Long Lasting Effects of an Instructional Intervention on Interleaving Preference in Inductive Learning and Transfer

All data and experimental stimuli contained in this study are publicly available at OSF, and the web-link is provided on the cover page.

Number of words: 11,961 words (journal requirement: ≤ 1,2000 words)

Number of pages: 50 pages (journal requirement: ≤ 50 pages)

Abstract length: 176 words (journal requirement: 150 – 250 words)

Number of keywords: 5 keywords (journal requirement: ≤ 6 keywords)

Abstract

Observing category exemplars in an interleaved manner is more beneficial for inductive learning than blocked (massed) presentation, a phenomenon termed the *interleaving effect on inductive learning*. However, people tend to erroneously believe that massed is more beneficial than interleaved learning, and learners prefer the former during self-regulated learning. We report four experiments designed to investigate whether explicit instructions, which include individual performance feedback and the interleaving effect results from previous research, can (1) correct metacognitive illusions regarding the interleaving effect, (2) promote self-employment of interleaving, and (3) facilitate category learning. In addition, the current study explored (4) whether the intervention effect is long-lasting and (5) transferable to learning of categories in other domains. Experiments 1-4 established the effectiveness of the instruction intervention to enhance metacognitive appreciation of the interleaving effect, to promote self-employment of interleaving, and to facilitate learning of new categories. The intervention effect was long-lasting (at least 24 hours; Experiment 2), and transferable to learning of categories in different domains (Experiments 3 & 4). These findings support the practical use of the instruction intervention.

Keywords: inductive learning; interleaving effect; instruction intervention; transfer; metacognitive awareness

Inductive learning (including category learning) is an essential vehicle for people to acquire general knowledge about the world (Holland, Holyoak, Nisbett, & Thagard, 1989). Numerous studies have explored potential strategies for promoting inductive learning (e.g., Choi & Lee, 2020; Jacoby, Wahlheim, & Coane, 2010; Yang & Shanks, 2018). Many of these have found that interleaved learning (i.e., arranging exemplars to be interleaved and temporally separated) is a more effective strategy for inductive learning than massed learning (i.e., arranging exemplars from the same category in blocks), a phenomenon termed the *interleaving effect on inductive learning* (for recent reviews, see Brunmair & Richter, 2019; Firth, Rivers, & Boyle, 2021). However, a dismaying conclusion from previous research is that people overwhelmingly tend to erroneously believe that interleaved learning is less beneficial than massed learning, reflecting a metacognitive illusion (Kornell & Bjork, 2008; Yan, Bjork, & Bjork, 2016), and people prefer to schedule category exemplars in a blocked rather than interleaved format during self-regulated learning, implying underemployment of the interleaved strategy (Tauber, Dunlosky, Rawson, Wahlheim, & Jacoby, 2013; Yan, Soderstrom, Seneviratna, Bjork, & Bjork, 2017).

Below, we briefly summarize empirical findings about the interleaving effect on inductive learning, then discuss metacognitive (un)awareness and practical (under)employment, and finally introduce the rationale of the current study.

Interleaving effect on inductive learning

It seems reasonable that items (or exemplars) sharing similar features should be treated as a coherent whole. Related knowledge points (or concepts) in textbooks are typically organized according to their similarities, which makes it easy for learners to recognize the similarities among exemplars. For instance, Rohrer, Dedrick, and Hartwig (2020) found that about 90% of related mathematic problems in 7th grade textbooks were arranged in a massed manner, with only about 10% presented in an interleaved way.

But is massed learning really more powerful than interleaved learning for category induction? Kornell and Bjork (2008) conducted a classic study to investigate whether interleaving is the "enemy" or "friend" of inductive learning. Participants were instructed to study 12 artists' painting styles by viewing 72 paintings, with 6 paintings from each artist. For 6 artists, their paintings were presented in a massed format (i.e., participants first studied the 6 paintings from one artist, then studied the 6 paintings from another artist, and so on), whereas for the other 6 artists, their paintings were presented in an interleaved format (i.e., participants first studied 6 paintings from the 6 artists, with one painting from each artist, then studied another 6 paintings from the 6 artists, and so on). After viewing all paintings, they undertook an induction test, in which they were shown 48 new paintings from the 12 artists (with 4 paintings from each artist), and were instructed to indicate which artist was the author of each painting. The results showed substantially better test performance for artists studied in the interleaved than in the massed condition, demonstrating an interleaving effect on inductive learning. Hence, Kornell and Bjork concluded that spacing (i.e., interleaving) is a "friend" instead of an "enemy" of inductive learning.

Note that Kornell and Bjork (2008) and many subsequent researchers termed the enhancement of interleaved learning as the *spacing effect* (e.g., Kornell, Castel, Eich, & Bjork, 2010; Kornell & Vaughn, 2018; Metcalfe & Xu, 2016), whereas others have used the term interleaving effect (e.g., Firth et al., 2021; Yan et al., 2016; Zulkiply & Burt, 2013). Spacing typically refers to a temporal window between the initial study and restudy of the same items (for review, see Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006). In contrast to spacing, interleaving is a more appropriate term to describe the sequential juxtaposition of exemplars from different categories. Indeed, a possible explanation of the interleaving effect is the *discriminative-contrast hypothesis*, which claims that interleaved presentation of exemplars facilitates the noticing of differences among categories (for detailed discussion, see Birnbaum, Kornell, Bjork, & Bjork, 2013).

The interleaving effect on inductive learning is robust, especially when the similarities among different categories are high (Brunmair & Richter, 2019), and the effect is generalizable to different types of study materials and populations (Firth et al., 2021). For instance, the effect has been documented on learning of artists' painting styles (Kang & Pashler, 2012; Metcalfe & Xu, 2016), species of butterflies, birds, and penguins (Birnbaum et al., 2013; Kornell & Vaughn, 2018), mathematical and statistical concepts (Foster, Mueller, Was, Rawson, & Dunlosky, 2019; Sana, Yan, & Kim, 2017), social science concepts (Rawson, Thomas, & Jacoby, 2015), composers' musical styles (Wong, Low, Kang, & Lim, 2020), foreign languages (Carpenter & Mueller, 2013), chest radiography scans (Rozenshtein, Pearson, Yan, Liu, & Toy, 2016), and organic chemistry compounds (Eglington & Kang, 2017).

The interleaving effect is generalizable to a variety of populations, including children (Vlach & Sandhofer, 2012), college students (Kornell & Bjork, 2008), and older adults (Kornell et al., 2010). This effect survives across a variety of study-test intervals, such as a 15-sec interval (Kornell & Bjork, 2008), a one-day interval (Taylor & Rohrer, 2010), a two-day interval (Pan, Tajran, Lovelett, Osuna, & Rickard, 2019), a one-week interval (Foster et al., 2019), and a two-week interval (Rohrer, Dedrick, & Burgess, 2014).

Overall, the interleaving effect on inductive learning is a highly robust phenomenon, generalizable to a variety of study materials, populations, and study-test intervals. The (meta)cognitive underpinnings of the interleaving effect are outside the scope of the current study, and hence are not fully discussed here. Interested readers can consult recent reviews (e.g., Firth et al., 2021; Yan, Schuetze, & Eglington, 2020).

Metacognitive unawareness and underemployment

Although, as discussed above, the benefits of interleaved learning are substantial, people tend to lack metacognitive appreciation of the interleaving effect, and they frequently choose the massed rather than interleaved strategy during self-regulated learning. For instance, in the study by Kornell and Bjork (2008),

after completing the induction test, participants were asked to report which strategy they thought helped them learn more by selecting one from three options: *A. Massed; B. About the same; C. Spaced* (interleaved). Strikingly, even though their test performance clearly demonstrated a reliable interleaving effect, about 80% of participants erroneously believed that the massed strategy was more effective, demonstrating a metacognitive illusion. This metacognitive illusion has been repeatedly documented in many subsequent studies (e.g., Birnbaum et al., 2013; Kornell et al., 2010; Tauber et al., 2013; Yan et al., 2016; Yan et al., 2017; Zulkiply, McLean, Burt, & Bath, 2012).

It is well-known that metacognitive control (e.g., study strategy selection, study time allocation) is intimately related to metacognitive monitoring (Thule, 2005; Yang, Potts, & Shanks, 2017; Yang, Sun, & Shanks, 2018). Correspondingly, a metacognitive illusion regarding the interleaving effect is likely to lead to underemployment of interleaving during self-regulated learning. Consistent with this hypothesis, in their Experiments 1 and 2, Yan et al. (2017) asked participants to indicate whether they preferred to schedule category exemplars in an interleaved or massed format in a category learning task. A majority (86% in Experiment 1 and 82% in Experiment 2) of participants chose the massed format. In the study by Tauber et al. (2013), participants were allowed to self-schedule exemplar photographs of bird families. Tauber et al. classified a participant as a "blocker" if he or she scheduled > 50% of photographs in the blocked format. Across four experiments, 78% - 100% of participants were classified as "blockers".

Overall, although interleaved learning is a more powerful study strategy, people erroneously believe that massed learning is superior, and massed learning is overwhelmingly more popular in self-regulated study (also see Kornell & Vaughn, 2018).

Rationale and overview of the current research

The interleaving effect on inductive learning, its associated metacognitive illusion and underemployment jointly highlight the urgent need to develop effective interventions to overcome the

metacognitive illusion and promote self-employment of interleaved learning. To our knowledge, thus far only one study has investigated potential interventions to correct the metacognitive error (Yan et al., 2016). In this study, Yan et al. found that (1) allowing participants to experience the interleaving effect at first-hand, and (2) informing them about the reasons why interleaved learning is more effective than massed learning, are insufficient to fully change people's erroneous belief that massed learning is better. Hence, Yan et al. concluded that it is difficult to overcome metacognitive misalignment about the interleaving effect.

Yan et al. also proposed two potential reasons why the illusion may persist. The first is that blocked encoding of exemplars is accompanied with greater ease of processing (i.e., processing fluency) than spaced encoding, and numerous studies have shown that ease of processing is positively related to judgments of learning (JOLs) (that is, people may use the heuristic that "easily learned means easily remembered", see Ball, Klein, & Brewer, 2014; Yang, Huang, & Shanks, 2018; Yang, Yu, et al., 2021). Instructional interventions may be inadequate to override this experience of processing fluency. The second mechanism concerns the *a priori* beliefs that people might have. Specifically, people might have pre-existing beliefs that massed learning is superior to interleaved learning (see Yan et al., 2016, for detailed discussion of the underlying mechanisms).

Going beyond Yan et al. (2016), the current study investigates whether explicit instructions about the interleaving effect can calibrate metacognition. The instructions in our intervention deliver individual feedback (that is, feedback about participants' own test performance, which clearly demonstrates the interleaving effect) and include a description of the interleaving effect results from Kornell and Bjork (2008). Our hypothesis is that these direct and explicit instructions, which contain counterintuitive information, will be sufficient to overcome people's natural tendency to believe that massing is superior to interleaving.

Previous studies did show that providing explicit instructions is a viable method to correct another metacognitive illusion, namely relating to the *testing effect* (Ariel & Karpicke, 2018; Hui, de Bruin, Donkers, & van Merriënboer, 2021). The testing effect refers to the phenomenon that retrieval practice consolidates long-term retention and facilitates learning efficiency more effectively than restudying (for reviews, see Roediger & Karpicke, 2006; Yang, Luo, Vadillo, Yu, & Shanks, 2021; Yang, Potts, & Shanks, 2018). Similar to the interleaving effect, people tend to erroneously believe that restudying is more beneficial than practice testing (for review, see Rivers, 2021). Recently, Hui et al. (2021) found that providing explicit instructions, which contain individual feedback (i.e., feedback about their own test performance for restudied and tested items in a previous learning task) and the testing effect results from previous studies, is sufficient to improve metacognitive appreciation of the testing effect and promote self-use of practice testing during a subsequent learning task (for related findings, see Ariel & Karpicke, 2018).

Overall, the first aim of the current study is to explore whether an instruction intervention can override the typical metacognitive illusion about the spacing effect. To foreshadow, our Experiment 1 showed that the instruction intervention successfully improved participants' awareness about the benefits of interleaving. Then, another important question comes to the fore: Does enhanced metacognitive awareness lead to greater use of the interleaved strategy when learning new categories? To test this question, we instructed participants to make item-by-item choices about whether they preferred to study new categories via the massed or interleaved strategy.

To foreshadow, Experiment 1 found that the intervention group chose to study more categories using the interleaved strategy than the control group, which received placebo instructions unrelated to the interleaving effect. Experiment 1 next addressed a third important question: Does enhanced use of interleaving produce superior mastery of new categories? To answer this question, we allowed both the

intervention and control groups to study new categories in the formats that honored their strategy choices, and then assessed their test performance on these categories.

The perceived value of an intervention is likely to be heavily determined by its long-lasting effectiveness and success in transferring to new materials (McDaniel & Einstein, 2020). If the effect of a given intervention is short-lived and not transferable, its practical value might be perceived as weak. Accordingly, the last two aims of the current study are to explore whether the instruction intervention effect is long-lasting and transferable. A long interval (24 hours) was inserted between the intervention phase and the subsequent learning task in Experiment 2. In Experiments 3 and 4, participants were instructed to study another type of materials (i.e., butterfly species) during a subsequent learning task, which was different from the materials in the prior learning task (i.e., artists' painting styles).

In summary, the current study targets to address five important questions: (1) Whether the instruction intervention is able to override the standard metacognitive illusion regarding the interleaving effect, (2) whether enhanced metacognitive appreciation promotes self-employment of the interleaved strategy when learning new categories, (3) whether enhanced use of interleaving boosts mastery of new categories, (4) whether the intervention effect is long-lasting, and finally (5) whether the intervention effect is transferable to learning of categories in different domains.

Experiment 1

Experiment 1 was designed to explore the effects of the instruction intervention on metacognitive awareness of the interleaving effect, self-employment of the interleaved strategy (i.e., strategy choices), and learning outcomes (i.e., whether the instruction intervention facilitates learning of new categories).

Method

Participants

A pilot study was conducted to determine the required sample size, which observed Cohen's *d* = 0.93 for the intervention effect on strategy choices. A power analysis, conducted via G*power (Faul, Erdfelder, Lang, & Buchner, 2007), revealed that 26 participants in each group were required to observe a significant (2-tailed, α = .05) intervention effect at 0.90 power. To be more conservative, we decided to increase the sample size to 30 in each group, a widely-used sample size in Psychology research.

Accordingly, 60 participants (*M* age = 22.15, *SD* = 2.18; 83.3% female) were recruited from XXX University (institution information is masked), and randomly allocated into the control and intervention groups, with 30 participants in each group. They were native Chinese speakers, were not Psychology major students, had normal or corrected-to-normal vision, signed informed consent, were tested individually in a sound-proofed cubicle, and received monetary compensation. The study was approved by the Ethics Committee of the Faculty of Psychology, XXX (institution information is masked).

Materials

The experiment consisted of two learning tasks. In the first task, 10 paintings from each of 12 relatively unfamiliar artists (Georges Braque, Henri-Edmond Cross, Judy Hawkins, Philip Juras, Ryan Lewis, Marilyn Mylrea, Bruno Pessani, Ron Schlorff, Georges Seurat, Ciprian Stratulat, George Wexler, YieMei) were taken from Kornell and Bjork (2008). The 12 artists were divided into two sets, and learning difficulty between these two sets was matched according to Yan et al. (2016). For each participant, set assignment to conditions (massed *vs*. interleaved) was randomly determined by the computer.

For each artist, six paintings were used for study, with the remaining four presented in the induction test. To avoid potential confounding of prior knowledge that participants might have about these artists, we selected 12 popular Chinese names, extracted from the website named *The most commonly used boy*

names (available at https://jingyan.baidu.com/article/e6c8503c149adce54f1a1899.html), to replace the original artists' names. Thus, participants learned fake (rather than actual) artists' names.

In the second learning task, 10 paintings from each of 12 further artists (Carla Bosch, David Grossmann, Wassily Kandinsky, Peder Mork Monsted, Grandma Moses, Roger Mühl, Georgia O'Keeffe, Pierre-Auguste Renoir, Henri Rousseau, Egon Schiele, Maurice Utrillo, Guim Tio Zarraluki) were selected following Kornell and Bjork's (2008) selection criteria (e.g., all paintings are landscapes or skyscapes). For each artist, 6 paintings were used in the study phase, with the remaining 4 presented in the induction test. These artists' names were replaced by 12 further Chinese names, selected from the same website. All paintings were cropped to remove identifying characteristics (e.g., signatures) and then resized to 900 × 750 pixels. All stimuli were presented via Matlab *Psychtoolbox*.

The paintings and Chinese names are publicly available at OSF (web-link is masked).

Experimental design and procedure

Experiment 1 involved a between-subjects design (group: control *vs*. intervention). The procedure is schematically depicted in Figure 1A.

In the first learning task, participants were instructed to study 12 artists' painting styles in preparation for a later test. The procedure was adapted from Kornell and Bjork (2008, Experiment 1b). Specifically, six artists' paintings were studied in a massed format, with the other six artists' paintings studied in an interleaved format. In the massed condition, the 36 paintings were presented one-by-one and artist-by-artist (i.e., A1, A2, …, A6; B1, B2, …, B6; …; F1, F2, …, F6), where the letters refer to artists and the numbers to that artist's paintings. Thus the 36 paintings were divided into six massed blocks, with each block comprising six paintings from the same artist. For each participant, the sequence of the six massed blocks and the sequence of paintings in each block were randomly determined by the computer. Each painting was presented on screen for 5 sec, with the corresponding artist's (Chinese) name presented

below it. Following the presentation of each painting, a crosshair was presented for 0.5 sec to mark the interstimulus interval.

The procedure in the interleaved condition was the same as that in the massed condition, except that the six artists' paintings were divided across six interleaved blocks, with each block comprising one painting from each artist (i.e., G1, H1, …, L1; G2, H2, …, L2; …; G6, H6, …, L6). Hence, participants first studied one painting from each of the six artists, then another six paintings from the six artists, and so on.

Overall, there were six massed and six interleaved blocks of paintings in the study phase. Following Kornell and Bjork (2008), the order of the 12 blocks was MIIMMIIMMIIM (I represents an interleaved block and M refers to a massed block). After viewing all 72 paintings, participants engaged in a 15-sec distractor task, in which they subtracted 3 from 537 in succession. Immediately after that, participants completed an induction test. In the test, 48 new paintings, with 4 paintings from each artist, were presented one-by-one in a random order. Below each painting, the 12 artists' (Chinese) names were presented as 12 response options. Participants were asked to decide which artist was associated with the on-screen painting by using the mouse to click one of the names. There was no time pressure and no feedback during the test.

After the induction test, participants were asked: *Which strategy do you think is more effective? A. Massed; B. About the same; C. Spaced*. This metacognitive judgment question measured participants' awareness of the spacing effect. Note that, following Kornell and Bjork (2008), we framed "interleaved learning" as "spaced learning" in the instructions and metacognitive questions. We followed Kornell and Bjork's (2008) procedure and terminology as closely as we could to enhance the likelihood of replicating the interleaving effect and its associated metacognitive illusion. It should be highlighted that (1) none of the participants in the current study were Psychology major students (which means that they had little

prior knowledge about interleaved or spaced learning), and (2) the metacognitive judgments were made after the first learning task (which means that all participants knew what the strategy names referred to, regardless of how the strategies were labelled).

Next, the intervention phase began. For the intervention group, participants received instructions about the interleaving effect. There were two histogram graphs presented on screen, with one depicting the participant's own test performance and the other showing the results from Kornell and Bjork (2008, Experiment 1b) (see Figure 1B). The on-screen instructions were adapted according to the participant's actual test performance.

For participants who correctly classified more paintings in the interleaved than in the massed condition, the instructions were as follows: *As shown in the left figure, you correctly classified X/24 (X%) paintings for the artists studied through the massed strategy, and X/24 (X%) paintings for the artists studied through the spaced strategy. These results suggest that the spaced strategy is more effective for you than the massed strategy. Indeed, in a classic study which employed the same paintings and experimental procedure, Kornell and Bjork (2008) found that participants correctly classified more paintings for the artists studied through the spaced strategy than they did for the artists studied through the massed strategy (see the right figure). Overall, the spaced strategy is more effective than the massed strategy.*

For participants who showed superior test performance in the massed than in the interleaved condition, the instructions were as follows: *As shown in the left figure, you correctly classified X/24 (X%) paintings for the artists studied through the massed strategy, and X/24 (X%) paintings for the artists studied through the spaced strategy. These results suggest that the massed strategy seems to be more effective for you than the spaced strategy. However, in a classic study which employed the same paintings and experimental procedure, Kornell and Bjork (2008) found that participants correctly classified more*

paintings for the artists studied through the spaced strategy than they did for the artists studied through the massed strategy (see the right figure). It is possible that the fact that your results are inconsistent with the findings from other prior studies might be due to random noise (or random error). Overall, many previous studies have consistently demonstrated that the spaced strategy is more effective than the massed strategy.

For participants who correctly classified an equal number of paintings in the massed and interleaved conditions, the instructions were as follows: *As shown in the left figure, you correctly classified X/24 (X%) paintings for the artists studied through the massed strategy, and X/24 (X%) paintings for the artists studied through the spaced strategy. These results suggest that the massed and spaced strategies seem to be equally effective for you. However, in a classic study which employed the same paintings and experimental procedure, Kornell and Bjork (2008) found that participants correctly classified more paintings for the artists studied through the spaced strategy than they did for the artists studied through the massed strategy (see the right figure). It is possible that the fact that your results are inconsistent with the findings from other prior studies might be due to random noise (or random error). Overall, many previous studies have consistently demonstrated that the spaced strategy is more effective than the massed strategy.*

In contrast to the intervention group, the control group received a placebo intervention involving instructions about the testing effect (unrelated to the interleaving effect). A histogram graph was presented, depicting the testing effect results from Roediger and Karpicke (2006, Experiment 1) (see Figure 1C). The instructions were as follows: *Over the last century, numerous studies have consistently demonstrated that testing is a more effective strategy to enhance learning than restudying. For instance, Roediger and Karpicke (2006) asked participants to study two texts, with one text studied once and tested once and the other text repeatedly studied twice. In a test administered one week later, participants were*

asked to recall as much information as they could from the two texts. As shown in the below figure, test performance for the tested text was much better than that for the restudied text, implying that testing is a more effective strategy than restudying. We chose to describe the testing effect in the placebo instructions with the aim of making these instructions roughly comparable with the intervention instructions (that is, instructions in both groups described a well-established learning effect). To foreshadow, the placebo instructions had no statistically detectable influence on participants' appreciation of the interleaving effect.

Following the intervention phase, both groups were re-asked the question: *Which learning strategy do you think is more effective? A. Massed; B. About the same; C. Spaced*. This question measured whether the intervention successfully updated participants' metacognitive awareness.

Next, participants initiated the second learning task, in which they were instructed to study 12 new artists' painting styles. Before the study phase, the 12 artists' (Chinese) names were presented one-by-one in a random order, with two options (i.e., *A. Massed; B. Spaced*) presented below the name. Participants were asked to make a strategy choice for each artist.

During the study phase, the 72 paintings (6 paintings from each artist) were presented in the format that honored the participant's selections. For instance, if a participant chose to study nine artists' paintings using the massed strategy, with the other 3 artists' paintings studied via the spaced (interleaved) strategy, the computer would first randomly decide whether to present the massed or interleaved paintings first. If it decided to present the massed paintings first, the 54 (6×9) paintings from the nine relevant artists were presented one-by-one, for 5 sec each, in a blocked schedule (i.e., A1, A2, …, A6; B1, B2, …, B6; …; I1, I2, …, I6). The presentation order of each artist's paintings in each block and the block order were randomly determined by the computer.

Next, the remaining three artists' 18 paintings were presented one-by-one, for 5 sec each, in an interleaved schedule (i.e., J1, K1, L1; J2, K2, L2; …; J6, K6, L6). The presentation order of each artist's paintings and the artist order were randomly determined by the computer.[1] The interleaved items were presented prior to the massed ones for participants for whom the randomization was the other way round.

After the study phase, participants engaged in the same 15-sec distractor task as before, and then undertook an induction test. In the test, 48 new paintings, comprising 4 paintings from each artist, were presented one-by-one in a random order, and participants were asked to report which was the author of each painting.

After this criterial induction test, participants completed two sets of questionnaires to measure their intelligence mindset (Dweck & Yeager, 2020) and effort beliefs (Blackwell, 2002). These questionnaires were included for exploratory purposes and the results are not of substantive research interest. It is also noteworthy that there is little reason to suspect that the inclusion of these questionnaires might affect the main results, because they were administered at the end of the experiment. We hence do not discuss the questionnaire results in the current article.

Results

Test performance in the first learning task

A mixed analysis of variance (ANOVA) found a main effect of study strategy, $F(1,58) = 23.32$, $p < .001$, $\eta_p^2 = .29$ (see Figure 2A). There was no main effect of group, $F(1,58) = 0.21$, $p = .65$, $\eta_p^2 = .004$, and no statistically detectable interaction, $F(1,58) = 0.27$, $p = .61$, $\eta_p^2 = .005$. Pre-planned paired *t*-tests showed that participants performed better in the interleaved than in the massed condition in both groups (see Table 1), replicating the classic interleaving effect on inductive learning.

[1] A noteworthy point is that if a given participant only selected one artist to be studied using the interleaved strategy, all 12 artists' paintings would be presented in the massed format because there were no other artists' paintings to be inserted for the interleaved artist.

Awareness before intervention

Figure 2B shows participants' awareness of the interleaving effect in the control and intervention groups, respectively, measured before intervention. In both groups, a majority of participants believed that massed learning is more effective than interleaved learning (M > I), with a minority believing that interleaved learning is better than massed learning (M < I) or that the effectiveness of these two strategies is about the same (M = I). A Chi-square test showed no statistically detectable difference in metacognitive awareness between the two groups, $\chi^2(2) = 1.29$, $p = .52$. Hence, participants' judgments were collapsed across groups to increase statistical power.

Across the 60 participants, 75.0% believed M > I, greater than the proportion (18.3%) believing M < I, $\chi^2(1) = 36.46$, $p < .001$, and greater than the proportion (6.7%) believing M = I, $\chi^2(1) = 55.19$, $p < .001$. There was no statistically detectable difference between the proportions believing M < I and M = I, $\chi^2(1) = 2.74$, $p = .10$. These results successfully replicate the standard metacognitive bias regarding the interleaving effect.

Awareness after intervention

Figure 2B shows metacognitive awareness measured after intervention. A Chi-square test found a significant difference in awareness between groups, $\chi^2(2) = 17.20$, $p < .001$. Hence, below we report results for the two groups separately.

For the control group, 73.3% (22 out of 30) of participants believed M > I, greater than the proportion (23.3%) believing M < I, $\chi^2(1) = 13.80$, $p < .001$, and the proportion (3.3%) believing M = I, $\chi^2(1) = 28.20$, $p < .001$. The proportion believing M < I tended to be greater than the proportion believing M = I, $\chi^2(1) = 3.61$, $p = .056$. These results imply that the placebo intervention was ineffective, and a majority of participants (approximately the same proportion as in the first measurement) still believed that massed learning is better.

By contrast, for the intervention group, a majority (66.7%) of participants now reported M < I, greater than the proportion (20.0%) believing M > I, $\chi^2(1) = 11.47$, $p < .001$, and the proportion (13.3%) believing M = I, $\chi^2(1) = 15.63$, $p < .001$. There was little difference between the proportions believing M > I and M = I, $\chi^2(1) = 0.12$, $p = .73$. These results imply that the instruction intervention was successful, and a majority of participants in the intervention group came to believe that interleaved learning is better.

For each group, further tests were performed to compare participants' metacognitive awareness before and after the intervention. The corresponding results are reported in Table 2. For the intervention group, there was a significant difference between awareness measured before and after the intervention, $\chi^2(2) = 16.02$, $p < .001$. By contrast, there was no statistically detectable difference in the control group, $\chi^2(2) = 0.42$, $p = .81$. These results again suggest that the instruction intervention, but not the placebo intervention, was effective in improving awareness of the interleaving effect.[2]

Overall, the above results indicate that the instruction intervention successfully enhanced participants' appreciation of the interleaving benefit, whereas the placebo intervention failed to alter participants' erroneous beliefs.

Strategy choices

As shown in Figure 2C, the intervention group chose to study a substantially greater (almost doubled) proportion (M = 61.7%, SD = 28.5%) of artists via the interleaved strategy than the control group (M = 31.1%, SD = 35.1%), difference = 30.6% [14.0%, 47.1%], $t(58) = 3.70$, $p < .001$, $d = 0.96$, implying that the instruction intervention can be applied as a practical strategy to promote self-employment of the interleaved strategy.

[2] As shown in Table 2, the exact same patterns were observed in Experiments 2-4. For the sake of brevity, we do not discuss these results further.

Test performance in the second learning task

In the second induction test, performance in the intervention group (M = 56.0%, SD = 20.6%) was only numerically (but not significantly) greater than that in the control group (M = 51.7%, SD = 21.8%), difference = 4.3% [-6.7%, 15.3%], $t(58) = 0.79$, $p = .44$, $d = 0.20$ (see Figure 2D), implying that the intervention effect on learning outcomes was modest.

It is reasonable to assume that the non-significant intervention effect on test performance might be a false negative, resulting from low statistical power. In addition, to foreshadow, Experiments 2-4 consistently found numerical (non-significant) intervention effects on the equivalent tests. Hence, after Experiment 4, we report a multilevel regression analysis to estimate the magnitude of this aspect of the intervention, in which the results from all four experiments were combined to increase statistical power.

Further tests were performed to investigate whether test performance on artists studied via the interleaved strategy was better than that on artists studied via the ~~interleaved~~ massed strategy. The answer was affirmative. Because these results are not central to our research interest, they are reported in the Supplemental Information (SI) (the SI file has been submitted via the journal review system to make it available to reviewers).

After reporting Experiment 4, we discuss individual differences in the intervention effect, in which participants in each group were split into two sub-groups according to their test performance in the first learning task (see below for details).

Discussion

Experiment 1 observed that the instruction intervention was successful in (1) correcting participants' metacognitive illusion regarding interleaving, and (2) promoting self-employment of the interleaved strategy, but (3) the intervention effect on learning outcomes was minimal. Also, the results demonstrate that direct experience of the contrast between massing and interleaving is not in itself sufficient to

promote adoption of the interleaved strategy: participants had this direct experience of the interleaving effect induced by the induction test, but a majority reported in the pre-intervention assessment that massing is superior to interleaving (Yan et al., 2016). It was only when this experience was combined with explicit information about the interleaving effect (i.e., individual feedback plus the interleaving effect results from previous research) that it translated into a change in preference.

Experiment 2

Experiment 1 established the short-term success of the instruction intervention in modifying self-employment of the interleaved strategy. Participants made their strategy choices immediately following the intervention phase. Experiment 2 asked whether the efficacy of the intervention is long-lasting: a 24-hour interval was inserted between the intervention phase and the second learning task.

Method

Participants

Following Experiment 1, 60 participants (*M* age = 21.78, *SD* = 1.82; 91.7% female) were recruited from XXX (institution information is masked), and randomly allocated into the control and intervention groups, with 30 participants in each group. All participants were native Chinese speakers, were not Psychology major students, had normal or corrected-to-normal vision, signed informed consent, were tested individually in a sound-proofed cubicle, and received monetary compensation.

Materials, experimental design, and procedure

The stimuli, experimental design, and procedure were identical to those in Experiment 1, but with one exception. Specifically, following the administration of the intervention and the second awareness assessment, participants were dismissed and invited to come back 24 h later. One day later, they returned to the same laboratory cubicle to complete the second learning task. As before, in the second learning

task, participants first made a strategy decision for each artist, then studied the paintings according to their decisions, and finally took the induction test.

Results

Test performance in the first learning task

A mixed ANOVA showed a main effect of study strategy, $F(1,58) = 15.60$, $p < .001$, $\eta_p^2 = .21$, but no main effect of group, $F(1,58) < 0.001$, $p = .98$, $\eta_p^2 < .001$, nor interaction, $F(1,58) = 0.31$, $p = .58$, $\eta_p^2 = .005$ (see Figure 3A). In both groups, participants performed better in the interleaved than in the massed condition (see Table 1).

Awareness before intervention

Figure 3B shows participants' awareness measured before intervention. A Chi-square test found minimal difference in awareness between groups, $\chi^2(2) = 0.87$, $p = .65$. Hence, judgments were collapsed. Across all 60 participants, 75.0% believed M > I, greater than the proportion (20.0%) believing M < I, $\chi^2(1) = 34.20$, $p < .001$, and the proportion (5.0%) believing M = I, $\chi^2(1) = 58.37$, $p < .001$. The proportion believing M < I was greater than the proportion believing M = I, $\chi^2(1) = 4.88$, $p = .03$. Overall, these results again replicate the well-known metacognitive bias regarding the interleaving effect.

Awareness after intervention

Figure 3B shows participants' awareness measured after intervention. A Chi-square test found a significant difference between groups, $\chi^2(2) = 13.81$, $p = .001$. For the control group, 63.3% of participants believed M > I, greater than the proportion (26.7%) believing M < I, $\chi^2(1) = 6.73$, $p = .009$, and the proportion (3.3%) believing M = I, $\chi^2(1) = 16.15$, $p < .001$. There was little difference between the proportions believing M < I and M = I, $\chi^2(1) = 1.78$, $p = .18$. These results imply that a majority of participants in the control group still believed that massed learning is better than interleaved learning.

By contrast, for the intervention group, the majority (66.7%) now believed M < I, greater than the proportion (16.7%) believing M > I, $\chi^2(1) = 13.44$, $p < .001$, and the proportion (16.7%) believing M = I, $\chi^2(1) = 13.44$, $p < .001$. These results confirm that the instruction intervention successfully improved participants' awareness of the interleaving effect (also see Table 2).

Strategy choices

The intervention group chose to study a substantially greater proportion (M = 66.7%, SD = 27.4%) of artists using the interleaved strategy than the control group (M = 38.9%, SD = 38.7%), difference = 27.8% [10.5%, 45.1%], $t(58) = 3.21$, $p = .002$, $d = 0.83$ (see Figure 3C), demonstrating the effectiveness of the instruction intervention to promote self-employment of interleaving after a 24 h delay.

Test performance in the second learning task

In the second induction test, performance in the intervention group (M = 50.6%, SD = 18.5%) was only numerically greater than in the control group (M = 47.8%, SD = 15.5%), difference = 2.8% [-6.1%, 11.6%], $t(58) = 0.63$, $p = .53$, $d = 0.16$ (see Figure 3D), implying that the intervention effect on learning outcomes was small at best.

Discussion

Experiment 2 showed that the intervention effect on strategy choices was long-lasting (at least 24 h). Like Experiment 1, Experiment 2 observed that the intervention effect on learning outcomes was modest.

Experiment 3

Both Experiments 1 and 2 observed an intervention effect on self-employment of the interleaved strategy during learning of new artists' painting styles. Experiment 3 investigated the transferability of the intervention effect. Specifically, it asked whether the instruction intervention – administered in the context of learning artists' painting styles – can promote self-employment of the interleaved strategy during subsequent learning of categories in a different domain (e.g., natural butterfly species).

Method

Participants

Sixty participants (*M* age = 20.95, *SD* = 2.05; 93.9% female) were recruited from XXX (institution information is masked), and randomly allocated to the control and intervention groups, with 30 participants in each group. All participants were native Chinese speakers, were not Psychology major students, had normal or corrected-to-normal vision, signed the informed consent, were tested individually in a sound-proofed cubicle, and received monetary compensation.

Materials, experimental design, and procedure

The stimuli in the first learning task were identical to those in Experiment 1. The stimuli in the second learning task were 80 photographs of 16 species of butterflies, with 5 photographs for each species, taken from Birnbaum et al. (2013). All photographs were resized to 900 × 750 pixels. The resized photographs are publicly available at OSF (web-link is masked). The names of the 16 species were Admiral, American, Baltimore, Cooper, Eastern Tiger, Hairstreak, Harvester, Mark, Painted Lady, Pine Elfin, Pipevine, Sprite, Tipper, Tree Satyr, Viceroy, and Wood Nymph. In the second learning task, 64 photographs, with 4 photographs for each species, were used for study, and the remaining 16 photographs were used in the induction test.

The experimental design and procedure were identical to those in Experiment 1, but with several minor changes in the second learning task. Specifically, participants were asked to study 16 species of butterflies, the butterfly names were presented one-by-one in a random order, and participants decided which strategy they wanted to use to study each species' photographs. During the study phase, 64 photographs were presented one-by-one, for 5 sec each, in a manner that honored the participant's decisions. Next, participants engaged in the same 15-sec distractor task. Finally, they completed an induction test. In the test, 16 new photographs (comprising one photograph from each species) were

presented one-by-one in a random order, with the 16 species' names presented below each photograph. Participants were asked to report which species the on-screen butterfly belonged to.

Results

Test performance in the first learning task

A mixed ANOVA showed a main effect of study strategy, $F(1,58) = 24.47$, $p < .001$, $\eta_p^2 = .30$, but no main effect of group, $F(1,58) = 0.05$, $p = .83$, $\eta_p^2 = .001$, nor interaction, $F(1,58) = 0.003$, $p = .96$, $\eta_p^2 < .001$ (see Figure 4A). In both groups, participants performed better in the interleaved than in the massed condition (see Table 1).

Awareness before intervention

Figure 4B shows participants' awareness measured before intervention. A Chi-square test found little difference in awareness between groups, $\chi^2(2) = 2.10$, $p = .35$. Hence, judgments were collapsed. Across all 60 participants, 75.0% believed M > I, greater than the proportion (21.7%) believing M < I, $\chi^2(1) = 32.07$, $p < .001$, and the proportion (3.3%) believing M = I, $\chi^2(1) = 61.70$, $p < .001$. The proportion believing M < I was greater than that believing M = I, $\chi^2(1) = 5.84$, $p = .02$. Overall, before intervention, a majority of participants erroneously believed that the massed strategy is superior.

Awareness after intervention

Figure 4B shows participants' metacognitive awareness measured after intervention. A Chi-square test found a significant difference in awareness between groups, $\chi^2(2) = 15.59$, $p < .001$. For the control group, 73.3% of participants believed M > I, greater than the proportion (26.7%) believing M < I, $\chi^2(1) = 11.27$, $p < .001$. No one believed M = I. By contrast, for the intervention group, a majority (70.0%) came to believe M < I, greater than the proportion (23.3%) believing M > I, $\chi^2(1) = 11.32$, $p < .001$, and the proportion (6.7%) believing M = I, $\chi^2(1) = 22.84$, $p < .001$. There was little difference between the proportions believing M > I and M = I, $\chi^2(1) = 2.09$, $p = .15$. These results confirm again that the

instruction intervention, but not the placebo intervention, was successful in enhancing awareness about the interleaving effect (also see Table 2).

Strategy choices

The intervention group chose to study a substantially greater proportion (M = 67.7%, SD = 18.2%) of butterfly species via interleaving than the control group (M = 29.8%, SD = 18.2%), difference = 37.9% [25.6%, 50.2%], $t(58) = 6.16$, $p < .001$, $d = 1.59$ (see Figure 4C), establishing the effectiveness of the instruction intervention to promote self-employment of the interleaved strategy in a transfer domain.

Test performance in the second learning task

In the second induction test, test performance for the intervention group (M = 40.6%, SD = 19.5%) was only numerically greater than that for the control group (M = 35.4%, SD = 17.2%), difference = 5.2% [-4.3%, 14.7%], $t(58) = 1.10$, $p = .28$, $d = 0.28$ (see Figure 4D), again implying that the intervention effect on learning outcomes was small.

Discussion

Experiment 3 found that the intervention effect on study strategy regulation transferred to learning of categories in a different domain.

Experiment 4

Experiment 4 investigated the long-lasting transferability of the intervention effect. To achieve this aim, it replicated Experiment 3 but with a 24 h interval inserted between the intervention phase and the second learning task.

Method

Participants

Sixty participants (M age = 20.77, SD = 2.13; 93.3% female) were recruited from XXX (institution information is masked), and were randomly allocated into the control and intervention groups, with 30

participants in each group. All participants were native Chinese speakers, were not Psychology major students, had normal or corrected-to-normal vision, signed the informed consent, were tested individually in a sound-proofed cubicle, and received monetary compensation.

Materials, experimental design, and procedure

The stimuli, design, and procedure were identical to those in Experiment 3, but with one exception. Specifically, after the intervention phase and the second awareness assessment, participants were dismissed and invited to come back 24 h later. One day later, they returned to the same laboratory cubicle to complete the second learning task.

Results

Test performance in the first learning task

A mixed ANOVA found a main effect of study strategy, $F(1,58) = 29.55$, $p < .001$, $\eta_p^2 = .34$, but no main effect of group, $F(1,58) = 0.12$, $p = .73$, $\eta_p^2 = .002$, nor interaction, $F(1,58) = 0.70$, $p = .41$, $\eta_p^2 = .01$ (see Figure 5A). Both groups performed better in the interleaved than in the massed condition (see Table 1).

Awareness before intervention

Figure 5B shows participants' awareness measured before intervention. Note that because there was no difference in metacognitive awareness between groups, judgments were collapsed across groups. Across all 60 participants, 83.3% believed M > I, greater than the proportion (13.3%) believing M < I, $\chi^2(1) = 50.10$, $p < .001$, and the proportion (3.3%) believing M = I, $\chi^2(1) = 74.17$, $p < .001$. The proportions believing M < I and M = I were not statistically different, $\chi^2(1) = 2.73$, $p = .10$. The results again replicate the illusory belief that participants hold about interleaving.

Awareness after intervention

Figure 5B shows participants' awareness measured after intervention. A Chi-square test found a significant difference in awareness between groups, $\chi^2(2) = 17.15$, $p < .001$. In the control group, 80.0% (24 out of 30) of participants believed M > I, greater than the proportion (16.7%) believing M < I, $\chi^2(1) = 21.62$, $p < .001$, and the proportion (3.3%) believing M = I, $\chi^2(1) = 33.19$, $p < .001$. There was little difference between the proportions believing M < I and M = I, $\chi^2(1) = 1.67$, $p = .20$. By contrast, in the intervention group, a majority (60.0%) now reported M < I, greater than the proportion (26.7%) believing M > I, $\chi^2(1) = 5.50$, $p = .02$, and the proportion (13.3%) believing M = I, $\chi^2(1) = 12.13$, $p < .001$. There was little difference between the proportions believing M > I and M = I, $\chi^2(1) = 0.94$, $p = .33$. Once more the instruction intervention, but not the placebo intervention, was successful in driving participants to appreciate the interleaving effect (also see Table 2).

Strategy choices

The intervention group chose to study a substantially greater proportion (M = 64.6%, SD = 22.3%) of butterfly species via interleaving than the control group (M = 27.9%, SD = 24.9%), difference = 36.7% [24.4%, 48.9%], $t(58) = 6.01$, $p < .001$, $d = 1.55$ (see Figure 5C).

Test performance in the second learning task

In the second induction test, performance in the intervention group (M = 44.0%, SD = 15.9%) was only numerically greater than in the control group (M = 37.7%, SD = 16.2%), difference = 6.3% [-2.0%, 14.5%], $t(58) = 1.51$, $p = .14$, $d = 0.39$ (see Figure 5D), implying that the intervention effect on learning outcomes was modest.

Discussion

Extending the results of Experiment 3, Experiment 4 showed that the transferability of the intervention effect on study strategy regulation is long-lasting (at least 24 hours).

Multilevel Analyses

Experiments 1-4 consistently found that the instruction intervention successfully changed participants' beliefs and promoted self-adoption of interleaving. Although all four experiments observed that the instruction intervention numerically improved subsequent test performance, none of them detected significant evidence. We therefore conducted three multilevel regression analyses, via the R *lme4* package, to integrate results across experiments to increase statistical power. Besides these regression analyses, a multilevel mediation analysis was conducted, via the R *mediation* package, to determine whether the intervention effect on final test performance was mediated by its effect on strategy choices. Note that, in all multilevel analyses reported below, participants were treated as the first level variable, with experiments as the second level variable.

Results

The first multilevel analysis, which regressed test performance in the second learning task onto group (intervention = 0.5; control = -0.5), found a significant intervention effect on learning outcomes, b = 4.6% [0.1%, 9.2%], p = .048, indicating that the instruction intervention improved test performance by about 4.6%.

The second analysis, which regressed strategy choices onto group, found a significant intervention effect on self-employment of interleaving, b = 33.2% [22.1%, 40.4%], p < .001, implying that the instruction intervention enhanced the likelihood of selection of the interleaved strategy by about 33%.

In the third analysis, test performance in the second learning task was regressed onto strategy choices. This revealed a significant relation between these two variables, b = 0.18% [0.12%, 0.25%], p < .001. Every increase of 10% in choice of interleaving increased test performance by about 1.8%.

Finally, a multilevel mediation analysis was conducted (see Figure 6). The results re-confirmed that the total intervention effect on test performance was significant, 4.6% [0.05%, 9.23%], p = .046. Critically, the indirect effect through strategy choices was significant, 6.6% [3.84%, 9.63%], p < .001.

Moreover, the direct effect was minimal, -1.9% [-7.01%, 3.01%], p = .438. Overall, these results imply that the intervention effect on learning outcomes was completely mediated by its effect on self-employment of the interleaved strategy.

Discussion

The instruction intervention developed and tested here was successful in enhancing metacognitive awareness of the interleaving effect, promoting self-employment of the interleaved strategy, and facilitating inductive learning. Importantly, the intervention effect on test performance was fully mediated by its effect on self-employment of the interleaved strategy.

Individual Differences Analyses

To explore individual differences in the intervention effect, participants in each of Experiment 1-4's intervention and control groups were divided into two sub-groups according to their test performance in the first learning task: (1) an interleaving effect set (comprising participants whose test performance demonstrated an interleaving effect in the first learning task) and (2) a no interleaving effect set (consisting of participants whose test performance did not exhibit an interleaving effect in the first learning task).

The detailed results are reported in the SI. Overall, the results show that, regardless of whether participants demonstrated an interleaving effect or not in the first learning task, the instruction intervention always significantly improved their awareness of the interleaving effect, significantly enhanced their choices of the interleaved strategy, and numerically improved their accuracy in the second induction test.

General Discussion

The interleaving effect on inductive learning is a robust phenomenon, generalizable to different types of materials, populations, and study-test intervals (see the Introduction). Unfortunately, learners

overwhelmingly erroneously believe that massed learning is superior (Kornell & Bjork, 2008; Yan et al., 2016), a metacognitive illusion that leads to underemployment of interleaving during self-regulated learning (Tauber et al., 2013; Yan et al., 2017). These findings were successfully replicated in the current study. For instance, all four experiments observed superior test performance in the interleaved than in the massed condition in the first learning task; across Experiments 1-4, 75.5% - 83.3% (similar to the proportions reported by Kornell & Bjork, 2008) of participants suffered from the illusion that massed learning is more effective; and in Experiments 1-4, the control group only chose to study 27.9% - 38.9% of artists or butterfly species via the interleaved strategy.

These findings highlight the need to develop effective interventions to override people's natural bias towards massed learning and to promote self-employment of the interleaved strategy. In the only previous effort in this direction, Yan et al. (2016) investigated whether two kinds of interventions (i.e., personal experience of the interleaving effect and informing learners why interleaved learning is more beneficial) are able to correct the massing-is-better-than-interleaving illusion, and their answer was largely negative. At least one aspect of Yan et al.'s findings was conceptually replicated here. Specifically, even after the induction test in the first learning task (in which participants' test performance did clearly show the interleaving effect), there were still 75.5% - 83.3% of participants holding the erroneous belief that massed learning is superior. Thus, direct experience of the benefits of interleaving is insufficient to overcome the massing bias.

Different from Yan et al. (2016), the current research explored whether explicit (counterintuitive) instructions, which refer to individual feedback and the interleaving effect results from Kornell and Bjork (2008), can reverse this metacognitive illusion. The answer was positive: across Experiments 1-4, after the instruction intervention, 60.0% - 70.0% of participants in the intervention group explicitly acknowledged the benefits of interleaving. Noteworthy is that, after the "placebo" intervention, 66.3% -

80.0% of participants in the control group continued to erroneously believe that massed learning is superior. These results establish the efficacy of the instruction intervention in mending metacognitive unawareness of the interleaving effect.

Given that metacognitive control is strictly related to metacognitive monitoring (Finn, 2008; Yang et al., 2017), the enhanced appreciation, induced by the instruction intervention, correspondingly increased self-adoption of the interleaved strategy during the second learning task. Specifically, in Experiments 1-4, the intervention group selected 61.6% - 67.7% of artists or butterfly species to be studied via the interleaved strategy, in contrast to 27.9% - 31.1% in the control group (Tauber et al., 2013; Yan et al., 2017).

Although all four experiments observed that enhanced use of interleaving only numerically improved test performance in the second learning task, when all results across Experiments 1-4 were integrated in the multilevel regression analyses, we did detect a significant enhancing effect of the instruction intervention on test performance, although the magnitude of the enhancement (about 4.6%) was modest (see below for further discussion). Furthermore, the multilevel analyses showed that the greater use of the interleaved strategy in the second learning task, the superior the test performance was. In addition, the intervention effect on test performance was fully mediated by its effect on strategy choices.

More importantly, Experiment 2 demonstrated that the intervention effect on strategy choices was long-lasting (at least 24 h), Experiment 3 found that the effect was transferable to learning of categories in a different domain (i.e., natural butterfly species), and Experiment 4 observed that the transfer effect is also long-lasting. These findings further establish the practical value of the instruction intervention developed here (McDaniel & Einstein, 2020).

Individual differences results (see the SI) demonstrated that, regardless of whether a given participant showed an interleaving effect or not in the first learning task, the instruction intervention always significantly improved metacognitive awareness of the effect, enhanced choices of the interleaved strategy, and numerically boosted test performance in the second learning task. These findings suggest little need to worry about individual differences issues.

Overall, explicit instructions, which include individual feedback and a description of the interleaving effect from previous research, can be employed as a practical strategy to reverse and overcome the massed-interleaved metacognitive illusion, promote self-adoption of the interleaved strategy, and facilitate inductive learning.

Some divergences between the study by Yan et al. (2016) and the current research should be elaborated. In Yan et al.'s Experiments 3-5, after completing the induction test, participants in a percentage group were informed that "*previous research has shown that 90% of individuals learn better when the paintings of an artist are presented intermixed with paintings by other artists*". After that, they indicated whether the interleaved or massed schedule led to better learning. The results showed that the proportion (40%) of participants believing M < I was roughly equal to the proportion (35%) believing M > I. Yan et al. proposed a possible explanation for why their intervention was insufficient to fully reverse the massing-is-better-than-interleaving belief. Specifically, they conjectured that their participants might regard themselves as the minority 10% of individuals for whom interleaving was not superior to massing. McDaniel and Einstein (2020) claimed that to promote self-regulated strategy use, learners must be convinced that the target strategy works for them.

Different from the percentage information presented in Yan et al. (2016), the instructions developed here did not involve uncertainty that interleaved learning only works for some (but not all) individuals. Instead, the current study directly showed participants their own test performance and the interleaving

effect results from previous research. In addition, for participants whose test performance did not exhibit an interleaving effect in the first learning task, the instructions explained that the inconsistency between their own results and those from previous research might be due to random error. More importantly, at the end of the instructions, we explicitly emphasized that interleaved (spaced) learning is more effective than massed learning. Such strong highlighting should be sufficient to convince learners about the benefits of interleaving.

In summary, we suggest that the reason why Yan et al.'s intervention was relatively less effective than the instruction intervention developed here is that their percentage information implied that interleaving is only beneficial for some individuals, and participants regarded themselves as the special individuals for whom interleaved learning does not work (for related discussion, see McDaniel & Einstein, 2020, p. 7).

Besides the above discussed practical implications, the present findings also bear theoretical importance. McDaniel and Einstein (2020) recently proposed a *knowledge, belief, commitment, and planning* (KBCP) framework to guide learning strategy training. A key component of the KBCP framework is the beliefs that learners hold about a given strategy, and McDaniel and Einstein (2020) proposed that, to train learners to use a given strategy, it is necessary to convince them that the strategy is truly helpful. Our results support the assumption of the KBCP framework about the important role of beliefs in study strategy use. Furthermore, the observed results are in line with McDaniel and Einstein's (p. 9) claim that "*participatory demonstrations that include feedback likely help students realize that there is a connection between their study strategies and their learning outcomes, making it more likely that they will expend the effort to use trained strategies in relevant learning contexts*".

Limitations and future research directions

Several limitations should be considered when interpreating the findings reported here, and more research to explore valid interventions is called for. For instance, all four experiments observed that the intervention effect on test performance was modest. This is hardly surprising because, after the intervention, not all participants in the intervention group believed that interleaved learning is more beneficial than massed learning, and the rate (61.6% - 67.7%) of self-employment of the interleaved strategy during the second learning task was far from perfect (100%) in the intervention group. Hence, future research could aim to develop and test more effective interventions.

Another noteworthy point is that it is still premature to recommend the instruction intervention to all populations before assessing its generalizability. Readers should bear in mind that all participants employed in the current study were Chinese university students, and all stimuli and instructions were presented in Chinese. Future research is encouraged to test the generalizability of the effects of the instruction intervention on metacognitive awareness, strategy choices, and learning outcomes to other populations (e.g., children, adolescents, and older adults) and with different languages in different countries or cultures.

In Experiments 1-4, all stimuli used in the first and second learning tasks were pictorial (i.e., artists' paintings or photographs of butterflies) and non-verbal. Future research needs to investigate whether the instruction intervention effect is generalizable to learning of verbal knowledge concepts, such as mathematical and statistical concepts (Foster et al., 2019; Sana et al., 2017) and social science concepts (Rawson et al., 2015).

Conclusion

Explicit instructions, which incorporate individual feedback and the findings from prior research about the interleaving effect on inductive learning, are valid to overcome the metacognitive illusion that interleaving is inferior to massing, promote self-employment of the interleaved strategy, and facilitate

inductive learning. Furthermore, the intervention effect on categorization accuracy is mediated by its effect on study strategy regulation. Most importantly, the intervention effect is long-lasting and transferable to learning of categories in other domains.

References

Ariel, R., & Karpicke, J. D. (2018). Improving self-regulated learning with a retrieval practice intervention. *Journal of Experimental Psychology: Applied, 24*, 43-56. doi:10.1037/xap0000133

Ball, B. H., Klein, K. N., & Brewer, G. A. (2014). Processing fluency mediates the influence of perceptual information on monitoring learning of educationally relevant materials. *Journal of Experimental Psychology: Applied, 20*, 336-348. doi:10.1037/xap0000023

Birnbaum, M. S., Kornell, N., Bjork, E. L., & Bjork, A. R. (2013). Why interleaving enhances inductive learning: The roles of discrimination and retrieval. *Memory & Cognition, 41*, 392-402. doi:10.3758/s13421-012-0272-7

Blackwell, L. S. (2002). *Psychological mediators of student achievement during the transition to junior high school: The role of implicit theories*: Columbia University.

Brunmair, M., & Richter, T. (2019). Similarity matters: A meta-analysis of interleaved learning and its moderators. *Psychological Bulletin, 145*, 1029-1052. doi:10.1037/bul0000209

Carpenter, S. K., & Mueller, F. E. (2013). The effects of interleaving versus blocking on foreign language pronunciation learning. *Memory & Cognition, 41*, 671-682. doi:https://doi.org/10.3758/s13421-012-0291-4

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:https://doi.org/10.1037/0033-2909.132.3.354

Choi, H., & Lee, H. S. (2020). Knowing is not half the battle: The role of actual test experience in the forward testing effect. *Educational Psychology Review, 32*, 765-789. doi:10.1007/s10648-020-09518-0

Dweck, C. S., & Yeager, D. S. (2020). A growth mindset about intelligence. In G. M. Walton & A. J. Crum (Eds.), *Handbook of Wise Interventions: How social psychology can help people change* (pp. 9-35). New York: The Guilford Press.

Eglington, L. G., & Kang, S. H. K. (2017). Interleaved presentation benefits science category learning. *Journal of Applied Research in Memory and Cognition, 6*, 475-485. doi:https://doi.org/10.1016/j.jarmac.2017.07.005

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191. doi:10.3758/BF03193146

Finn, B. (2008). Framing effects on metacognitive monitoring and control. *Memory & Cognition, 36*, 813-821. doi:https://doi.org/10.3758/MC.36.4.813

Firth, J., Rivers, I., & Boyle, J. (2021). A systematic review of interleaving as a concept learning strategy. *Review of Education, 9*, 642-684. doi:https://doi.org/10.1002/rev3.3266

Foster, N. L., Mueller, M., Was, C., Rawson, K. A., & Dunlosky, J. (2019). Why does interleaving improve math learning? The contributions of discriminative contrast and distributed practice. *Memory & Cognition, 47*, 1088-1101. doi:10.3758/s13421-019-00918-4

Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. R. (1989). *Induction: Processes of inference, learning, and discovery.*: Mit Press.

Hui, L., de Bruin, A. B. H., Donkers, J., & van Merriënboer, J. J. G. (2021). Does individual performance feedback increase the use of retrieval practice? *Educational Psychology Review, Advance online publication*. doi:10.1007/s10648-021-09604-x

Jacoby, L. L., Wahlheim, C. N., & Coane, J. H. (2010). Test-enhanced learning of natural concepts: effects on recognition memory, classification, and metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 1441-1451. doi:10.1037/a0020636

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:https://doi.org/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:https://doi.org/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

Kornell, N., & Vaughn, K. E. (2018). In inductive category learning, people simultaneously block and space their studying using a strategy of being thorough and fair. *Archives of Scientific Psychology, 6*, 138-147. doi:http://dx.doi.org/10.1037/arc0000042

McDaniel, M. A., & Einstein, G. O. (2020). Training learning strategies to promote self-regulation and transfer: The knowledge, belief, commitment, and planning framework. *Perspectives on Psychological Science, 15*, 1363-1381. doi:10.1177/1745691620920723

Metcalfe, J., & Xu, J. (2016). People mind wander more during massed than spaced inductive learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*, 978. doi:10.1037/xlm0000216

Pan, S. C., Tajran, J., Lovelett, J., Osuna, J., & Rickard, T. C. (2019). Does interleaved practice enhance foreign language learning? The effects of training schedule on Spanish verb conjugation skills. *Journal of Educational Psychology, 111*, 1172-1188. doi:10.1037/edu0000336

Rawson, K. A., Thomas, R. C., & Jacoby, L. L. (2015). The power of examples: Illustrative examples enhance conceptual learning of declarative concepts. *Educational Psychology Review, 27*, 483-504. doi:10.1007/s10648-014-9273-3

Rivers, M. L. (2021). Metacognition about practice testing: A review of learners' beliefs, monitoring, and control of test-enhanced learning. *Educational Psychology Review, 33*, 823-662. doi:10.1007/s10648-020-09578-2

Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 17*, 249-255. doi:https://doi.org/10.1111/j.1745-6916.2006.00012.x

Rohrer, D., Dedrick, R. F., & Burgess, K. (2014). The benefit of interleaved mathematics practice is not limited to superficially similar kinds of problems. *Psychonomic Bulletin & Review, 21*, 1323-1330. doi:10.3758/s13423-014-0588-3

Rohrer, D., Dedrick, R. F., & Hartwig, M. K. (2020). The scarcity of interleaved practice in mathematics textbooks. *Educational Psychology Review, 32*, 873-883. doi:10.1007/s10648-020-09516-2

Rozenshtein, A., Pearson, G. D. N., Yan, S. X., Liu, A. Z., & Toy, D. (2016). Effect of massed versus interleaved teaching method on performance of students in radiology. *Journal of the American College of Radiology, 13*, 979-984. doi:https://doi.org/10.1016/j.jacr.2016.03.031

Sana, F., Yan, V., & Kim, J. A. (2017). Study sequence matters for the inductive learning of cognitive concepts. *Journal of Educational Psychology, 109*, 84-98. doi:10.1037/edu0000119

Tauber, S. K., Dunlosky, J., Rawson, K. A., Wahlheim, C. N., & Jacoby, L. L. (2013). Self-regulated learning of a natural category: Do people interleave or block exemplars during study? *Psychonomic Bulletin & Review, 20,* 356-363. doi:10.3758/s13423-012-0319-6

Taylor, K., & Rohrer, D. (2010). The effects of interleaved practice. *Applied Cognitive Psychology, 24,* 837-848. doi:https://doi.org/10.1002/acp.1598

Thule, E. J. (2005). *Accuracy of metacognitive monitoring and learning of texts.* (Master dissertation). University of Toronto (Canada), Ann Arbor. Retrieved from https://search.proquest.com/docview/305366978?accountid=8554

Vlach, H. A., & Sandhofer, C. M. (2012). Distributing learning over time: The spacing effect in children's acquisition and generalization of science concepts. *Child Development, 8,* 1137-1144. doi:https://doi.org/10.1111/j.1467-8624.2012.01781.x

Wong, S. S. H., Low, A. C. M., Kang, S. H. K., & Lim, S. W. H. (2020). Learning music composers' styles: To block or to interleave? *Journal of Research in Music Education, 68,* 156-174. doi:10.1177/0022429420908312

Yan, V. X., Bjork, E. L., & Bjork, R. A. (2016). On the difficulty of mending metacognitive illusions: A priori theories, fluency effects, and misattributions of the interleaving benefit. *Journal of Experimental Psychology: General, 145*, 918-933. doi:10.1037/xge0000177

Yan, V. X., Schuetze, B. A., & Eglington, L. G. (2020). A review of the interleaving effect: Theories and lessons for future research. *PsyArXiv*. doi:https://doi.org/10.31234/osf.io/ur6g7

Yan, V. X., Soderstrom, N. C., Seneviratna, G. S., Bjork, E. L., & Bjork, R. A. (2017). How should exemplars be sequenced in inductive learning? Empirical evidence versus learners' opinions.

Journal of Experimental Psychology: Applied, 23, 403-416. doi:http://dx.doi.org/10.1037/xap0000139

Yang, C., Huang, T. S. T., & Shanks, D. R. (2018). Perceptual fluency affects judgments of learning: The font size effect. *Journal of Memory and Language, 99*, 99-110. doi:https://doi.org/10.1016/j.jml.2017.11.005

Yang, C., Luo, L., Vadillo, M. A., Yu, R., & Shanks, D. R. (2021). Testing (quizzing) boosts classroom learning: A systematic and meta-analytic review. *Psychological Bulletin, 147*, 399-435. doi:10.1037/bul0000309

Yang, C., Potts, R., & Shanks, D. R. (2017). Metacognitive unawareness of the errorful generation benefit and its effects on self-regulated learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*, 1073-1092. doi:10.1037/xlm0000363

Yang, C., Potts, R., & Shanks, D. R. (2018). Enhancing learning and retrieval of new information: A review of the forward testing effect. *npj Science of Learning, 3*, 8. doi:10.1038/s41539-018-0024-y

Yang, C., & Shanks, D. R. (2018). The forward testing effect: Interim testing enhances inductive learning. *Journal of Experimental Psychology: Learning, Memory & Cognition, 44*, 485-492. doi:10.1037/xlm0000449

Yang, C., Sun, B., & Shanks, D. R. (2018). The anchoring effect in metamemory monitoring. *Memory & Cognition, 46*, 384–397. doi:https://doi.org/10.3758/s13421-017-0772-6

Yang, C., Yu, R., Hu, X., Luo, L., Huang, T., & Shanks, D. R. (2021). How to assess the contributions of processing fluency and beliefs to the formation of judgments of learning: methods and pitfalls. *Metacognition and Learning, 16*, 319-343. doi:10.1007/s11409-020-09254-4

Zulkiply, N., & Burt, J. S. (2013). The exemplar interleaving effect in inductive learning: Moderation by the difficulty of category discriminations. *Memory & Cognition, 41*, 16-27. doi:10.3758/s13421-012-0238-9

Zulkiply, 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:https://doi.org/10.1016/j.learninstruc.2011.11.002

Table 1. Test performance results (i.e., interleaving effect) in Experiment 1-4's first learning task.

Experiment	Interleaved	Massed	Difference & 95% CI	*t*(29)	*p*	Cohen's *d*
Experiment 1						
Intervention	52.4% (18.1%)	39.4% (18.0%)	12.9% [6.6%, 19.3%]	4.15	< .001	0.76
Control	49.2% (20.0%)	38.8% (18.9%)	10.4% [2.9%, 18.0%]	2.82	.009	0.51
Experiment 2						
Intervention	52.4% (18.9%)	40.8% (16.6%)	11.5% [4.1%, 18.9%]	3.18	.003	0.58
Control	50.4% (17.7%)	38.8% (18.9%)	11.7% [2.2%, 21.1%]	2.52	.020	0.46
Experiment 3						
Intervention	50.1% (18.0%)	37.8% (22.3%)	12.4% [5.8%, 18.9%]	3.84	< .001	0.70
Control	51.0% (16.4%)	38.9% (21.5%)	12.1% [4.4%, 19.8%]	3.22	.003	0.59
Experiment 4						
Intervention	56.0% (17.0%)	39.9% (19.3%)	16.1% [9.4%, 22.9%]	4.87	< .001	0.89
Control	52.5% (16.0%)	40.7% (19.0%)	11.8% [3.8%, 19.8%]	3.00	.005	0.55

Note: The second and third columns list *M* (*SD*) of test performance.

Table 2. Intervention effects on metacognitive awareness in Experiments 1-4.

	Intervention group		Control group	
Experiments	$\chi^2(2)$	*p*	$\chi^2(2)$	*p*
Experiment 1	16.02	< .001	0.42	.81
Experiment 2	17.39	< .001	2.27	.32
Experiment 3	16.09	< .001	0.00	1.00
Experiment 4	19.47	< .001	0.13	.94

Note: The relevant tests compare judgments about the interleaving effect before and after the intervention.

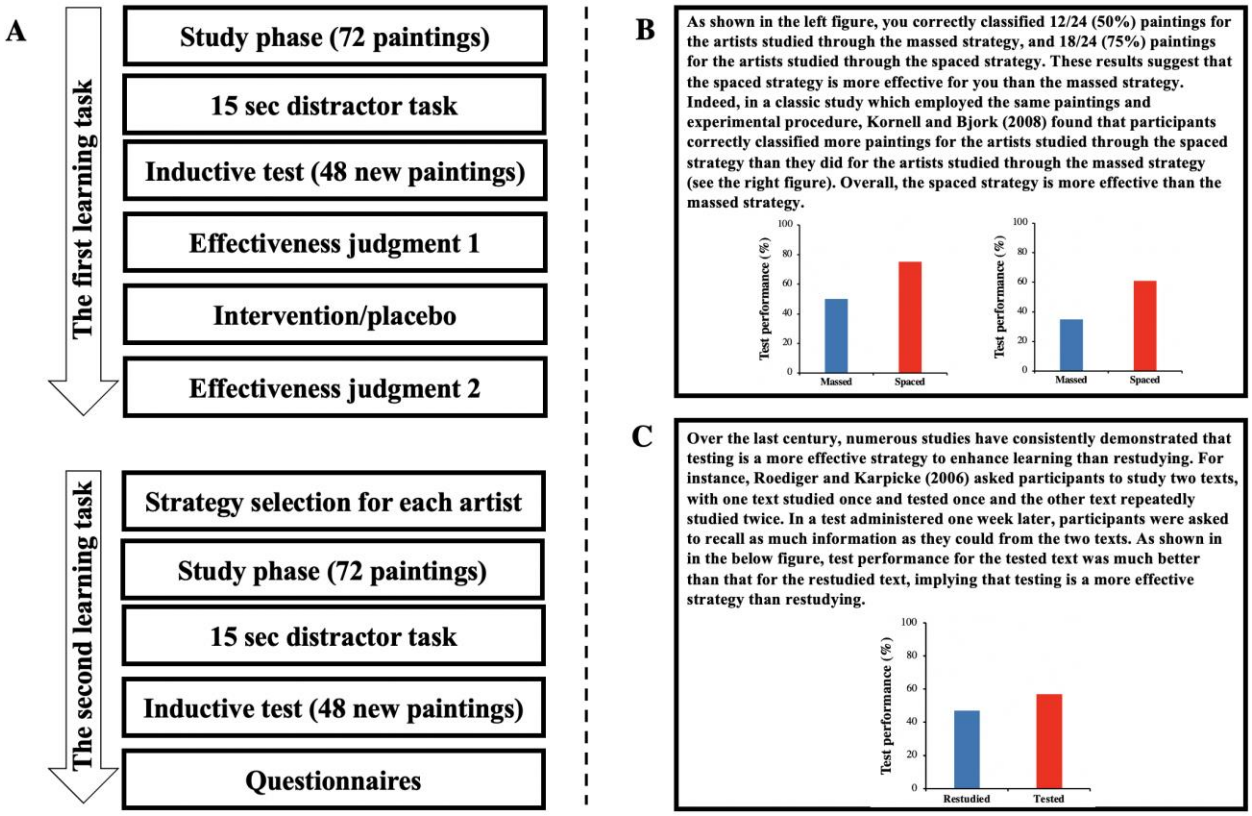

Figure 1. A: Schematic illustration of the task procedure in Experiment 1. B: Intervention instructions and figures for the intervention group. C. Intervention instructions and figure for the control group. Note that these are English translations of the original (Chinese) materials.

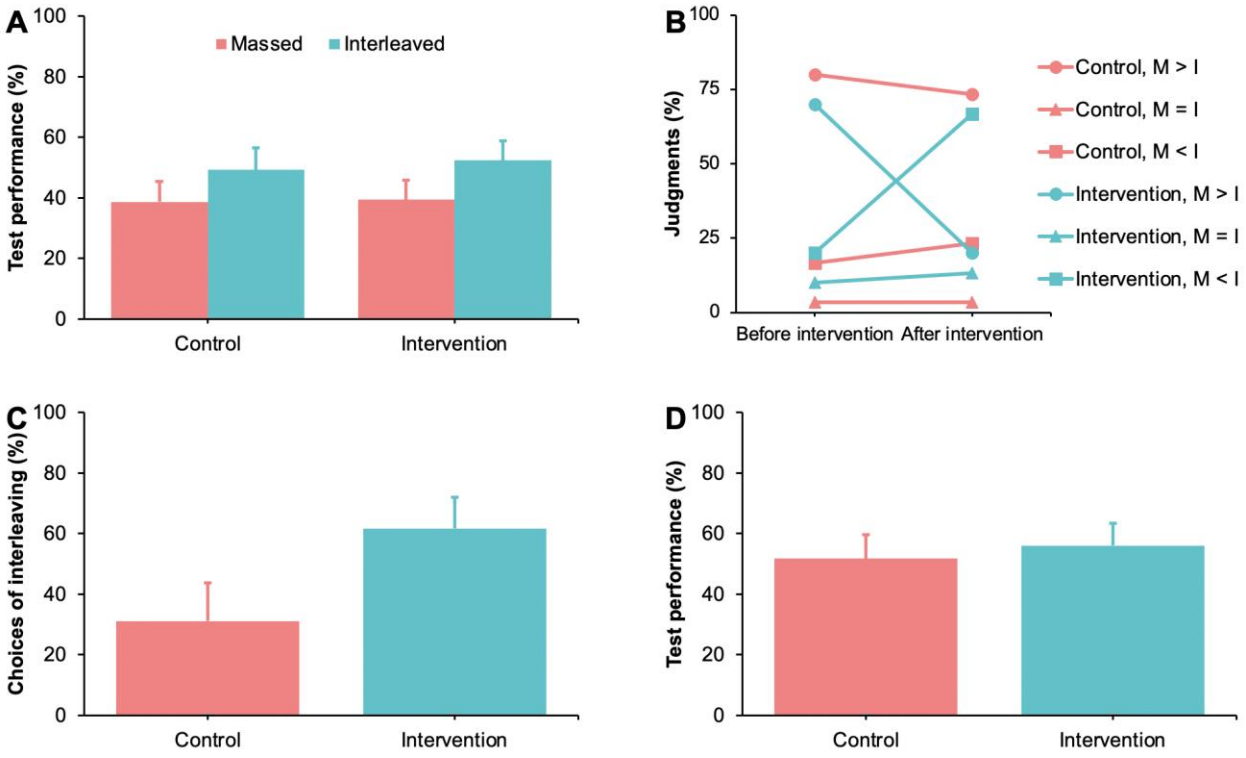

Figure 2. Results of Experiment 1. A: Test performance in the first learning task as a function of group and study strategy. B: Metacognitive judgments as a function of group and judgment period. C: Proportion of artists selected to be studied using the interleaved strategy. D: Test performance in the second learning task. Error bars represent 95% CI.

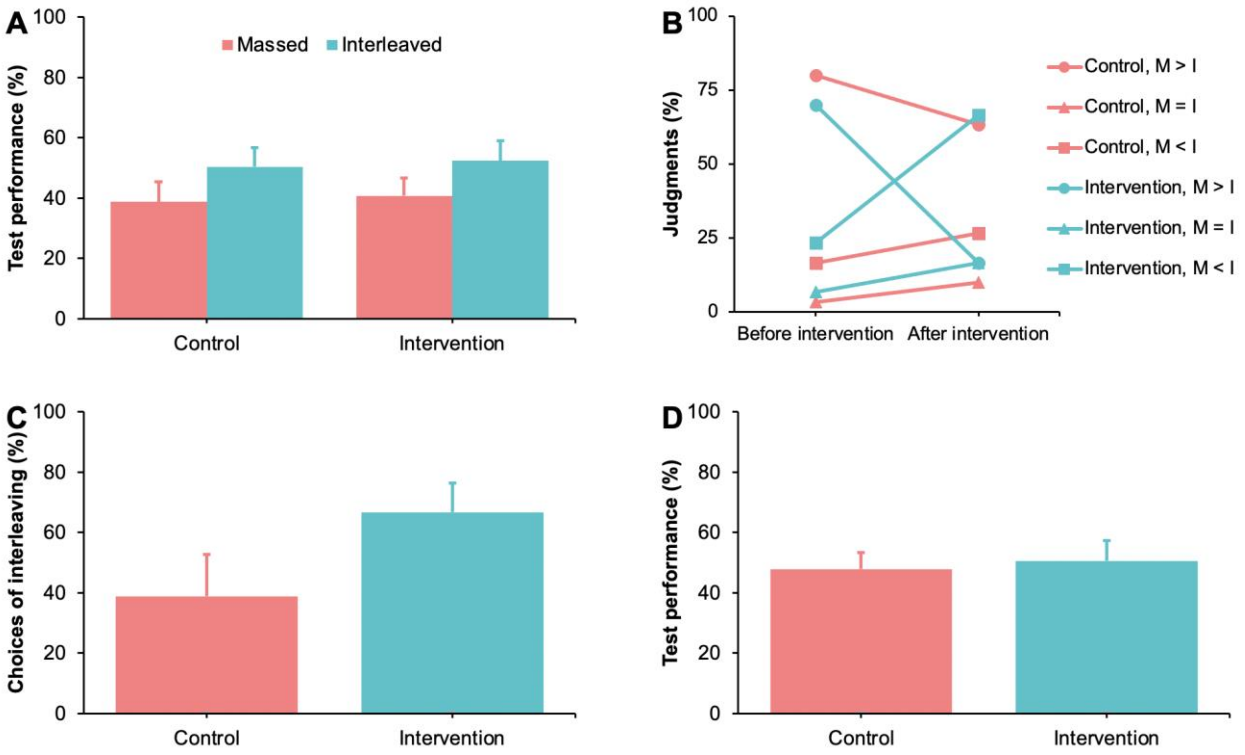

Figure 3. Results of Experiment 2. A: Test performance in the first learning task as a function of group and study strategy. B: Metacognitive judgments as a function of group and judgment period. C: Proportion of artists selected to be studied using the interleaved strategy. D: Test performance in the second learning task. Error bars represent 95% CI.

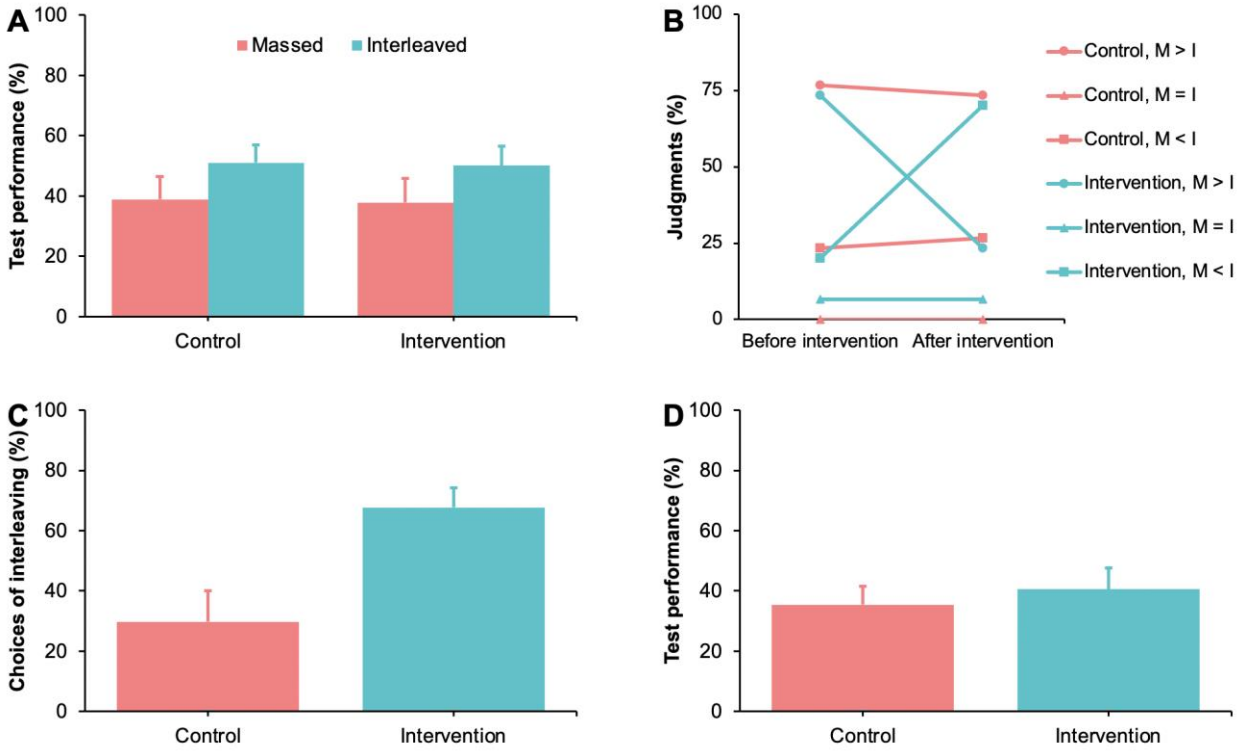

Figure 4. Results of Experiment 3. A: Test performance in the first learning task as a function of group and study strategy. B: Metacognitive judgments as a function of group and judgment period. C: Proportion of butterfly species selected to be studied using the interleaved strategy. D: Test performance in the second learning task. Error bars represent 95% CI.

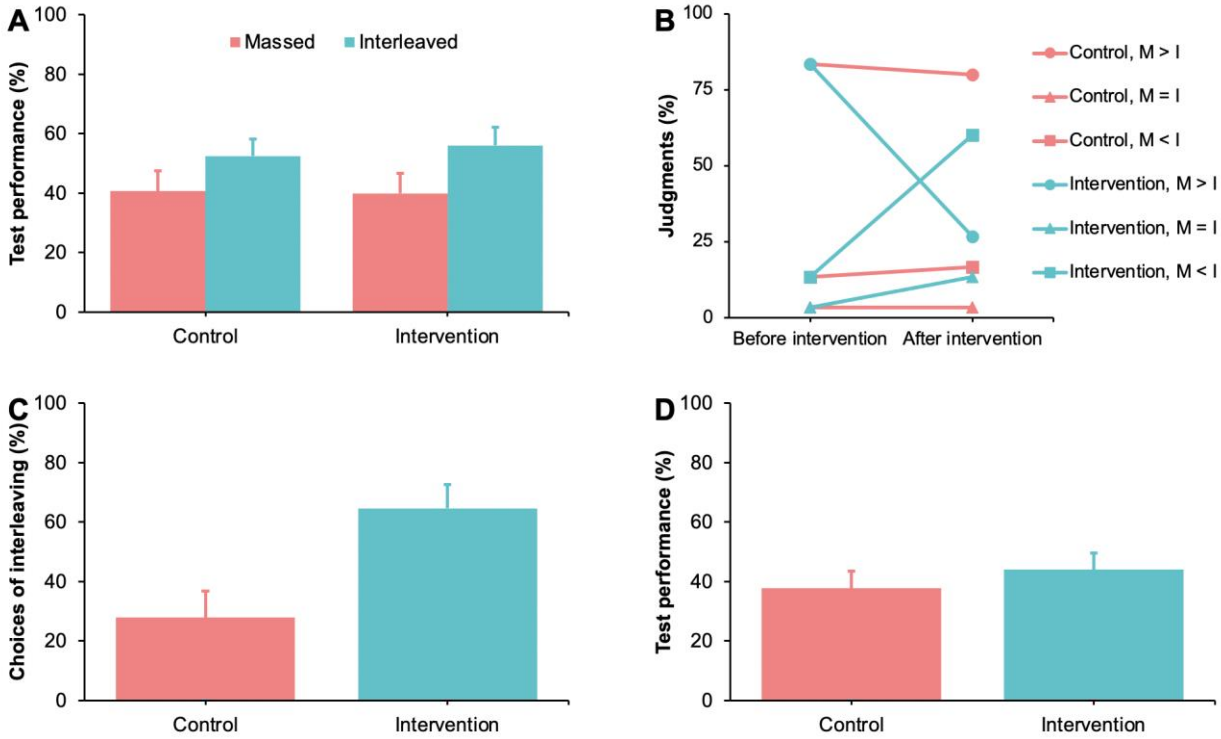

Figure 5. Results of Experiment 4. A: Test performance in the first learning task as a function of group and study strategy. B: Metacognitive judgments as a function of group and judgment period. C: Proportion of butterfly species selected to be studied using the interleaved strategy. D: Test performance in the second learning task. Error bars represent 95% CI.

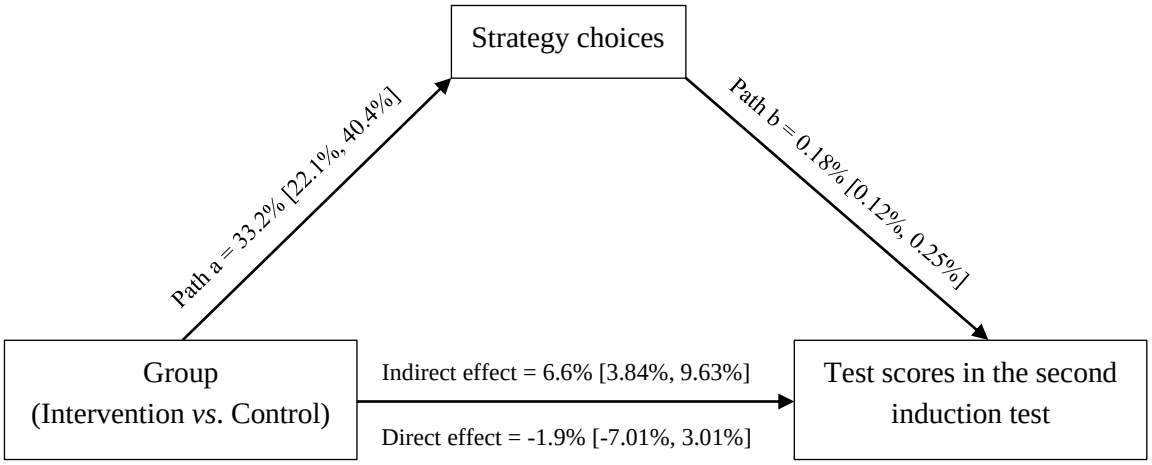

Figure 6. Results from the multilevel mediation analysis.