Title
The Interactions of Category Structure and Supervision in Category Learning: a comparative approach

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 36(36)

ISSN
1069-7977

Authors
Yim, Hyungwook
Castro, Leyre
Wasserman, Edward
et al.

Publication Date
2014

Peer reviewed

The Interactions of Category Structure and Supervision in Category Learning : a comparative approach

Hyungwook Yim (yim.31@osu.edu)
Department of Psychology, The Ohio State University
1835 Neil Avenue, Columbus, OH 43210, USA

Leyre Castro (leyre-castroruiz@uiowa.edu)
Edward A. Wasserman (ed-wasserman@uiowa.edu)
Department of Psychology, The University of Iowa
E11 Seashore Hall, Iowa City, IA 52242, USA

Vladimir M. Sloutsky (sloutsky.1@osu.edu)
Department of Psychology, The Ohio State University
1835 Neil Avenue, Columbus, OH 43210, USA

Abstract

The current study investigated the interactions among category structure, supervision, and the ability to selectively attend during category learning. Specifically, we compared pigeons' with human adults' category learning using artificial categories to examine the role of selective attention in category learning. Results showed that pigeons benefit more from supervision, and unlike human adults, the benefit is stronger for sparse categories. Moreover, whereas supervision did not affect human adults' generalization performance, low-supervision resulted in lower generalization for pigeons. The results were discussed in terms of the difference in utilizing the supervisory signal, and the ability to selectively attend to category relevant information.

Keywords: category learning; selective attention; supervision; category structure; comparative study.

Introduction

Selective attention is one of the crucial components in category learning (Shepard, Hovland, & Jenkins, 1961). Importantly, selective attention supports efficient category learning and generalization since it enables one to focus on category-relevant information while ignoring category-irrelevant information (Mackintosh, 1965). For example, when learning how to distinguish halibuts from flounders, the shape of the tail is a relevant dimension. Once one learns to focus on the shape of the tail and ignore other parts of the fish, distinguishing the two fishes becomes efficient when encountering new instances. Moreover, since selective attention undergoes marked development due to the maturation of the prefrontal cortex (PFC) (Diamond, 2002; Hanania & Smith, 2010), selective attention has been considered to be one of the main factors in explaining the development of category learning (Sloutsky, 2010).

Along with selective attention, the structure of the category is another important factor that shapes the process of category learning. Among many, one way to view category structure is by its density (Gentner, 1981; Kloos & Sloutsky, 2008). Specifically, Kloos & Sloutsky (2008) proposed a formal measurement for category density using entropy values. Category density was defined by ”the ratio of variance relevant for category membership to the total variance across members and nonmembers of the category”. Therefore, dense categories have low within variability and high between variability, whereas sparse categories have low between variability and high within variability. For example, dogs and cats are dense categories since they have many overlapping features within their category, but are also distinguishable from other categories. Due to the redundancy of the category relevant information it is known that even infants are able to learn dense categories (e.g., Quinn, Eimas, & Rosenkrantz, 1993). On the other hand, chest x-rays of different kinds of diseases could be considered as sparse categories. For example, it is hard to distinguish patient who have pneumonia from those who do not or even from patients who have other diseases. This is due to the fact that while many x-rays look alike, only a few features are relevant in diagnosing a certain disease. Also, having only little category relevant information makes category learning hard, and requires extensive training (e.g., Palmeri, Wong, & Gauthier, 2004).

Naturally, learning sparse categories requires selective attention, whereas dense categories could be learnable without selective attention. From this point of view, evidence has shown that 3- to 4-month-olds do not selectively attend to category relevant dimensions when learning dense categories (Best, Yim, & Sloutsky, 2013), and adults could learn dense categories using distributed attention (Best, Robinson, & Sloutsky, 2014).

Finally, supervision also interacts with the above two components during category learning, where supervision aids selective attention to find the category-relevant information. The distinction between unsupervised and supervised learning in category learning is well recognized (Love, 2002), and different underlying mechanisms have been also proposed (Love, Medin, & Gureckis, 2004). On the other hand, the interaction among the amount of supervision, category structure, and the ability to selectively attend is not extensively studied. Some evidence has shown that sparse categories benefit from supervision more than dense categories (Kloos & Sloutsky, 2008). It was argued that the use of supervisory sig-

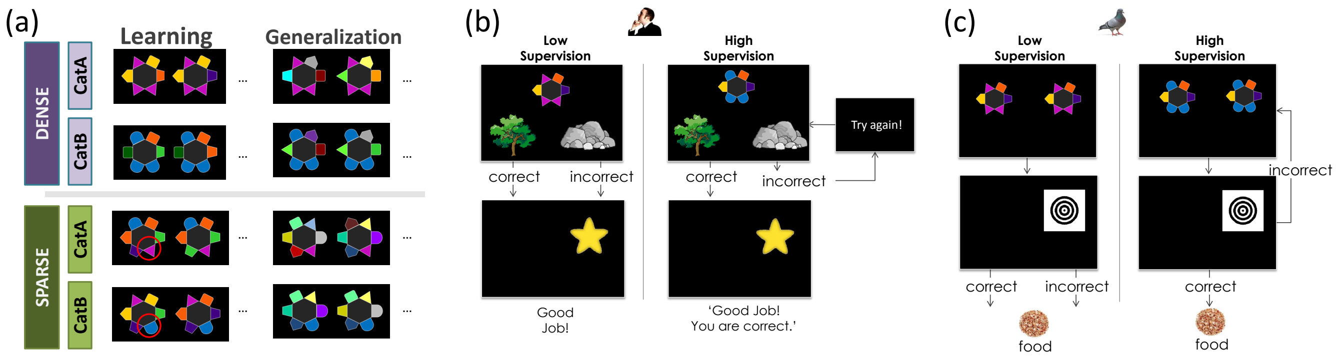

Figure 1: Stimuli structure and experimental design used in Experiment 1 and 2. (a) Dense and sparse stimuli structure used in the learning phase and generalization phase. (*note that the red circle depicted in the sparse categories was invisible to the subjects*). (b) Experimental design in Experiment 1. (c) Experimental design in Experiment 2.

nals were mediated by selective attention and the related brain areas such as the PFC. Therefore, sparse categories, which require selective attention, would benefit more than dense categories.

In the current study, we further investigated this view by manipulating (1) different kinds of category structure, (2) different amount of supervision, and (3) different ability in managing selective attention during category learning, which was achieved by cross-species comparisons. Since selective attention is closely linked to the PFC, we selected pigeons - species that are known to not have a comparable PFC to humans. Therefore, the cross-species comparison will show us the difference in the category learning due to the functionality of the PFC (or the ability to selectively attend).

Experiment 1

Experiment 1 examined the benefit of supervision during category learning in human adults. The effect of category density was manipulated by stimuli (i.e., Dense/Sparse), whereas the amount of supervision was manipulated through the learning procedure (i.e., High-supervised/Low-supervised).

Methods

Subjects Ninety six undergraduate students at The Ohio State University participated for course credit (56 females, M = 19.63 years, SD = 2.00). Each participant was randomly assigned to one of the four conditions, resulting in 24 subjects for each condition. Two additional participants from the Low-supervised Dense condition was excluded due to not understanding the instructions correctly.

Materials Flower-like artificial categories were used in the experiment (see Figure 1a). For the learning phase, there were two sets of categories (i.e., Category A, and Category B) for each density condition (i.e., Dense, and Sparse). Each exemplar had a gray hexagon in the middle with six colored shapes on every side. For Dense categories three out of six changed their color and shape, whereas three were constant in their shape, color, and location, which served as category relevant dimensions. For example, Category A in the Dense condition had three pink triangles that where constant throughout the exemplars, whereas Category B had three blue semicircles in the same location. For Sparse categories, five out of six dimensions changed their color and shape, whereas one was constant. For example, one category in the Sparse condition (i.e., Category A) had a pink triangle whereas another category (i.e., Category B) had a blue semi-circle. Stimuli for the generalization phase were identical to that of the learning phase. However, irrelevant dimensions changed their shape and color from the learning phase. Each category set in the learning and generalization phase had 18 unique exemplars.

Procedure The subjects were told that they are going to play a game where a star is hidden under the tree or the rocks. They were also told that the location of the star will depend on the artificial flower that will be presented in the middle of the screen, and that they should look at the flower carefully to find the star. There were 12 blocks, where each block contained 12 learning trials. Among the 12 trials, half contained exemplars from Category A, and the other 6 trials contained exemplars from Category B. The exemplars were presented in a random order.

On each trial, an exemplar was presented on the top center of the computer screen, and a picture of a tree and a rock was each presented on the bottom left and right side of the screen, which was randomized by subject (see Figure 2b). On every trial, subjects used the mouse to click on the picture of a tree or rock to indicate their decision about where the star is hidden. In the Low-supervised condition, when the subject made a decision, the star appeared on the correct side, which was counterbalanced for Category A and B for each subject, for 2000 msec regardless of their response accuracy. Also a "Good Job!" phrase was simultaneously presented on the bottom of the screen. In the High-supervised condition, the star was only presented when the subject made a correct response. If the subject made an incorrect response, a "Try again!" phrase was presented on the center of the screen for 2000 msec, and the subjects saw the same exemplar again.

When a subject had more than 10 trials correct in a block (i.e., 83% accuracy), a generalization phase with 6 trials was introduced before proceeding to the next block. The generalization trials were identical to the learning trials except: (1) stimuli from the generalization set were used, and (2) a blank screen was presented for 2000 msec after the subject's responses, which prevented subjects from getting any direct feedback. The experiment was terminated for each subject either when the subject responded to all 12 blocks, or when the subject's responses were perfect for 2 consecutive blocks. The experiments were written in MatLab, using Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997).

Results

We first classified learners and non-learners by setting a criterion of .85 accuracy in a block. Within 12 blocks, subjects who reached .85 accuracy and showed an asymptotic performance after reaching the criterion were considered as learners. Overall, there were more learners in the Dense conditions, and in the High-supervised conditions (see Table 1). A binary logistic regression with Supervision and Density as covariates show that there was a statistically significant effect for Supervision (*Wald* = 7.76, $p < .005$), and Density (*Wald* = 5.77, $p < .05$). The non-learners were excluded from all following analysis.

Table 1: Proportion of learners in Experiment 1.

	Dense Category	Sparse Category
High-Supervised	1.0	.83
Low-Supervised	.87	.67

Trials-to-criterion (TTC) was calculated by counting the number of trials to reach the criterion (.85) for each subject. The High-supervised Dense condition showed the fastest learning (M = 22.50, SD = .99), followed by the Low-supervised Dense condition (M = 34.80, SD = 2.49), High-supervised Sparse condition (M = 47.43, SD = 1.94), and the Low-supervised Sparse condition (M = 75.43, SD = 2.70) (see Figure 2a). The TTC data was submitted to a 2 $\times$ 2 (Supervision $\times$ Density) between-subjects ANOVA. Results showed a statistically significant main effect of Supervision (F = 12.94, $p < .001$), and Density (F = 34.24, $p < .001$), but no interaction (F = 1.96, $p = .17$). By comparing the ratio of the High-supervised condition to the Low-supervised condition, the benefit of supervision could be calculated for the Dense and Sparse conditions. The calculations showed that the Sparse condition benefited slightly more (1.59 = 75.43 trials / 47.43 trials) than the Dense condition (1.54 = 34.8 trials / 22.5 trials).

Finally, generalization trials were analyzed. The average of the first 2 generalization trials after each subject reached the criterion (.85) was used in the analysis. The two Dense conditions showed similar accuracy on the generalization trials with both conditions having an average of .97 (High-supervised, SD = .12; Low-supervised, SD = .07). The High-supervised Sparse condition (M = .81, SD = .18) was slightly less accurate than the Low-supervised Sparse condition (M = .89, SD = .18). Conducting a 2 $\times$ 2 (Supervision $\times$ Density) between-subject ANOVA showed that there was a statistically significant main effect for Density (F = 12.86, $p < .001$), but not for Supervision (F = 1.46, $p = .23$) or interaction (F = 1.63, $p = .21$).

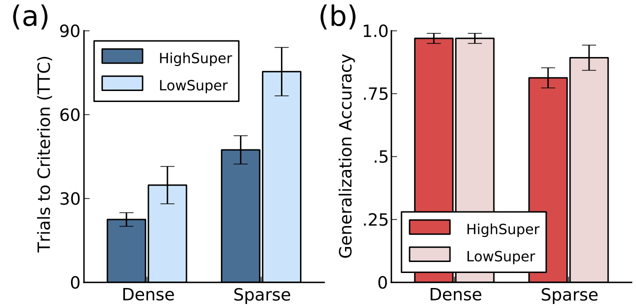

Figure 2: Results from Experiment 1 with humans. (a) Trials-to-criterion, and (b) generalization accuracy. (*error bars represent +/- 1 SEM*)

In sum, results from the number of learners, and TTC showed that Dense conditions are overall easier than Sparse conditions, and that supervision aided learning in both conditions. However, there was no significant interaction between category structure and the amount of supervision. That is, the benefit of supervision did not significantly differ across different densities. Moreover, generalization was only affected by density and not by the amount of supervision.

Experiment 2

In Experiment 2, pigeons were introduced to compare the effect of the ability to selectively attend during category learning. The manipulation of the two other factors (i.e., category density, and amount of supervision) was similar to Experiment 1.

Subjects

The subjects were 16 feral pigeons maintained at 85% of their free-feeding weights by controlled daily feedings. The pigeons were randomly assigned into one of the conditions, resulting in 4 pigeons for each condition. The pigeons had served in unrelated studies prior to the present project.

Materials

The experiment used four 36 $\times$ 36 $\times$ 41 cm operant conditioning chambers detailed by Gibson et al. (2004). The chambers were located in a dark room with continuous white noise. Each chamber was equipped with a 15-in LCD monitor located behind an AccuTouch resistive touchscreen (Elo TouchSystems, Fremont, CA). The portion of the screen that was viewable by the pigeons was 28.5 cm $\times$ 17 cm. Pecks to the touchscreen were processed by a serial controller board outside the box. A rotary dispenser delivered 45-mg pigeon pellets through a vinyl tube into a food cup located in the center of the rear wall opposite the touchscreen. Illumination during the experimental sessions was provided by a houselight mounted on the upper rear wall of the chamber. The pellet dispenser and houselight were controlled by a digital I/O interface board. Each chamber was controlled by a personal computer, and the experiments were developed in MatLab. The category exemplars used in the experiment were identical to those used in Experiment 1.

Procedure

The learning phase consisted of 144 trials each day. Among the 144 trials half contained exemplars from Category A, and the other half contained exemplars from Category B. On each trial, two identical exemplars were presented on the right and left side of the computer screen (see Figure 2c). In the Low-supervised condition, after five responses to either of these two images, a target stimulus (i.e., black concentric rings on a white background) was presented on the correct side of the screen, where the correct side for Category A and Category B was counterbalanced among subjects. After one response to this target stimulus, food reward was delivered. Completing the sequence of category exemplars-target stimulus always ended with food reinforcement, regardless of the specific side that the pigeons chose when the exemplars were presented. Inter-trial intervals were 5000 msec in duration. The High-supervised condition was very similar to the unsupervised procedure. However, the five responses to the category exemplars must be on the correct side for the target stimulus to be presented. If the last peck was on the wrong side, the screen would go black for 1000 msec and the category exemplars were presented again. The pigeons had to complete the 5-pecks response requirement again until the correct response was made. Responses from the correction trials were not analyzed. When a bird reached a learning criterion of .85, generalization phase was introduced the next day for 4 days.

The generalization phase consisted of 162 trials a day which included 144 training trials and 18 testing trials randomly intermixed. The trials presented the generalization stimuli; on these trials, no correction procedure was given (to avoid teaching the birds the correct response) so that pigeons' choices were always reinforced, regardless of the pigeons being in the Low- or High-supervised conditions.

Results

One bird in the Low-supervised sparse condition was excluded from all analysis due to not reaching the learning criterion after 76 days (10944 trials). Trials-to-criterion (TTC) was calculated by counting the number of trials to reach the criterion (.85) for each subject. Overall, birds in the Dense conditions learned the categories faster (M = 1062.00, SD = 1378.92) than those in the Sparse conditions (M = 2777.14, SD = 3262.29), and birds in the High-supervised conditions learned the categories faster (M = 450.00, SD = 179.48) than those in the Low-supervised conditions (M = 3476.57, SD = 2992.32) (see Figure 3). The TTC data was submitted to a 2 $\times$ 2 (Supervision $\times$ Density) between-subject ANOVA. Results showed a statistically significant main effect of Supervision (F = 15.46, $p < .005$), Density (F = 6.16, $p < .05$), and a marginal interaction (F = 4.36, p = .061). The benefit of supervision was also calculated as in Experiment 1. The benefit in the Dense condition was 6.38 (= 1836 trials / 288 trials), and 9.25 (= 5664 trials / 612 trials) in the Sparse condition, which indicated a greater benefit in the Sparse condition.

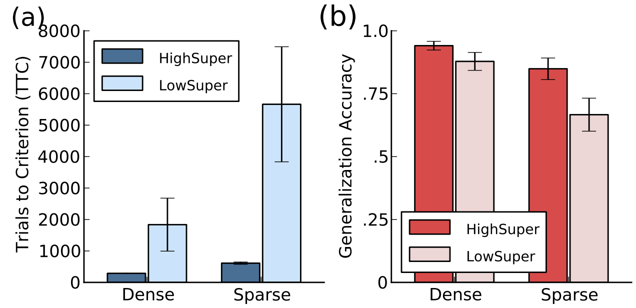

Figure 3: Results from Experiment 2 with pigeons. (a) Trials-to-criterion, and (b) generalization accuracy. (*error bars represent +/- 1 SEM*)

In the generalization trials, the High-supervised Dense condition showed the highest accuracy (M = .94, SD = .03) followed by the Low-supervised Dense condition (M = .88, SD = .17), High-supervised Sparse condition (M = .85, SD = .09), and the Low-supervised Sparse condition (M = .67, SD = .13). A 2 $\times$ 2 (Supervision $\times$ Density) between-subjects ANOVA showed a statistically significant main effect of Density (F = 14.03, $p < .005$) and Supervision (F = 9.14, $p < .01$), but no interaction (F = 2.19, p = .17). Interestingly, the Low-supervised Sparse condition did not show above chance performance (t = 2.54, p = .13).

Based on the TTC results, pigeons showed a greater benefit from supervision in learning the categories than adults. Moreover, pigeons benefited more in the Sparse condition than the adults. Interestingly, supervision also interacted in the generalization phase unlike human adults - lower supervision was accompanied by worse generalization. This generalization difference between humans and pigeons is further investigated in the next section using a simulation approach.

ALCOVE simulation

In this section we further investigate the mechanism underlying the generalization difference between the two species. When comparing the two species regarding generalization, pigeons needed more learning trials to reach the criterion (i.e., .85) while showing less accuracy in generalization. This is particularly noticeable in the Low-supervised Sparse condition where pigeons showed chance level generalization. The results could be explained by the difference in the ability to selectively attend between pigeons and humans. If humans have a better ability to selectively attend to category relevant information, it would take less time to learn the category, and would be easier to apply it to new instances since one would only need to pin point the relevant information while ignoring irrelevant information. On the other hand, if selective attention is less functional, category learning would have to rely on rote memory. Therefore when encountering new instances, generalization would rely on the overall similarity between the learned (or memorized) exemplars and the new instances. We could easily simulate this account by utilizing computational models in category learning. Here we used ALCOVE (Kruschke, 1992) to identify the contribution of selective attention, and memory to category learning.

ALCOVE is an exemplar based network model that represents the exemplars on a multidimensional space, and incorporates attentional learning. Compared to other exemplar based category learning models, it is able to simulate learning trajectories (c.f., Nosofsky, 1986), and simple in that it only deals with supervised learning (c.f., Love, Medin, & Gureckis, 2004), which is suitable to our data. However, any exemplar based category learning model that incorporates selective attention would be able to simulate the same results.

We took a simulation approach where we changed λ_α (attention weight learning rate) systematically while fixing all other parameters. The λ_α controls how well the model is able to shift attention throughout learning, and has been an useful parameter to explaining various category learning studies (e.g., Nosofsky & Kruschke, 2002). The simulation used the sparse categories used in the study, and the λ_α was decreased from 1 to 1e-5 while other free parameters were fixed (i.e., λ_w = .01, c = .5, ϕ = 4.5). The model learned the exemplars until it reached the learning criterion (.85), and was tested on the generalization stimuli without learning.

Results are shown in Figure 4. The first row shows trials to criterion (TTC) as λ_α decreases. As attentional flexibility decreased learning slowed down. The second row shows generalization accuracy as λ_α decreased. It shows that even though more learning was required under low λ_α, generalization accuracy decreased. The third row shows how attention was deployed in the model, where dimension 1 is the relevant dimension in the sparse category (note that there was only one relevant dimension in the sparse category). It shows that the model learned to attend to the relevant dimension when λ_α was high, whereas attention was diffused across all dimensions when λ_α was low. In sum, simulation results using ALCOVE were consistent with our argument where the difference in the ability to selectively attend affected both TTC and generalization accuracy.

General Discussion

The current study investigated the interactions among category density, supervision, and the ability to selectively attend during category learning. Especially, we compared pigeons with human adults to examine the effect of the ability to selectively attend to relevant dimensions during category learning. Overall, results showed that learning was faster in the

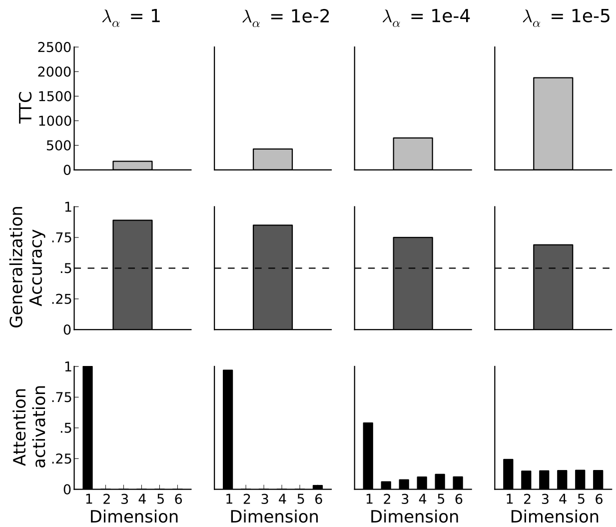

Figure 4: ALCOVE simulation results. The first column (λ_α = 1) resembles the performance of human adults whereas the fourth column (λ_α = 1e-5) resembles the performance of pigeons. All parameters where fixed where only the λ_α parameter (attention weight learning rate) was systematically changed (see text for details). The top row in each subplot represents TTC where the criterion was .85, the middle row represents accuracy in the generalization trials, and the bottom row represents attention weights where the first value on the x-axis represents the attention weight for the relevant dimension.

Dense condition than in the Sparse condition, and pigeons required more trials to learn the categories than adults. Interesting findings were that only density but not supervision affected generalization for human adults, whereas both factors affected generalization for pigeons. Moreover, adults did not have any difference in the benefit of supervision between the Dense and Sparse conditions, whereas pigeons had a greater benefit in the Sparse condition.

First, pigeons' greater benefit in the Sparse condition is worth noting. We had hypothesized that since pigeons do not have a PFC as humans do, pigeons' ability to focus on the relevant dimension (i.e., selective attention) would not benefit from the supervisory signal. There could be two possible explanations of the results. One is that pigeons do have a brain structure subserving selective attention (Shimizu, 2009). Therefore, the supervisory signal may aid the pigeons with focusing on the relevant dimension during learning. Another possibility is that the supervisory signal helps pigeons to learn the contingency between the studied exemplars and the response action (i.e., the location of pecking). In this case, pigeons would benefit from supervision during category learning without any involvement of selective attention. The two explanations are both possible, and further research would be required. Especially, the later possibility could be tested by conducting a latent learning paradigm (e.g., Tolman, 1948), where pure category learning could be decomposed from contingency learning.

Second, it was interesting that pigeons needed more learning trials to reach the learning criterion (i.e., .85) while showing less accuracy in generalization. We used ALCOVE (Kruschke, 1992) to argue that this could be due to the difference in the ability to selectively attend to category relevant information. The simulation results are also consistent with previous eye-tracking studies where infants learned categories with diffused attention whereas adults showed evidence of selective attention (Best, Yim, & Sloutsky, 2013). The two studies together imply that with higher functionality of the PFC, selective attention becomes more flexible during category learning. Moreover, with flexible selective attention category learning becomes easier, and more generalizable. This fact is particularly interesting regarding the development of category learning, since the ability to control selective attention (or the maturation of PFC) could explain the development of the speed and flexibility of category learning.

Finally, the amount of supervision between the two groups should be further investigated. Regarding the TTC results, human adults had less benefit than the pigeons. Though the results could directly indicate performance difference between the two species, it is possible that the amount of supervision between the Low-supervised and High-supervised conditions were different between the two groups. For example, it is possible that for human adults, the Low-supervised condition provided enough supervision, whereas for pigeons the amount of supervision in the High-supervised condition was tantamount to that. Future research would be required to investigate the comparable amount of supervision between adults and pigeons.

Acknowledgments

This research was supported by a National Science Foundation (NSF) grant (BCS-0720135) to Vladimir Sloutsky. We would like to thank John Opfer, and Keith Apfelbaum for their helpful comments.

References

Best, C. A., Robinson, C. W., & Sloutsky, V. M. (2014). *Attention optimization in the course of category learning.* Manuscript submitted for publication.

Best, C. A., Yim, H., & Sloutsky, V. M. (2013). The cost of selective attention in category learning: Developmental differences between adults and infants. *Journal of Experimental Child Psychology*, *116*(2), 105–119.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436.

Diamond, A. (2002). Normal development of prefrontal cortex from birth to young adulthood: Cognitive functions, anatomy, and biochemistry. In D. Stuss, & R. Knight (Eds.) *Principles of frontal lobe function*, (pp. 466–503). London, UK: Oxford University Press.

Gentner, D. (1981). Some interesting differences between verbs and nouns. *Cognition and Brain Theory*, *4*, 161–178.

Gibson, B. M., Wasserman, E. A., Frei, L., & Miller, K. (2004). Recent advances in operant conditioning technology: a versatile and affordable computerized touchscreen system. *Behavior Research Methods, Instruments, & Computers*, *36*(2), 355–62.

Hanania, R., & Smith, L. B. (2010). Selective attention and attention switching: towards a unified developmental approach. *Developmental Science*, *13*(4), 622–635.

Kloos, H., & Sloutsky, V. M. (2008). What's Behind Different Kinds of Kinds: Effects of Statistical Density on Learning and Representation of Categories. *Journal of Experimental Psychology: General*, *137*(1), 52–72.

Kruschke, J. K. (1992). ALCOVE: An Exemplar-Based Connectionist Model of Category Learning. *Psychological Review*, *99*(1), 22–44.

Love, B. C. (2002). Comparing supervised and unsupervised category learning. *Psychonomic bulletin & review*, *9*(4), 829–835.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: a network model of category learning. *Psychological review*, *111*(2), 309–32.

Mackintosh, N. J. (1965). Selective Attention in Animal Discrimination Learning. *Psychological Bulletin*, *64*(2), 124–150.

Nosofsky, R. M. (1986). Attention, Similarity, and the Identification-Categorization Relationship. *Journal of Experimental Psychology: General*, *115*(1), 39–57.

Nosofsky, R. M., & Kruschke, J. K. (2002). Single-system models and interference in category learning: Commentary on Waldron and Ashby (2001). *Psychonomic Bulletin & Review*, *9*(1), 169–174.

Palmeri, T. J., Wong, A. C.-N., & Gauthier, I. (2004). Computational approaches to the development of perceptual expertise. *Trends in cognitive sciences*, *8*(8), 378–86.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442.

Quinn, P. C., Eimas, P. D., & Rosenkrantz, S. L. (1993). Evidence for representations of perceptually similar natural categories by 3-month-old and 4-month-old infants. *Perception*, *22*(4), 463–75.

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs*, *75*(13), 1–41.

Shimizu, T. (2009). Why can birds be so smart? Background, significance, and implications of the revised view of the avian brain. *Comparative Cognition & Behavior Reviews*, *4*, 103–115.

Sloutsky, V. M. (2010). From perceptual categories to concepts: What develops? *Cognitive Science*, *34*, 1244–86.

Tolman, E. (1948). Cognitive maps in rats and men. *Psychological Review*, *55*, 189–208.