schedules (shown in Appendix B), were reminded of the schedule in which they studied the paintings and then asked which schedule they believed would lead to the best learning of artist styles, and finally, were asked to make a choice between either purely blocked or purely interleaved study.

Table 3.1

Comparison of the five schedule conditions across Experiments 1 and 2

<table>
<thead>
<tr>
<th>Schedule</th>
<th>Experiment 1</th>
<th>Experiment 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocked</td>
<td>6 in a row</td>
<td>12 in a row</td>
</tr>
<tr>
<td>Mini-blocks</td>
<td>3-3</td>
<td>3-3-3-3</td>
</tr>
<tr>
<td>Blocked-to-interleaved</td>
<td>3-2-1</td>
<td>6-3-2-1</td>
</tr>
<tr>
<td>Interleaved-to-blocked</td>
<td>1-2-3</td>
<td>1-2-3-6</td>
</tr>
<tr>
<td>Interleaved</td>
<td>1 (x6)</td>
<td>1 (x12)</td>
</tr>
</tbody>
</table>

Note. Numbers represent the number of paintings by a given artist studied in a row before moving onto a new artist
Results and discussion

Classification performance. The results of the classification test, by test block, are presented in Figure 3.2. As is expected from tests with feedback (which amounts to interleaved learning), performance increased across the test blocks. A 4x5 mixed ANOVA showed that there was a significant main effect of test block, $F(3,453) = 13.16, MSE = .02, p < .001, \eta^2_p = .08$, and a marginal main effect of schedule, $F(1,151) = 2.12, MSE = .11, p = .08, \eta^2_p = .05$, and no interaction, $F < 1$. Overall, performance increased across test blocks with a mean performance of .44 ($SD = .21$) on test block 1 and a mean performance of .51 ($SD = .21$) on test block 4.

![Figure 3.2](image_url)

Figure 3.2. Performance on the classification test across test blocks, by study schedule, in Experiment 1. Error bars represent standard error of the mean.

As the test phase essentially amounted to interleaved learning, we analyzed only the results from the first test block to examine the effect of the different schedules. A one-way between-subjects ANOVA revealed a significant effect of schedule, $F(4,151) = 2.54, p < .05, \eta^2_p$
Post-hoc comparisons revealed that interleaved study \((M = .52, SD = .20)\) led to significantly better performance than did blocked study \((M = .37, SD = .21)\), \(t(64) = 2.96, p < .01, d = .74\), and interleaved-to-blocked study \((M = .41, SD = .17)\), \(t(59) = 2.29, p < .05, d = .60\). Blocked-to-interleaved study \((M = .47, SD = .21)\) was marginally better than blocked study, \(t(62) = 1.83, p = .07, d = .46\). There was only a trend for the mini-blocks condition \((M = .45, SD = .23)\) to be better than the blocked study condition, \(t(64) = 1.48, p = .14, d = .37\). No other differences between conditions were revealed, \(ps > .05\).

**Metacognitive judgments.** When asked which of the five schedules would be most effective for learning artist painting styles, the most popular responses were mini-blocks (36%), blocked (24%), and blocked-to-interleaved (21%). Interleaved-to-blocked received only 16% of the votes, and interleaved study was judged most effective by only 3% of participants. The pattern of responses, however, differed by the conditions that participants had experienced, \(\chi^2(16) = 27.54, p < .05\). Figure 3.3 displays the proportion of times that each schedule was judged most effective within each condition. While mini-blocks was generally most popular overall, those in the interleaved-to-blocked condition were especially likely to favor mini-blocking (53%). Those who had experienced interleaving heavily favored both mini-blocks (42%) and pure blocking (32%), as did those who had experienced blocked study (37% and 31%, respectively). Those who had experienced blocked-to-interleaved schedules, however, were most likely to choose their experienced schedule to be the most effective (38%).

When forced to choose between a blocked or interleaved schedule, the majority 86% of participants chose the blocked schedule. Again, however, patterns differed significantly between conditions, \(\chi^2(1) = 14.25, p < .01\): While not a single person who had experienced blocking chose interleaving, 27% of those in the mini-blocks condition and 26% of those in the
interleaved condition thought that interleaving would be more effective than blocking. Only 14% and 7% of participants from the blocked-to-interleaved and interleaved-to-blocked conditions, respectively, indicated that interleaving would be more effective. Participants’ judgments, therefore, were still categorically wrong when forced to make a choice between blocking and interleaving.

*Figure 3.3.* Proportion of participants within each experimental condition (schedule) that judged each schedule to be the most effective in Experiment 1.
Experiment 2

Experiment 1 replicated the now-common finding that interleaved study leads to better inductive learning than does blocked study. It failed, however, to show that combining blocked and interleaved study schedules led to any greater benefit than pure interleaved study. With only six exemplars per artist however, “blocking” in the hybrid schedules meant that only three exemplars were studied consecutively, and it could be that three is simply not enough to obtain the benefits of blocking. In Experiment 2, participants were therefore presented with 12 exemplars per artist to give ample opportunity to learn throughout all blocked and interleaved components of the hybrid schedules.

Method

Participants and design. One hundred and seventy-four UCLA undergraduates (123 females; one undisclosed; mean age = 19.9 years, SD = 1.40 years, age range = 18-26) participated for partial course credit. Again, participants experienced one of five schedules: blocked (n = 32), interleaved (n = 38), mini-blocks (n = 34), blocked-to-interleaved (n = 35), and interleaved-to-blocked (n = 35). The design of these conditions are shown in Table 3.1.

Materials and procedure. Experiment 2 was the same as Experiment 1, except for three differences: twelve paintings per artist were studied, only two paintings per artist were tested in the test phase, and no feedback was given during the test. The blocked and interleaved study conditions, therefore, remained very similar to those in the Experiment 1—in the blocked condition, twelve paintings by each artist were presented consecutively; and in the interleaved condition, the stimuli were presented in twelve randomized blocks of twelve paintings (where again, each block consisted of one painting per artist). In the mini-blocks condition, three paintings per artist were shown consecutively in each of four cycles (i.e., 3-3-3-3).
blocked-to-interleaved condition, there were also four cycles—participants first saw six paintings per artist consecutively, then three, then two and finally, one painting per artist (e.g., AAAAAABBBBB… AAABBB… AABCC… ABCDEF…; notated as 6-3-2-1), and the order of these four cycles were reversed for the blocked-to-interleaved condition (i.e., 1-2-3-6).

![Figure 3.4](image)

*Figure 3.4. Average classification performance under each schedule in Experiment 2. Error bars represent standard error of the mean.*

**Results and discussion**

**Classification performance.** The average classification score for each condition is presented in Figure 3.4, which shows that we replicated the standard benefit of interleaved ($M = .58$, $SD = .19$) over blocked study ($M = .42$, $SD = .16$). The mini-blocks ($M = .55$, $SD = .18$) and blocked-to-interleaved schedules ($M = .55$, $SD = .17$) led to performance that did not differ
significantly from the pure interleaved schedule. The interleaved-to-blocked schedule ($M = .48$, $SD = .19$), however, led to performance that fell somewhere between the blocked and interleaved schedules.

A one-way between-subjects ANOVA confirmed that there were significant differences between the conditions, $F(1,4) = 5.13, MSE = .03, p < .001, \eta^2_p = .11$. Post-hoc comparisons revealed that the interleaved, mini-blocks, and blocked-to-interleaved conditions did not differ from each other, $ps > .05$, and all three led to significantly better classification performance than did the blocked condition, $t(68) = 3.99, p < .001, d = .97$, $t(64) = 3.35, p < .001, d = .84$, and $t(65) = 3.49, p < .01, d = .87$, respectively. The interleaved-to-blocked condition was significantly worse than the interleaved condition, $t(71) = 2.41, p < .05, d = .57$, and not significantly better than the blocked condition, $t(65) = 1.38, p > .05, d = .34$. Additionally, the interleaved-to-blocked condition was marginally worse than both the mini-blocks, $t(67) = 1.78, p = .08, d = .43$, and the blocked-to-interleaved conditions, $t(68) = 1.86, p = .07, d = .45$.

**Metacognitive judgments.** Participants’ metacognitive judgments, by their experienced schedule, are presented in Figure 3.5 (two participants failed to answer this question). Participants’ judgments for what schedule of study leads to the most effective learning largely converged upon the schedules that blocked study, or started with blocked study—the pure blocked (32%), mini-blocks (24%) and the blocked-to-interleaved (33%) schedules. Only 7% of participants chose the interleaved-to-blocked schedule and only 3.5% chose the purely interleaved schedule.

We conducted a chi-square test of independence to test whether the efficacy judgment (eliminating the interleaved and interleaved-to-blocked responses owing to insufficient numbers) was affected by the study condition that participants had experienced. This analysis showed that
judgment and experienced schedule were independent, $\chi^2(8) = 12.25, p = .14$. The patterns are nonetheless interesting, however: Reports that purely blocked schedules were most effective was the most prevalent among those who experienced the purely interleaved schedule (41%); on the other hand, those who had experienced a blocked schedule were most likely to report that a mini-blocks schedule was the most effective (44%).

When participants were forced to choose between either blocked or interleaved learning, 141 out of 171 (82%) respondents indicated they would want to study the artists in a blocked schedule, and these responses did not differ by experienced schedule, $\chi^2(4) = 3.78, p > .05$.

![Figure 3.5](image)

**Figure 3.5.** Proportion of participants within each experimental condition (schedule) that judged each schedule to be the most effective in Experiment 2.
Combined experiments 1 and 2 results

**Classification performance.** A 2 (experiment) x 5 (schedule) between-subjects ANOVA revealed a significant main effect of experiment, $F(1, 320) = 12.04, p < .01, \eta^2_p = .04$, a significant main effect of schedule $F(4, 320) = 7.13, p < .001, \eta^2_p = .08$, and no interaction, $F < 1$. Classification performance in Experiment 2 ($M = .52, SD = .20$) was significantly greater than that in Experiment 1 ($M = .44, SD = .20$), as would be expected from studying more exemplars. It is interesting to note, however, that the benefit of studying twice the number of exemplars is not large: performance increased by only an average of 8%.

Post-hoc t-tests revealed much the same pattern as the individual Experiment 1 and Experiment 2 results: Interleaved ($M = .56, SD = .20$), blocked-to-interleaved ($M = .52, SD = .19$), and mini-blocks ($M = .50, SD = .21$) all led to significantly greater induction performance than did pure blocking ($M = .39, SD = .19$), $t(134) = 4.97, p < .001, d = .84, t(129) = 3.72, p < .001, d = .66$, and $t(130) = 3.27, p < .01, d = .57$, respectively. No significant differences between interleaved, blocked-to-interleaved, and mini-blocks schedules emerged, $p_s > .05$. Finally, performance in the interleaved-to-blocked schedule ($M = .45, SD = .18$) was significantly worse than that in both the blocked-to-interleaved, $t(127) = 2.10, p < .05, d = .37$, and the interleaved study schedules, $t(132) = 3.33, p < .01, d = .58$, and was marginally worse than performance in the mini-blocks schedule, $t(128) = 1.71, p = .09, d = .30$, and marginally better than that in the purely blocked schedule, $t(130) = 1.67, p = .09, d = .29$.

**Effect of schedules on the induction of specific artists**

There are reasons to expect that the learning of artists’ styles would not always benefit from interleaving—Artists whose styles are more easily discriminable from those of other artists might benefit less from being juxtaposed with other artists. We found, however, consistent
effects of schedules across the different artists within our stimuli set. Figure 3.6 shows the results of each study schedule for each artist—the top half shows the results for Experiment 1 (test block 1 only), and the bottom half shows the results for Experiment 2. The order of artists along the x-axis represents the worst to best classified artists on the final test within each experiment. From the figure, it is clear that there are remarkable similarities in the pattern of results across the artists, and notably, interleaving led to better induction than blocking in just about every single case.

Each schedule was ranked from worst (1) to best (5) for the learning of each artist’s style, and the mean rank sum of the five schedules in the two experiments separately and combined are shown in Table 3.2. A Friedman test confirmed that there were reliable differences between the five schedules, Friedman $\chi^2(4) = 34.88, p < .001$, and Wilcoxon rank sum tests revealed that all pairwise comparisons were significant, $ps < .05$, except for the difference between mini-blocks and pure interleaved, $p = .09$, and between blocked-to-interleaved and mini-blocks, $p = 1.00$.

Table 3.2

*Mean rank sum of each schedule, when examined by artist, for Experiments 1 and 2 separately and combined*

<table>
<thead>
<tr>
<th>Schedule</th>
<th>Experiment 1</th>
<th>Experiment 2</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocked</td>
<td>1.79</td>
<td>1.21</td>
<td>1.21</td>
</tr>
<tr>
<td>Interleaved-to-blocked</td>
<td>2.58</td>
<td>2.21</td>
<td>2.08</td>
</tr>
<tr>
<td>Blocked-to-interleaved</td>
<td>3.21</td>
<td>3.54</td>
<td>3.5</td>
</tr>
<tr>
<td>Mini-blocks</td>
<td>3.29</td>
<td>3.83</td>
<td>3.79</td>
</tr>
<tr>
<td>Interleaved</td>
<td>4.13</td>
<td>4.21</td>
<td>4.42</td>
</tr>
</tbody>
</table>
Figure 3.6. Classification performance for each artist when studied in each of the five schedules in Experiment 1 (top) and Experiment 2 (bottom), with the artists from each study organized from most difficult to least difficult to learn.
General Discussion

Experiments 1 and 2 showed converging results: While combining blocked and interleaved schedules was better for learning than was purely blocked learning, no combination led to better overall inductive learning than pure interleaving. Consistently—across experiments and across specific artists—we also found that interleaved-to-blocked study is not as effective for learning as is pure interleaving or either of the other two hybrid schedules (blocked-to-interleaved and mini-blocks). This result is somewhat surprising from the perspective that blocking enhances comparisons and interleaving enhances contrasts, as the blocked-to-interleaved and interleaved-to-blocked schedules should afford the same opportunities for making comparisons and contrasts. The reasons for why interleaved-to-blocked study is less effective may lie in the explanations from the spacing effect literature.

In the present studies (as is also the case in many more general instances of category learning), category induction is only one part of the whole process—learners first have to induce each artist’s style (perhaps through noticing commonalities and differences within and between categories), learn the names, and then learn the association between the induced style and the corresponding artist’s name. Interleaving study of the paintings by different artists not only allows learners to spot the differences between the artists, but it also introduces inherent spacing between paintings by the same artist. Spacing may be beneficial if spaced presentations act as “reminders” and encourage retrieval of prior paintings (e.g., Appleton-Knapp, Bjork, & Wickens, 2005; Benjamin & Tullis, 2010; Thios & D’Agostino, 1976), and it may also enhance associative learning of the name and induced categories. This associative learning is not a trivial component of category learning, particularly when there are many categories to learn. Kang and Pashler (2012) concluded that spacing is advantageous in category learning only when it
promotes discriminative contrast, but their materials consisted of only three different artists, where associative learning should be relatively easy. Using 16 categories of butterflies, Birnbaum, Kornell, Bjork, and Bjork (2013) showed that when the number of discriminative contrasts was held constant, a greater number of intervening items between exemplars from the same category led to greater inductive learning.

In the present experiments, the blocked-to-interleaved schedule is analogous to an expanding schedule: Presentations of a given artist are increasingly spaced out, with no spacing between the first three paintings and an average interval of 11 intervening paintings (by different artists) by the end. On the other hand, the interleaved-to-blocked schedule is analogous to a contracting schedule in which the presentations of a given artist’s paintings become presented increasingly closer to each other. Research has shown that expanding schedules benefit memory (typically compared to uniform spaced conditions) particularly in cases when forgetting occurs rapidly between the first and second presentations (Landauer & Bjork, 1978; Storm, Bjork, & Storm, 2010). Additionally, assuming that it is impossible to associate an induced artist’s style to its corresponding artist’s name before the style has been induced, spacing should be relatively more effective when it occurs toward the end of study (after a style has been induced) rather than toward the beginning. It is important to note, however, that despite these apparent benefits of the blocked-to-interleaved schedule, it was still no better for learning than the purely interleaved schedule.

Participants’ judgments of what schedule of inductive learning is most effective show that—although participants do hold the overall, metacognitively inaccurate tendency to prefer blocked study—their metacognition is not as wildly inaccurate as previous studies might suggest. When forced to choose between the two extremes, 84% (across both experiments) chose
blocking over interleaving. Faced, however, with a range of options that lay between these two extremes, the majority 60% of participants chose one of the top three (and equally effective) schedules: interleaved, mini-blocks, or blocked-to-interleaved. In fact, given the five tested options, the proportion of participants who reported believing that pure blocking was the most effective dropped to 28%.

It remains unclear how these results—both the classification performance results and the metacognitive results—might be affected by different category structures. With respect to category induction, learning from exemplars is likely to involve multiple processes—the search for similarities among the exemplars of the same category (Goldstone, 1996), the search for differences across different categories, associative learning of the induced categories and category labels (particularly when there are a greater number of to-be-learned categories), and perhaps even retrieval or reloading of previously learned formulae or rules (e.g., as in the case of math learning, or other rule-based categories). While the first process—noticing commonalities (e.g., by spotting common features or discovering some rule to which all members of a category conform) within a category—is likely to benefit from blocking (Carvalho & Goldstone, 2013), the latter three processes may benefit more from interleaving and spacing. Manipulating within- and between-category similarity and exploring hybrid schedules using different category structures may prove a useful method for testing and teasing apart the multiple processes underlying category learning and contribute to a richer understanding of how we learn from examples. Finally, an interesting metacognitive question that follows is whether people are sensitive to these category structure differences in scheduling their own study or judging different schedules or whether they simply hold a general and universally applied theory to primarily block study.
Figure 3.1. A visual representation of the five schedules in Experiment 1. Paintings were presented sequentially (in the diagram, this order is to be read left to right, top to bottom). Each letter represents different image by a given artist; the specific order of the artists, however, was block randomized—each ‘block’ being marked by a black border.
Two main findings emerged from the two hybrid-schedule experiments in Chapter 3. First, that certain hybrid schedules—namely mini-blocks and blocked-to-interleaved schedules—can be just as effective for learning as a purely interleaved schedule. Second, unlike pure interleaving, which (consistent with prior results) was very unpopular among learners, the mini-blocks and blocked-to-interleaved schedules were accurately judged to be effective for learning. In fact, prior studies may have overestimated the reliance of learners on blocked study. Given only the two extremes (blocked and interleaved), 84% of participants choose blocking. When provided when a set of hybrid schedules, however, the percentage of participants who chose pure blocking dropped to 28%; and instead, the majority (60%) chose one of the three best performing schedules (interleaved, mini-blocks, or blocked-to-interleaved).

In order to examine the *a priori* theories about optimizing induction that learners might hold, the experiments reported in Chapter 4 gave participants hypothetical learning scenarios and asked them to self-schedule their own learning. Do learners use hybrid schedules? In addition, the four experiments in Chapter 4 explored a number of different factors that might influence the strategies learners use—namely, choice configuration (Exp 1), number of exemplars per category (Exps 1 and 2), interrelationships between the to-be-learned categories (Exp 3), and prior experience of blocking or interleaving and the familiarity of to-be-learned stimuli (Exp 4).
Do Learners Blindly Block Study? Self-Scheduling and the Role of Prior Exposure to Interleaving

While decades of research have focused on various types of category representations (e.g., prototype vs. exemplar based; Medin & Schaffer, 1978; Rosch, 1975; Posner & Keele, 1968), and multiple systems of category learning (e.g., rule-based vs. information-integration; Ashby, Alfonse-Reese, Turken, & Waldron, 1998; Ashby & Maddox, 2005), research into the schedules that optimize category learning has only recently become a focus of such research. Studies on the sequencing of study exemplars for inductive learning have typically focused on two extremes: blocked and interleaved study. In a blocked schedule, participants study all the exemplars of a given category before moving onto the next category. In an interleaved schedule, participants study the exemplars of different categories in a mixed up order.

Early studies on sequencing of category learning seemed to favor blocking over interleaving. For instance, Gagné (1950) found that blocking nonsense form ‘categories’ yielded better performance and fewer errors during the last two trials of learning than did interleaving the categories. Kurtz and Hovland (1956) also compared the effectiveness of the two schedules for learning four nonsense syllable categories of geometric patterns, which varied along four relevant dimensions (shape, color, size and position; categories were defined by combinations of two of these dimensions). Although participants were equally good at classifying exemplars into the appropriate category, those in the blocked condition were better able to verbalize the rules defining each category. A closer look at these studies thus shows that their claims about the benefits of blocking may be overstated: Stimuli were highly artificial categories that were defined strictly by rules, and the actual benefits of blocking over interleaving in the transfer of learning (as opposed to, for example, acquisition curves) was not demonstrated. Nevertheless,
the notion that “spacing in the enemy of induction” (Rothkopf, personal communication as reported in Kornell & Bjork, 2008) and that blocking is better because it allows learners to spot the similarities that tie a category together was hugely compelling.

The study by Kornell and Bjork (2008) that revived the recent interest in optimizing induction used more realistic materials, for which rules are not so easy to describe, and tested classification of new exemplars. In contrast to intuitions and the earlier studies, Kornell and Bjork found that interleaving exemplars from different categories was better for inductive learning than was blocking exemplars by category. Specifically, in their study, participants were instructed to learn the artist painting styles of 12 different landscape artists and were tested on their ability to correctly identify the artist responsible for new, never-before-studied, paintings. After the final test, they were asked which schedule (blocked or interleaved) they believed led to better learning of the artists’ styles.

Kornell and Bjork’s findings were surprising on two counts, and they have stimulated many follow-up studies across different labs. First, that interleaving study of different artists was significantly better (and by a rather large margin!) for inductive learning than was blocking study by artist ran in stark contrast to the intuition that “spacing is the enemy of induction” and other research demonstrating the importance of comparing exemplars from the same category (Gentner & Namy, 1999; Oakes & Ribar, 2005; Rittle-Johnson & Star, 2011). Second, and most pertinent to the present research, the Kornell and Bjork results showed that even after receiving feedback on the test, participants still overwhelmingly believed that blocking was more effective for learning artist styles’ than was interleaving. Both of these results have now been replicated and extended (e.g., Birnbaum, Kornell, Bjork, & Bjork, 2013; Kang & Pashler, 2012; Kornell, Castel, Eich, & Bjork, 2010; Wahlheim, Dunlosky, & Jacoby, 2011; Zulkiply, McLean, Bath, &
Burt, 2012), although studies of the mechanisms underlying the interleaving benefit outnumber the studies that have focused on the metacognitive failure to appreciate interleaving.

Kornell et al. (2010) found that although both younger and older adults benefited from interleaving, the majority (56% and 75%, respectively) judged blocking to be more effective. In fact, the experiments reported in the present Chapter 2 illustrate how compelling the illusion that blocking enhances inductive learning is: Despite going to great lengths to reveal to people that they had learned interleaved artists better to explain the interleaving benefit, the majority of participants still stubbornly clung to the belief that blocking is more effective for themselves.

Going beyond post-test self-reports of which schedule was considered to be more effective, Tauber, Dunlosky, Rawson, Wahlheim, and Jacoby (2013) examined the a priori theories about effective category learning that participants may bring with them into the lab. Across four different experiments, Tauber et al. (2013) asked participants to learn to recognize different bird families by studying six unique exemplars from each category. Participants were asked to choose on a trial-by-trial basis the category (bird families) they wanted to study next, and they were allowed to study for as long and as many trials as they wished. In their first three experiments, participants were additionally first presented with every exemplar in an interleaved manner (as a familiarization phase) before being asked to restudy the category exemplars in the order of their choosing. Participants’ schedules were analyzed: Blocking was classified as studying two or more exemplars from the same bird family in a row and interleaving strictly referred to switching between different bird families (switches between blocks, however, were not counted as interleaving). If more than half of the exemplars were blocked, the participant was designated as a “blocker”. Across the four studies, between 78-100% of participants were classified as blockers.
Tauber et al. (2013) also reported average run lengths (in other words, the average number of exemplars from the same bird family that participants chose to study successively) and average total number of exemplars studied. In the first three studies, where participants were given a familiarization phase, the average run length was 6, 7.6 and 6, respectively. In other words, participants essentially exhausted study of all the exemplars of a given category before moving on to the next one. Additionally, the average total number of trials only barely exceeded the total number of unique exemplars (for example, in their second experiment, there was a total of 48 unique exemplars—6 from each of 8 categories—and participants studied an average of 58 total exemplars). In their fourth experiment, however, where there was no familiarization phase, the average run length dropped to 3.7, and participants studied an average of 107 total exemplars. In other words, participants were likely studying multiple spaced runs of any given category, instead of studying the exemplars of a given category in one single long block.

In light of the results presented in Chapter 3—that certain hybrid schedules of learning can be as effective as pure interleaving—it seems prudent to examine learner’s study choices more closely before closing the book on participants’ metacognitive ineptitude. Thus, in the present Experiments 1-3, we took a closer look at how participants chose to schedule their own learning, when there were 6 exemplars per artist (Exp 1), 20 exemplars per artist (Exp 2), and for courses/modules as one might study in school (Exp 3). One particular difference between our studies and those of Tauber et al. (2013)—other than the coding and analysis of participants’ schedules—is that our participants were self-scheduling hypothetical future learning and were not concurrently engaging in learning and scheduling. In Experiment 4, we then examined whether prior experience in learning artist painting styles in either a blocked or interleaved
manner would affect subsequent self-scheduling, and whether self-scheduling would differ for new learning (never-before-studied artists) versus repeated study of previously studied artists.

**Experiment 1**

Drawing purely on their metacognitive beliefs and knowledge—without being presented with any examples of stimuli and without concurrently attempting to learn categories—how do participants choose to schedule their own learning? In Experiment 1, we presented participants with a hypothetical artist learning scenario and then asked them how they would schedule the learning of six artists’ painting styles, with a study set of six exemplars per artist.

**Method**

**Participants and design.** Twenty-nine undergraduates from the University of California, Los Angeles (UCLA) participated in this study for partial course credit. Flashcards were set out in six piles of six cards in front of the participants in one of two arrangements: Each pile represented the six paintings by each artist (block-promoting; \( n = 15 \)), or each pile contained one “painting” per artist (block-discouraging; \( n = 14 \)). Participants were alternately assigned into the block-promoting and block-discouraging conditions.

We aimed to recruit 30 participants, but terminated the study at the end of the academic quarter.

**Materials.** Thirty-six 3” x 4” index cards were used, with each card representing one painting by a given artist. No paintings, however, were actually shown. Instead, on the front of each card was an artist name, written large to fill up the majority of the blank space. Each artist name was written in a different color ink, so as to highlight the different names.
There were six cards for each of the six artist names: Hawkins, Lewis, Foster, Cross, Oliver, and Juras. These names were randomly chosen from the subset of artists that had been used in prior interleaving studies and were unfamiliar to all the participants.

These 36 cards were separated into six piles of six cards and arranged on the desk in front of the participant into two rows of three piles. In the block-promoting condition, each pile contained six cards with the same name. It would therefore be easy to use the “paintings” to study in a blocked schedule by simply picking up the whole pile. In the block-discouraging condition, each pile contained one of each of the six artist names. Furthermore, the order of the names in each pile was randomly shuffled, with the constraint that the top card must always be a different name. Therefore, in order to organize his/her study in a blocked schedule, a participant would have to root through each of the piles to find the appropriate name.

**Procedure.** The cards were first laid out in front of the participants, and then participants were given an instruction sheet that described the task (shown on the first page of Appendix C). They were told to pretend they were art students trying to learn painting styles of six different artists by studying six paintings by each artist. The instructions also listed the six names, and described that each index card represented one painting by one of the six artists. Participants were instructed that they were to create one single pile of these cards that would represent the order in which they would study them.

After the participants read the instruction sheet, the experimenter then re-explained the instructions verbally to make sure that the participants understood them. Participants were allowed to take as long as they needed to create their order of cards. Once finished, they gave their pile of cards to the experimenter who confirmed which index card was intended to be the first card and then recorded the order each participant had created.
Finally, participants filled out a questionnaire (shown on the remaining pages of Appendix C) that asked them to describe the strategy behind how they scheduled their study of artists. They were also provided with a list of possible study orders that they could have chosen and asked both to indicate which order their own most resembled and which order they thought would lead to the best learning of artists’ styles. Every part of the study was self-paced.

Table 4.1

*The number of participants who scheduled each order and judged each order to be the most effective for learning artist styles, in Experiment 1*

<table>
<thead>
<tr>
<th>Order</th>
<th>Categorized Schedule</th>
<th>Judged most effective order</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pure blocked</td>
<td>24</td>
<td>9</td>
</tr>
<tr>
<td>Familiarize with one painting by each artist, then block study</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Study two or three from each artist at a time</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Blocked-to-interleaved</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>Interleaved-to-blocked</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pure interleaved</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

**Results and discussion**

The results of Experiment 1—how participants scheduled their study (as categorized by the experimenters into the possible study orders provided in the post-test questionnaire) and which order participants thought would be most effective—are presented in Table 4.1, organized from the most blocked to the most interleaved schedules.
Scheduled orders. Across both card-arrangement conditions, participants showed an overwhelming tendency to schedule their study in a blocked order. Twenty-four out of the 29 participants—all 14 participants from the block-discouraging condition and 10 from the block-promoting condition—chose a purely blocked schedule, choosing to study all six paintings by a given artist before moving on to the next artist. In other words, even when the decks were arranged to make blocked scheduling more difficult, participants went out of their way and spent more time in order to create a less effective study schedule!

Of the remaining five participants who did not choose a purely blocked schedule, one chose to study three paintings by a given artist at a time, one chose to study two paintings by a given artist at a time, one chose to first study one painting by each artist before alternating study in pairs (i.e., ABCDEF-ABABABABAB-CDCDCDCDCD-DFDFDFDFDF, where each unique letter corresponds to one artist), and only two chose to study the artists all mixed up (i.e., a “pure” interleaved schedule).

Post-test questionnaire. When asked to pick (from the list) the order that would be most effective for learning, the most popular response (n =11, or 42%) was blocked-to-interleaved, followed closely by pure blocking (n = 9, or 29%). Only one person chose pure interleaving.

Consistent with the results of Tauber et al. (2013), whether we are looking at how people schedule their own learning or how they choose from a list of possible study orders, people have an overwhelming preference for blocked study. That a large proportion of participants judged the blocked-to-interleaved schedule as being the most effective suggests that learners may see some benefits to interleaving, even though most participants did not spontaneously include any form of interleaving in their own schedules.
Experiment 2

Experiment 1 showed that participants overwhelmingly chose to study all six exemplars by a given artist in a row—in other words, blocked their study. Their post-test questionnaire responses, however, indicate that perhaps learners have some understanding that purely blocking is not the optimal method of study. One possibility for the discrepancy between their scheduled order and what they report to be the most effective is that six exemplars is simply not enough for learners to feel comfortable in using a schedule that includes interleaving. It may be that participants would engage in more metacognitively sophisticated scheduling if they were given more exemplars with which to work. In Experiment 2, therefore, we gave participants the chance to schedule the study of 20 exemplars per artist. Will learners now show a greater tendency to interleave (and if so, when does interleaving first appear), or will learners be so adamant in their blocking preference that they will go so far as to block 20 exemplars in a row?

Method

Participants. Fifty UCLA undergraduates participated for partial course credit. The results from two participants were thrown out, as they did not follow instructions and scheduled an incorrect numbers of exemplars per artist.

Materials and procedure. Participants were first given a single page that contained the instructions for the task and 10 rows of 10 cells each (as shown in the first page of Appendix D). Participants were given the same cover story as in Experiment 1: they were preparing to take an art exam, for which they would have to identify the painter responsible for new paintings. In contrast to Experiment 1, there were only five artists in Experiment 2, and participants had to schedule the study of 20 paintings by each artist. Instead of organizing a stack of flashcards, participants had to write one of the letters A through E (where each letter represented one of the
five artists) and one number between 1 and 20 (where each number represented one of the 20 paintings by a given artist) in each cell. Participants were instructed to work from left to right, top to bottom, and gaps were printed between each row to clarify this order.

Once completed, participants were handed two more pages of paper containing a post-test questionnaire and demographics questions (as shown in the remaining pages of Appendix D), where they described the rationale behind their scheduled order, selected the sample schedule that best fit their scheduled order, the sample schedule that they believed would be most effective for learning, were forced to choose between either a pure blocked or pure interleaved schedule, and were required to explain why they would block and why they would interleave paintings. The latter two questions were added into the study late, and thus, were not asked of the first nine participants. Completion of this study was self-paced, but participants generally took between five to ten minutes.

Table 4.2 

The number of participants who scheduled each order and judged each order to be the most effective for learning artist styles in Experiment 2

<table>
<thead>
<tr>
<th>Order</th>
<th>Categorized Schedule</th>
<th>Judged most effective order</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocked</td>
<td>14</td>
<td>8</td>
</tr>
<tr>
<td>Familiarize, then block</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>Ten at a time</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>Five at a time</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>Block to interleave</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>Interleave to block</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Pairs or triples</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Interleaved</td>
<td>3</td>
<td>0</td>
</tr>
</tbody>
</table>
Results and discussion

Scheduled orders. The results of Experiment 2—how participants scheduled their study (as categorized by the experimenter into the possible study orders provided in the post-test questionnaire) and the order they thought would be most effective—are presented in Table 4.2, organized from the most blocked to the most interleaved schedules.

![Graph](image)

*Figure 4.1.* The maximum run length (i.e., maximum number of paintings by the same artist in a row) in the presentation sequences constructed by the participants in Experiment 2.

Most striking is the fact that even with 20 exemplars per artist, 14 (29%) of participants still chose to study all 20 paintings in a row, and an additional two participants chose to study either 18 or 19 paintings by one artist in a row (“familiarize, then block”). As is evident from Table 4.2 and from Figure 4.1—which illustrates the maximum run lengths of the participants (i.e., the longest consecutive block of paintings by the same artist)—participants had a tendency to block study paintings by the same artist. Very few people chose a pure interleaving strategy.
(where two paintings by the same artist are never studied consecutively) or even chose to switch up artists every two or three paintings.

It is interesting, however, that when asked to select from the list of possible orders the one that would be most effective for learning artist styles, 52% (25 out of 48) of participants chose the blocked-to-interleaved combination. That is, even if they did not spontaneously generate a schedule that included interleaving, people still seemed to realize that there could be benefits to both blocking and interleaving.

In fact, despite an overall tendency to block paintings, the 34 participants who did not simply block all 20 paintings of an artist were strategic in their scheduling. For instance, some chose to see one painting by each artist at the beginning to familiarize themselves with the range of painting styles. Others chose to interleave paintings by different artists in the middle or at the end in order to test themselves or to examine the differences between the artists. Examples of the rationale that participants gave for scheduling their study are shown in Appendix E.

**Blocking vs. interleaving.** When required to make a forced choice between studying all the paintings in a blocked manner or in a fully interleaved manner, 32 out of 39 (82%) participants responded that it would be better to study the paintings blocked by artist, highlighting the fact that participants have an overwhelming tendency to believe that blocking is more useful for learning concepts than is interleaving.

Responses to the questions, “If/When you put paintings by the same artist next to each other, why did you do so?” and “If/When you mixed up the paintings by different artists, why did you do so?” highlighted again, however, that participants are not simply blindly blocking their study, but rather hold some metacognitively more sophisticated theories regarding the benefits of blocking and interleaving: 30 (77%) participants responded that they blocked in order
to see the similarities within an artist’s paintings, and 20 (51%) responded that they interleaved in order to learn the differences between the different artists’ styles—these responses are consistent with hypotheses that cognitive researchers have advanced as being the relative benefits of blocking and interleaving (e.g., Carvalho & Goldstone, 2013; Kang & Pashler, 2012; Kornell & Bjork, 2008). Seven (18%) participants also reported that interleaving paintings by different artists was useful as a way of self-testing whether the styles had been learned.

Taken together with the results of Experiment 1, the present findings demonstrate that although participants hold a general belief that blocking exemplars by category is more effective for learning—a finding consistent with that of previous studies (e.g., Tauber et al., 2013)—they still have some measure of metacognitive sophistication in that they tend to understand that interleaving highlights differences between artists and generally prefer to have some degree of interleaving alongside blocking. The interleaving of paintings by different artists, however, is most likely to be introduced after first studying paintings blocked by artist: Across both Experiments 1 and 2, only two out of 77 participants judged the interleaved-to-blocked schedule as being the most effective, whereas 36 participants judged the blocked-to-interleaved schedule as being the most effective. Similarly, with the exception of a handful ($n = 14$) of participants, almost all ($n = 65$) participants began their study with a block of at least four painting by the same artist.
Experiment 3

Presumably, most of our participants have not had extensive experience in studying and learning artists’ painting styles, and learners may choose to schedule perceptual and cognitive tasks differently. In Experiment 3, we examined how learners would schedule learning in scenarios that were likely to be more similar to their own experiences—namely, preparing for course exams. Rohrer and Taylor (2007) and Taylor and Rohrer (2010) demonstrated in classroom studies that students learn math formulae better when they are practiced intermixed rather than blocked by problem type, in large part because the intermixed practice enhanced discrimination between the formulae and afforded repeated retrieval practice as participants switched back and forth between formulae. Would participants appreciate the benefits of interleaving related modules? Additionally, given that intermixing even unrelated topics introduces spacing (for a review of the literature on spacing effects, see e.g., Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Dempster, 1989), would participants intermix study of unrelated courses?

Method

Participants and design. Ninety-one undergraduate students (58 females; age range: 18-36, mean age = 20.4) from UCLA participated for partial course credit. Participants were asked to schedule study for either four unrelated courses ($n = 49$) or for four related modules within one course ($n = 42$). The data from ten participants were eliminated as they did not follow instructions of scheduling three hours per course or module, leaving 41 participants in the unrelated courses condition and 40 in the related modules condition.

Materials and procedure. Experiment 3 was conducted in a three-page paper packet (as shown in Appendix F). Participants were presented with a hypothetical scenario in which they
were told that over the next five days they would need to prepare to take an exam on either one course with four modules (e.g., if it were a mathematics course, the four modules might be integration, volume, geometry and factorial equations) or on four unrelated courses (e.g., geography, mathematics, psychology and history). These four courses or modules were arbitrarily labeled A, B, C, and D. Participants were further told that 16 hours of study time (i.e., four hours per course/module) were to be allocated over the five days with the constraint that each of the courses/modules were to be studied for at least one hour on the final day (the fifth day). Their task was to schedule the remaining three hours per course/module, with one letter representing one hour of that topic. Participants were reminded that they did not need to schedule study on every single day and they could schedule more study on the fifth day. They were told to assume that all modules or courses were equally difficult and equally important.

Participants were presented with a five-column table, with each column representing the study for each of the five days. One of each letter, A, B, C, and D, was printed in the fifth column. They were allowed to complete this first page at their own pace and were not given the remaining two pages until they had completed this first one.

On the second page, participants were presented with three sample schedules, representing blocked study (i.e., studying all three hours of one subject per day for each of the first four days), interleaved study (i.e., studying one hour on each subject on each of days 2-4) or crammed study (i.e., studying only on the last two days). Finally, on the third page, participants provided demographics details, including age, sex, year in college, and GPA range.

**Results and discussion**

The variance of each course or module was calculated for each participant, and the variance of all four courses/modules averaged (using the numbers of the days on which each
course/module was scheduled). Results showed that the average variance in the related modules condition \((M = 0.91, SD = .80)\) was significantly lower than the average variance in the unrelated courses condition \((M = 1.31, SD = .56)\), \(t(79) = 2.62, p < .05, d = .59\), indicating that participants were more likely to interleave their study in the unrelated courses condition than in the related modules condition.

We categorized each participant’s schedule into one of four schedules: Pure blocked, 2-then-1 (study for two hours, then review for one hour), 1-then-2 (study for one hour, then study for two hours straight), and pure interleaved. The proportion of participants whose schedules fell into each of these four categories are represented in the left half of Figure 4.2. The most commonly used schedule in the unrelated courses condition was pure interleaving (68% of participants) while the majority of participants in the related modules condition were split between pure interleaving (43%) and pure blocking (43%). In other words, people were more likely to block in the related modules condition and more likely to interleave in the unrelated courses condition.

The same pattern was shown in participants’ sample schedule selection (shown in the right side of Figure 4.2): 69% of those in the unrelated courses condition selected the interleaved schedule with only 26% selecting the blocked schedule. Conversely, in the related modules condition, 71% chose the blocked schedule and only 26% chose the interleaved schedule. This reversal of the pattern of responses was significant, \(\chi^2(1) = 15.62, p < .001\). Only three participants (two from the unrelated courses condition and one from the related modules condition) chose the crammed schedule, and for this reason, the cram response was removed from the chi-square analysis.
Figure 4.2. Proportion of participants—by condition—whose schedules fell into each of the categories (left), and who selected each sample schedule (right).

It appears that participants have a strong tendency to believe that unrelated courses should be interleaved while related modules should be blocked. The desire to interleave unrelated courses does not, however, necessarily underscore an appreciation for any benefits of spaced or interleaved learning, but rather, based upon participants’ explanations about their scheduling, seems to reflect a desire to take “breaks.” On the other hand, participants’ blocking of modules seems to underscore a theory that reflects traditional classroom and textbook organization: focusing on one thing before learning the next.
Experiment 4

Experiments 1-3 demonstrated that learners, by and large, have a strong preference to block to-be-learned materials. If the fact that blocking is prevalent in the real world is a part of the reason why learners show such a strong preference for blocking, then it is plausible that simply exposing participants to a different schedule of study might attenuate the blocking preference. Given that learning artist paintings styles is a relatively unfamiliar task for the majority of learners, we returned to the painting stimuli, and tested whether simply exposing participants to one study-test cycle of interleaved study would shift their “default” for scheduling their own future learning away from blocking. Additionally, given the preference (as observed in Experiments 1 and 2) for introducing some form of interleaving late in the course of studying, we examined whether participants strategies differed between studying artists for the first time compared to re-studying artists for which they should already have built up some knowledge.

Method

Participants and design. Ninety-four (72 females; mean age = 20.83, range = 17-42) UCLA undergraduates participated for partial course credit. The experiment was conducted in two phases. In the first phase, participants studied six paintings by each of eight artists in either a blocked or interleaved fashion and were then tested on them. In the second phase, participants were asked to self-schedule either the same set of eight artists or a new set of eight artists, studied the artists in that order, and were then tested on the studied artists. Initial schedule (blocked vs. interleaved) and second phase study set (repeated artists vs. new artists) were manipulated between-subjects.

Materials and apparatus. Two randomly assigned sets of eight artists were used: One set consisted of landscape paintings by Jamie Grossman, Judy Hawkins, Philip Juras, Ryan
Lewis, Richard Lindenberg, Richard McKinley, YieMei, and Bruno Pessani; the other set consisted of landscape paintings by Tom Brown, Maryanne Jacobsen, Karen Margulis, Marilyn Mylrea, Lori McNamara, Julie Ford Oliver, Marina Petro, and Ron Schlorff. Six paintings per artists were studied, and two paintings per artist were used in each of the tests. In the conditions where participants restudied the same set of artists in the second phase, the same six paintings were restudied. The tests, however, were always on never-before-seen paintings.

The entire first phase and the test of the second phase were conducted via an Internet browser (programmed using PHP-based psychology experiment platform, Collector, github.com/gikeymarcia/Collector). In the second phase, participants self-scheduled their order using 4x6 inch index cards. Six index cards per artist were created such that the names were written in large print on the backside, and one painting was printed and pasted on the other side, with the name re-written again above each image. As in prior studies, each artist name was written in a different colored ink. As people self-studied these flashcards, a PowerPoint was created to chime every four seconds, alerting participants to move onto the next painting.

**Procedure.** The study took part in two phases. In the first phase, participants were told that they would study six paintings by each of eight artists and that their goal was to learn each artist’s painting style. They were informed that on the final test, they would be shown new paintings by the eight artists and would have to identify, from a list of names, which artist painted each painting. Each painting was presented for three seconds and presented in the middle of the computer screen with the artist’s last name written underneath each image. Paintings were either presented blocked by artist (the order of the specific paintings and the order of the artists were randomized between participants) or the paintings by different artists were interleaved (the paintings were divided into six blocks of eight images—with each block consisting of one
painting per artist, and the order of paintings within each block randomized). After studying all 48 images, participants were tested: Two new paintings by each artist were presented (paintings randomized in two blocks of eight paintings, where each block consisted of one painting per artist), and the eight names were presented below each painting in two rows of four names (randomized order between participants). Participants selected the name of the artist they believed was responsible for each painting by clicking on their name, and they were given feedback on the correct answer.

After they completed the test for the first phase, participants were instructed to call the experimenter for instructions on the next phase. Participants were handed an instruction sheet for the second phase (see Appendix G), and the experimenter also repeated instructions verbally, checking that participants understood instructions. Participants were told that they would study the same artists as in the previous phase (repeated study condition) or a new set of eight artists (new study condition), but that this time, they would be able to choose the order in which the paintings would be studied. Eight piles (one for each artist) of index cards were placed image side down in front of participants. Without turning the cards over to look at the images, participants were asked to create a single pile representing the order in which they wanted to study the paintings (with the first to-be-studied painting on the bottom—that way, when the whole pile was flipped around, the painting would be on the “front”). To incentivize strategic thinking, participants were told that they would study the paintings in the order they had arranged and would score one point for every painting that they correctly classified on the final tests and to aim to score as many points as possible.

Once they had created their single pile, participants were told that they would now study the painting on each card for three seconds each, and would be alerted by a chime when it was
time to put the current painting down and to pick up the next one. They pressed “Enter” on the keyboard when they were ready to begin. In reality, a chime occurred every four seconds to allow time to put one card down and focus attention on the next image. Participants then returned to the Internet browser and completed the final test on their second set of artists.

Finally, participants completed a post-test questionnaire (shown in Appendix G) that asked the same questions that were asked in Experiment 2: The strategy they used in scheduling their study, selecting from a list of possible orders the one they that most resembled their own and the one that would be most effective, a forced choice between blocking and interleaving, and the reasons why they would block or interleave paintings.

Results and discussion

Phase 1 classification performance. In the first phase, participants were randomly assigned to either a blocked or interleaved study schedule. Those who studied the artists interleaved classified significantly more paintings correctly ($M = .59, SD = .19$) than those who studied the artists blocked by painting ($M = .44, SD = .20$), $t(93) = 3.83, p < .001, d = .79$.

Phase 2 scheduling. Participants’ schedules were classified in two levels of specificity. In the broader strategy classification, participants’ self-scheduled orders were classified according to whether they predominantly blocked study or predominantly interleaved study. If the average run of paintings by the same artist was less than three, the schedule was classified as predominantly interleaved. If the average run was three or greater, then the schedule was classified as predominantly blocked.

In the more specific strategy classification, the self-scheduled orders were classified according to whether they were purely blocked, purely interleaved (as defined by never placing two paintings by the same artist in a row), whether the average run of paintings by the same artist
was equal to or greater than 3 (but less than 6; in other words, mostly blocked), whether the maximum run of artists was 2 (in other words; mostly interleaved), or whether participants used runs of three, two and one (i.e., some combination of blocking and interleaving—either blocked-to-interleaved or vice-versa). Examples of the schedules that fall into each of these broad and specific categories and the number of participants whose schedules fell into each category are shown in Table 4.3.

Table 4.3

*Broad and specific strategy classifications in Experiment 4, examples of schedules, and the number of participants whose schedules fell into each category*

<table>
<thead>
<tr>
<th>Broad Strategy</th>
<th>Specific Strategy</th>
<th>Example Schedules</th>
<th>Number of Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predominantly blocked</td>
<td>Pure blocking</td>
<td>Ax6, Bx6, Cx6…</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>Average run 3+</td>
<td>(AAABBBCC…) x2 ABCD…Ax5, Bx5…</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ax5, Bx5…ABCD…</td>
<td></td>
</tr>
<tr>
<td>Predominantly</td>
<td>Maximum two in a</td>
<td>(AABBCC…) x 3 AABBCC…(ABC…) x4</td>
<td>8</td>
</tr>
<tr>
<td>interleaved</td>
<td>row</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Combination</td>
<td>AAABBBCCC…AABBCC…ABC ABC…AABBBCC…AAABBCCC</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Pure interleaving</td>
<td>(ABCDEFGH) x6 (AB) x6, CD(x6)…</td>
<td>11</td>
</tr>
</tbody>
</table>
Figure 4.3. Proportion of participants within each initial schedule x second phase study set condition that predominantly blocked or predominantly interleaved their phase 2 study, in Experiment 4.

Figure 4.3 shows how the broad strategy classification of the self-schedules was affected by experimental condition. Within the participants who had been presented with paintings in a blocked schedule in the first phase, the 34 out of 46 participants (or, 74%; evenly split between the “new” and “repeated” study conditions) also predominantly blocked their study in the self-scheduled second phase. Only six from each of the new and repeated study set conditions predominantly interleaved their study. Within the participants in the initially blocked condition then, phase 2 scheduling and study set were independent, $\chi^2(1) = 0.00, p = 1.00$.

Within the participants who had been presented with paintings in an interleaved schedule in the first phase, 35 out of 48 participants (73%) chose to predominantly block their study in the second phase. The patterns of predominantly blocking versus interleaving, however, differed
depending on the study set condition, $\chi^2(1) = 8.47, p < .01$: Whereas 19 out of 20 participants (95%) chose to predominantly block in the repeated study condition, only 16 out of 28 (57%) chose to predominantly block in the new study condition. It appears therefore, that having studied the artists in an interleaved manner, those in the repeated study condition now want to study it differently (blocked), whereas those in new study condition are less likely to anchor their “default” schedule from a predominantly blocked schedule. There was no difference in phase 1 classification performance between the participants who predominantly blocked ($M = .51, SD = .20$) and those who predominantly interleaved ($M = .54, SD = .23$), $t(92) = .67, p > .05, d = .15$. There were also no differences in reported GPA, number of classes skipped per week, hours of study per day, or theory of intelligence (the full individual differences data for blockers and interleavers are found in Appendix H1).

Analyzing the specific schedule classification data by initial schedule and study set conditions tells a similar story: Within the blocked initial schedule condition, participants in the new and repeated study conditions used pure blocking (14 out of 23, or 61%, in both study set conditions) or scheduled with an average run length of at least three paintings (three out of 23 in both study set conditions, or, 13%) equally often. On the other hand, within the interleaved initial schedule condition, fewer participants in the new study set condition ordered the paintings in a purely blocked manner (46% vs. 75%) or scheduled with an average run length of at least 3 paintings (11% vs. 20%) as compared to those in the repeated study set condition. Cell counts were too low to conduct statistical analyses, but the full data of specific schedule categories by experimental conditions are presented in Appendix H2.

In the post-test questionnaire, participants were asked which of the listed schedules they thought would lead to the best learning. Consistent with the results of Experiments 1 and 2,
participants tended to favor the blocked schedules with more blocked study: By far the most popular option was the one that combined blocking and interleaving (the specific description described blocked-to-interleaved study, but the interleaved-to-blocked counterpart was not provided as an option), with 40% of participants choosing this schedule. Pure blocking was selected by 24% of participants, and familiarization with one painting per artist and then a block of five paintings by the same artist in a row was selected by 14% of participants. Overall then, these three schedules accounted for 78% of responses. Only 8% chose a pure interleaved schedule and 14% wanted to study the paintings in pairs or triples.

**Blocking versus interleaving.** When forced to make a choice between pure blocking and pure interleaving, a majority 71 out of 93 respondents (one person failed to answer this question) indicated pure blocking. Similar to the self-scheduling results, however, these responses interacted with experimental condition: among the participants in the blocked initial schedule condition, 83% of both study set conditions selected pure blocking, whereas among participants in the interleaved initial schedule condition, 52% of the new study set and 95% of the repeated study set conditions selected pure blocking.

Finally, when we asked participants why they blocked (put two paintings by the same artist in a row), 66 participants (70%) said they liked to see paintings blocked by artist because it allowed them to see similarities within a given artist’s paintings, and 8 participants (9%) said that blocking was just easier for memorization. When we asked participants why they interleaved (put paintings by different artists next to each other), 24 participants (26%) said that interleaving helped highlight differences between artists (a couple of participants even explicitly expressed wanting to interleave to contrast the more confusable artists), 10 (11%) stated that they interleaved to test themselves (typically at the end of study), and 8 (9%) used interleaving to
space out study or used interleaving as a way of refreshing their memories. Forty-seven participants, however, either left the space for this interleaving question blank, wrote “N/A” or “I didn’t”—suggesting that they do not think that interleaving is helpful (in contrast, only 11 participants left the space for the blocking question blank).

**Phase 2 classification performance.** Although participants self-scheduled in this phase and were not randomly assigned to the predominantly blocked or predominantly interleaved conditions, the pattern of performance, by phase 2 schedule, and experimental conditions (initial schedule and study set) are shown in Figure 4.4 (note that the right-most bar has no error bar, as only one participant in the initially blocked-repeated study set condition predominantly interleaved his or her study in phase 2). A 2x2x2 between-subjects ANCOVA analyzing the effects of initial study schedule, phase 2 study set condition and phase 2 schedule, with phase 1 classification performance covaried out, revealed only a main effect of study set condition and a main effect of phase 2 schedule. Those in the repeated study set condition (\(M_{adj} = .65, SE = .05\)) performed significantly higher than those in the new study set condition (\(M_{adj} = .51, SE = .03\)), \(F(1,85) = 7.57, MSE = .03, p < .01, \eta^2_p = .08\), and the interleavers (\(M_{adj} = .64, SE = .05\)) significantly outperformed the blockers (\(M_{adj} = .52, SE = .02\)), \(F(1,85) = 5.72, MSE = .03, p < .05, \eta^2_p = .06\).
Figure 4.4. Experiment 4’s phase 2 classification performance (adjusted means, after covarying out phase 1 performance), by initial study schedule, phase 2 study set and phase 2 schedule.

General Discussion

Experiments 1 and 2 revealed that although participants generally believe that blocking is effective, a good proportion of participants spontaneously shy away from pure blocking when there are a lot of study exemplars (in Experiment 2, only a third of participants purely blocked study of 20 exemplars per artist, as compared to the 83% that purely blocked the study of six exemplars per artist in Experiment 1). The post-test questionnaire results of Experiments 2 and 4 also show that the majority of participants—once presented with a list of possible schedule options—will report that some combination of blocked and interleaved study (in particular, mini-blocks or blocked-to-interleaved) is in fact the most effective study schedule. When asked directly why they would block or interleave study, participants were very much able to express
that blocking helps them spot the commonalities, and a sizeable proportion was additionally able to indicate that interleaving would help them distinguish between the artists. These questions were framed as open-ended questions rather than as multiple-choice questions, making the number of participants who were able to express these theories more impressive. Taken together, these results provide a more nuanced perspective of what learners understand about their own learning as compared to what prior studies might suggest. In fact, the two hybrid schedules experiments reported in Chapter 3 show that choosing a blocked-to-interleaved or mini-blocks schedule leads to learning that is just as effective as purely interleaving study. The participants who choose either of these hybrid schedules might then, be considered to be metacognitively accurate.

Interestingly—and consistent with the test performance results reported in Chapter 3—participants show a large preference for blocked-to-interleaved study over interleaved-to-blocked study. The reasons participants preferred the former over the latter may not necessarily align with the reasons why the former is empirically more effective than the latter. Although we did not directly question participants about their choice between blocked-to-interleaved and interleaved-to-blocked schedules, the impression we got from their open-ended response about their own schedules is that many participants used interleaving at the end to test themselves (and thus, interleaving should only be introduced after “learning). This reasoning would be consistent with findings from Kornell and Son (2009) that showed that people choose to test themselves at the end of study (instead of restudying) out of a desire to diagnose their level of learning, rather than out of an appreciation of the testing effect.

Experiment 3 showed that participants’ scheduling is sensitive to the relationships between the to-be-learned topics: when preparing for related modules, participants were more
likely to block study by module, but when preparing for unrelated courses, to interleave study of unrelated courses. The research thus far on sequencing effects in category learning and contextual interference, however, would support the opposite—that related modules should be interleaved. The research on how best to sequence unrelated topics, however, is yet unclear. One speculation is that the learning of each topic should still benefit from the spacing inherent to interleaving, although spacing studies typically examine restudying the same content between presentations, rather than the studying of different/new portions of a given topic. Nonetheless, Experiment 3 revealed that even with a very different nature of learning tasks, there is a general tendency to block study of related concepts—potentially as a result of how formal education tends to structure instruction.

To examine the role of prior experiences, Experiment 4 showed that a prior exposure to interleaved study of one set of artists can change how learners go about learning a new set of artists: Those who had experienced interleaved study were less likely to block their self-scheduled study of a new set of artists than those who had experienced blocked study. Those who were most likely to block their self-schedules, however, were those who were re-studying the same sets of artists that had been initially interleaved. The reasons for this latter finding are unclear, though we might make some speculations: Perhaps they wanted to vary how they studied the artists, and therefore switched from interleaved to blocked—but this explanation would not explain why the people who had initially studied artists in a blocked schedule would stick with a blocked schedule in the self-schedule phase. Perhaps they simply decided that interleaving felt too difficult and therefore switched to more of a blocked strategy—but this explanation would not account for why those scheduling learning for a new set of artists after
exposure to interleaving were more likely to then begin new learning with a predominantly interleaved schedule.

Regardless, these studies illustrate that learners do not blindly block their study. Rather, they suggest that we can better appreciate what learners understand by more closely examining learners’ metacognitive beliefs, and that learners may be further pushed away from blocked study when prior learning begins with interleaving.
CHAPTER 5
Concluding Remarks

Summary of Results

The ten experiments presented in Chapters 2 through 4 reveal first that it is difficult to convince people that interleaving study of exemplars from different categories is better for category induction than is blocking study of exemplars by category (Chapter 2). There are, however, other schedules that fall between pure blocking and pure interleaving that learners are very much willing to accept, some of which may be just as effective for inductive learning as interleaving (Chapter 3). In fact, this series of studies shows that people tend not to believe that only blocking study is beneficial. Even in cases where learners do not spontaneously introduce any level of interleaving into their studies, they are very willing to accept that blocked-to-interleaved schedules or spacing out repeated blocks of exemplars from the same category are more effective than purely blocking study. Together, the metacognitive results from Chapters 3 and 4 reveal a more tempered and fine-grained perspective on learners’ metacognitive strategies for category learning—while the majority of participants do state that blocking highlights the commonalities across different exemplars of a given category, a substantial proportion also were able to elucidate that interleaving would be beneficial for learning to distinguish between artists and/or as a way of testing their learning. Finally, Chapter 4 further showed that people were less likely to purely block study and to be more strategic in their scheduling when given more exemplars to work with, and when they were exposed to a prior interleaving phase for the scheduling of new learning (but not repeated study).
Is Interleaving Always Better Than Blocking?

Throughout the studies presented in this dissertation, there has been an assumption that the interleaving benefit is very robust, and for at least a large range of rich, realistic categories, interleaving is better. Studies using artificial categories, however, have not always found this superiority of interleaving to be the case. Below, I review the materials used and the results of various studies that have compared blocking and interleaving.

**Learning Realistic and Naturalistic Categories.** To date, only a handful of studies have used naturalistic or real-life perceptual categories to examine the relative benefit of interleaving over blocking. These categories include artists’ paintings styles (Kang & Pashler, 2012; Kornell & Bjork, 2008; Kornell, Castel, Eich, & Bjork, 2010), bird families (Birnbaum, Kornell, Bjork, & Bjork, 2013; Wahlheim, Dunlosky, & Jacoby, 2011), butterfly species (Birnbaum et al., 2013) and women’s voices (Yan, Vetter, & Bjork, 2012). These studies have typically used 12 categories (6 studied exemplars per category)—following from Kornell and Bjork’s (2008) original experiment, but have used as few as three categories (with 20 to 24 studied exemplars per category; Kang & Pashler, 2012) or as many as 16 categories (with 4 studied exemplars per category; Birnbaum et al., 2013). In all of these studies, participants passively studied (i.e., simply saw the exemplar with the name written below the exemplar) the categories in either a purely blocked or a purely interleaved schedule; in all these studies, interleaving outperformed blocking.

There is one instance in which interleaving is not better than blocking for the learning of realistic perceptual categories: Carpenter and Mueller (2013) found no difference between blocked and interleaved study in the learning of eight French pronunciation rules (e.g., the pronunciation of the italicized portions in the following words: bateau, fardeau, tonneau,
chardon, bouton, osseux, genou, compris). Participants were tested on their ability to select the correct pronunciation of new French words—participants heard three versions of each new word, one which was completely correct (for both the stem and rule portions of the word), one that was rule-correct (correct pronunciation of the rule portion; incorrect pronunciation of the stem portion) and one that was completely incorrect. Blocking led to better identification of the completely correct pronunciation, but there was no overall difference between the two schedules in rule learning (i.e., combining how often participants picked either the completely correct or the rule-correct option).

**Learning Artificial Categories.** It has been easier to vary, in a controlled and quantitative manner, the category structures of artificial categories, and in doing so, research has revealed factors that moderate the interleaving benefit. Zulkiply and Burt (2013) created two sets of 12 categories, one in which the defining category characteristics were relatively easier to spot (only one irrelevant shape) and one in which the defining category characteristics were relatively difficult to spot (five irrelevant shapes). All categories were defined by three relevant shapes: a circle or a square, and two other shapes that were outlined in one color (blue, green, or red) and filled with another color (green, red, or yellow). Using the same blocking versus interleaving procedure as Kornell and Bjork (2008), Zulkiply and Burt found that there was a blocking benefit for the easier category set and an interleaving benefit for the harder category set.

Garcia, Kornell, and Bjork (in prep) created a set of cartoon fish stimuli in which the categories were defined either by a set of salient features (e.g., body shape, color, and presence of stripes) or by a set of less salient features (e.g., body shape, presence of a dot, and presence of a lower fin). Each participant was tasked with learning to distinguish between six categories of fish (from either the easier set or the harder set). In contrast to Zulkiply and Burt, where the
"easier" set of categories benefited from blocked study, Garcia et al. found that their "easier" set was learned better (at least numerically) in the interleaved condition, while the "harder" set benefited from blocked study.

Finally, Carvalho and Goldstone (2013) similarly investigated the effect of category structure on optimal scheduling. Participants studied six categories of blobs—three interleaved and three blocked. Blocked and interleaved learning took place in two separate study-test cycles, so at any one point, participants were learning to distinguish between just three categories. In one category structure condition, the three categories shared high within- and between-category similarity. In the other condition, there was low within- and between-category similarity. What defined each category was one particular notch in the outline of each blob. As the study phase involved active/feedback training (i.e., participants had to classify the stimuli throughout the study phase and learned via feedback), frequent (75% alternations) and infrequent (25% alternations) alternation schedules were compared (as opposed to purely interleaved and blocked schedules). Results showed that frequent alternation (or, "interleaving") was better for learning the high similarity categories while low alternation (or, "blocking") was better for learning the low similarity categories.

**When and Why does Interleaving Enhance Inductive Learning?**

There appear to be three different, but not mutually exclusive, mechanisms by which interleaving confers benefits:

1. Interleaving exemplars from different categories highlights the differences between categories, thereby enhancing discrimination. Enhanced discrimination is particularly useful when categories are similar and is perhaps less useful when the main
variability among the stimuli arise from within-category differences. In contrast, blocking may enhance the noticing of commonalities within a category.

2. Interleaving may discourage the use of an explicit, rule-based, hypothesis-testing learning strategy by making it difficult for learners to search for common features among the exemplars of each category and to hold these hypotheses in working memory; in doing so, interleaving may allow for more effective information-integration learning. In contrast, blocking should make the discovery of rules that define each category easier and promote the explicit search for commonalities.

3. The spacing inherent to interleaving will benefit any components of memorization—for example, learning the association between the induced category and its label, or the learning (but not discovery) of rules.

How might we tie all of these mechanisms of category learning (within/between category similarity, learning strategy, and memory load) together? Table 5.1 summarizes the existing studies, and how their materials might be described along the lines of category structure, learning strategy, and memory load. It is clear from this table, however, that more research investigating the dynamics between these mechanisms, and how they contribute to optimal scheduling is needed.

---

1 Ashby, Maddox, and colleagues (e.g. Ashby & Maddox, 2005; Ashby, Alfonso-Reese, Turken, & Waldron, 1998) have distinguished between two category learning systems—explicit, rule-based learning (in which categories can be distinguished with verbalizable rules) and implicit, information-integration learning (where categories are distinguishable only through the integration of multiple dimensions, and which is not easily verbalizable).
Some speculations of when interleaving is beneficial might be proposed. When category stimuli are very rich and require integration of multiple dimensions and when between-category contrasts are relatively more important (compared to within-category comparisons), interleaving should be better than blocking. Most of the studies using rich stimuli do indeed show an interleaving benefit. The only one that does not—Carpenter and Mueller (2013)—is a case where it is less clear-cut as to whether the categories (French pronunciation) require information-integration or can be learned through rules (e.g. “eau” sounds like “o”).

The study that needs to be conducted is one in which within- and between-category similarity of rich, real, and information-integration categories is manipulated within the same study. How could similarity be manipulated in such a way using real-life categories? The existing set of artist stimuli—in which all paintings are of landscapes—might represent a case in which there is both high within- and between-category similarities. One way perhaps of reducing within-artist similarity while keeping between-artists similarity relatively high is to create a set of paintings where every artist paints the same set of different objects. For example, every artist’s portfolio could feature a bicycle, a flower, a building, and a face, creating low within-artist similarity, but because every artist has a painting of each of these objects, the between-artist similarity is relatively high. In this case, noticing the commonalities across the different paintings by the same artist might become relatively important. Whether this type of category structure will eliminate the interleaving benefit remains, however, an empirical question. It is not obvious that blocking would be better than interleaving. Particularly if there are a large number of to-be-learned categories, then spacing would benefit the learning of the category-label associations, and thus, the benefits of blocking and the benefits of spacing may balance out. In
the case that interleaving and blocking are equally effective for categories with high within-category similarity, the use of hybrid schedules may tease apart these different processes.

When categories are rule-based, we might make different predictions. Blocking study would allow an easy way for learners to notice common features between exemplars from the same category and to hypothesis test as they generate rules. If the rules are very easy to notice and are not too difficult to keep in working memory, then once the rules have been discovered, maintaining and memorizing these rules should benefit from spacing. This mechanism might explain why the easier categories (where the six cartoon fish categories were defined by very salient features) in Garcia, Kornell, and Bjork (in prep) benefited numerically from interleaved study. When rule discovery is moderately difficult and requires study of all the exemplars, then blocking should be more effective for learning than interleaving—this process is potentially why the “easier” categories in Zulkiply and Burt (2013) benefited from blocked study (note that their “easy” categories were much more difficult, more numerous, and therefore also more difficult to keep in working memory than even the “difficult” fish categories in Garcia et al.). Another logical extension of this speculation, however, is that if blocking is only truly beneficial for rule discovery, then if we tell learners what features are important, we might recover a benefit of interleaving (as a result of spaced retrieval of rules). Finally, if learning the rules is too difficult, then learning might be better served using an implicit, information-integration process—and this hypothesis might explain why Zulkiply and Burt’s more difficult categories benefited from interleaving.

These hypotheses are all very testable, and future research should be aimed at not just exploring what schedules are effective for study, but at finding a way of creating a unifying theory that predicts optimal scheduling.
Future Directions in Self-Regulated Category Learning

While there are many situations in which interleaving is more effective than blocking, there are a few cases in which blocking may be more effective. Are participants sensitive to differences in the nature of to-be-learned categories and adjust their learning strategies, or do they universally apply an *a priori* theory that study of exemplars should be blocked (or blocked, then interleaved)? Experiment 3 in Chapter 4 showed some hint that participants may adjust learning—participants were more likely to space out study of unrelated courses than of related module (whether this strategy difference is accurate is another matter entirely). Zulkiply and Burt (2013), however, found that for both their easier categories (where blocking was more effective) and more difficult categories (in which interleaving was more effective), participants overwhelmingly judged blocking to be better for inductive learning.

Finally, how can we enhance learners’ metacognitive understanding of their own learning? The studies in Chapter 2 revealed how difficult it can be to convince people through both experience- and theory-based debiasing. There are multiple reasons why learners believe that blocking is more effective for learning: Blocking may lead to a sense of fluency during the course of studying, participants hold *a priori* beliefs that blocking is more effective, and blocking is prevalent in the real world. Persuading learners of interleaving benefits—and indeed of many of the other metacognitive illusions that learners hold (for example, a general belief that learning should be easy, and is benefited by methods that make learning feel easy as opposed to “desirable difficulties,” Bjork, 1994)—will require addressing these multiple sources of metacognitive failure.
Table 5.1

Summary of the features and results of existing studies comparing interleaving and blocking

<table>
<thead>
<tr>
<th>Study</th>
<th>Condition</th>
<th>Categories /Materials</th>
<th># Cats</th>
<th>Result</th>
<th>Within/ Between Category Similarity</th>
<th>RB vs. II</th>
<th>Associative Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Various]</td>
<td>N/A</td>
<td>Artists, birds, butterflies, voices</td>
<td>Mode</td>
<td>I &gt; B</td>
<td>Relatively high between- category similarity</td>
<td>No easy “rules”, integration of multiple dimensions</td>
<td>Need to associate label to induced category</td>
</tr>
<tr>
<td>Carpenter &amp; Mueller (2013)</td>
<td>N/A</td>
<td>French pronunciation rules (word endings)</td>
<td>8</td>
<td>I = B</td>
<td>Low between- category similarity</td>
<td>Unclear…perhaps one could generate rules, like “eau” sounds like “o”</td>
<td>Unclear…if rules are generated, they would benefit from spacing</td>
</tr>
<tr>
<td>Zulkiply &amp; Burt (2013)</td>
<td>Easier</td>
<td>3 defining shapes + 1 irrelevant shapes</td>
<td>12</td>
<td>B &gt; I</td>
<td>More within- category similarity</td>
<td>Rules that are not trivially easy, but manageable</td>
<td>Need to associate rules to label; high WM load (12 rules)</td>
</tr>
<tr>
<td>Zulkiply &amp; Burt (2013)</td>
<td>Harder</td>
<td>3 defining shapes + 5 irrelevant shapes</td>
<td>12</td>
<td>I &gt; B</td>
<td>Less within- category similarity</td>
<td>Difficult rule discovery…may fall back on information-integration learning?</td>
<td>Need to associate rules to label; high WM load (12 rules)</td>
</tr>
<tr>
<td>Garcia, Kornell, &amp; Bjork (in prep)</td>
<td>Easier</td>
<td>Cartoon fish defined by salient features (with less salient features randomly varied)</td>
<td>6</td>
<td>I = B</td>
<td>N/A</td>
<td>Very easy rule discovery</td>
<td>Need to associate rules to labels; manageable WM load</td>
</tr>
<tr>
<td>Study</td>
<td>Condition</td>
<td>Description</td>
<td>N</td>
<td>B &gt; I</td>
<td>I &gt; B</td>
<td>Rule</td>
<td>WM Load</td>
</tr>
<tr>
<td>---------------------------</td>
<td>-----------</td>
<td>--------------------------------------------------------------</td>
<td>----</td>
<td>-------</td>
<td>-------</td>
<td>--------------------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>Garcia, Kornell, &amp; Bjork</td>
<td>Harder</td>
<td>Cartoon fish defined by less salient features</td>
<td>6</td>
<td>B &gt; I</td>
<td>N/A</td>
<td>Manageable rule discovery</td>
<td>Need to associate rules to labels; manageable WM load</td>
</tr>
<tr>
<td>Carvalho &amp; Goldstone</td>
<td>High sim.</td>
<td>All blobs share a general shape; categories defined by a “notch” in a certain location</td>
<td>12</td>
<td>I &gt; B</td>
<td></td>
<td>High within and between category similarity</td>
<td></td>
</tr>
<tr>
<td>Carvalho &amp; Goldstone</td>
<td>Low sim.</td>
<td>All blobs with different general shapes; categories defined by a “notch” in a certain location</td>
<td>12</td>
<td>B &gt; I</td>
<td></td>
<td>Low within and between category similarity</td>
<td>Rule = notch</td>
</tr>
</tbody>
</table>
Appendix A

Judgments of Learning during Blocked and Interleaved Study

This experiment was designed to investigate participants’ feelings of fluency as they experienced blocked and interleaved study. We collected judgments of learning (JOLs) halfway through study and at the end of study for half of the blocked and half of the interleaved artists. Soliciting JOLs for half of the artists allowed examination of the effect of making JOLs during study, and also reduced the number of interruptions during study.

Method

Participants and design. One-hundred and nine undergraduate from the University of California, Los Angeles, participated for partial course credit. The experiment manipulated schedule (blocked vs. interleaved), JOL (JOL vs. no-JOL) in a 2x2 within-subjects design. Two judgments of learning were made for the half of the artists—one immediately after study of the third painting, and one immediately after study of the sixth (last) painting.

Materials. The materials were taken from Kornell and Bjork (2008)—participants studied six paintings by each of 12 artists, and were tested on their ability to classify one new painting by each artist. The assignment of painting to study and test phases was randomized for each individual, as was the assignment of artists to blocked and interleaved conditions, and assignment of artists to JOL and no-JOL conditions.

Procedure. Participants were instructed that they were to learn to recognize the painting styles of 12 different artists so that they could classify new, never-before-studied paintings by these artists on a final test. Participants studied six paintings by each of the 12 artists, half in a blocked manner (all six paintings by the artist were presented consecutively) or in an interleaved manner (the paintings by six different artists were intermixed). The order of presentation was the
same as that of Kornell and Bjork’s Experiment 1A: B I I B B I I B B I I B, where each “B” represents six paintings by the same artist and “I” represents six paintings by six different artists. The order of paintings within each of these sets was randomized. All paintings were presented on the center of the computer screen for 3 seconds, with the artist name written underneath.

Participants were also informed that they would occasionally be asked to judge how likely they would be able to recognize a new painting by an artist on the final test, rating on a scale of 0 (not at all) to 100 (definitely). These JOLs were solicited for half of the blocked and half of the interleaved artists, after the third and sixth paintings had been studied. For half of the participants, these were the 2nd, 4th and 6th blocked artists that were presented (and a random three of the interleaved artists), and for the other half of the participants, these were the 1st, 3rd and 5th blocked artists that were presented.

A 45-second Tetris distracter intervened between the study and test phase. One new painting by each artist was presented in a random order on the final test. For each, participants were asked to select the name (from 12 buttons arranged beneath each painting) the artist they believed responsible for each painting. Feedback (correct artist name) was given for each trial.

Finally, participants were told that they had studied half the artists in a blocked sequence (all six examples by one artist in a row) and half of the artists in an interleaved sequence (all six of that artist’s paintings appeared spread throughout the experiment), and asked: Which schedule do you think led to better learning: Blocked, Interleaved, or They were equally effective.

**Results and discussion**

**Classification performance.** A 2x2 (schedule x JOL) within-subjects ANOVA revealed only a significant main effect of schedule, $F(1,121) = 71.77$, MSE = .08, $p < .001$, $\eta^2_p = .37$. 

93
Participants were better able to classify new paintings by artists that had been studied interleaved (M = .63, SE = .02) than paintings by artists that had been studied blocked (M = .42, SE = .02). There was no main effect of JOL, F(1, 121) = 2.10, MSE = .06, p = .15, $\eta^2_p = .02$, although the pattern was that those artists for whom JOLs had been made were classified at a numerically lower rate (M = .51, SE = .02) than those for whom no JOLs had been made (M = .54, SE = .02). There was also no interaction between schedule and JOL, F < 1.

![Figure A1](image_url)

**Figure A1.** Averaged judgments of learning for the blocked and interleaved artists after study of the third and sixth paintings of each given artist.

**Judgments of learning.** Participants’ judgments of learning for the blocked and interleaved artists after the third and sixth paintings are shown in Figure A1. A 2x2 (schedule x judgment position) within-subjects ANOVA revealed a significant main effect of schedule, $F(1,108) = 19.74, p < .001, \eta^2_p = .16$, a significant main effect of position, $F(1,108) = 139.77, p < .001, \eta^2_p = .56$, and a significant interaction, $F(1,108) = 11.63, p < .01, \eta^2_p = .10$. Participants overall gave higher JOLs to the blocked artists ($M = 66.74, SD = 17.93$) than they did to the
interleaved artists ($M = 59.80$, $SD = 21.18$), and gave higher JOLs after the studying the sixth painting ($M = 69.36$, $SD = 19.11$) than after studying the third painting ($M = 57.18$, $SD = 20$). Post-hoc comparisons showed that the JOLs were greater for the blocked artists than they were for the interleaved artists at both time points, although the difference was much smaller after the sixth painting ($t(108) = 5.36$, $p < .001$, $d = .51$, after the study of the third exemplar; $t(108) = 2.10$, $p < .05$, $d = .19$, after the study of the sixth exemplar).

**Metacognitive judgment.** Despite the fact that judgments of learning were very close by the end of study for the blocked and interleaved artists, 72% of participants believed that blocking was better, with the remaining participants equally split between believing (accurately) that interleaving was better or that the two schedules were equally effective.
Appendix B

Post-test Description of the Five Schedules in Chapter 3, Experiment 1

In this study, there were in fact five different types of study schedules. You just experienced [PARTICIPANTS’ SCHEDULE] schedule of studying.

The five schedules were:

1. **Blocked**: Study one artist at a time (see six paintings by an artist in a row).

2. **Mini-blocks**: Study three paintings by each artist at a time.

3. **Interleaved**: Study all the paintings and artists all mixed up together.

4. **Decreasing blocks**: Start with three paintings by an artist in a row, followed by two in a row, and then all 12 mixed up together.

5. **Increasing blocks**: Start with all 12 artists mixed up, then see two paintings by the same artist in a row, and then see three by the same artist in a row.
Appendix C

Instructions for Chapter 4, Experiment 1

Pretend that you are an art student, and that for your first art exam, you have to learn the painting styles of six different artists.

You will study six paintings by each of the six artists. On the final test, you will be shown new paintings you have never seen before by the same artists. Your job on this test will be to look at the painting and be able to say which of the six artists painted it.

The six artists are:
1. Hawkins
2. Lewis
3. Foster
4. Cross
5. Oliver
6. Juras

Each of the index cards in front of you represents one painting by one of the six artists. You have 36 (6 paintings x 6 artists) index cards in front of you.

The cards are sorted into 6 piles. Each pile (of 6 cards) contains one painting by each artist/the six paintings by one artist.

Create one single pile of these cards to represent the order in which you would study them, with your first pick on the top and your last pick on the bottom.

(One way to do this is to pick your cards one at a time, and place them face down in front of you as a pile -- that way, when you are done, you can just pick up the entire pile in front of you and flip it over).
Appendix D

Instructions and Experiment Packet for Chapter 4, Experiment 2

Pretend that you are an art student, and that for your first art exam, you have to learn the painting styles of five different artists. On this art exam, you will be shown new paintings you have never seen before by the same artists. Your job on this test will be to look at the painting and be able to say which of the five artists painted it.

You have the opportunity to study 20 paintings (that will not be on the exam) by each artist to learn their specific style. These paintings can be studied in any order you’d like, but they have to be studied one at a time, and once you have studied one, you cannot go back.

In the boxes below, please write out how you would schedule the study of these paintings. Please use letters A, B, C, D and E to refer to the five different artists, and numbers 1-20 to refer to each of the 20 paintings by each artist. For example, A1 refers to the first painting by artist “A”, and E20 refers to the 20th painting by artist “E”. Work from left to right, top to bottom.

[x 10 rows]
Scheduling study of artists — Post-test questionnaire

Please describe any rationale or strategy behind the way you scheduled your study of the artists:

Which of the following orders does your schedule most resemble (select one):

A. Study all paintings by one artist consecutively (e.g. AAA…BBB…CCCC…)
B. Study all the artists mixed up together (e.g., ABCDE, BCDEA, CAEDB…)
C. Familiarize self with one painting by each of the 5 artists first, then study one artist at a time (e.g. ABCDE, Ax19, Bx19…)
D. Study two or three from each artist consecutively (e.g., AABBCC… or, AAABBB…)
E. Study five from each artist consecutively
F. Study 10 from each artist consecutively
G. Start by studying one at time and then mix them up (e.g. starting with blocks of AAAAA, BBBBB, etc, and then end with ABCDE, BCDEA, CAEDB…)
H. Start by mixing them up and then one at a time (e.g. starting with ABCDE, BCDEA, CAEDB, etc. and then ending with blocks of AAAAA, BBBBB, etc)

Which of the above listed orders (A through H) do you think would lead to the best learning of artists’ styles?

____________________

[please turn page over to continue]
Given the choice between options A (studying one artist at a time) and B (mixing the artists up), which would you prefer?

If/When you put paintings by the same artist next to each other, why did you do so?

If/When you mixed up the paintings by different artists, why did you do so?

What year in college are you? Freshman Sophomore Junior Senior+

What is your GPA? ___________________ What is your age? ______________

What is your gender? M / F

Please indicate how much you agree with the following statements on a scale of 1-6, with 1 being strongly agree and 6 being strongly disagree:

**You have a certain amount of intelligence and you really cannot do much to change it.**

1 2 3 4 5 6

Strongly agree Strongly disagree

**Your intelligence is something about you that you cannot change very much.**

1 2 3 4 5 6

Strongly agree Strongly disagree

**You can learn new things, but you cannot really change your basic intelligence.**

1 2 3 4 5 6

Strongly agree Strongly disagree
Appendix E

Examples of Study Rationales in Chapter 4, Experiment 2

Below is a sample of participants’ responses to “Describe any rationale or strategy behind the way you scheduled your study of the artists”, sorted by how they self-scheduled.

**Pure blocking:**

“I’d study each 20 paintings by one artist at a time. Mixing them, I think would confuse me.”

“It is better to master an artist’s method of painting by looking at different art works. Then, study another artist’s paintings and analyze what his or her style is.”

“You can’t really determine the style of a painter or artist until you have looked at several pieces of their work. And so, by looking at all of the work of one artist at a time, you’re better able to pick up on the pattern.”

**Blocks of ten:**

“I like organizing material into big groups. I normally will study one group of related materials at a time and then move onto the next group.”

“I would try to recognize 10 different paintings by for each artist, and by then I’d hope to get a little idea of their style, so when I do the next 10, it would further reinforce the little ideas I already had.”

**Interleave-Block (x2):** “I think first, I have to know the difference between 5 artists, so a pic by each artist will be shown first. Then to know more about each artist, I’ll be shown 9 of their
paintings to know more about their style. To make sure I got the differences between their paintings, one painting from each artist is shown again.

**Block of 15, then five:** “I’ll examine each artist individually first by looking at 15 paintings done by each artist. I think that will definitely help me grasp their style and technique. Then I will cycle back through each artist’s paintings, examining the last five to test my knowledge and make comparisons between different artistic styles.”

**Blocks of five:**

“I did it so you get exposure to each artist, can then test yourself a couple of times before finally reviewing. Also, you can compare them to other styles.”

“Every five I chose to change an artist since I think it might be better to view the same artist’s paintings again after a little break to test if I can recognize them immediately and refresh my memory.”

“Studying the artist’s paintings in a group to get an idea of the style; Mixing it up with the other artists to have a chance for review later on.”

**Blocked-to-interleaved:**

“Well, I would spend some time on each artist to figure out their style and then I would mix it up to test if I could see patterns to prepare.”

“I wanted to mix the artists up as during the test, it is unlikely they would be arranged side to side. I also wanted to leave one artist at the end to heavily study so I could distinguish that
specific artist from the rest, allowing me to note differences between Artist E and the rest, as well as differences between Artists A-D.”

“I first thought that going through 10 paintings of each artist would help get a feel for each one’s style. Then I did consecutive pairs, thinking it would highlight differences between artists. I did it backwards to try and mix it up a little, but not too varied at first. Then I mixed it up, first in order (ABCDE), then at the end, random.”

**Pairs or triples:**

“I feel like if you familiarize yourself with several paintings of one artist, then several of another, you’re more likely to recognize distinct patterns in one artist in comparison to the other artists.”

“First I wanted to get a sense of the artist’s style by just studying “bigger” groups of one artists. Then I would start comparing them to each other, by alternating the artist’s paintings.”

**Interleaved:**

“I believe it’s more effective to study that way because I will be able to distinguish between the paintings better instead of learning them in order. The teacher will most likely mix up the paintings on the test, so why not study that way?”
Appendix F

Instructions and Experiment Packet for Chapter 4, Experiment 3

**RELATED MODULES CONDITION PROMPT:**
You are a student in a class with a final exam, which will take place in five days time. This class has four modules, A, B, C, and D (for example, if this were a mathematics course, the four modules might be integration, volume, geometry and factorial equations.) Assume these modules are equally difficult for you, and equally important for the course.

In these next five days, you will have 16 hours to study for the exam (i.e. 4 hours per module). Let's say that you will study each of these modules for at least one hour on the final day.

**UNRELATED COURSES CONDITION PROMPT:**
You are a student enrolled four classes – A, B, C, and D – that all have final exams, which will all take place in five days time. Unfortunately, all four exams not only occur on the same day, but back-to-back so you will not have the chance to study in between exams.

These four classes are unrelated (for example, geography, mathematics, psychology and history). Assume these classes are equally difficult for you, and you consider getting top grades in each class equally important.

——

How do you schedule the remaining 12 hours (i.e. **3 hours per module/class**), so as to maximize learning? Please schedule hour by hour. For example, each time you write ‘A’, it represents one hour of study for module/class A.

(You do not need to schedule study on every single day, and you may schedule more study on day 5, if you wish).

<table>
<thead>
<tr>
<th>Day 1</th>
<th>Day 2</th>
<th>Day 3</th>
<th>Day 4</th>
<th>Day 5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>B</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>C</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>D</td>
</tr>
</tbody>
</table>

In the space below, please explain the reasons for scheduling your study the way you did:
Shown below are three sample schedules. Please CIRCLE the one that you believe would lead to the best learning.

**Schedule 1**

<table>
<thead>
<tr>
<th>Day 1</th>
<th>Day 2</th>
<th>Day 3</th>
<th>Day 4</th>
<th>Day 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>B</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>D</td>
</tr>
</tbody>
</table>

**Schedule 2**

<table>
<thead>
<tr>
<th>Day 1</th>
<th>Day 2</th>
<th>Day 3</th>
<th>Day 4</th>
<th>Day 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>B</td>
<td>D</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>C</td>
<td>A</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>D</td>
<td>B</td>
<td>D</td>
<td></td>
</tr>
</tbody>
</table>

**Schedule 3**

<table>
<thead>
<tr>
<th>Day 1</th>
<th>Day 2</th>
<th>Day 3</th>
<th>Day 4</th>
<th>Day 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>AA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BB</td>
<td>BB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CC</td>
<td>CC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DD</td>
<td>DD</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Demographic questionnaire:**

<table>
<thead>
<tr>
<th>Age:</th>
<th>Sex:</th>
<th>Year:</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 2.0</td>
<td>&lt; 2.0</td>
<td>2.0-2.5</td>
</tr>
<tr>
<td>2.5-3.0</td>
<td>3.0-3.5</td>
<td>3.5-4.0</td>
</tr>
</tbody>
</table>

GPA: Not estab.
Please indicate how much you agree with the following statements on a scale of 1-6, with 1 being strongly agree and 6 being strongly disagree.

**You have a certain amount of intelligence and you really cannot do much to change it.**

1 2 3 4 5 6

Strongly agree  Strongly disagree

**Your intelligence is something about you that you cannot change very much.**

1 2 3 4 5 6

Strongly agree  Strongly disagree

**You can learn new things, but you cannot really change your basic intelligence.**

1 2 3 4 5 6

Strongly agree  Strongly disagree
Appendix G
Instructions and Experiment Packet for Chapter 4, Experiment 4

In the next part of our experiment, you again be asked to learn the painting styles of eight artists. They will be the same artists as in the previous task/new artists, who you have not studied previously.

As a reminder, the artists are:

Brown/Grossman
Jacobsen/Hawkins
Margulis/Juras
Mylrea/Lewis
McNamara/Lindentberg
Oliver/McKinley
Petro/Mei
Schlorff/Pessani

This time, however, you will get to choose the order that the paintings appear for the next study phase.

Each of the index cards in front of you represents one painting by one of the eight artists. You will have 48 (6 paintings x 8 artists) cards. The cards are sorted into 8 piles. Each pile (of 6 cards) contains the paintings by one artist.

The top (which you can see) has the artist name. The other side has a painting by the artist (which you will study later).

WITHOUT TURNING THE CARDS OVER, create one single pile of these cards to represent the order in which you would like to study them, on the next part of the task. Please try to choose the order that you think will allow you to best learn the painting styles of these artists.

After you create your order, you will study the paintings in your chosen order. Then, you will be again tested on how well you learned the artists—again, you will see new paintings by these artists and be asked to identify which artist painted each image.

You will score 1pt for every painting that you correctly classify. Your goal is to score as many points as possible in this phase of the study.
Appendix H

Additional Results for Chapter 4, Experiment 4

Table 1

Individual differences between the participants who chose to predominantly block their schedules and those who chose to predominantly interleave their schedules

<table>
<thead>
<tr>
<th>Variable</th>
<th>Predominantly Blockers (SD)</th>
<th>Predominantly Interleavers (SD)</th>
<th>t-statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase 1 test performance</td>
<td>.51 (.20)</td>
<td>.54 (.23)</td>
<td>.67</td>
<td>.51</td>
</tr>
<tr>
<td>Year in college</td>
<td>2.68 (1.15)</td>
<td>2.68 (1.25)</td>
<td>.01</td>
<td>.99</td>
</tr>
<tr>
<td>GPA</td>
<td>2.65 (1.43)</td>
<td>2.60 (1.55)</td>
<td>.17</td>
<td>.87</td>
</tr>
<tr>
<td>Classes skipped per week</td>
<td>.76 (1.06)</td>
<td>.68 (1.05)</td>
<td>.33</td>
<td>.74</td>
</tr>
<tr>
<td>Hours of study daily</td>
<td>2.59 (1.69)</td>
<td>3.37 (2.18)</td>
<td>1.86</td>
<td>.07</td>
</tr>
<tr>
<td>Average theory of intelligence score</td>
<td>4.04 (1.25)</td>
<td>4.02 (1.20)</td>
<td>.05</td>
<td>.96</td>
</tr>
</tbody>
</table>

Table 2

Proportion of participants from each experimental condition (initial schedule x phase 2 study set) for each specific strategy classification

<table>
<thead>
<tr>
<th>Specific (Self-)Schedule</th>
<th>Blocked</th>
<th></th>
<th>Interleaved</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>New</td>
<td>Repeated</td>
<td>New</td>
<td>Repeated</td>
</tr>
<tr>
<td>Pure blocking</td>
<td>0.61</td>
<td>0.61</td>
<td>0.46</td>
<td>0.75</td>
</tr>
<tr>
<td>Average run 3+</td>
<td>0.13</td>
<td>0.13</td>
<td>0.11</td>
<td>0.20</td>
</tr>
<tr>
<td>Maximum two in a row</td>
<td>0.09</td>
<td>0.04</td>
<td>0.18</td>
<td>0.00</td>
</tr>
<tr>
<td>Combination</td>
<td>0.13</td>
<td>0.00</td>
<td>0.11</td>
<td>0.00</td>
</tr>
<tr>
<td>Pure interleaving</td>
<td>0.04</td>
<td>0.22</td>
<td>0.14</td>
<td>0.05</td>
</tr>
</tbody>
</table>
References


Zechmeister, E. B., & Shaughnessy, J. J. (1980). When you know that you know and when you think that you know but you don't. *Bulletin of the Psychonomic Society, 15*, 41-44.
