

The exemplar interleaving effect in inductive learning: Moderation by the difficulty of category discriminations

Norehan Zulkipli · Jennifer S. Burt

© Psychonomic Society, Inc. 2012

Abstract Recent research demonstrates a spacing effect in inductive learning. Spacing different individual exemplars apart in time, rather than massing them together, aids in the learning of categories. Experiment 1 examined whether it is interleaving or temporal spacing that is critical to the spacing effect in the situation where the memory load is high, and the results favored interleaving. Experiment 2 examined the effect of the difficulty of the category discrimination on presentation style (massed vs. spaced) in inductive learning, and the results demonstrated that spacing (i.e., interleaving of exemplars from different categories) is advantageous for low-discriminability categories, whereas massing is more effective for high-discriminability categories. In contrast to these performance measures, massing was judged by participants to be more effective than spacing in both discriminability conditions, even when performance for low-discriminability categories showed the opposite.

Keywords Spacing effect · Inductive learning · Category learning · Category induction · Category discrimination

It has long been known that repetitions of items further apart in time produce better memory than do repetitions close together in time (e.g., Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006;

Childers & Tomasello, 2002; Donovan & Radosevich, 1999; Ebbinghaus, 1964; Melton, 1970; Rea & Modigliani, 1987; Toppino, 1993). This finding, known as the *spacing effect*, is well documented in the memory literature (e.g., Cepeda et al., 2006; Dempster, 1996; Donovan & Radosevich, 1999). The effect is robust and has been demonstrated using a wide range of materials, including nonsense syllables (e.g., Ebbinghaus, 1985/1913), words (e.g., Glenberg & Lehmann, 1980), sentences (e.g., Rothkopf & Coke, 1966), pictures (e.g., Hintzman & Rogers, 1973), and faces (e.g., Cornell, 1980). The effect also applies to various contexts, from learning simple lists (e.g., Glenberg, 1979) to learning complex judgment tasks (e.g., Helsdingen, Van Gog, & Van Merriënboer, 2011).

Recent research also demonstrates a spacing effect in inductive learning—in particular, category learning. A typical study examining a spacing effect in inductive learning compares a massed condition in which exemplars from each category are presented contiguously with a spaced condition in which exemplars from each category are presented apart in time (e.g., Kornell & Bjork, 2008; Kornell, Castel, Eich, & Bjork, 2010). Several earlier studies showed that massing facilitates induction (e.g., Gagne, 1950; Kurtz & Hovland, 1956), with other studies providing less direct evidence that massing facilitates induction (e.g., Appleton-Knapp, Bjork, & Wickens, 2005; Dellarosa & Bourne, 1985; Glover & Corkill, 1987; Melton, 1970; Wulf & Shea, 2002). Nonetheless, as was noted, there is growing evidence from recent research suggesting that spacing results in better learning of categories and concepts (i.e., Kang & Pashler, 2012; Kornell & Bjork, 2008; Kornell et al., 2010; Vlach, Sandhofer, & Kornell, 2008; Wahlheim, Dunlosky, & Jacoby, 2011). Various types of learning materials have been used in recent studies, such as paintings from several artists (e.g., Kang & Pashler, 2012; Kornell & Bjork, 2008; Kornell et al., 2010), different categories of novel objects that were constructed from arts and craft supplies and objects from hardware stores (e.g., Vlach et al., 2008), and different categories of bird families (e.g.,

N. Zulkipli (✉) · J. S. Burt
School of Psychology, University of Queensland,
Brisbane, QLD 4072, Australia
e-mail: norehan.zulkipli@uqconnect.edu.au

N. Zulkipli
e-mail: znorehan@fcs.unimas.my

N. Zulkipli
Faculty of Cognitive Sciences and Human Development,
Universiti Malaysia Sarawak,
94300 Kota Samarahan,
Sarawak, Malaysia

Wahlheim et al., 2011). In our recent work (Zulkipli, McLean, Burt, & Bath, 2012), we also found a result similar to that obtained by previous studies—that is, spaced presentation facilitated the learning of categories in inductive learning—and the finding was extended to textual material. Thus, recent studies seem to demonstrate generalization of the spacing effect in the context of inductive learning.

Despite the fact that there is growing evidence that induction profits from spacing in category learning, there are still some questions about the mechanisms of the effect as it applies to induction. The first question concerns the roles of interleaving versus spacing: Which one is more critical to the effect? In Kornell and Bjork's (2008) experiment, participants viewed examples of paintings by 12 artists, with the artists' names displayed underneath the paintings, and they were asked to decide which of the studied artists were responsible for new paintings that were presented at test. In the massed condition, paintings were blocked by artists, and in the spaced condition, exemplars of different artists were interleaved. It is argued that it could be the interleaving of the artists (and not the spacing itself) that might have enhanced discrimination learning, allowing participants to differentiate the styles of each artist from their paintings, thus giving an advantage to spacing (Kornell & Bjork, 2008).

A recent study by Kang and Pashler (2012) suggested the importance of discriminative contrast in the learning of painting styles. In their first experiment, they presented 72 paintings of three artists, together with the artists' names, in four study conditions (massed, interleaved, temporal spaced, and simultaneous massed). In the massed and interleaved conditions, paintings were presented one at a time, separated by a 0.5-s blank screen after each painting. In the massed condition, the paintings were blocked by artist, whereas in the interleaved condition, the paintings by an artist were interleaved with paintings by other artists. In the temporal spaced condition, the order of the paintings was similar to that in the massed condition, but 11.5 s of unrelated filler materials was inserted between presentations of paintings. In the simultaneous massed condition, the paintings were presented four at a time (instead of singly), with a 2-s blank screen after each set of four. The paintings were blocked by artist, in terms of both each set of four and the sequencing of the sets. The results of the first experiment demonstrated that the critical factor that enhanced learning of the artists' styles was not the temporal spacing of paintings during study but, rather, the interleaving of different artists. In their second experiment, Kang and Pashler included another study condition—that is, simultaneous different—which allowed for the simultaneous viewing of paintings painted by different artists, and they found that this condition yielded the best performance but was not significantly different from the interleaved condition.

The aim of the present study was to contribute to the spacing versus interleaving question. In Experiment 1, 12, rather than 3,

artist categories (as in Kang & Pashler, 2012) were used. The increase in the number of categories was likely to increase memory demands and perhaps also discriminability difficulty; thus, it would be interesting to investigate whether the interleaving effect could still be observed in such a situation.

The second question motivating the present experiments was whether the effect of presentation style (massed vs. spaced) in inductive learning varies over different levels of discriminability of the learning materials. In an earlier study of induction in which massed presentation was superior to spaced presentation, Kurtz and Hovland (1956) argued that “when the degree of discriminability is low it might be expected that placing of exemplars from different concepts in juxtaposition would facilitate discrimination learning, whereas with greater discriminability, like that obtaining in the present study, the reverse might obtain” (p. 242). Massing allows one to notice the similarities between successive exemplars within a category, whereas spacing makes doing so more difficult (Rothkopf, as cited in Kornell & Bjork, 2008). Conversely, spacing (interleaving) facilitates comparison and contrast of exemplars of different categories, which may be useful when the features that differentiate categories are difficult to detect. Kornell and Bjork (2008) endorsed the argument of Kurtz and Hovland and seemed to agree that the advantage for spacing may have occurred because the discrimination among categories in their experiment was difficult. Nevertheless, to date, there is no direct experimental evidence that clarifies the issue of different levels of discriminability of the learning materials and how this affects massed and spaced presentation in inductive learning.

Additionally, Kornell and colleagues (i.e., Kornell & Bjork, 2008; Kornell et al., 2010) examined participants' judgments about which presentation style they thought helped them learn better (i.e., massed or spaced). Interestingly, they found that the majority of their participants reported massing to be more effective, even though their actual performance showed the opposite. The first aim of the present study was to investigate whether it is interleaving or temporal spacing that is critical to the spacing effect in induction when the memory load is high (Experiment 1). A second aim was to investigate the effect of different degrees of learning material discriminability on the presentation style (massed vs. spaced) in inductive learning (Experiment 2).

Experiment 1

There were two specific research questions for Experiment 1: First, which is more critical to the spacing effect, interleaving or temporal spacing, and second, does increasing temporal spacing between exemplars arranged in a spaced (interleaved) presentation sequence enhance or reduce the learning of the categories, as compared with the typical interleaved condition,

which has no delay between exemplars? Experiment 1 examined four presentation (study) conditions. The first two conditions—*massed immediate* and *interleaved immediate*—were basically similar to the ones used by Kornell and Bjork (2008) and Kang and Pashler (2012). The third study condition—*massed temporally spaced*—which used the same sequence of paintings as the massed immediate condition, was similar to another condition labeled as temporal spaced in Kang and Pashler (2012), except that in the present study, 30 s of unrelated filler material was inserted between exemplars, whereas Kang and Pashler inserted 11.5 s of unrelated filler material. Given that 30 s of filled distraction should eliminate any residual effects of short-term memory, one might anticipate that the interleaving effect (or the compare and contrast effect) would be reduced. Additionally, one new study condition was included in the present study—that is, *interleaved temporally spaced*—which had the same sequence of paintings as the interleaved immediate condition but featured the insertion of 30 s of unrelated filler material between exemplars. If memory is important in the induction task, temporal spacing may further enhance category learning in this condition. In contrast, if temporal spacing impairs comparison of exemplars, the spacing effect may be eliminated in the interleaved temporally spaced condition.

Method

Participants and design

Eighty students (55 of them female, 25 male) from an introductory psychology class earned course credit in exchange for their participation. The design of the experiment was a 2 (presentation style: massed vs. interleaved) \times 2 (temporal spacing: immediate vs. temporally spaced) \times 4 (test block: blocks 1–4) mixed factorial design. Presentation style and temporal spacing were varied between participants, while test block was varied within participants.

Materials

The materials were 120 paintings (landscapes or skyscapes) by 12 different artists and were taken from Kornell and Bjork's (2008) study. The artists were Yie Mei, Ciprian Stratulat, Bruno Pessani, Georges Braque, Judy Hawkins, George Wexler, Georges Seurat, Marilyn Mylrea, Ron Schlorff, Ryan Lewis, Philip Juras, and Henri-Edmond Cross. All of the painting files were in the format of JPEG files and were resized to fit into a 19 \times 29 cm rectangle on the computer screen.

Procedure

Participants were tested individually at computers within a multistation lab, with a maximum of 4 persons per session.

Participants first were instructed about the nature of the experiment; in particular, they were told that they had to study 72 paintings from 12 artists that would be presented to them for 3 s each and that later they would be shown 48 new paintings from the 12 learned artists and that they had to identify who painted each one. Participants were randomly assigned to one of these four experimental conditions: massed immediate, interleaved immediate, massed temporally spaced, and interleaved temporally spaced. There were three steps involved in the experimental manipulation: presentation (study) phase, distractor task, and test phase. During the presentation phase, in the massed immediate and interleaved immediate conditions, 72 paintings (6 paintings per artist), were presented 1 at a time on a computer screen for 3 s each, with the last name of the artist displayed underneath the painting. Each painting was followed by a blank screen with a duration less than 100 ms. The difference was that in the massed immediate condition, the paintings were blocked by artists, whereas in the interleaved immediate condition, the paintings of each of the 12 artists were intermingled with the paintings by other artists. The sequence of the 72 paintings in the massed temporally spaced and the interleaved temporally spaced conditions was identical to that in the massed immediate and the interleaved immediate conditions, respectively, except that the presentation of each painting was followed by a 30-s unrelated filler task, during which participants were asked to answer as many trivia questions as possible, which were automatically generated on the computer screen. This task was meant to prevent deeper processing of the paintings during the interval and to allow time for information to be lost from short-term memory.

In all four presentation conditions, the paintings of each of the 12 artists were arranged in 12 learning blocks. Next, participants were asked to complete a distractor task, during which they were asked to count backward by three starting from 547 for 15 s and to type the numbers in a given box on the computer. In the subsequent test phase, participants were shown 48 new paintings (from the 12 learned artists, 4 paintings per artist), and they then had to identify the artist who created the painting. Participants saw 1 painting at a time on the computer screen, with 13 buttons displayed underneath the painting. Twelve of the buttons were labeled with the artists' names, and 1 button was labeled "I don't know." Participants responded according to who they thought had created each painting by clicking the computer mouse on the corresponding button. Feedback was given after each response. If participants responded correctly, the word "correct" appeared on the computer screen. If they responded incorrectly, the correct artist's name was presented on the computer screen. Participants completed the test phase at their own pace. Participation in the massed immediate and interleaved immediate conditions took approximately 30 min, whereas

participation in the interleaved temporally spaced and massed temporally spaced conditions took approximately 60 min. Participants were debriefed about the experiment before they left the experimental room.

Results

Figure 1 shows the proportions of pictures selected correctly on the test in Experiment 1. As can be seen from the figure, the interleaved immediate condition produced the best performance, followed by the interleaved temporally spaced condition, while the massed immediate and massed temporally spaced conditions produced poorer performance. A three-way mixed factorial ANOVA was conducted on the data for Experiment 1, with between-participants factors of presentation style and temporal spacing. The main effect of presentation style was significant, reflecting higher accuracy in the interleaved conditions over the massed conditions, $F(1, 76) = 15.98$, $MSE = 3.33$, $p < .001$, $\eta_p^2 = .174$. Interleaving seems to be critical for the benefit of spacing observed in induction under the present high memory load. There was also a significant effect of test block, $F(3, 228) = 22.47$, $MSE = 1.80$, $p < .001$, $\eta_p^2 = .228$. The main effect of temporal spacing was not significant, $F(1, 76) = 1.32$, $MSE = 3.33$, $p = .25$, $\eta_p^2 = .017$. The two-way interactions and the three-way interaction were not significant [presentation style \times temporal spacing, $F(1, 76) = 0.56$, $MSE = 3.33$, $p = .455$, $\eta_p^2 = .007$; presentation style \times test block, $F(3, 228) = 1.49$, $MSE = 1.80$, $p = .217$, $\eta_p^2 = .019$; temporal spacing \times test block, $F(3, 228) = 1.64$, $MSE = 1.80$, $p = .181$, $\eta_p^2 = .021$; presentation style \times test block \times temporal spacing, $F(3, 228) = 1.56$, $MSE = 1.80$, $p = .200$, $\eta_p^2 = .020$]. In particular, the nonsignificant interaction between the effect of presentation style and the effect of temporal spacing suggests that there were no differences in the interleaving effect over immediate and temporally spaced conditions. From these results, it can be concluded that the spacing effect obtained in this inductive learning experiment is due to interleaving, and not temporal spacing.

Even though the interactions were not significant, there was some indication in the data that temporal spacing might have had a selective effect in the interleaving conditions. There was a trend toward lower accuracy (by 7 percentage points) in the interleaved temporally spaced condition (see Fig. 1). Consequently, a separate two-way mixed factorial ANOVA was conducted for the two conditions with interleaving—interleaved immediate and interleaved temporally spaced—to determine whether increasing temporal spacing between interleaved exemplars enhanced or reduced the learning of the artists' styles, as compared with the interleaved immediate condition. The main effect of temporal spacing was not significant, indicating that there was no difference in performance between the interleaved

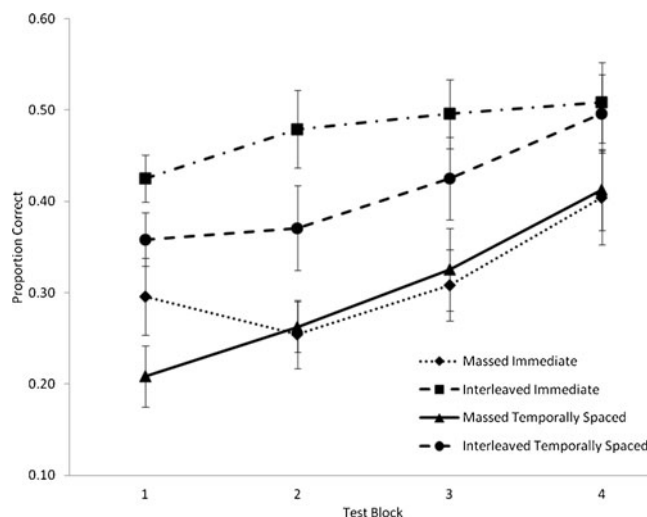


Fig. 1 Proportion of artists selected correctly on the test in Experiment 1, as a function of presentation condition and test block. Error bars represent standard errors

temporally spaced ($M = 0.41$) and interleaved immediate ($M = 0.48$) conditions and, thus, that spacing did not have any effect on the benefit of interleaving, $F(1, 38) = 1.78$, $MSE = 3.37$, $p = .190$, $\eta_p^2 = .045$. The effect of test block was significant, $F(3, 114) = 8.04$, $MSE = 1.61$, $p < .001$, $\eta_p^2 = .175$; however, the interaction between temporal spacing and test block was not significant, $F(3, 114) = 1.39$, $MSE = 1.61$, $p = .249$, $\eta_p^2 = .035$.

Discussion

Experiment 1 showed that learning of the artists' styles was enhanced when the paintings by the 12 artists were interleaved during the presentation (study) phase. This result is consistent with Kang and Pashler (2012), and the present study extended this finding to a situation where memory load was high and the number of category discriminations to be learned was increased from 3 to 12. The mean test accuracy for the three conditions in Experiment 1 (i.e., massed immediate [$M = 0.32$], interleaved immediate [$M = 0.48$], and interleaved temporally spaced [$M = 0.41$]) are slightly lower than in the similar three conditions used by Kang and Pashler (i.e., massed [$M = 0.60$], interleaved [$M = 0.68$], and temporal spaced [$M = 0.61$], respectively), but that could possibly be due to the fact that memory load is higher in the present study than in Kang and Pashler's study, with 12 painters in the present study and 3 in that of Kang and Pashler. Also, associated with the larger number of painters used in the present study, the number of response alternatives was higher; thus, chance performance was obviously lower in the present experiment.

Additionally, there was no concrete indication that increasing temporal spacing between painting exemplars

arranged in spaced presentation sequence (as in the interleaved temporally spaced condition) enhanced or reduced the learning of the artists' styles, as compared with the interleaved immediate condition. Temporal spacing can be expected to improve memory for the exemplars. As was noted by Vlach et al. (2008), category learning must rely on participants' ability to remember previous category exemplars. However the absence of an effect of temporal spacing suggests that, at least with the paintings used here, the memory of exemplars was not a limiting factor in category learning. Alternatively, it is possible that the spacing manipulation did not appreciably affect memory for the paintings. The small decrement for the interleaved temporally spaced condition, relative to the interleaved immediate condition, was not significant, but it is suggestive evidence for the role of category comparisons in the benefit of interleaving, in that comparisons would be more difficult in the interleaved temporally spaced condition. In summary, the results of Experiment 1 strongly suggest that temporal spacing between exemplars does not seem to have any effect on the benefit of interleaving. It is the interleaving of the category members that is critical in enhancing inductive learning.

Experiment 2

Experiment 1 revealed that the main factor that is critical to the spacing effect in inductive learning is interleaving, and not temporal spacing. The term *spacing* that Kornell and Bjork (2008) used to describe their manipulation will hereafter be replaced by the term *interleaving*. The results of Experiment 1 suggest that participants may have taken advantage of interleaving to compare and contrast exemplars of different categories. As was noted previously, such a strategy may be more important when the category discrimination is difficult (Kurtz & Hovland, 1956). This issue was investigated in Experiment 2 with the interleaved immediate versus massed immediate conditions, in line with Kornell and Bjork. The stimuli were squares containing familiar and nonsense shapes in varying colors. Two research questions were addressed. First, what effect does interleaving have on inductive learning as a function of the discriminability of materials used? Specifically, we were interested in finding out whether the interleaving benefit is observed only when category discrimination is difficult. Second, which one is judged to be more helpful in the learning of categories in inductive learning—massing or interleaving? On the basis of previous findings when presentation style was varied within participants (e.g., Kornell & Bjork, 2008; Kornell et al., 2010), it was hypothesized that the majority of participants in Experiment 2 would report massing to be more helpful

than interleaving in learning both high- and low-discriminable categories.

Method

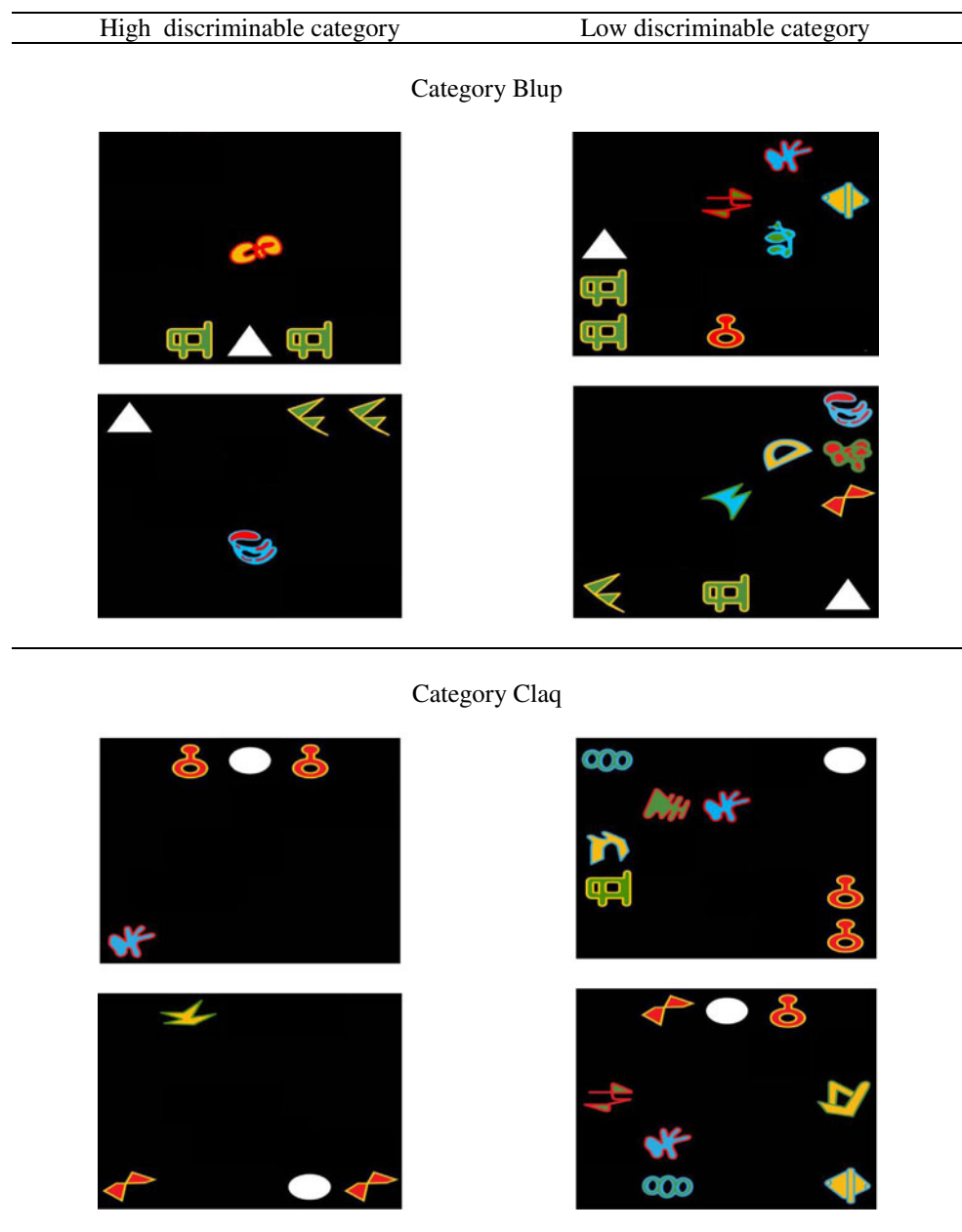
Participants and design

Forty students (28 of them female, 12 male) from an introductory psychology class participated in the experiment for course credit. The design of the experiment was a 2 (presentation style: interleaved vs. massed) \times 2 (degree of discriminability: high vs. low) \times 4 (test block: blocks 1–4) mixed factorial design. Category discriminability was varied between participants, while presentation style and test block were varied within participants. There were four steps involved in the experimental manipulation: presentation (study) phase, distractor task, test phase, and question phase. The four steps introduced in this experiment followed the approach in Kornell and Bjork (2008). In the presentation phase, the picture exemplars from the 12 categories were arranged in 12 learning blocks. The order of the blocks was MIIMMIIMMIIM (M for massed; I for interleaved). In particular, in the interleaved learning blocks, the pictures from a category were interleaved with pictures from other categories. The assignment of picture categories to condition (massed vs. interleaved) was counterbalanced over two lists. Two versions of each list were produced in which there was a different assignment of categories to blocks. Thus, there were four lists in total. In the test phase, 48 new picture exemplars from the 12 categories learned in the presentation phase were arranged in four test blocks. Each block consisted of one new picture from each category presented in a fixed order across participants.

Materials

The materials were 240 pictures drawn and prepared using Adobe Illustrator and Adobe Photoshop software (see Fig. 2). The pictures were divided into two sets, each having 120 pictures from 12 categories. Specifically, the first set consisted of pictures that were designated to have a high degree of discriminability, while the second set consisted of pictures that were designated to have a low degree of discriminability. Furthermore, in each set, 72 pictures were used in the presentation/study phase (6 pictures per category), and 48 pictures were used in the test phase (4 pictures per category). Each category was assigned a distinctive nonsense-syllable name. Additionally, each picture consisted of a black square frame containing three elements that defined the category and one or more distractors. Two elements were nonsense shapes (termed *category icons*) that shared a color composition, and the third element was a shape (circle or triangle; see Table 1). The features that were

Fig. 2 Examples of exemplars from the high-discriminable category and the low-discriminable category used in Experiment 2



diagnostic of the category were (1) color composition of the category icons and (2) the shape that accompanied the category icons (circle or triangle). The nondiagnostic features (those that did not distinguish categories) were (1) the shape of the category icons, (2) the position of the category icons and the accompanying shape, and (3) the number of distractors in the square frame. The high-discriminable categories and the low-discriminable categories differed in the three nondiagnostic features: (1) For high-discriminable categories, the category icons had the same shape in each picture, and for low-discriminable categories, the category icons were of either the same or a different shape in each picture; (2) for high-discriminable categories, the category icons and the shape were all positioned at either the top or

the bottom, and for low-discriminable categories, the category icons and the shape were all positioned top, bottom, right, or left; and (3) for high-discriminable categories, there was only one distractor, and for low-discriminable categories, there were five distractors, one distractor from each category.

Procedure

The presentation phase (72 pictures and their names), distractor task, and test phase with 48 new pictures were conducted as in Experiment 1. The only notable difference was that the pictures and their category names were displayed for 5 s during the presentation phase. Additionally, following

Table 1 Categories of the pictures used in both conditions (high-discriminable and low-discriminable categories) in Experiment 2 and the elements that define each category

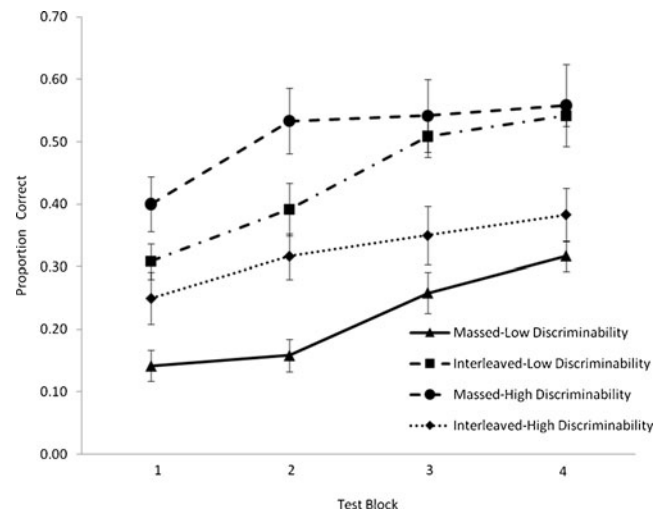
Categories of pictures	Elements that define each category
Zas	2 blue–green category icons with a circle
Jed	2 blue–red category icons with a circle
Vix	2 blue–yellow category icons with a circle
Foy	2 green–yellow category icons with a circle
Guj	2 red–green category icons with a circle
Clap	2 red–yellow category icons with a circle
Drem	2 blue–green category icons with a triangle.
Flis	2 blue–red category icons with a triangle
Kros	2 blue–yellow category icons with a triangle
Blup	2 green–yellow category icons with a triangle
Yack	2 red–green category icons with a triangle
Wex	2 red–yellow category icons with a triangle

Kornell and Bjork's (2008) approach, after the test phase, participants read a description about the meanings of the terms “massed” and “spaced” (interleaved) on the computer screen. They were asked, “Which option do you think helped you learn more?” and were provided with three possible answers—“massed,” “about the same,” or “spaced”—which ended the experimental manipulation. Participation in the experiment took approximately 40 min, and participants were debriefed about the experiment before they left the experimental room.

Results

The results of main interest are shown in Fig. 3. As can be seen from the figure, when the degree of discriminability was high, performance was higher in the massed condition ($M = 0.51$) than in the interleaved condition ($M = 0.33$), whereas when the degree of discriminability was low, performance was higher in the interleaved condition ($M = 0.44$) than in the massed condition ($M = 0.22$).

A three-way mixed ANOVA was conducted on the data for Experiment 2. The main effect of degree of discriminability was significant, $F(1, 38) = 6.09$, $MSE = 3.71$, $p = .018$, $\eta_p^2 = .138$, which indicated that, on average, participants' accuracy was significantly higher in the high-discriminability condition than in the low-discriminability condition. Participants' performance also significantly increased across test block, $F(3, 114) = 25.87$, $MSE = .66$, $p < .001$, $\eta_p^2 = .405$ (as illustrated in Fig. 3). However, the effect of presentation style was not significant, $F(1, 38) = 0.48$, $MSE = 1.89$, $p = .494$, $\eta_p^2 = .012$. Critically, the interaction between the two main effects—degree of discriminability and presentation style—was significant, $F(1, 38) = 61.59$, $MSE = 1.89$, $p < .001$, $\eta_p^2 = .618$, which indicates that the effect of presentation style

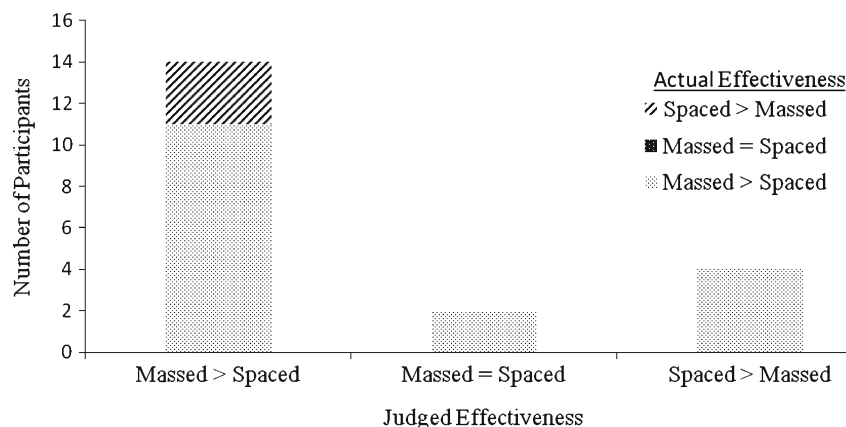
**Fig. 3** Proportion of pictures selected correctly on the test in Experiment 2, as a function of study condition and test block. Error bars represent standard errors

differed at each level of discriminability (see Fig. 3). Follow-up t -tests revealed significant differences between interleaved and massed conditions in both the high-discriminability condition, $t(19) = 4.14$, $p = .001$, and the low-discriminability condition, $t(19) = -8.48$, $p < .001$. The remaining two-way interactions and the three-way interaction were not significant [test block \times degree of discriminability, $F(3, 114) = 2.47$, $MSE = .66$, $p = .066$, $\eta_p^2 = .061$; presentation style \times test block, $F(3, 114) = 0.14$, $MSE = .61$, $p = .934$, $\eta_p^2 = .004$; presentation style \times test block \times degree of discriminability, $F(3, 114) = 1.11$, $MSE = .061$, $p = .350$, $\eta_p^2 = .028$].

With regard to participants' judgments of which particular study presentation had helped them most, a similar pattern of preference for massed presentation was observed in both discriminability conditions. In the high-discriminability condition, of a total of 20 participants, a majority of 14 (70 %) claimed massed, 4 (20 %) claimed spaced, and another 2 (10 %) judged massed and spaced to have contributed about the same (as illustrated in Fig. 4). A one-way chi-square analysis compared the proportion of participants who judged massed to be most useful with the proportion not preferring massed presentation and confirmed that the majority of participants reported massing to be more helpful than spacing in learning the high-discriminable categories, $\chi^2(2, N = 20) = 12.40$, $p = .002$. In terms of categorization performance, 17 (85 %) of the participants performed better in the massed condition, and 3 (15 %) performed better in the spaced (interleaved) condition.

Similarly, in the low-discriminability condition, of a total of 20 participants, a majority of 12 (60 %) participants claimed that massed was most effective, 4 (20 %) claimed spaced, and another 4 (20 %) judged the two conditions equally effective, regardless of their performance in the two

Fig. 4 Number of participants (out of 20) who judged massing as more effective than, equally effective as, or less effective than spacing in Experiment 2 (for high degree of discriminability category). For each judgment, the number of participants is divided according to their actual performance in the spaced condition, relative to the massed condition

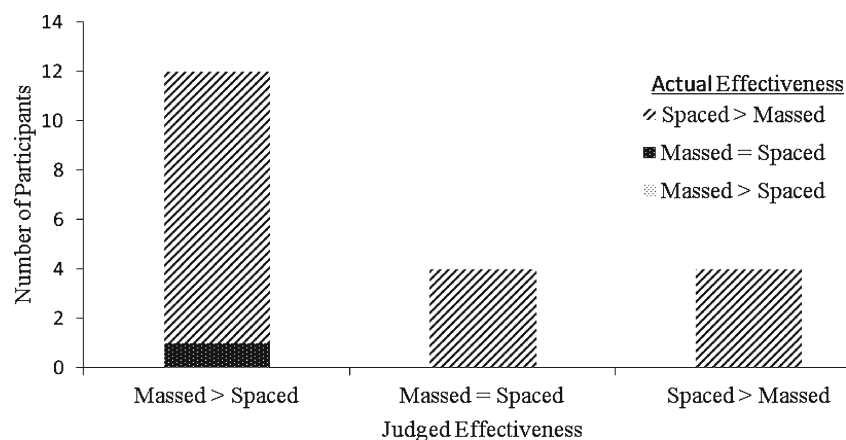


conditions (massed and spaced; as illustrated in Fig. 5). The result of a one-way chi-square analysis conducted on the judgment data revealed that most participants judged massing to be most helpful in learning the low-discriminable categories, $\chi^2(2, N = 20) = 6.40, p = .041$. In terms of categorization performance, 19 (95 %) of the participants performed better in the spaced (interleaved) condition, and 1 (5 %) performed equally in the two conditions.

Discussion

Experiment 2 provided experimental evidence in support of Kurtz and Hovland's (1956) argument that, in inductive or category learning, massing is more effective in learning the high-discriminable categories, whereas in learning the low-discriminable categories, interleaving the exemplars from different categories is critical in enhancing discrimination learning. These results suggest that the interleaving effect observed with paintings in Experiment 1 and elsewhere (e.g., Kornell & Bjork, 2008) reflects the high difficulty of discriminating the painter categories employed. It must be noted that there were some differences in the materials between the two experiments. The implications of these differences will be considered in the General Discussion section.

Fig. 5 Number of participants (out of 20) who judged massing as more effective than, equally effective as, or less effective than spacing in Experiment 2 (for low degree of discriminability category). For each judgment, the number of participants is divided according to their actual performance in the spaced condition, relative to the massed condition



On the postexperimental questionnaire, the majority of participants in both discriminability conditions (i.e., high and low) appeared to believe that massed presentation made it easier to recognize the nature of each category of picture exemplars during the study phase. In line with Kornell and colleagues' observation (Kornell & Bjork, 2008; Kornell et al., 2010) and our previous study (Zulkiply et al., 2012), it is interesting to discover that in the low-discriminability condition, the majority of the participants rated massing as more effective than interleaving in response to the administered questionnaire even though their own test performance had demonstrated the opposite. This finding is consistent with the view that people's access to their complex mental processes is not very effective (Kahneman, 2011; Nisbett & Wilson, 1977). Nisbett and Wilson argued that we typically are conscious of the products of our thinking but are only vaguely conscious of the process of our thinking. The impressions, intuitions, and feelings that guide us are not always justified, and we are often confident even when we are wrong (Kahneman, 2011). It is likely that massed exemplars of a category that were presented to the participants consecutively caused them to develop a sense of familiarity toward each category regardless of whether the categories were difficult or easy, which then influenced their judgment toward massing. The sense of familiarity is one of the causes of cognitive ease,

and we are likely to trust our intuitions when we are in a state of cognitive ease (Kahneman, 2011). In contrast, interleaving the exemplars of a category is unlikely to have created an equivalent sense of familiarity with the categories in participants, since the exemplars of each category were spaced apart throughout the presentation phase.

General discussion

The present experiments further examined the spacing effect in inductive learning in terms of the role of interleaving versus spacing and as a function of the degree of discriminability of the learning materials.

Consistent with Kang and Pashler (2012), the present results revealed that the spacing advantage observed in inductive learning was due to the interleaving of artists during the study phase and not to the temporal spacing between painting exemplars, and crucially, this finding was extended to a condition in which memory load was higher, in that more artists' styles had to be learned (Experiment 1). Interleaved presentation appeared to have highlighted the differences among artists' styles and facilitated comparison of painting exemplars, which made it easier to detect the features that differentiated the style of one artist from those of the others, thus enhancing discrimination learning. It is also likely that interleaved presentation in Experiment 1 made it more difficult to recognize the style of each of the artists during the presentation/study phase, thus resulting in a deeper processing of the interleaved exemplars. The extra attention given to the interleaved exemplars might have resulted in better learning of the interleaved exemplars.

In contrast, massed presentation possibly allowed the participants to ignore how a painting's style compared with those of other artists, because all painting exemplars that were presented were blocked by artists and, thus, had the same style (e.g., colors and themes used across painting exemplars). Massing may impair learning by reducing the amount of attention people pay to repeated presentations because the massed items become highly familiar (Hintzman, 1974). It is also likely that attention was less because participants overestimated the degree to which they would remember massed items (Zechmeister & Shaughnessy, 1980), and they may have thought that massed presentation made it easier for them to discern a painter's style (as compared with interleaved presentation). Similarly, previous authors have suggested attenuation of attention (Kornell et al., 2010; Wahlheim et al., 2011), with attention more likely to weaken across exemplars from a category when they are presented massed rather than when they are interleaved. Thus, massed presentation of paintings was detrimental to inductive learning, as also was observed in Kang and Pashler (2012).

Another outcome of Experiment 1 was the finding that for the interleaved conditions, accuracy was higher in the immediate than in the temporally spaced condition, but not significantly so. Kang and Pashler (2012, Experiment 2) found a small (nonsignificant) advantage for simultaneous presentation of exemplars from three categories over the standard interleaved condition with successive presentation of exemplars. These results suggest that minimizing the separation between exemplars may facilitate exemplar comparisons. Nevertheless, it appears that comparing and contrasting in interleaved presentation can be performed reasonably effectively when the exemplars are not presented together and even when there is a substantial interval (30 s) separating them.

Crucially, Experiment 2 revealed that the benefit with interleaved categories was observed only when category discriminations were difficult. In light of the results of Experiment 1 and Kang and Pashler (2012), it can be assumed that this benefit for difficult discriminations was due to interleaving, and not temporal spacing. In contrast, learning of the high-discriminable categories was superior when the categories were massed. The interaction with discrimination difficulty provided experimental evidence in support of Kurtz and Hovland's (1956) argument that placing of exemplars from different concepts in juxtaposition would facilitate discrimination learning only when category discriminability was low. The finding observed for the low-discriminability condition is consistent with Kornell and Bjork (2008) and Kornell et al. (2010); both studies used paintings as stimulus materials, which have low discriminability, as was argued by Kornell and Bjork. On the other hand, the finding observed in the high-discriminability condition in the present study accords with Kurtz and Hovland.

With respect to attention during learning in Experiment 2, the suggestion that participants might have devoted more attention to exemplars in the interleaved condition does not seem applicable to the interleaved exemplars from the high-discriminable categories. The fact that massed exemplars were learned better than interleaved exemplars indicates that when category assignment was easy, the interleaved exemplars did not receive extra attention during the study phase or that extra attention was not required for successful category learning.

In Experiment 2, the picture exemplars from the low-discriminable categories involved a larger number of features and feature variations than did the exemplars from the high-discriminable categories. Interleaving picture exemplars from the different categories might have encouraged participants to compare and contrast different categories to ascertain which items or features were distractors and which were diagnostic of the categories. On the contrary, for the picture exemplars from the high-discriminable categories, there were fewer feature variations to consider (e.g., the category icons were always the same shape, there was only one distractor, and there was less variation in other dimensions). In this situation, massing

may have facilitated the identification of the common features that defined the category (e.g., Kurtz & Hovland, 1956; Underwood, 1952). Additionally, it is likely that opportunities for comparison and contrast among exemplars provided by the interleaved presentation did not produce much benefit for the easy discriminations. For example, participants may have remembered the defining features of other categories during each block.

The above explanation for the findings of Experiment 2 seems to be supported by other research. Goldstone (1996; Goldstone, Steyvers, & Rogosky, 2003) proposed a distinction between interrelated and isolated concepts. When participants form isolated concepts, the characterization of one category is somewhat independent of other categories and relies on a relatively large use of nondiagnostic features (those shared by members of the category, but an imperfect basis for distinguishing categories). Categorization performance is better for the concept's prototype. By contrast, the characterization of interrelated concepts is influenced by other categories and is more dependent on features that are diagnostic of category membership. Relative to a prototype, categorization performance is better on a caricature, an exemplar that is more distant than the prototypes from other categories. Goldstone found that interleaved presentation of exemplars was more likely to produce interrelated concepts and heavier reliance on diagnostic features. In contrast, infrequent alternation of categories (as in the massed presentation) has the advantage of highlighting information that remains constant across the members within a category (Medin, Wattenmaker, & Michalski, 1987). Thus, massing favors isolated concepts. On the basis of Goldstone's work (1996; Goldstone et al., 2003), it appears that the paintings in Experiment 1 and the low-discriminability patterns in Experiment 2 supported relatively interrelated category concepts whose development was enhanced by interleaving.

It is likely that the structure of the categories also affects the nature of the category learning and, potentially, the impact of interleaving. The pictures devised in Experiment 2 have a rule-based category structure that can be learned via some explicit reasoning process (see Ashby, Maddox, & Bohil, 2002). In both discriminability conditions (high and low), it is possible to list the rule that defines the categories. For such category structures, often the rule that maximizes accuracy (i.e., the optimal rule) is easy to describe verbally (Ashby, Alfonso-Reese, Turken, & Waldron, 1998). In another type of category structure, called information-integration categories, the optimal rule to define these types of categories often is hard to describe verbally (Ashby et al., 1998). Learning the categories that have information-integration category structures requires one to integrate information from two or more stimulus dimensions at some predecisional stage in order to maximize accuracy (Ashby & Gott, 1988). It is suggested that perceptual integration could take many

forms, from computing a weighted linear combination of the dimensional values to treating the stimulus as a gestalt (Ashby et al., 2002).

Arguably, the paintings used in Experiment 1 have information-integration category structures. Ashby et al. (2002) found that especially for information-integration structures, asking participants to give category responses and providing accuracy feedback was more helpful than the observational training procedure used here and by Kornell and Bjork (2008). Thus, it is possible that the advantage of interleaving for the paintings depended on the use of a less optimal training regime. Nevertheless, it is likely that the effect of category discrimination difficulty observed in Experiment 2 is also an important factor in the beneficial effect of interleaving for the paintings used in Experiment 1. Given the absence of obvious feature rules for classifying the paintings, the category discrimination in Experiment 1 appears to be difficult (cf. Kornell & Bjork, 2008). Also, the large number of painting categories plausibly favors the formation of interrelated concepts. It remains to be seen whether information-integration and rule-based categories are differentially susceptible to the formation of interrelated concepts. It has been suggested that different cognitive systems mediate learning of the two category structures (Ashby et al., 1998), but this issue is under debate (Stanton & Nosofsky, 2007).

An additional factor in the effect of interleaving may be the within-category similarity of exemplars, with recent work (Carvalho & Goldstone, 2011) indicating that massed presentation has a benefit when the within-category exemplar similarity is low. This factor is relevant to Experiment 2, in that the complex rules devised for the difficult discrimination may have reduced the within-category similarity, as well as the category differences. However, given that Carvalho and Goldstone found that low within-category similarity was associated with benefits from massing, this factor does not appear to have affected the results. As was suggested above, it is likely that the category rules and diagnostic features encouraged the formation of interrelated concepts in Experiment 2. Taken together, the present and previous results suggest that the effect of interleaving depends on a number of variables that affect the relative ease with which participants can infer what links items within a category and can infer what differentiates categories.

The findings of the present study are of both practical and theoretical importance for inductive learning, which is generally important in education. From a practical perspective, the benefits of interleaving imply that inductive learning of categories that are difficult to distinguish would be more effective if the exemplars from different categories are interleaved, rather than if the exemplars from a category are presented consecutively, one after another. In teaching category-based related topics, educators may want to consider interleaving in

presenting their materials (e.g., cases or pictures), which can be achieved by shuffling examples from different categories, for example, in one learning session. The interleaving of exemplars and strategies designed to encourage inductive learning may be particularly important when category membership cannot easily be specified in a set of rules. In addition, the findings imply that, practically, students may benefit from learning high-discriminable categories (i.e., categories that are easy to discriminate) when the categories are presented in massed fashion. With respect to theoretical importance, the finding that interleaving is advantageous when the degree of discrimination is low and massing is beneficial when the degree of discriminability is high supports the suggestion of Kurtz and Hovland (1956) that has never been empirically tested in past studies of the spacing effect in inductive learning. The present experiments show that spacing works through interleaving and that interleaving is more beneficial when the category discrimination is difficult. The latter finding provides evidence for the proposition that the nature of the category discrimination is critical in understanding the effects of interleaving in induction.

Author note Norehan Zulkiply, University of Queensland; Jennifer S. Burt, University of Queensland.

We thank Dr. Nate Kornell from the Department of Psychology, Williams College, Williamstown, Massachusetts, USA for assistance with the stimulus materials for our first experiment, and Associate Prof. Dr. John McLean from Department of Psychology, University of Queensland, Australia for his helpful comments on the earlier version of the manuscript.

References

- Appleton-Knapp, S. L., Bjork, R. A., & Wickens, T. D. (2005). Examining the spacing effect in advertising: Encoding variability, retrieval processes and their interaction. *Journal of Consumer Research*, *32*, 266–276. doi:10.1086/432236
- Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *105*, 442–481.
- Ashby, F. G., & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 33–53.
- Ashby, F. G., Maddox, W. T., & Bohil, C. J. (2002). Observational versus feedback training in rule-based and information-integration category learning. *Memory and Cognition*, *30*, 666–677.
- Carvalho, P. F., & Goldstone, R. L. (November, 2011). *Stimulus similarity relations modulate benefits for blocking and interleaving during category learning*. Paper presented at the 52nd Annual Meeting of the Psychonomic Society, Seattle, WA.
- Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*, 354–380. doi:10.1037/0033-2909.132.3.354
- Childers, J. B., & Tomasello, M. (2002). Two-year-olds learn novel nouns, verbs, and conventional actions from massed or distributed exposures. *Developmental Psychology*, *38*, 967–978. doi:10.1037//0012-1649.38.6.967
- Cornell, E. H. (1980). Distributed study facilitates infants' delayed recognition memory. *Memory and Cognition*, *8*, 539–542. doi:10.3758/BF03213773
- Dellarosa, D., & Bourne, L. E., Jr. (1985). Surface form and the spacing effect. *Memory and Cognition*, *13*, 529–537. doi:10.3758/BF03198324
- Dempster, F. N. (1996). Distributing and managing the conditions of encoding and practice. In E. L. Bjork & R. A. Bjork (Eds.), *Memory: Handbook of perception and cognition* (pp. 317–344). San Diego: Academic.
- Donovan, J. J., & Radosevich, D. J. (1999). A meta-analytic review of the distribution of practice effect: Now you see it, now you don't. *Journal of Applied Psychology*, *84*, 795–805. doi:10.1037/0021-9010.84.5.795
- Ebbinghaus, H. (1985/1913). *Memory: A contribution to experimental psychology* (H. A. Ruger, & C. E. Bussenius, Trans.). New York: Columbia University.
- Ebbinghaus, H. (1964). In H. A. Ruger, C. E. Bussenius, & E. R. Hilgard (Eds.), *Memory: A contribution to experimental psychology*. New York: Dover Publications, Inc. Original work published in 1885.
- Gagne, R. M. (1950). The effect of sequence of presentation of similar items on the learning of paired-associates. *Journal of Experimental Psychology*, *40*, 61–73. doi:10.1037/h0060804
- Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition*, *7*, 95–112.
- Glenberg, A. M., & Lehmann, T. S. (1980). Spacing repetitions over 1 week. *Memory and Cognition*, *8*, 528–538.
- Glover, J. A., & Corkill, A. J. (1987). Influence of paraphrased repetitions on the spacing effect. *Journal of Educational Psychology*, *79*, 198–199.
- Goldstone, R. L. (1996). Isolated and interrelated concepts. *Memory and Cognition*, *24*, 608–628.
- Goldstone, R. L., Steyvers, M., & Rogosky, B. J. (2003). Conceptual interrelatedness and caricatures. *Memory & Cognition*, *31*, 169–180.
- Helsdingen, A. S., Van Gog, T., & Van Merriënboer, J. J. G. (2011). The effects of practice schedule on learning a complex judgment task. *Learning and Instruction*, *21*, 126–136. doi:10.1016/j.learninstruc.2009.12.001
- Hintzman, D. L., & Rogers, M. K. (1973). Spacing effects in picture memory. *Memory and Cognition*, *1*, 430–434.
- Hintzman, D. L. (1974). Theoretical implications of the spacing effect. In R. L. Solso (Ed.), *Theories in cognitive psychology. The Loyola Symposium* (pp. 77–99). Potomac: Erlbaum.
- Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.
- Kang, S. H. K., & Pashler, H. (2012). Learning painting styles: Spacing is advantageous when it promotes discriminative contrast. *Applied Cognitive Psychology*, *26*, 97–103. doi:10.1002/acp.1801
- Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the “enemy of induction”? *Psychological Science*, *19*, 585–592. doi:10.1111/j.1467-9280.2008.02127.x
- Kornell, N., Castel, A. D., Eich, T. S., & Bjork, R. A. (2010). Spacing as the friend of both memory and induction in young and older adults. *Psychology and Aging*, *25*, 498–503. doi:10.1037/a0017807
- Kurtz, K. H., & Hovland, C. I. (1956). Concept learning with differing sequences of exemplars. *Journal of Experimental Psychology*, *51*, 239–243.
- Medin, D., Wattenmaker, W. D., & Michalski, R. S. (1987). Constraints and preferences in inductive learning. *Cognitive Science*, *11*, 299–339.

- Melton, A. W. (1970). The situation with respect to the spacing of repetitions and memory. *Journal of Verbal Learning and Verbal Behavior*, 9, 596–606.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84, 231–259. doi:10.1037/0033-295x.84.3.231
- Rea, C. P., & Modigliani, V. (1987). The spacing effect in 4- to 9- year-old children. *Memory and Cognition*, 15, 436–443. doi:10.3758/BF03197733
- Rothkopf, E. Z., & Coke, E. U. (1966). Variations in phrasing and repetition interval and the recall of sentence material. *Journal of Verbal Learning and Verbal Behavior*, 5, 86–91. doi:10.1016/S0022-5371(66)80111-1
- Stanton, R. D., & Nosofsky, R. M. (2007). Feedback interference and dissociations of classification: Evidence against the multiple-learning-systems hypothesis. *Memory and Cognition*, 35, 1747–1758. doi:10.3758/bf03193507
- Toppino, T. C. (1993). The spacing effect in preschool children's free recall of pictures and words. *Bulletin of the Psychonomic Society*, 31, 27–30.
- Underwood, B. J. (1952). An orientation for research on thinking. *Psychological Review*, 59, 209–220.
- Vlach, H. A., Sandhofer, C. M., & Kornell, N. (2008). The spacing effect in children's memory and category induction. *Cognition*, 109, 163–167. doi:10.1016/j.cognition.2008.07.013
- Wahlheim, C. N., Dunlosky, J., & Jacoby, L. L. (2011). Spacing enhances the learning of natural concepts: An investigation of mechanisms, metacognition, and aging. *Memory and Cognition*, 39, 750–763. doi:10.3758/s13421-010-0063-y
- Wulf, G., & Shea, C. H. (2002). Principles derived from the study of simple skills do not generalize to complex skill learning. *Psychonomic Bulletin and Review*, 9, 185–211. doi:10.3758/BF03196276
- 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.
- Zulkipli, N., McLean, J., Burt, J. S., & Bath, D. (2012). Spacing and induction: Application to exemplars presented as auditory and visual text. *Learning and Instruction*, 22, 215–221. doi:10.1016/j.learninstruc.2011.11.002