

INDUCTIVE LEARNING: DOES INTERLEAVING EXEMPLARS AFFECT LONG-TERM RETENTION?

Norehan Zulkipli¹

*Faculty of Cognitive Sciences and Human Development
Universiti Malaysia Sarawak*

Jennifer S. Burt

*School of Psychology
University of Queensland, Australia*

¹Corresponding author: znorehan@fcs.unimas.my

ABSTRACT

Purpose – The present study investigated whether or not the benefits of interleaving of exemplars from several categories vary with retention interval in inductive learning.

Methodology – Two experiments were conducted using paintings (Experiment 1) and textual materials (Experiment 2), and the experiments used a mixed factorial design. Forty students participated in each experiment for course credit. In each experiment, participants studied a series of exemplars from several categories which were presented massed and interleaved, and later their induction was tested either shortly after the study phase (short-term retention) or after a week's delay (long-term retention).

Findings – Consistent with findings from previous studies, the interleaving effect was found in the short-term retention condition, and crucially, the present study provided the initial evidence that interleaving of exemplars also affected long-term retention. Interestingly, massing was judged to be more effective than spacing (interleaving) in most groups, even when actual performance showed the opposite.

Significance – The present study shows that interleaved exemplars have considerable potential in improving inductive learning in the long term. For example, induction is used in case-based reasoning

which requires one to start with learning from specific cases, and then form generalizations of these cases by identifying the commonalities between them. In order to enhance long-term retention, educators may want to consider using interleaved presentation rather than massed presentation in teaching examples or cases from a particular category or concept.

Keywords: Interleaving effect, inductive learning, category learning, category induction, spacing effect

INTRODUCTION

It is an established experimental finding that memory retention for spaced repeated items is better than massed repeated items (e.g., Cepeda, Pashler, Vul, Wixted & Rohrer, 2006; Donovan & Radosevich, 1999; Melton, 1970). The finding of improved memory for spaced repetitions, called the spacing effect, has been documented in a broad range of memory tasks with many different types of study materials (Cepeda et al., 2006; Dempster, 1996; Donovan & Radosevich, 1999), including nonsense syllables (e.g., Ebbinghaus, 1985/1913), pictures (e.g., Hintzman & Rogers, 1973), words (e.g., Glenberg & Lehmann, 1980), sentences (e.g., Rothkopf & Coke, 1966) and faces (e.g., Cornell, 1980). The spacing effect is not only found when the test is given shortly after the study phase, which measures short-term memory retention (e.g., Rea & Modigliani, 1987; Toppino, 1993) but also when the test is given after a delay interval which measures long-term memory retention, ranging from days (e.g., Ausubel, 1966; Cepeda, Coburn, Rohrer, Wixted, Mozer & Pashler, 2009 (i.e., Experiment 1), to months (e.g., Bahrick & Phelps, 1987; Cepeda et al., 2009 (Experiment 2a)). Clearly, long-term memory retention benefits from spaced repetitions.

In inductive learning, it is generally not known whether or not spaced presentation of exemplars from the same categories affects long-term retention. In their influential book on induction, Holland, Holyoak, Nisbett, and Thagard (1986) define induction as ‘all inferential processes that take place in the face of uncertainty’ (p.1). In other words, induction is concerned with inferring knowledge from an incomplete set of observations, and this contrasts with deduction,

where the learner formulates regularities observed in a complete set of data (Murphy, 2002). Inductive learning of categories is the process of learning through examples, whereby students work from specific exemplars and derive general concepts or categories from those exemplars.

Several studies have been conducted to investigate the spacing effect on inductive learning to date. In a typical study that tests the spacing effect in inductive learning, participants are asked to learn exemplars from several categories which are presented with a variable degree of spacing between exemplars, and at the end of the session induction is tested on a set of novel exemplars from the same categories learnt in the study phase. Past studies have produced contradictory results. In two earlier studies, massing was superior to spacing (Gagne, 1950 ; Kurtz & Hovland, 1956). Less direct evidence that massing facilitates induction comes from experiments that compared exact and non-exact repetitions (e.g., Appleton-Knapp, Bjork, & Wickens, 2005; Dellarosa & Bourne, 1985; Glover & Corkill, 1987; Melton, 1970), and research on motor learning which involves learning complex motor skills (Wulf & Shea, 2002). Interestingly, a few recent studies showed the opposite finding, that spacing improves induction (Kang & Pashler, 2012; Kornell & Bjork, 2008; Kornell, Castel, Eich, & Bjork, 2010; Vlach, Sandhofer, & Kornell, 2008; Wahlheim, Dunlosky, & Jacoby, 2011; Zulkiply, McLean, Burt, & Bath, 2012). All of these previous research studies, including those that found the spacing effect in inductive learning, tested induction shortly after the study phase, thus measuring the spacing effect over short-term retention (e.g., Kornell & Bjork, 2008; Kornell et al., 2010). A recent study by Kang and Pashler (2012) and Zulkiply and Burt (2013) suggested that, in the learning of painting styles, it is interleaving of exemplars from different categories (and not temporal spacing itself) that is critical to the spacing effect.

Past studies that found the spacing effect in inductive learning tested induction over a short retention interval, whereby the test was given shortly after the study phase (e.g., Kang & Pashler, 2012; Kornell & Bjork, 2008; Kornell et al., 2010; Vlach et al., 2008). It is not known whether or not spacing different individual exemplars apart in time aids in the retained learning of categories over a longer time interval.

THE PRESENT STUDY

The primary aim of the present study was to investigate whether or not the benefits of interleaving varied with retention interval. As mentioned earlier, Kang and Pashler (2012) and Zulkipli and Burt (2013) provided evidence that, it is interleaving of exemplars from different categories (and not temporal spacing itself) that is critical to the spacing effect. Two experiments using paintings (Experiment 1) and textual materials (Experiment 2) were conducted. The methods followed those developed by Kornell and Bjork (2008), except that that the term *spacing* that Konell and Bjork (2008) used to describe their manipulation was replaced by the term *interleaving*. Both types of stimulus presentation used in the present study (i.e., paintings and texts) are important because of their educational relevance to university teaching and learning. Pictures and textual materials are commonly used to present information, as in a lecture or speech, and understanding under which circumstance (massed condition or interleaved condition) the material works better will be beneficial to students. The present study also aimed to examine participants' judgement towards massing and spacing (interleaving)— which one is judged to be more helpful in the learning of categories in inductive learning in particular, in the long-term. In light of previous findings (e.g., Kornell & Bjork, 2008; Kornell et al. 2010), it was hypothesized that the majority of participants would report massing to be more helpful to learning than spacing, even in the long term. The focus of the present study was on the inductive learning that occurs during category learning.

EXPERIMENT 1

Experiment 1 examined whether or not the interleaving of artists enhances learning of their paintings in the long term.

METHODOLOGY

Participants and Design

Forty students (29 females, 11 males) from an introductory psychology class participated in the experiment for course credit. The design of the experiment was a 2 (Presentation style: Massed

vs. Interleaved) x 2 (Retention type: Short-term vs. Long-term) x 4 (Test block: Blocks 1- 4) mixed-factorial design. Retention type 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 phase, test phase and question phase. In the presentation phase, 72 paintings by each of six of the artists were presented consecutively (massed), whereas the paintings by each of the other six artists were intermingled with paintings by other artists (interleaved). The paintings were arranged in 12 learning blocks (six blocks of massed presentation, six blocks of interleaved presentation). The order of the blocks was MIIMMIIMMIIM (M for massed; I for interleaved). The assignment of artists 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 artists to blocks. Thus, there were four lists in total. In the test phase, the 48 new paintings by the artists were arranged in four test blocks. Each block consisted of one new painting by each of the 12 artists, presented in a fixed order across participants.

Materials

The materials were 120 paintings from 12 different artists taken from Kornell and Bjork (2008) (See Appendix A for samples of paintings). As noted, 72 paintings were used in the presentation/study phase (six paintings per artist), and another 48 paintings were used in the test phase (four paintings per artist). 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 the painting files were either landscapes or skyscapes. The paintings were in the format of JPEG file and were resized to fit into a 19 cm x 29 cm rectangle on the computer screen.

Procedure

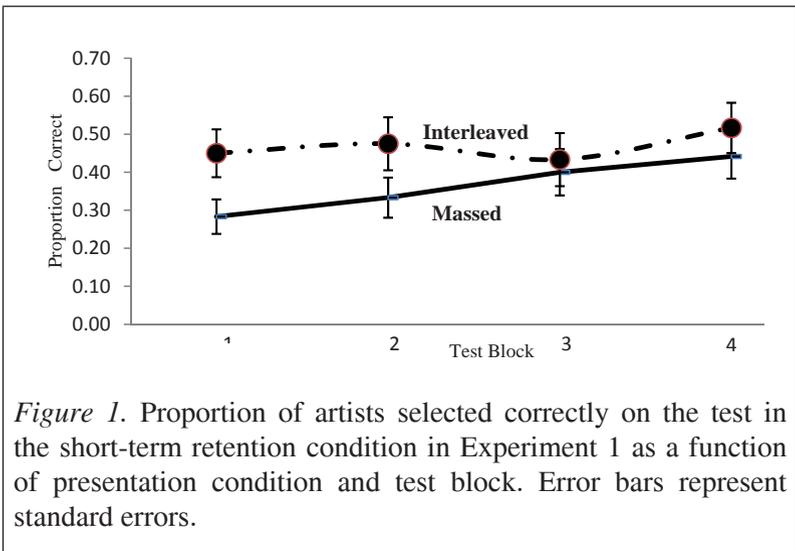
Participants were randomly assigned to either the short-term retention condition or the long-term retention condition. Participants first were instructed about the nature of the experiment before they entered the presentation phase which was subject to experimental manipulation. In the presentation phase, participants were asked to

study the 72 paintings by twelve artists. Each painting was shown on a computer screen for three seconds, with the last name of the artist displayed underneath. When participants studied the paintings, they were asked to learn to recognize which artist painted which picture, based on the artists' style. Next, participants were asked to complete a distractor task, during which they were asked to count backwards by 3s starting from 612, for 15 seconds, while typing the numbers in the designated box on the computer screen. For the test phase, participants in the short-term retention condition were tested immediately after the end of the distractor task, whereas participants in the long-term retention condition were tested a week later. During the test phase, participants were shown 48 new paintings, which they had not seen before, and they were required to identify the artist. Participants were shown one painting at a time on a computer screen, with 13 buttons below the painting. Twelve of the buttons were labelled with the artists' names and one button was labelled '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' would appear in the middle of the computer screen. If they responded incorrectly, the correct artist's name would be presented on the computer screen. Participants completed the test phase at their own pace. 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'. The question phase which followed Kornell and Bjork's (2008) approach ended the experimental manipulation. Participation in the experiment took approximately 30 minutes and participants were debriefed about the experiment they had just participated in before they left the experimental room.

Results

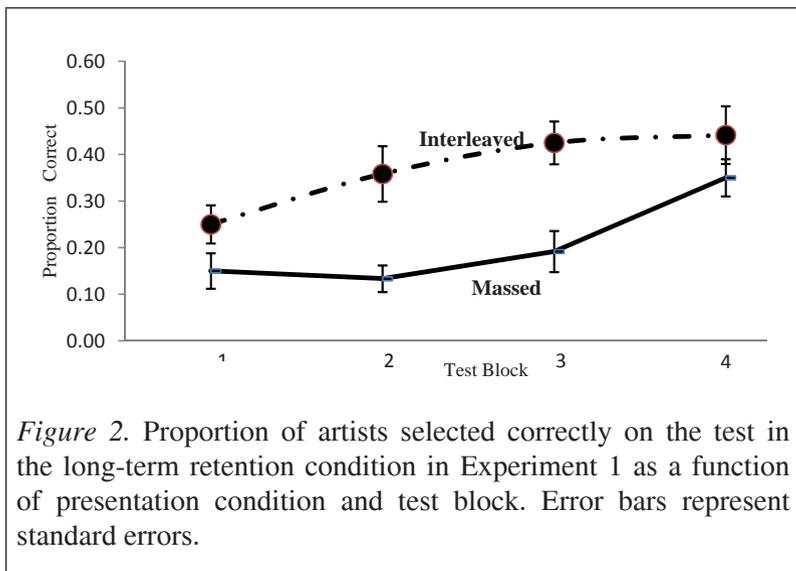
A three-way repeated measures ANOVA was conducted on the data for Experiment 1. As shown in both Figure 1 and Figure 2, the effect of presentation style was significant. Specifically, participants' performance in spaced (interleaved) study was significantly better than their performance in massed study in the short and long-term retention conditions, $F(1, 38) = 17.43$, $p < .001$, $\eta_p^2 = .31$, and participants' accuracy also increased significantly across test blocks,

$F(3,114) = 14.48, p < .001, \eta_p^2 = .28$. The main effect of retention interval was also significant, $F(1, 38) = 5.29, p = .027, \eta_p^2 = .12$. Performance was better when testing was made shortly after the study phase than after a week's delay. The three-way interaction between the variables was significant, $F(3, 114) = 2.92, p = .037, \eta_p^2 = .07$. This indicates that participants' performance was different over the short retention condition and the long retention condition, and across test blocks depending on whether the exemplars were presented massed or interleaved. Specifically, the difference between massed and interleaved was greatest during the initial blocks for the short-term retention group and greatest during the middle two blocks for the long-term retention group. None of the two-way interactions was significant (presentation style and retention type, $F(1, 38) = .83, p = .367$; test block and retention type, $F(3, 114) = 1.35, p = .262$; presentation style and test block, $F(3, 114) = 1.53, p = .210$)



With regards to participants' judgement of which study presentation helped them learn more, a similar preference for massed presentation was observed in both retention conditions. A one-way Chi-square analysis was conducted to compare the proportion of participants who judged massed to be more useful, with the proportion of participants preferring spaced (interleaved) and the proportion judging that the two conditions contributed equally in helping them to learn more during the study phase. As predicted, the result for the

short term retention condition is consistent with previous findings (e.g., Kornell & Bjork, 2008; Kornell et al., 2010), $\chi^2(2, N=20) = 6.70, p = .035$. Of a total of 20 participants, a majority of 12 (60%) claimed massed presentation was better, five (25%) claimed spaced and three (15%) judged that both massed and spaced presentations contributed equally in helping them to learn during the learning phase, regardless of their performance in the two conditions. In terms of categorisation performance, 13 (65%) of participants performed better in spaced (interleaved) condition, five (25%) performed better in massed condition and two (10%) performed equally in the two conditions.



A similar pattern of judgement was observed in the long retention interval condition. Of a total of 20 participants, a majority of 10 (50%) participants claimed massed was more effective, four (20%) claimed spaced and another six (30%) judged the two conditions equally effective, regardless of their performance in the two conditions, massed and spaced. Nevertheless, the result of a one-way Chi-square analysis conducted on the judgement data was not significant, $\chi^2(2, N=20) = 2.80, p = .247$. In terms of categorisation performance, 14 (70%) of the participants performed better in spaced (interleaved) condition, four (20%) performed better in massed condition and two (10%) performed equally in the two conditions.

Summary of Experiment 1

The benefits of interleaved presentation over short-term retention demonstrated in Experiment 1 are consistent with the findings from past studies (e.g., Kang & Pashler, 2012; Kornell & Bjork, 2008; Kornell et al., 2010; Zulkiply et al., 2012; Zulkiply & Burt, 2013). Interestingly, the advantage of interleaving was also observed over long-term retention. In both retention conditions, performance in the interleaved condition surpassed performance in the massed condition. However, performance after the short retention interval was generally better than performance after the long retention interval, the latter performance likely being due to forgetting. This findings are discussed further in the 'Discussion' section. Performance over the test blocks for both massed and interleaved conditions generally improved and this could be due to the accuracy feedback the participants received after each test trial.

On the post-experimental questionnaire, the majority of participants in both retention conditions appeared to believe that the massed presentation made it easier to recognise the style of each individual artist during the presentation or study phase. However this effect was not significant in the long-term retention group.

EXPERIMENT 2

Experiment 2 was conducted to investigate further whether or not interleaving affects long-term retention. For this purpose, textual materials were used as the stimuli in Experiment 2, as in Zulkiply et al. (2012).

Participants and Design

Participants were 40 students (22 females, 18 males) from an introductory psychology class who received course credit to participate in the experiment. All major aspects of the design of Experiment 2 were identical to those in Experiment 1 except that the stimuli used in the presentation phase were visually presented texts (i.e., case studies) and, with only 36 cases studies from six psychopathological categories (six case studies per category), a slightly different way of arranging the cases in the learning blocks

(i.e., in the presentation phase) and in the test blocks was implemented here. In the presentation phase, the 18 cases were arranged in six learning blocks (three blocks for massed presentation, three blocks for interleaved presentation). The order of the blocks was 'MIMIMI' (M for massed; I for interleaved). The assignment of psychological disorders 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 disorders to blocks. Thus, there were four lists in total. In the test phase, the 18 new cases from the six psychopathological categories were arranged in three test blocks. Each block consisted of one new case from each category, presented in a fixed order across participants.

Materials

The materials were 36 case studies developed from six categories of psychopathological disorders as in Zulkipli et al. (2012) (see Appendix B for samples of case studies). As noted, 18 cases were used in the presentation/study phase (three cases per category) and another 18 cases were used in the test phase (three cases per category). The psychopathological disorder categories used were identified by nonsense names to minimise the effects of participants' prior assumptions and expectations. Table 1 illustrates the six disorder categories chosen as the basis of the case studies as well as the novel names assigned to each of the categories. Each case study was between 100 and 120 words in length and incorporated a description of a few symptoms representative of the four factors of symptoms in general: Cognitive, Behavioural, Emotional, Physical. All the case studies used in Experiment 2 were pilot-tested by ten Clinical PhD students. Examples of case studies can be found in Zulkipli et al. (2012).

Procedure

The procedure for Experiment 2 was identical to that used in the previous experiment, except that, in the presentation phase, each case study was visually presented on the computer screen for approximately 30 seconds. While participants studied each of the case studies, the novel label of the corresponding category was displayed underneath each case study.

Table 1

Categories and Novel Names of Psychopathological Disorders Used to Develop the Case Studies for Experiment 2

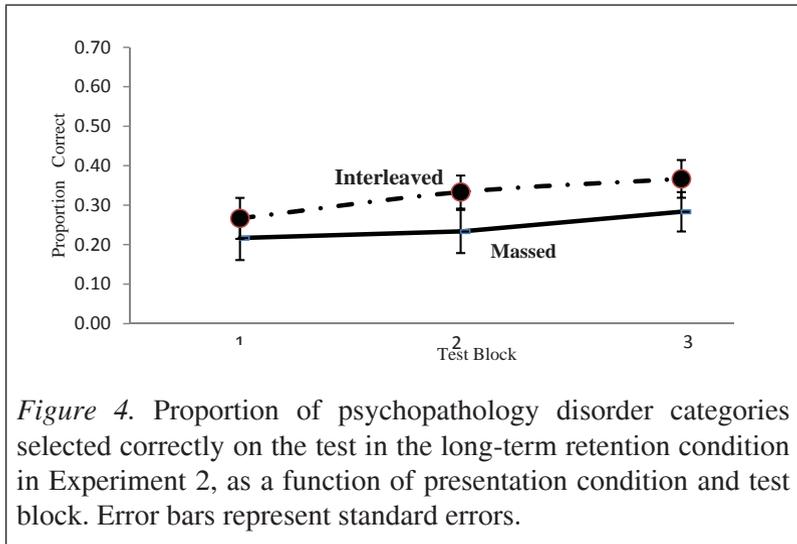
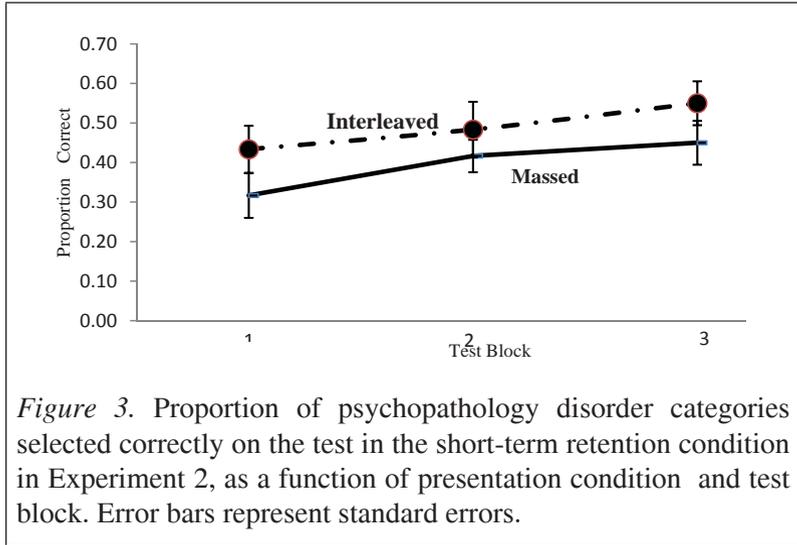
Categories of psychopathological disorders	Novel names assigned
Obsessive Compulsive Disorder	Duv
Phobia Disorder	Baj
Schizophrenia Disorder	Tem
Attention Deficit Disorder (Inattentive Type)	Pliq
Attention Deficit Disorder Hyperactive and Impulsive Type	Hix
Depression Disorder	Wos

Results

The data were analysed with a three-way repeated measure ANOVA. As shown in both Figure 3 and Figure 4, participants' performance in interleaved study was significantly better than their performance in massed study, $F(1, 38) = 5.61, p = .023, \eta_p^2 = .13$, and participants' accuracy also increased significantly across test blocks, $F(1.7, 64.66) = 4.58, p = .018, \eta_p^2 = .11$. The main effect of retention interval was also significant, $F(1, 38) = 19.443, p < .001, \eta_p^2 = .34$. Consistent with the findings in Experiment 1, performance over short retention interval was better than performance over long retention interval. All the two-way and three-way interaction effects were not significant (presentation style and retention type, $F(1, 38) = .053, p = .820$; test block and retention type, $F(1.7, 64.66) = .20, p = .780$; presentation style and test block, $F(2, 76) = .01, p = .990$; presentation style, test block and retention type, $F(2, 76) = .26, p = .773$;

Interestingly, participants' responses to the questionnaire administered after the test in both retention conditions revealed consistency with the previous experiment. In a short-term retention condition, a one-way Chi-square analysis showed a significant difference among the three judgement options, $\chi^2(2, N=20) = 10.9, p = .004$. Of a total of 20 participants, a majority of 13 (65%) claimed massed presentation was better, six (30%) claimed spaced and one (5%) judged that both massed and spaced contributed equally in helping them to learn during the learning phase, regardless of their performance in the two conditions. In terms of categorisation

performance, 12 (60%) of the participants performed better in spaced (interleaved) condition, seven (35%) performed better in massed condition and one (5%) performed equally in the two conditions.



Similarly, in the long-term retention condition, of a total of 20 participants, a majority of 14 (70%) participants claimed massed was more effective, one (5%) claimed spaced (interleaved) and another five (25%) judged the two conditions equally effective, regardless

of their performance in the two conditions. The result of a one-way Chi-square analysis conducted on the judgement data confirmed our prediction, $\chi^2(2, N=20) = 13.3, p = .001$. In terms of categorisation performance, nine (45%) of the participants performed better in spaced (interleaved) condition, three (15%) performed better in massed condition and eight (40%) performed equally in the two conditions.

Summary of Experiment 2

Using visually presented texts as the learning material, Experiment 2 revealed findings that are consistent with Experiment 1—performance was significantly better in the interleaved condition over both retention conditions and significantly increased over the test block. Performance was generally better when induction was tested shortly after the study phase than after a one week's delay. Participants also demonstrated a similar pattern of judgement towards massed presentation in both retention conditions.

DISCUSSION

Consistent with findings from past studies (e.g., Kang & Pashler, 2012; Kornell & Bjork, 2008; Kornell et al., 2010; Zulkipli et al., 2012; Zulkipli & Burt, 2013), the present study revealed that short-term retention benefits from interleaved presentation in inductive learning. Most importantly, the present study provides initial evidence that interleaved presentation also affects long-term retention (i.e., when induction was tested after a week). Interleaving had similar effects on induction in the long- and short-retention groups, although overall accuracy at the beginning of the test session was lower for the long-retention group.

There are a number of possible reasons for the advantage of interleaved presentation in induction. The first possible account for the benefits of interleaving found in the present study concerns attention and the way it affects massed and interleaved exemplars. Hintzman (1974) suggests that massing impairs learning by reducing the amount of attention people pay to repeated presentations, because the massed items become highly familiar. Further, according to the attention attenuation hypothesis (from Kornell et al., 2010;

Walheim et al., 2011), recall of massed items is impaired because it is difficult to pay full attention to the second (and subsequent) presentations of massed items. In inductive learning, there is a possibility that participants might have believed that massed presentation of exemplars from a particular category made it easier for them to discern a painter's style (Experiment 1) or to identify a particular psychopathological category (Experiment 2), resulting in those massed exemplars being given less attention and receiving less processing time compared to spaced (interleaved) exemplars. It is likely also that attention is weakened across massed exemplars from the same category (Wahlheim et al., 2011). In contrast, when the exemplars were interleaved with exemplars from several other categories, participants may have paid more attention to and more deeply processed the exemplars.

The second possible account for the benefits of interleaving in inductive learning lies in the association between induction and discrimination processes (e.g., Kornell & Bjork, 2008). The interleaving of exemplars from different categories is thought to have facilitated comparison and contrast, as well as fostered and enhanced discrimination learning, allowing participants to notice the different characteristics of the painting exemplars (Experiment 1) and the different natures of the case studies (Experiment 2)—this would be expected to assist participants in understanding the style of each artist and the nature of each psychopathological category. Several authors have suggested that interleaving promotes the apprehension of points of contrasts among exemplars, making these differences among the categories more salient (e.g., Goldstone, 2003; Kang & Pashler, 2012).

The third possible explanation for the benefits of interleaving in induction is the likelihood of study-phase retrieval. When an item is presented, previous presentations of the same item may be retrieved from memory and this retrieval process enhances learning (from Kornell et al., 2010). It is argued that the more difficult the retrieval, the more learning is enhanced (Bjork & Allen, 1970; Krug, Davis & Glover, 1990). In inductive learning, the exemplars from the same category are different from each other, thus likely are more difficult to be learnt. Interleaving of exemplars from several categories might have increased the difficulty of retrieval of exemplars from the same category which then enhanced induction.

In both experiments of the present study, participants' performance in the short-term retention condition was better than in the long-term retention condition. A comparison of the proportion mean scores of the first test block between the two retention conditions (see Figure 1 and Figure 2 for Experiment 1, and see Figure 3 and Figure 4 for Experiment 2), shows a decline in participants' performance when testing was delayed a week— score decreases from 0.45 to 0.25 in Experiment 1 and score decreases from 0.43 to 0.27 in Experiment 2, indicating forgetting for the long retention group.

The test with feedback introduced in the present study supported some new learning in all groups. In the test, participants saw a series of novel exemplars from the same categories learnt in the study phase and they received feedback on each test trial, including the correct category name when they made an error. Apparently, for the long-term groups, at test a week later, the participants showed a benefit of returning to the learning context as their performance improved over the test blocks. Memory is enhanced when the situation during study closely resembles the situation during test (Fisher & Craik, 1977). Reinstating experiences such as a physical context (e.g., Smith, 1979) and emotional mood (e.g., Bower, Monteiro, & Gilligan, 1978) can enhance remembering. It is likely that context reinstatement might have helped the participants, particularly in the long-term groups, to retrieve forgotten material and further enhance their performance during the test session. Specifically, participants may have used context information as a source of memory cues to enhance their memory about the exemplar categories. There was some support for this idea in Experiment 1 with the three-way interaction between the variables showing a bigger involvement later in the test session for the long-term group. However this result was not observed in Experiment 2 with texts. It is unclear whether the difference reflects greater encoding difficulty with texts or the smaller number of categories in Experiment 2.

Furthermore, comparing participants' performance for spaced exemplars as in Figure 1 and Figure 2 (for Experiment 1), and as in Figure 3 and Figure 4 (for Experiment 2), it is noted that the mean score of the last test block for the long-term retention condition is approaching the mean score of the first test block in the short retention condition (i.e., 0.44 to 0.45 in Experiment 1, and 0.37 to 0.43 in Experiment 2). This pattern of results seems to describe a

limit on the learning achievable after one study session, whereby it is likely that no further score could be reached upon completing the final test block in the long-term retention condition in both experiments. It is possible that by giving the participants more training such as via a repeated study phase (i.e., more than one study session) or a repeated test block (either in a single study session or more), participants may be able to achieve near perfect accuracy. In the present studies they were reaching a plateau at well below 100%. In addition, tests with feedback are beneficial as they reveal which items have been sufficiently learnt and which ones require further study (Carpenter, Pashler, Wixted, & Vul, 2008). Retention can be enhanced when the same information is recalled more than once (Carpenter et al., 2008; Kuo & Hirshman, 1996; Karpicke & Roediger, 2007). Thus, giving a test with feedback regularly in a series of successive tests (e.g., in the first, second and third weeks after the study phase) may also improve accuracy in induction.

Consistent with past studies (e.g., Kornell and Bjork, 2008; Kornell et al., 2010, Zulkiply et al., 2012), in both experiments of the present study, specifically in the short-term retention condition, the majority of the participants judged massing to be more effective than spacing, although their actual performance revealed the opposite. It is likely that because of the consecutive presentations in the massed condition, participants developed a sense of familiarity for the exemplars of a similar nature from a particular category. This could have led participants to perceive that the massed presentation required less effort and to infer that the learning task was easier in the massed condition. The participants then may have believed that their learning outcomes in the massed condition were better than in the spaced condition and perceived massed presentation as more helpful in learning the categories. Interestingly, a similar preference for massed presentation was also observed in the long-term retention condition in both experiments.

CONCLUSION

Induction is used in everyday life, for instance, to make predictions and choices based on our observations or provided facts. Induction is also used to discover something new. For example, in science, induction is a basic procedure followed to make scientific

discoveries, and this is achieved by making systematic observations (which can include observations of a real event or phenomenon and observations from laboratory experiments). In category learning, induction is used to draw conclusions about the category in general (Murphy, 2002). For example, one decides that ripened bananas generally have dark spots or patches on their skin based on a few observed examples – this is induction rather than deduction because it involves drawing an uncertain inference to the category as a whole. Induction is also the kind of reasoning that one uses to solve problems such as in case-based reasoning, which requires one to start from specific cases, and then form generalization of these cases by identifying the commonalities between them. To identify a novel case (i.e., the target problem), one compares it to the examples of cases previously learned (i.e., retrieved cases). The present study shows that interleaved exemplars have considerable potential in improving inductive learning in the long term. Given the initial evidence from the present study, that induction benefits from interleaved presentation over long-term retention, a systematic approach in inductive learning of categories could be planned and implemented by educators in order to achieve the optimal benefits of the interleaving effect in inductive learning for long-term retention. In particular, educators may want to consider using interleaved presentation rather than massed presentation in teaching examples or cases (as in case-based reasoning) from a particular category or concept.

ACKNOWLEDGEMENT

We thank Dr. Nate Kornell from the Department of Psychology, Williams College, Williamstown, Massachusetts, USA for assistance with the stimulus materials for Experiment 1.

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
- Ausubel, D. P. (1966). Early versus delayed review in meaningful learning. *Psychology in the Schools*, 3, 195-198.

- Bahrick, H.P., & Phelps, E. (1987). Retention of Spanish vocabulary over 8 years. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 344–349.
- Bower, G. H., Monteiro, K. P., & Gilligan S. G. (1978). Emotional mood as a context for learning and recall. *Journal of Verbal Learning and Verbal Behavior*, *17*, 573–585.
- Bjork, R. A., & Allen, T. W. (1970). The spacing effect: Consolidation or differential encoding. *Journal of Verbal Learning & Verbal Behavior*, *9*, 567–572.
- Brainerd, C.J., & Reyna, V.F. (2002). Fuzzy-trace theory and false memory. *Current Directions in Psychological Science*, *11*, 164–169. doi: 10.1111/1467-8721.00192
- Carpenter, S. K., Pashler, H., Wixted, J. T., & Vul, E. (2008). The effects of tests on learning and forgetting. *Memory and Cognition*, *36*, 438–448. doi: 10.3758/MC.36.2.438
- 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
- Cepeda, N.J., Coburn, N., Rohrer, D., Wixted, J.T., Mozer, M.C., & Pashler, H. (2009). Optimizing distributed practice: Theoretical analysis and practical implications. *Experimental Psychology*, *56*, 236–246.
- Cornell, E.H. (1980). Distributed study facilitates infants' delayed recognition memory. *Memory and Cognition*, *8*, 539–542. doi:10.3758/BF03213773
- Dellarosa, D., & Bourne Jr, L.E. (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 and R. A. Bjork (Eds.), *Memory: Handbook of perception and cognition* (pp.317–344). San Diego, CA: Academic Press.
- 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.
- Fisher, R. P., & Craik, F. I. (1977). Interaction between encoding and retrieval operations in cued recall. *Journal of Experimental Psychology: Human Learning and Memory*, *3*, 701–711.

- 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., & 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. (2003). Learning to perceive while perceiving to learn. In R. Kimchi, M. Behrmann, & C. Olson (Eds.), *Perceptual organization in vision: Behavioral and neural perspectives* (pp. 233–278). Mahwah, NJ: Lawrence Erlbaum Associates.
- 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, MD: Erlbaum.
- 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
- Karpicke, J. D., & Roediger, H. L. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language*, 57, 151–162. doi: 10.1016/j.jml.2006.09.004
- 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
- Krug, D., Davis, B., & Glover, J. A. (1990). Massed versus distributed repeated reading: A case of forgetting helping recall? *Journal of Educational Psychology*, 82, 366–371.
- Kurtz, K.H., & Hovland, C.I. (1956). Concept learning with differing sequences of exemplars. *Journal of Experimental Psychology*, 51, 239-243.
- Kuo, T., & Hirshman, E. (1996). Investigations of the testing effect. *American Journal of Psychology*, 109, 451–464. doi: 10.2307/1423016

- 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.
- 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
- Smith, S. T. (1979). Remembering in and out of context. *Journal of Experimental Psychology: Human Learning and Memory*, 5, 460-471.
- 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.
- 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
- Zulkipli, N., & Burt, J. S. (2013). The exemplar interleaving effect in inductive learning: Moderation by the difficulty of category discriminations. *Memory and Cognition*, 41, 16-27. doi:10.3758/s13421-012-0238-9.
- 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

Appendix A

Examples of paintings from Experiment 1 (as in Kornell & Bjork, 2008)

Artist: Yie Mei

Sample 1:



Sample 2:



Artist: Pessani

Sample 1:



Sample 2:



Appendix B
Examples of Case Studies from Experiment 2
(as in Zulkipli et al., 2012)

Category TEM (Schizophrenia Disorder)

Sample 1:

Wills, 35 years old, is a successful businessman but lately, his behavioural changes seemed to affect his relationship with clients. Since 6 months ago, he had begun to hear voices that tell him he is not a good man. He has begun to talk to himself about how bad he is during meetings with clients. This has affected his relationship with clients. At the office, his workers were shocked by his very rapid changing mood, from happy to sad to angry, for no apparent reason. When he talked, it seemed that he was having thought disturbances, as he mixed up unrelated issues and could not connect his thoughts logically. He also keeps rolling up his tongue and that is somewhat annoying to his workers.

Sample 2:

Melinda, a 40-year old lady complained to her neighbours that she was fearful, depressed, and couldn't get off to sleep at night. She said she had been seeing her late mother lately, and her mother told her that her husband was going to hurt her badly. Melinda's husband was confused with Melinda's unusual behaviour, such as staring at him and locking herself in another room at night to avoid him. Two weeks later, Melinda ran away and stayed with her friend. While there, she wrote a letter to her husband saying that she was protected by a superpower and can never be hurt by anybody.

Category PLIQ (Attention Deficit Disorder- hyperactive and impulsive type)

Sample 1:

Ben, 11 years old, is a cheerful child, who often has problems in concentrating and following instructions by his teacher at school. When he does his schoolwork, he will make one or two scribbles on it and then he will start to giggle and whisper with his classmates. At

home, Ben often fails to complete the house work assigned to him by his parents. For exemplar, when asked to clean up his room, he does it for a minute and then does something else which will also be left unfinished. Lately, Ben also complains that he often feels hot and he drinks more water than he usually does.

Sample 2:

Maria, age 4, had problems at preschool. Her teacher said that she seemed disorganized and inattentive when performing school activities. When she drew something, her teacher had to repeat instructions, and Maria always left half-finished drawings all over her classroom. When she played at a puzzle, she did it half way, and then she left the incomplete work for other activities. Another thing that has become apparent in Maria lately is that her hands sweat a lot though she was just doing relaxing activities. Her teacher also found that she can get very angry if she can't get what she wants, e.g., when she wants to play on the swing but the swings are all occupied.

