Mem Cogn
DOI 10.3758/s13421-010-0063-y

Spacing enhances the learning of natural concepts: an investigation of mechanisms, metacognition, and aging

Christopher N. Wahlheim · John Dunlosky · Larry L. Jacoby

Abstract In two experiments, we examined spacing effects on the learning of bird families and metacognitive assessments of such learning. Results revealed that spacing enhanced learning beyond massed study. These effects were increased by presenting birds in pairs so as to highlight differences among families during study (Experiment 1). Self-allocated study time provided evidence that more attention was paid during spaced than during massed study and resulted in no age differences in learning (Experiment 2). Metacognitive measures revealed sensitivity to the processing advantage of spaced study and to differences in classification difficulty across categories. No difference occurred in monitoring accuracy for young versus older adults. These findings provide evidence for discrimination- and attention-based accounts of the spacing effect in natural concept learning.

Keywords Concept learning · Categorization · Spacing effects · Metacognition · Aging

The experiments reported in this article follow up on the spacing effects found by Kornell and colleagues (Kornell & Bjork, 2008; Kornell, Castel, Eich, & Bjork, 2010) in their investigations of natural concept learning. In their experiments, paintings by various artists were studied under massed or spaced study conditions. For the massed study condition, blocks of paintings by the same artist were presented sequentially. In contrast, for spaced study, paintings by different artists were intermixed in blocks, and additional paintings by the same artist were not presented until after some number of paintings by other artists had occurred. At the time of test, participants were instructed to classify new paintings from studied artists. The results revealed that both young and older adults classified new paintings more accurately when the originals were studied in spaced, rather than massed, fashion. However, posttest questionnaires revealed that participants believed their learning to be better after massed than after spaced practice.

Consistent with participants' beliefs, Kornell and Bjork (2008) did not expect spacing to benefit natural concept learning. In fact, Kornell and Bjork predicted just the opposite, in that they expected massed study to enhance concept learning beyond that of spaced study. Their prediction was based on the notion that massed presentation of exemplars encourages one to notice similarities among exemplars of a category, whereas spaced presentation makes it more difficult to abstract the common elements across exemplars (cf. Rothkopf, as quoted in Kornell & Bjork, 2008). This line of reasoning is consistent with early findings in which an advantage occurred for massed study in artificial concept learning (e.g., Gagné, 1950; Kurtz & Hovland, 1956).

The experiments reported in this article extended investigation of the effects of massed versus spaced presentation on the learning of natural concepts to effects on the classification of bird families. To account for the disparate effects of spacing with natural and artificial concepts, Kornell and Bjork (2008) posited that spacing effects in concept learning might be

C. N. Wahlheim (✉)
Department of Psychology, Washington University in St. Louis, St. Louis, MO 63130, USA
e-mail: cnwahlheim@gmail.com

J. Dunlosky
Kent State University, Kent, OH, USA

L. L. Jacoby
Washington University in St. Louis, St. Louis, MO, USA

moderated by the level of discriminability between concepts. This *discrimination* hypothesis holds that because artificial concepts can often be easily discriminated from one another, no additional learning benefits are gained from the examination of differences among exemplars of a category. In contrast, because natural concepts are presumed to have more between-category similarity, learning how to discriminate among categories may be beneficial to learning. Spacing may be ideal for learning natural concepts for which there is likely to be high intercategory similarity for the novice (e.g., bird families). As compared with massed practice, spacing results in the intermixing of birds from different families, which makes it easier for participants to search for differences among exemplars. In line with the discrimination hypothesis, bird experts hold that learning to identify birds is best accomplished by searching for differences across birds. According to expert David Sibley (2002), "Whether you are looking at two birds side by side in the field or comparing a bird in the field to pictures in a book, you must make comparisons and search for differences" (p. 22).

An alternative account of the spacing effect in natural concept learning is provided by the *attention attenuation* hypothesis (from Kornell et al., 2010). According to this hypothesis, spacing effects obtain in natural concept learning because attention is more likely to diminish across exemplars in a study block if those exemplars are from the same category. In the present experiments, as in Kornell and Bjork (2008), massed presentation involved presenting six exemplars from a given category in a block, with each exemplar being presented for a fixed amount of time. After several exemplars from one category were presented, participants might believe that they had learned the category and, so, spend less time studying subsequent exemplars within that block. In contrast, spaced presentation resulted in six exemplars within a block being from different categories, rendering such attenuation of attention less likely. By this account, functional study times may actually be longer for spaced than for massed study, which may contribute to the spacing effect.

In the present experiments, we sought to extend the spacing effect observed by Kornell and Bjork (2008) to the classification of birds into families and to further evaluate the two theoretical accounts described above. Birds were chosen from 12 families, and exemplars were presented in massed or spaced fashion, using the same list structure as did Kornell and Bjork. For massed study, birds from the same family were presented sequentially in blocks, whereas for spaced study, birds from different families were intermixed in blocks and additional exemplars from those families were presented after intervening massed study of members of different families.

To further test the discrimination hypothesis, birds were presented either individually or side by side in pairs in our experiments. This manipulation was used to enhance the salience of similarities and differences. As compared with presenting exemplars individually, presentation in pairs was expected to increase the extent to which participants noticed similarities (massed) or differences (spaced) between exemplars. According to the discrimination hypothesis, the spacing effect will be further enhanced when exemplars are presented in pairs, as compared with when they are presented individually.

Concerning the attention attenuation hypothesis, Kornell et al. (2010) evaluated this account in an ad hoc fashion and found no supporting evidence for it, whereas we provide a more direct test of the hypothesis. In particular, examination of classification performance for studied exemplars in each position within a block might reveal that performance decreases as a function of relative position for massed exemplars, whereas no such decrease will arise for spaced exemplars. Furthermore, allowing participants to allocate their study time would provide an index of functional attention for exemplars at each position. Thus, one might expect a similar pattern of results for classification performance and study time. To examine these possibilities, we analyzed both classification performance for studied exemplars (Experiment 1) and study time allocation (Experiment 2) as a function of relative position within each block.

As was mentioned above, Kornell and colleagues (2010) found that people's metacognitive beliefs were largely inconsistent with the effects of spacing on classification performance. In particular, the majority of people reported that massed study was more effective than spaced study on posttest questionnaires. Although questionnaires are a valid means of assessing one's beliefs, questionnaire responses may not necessarily reflect participants' sensitivity to the effects of spacing on category learning. Participants' judgments on a questionnaire may be based on the fluency with which they had processed the *group* of exemplars in each of the massed and spaced conditions. In contrast, judgments made for specific exemplars or even specific categories may be a better measure of participants' sensitivity to these effects because the fluency on which these judgments are based is specific to particular exemplars and categories.

In line with this reasoning, Dunlosky and Hertzog (2000) have shown that posttest questionnaires can differ from individual judgments of learning (JOLs) in terms of their sensitivity to imagery effects on associative learning. This difference is presumably due to questionnaire responses being based on declarative knowledge about a group of items, whereas individual judgments are based on the processing of particular items. Results such as these suggest that judgments made at the individual or category levels may be sensitive to spacing effects in concept learning. To explore this possibility, participants in the present study were asked to make JOLs for each studied

item and retrospective confidence judgments for all test items. In addition, participants made category-learning judgments (CLJs) for each family (see Jacoby, Wahlheim, & Coane, 2010). For the CLJs, family names were presented individually following the study phase, and participants predicted the likelihood that they could correctly classify new birds from those families when subsequently tested.

Although item-level JOLs and confidence judgments have been studied extensively in the metacognition literature, little is known about how people judge their learning of categories (but see Jacoby et al., 2010). Examination of the sensitivity of CLJs to differences in learning across categories has the potential to inform theory about the monitoring of concept learning. In particular, examination of CLJs will indicate how well participants believe particular categories have been learned. Thus, besides evaluating whether people's metacognitive judgments are sensitive to spacing effects, our experiments are meant to motivate theoretical development regarding metacognition and concept learning.

We also sought to evaluate age differences in the learning of members of bird families and whether allowing participants to control their study strategies would eliminate any age differences that were found. On the basis of prior research, older adults are less likely to self-initiate effective learning strategies (see Castel, 2008), show poorer concept learning than young adults (see, e.g., Griego & Kliegel, 2008; Hess & Slaughter, 1986; Kornell et al., 2010), and show poorer metacognitive control over their study than do young adults (e.g., Dunlosky & Connor, 1997). Thus, one might expect young adults to outperform older adults, even when both groups are given unlimited time to study. In contrast, learning bird families is an intrinsically engaging task that could motivate older adults to use more effective learning strategies, leading to no age-related differences in the learning of members of bird families. We examined these possibilities in Experiment 2 by allowing participants to allocate their study time and by comparing the monitoring accuracy of young and older adults.

Experiment 1

Experiment 1 examined mechanisms of the spacing effect in natural concept learning and metacognitive evaluations of such effects. Birds were studied in massed or spaced fashion, following the procedure employed by Kornell and Bjork (2008). JOLs, CLJs, and confidence judgments were recorded. Presenting birds in pairs was expected to highlight the differences among exemplars and enhance the positive effects of spacing, providing support for the discrimination hypothesis. Also, in accord with the attention attenuation hypothesis, classification of studied exemplars in massed blocks was expected to be greatest for the first exemplar, with performance declining for subsequent exemplars, and with little change being expected in the spaced blocks. Finally, in contrast to questionnaire measures, our metacognitive measures were expected to reflect processing differences between massed and spaced study as a result of judgments being influenced by item- and category-level processing.

Method

Participants

Forty-eight Washington University undergraduates participated in exchange for course credit or $10 per hour. Twenty-four participants were randomly assigned to the singles and pairs groups. Participants were tested individually.

Design and materials

A 2 (study: massed vs. spaced) × 2 (presentation: singles vs. pairs) × 2 (exemplar: studied vs. novel) mixed factorial design was used. The study variable refers to whether study blocks included birds from the same family (massed) or from different families (spaced). In addition, the presentation variable refers to whether one (singles) or two (pairs) exemplars were presented during study. Finally, the exemplar variable refers to whether test items were presented during the study phase (studied) or whether they were new exemplars from the studied families (novel).

Pictures of perching birds from the taxonomic order Passeriformes were chosen to represent natural concepts (for examples, see Fig. 1). We selected families from the same taxonomic order to provide enough between-family similarity to avoid ceiling effects. Also, families were chosen such that there was enough within-family similarity to produce reasonable levels of discriminability between families. Exemplars were chosen from the following 12 families: chickadees, finches, flycatchers, grosbeaks, jays, orioles, sparrows, swallows, thrashers, thrushes, vireos, and warblers.

Of the 12 families chosen, 10 species (exemplars) were selected from each, for a total of 120 exemplars. For counterbalancing purposes, the 12 families were divided into two groups of 6 families and served equally often in the massed and spaced study conditions. Each group was matched on classification difficulty from previous experiments. Of the 10 exemplars in each family, 6 were randomly assigned to be studied exemplars, and 4 were randomly assigned to be novel exemplars. Items in the studied and novel exemplar conditions remained constant across experimental formats.

Fig. 1 Examples of exemplars from each of the 12 families used in Experiments 1 and 2

During study, exemplars were presented in massed fashion (M) by presenting the six studied items from the same family in a block. In contrast, exemplars in the spaced condition (S) were distributed across blocks such that each block contained one exemplar from each family. The block order was the same as that used by Kornell and Bjork (2008; i.e., MSSMMSSMMSSM). Families and exemplars were assigned to massed and spaced blocks in a fixed random order. Exemplars in each block were presented in a new random order for each participant. The counterbalancing of assignment of groups to massed and spaced study conditions produced two experimental formats.

In the singles group, each block contained six individually presented exemplars. In contrast, each block in the pairs group included three presentations of two exemplars side by side. On each presentation in the pairs group, two exemplars from different families were shown in the spaced condition, and two exemplars from the same family were shown in the massed condition.

Procedure

Participants first completed the study phase. All stimuli were presented on a computer monitor against a black background. Six exemplars from each of the 12 families were presented in random order within each study block (72 exemplars total). In the singles group, each exemplar was presented for 8 s, with the corresponding family name below. Participants were told to say the family name aloud and to study the picture in anticipation of classification tests of novel and studied exemplars. After 8 s expired, the JOL query "Likelihood of Correct Classification (8%–100%)" replaced the family name. Participants were told that they would be required to select the correct family name of each bird on the final classification test from a list of 12 families and that their task was to predict the likelihood of correctly classifying each studied bird. Participants' ratings were made on a scale from 8% (*guessing*) to 100% (*certain correct*). They were told to use the full range of the scale to make their predictions as precise as possible. In the pairs group, two exemplars were presented side by side for 16 s, with their respective family names below. Participants were told that the birds would appear for 16 s and that they should spend an equal amount of time studying each bird. After 16 s expired, the left member of the pair was presented alone, and the JOL query appeared below it. Once the JOL was recorded for the left member, the right member appeared alone with the same query.

After the study phase, participants were told to make predictions of future classification performance for novel exemplars from the 12 studied families at the category level (CLJs). Each family name was presented individually in random order, with the same judgment query and scale (8%–100%) as in the study phase.

Following the CLJs, participants were given a classification test for the novel exemplars. Four novel exemplars from each of the 12 studied families (48 total) were presented individually and in random order. The names of the 12 families appeared below each exemplar. Participants were told that they should say the name of the family to which each bird belonged and that their responses would be recorded by the experimenter. After each response, participants made confidence judgments regarding their classification accuracy. The confidence scale was the same as that used in earlier judgments.

Once the classification test of novel exemplars had been completed, participants were given the same test for studied exemplars. The original six exemplars from each of the 12 families (72 total) were presented individually and in

random order, and the same instructions were given as in the test of novel exemplars. Tests of novel and studied exemplars were both self-paced. All effects in the following experiments were significant below $p = .05$, unless otherwise noted.

Results and discussion

Classification performance

Novel exemplars To examine the role of discriminability in the effects of spacing on natural concept learning, classification performance was compared for novel exemplars (Table 1, upper left section). A spacing effect was revealed in that classification performance was better for spaced (.43) than for massed (.35) study, $F(1, 46) = 10.68$, $\eta_p^2 = .19$. The singles and pairs groups did not differ, $F < 1$. Most important, a study × presentation interaction, $F(1, 46) = 4.81$, $\eta_p^2 = .10$, indicated that there was an advantage of spaced, as compared with massed, study in the pairs group, $t(23) = 3.40$, but not in the singles group, $t < 1$. Consistent with the discrimination account, these results suggest that structuring the study trials so as to highlight differences between families benefited participants' learning by enhancing their ability to discriminate among families.

Studied exemplars A spacing effect was also obtained for studied exemplars presented in pairs, whereas there was not a significant effect of spacing for individually presented exemplars (Table 1, upper right section). Classification performance was better for spaced (.47) than for massed (.39) study, $F(1, 46) = 17.83$, $\eta_p^2 = .28$, and the study × presentation interaction was also significant, $F(1, 46) = 7.21$, $\eta_p^2 = .14$. Classification performance was better for spaced than for massed exemplars in the pairs group, $t(23) = 4.33$, but not in the singles group, $t(23) = 1.27$, $p = .22$. These results provide further evidence that spacing benefits the learning of natural concepts when differences among families are sufficiently highlighted.

Table 1 Probability of correct classification as a function of study, presentation, exemplar, and age: Experiments 1 and 2

	Novel		Studied	
	Massed	Spaced	Massed	Spaced
Experiment 1				
Young				
Singles	.36 (.03)	.39 (.03)	.41 (.03)	.44 (.03)
Pairs	.33 (.03)	.46 (.03)	.38 (.03)	.50 (.03)
Experiment 2				
Young				
Singles	.33 (.03)	.39 (.03)	.31 (.03)	.38 (.03)
Pairs	.32 (.02)	.35 (.03)	.32 (.03)	.37 (.03)
Older				
Singles	.31 (.03)	.35 (.03)	.32 (.03)	.34 (.03)
Pairs	.36 (.03)	.42 (.03)	.38 (.03)	.44 (.03)

Standard errors of the means are presented in parentheses

The inclusion of a test of studied exemplars also allowed for a test of the attention attenuation hypothesis (see Kornell et al., 2010). Classification performance was examined as a function of relative position within a block (six birds per block), collapsed across presentation groups (singles and pairs) in the massed and spaced study conditions (Fig. 2). Exemplars in the pairs group are separated for display in Fig. 2 by assigning the left exemplars to positions earlier than those for right exemplars. For example, relative position 1 includes the left member of the first pair, relative position 2 includes the right member, and so forth.

Consistent with the attention attenuation hypothesis, classification performance in the massed condition was higher in the first than in the second position, with each subsequent position being numerically lower, with the exception of position 5 (perhaps because of random error). In contrast, performance in the spaced condition increased from the first to the second position, with there being no systematic change in performance across positions. This pattern of results was confirmed by a significant study × position interaction, $F(5, 230) = 3.69$, $\eta_p^2 = .07$. Follow-up *t*-tests revealed significant differences between adjacent positions in the massed condition in positions 1 and 2 and in positions 5 and 6, $ts(47) > 2.23$. The only significant difference between adjacent positions in the spaced condition was between positions 1 and 2, $t(47) = 2.46$. These results provide preliminary evidence that spacing effects were partially due to less attention being given to exemplars

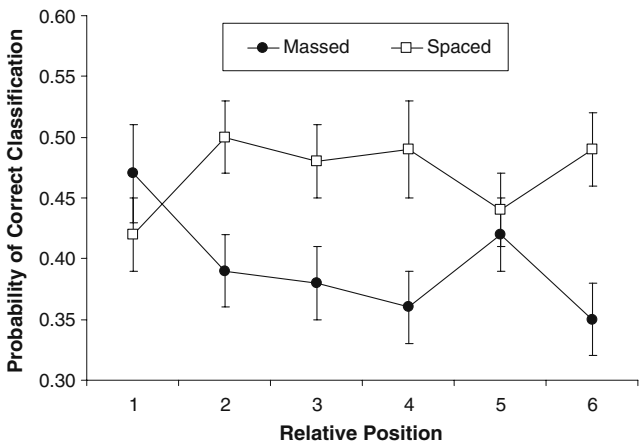

Fig. 2 Mean classification performance for studied exemplars collapsed across presentation groups (singles and pairs) as a function of relative position and study condition in Experiment 1. The bars represent standard errors of the means

studied later in massed blocks. We examined this issue further in Experiment 2.

Metacognitive judgments

The accuracy of metacognitive judgments was assessed in two ways. First, the magnitudes of each measure were compared as a function of study and presentation conditions to examine participants' sensitivity to the effects of spacing. Second, monitoring resolution was examined for each measure by computing mean within-participant gamma correlations between the measure and classification performance (see Nelson, 1984). Monitoring resolution was computed to index the extent to which people's judgments predicted classification performance for individual items or categories.

Judgments of learning The top rows of Table 2 show that JOLs did not differ for spaced and massed study conditions or for singles and pairs groups, and there was not a significant study × presentation interaction, Fs < 3.06, ps > .09. However, spacing did enhance monitoring resolution of JOLs. The top rows of Table 3 show that resolution was better for spaced (.48) than for massed (.35) study, $F(1, 46) = 6.03$, $\eta_p^2 = .12$. No other effects were significant, Fs < 1.07.

Category-learning judgments CLJs were sensitive to the effects of spacing (Table 2, top rows), as revealed by their being higher for spaced (.50) than for massed (.43) study, $F(1, 46) = 34.24$, $\eta_p^2 = .43$. No other effects were significant, Fs < 2.58, ps > .12. It is noteworthy that CLJs were more sensitive to the effects of spacing than were JOLs, given that previous work has shown global judgments to be insensitive to spacing effects (see Kornell & Bjork, 2008; Kornell et al., 2010).

Monitoring resolution for CLJs was computed by correlating, for each participant, the CLJs given for each family with the mean overall classification performance on novel exemplars for each of those families (Table 3, top rows). Resolution did not differ between massed (.47) and spaced (.48) study conditions, Fs < 1. This result is understandable in that spacing was expected to benefit the encoding of individual exemplars, which would not necessarily benefit relative monitoring accuracy at the category level. If spacing had a relatively uniform effect on performance across categories, its increase would not enhance discrimination of difficulty differences among categories. Moreover, both specific and general category representations likely contribute to CLJs (see Jacoby et al., 2010), and general information could have been used in the massed condition to offset the memory disadvantage of individual exemplars.

The magnitude of gamma correlations for CLJs indicates that participants were sensitive to differences in classification difficulty across categories, which replicates the results found by Jacoby et al. (2010).

Confidence judgments Confidence judgments were sensitive to the effects of spacing (Table 2, top rows). Confidence in classification of studied and novel exemplars was higher for spaced (.45) than for massed (.43) study, $F(1, 46) = 5.20$, $\eta_p^2 = .10$. Confidence was also higher for studied (.48) than for novel (.40) exemplars, $F(1, 46) = 79.49$, $\eta_p^2 = .63$. Finally, confidence tended to be higher for pairs (.48) than for singles (.39), $F(1, 46) = 4.04$, $p = .05$, $\eta_p^2 = .08$. No other effects were significant, Fs < 1.

Monitoring resolution of confidence judgments benefited from spaced study, but only for studied exemplars (Table 3, top rows). This finding was confirmed by a significant exemplar × study interaction, $F(1, 46) = 4.16$, $\eta_p^2 = .08$.

Table 2 Metacognitive judgment magnitudes as a function of measure, study, presentation, exemplar, and age: Experiments 1 and 2

	JOLs		CLJs		Confidence			
					Novel		Studied	
	Massed	Spaced	Massed	Spaced	Massed	Spaced	Massed	Spaced
Experiment 1								
Young								
Singles	.44 (.04)	.46 (.04)	.39 (.03)	.46 (.03)	.34 (.03)	.36 (.03)	.43 (.03)	.45 (.03)
Pairs	.51 (.04)	.53 (.04)	.47 (.03)	.53 (.03)	.43 (.03)	.45 (.03)	.51 (.03)	.53 (.03)
Experiment 2								
Young								
Singles	.44 (.04)	.45 (.04)	.43 (.04)	.46 (.04)	.44 (.03)	.45 (.04)	.44 (.04)	.46 (.04)
Pairs	.46 (.03)	.44 (.03)	.40 (.03)	.43 (.03)	.41 (.03)	.42 (.03)	.46 (.04)	.48 (.04)
Older								
Singles	.37 (.04)	.43 (.04)	.33 (.04)	.37 (.04)	.31 (.04)	.33 (.04)	.33 (.04)	.35 (.04)
Pairs	.43 (.04)	.46 (.04)	.35 (.04)	.41 (.04)	.41 (.04)	.42 (.04)	.47 (.04)	.48 (.04)

Standard errors of the means are presented in parentheses. JOL, judgment of learning; CLJ, category-learning judgment

Table 3 Monitoring resolution as a function of measure, study, presentation, exemplar, and age: Experiments 1 and 2

	JOLs		CLJs		Confidence			
					Novel		Studied	
	Massed	Spaced	Massed	Spaced	Massed	Spaced	Massed	Spaced
Experiment 1								
Young								
Singles	.30 (.06)	.48 (.04)	.44 (.07)	.47 (.08)	.47 (.06)	.48 (.05)	.52 (.05)	.62 (.04)
Pairs	.41 (.06)	.48 (.04)	.50 (.07)	.50 (.08)	.52 (.06)	.48 (.05)	.53 (.05)	.67 (.04)
Experiment 2								
Young								
Singles	.36 (.06)	.42 (.05)	.51 (.08)	.48 (.07)	.45 (.06)	.39 (.05)	.51 (.05)	.56 (.04)
Pairs	.35 (.06)	.51 (.04)	.54 (.07)	.54 (.07)	.39 (.06)	.57 (.05)	.52 (.04)	.59 (.04)
Older								
Singles	.40 (.07)	.46 (.05)	.48 (.09)	.40 (.08)	.48 (.07)	.57 (.06)	.58 (.05)	.62 (.05)
Pairs	.29 (.07)	.38 (.05)	.53 (.08)	.48 (.08)	.51 (.07)	.52 (.06)	.55 (.05)	.56 (.04)

Standard errors of the means are presented in parentheses. JOL, judgment of learning; CLJ, category-learning judgment

Follow-up tests revealed that resolution for studied exemplars was higher for spaced than for massed study (.64 vs. .53), $F(1, 46) = 6.80$, $\eta_p^2 = .13$, whereas resolution did not differ between study conditions for novel exemplars (.50 vs. .48), $F < 1$. Resolution was higher for studied (.58) than for novel (.49) exemplars, $F(1, 46) = 8.70$, $\eta_p^2 = .16$. No other effects were significant, Fs < 1.81, ps > .18.

Summary of metacognition results In contrast to the findings of Kornell and colleagues (Kornell & Bjork, 2008; Kornell et al., 2010), our results show that participants were sensitive to the effects of spacing on classification of individual exemplars during encoding and at the time of test. However, this sensitivity does not necessarily demonstrate that participants were aware that spacing was a more effective means of studying than was massing. That is, people may not have declarative knowledge that spacing is superior to massing practice, and instead the effects of spacing on the metacognitive judgments may reflect sensitivity to differences in processing or ease of classifying massed and spaced items. Nonetheless, it is interesting that participants seem to have been sensitive to these effects and, in particular, that participants were sensitive to the benefits of spacing at an intermediate level (i.e., the category level). These results suggest that judgments made on questionnaires reflect different bases than do judgments made at the item and category levels (see Dunlosky & Hertzog, 2001). We consider this distinction further in the "General discussion" section.

Experiment 2

The results from Experiment 1 revealed that spacing effects in natural concept learning can be generalized to the classification of birds. The results from presenting birds in pairs versus singles provide support for the notion that differentiation among families played a role in the effects of spacing on natural concept learning. Furthermore, the effects of relative position observed in classification of studied exemplars suggest that attention also played a role in producing spacing effects.

Experiment 2 was identical to Experiment 1, with the exception that study time was self-allocated. Self-paced study was used to provide an index of the functional attention paid to exemplars in each position of each block. In accord with the attention attenuation hypothesis, we expected study time to decrease across exemplars in the massed blocks, whereas spaced blocks should exhibit no change across exemplars.

We were also interested in whether age differences exist in the benefits of spacing, overall learning, and in the accuracy of metacognitive monitoring (see Kornell et al., 2010). Given that Kornell et al. showed that older adults benefited from spacing in their learning of natural concepts, we expected a similar pattern of results. Experiment 2 included the standard measures of metacognition used in Experiment 1. This allowed for examination of potential age differences in monitoring accuracy in the context of concept learning. Finally, it was an open question as to whether allowing older adults to allocate their study time would eliminate age-related differences in classification performance.

Method

Participants

Sixty-two older adults (mean age = 69.8 years) were recruited using a newspaper advertisement from the greater Akron (OH) area, and 76 young adults (mean age =

19.5 years) were recruited from the Kent State University campus. Participants were randomly assigned to the singles group (35 young, 30 older) and to the pairs group (41 young, 32 older). Both young ($15) and older ($25) adults were paid for participating. Vocabulary scores were higher for older (.68) than for young (.45) adults, $t(136) = 9.5$. Self-reported health (1 = *excellent* to 5 = *poor*) indicated that both older ($M = 2.2$) and young ($M = 1.9$) adults were in relatively good health.

Design, materials, and procedure

The design, materials, and procedure were identical to those in Experiment 1, except that participants' study was self-paced. During study, items were presented until participants pressed a computer key to indicate that they were finished. The JOL prompt was presented immediately thereafter.

Results and discussion

Before considering classification performance results, it is important to note that young adults spent less time studying in Experiment 2 than in Experiment 1 (singles, 4.78 vs. 8 s; pairs, 8.68 vs. 16 s), and older adults spent more time studying in Experiment 2 than did young adults (singles, 11.11 vs. 4.78 s; pairs, 19.71 vs. 8.68 s). Thus, differences in patterns of classification performance between Experiments 1 and 2 cannot be considered a failure to replicate, because the dramatic differences in study time are likely to have moderating effects. We return to this issue later when we describe the relationship between study time and classification performance.

Classification performance

Novel exemplars Classification performance for novel exemplars (Table 1, bottom section) revealed a spacing effect for both young and older adults; classification performance was higher for spaced (.38) than for massed (.33) study, $F(1, 134) = 12.46$, $\eta_p^2 = .09$. No other effects were significant, Fs < 3.72, ps > .05. These results provide further evidence that spacing is an effective means of enhancing natural concept learning for both young and older adults.

Studied exemplars As was found for novel exemplars, spacing enhanced classification of studied exemplars for both young and older adults (Table 1, bottom section). Classification was better for spaced (.38) than for massed (.33) exemplars, $F(1, 134) = 18.36$, $\eta_p^2 = .12$. No other effects were significant, Fs < 2.34, ps > .11. Although the benefits of pairing found in Experiment 1 were not found in Experiment 2, there was a trend indicating that the spacing effects observed for older adults were larger for pairs than for singles. Another difference between Experiments 1 and 2 was that there were significant spacing effects in the singles groups in Experiment 2, whereas the same was not true in Experiment 1. It is likely that these differences in performance were due to differences in study time. We return to this issue in the "General discussion" section.

Given that the interpretation of classification performance as a function of relative position is complicated by self-paced study, we do not report those analyses here. Instead, we examine self-allocated study time as a function of relative position in the next section to more directly test the attention attenuation hypothesis. The lack of an effect of age on classification of studied and novel exemplars suggests that older adults were able to monitor their learning in ways that allowed them to eliminate any age differences in memory. We examine the accuracy of young and older adults' metacognitive monitoring and control in the following sections.

Self-allocated study time

As is shown in Table 4, more time was spent studying spaced (11.54 s) than massed (10.59 s) exemplars, $F(1, 134) = 26.45$, $\eta_p^2 = .17$. In addition, more time was spent studying pairs (14.19 s) than singles (7.94 s), $F(1, 134) = 26.14$, $\eta_p^2 = .16$. Also, older adults studied for more than twice as long as young adults (15.41 vs. 6.73 s), $F(1, 134) = 50.43$, $\eta_p^2 = .27$. It is interesting that older adults spent more time studying than did young adults. Older adults typically show slower processing speed, as compared with young adults (e.g., Salthouse, 1996), and there are typically no age differences in metacognitive monitoring. Thus, older adults may have spent more time studying than did young adults because the subjective experience of completing learning did not occur

Table 4 Study time allocation (in seconds) as a function of age, presentation, and study condition: Experiment 2

	Study Condition	
	Massed	Spaced
Young		
Singles	4.60 (.48)	4.95 (.59)
Pairs	8.01 (.75)	9.34 (.88)
Older		
Singles	10.82 (1.13)	11.39 (1.31)
Pairs	18.92 (2.08)	20.49 (2.05)

Standard errors of the means are presented in parentheses

until later. Finally, the difference in study time between spaced and massed exemplars was larger for pairs (14.91 vs. 13.47 s) than for singles (8.17 vs. 7.71 s), $F(1, 134) = 7.07$, $\eta_p^2 = .05$. No other effects were significant, Fs < 3.69, ps > .05.

More germane to the attention attenuation hypothesis, study time was examined as a function of relative position for each study condition. In these analyses, study time was examined for the six presentations in the singles group (Fig. 3) and the three presentations in the pairs group (Fig. 4). Examination of Figs. 3 and 4 reveals patterns of results similar to classification of studied exemplars in Experiment 1. The general pattern for the massed conditions in the singles groups is that both young and older adults studied the first exemplar longest, with study time decreasing most rapidly from the first to the second position. After the initial sharp decline, there was a trend indicating that study time continued to decrease at a slower rate thereafter. The pattern for the massed conditions was similar in the pairs group; however, the decline from the first to the second position was not larger than that from the second to the third position. In contrast, study time in the spaced blocks remained relatively constant across positions

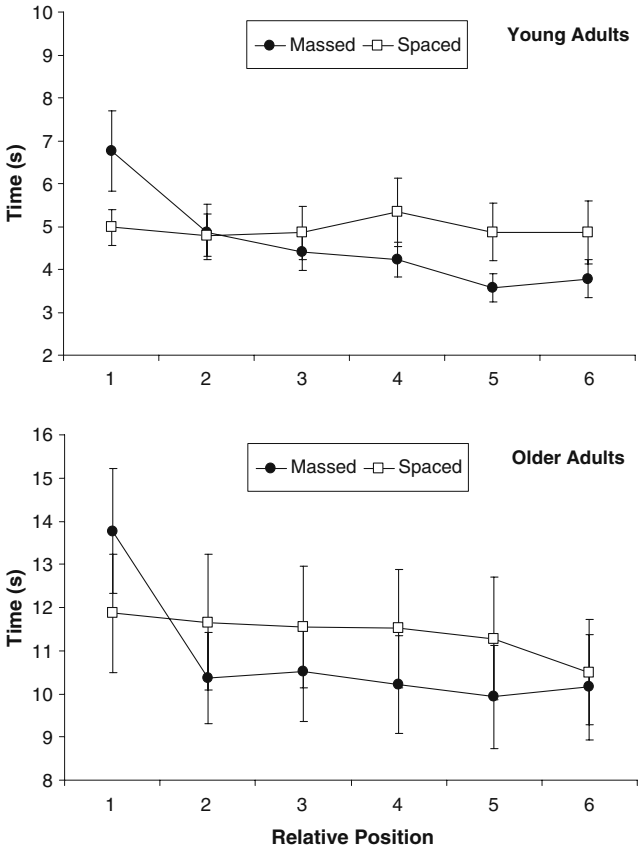

Fig. 3 Mean study time as a function of age group, study condition, and relative position in the singles group. The bars represent standard errors of the means

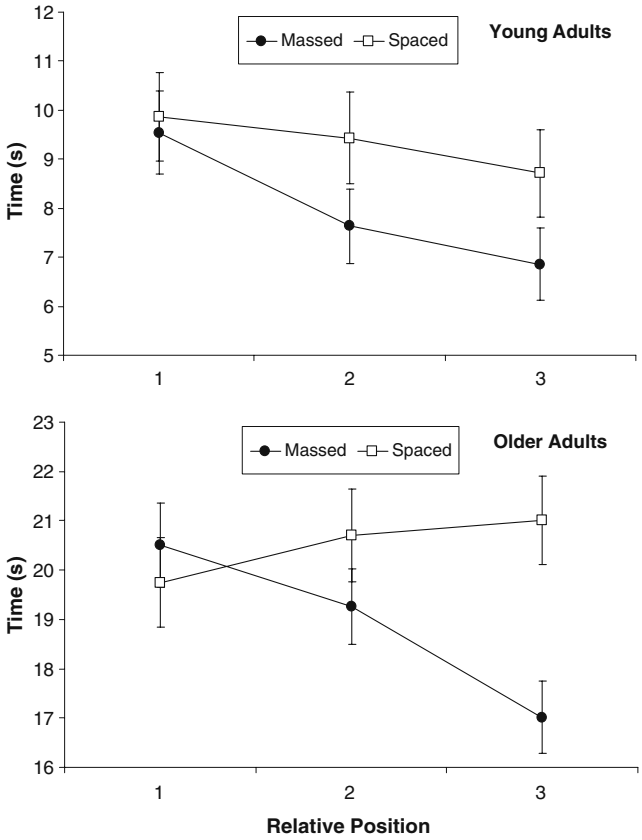

Fig. 4 Mean study time as a function of age group, study condition, and relative position in the pairs group. The bars represent standard errors of the means

in both presentation groups. These effects were confirmed by study × position interactions in the singles, $F(5, 315) = 6.58$, $\eta_p^2 = .10$, and pairs, $F(2, 142) = 8.21$, $\eta_p^2 = .10$, groups. The most striking difference in position curves for massed and spaced families is in the first two positions within a block. This difference is understandable in terms of a difference in the novelty of exemplars from a family. For massed blocks, the first exemplar in a block was always the first encounter with an exemplar from that particular family and, so, may have been given additional study because of its novelty. In contrast, the first exemplar from each family was presented in the first block of spaced study, and, so, exemplars in those blocks did not differ in their novelty.

Given that the attention attenuation hypothesis predicts diminishing attention across positions in the massed condition, follow-up t-tests were first conducted on adjacent positions for each age group in each study condition. To anticipate, few differences were found between adjacent positions following the comparison of the first and second positions. Because the differences following the second position were numerically small, the extent to which attention continued to diminish was then examined by comparing study time in the second position with that in later positions.

For young adults in the singles group, results revealed significant differences in the massed condition between positions 1 and 2 and between positions 4 and 5, *t*s(34) > 2.71. There were also significant differences between positions 2 and 5 and between positions 2 and 6, *t*s(34) > 3.41. There were no significant differences in the spaced condition. Older adults in the singles group showed differences in the massed condition between positions 1 and 2, *t*(29) = 4.19, and no differences in the spaced condition. Young adults in the pairs group showed significant differences in the massed condition between positions 1 and 2, 2 and 3, and 1 and 3, *t*s(40) > 2.05. They also showed a difference in the spaced condition between positions 1 and 3, *t*(40) = 3.09. Older adults in the pairs group showed differences in the massed condition between positions 1 and 3 and between positions 2 and 3, *t*s(31) > 2.79, and no differences in the spaced condition. Together, these results suggest that attention in the massed condition diminished most rapidly from the first to the second position and tended to diminish beyond the second position, but at a slower rate. These results are consistent with the prediction from the attention attenuation hypothesis.

As was mentioned earlier, young adults spent much less time studying in Experiment 2. It is noteworthy that this difference in study time did not dramatically change overall classification performance (Experiment 1 = .41 vs. Experiment 2 = .36). These results suggest that allowing self-allocated study allowed more efficient learning. However, there was still a tendency for performance to be lower in Experiment 2, suggesting that participants did not take full advantage of the opportunity to self-regulate their learning. Perhaps young adults were sensitive to the diminishing returns of additional study time and, consequently, terminated their study before learning the materials completely (see Nelson & Leonesio, 1988).

Classification performance and study time

To examine the link between study time and performance, we correlated study time and classification performance, separately as a function of exemplar (novel or studied) and presentation format (spaced vs. massed practice). As was expected, study time was positively correlated with classification of novel (massed, r = .15; spaced, r = .23) and studied (massed, r = .23; spaced, r = .14) exemplars, ps < .05. A one-tailed test was used to test this difference, because self-paced study time and performance are often positively related. Given these results, we included study times (for massed and spaced items) as covariates in an ANOVA that reexamined the influence of spacing on the classification of novel and studied items. In both cases, the influence of spacing was not significant, Fs < 1.50, when study time was a covariate. These outcomes further establish the link between study time and classification performance and provide supporting evidence for the attention attenuation hypothesis.

Metacognitive judgments

Metacognitive judgments were examined in the same manner as in Experiment 1.

Judgments of learning JOLs were sensitive to the effects of spacing for older adults, but not for young adults (Table 2, bottom section). JOLs were higher for spaced than for massed study (.44 vs. .43), $F(1, 134) = 6.59$, $\eta_p^2 = .05$. The interaction with age was significant, $F(1, 134) = 12.77$, $\eta_p^2 = .09$, showing that older adults' JOLs were higher for spaced (.44) than for massed (.40) study, $t(61) = 3.93$, whereas young adults' JOLs did not differ (.45 vs. .45), $t < 1$. Note that although young adults' JOLs were not sensitive to the benefits of spacing on classification performance, they did not indicate that massed exemplars were learned better than spaced exemplars (see Kornell & Bjork, 2008). No other effects were significant, Fs < 3.65, ps > .06.

Resolution for JOLs could not be computed for 7 participants (2 young, 5 older), due to constant values on one variable in at least one study condition (i.e., zero or perfect classification performance, or the same JOL for every item). This explains why the degrees of freedom in the following analyses differ from those in previous analyses. As was found in Experiment 1, resolution was enhanced by spacing (Table 3, bottom section). Gammas were higher for spaced (.44) than for massed (.35) study, $F(1, 127) = 7.16$, $\eta_p^2 = .05$. No other effects were significant, Fs < 2.17, ps > .14.

Category-learning judgments CLJs were also sensitive to the effects of spacing (Table 2, bottom section). CLJs were higher for spaced (.41) than for massed (.38) study, $F(1, 134) = 18.95$, $\eta_p^2 = .12$. No other effects were significant, Fs < 3.38, ps > .07.

Resolution for CLJs (Table 3, bottom section) could not be computed for 13 participants (4 young, 9 older), due to constant values on at least one variable. No effects were significant, Fs < 1. However, the grand mean of the correlations was greater than zero (.49), indicating that participants were sensitive to differences in classification difficulty across categories.

Confidence judgments Confidence judgments were also sensitive to the effects of spacing (Table 2, bottom section). Confidence was higher for spaced (.42) than for massed (.41) exemplars, $F(1, 134) = 7.92$, $\eta_p^2 = .06$. Confidence was also higher for studied (.43) than for novel (.40) exemplars, $F(1, 134) = 20.32$, $\eta_p^2 = .13$. This effect was qualified by an exemplar × presentation interaction, $F(1, 134) = 9.16$,

η_p^2 = .06, indicating that this difference was significant for pairs (.47 vs. .41), $F(1, 72) = 30.39$, η_p^2 = .30, but not for singles (.40 vs. .38), $F < 1$. No other effects were significant, Fs < 2.17, ps > .08.

Resolution for confidence judgments (Table 3, bottom section) could not be computed for 10 participants (3 young, 7 older) due to constant values on at least one variable. Resolution was better for studied (.56) than for novel (.49) exemplars, $F(1, 124) = 9.36$, η_p^2 = .07. No other effects were significant, Fs < 2.48, ps > .07.

In sum, the results from the metacognitive measures in Experiment 2 largely replicated the effects found in Experiment 1. That is, participants' metacognitive judgments were largely sensitive to the increased ease of classification produced by spacing exemplars (although these effects were relatively small in magnitude), as well as differences in classification difficulty across categories. Also, young and older adults did not differ in their overall judgment accuracy.

General discussion

The results from the present experiments extend the facilitative effects of spacing on the learning of natural concepts to the learning of bird families. Further examination of spacing effects revealed that they were accomplished in two ways. First, juxtaposing exemplars from different families enhanced concept learning by highlighting differences between exemplars. Second, less functional study time was devoted to exemplars during study when exemplars from the same family were presented consecutively than when they were intermixed with exemplars from different families. Metacognitive measures revealed sensitivity to differences in classification difficulty between massed and spaced study. Judgments were also sensitive to differences in classification difficulty across categories. Finally, older adults were able to learn to classify birds as well as young adults when allowed to allocate their study time.

Evaluating theories of the spacing effect

As was described in the introduction, one goal of our experiments was to further evaluate two accounts of the spacing effect in natural concept learning. The discrimination account held by Kornell and colleagues (Kornell & Bjork, 2008; Kornell et al., 2010) emphasizes the importance of comparing exemplars from different categories. In particular, noticing similarities and differences among exemplars of different categories presumably aids people in distinguishing among those categories at the time of test (Kornell et al., 2010). In our experiments, exemplars were presented in pairs to further highlight the differences between categories by reducing the memory demands associated with comparing individually presented exemplars. In accord with the discrimination account, the results from Experiment 1 showed that the positive effects of spacing were enhanced by presenting exemplars in pairs.

To further understand the way in which spacing effects were accomplished, we explored the possibility that differences in attention also contributed to these effects. In particular, we evaluated the attention attenuation hypothesis, which holds that attention decreases across subsequent exemplars in massed blocks but does not do so in spaced blocks. Kornell et al. (2010) examined this possibility by comparing the magnitude of spacing effects in their induction and repetition conditions. The critical difference between these conditions was in the content of the massed blocks. In the induction condition, various exemplars from the same family were presented, whereas in the repetition condition, the same exemplar was repeatedly presented. They argued that it should be more difficult to sustain attention for repeated presentations of an exemplar (repetition condition) than for different exemplars from the same family (induction condition). Consequently, the attention attenuation hypothesis would predict a larger spacing effect in the repetition than in the induction condition. Results revealed no support for this account, in that the magnitude of spacing effects did not differ in the induction and repetition conditions.

Although Kornell et al. (2010) did not find support for the attention attenuation hypothesis, their evaluation of it was ad hoc. In the present experiments, we directly tested it by examining classification performance for studied exemplars and time spent studying those exemplars in each relative position. According to the attention attenuation hypothesis, both classification performance and study time should decrease across positions in the massed blocks, with little, if any, change in these measures arising in the spaced blocks. The results from the present experiments were consistent with these predictions, and across participants, study time was significantly related to classification. Thus, we can conclude that both attention and discrimination played a role in producing spacing effects in the learning of bird families.

Finally, in Experiment 2, the spacing effect on younger adults' classification performance was not enhanced by pairing birds during study. Why do the classification results differ in Experiments 1 and 2? Although speculative, younger adults in Experiment 2 may have failed to benefit from presentation of pairs because they did not persist in studying the pairs long enough to fully process the differences between the birds within a pair. Thus, one possibility is that when people are encouraged to study

longer, as in Experiment 1, the benefits of pairing become evident.

Metacognition, spacing, and category-learning judgments

Another goal of our experiments was to examine the sensitivity of participants' metacognitive judgments to the effects of spacing and to differences in classification difficulty across families. Kornell and colleagues (Kornell & Bjork, 2008; Kornell et al., 2010) found that the majority of people reported that massed study produced better classification performance than did spaced study on posttest questionnaires. In contrast, the results from the present experiments revealed that participants' item- and category-level judgments were sensitive to the effects of spacing. The important difference between the present experiments and those conducted by Kornell and colleagues is that most of our metacognitive measures were concurrent reports (e.g., JOLs), whereas they used retrospective reports. The accuracy of these reports likely differs because concurrent reports reflect the current processing of items, whereas the accuracy of retrospective reports can be impaired by forgetting or by task beliefs (see Dunlosky & Hertzog, 2001).

Given the differences between the measures used in our study and those used by Kornell and colleagues (Kornell & Bjork, 2008; Kornell et al., 2010), it was not surprising that item-level judgments revealed sensitivity to spacing effects, whereas global retrospective reports did not. That is, if encoding benefits from the juxtaposition of exemplars in the spaced conditions, and if more attention is being paid to those exemplars, it seems reasonable that participants would be sensitive to the processing advantage of spaced, relative to massed, exemplars. However, it is interesting to note that the spacing advantage in JOLs was small in both experiments. One possibility is that for some items, participants' judgments reflected the fluency-based illusion that massed exemplars were learned more easily, whereas for other items, judgments reflected the superior processing of spaced exemplars. Reliance on both of these bases would produce opposing effects on JOL magnitudes. Consistent with this possibility, confidence judgments, which do not suffer from this problem, generally showed a larger advantage for spaced presentation than did JOLs. Spaced presentation increased both the magnitude and the resolution of confidence, particularly for studied exemplars.

Perhaps more interesting, CLJs were also sensitive to the effects of spacing. These results are interesting in that CLJs are intermediate-level judgments that fall between item- and global- (e.g., retrospective reports) level judgments. Although CLJs are global judgments of a sort, it is possible that they were not susceptible to the fluency-based illusion that massed study was better than spaced study, because requiring participants to make judgments about categories might bring to mind processing differences for the items in each group. It is also possible that CLJs tap into participants' memory for individual exemplars, which are presumably better for spaced than those for massed study. Thus, aggregate judgments might be less susceptible to metacognitive illusions when the queried dimensions draw upon the processing of, or memory for, individual items within categories.

CLJs were also sensitive to differences in classification difficulty across families, as evidenced by strong correlations between CLJs and classification performance at the category level. These results replicate the results in Jacoby et al. (2010), who interpreted these correlations as indicating that participants could use memory for both specific exemplars and their general characteristics as bases for relatively accurate predictions of classification performance on novel exemplars from those categories. These findings have practical import for the learning of concepts in educational settings. For example, students' ability to identify which topics are better learned than others (e.g., memory vs. attention in a cognitive psychology class) will determine the manner in which they approach their studying. Examination of students' sensitivity to topics in a classroom setting and their study choices can also inform educators' guidance of student study strategies.

Aging and the strategic regulation of natural concept learning

Kornell et al. (2010) hypothesized that older adults would benefit less from spacing than would young adults as a result of older adults being more forgetful of previously presented instances in the spaced condition. Contrary to this prediction, older adults demonstrated spacing effects that were similar in magnitude to those of young adults. The only age difference that they observed was a main effect indicating that young adults outperformed older adults. The present experiments were designed to further verify this lack of an age difference in the effects of spacing and to evaluate whether overall age differences could be eliminated by allowing participants to control their study time. The results revealed that both young and older adults showed positive effects of spacing, and no difference in overall performance was evident between the groups. The latter result is surprising in that older adults have been shown to have a deficit in metacognitive control strategies, relative to young adults, and these deficits result in lower levels of associative learning (see Dunlosky & Connor, 1997).

Although the mechanism by which older adults were able to perform as well as young adults is unclear, one possibility is that older adults may have been more motivated to self-initiate successful learning strategies

because of their intrinsic interest in the materials (see Castel, 2005, 2008). In line with this suggestion, older adults may have made use of strategies that maximized their ability to remember general information. According to previous accounts, older adults' memory for specific information is thought to be impaired, relative to young adults, whereas both groups are thought to have similar abilities in remembering general information (e.g., Adams, 1991; Castel, Farb, & Craik, 2007; Craik, 2002; Koutstaal, 2003). Perhaps young and older adults accomplished classification in different ways. Older adults may have focused on general characteristics during study by thinking back to previous exemplars, resulting in increased study time. In contrast, young adults may have focused more on specific features of exemplars than did older adults.

Yet another possibility is that the lack of age differences was due to older adults having more prior knowledge of birds than did young adults. Against this possibility, confidence judgments tended to be higher for young than for older adults, whereas one would expect the reverse to be true if older adults had more background knowledge. Moreover, we included a brief questionnaire in Experiment 2 that asked participants about their involvement with bird clubs (no participants reported belonging to one), and they rated their expertise in identifying birds (1 = *entirely incapable*, 4 = *novice*, 7 = *expert*). On average, both young (M = 2.4) and older (M = 3.0) adults rated themselves below the novice level, and although older adults rated themselves significantly higher, t (134) = 2.70, when level of expertise was included in analysis of classification performance (i.e., as a covariate), it did not influence the results. Thus, differential expertise cannot account for the present lack of age-related differences in classification performance.

Concluding comments

The findings presented here indicate that spacing enhances the learning of bird families. In addition to enhancing associative learning (e.g., Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Dempster, 1996), spacing has now been shown to enhance the learning of natural concepts in several studies (e.g., Kornell & Bjork, 2008; Kornell et al., 2010). A major goal of our ongoing research has been to discover ways in which to optimize the learning of bird families. So far, we have demonstrated that testing (Jacoby et al., 2010) and spacing are effective means by which to do so. Taken together, these findings suggest that well-known methods of enhancing associative learning can be extended to the learning of natural concepts. Further exploration along these lines will be important for informing theories of concept learning and for practical applications such as the learning of natural concepts in educational settings and the preservation of cognitive function in older adults.

Acknowledgements This research was supported by the James S. McDonnell Foundation 21st Century Science Initiative in Bridging Brain, Mind and Behavior Collaborative Award. We thank Sarah Arnspiger and Rachel Teune for their assistance with data collection. Correspondence concerning this article should be addressed to Christopher N. Wahlheim, Department of Psychology, Washington University, St. Louis, MO 63130. E-mail: cnwahlheim@gmail.com.

References

Adams, C. (1991). Qualitative age differences in memory for text: A lifespan developmental perspective. *Psychology and Aging, 6*, 323–336.

Castel, A. D. (2005). Memory for grocery prices in younger and older adults: The role of schematic support. *Psychology and Aging, 20*, 718–721.

Castel, A. D. (2008). The adaptive and strategic use of memory by older adults: Evaluative processing and value-directed remembering. In A. S. Benjamin & B. H. Ross (Eds.), *The psychology of learning and motivation* (Vol. 48, pp. 225–270). London: Academic Press.

Castel, A. D., Farb, N., & Craik, F. I. M. (2007). Memory for general and specific value information in younger and older adults: Measuring the limits of strategic control. *Memory & Cognition, 35*, 689–700.

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.

Craik, F. I. M. (2002). Human memory and aging. In L. Bäckman & C. von Hofsten (Eds.), *Psychology at the turn of the millennium* (pp. 261–280). Hove, U.K.: Psychology Press.

Dempster, F. N. (1996). Distributing and managing the conditions of encoding and practice. In E. L. Bjork & R. A. Bjork (Eds.), *Memory* (pp. 317–344). San Diego: Academic Press.

Dunlosky, J., & Connor, L. T. (1997). Age differences in the allocation of study time account for age differences in memory performance. *Memory & Cognition, 25*, 691–700.

Dunlosky, J., & Hertzog, C. (2000). Updating knowledge about strategy effectiveness: A componential analysis of learning about strategy effectiveness from task experience. *Psychology & Aging, 15*, 462–474.

Dunlosky, J., & Hertzog, C. (2001). Measuring strategy production during associative learning: The relative utility of concurrent versus retrospective reports. *Memory & Cognition, 29*, 247–253.

Gagné, 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.

Griego, J. A., & Kliegel, M. (2008). Adult age differences in function concept learning. *Aging, Neuropsychology, & Cognition, 15*, 1–30.

Hess, T. M., & Slaughter, S. J. (1986). Aging effects on prototype abstraction and concept identification. *Journal of Gerontology, 41*, 214–221.

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, & Cognition, 36*, 1441–1451.

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? *Psychological Science, 19*, 585–592.

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.

Koutstaal, W. (2003). Older adults encode—but do not always use—perceptual details: Intentional versus unintentional effects of

detail on memory judgments. *Psychological Science, 14*, 189–193.

Kurtz, K. H., & Hovland, C. I. (1956). Concept learning with differing sequences of instances. *Journal of Experimental Psychology, 51*, 239–243.

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95*, 109–133.

Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the "labor-in-vain effect". *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 676–686.

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review, 103*, 403–428.

Sibley, D. A. (2002). *Sibley's birding basics*. Toledo, Spain: Knopf.