



Original Articles

Cognitive flexibility and memory in pigeons, human children, and adults

Kevin P. Darby^{a,1}, Leyre Castro^b, Edward A. Wasserman^b, Vladimir M. Sloutsky^{a,*}^a Department of Psychology, The Ohio State University, 1835 Neil Avenue, Columbus, OH 43210, United States^b Department of Psychological & Brain Sciences, The University of Iowa, W311 Seashore Hall, Iowa City, IA 52242, United States

ARTICLE INFO

Keywords:

Memory

Perseveration

Cognitive flexibility

Animal cognition

Cognitive development

ABSTRACT

This work examines cognitive flexibility using a comparative approach. Pigeons (Experiment 1), human children (Experiment 2a), and human adults (Experiment 2b) performed a task that required changing responses to the same stimuli twice across the experiment. The results indicate that all three groups demonstrated robust memory for learned information. In addition, pigeons showed comparable and substantial perseveration following both response shifts. In contrast, both children and adults exhibited some perseveration following a first response shift, while exhibiting no cost following the second response shift. These findings are discussed in relation to memory-based theories of cognitive flexibility, according to which perseveration occurs as a result of competition between long-term and working memory, revealing important differences in memory and cognitive flexibility between species.

1. Introduction

Cognitive flexibility is an important aspect of executive function that may be defined as the ability to efficiently adapt to changing task demands. As the environment, needs, and goals change frequently for both non-human animals and humans, cognitive flexibility is often critical for survival. The current work examines one form of cognitive flexibility – changing responses to the same stimuli – using a comparative approach. Specifically, we examined the role of memory in this process in pigeons as well as human children and adults.

One aspect of cognitive flexibility is responding differently to the same stimuli, depending on the current situation. For example, the Wisconsin Card Sorting Test (Berg, 1948) provides a measure of cognitive flexibility by requiring participants to learn to sort cards by different dimensions (i.e., shape, color, or number) based on feedback. Crucially, the sorting rule periodically changes without warning, such that participants must learn to switch responses to the same stimuli based on shifting rules. This task is commonly used as a clinical measure of frontal lobe functioning; patients with frontal lobe damage have greater difficulty shifting between sorting rules (Robinson, Heaton, Lehman, & Stilson, 1980).

Variants of this task are commonly used to investigate the development of flexibility. For example, in the Dimension Change Card Sort (DCCS) task, young children are asked to sort cards according to shape or color (Zelazo, 2006). After a number of trials of sorting cards according to one dimension, children are told to shift and sort by the other

dimension. Three-year-old children have substantial difficulty shifting to the second dimension; they perseverate by continuing to sort by the prior dimension (Zelazo et al., 2003). Five-year-olds, in contrast, succeed at shifting to the new rule in the standard task, but struggle in an advanced version of the task, in which they are required to shift between sorting rules on a trial-by-trial basis conditional on a contextual cue, such as the color of a border surrounding the object (Hongwanishkul, Happaney, Lee, & Zelazo, 2005).

Why does perseveration occur? And what factors account for developmental change? One account stipulates that perseveration occurs due to “attentional inertia” (Kirkham, Cruess, & Diamond, 2003). According to this theory, participants learn to attend to a particular aspect of stimuli, such as shape; perseveration happens when the contingencies change and participants struggle to inhibit the now-irrelevant dimension and shift attention to another, now-relevant dimension. Hence, the primary locus of developmental change is inhibition of established attentional patterns.

An alternative account (Morton & Munakata, 2002) stipulates that perseveration occurs due to competition between latent (or long-term) memory and active (or working) memory. Specifically, participants first learn (over multiple repetitions) a given contingency, which eventually becomes part of their long-term memory. Then, after the shift, the contingencies change, yet the stimuli do not. As a result, these old stimuli re-activate the learned contingencies in long-term memory, thus triggering the learned response. At the same time, the new contingency has to be actively maintained in working memory. This co-existence of

* Corresponding author.

E-mail address: sloutsky@psy.ohio-state.edu (V.M. Sloutsky).¹ Present address: Department of Psychology, University of Virginia, 485 McCormick Road, Charlottesville, VA 22903, United States.

conflicting memory traces creates competition between the re-activated contingencies and the new contingencies. Any failure to inhibit the old contingency, while maintaining the new one, results in perseveration. Therefore, according to this memory-based account, perseveration stems from young children's working memory limitations (Gathercole, Pickering, Ambridge, & Wearing, 2004); previously learned contingencies overwhelm children's ability to maintain new contingencies in working memory.

The current work attempts to contribute to understanding of cognitive flexibility by using a new task in which participants learn to make different responses to the same stimuli (with no changing dimensions) across different experimental phases. In contrast to dimensional shift tasks (such as the Wisconsin Card Sorting Test or DCCS), shifts in the present task pertain only to the assignment of stimuli to responses, not to the dimensions of stimuli. Therefore, because neither stimuli nor dimensions of interest change across the phases, attention should not differ across the phases. As a result, perseveration on this task would present a challenge to the attentional inertia theory. At the same time, because working and long-term memory traces conflict across the phases, the memory competition theory has a natural way of explaining such perseverations. Specifically, participants would remember the pre-shift stimulus-response mappings, and, when contingencies change, these "old" mappings interfere with the current mappings that need to be maintained in working memory.

To increase the probability of high within-phase learning coupled with response perseveration, we decided to address these issues using a comparative approach. In addition to including human children and adults in the study, we also included pigeons. We did so because avian species (and pigeons in particular) exhibit remarkable learning and memory abilities (Castro & Wasserman, 2016; Emery, 2006; Güntürkün & Bugnyar, 2016; Kirsch, Güntürkün, & Rose, 2008; Wasserman, Brooks, & McMurray, 2015), despite possessing rather limited (compared to humans) working memory capacity (Gibson, Wasserman, & Luck, 2011). Therefore, examining these organisms and comparing their performance to that of human children and adults might prove to be especially informative.

Previous research suggests that pigeons display many cognitive similarities, and some key differences, compared to humans. Pigeons have excellent visual abilities (Gibson, Wasserman, Gosselin, & Schyns, 2005; Levenson, Krupinski, Navarro, & Wasserman, 2015; Watanabe, Sakamoto, & Wakita, 1995) and are able to memorize and categorize a wide range of visual objects (Soto & Wasserman, 2014). For example, Fagot and Cook (2006) found that pigeons could memorize 800–1200 different images associated with different responses learned over a period of 3–5 years. Pigeons are not limited to memorizing individual items; they also show clear evidence of category learning (Lazareva & Wasserman, 2010; Wasserman, 2016) and even relational learning (Wasserman & Young, 2010). In addition, these birds exhibit primacy and recency effects in memory (Wright, Santiago, Sands, Kendrick, & Cook, 1985), as well as episodic-like memory (Zentall, Clement, Bhatt, & Allen, 2001), long-term memory (Cook, Levison, Gillett, & Blaisdell, 2005), and working memory (Diekamp, Kalt, & Güntürkün, 2002).

Although pigeons display impressive memory abilities, there are some critical differences in working memory between pigeons and humans. Perhaps not surprisingly, pigeons have lower working memory capacity than human adults (Gibson et al., 2011). More importantly, these capacity differences may stem from differences in how pigeons and humans bind visual features in visual working memory. In a change detection task (Lazareva & Wasserman, 2016), pigeons were sensitive to the total number of changed *features* rather than the number of changed *objects*, suggesting an absence of feature binding. In contrast, human adults did exhibit evidence of binding: namely, they were as likely to detect change in a single *feature* within a single-feature object as they were to detect change in any of multiple features within a multi-feature object (Vogel, Woodman, & Luck, 2001).

These results suggest that humans may possess object-based

working memory, whereas pigeons' working memory may be feature-based; this disparity may explain the lower working memory capacity in pigeons compared to humans, especially when multi-feature objects are involved. If cognitive flexibility depends on working memory, then pigeons should exhibit substantially poorer cognitive flexibility than humans.

Yet, prior work does suggest that pigeons, at least under some circumstances, can exhibit impressive cognitive flexibility. In one recent study (Castro & Wasserman, 2016), pigeons were trained to categorize sets of multi-item visual stimuli according to two dimensions: numerosity and variability. Impressively, pigeons were able to shift back and forth between these categorization tasks on a trial-by-trial basis, using the same set of stimuli, with only the color of the background indicating which task and response rule should be performed on any given trial.

It should be noted, however, that, because the categorization tasks were presented in an interleaved manner, these pigeons had received extensive training (144 trials per day for 50 days), which resulted in extensive practice *shifting* between tasks from one trial to the next. Therefore, these findings attest more strongly to the impressive ability of pigeons to *learn* to shift from one task to another based on an external contextual stimulus than to their spontaneous (i.e., not requiring training) cognitive flexibility. This kind of training regimen is likely to have established strong long-term memory representations for each stimulus-response contingency, thus obviating the need to maintain the current contingency in working memory. This idea is similar to considerable work in human learning showing that attentional processes which at first must be highly controlled can eventually become automatic through extensive experience (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977).

An earlier study also found evidence of cognitive flexibility in pigeons (Cook & Rosen, 2010); in this study, pigeons were trained in a task-shifting procedure in which it was necessary to change response strategies halfway through each session. Without the support of external discriminative stimuli, pigeons ably learned to shift from a match-to-sample task to an oddity-from-sample task, suggesting that they used temporal cues to flexibly change their behavior. Here too, pigeons were trained over many daily sessions, again suggesting that pigeons can demonstrate flexible task-shifting behavior supported by long-term memories. In contrast, in the current work, we examine *spontaneous* cognitive flexibility when: (1) little training to shift is provided, and (2) long-term memories are likely to compete with new information that must be maintained in working memory.

In addition to pigeons, we also studied 5-year-old human children and adults. We included 5-year-olds because this age group demonstrates flexible task-shifting behavior in the DCCS task reviewed above, although they struggle with an advanced version of the task in which they are required to shift between tasks on the basis of contextual cues differing between trials (Hongwanishkul et al., 2005). Finally, we included adults as a comparison group to examine any developmental differences in humans.

To examine the role of memory in cognitive flexibility, as well as potential species differences in this domain, we used a new variant of a task previously used to measure proactive and retroactive interference in human memory across development (Darby & Sloutsky, 2015a, 2015b). In the original task, participants learned to associate *pairs* of objects with cartoon characters across three phases. In the first phase, participants learned one set of contingencies; in the second phase, a different set of contingencies was learned; and in the third phase, the original set of Phase 1 contingencies was again presented. In the prior version of this task, the second set of object pairs (presented in Phase 2) was a recombination of the same objects that had been presented in Phase 1. This recombination of objects required *new learning* in Phase 2 that was subject to proactive interference and *retrieval of previously learned* information in Phase 3 that was subject to retroactive interference.

In contrast, in the current version of the task (schematically

presented in Fig. 1), object pairs in the experimental condition remained *identical* across phases. However, the assignment of the pairs to responses changed across phases. This new version of the task could be considered a variant of the reversal shift (Kendler & Kendler, 1969): an experimental procedure that requires participants to switch responses to the same stimuli across different phases. Although different from the dimension shifting required in the Wisconsin Card Sort Task and the DCCS, such response shifting also requires cognitive flexibility (Kehagia, Murray, & Robbins, 2010). Because the object pairings themselves did not change, this task minimized new learning across phases. At the same time, the task required cognitive flexibility to respond differently to the same stimuli when a new phase was introduced.

Pairs that were the same across phases are referred to as *overlapping pairs*. As a control condition, we also included *unique pairs* that were the same in Phases 1 and 3, but that changed from Phase 1 to Phase 2 and from Phase 2 to Phase 3.

To investigate the role of memory in cognitive flexibility, we analyzed the two *shifts* in required responses: between Phases 1 and 2 as well as between Phases 2 and 3. For overlapping pairs, these shifts measured cognitive flexibility following a change in responses to the same stimuli, whereas for unique pairs, these shifts measured new learning (when Phase 2 was introduced), and memory for previously learned pairs (when Phase 3 was introduced).

We examined cognitive flexibility and memory across shifts in pigeons (Experiment 1), human children (Experiment 2a), and human adults (Experiment 2b). Given that pigeons' working memory capacity is lower than that of humans (Gibson et al., 2011), we expected pigeons to exhibit strong perseveration across response shifts, whereas we expected humans to show greater evidence of cognitive flexibility, with adults perhaps exhibiting greater flexibility than young children. Furthermore, we expected substantially stronger perseveration for the overlapping pairs than for the unique pairs.²

2. Experiment 1

2.1. Method

2.1.1. Subjects

The subjects were eight pigeons (*Columba livia*) maintained at 85% of their free-feeding weights by controlled daily feedings. The pigeons had served in unrelated projects prior to the present experiment.

2.1.2. Apparatus

The experiment used four 36 × 36 × 41 cm operant conditioning chambers detailed by Gibson, Wasserman, Frei, and Miller (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 × 17.0 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. A houselight mounted on the upper rear wall of the chamber provided illumination during the experimental sessions. The pellet dispenser and the houselight were controlled by a digital I/O interface board. Each chamber was controlled by its own Apple® iMac® computer. Programs to run this experiment were developed in MATLAB® with Psychtoolbox-3 extensions (Brainard, 1997; Pelli, 1997).

² Note that all stimuli and data for this paper are available at Harvard Dataverse: <https://dataverse.harvard.edu/dataset.xhtml?persistentId=http://dx.doi.org/10.7910/DVN/QPNN86>.

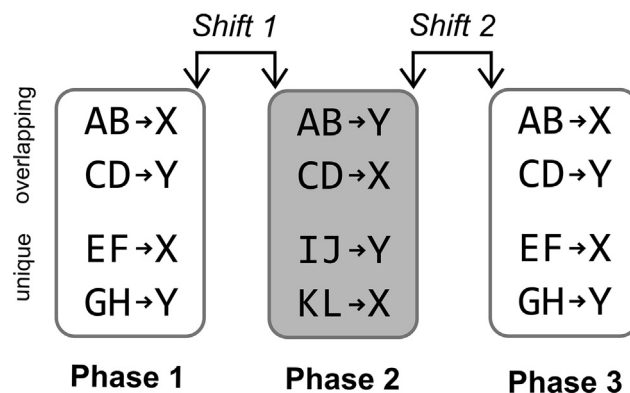


Fig. 1. Structure of all experiments. Objects are denoted by the letters A – L; response options by the letters X and Y. The shaded boxes represent the visual contextual difference among the three phases.

2.1.3. Stimuli and design

Stimuli included 12 illustrations of common objects (such as a boat, turtle, and baseball cap) that served as the objects to be sorted, as well as two cartoon characters (Mickey Mouse and Winnie the Pooh) that served as the response buttons. For half of the pigeons, Mickey Mouse always appeared on the left and Winnie the Pooh always appeared on the right; the assignment was reversed for the other half of the pigeons. Stimuli were presented on a gray background, the shade of which varied across the three experimental phases. For half of the pigeons, stimuli were presented on a light gray background (RGB color code: 200, 200, 200) in Phases 1 and 3, whereas in Phase 2, stimuli were presented on a dark gray background (RGB color code: 100, 100, 100); these assignments were reversed for the other half of the pigeons. This difference in the brightness of the background could act as a contextual cue to response shifts across phases.

The experiment comprised three phases. In each phase, pigeons learned to associate pairs of objects with different responses (instantiated as images of Mickey Mouse and Winnie the Pooh). Phases 1 and 3 involved the same set of pair-response contingencies, whereas a different set was learned in Phase 2, creating two response *shifts* (between Phases 1 and 2 as well as between Phases 2 and 3). Each set contained two types of pairs: overlapping and unique. The *overlapping pairs* served as experimental stimuli; they consisted of four objects that were identical across phases, although they were associated with different responses across phases. In contrast, the *unique pairs* served as control stimuli; they consisted of eight objects that were different across phases (see experimental design in Fig. 1). Assignment of the objects to each pair was counterbalanced across pigeons.

2.2. Procedure

Familiarization. Pigeons were exposed for 4 days to all of the individual objects that would serve as categorization stimuli and all of the individual cartoon characters that would serve as response buttons. Each of the images was presented equally often, 10 times, for a total of 140 familiarization trials per day. Each of the individual objects appeared half of the time on the top part and half of the time on the bottom part of the area that in training would be occupied by the stimulus pair. Pigeons had to peck each of the images a number of times (2 pecks on the first day, a requirement that was gradually increased to 5 pecks on the fourth day); once the requirement was completed, food reward was delivered. This familiarization phase allowed the birds to acclimate to the new stimuli so that neophobic or neophilic responses would be less likely to occur later during training.

Training. Daily sessions comprised 144 trials, divided into 6 blocks of 24 trials each; half of the trials presented overlapping pairs and half of the trials presented unique pairs, in a random order. At the beginning

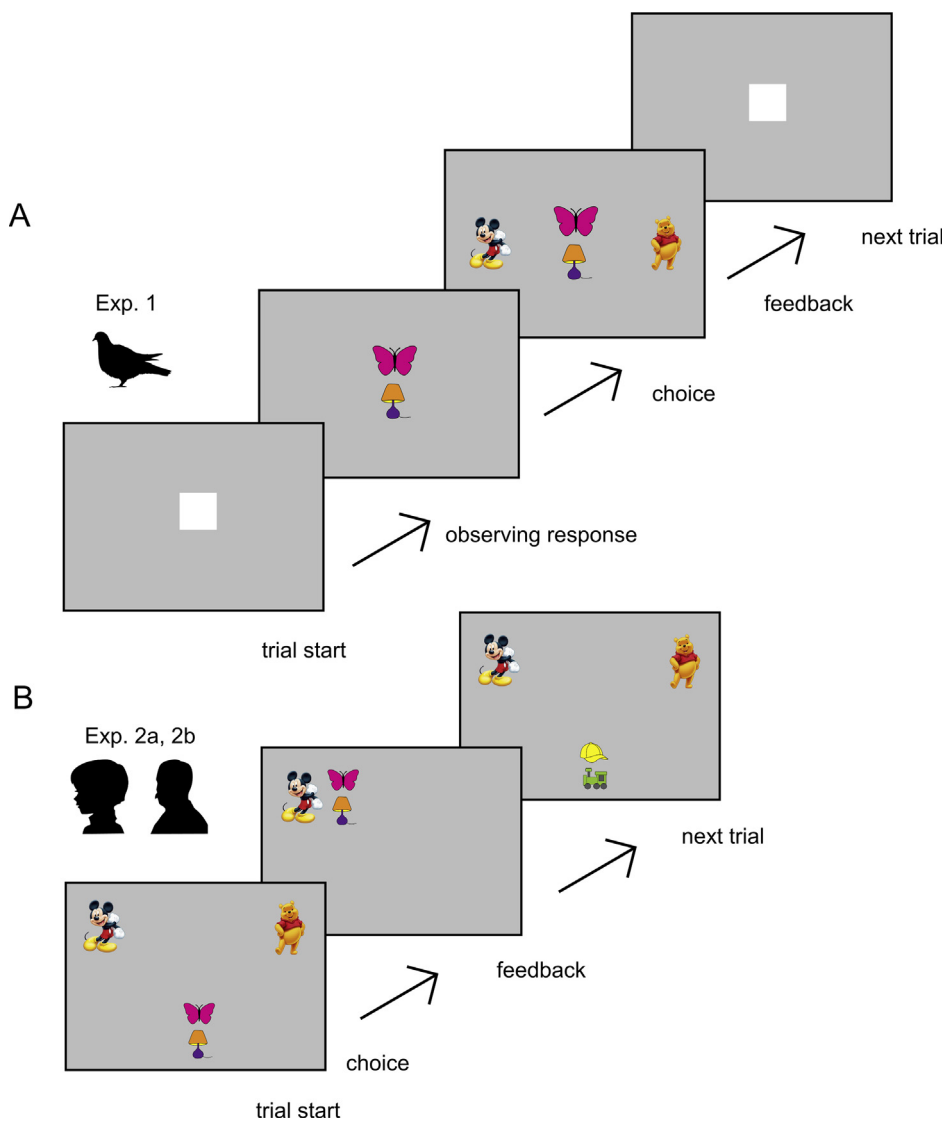


Fig. 2. Trial procedure for pigeons (Experiment 1; A) and humans (Experiments 2a and 2b; B). Upon a correct choice response, pigeons were provided with food reinforcement, whereas upon an incorrect response no food was provided, the houselight was darkened, and correction trials were presented until pigeons made the correct choice. Children (Experiment 2a) were given verbal feedback (e.g., “Great job, that pair does go to Mickey!” for correct responses or “Uh oh, that pair goes to Mickey!” for incorrect responses) as well as auditory feedback (i.e., a high- or a low-toned beep for correct and incorrect responses, respectively), whereas adults (Experiment 2b) were given auditory feedback only. Regardless of response accuracy, children and adults were shown the correct pair-character association prior to presentation of the next trial. (This figure appears in color in the on-line version.)

of a trial, the pigeons were presented with an orienting stimulus: a white square (3×3 cm) in the middle of the computer screen. After one peck anywhere on this white square, one stimulus pair was displayed in the center of the screen. The pigeons had to satisfy an observing response requirement (from 8 to 12 pecks, depending on the pigeon) to the stimuli. On completion of the observing response requirement, the two response buttons (Mickey Mouse and Winnie the Pooh) appeared to the left and right of the stimulus pair, aligned in the center of the stimulus array (see Fig. 2). The pigeons had to select one of the two buttons to indicate their response, depending on the stimulus presented. If the choice response was correct, then food reinforcement (1–2 pellets) was delivered and the intertrial interval (from 6 to 10 s) ensued. If the choice response was incorrect, then food was not delivered, the houselight darkened, and a correction trial was given. On correction trials, pigeons were presented with the same stimuli again until the correct response was made. No data were analyzed from correction trials. For each of the training phases, we trained the birds until they reached an accuracy level of 85% on 2 consecutive days, for each overlapping and unique pair, to ensure that performance had reached a high stable level. Then, the birds were moved to the next phase.

For half of the pigeons the shifts between phases took place within the same training session (halfway into the session), whereas for the other half of the pigeons the shifts between phases always occurred at the beginning of a new session (24 h apart). This manipulation was

done to determine whether perseveration might vary based on the amount of time between response shifts. However, no variables of interest were affected by this procedural difference, so we included all 8 pigeons as a single sample in all of the analyses reported below.

2.3. Results and discussion

Pigeons took an average of 6.13 days to reach criterion in Phase 1, 5.50 days to reach criterion in Phase 2, and 4.25 days to reach criterion in Phase 3.

To examine pigeons' ability to flexibly adapt their responses to new contingencies, we measured accuracy before and after transitions between experimental phases. We refer to these transitions as Shift 1 (between Phases 1 and 2) and Shift 2 (between Phases 2 and 3). Note that for overlapping pairs, these shifts required changing responses to the same stimuli from one cartoon character to the other, whereas for unique pairs, because the stimuli were different across phases, these shifts required responding to different stimuli.

We expected these shifts to result in decreased accuracy for overlapping pairs, since these pairs were the same, but were now associated with different responses across phases. For example, following the first shift (i.e., when starting Phase 2), it would be reasonable to respond to overlapping pairs in the same way as in Phase 1, initially leading to below chance performance. For unique pairs, the first shift would be

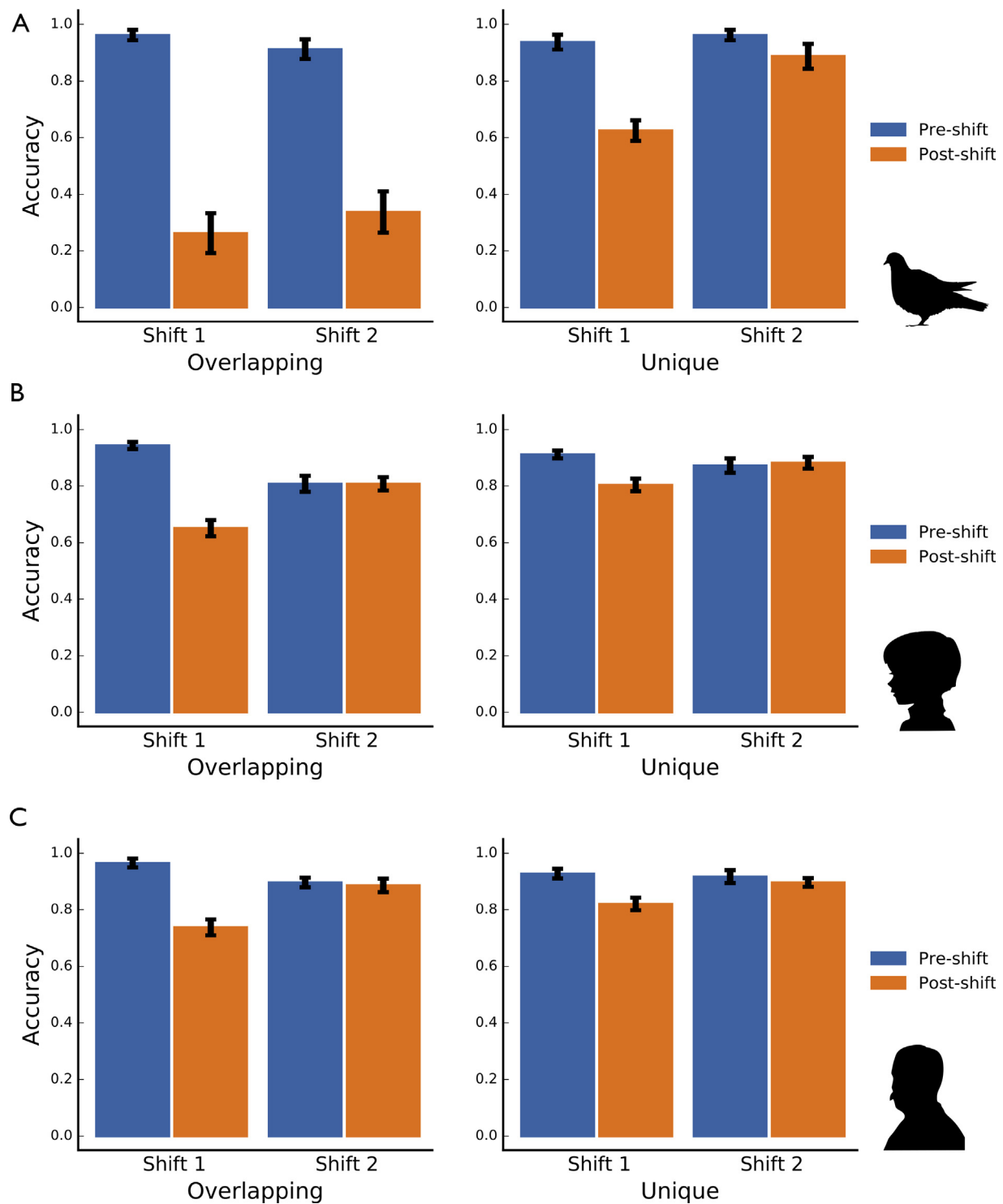


Fig. 3. Accuracy before and after each shift for overlapping (left) and unique (right) pairs, Experiments 1 (A) with pigeons, 2a (B) with children, and 2b (C) with adults.

expected to result in lower accuracy, given that new pairs were introduced after the shift. If pigeons retained long-term memory of what was learned in Phase 1, however, there should have been little or no cost for unique pairs after the second shift, when the initially learned pairs were re-introduced.

We calculated the costs of the two shifts by comparing accuracy in the task prior to the shift and after the shift for both overlapping and unique pairs. For each shift, we calculated pre- and post-shift accuracy using five repetitions of stimuli for each pair type. For example, Shift 1

was calculated as the average change in accuracy between the final 5 repetitions of each pair in Phase 1 versus the first 5 repetitions of each pair in Phase 2. To determine the generality of our findings, we also made comparisons using set sizes of 1 and 3 repetitions. As shown in the [Supplementary Materials \(Fig. S1\)](#), the pattern of results was the same regardless of the number of trials analyzed.

The critical results (i.e., accuracy before and after Shifts 1 and 2 for overlapping as well as unique pairs) are shown in [Fig. 3A](#). We first analyzed these results using a 2 (Order: Shift 1 vs. Shift 2) × 2 (Shift:

Pre-Shift vs. Post-Shift) \times 2 (Pair Type: Overlapping vs. Unique) repeated-measures ANOVA.

We expected that any changes in accuracy across the response shifts would vary between overlapping and unique pairs. This expectation was confirmed by a significant interaction between Shift and Pair Type, $F(1,7) = 102.88$, $p < .001$, $\eta_p^2 = 0.94$, indicating a larger decrease in performance following the shift for overlapping pairs compared to unique pairs. Perhaps most importantly, for the unique pairs, pigeons' performance was above chance following the first shift ($M_{\text{accuracy}} = 0.63$, $SD = 0.10$), $t(7) = 3.42$, $p = .01$, $CI_{\text{accuracy}} [0.54, 0.71]$, $d = 1.21$, $CI_d [0.26, 2.12]$, as well as the second shift ($M_{\text{accuracy}} = 0.89$, $SD = 0.13$), $t(7) = 8.88$, $p < .001$, $CI_{\text{accuracy}} [0.78, 0.99]$, $d = 3.11$, $CI_d [1.36, 4.83]$, whereas for the overlapping pairs, their performance was significantly below chance following the first shift ($M_{\text{accuracy}} = 0.26$, $SD = 0.20$), $t(7) = 3.37$, $p = .01$, $CI_{\text{accuracy}} [0.10, 0.43]$, $d = -1.19$, $CI_d [-2.09, -0.24]$, and marginally below chance following the second shift ($M_{\text{accuracy}} = 0.34$, $SD = 0.17$), $t(7) = -2.23$, $p = .06$, $CI_{\text{accuracy}} [0.17, 0.51]$, $d = -0.79$, $CI_d [-1.57, 0.04]$. Therefore, whereas pigeons adjusted their responses following the shifts for the unique pairs, they exhibited measurable perseveration after the shifts for the overlapping pairs.

Planned post hoc analyses indicated that there was a significant post-shift decrease in accuracy for overlapping pairs for both Shift 1 ($M_{\text{decrease}} = 0.70$, $SD = 0.21$), $t(7) = 9.26$, $p < .001$, $CI_{\text{decrease}} [0.52, 0.88]$, $d = 3.27$, $CI_d [1.45, 5.08]$, and Shift 2 ($M_{\text{decrease}} = 0.58$, $SD = 0.27$), $t(7) = 6.12$, $p < .001$, $CI_{\text{decrease}} [0.35, 0.80]$, $d = 2.16$, $CI_d [0.84, 3.45]$, indicating strong perseveration across both response shifts. For unique pairs, there was a significant reduction in accuracy across the first shift ($M_{\text{decrease}} = 0.31$, $SD = 0.16$), $t(7) = 5.69$, $p = .001$, $CI_{\text{decrease}} [0.18, 0.44]$, $d = 2.01$, $CI_d [0.75, 3.24]$, but there was no significant change in accuracy across the second shift ($M_{\text{decrease}} = 0.08$, $SD = 0.15$), $p = .20$; these findings indicate that although pigeons were less accurate at the beginning of Phase 2 than at the end of Phase 1 (which was expected given that new stimulus pairs were introduced in Phase 2), pigeons' long-term memory for unique pairs acquired in Phase 1 was largely intact.

There was also a marginally significant interaction between Order and Shift, $F(1,7) = 5.58$, $p = .05$, $\eta_p^2 = 0.44$, suggesting a smaller drop in overall performance following the second response shift compared to the first. No other interactions were significant, $ps > .08$.

Experiment 1 presents two important findings. First, pigeons retained memory of the stimulus-response mappings for the *unique* pairs learned in Phase 1 until Phase 3, as evidenced by their high accuracy after Shift 2 for the unique pairs. Second, pigeons exhibited strong perseveration for the *overlapping* pairs, indicating that they failed to promptly adjust their responses when the stimulus-response mappings changed following both shifts. The memory competition account (Morton & Munakata, 2002) suggests that perseveration occurred because pigeons' working memory for new stimulus-response mappings was overwhelmed by stronger long-term memory representations formed in the immediately preceding phase.

One potential counterargument is that perhaps our pigeons were trained too extensively on the contingencies in each phase, so that overlearning may have inflated levels of perseveration when contingencies shifted. However, there was no relationship (all $ps > .10$) between the number of trials to criterion in the first two phases and post-shift accuracy (or the difference in pre- and post-shift accuracy) for either pair type. These results suggest that the amount of training did not have a substantial impact on the amount of perseveration (see Fig. S2 in Supplementary Materials for the relationship between pre-shift number of trials to criterion and post-shift accuracy for overlapping pairs).

Although pigeons performed quite poorly following response shifts for overlapping pairs across phases, an important question is whether humans (who presumably, compared to pigeons, have more robust

ability to maintain information in working memory) would be able to avoid such prolonged perseverative responding. To investigate this question, we measured performance across response shifts in 5-year-old children (Experiment 2a) and adults (Experiment 2b). Recall that we selected this age group because 5-year-olds, although shifting flexibly in the standard version of the DCCS task (i.e., where the shift is announced), struggle with a more advanced version of the task (i.e., where unannounced shifts are presented on a trial-by-trial basis).

3. Experiment 2a

3.1. Method

3.1.1. Participants

Seventy preschool-aged human children ($M_{\text{age}} = 5.2$ years; $SD = 0.3$; range = 4.6–6.0; 31 boys; 39 girls) participated in this experiment. Children were recruited from preschools and day care centers in the area of Columbus, OH, on the basis of the child's verbal consent and a parent's or guardian's signed consent. Children received stickers for participating.

3.1.2. Stimuli, design, and procedure

The stimuli were identical to those used in Experiment 1. The procedure was also similar, except for some key differences. First, all children completed five blocks of training in all phases (i.e., the amount of training completed in a given phase was not based on a learning criterion), with 8 trials per block, for a total of 120 trials across the experiment. As in Experiment 1, the brightness of the background varied across phases. In Phases 1 and 3 the background was a light gray (RGB color code: 183, 188, 187), whereas in Phase 2 the background was a darker gray (RGB color code: 128, 131, 130).

The objects presented on each trial were shown near the bottom of the screen, with the two characters appearing in the two upper corners (see Fig. 2). The spatial position of each character (left or right) was randomized for each participant but was consistent within the session. Stimuli were presented to children using OpenSesame software (Mathôt, Schreij, & Theeuwes, 2012).

Children were instructed before the experiment began that they would see a pair of objects and would be asked to predict whether the pair belonged to Winnie the Pooh or to Mickey Mouse. On each trial, children made a response using a touchscreen and received visual, auditory, and verbal feedback: the pair of objects disappeared from the bottom of the screen and reappeared by the correct character (the incorrect character disappeared during feedback to emphasize the correct contingency), a high or low tone was heard for correct and incorrect responses, respectively, and the experimenter gave additional verbal feedback (e.g., "Great job, that pair does go to Mickey!" or "Uh oh, that pair goes to Pooh Bear!").

Approximately half of the children completed all three phases in a single session, whereas the other half of children completed Phases 1 and 2 in one session, and Phase 3 in a second session with the same experimenter 48 h later. This manipulation was done to determine if a long delay between phases might affect switching behavior (cf. Darby & Sloutsky, 2015b). However, no results of interest varied between these conditions, so we collapsed data from these conditions for all analyses presented below.

3.2. Results and discussion

Five children were excluded due to failure to complete all phases of the experiment. We also implemented a learning criterion of at least 60% accuracy for both the overlapping and the unique pairs, as well as at least 70% accuracy overall, averaged across Blocks 2–5 of the first phase. These criteria were implemented to ensure that children learned the first set of contingencies; without clear evidence of learning, it is difficult to interpret changes in accuracy later in the experiment. The

learning criteria resulted in excluding an additional 14 children. The final sample included 51 children ($M_{\text{age}} = 5.2$ years, $SD = 0.3$, min = 4.6, max = 6.0, 21 boys and 30 girls). Given the differences in the learning criteria between pigeons and children, we also analyzed a subset of children who exhibited accuracy of 85% or above for overlapping pairs in Phase 1 ($n = 28$), to examine whether success at learning the task might influence perseveration, and to better equate the learning criteria with those of pigeons; but, the pattern of results was the same as that of the larger sample (see Fig. S3 and related analyses in the Supplementary Materials).

Fig. 3B shows children's accuracy before and after each stimulus-response shift. To analyze children's performance, we first performed a 2 (Order: Shift 1 vs. Shift 2) \times 2 (Shift: Pre-Shift vs. Post-Shift) \times 2 (Pair Type: Overlapping vs. Unique) repeated-measures ANOVA. The three-way interaction among these factors was significant, $F(1,50) = 7.91$, $p = .007$, $\eta_p^2 = 0.14$. To better understand this interaction, we separately analyzed children's performance for unique and overlapping pairs.

Unique Pairs. We first performed a 2 (Order: Shift 1 vs. Shift 2) \times 2 (Shift: Pre-Shift vs. Post-Shift) repeated-measures ANOVA on children's accuracy for unique pairs. There was a significant interaction, $F(1,50) = 6.42$, $p = .01$, $\eta_p^2 = 0.11$, indicating some decrease in accuracy on the first shift ($M_{\text{decrease}} = 0.11$, $SD = 0.19$), $t(50) = 4.16$, $p < .001$, $CI_{\text{decrease}} [0.06, 0.16]$, $d = 0.58$, $CI_d [0.28, 0.88]$, but not on the second shift ($M_{\text{decrease}} = -0.01$, $SD = 0.19$), $p = .71$. Furthermore, accuracy was above chance following both the first shift ($M_{\text{accuracy}} = 0.80$, $SD = 0.16$), $t(50) = 13.68$, $p < .001$, $CI_{\text{accuracy}} [0.76, 0.85]$, $d = 1.92$, $CI_d [1.45, 2.38]$, and the second shift ($M_{\text{accuracy}} = 0.88$, $SD = 0.15$), $t(50) = 18.46$, $p < .001$, $CI_{\text{accuracy}} [0.84, 0.92]$, $d = 2.59$, $CI_d [2.01, 3.16]$. Therefore, children adjusted their responses following both shifts.

Overlapping Pairs. We also performed a 2 (Order: Shift 1 vs. Shift 2) \times 2 (Shift: Pre-Shift vs. Post-Shift) repeated-measures ANOVA on accuracy for overlapping pairs. Importantly, similar to the Unique pairs there was a significant interaction, $F(1,50) = 29.75$, $p < .001$, $\eta_p^2 = 0.37$, indicating that the change in accuracy was different between the shifts. There was a measurable decrease in performance following the first shift ($M_{\text{decrease}} = 0.29$, $SD = 0.22$), $t(50) = 9.49$, $p < .001$, $CI_{\text{decrease}} [0.23, 0.35]$, $d = 1.33$, $CI_d [0.95, 1.70]$, but not the second shift ($M_{\text{decrease}} = 0$, $SD = 0.23$), $p = 1.0$. Furthermore, accuracy was above chance following both Shift 1 ($M_{\text{accuracy}} = 0.65$, $SD = 0.20$), $t(50) = 5.33$, $p < .001$, $CI_{\text{accuracy}} [0.59, 0.71]$, $d = 0.75$, $CI_d [0.43, 1.05]$, and Shift 2 ($M_{\text{accuracy}} = 0.81$, $SD = 0.17$), $t(50) = 12.97$, $p < .001$, $CI_{\text{accuracy}} [0.77, 0.86]$, $d = 1.82$, $CI_d [1.36, 2.26]$. Therefore, in contrast to pigeons, children adjusted their responses following both shifts, perhaps more so after the second shift (i.e., in Phase 3).

This pattern of results suggests that children clearly remembered information learned in previous phases. Although they exhibited some perseveration in Phase 2 due to the different responses for the overlapping pairs, when a second shift occurred they exhibited no shift cost. This result is in marked contrast to the performance of pigeons in Experiment 1, who exhibited strong perseveration for overlapping pairs after both shifts. Whereas pigeons' responses were largely dominated by memory for the previous phase, children were able to learn to quickly and flexibly override responses learned in the previous phase. In Experiment 2b, we deployed the same paradigm to examine response shifting in adults.

4. Experiment 2b

4.1. Method

4.1.1. Participants

Twenty-nine human adults (21 females) participated in this experiment in exchange for partial undergraduate course credit.

4.1.2. Stimuli, design, and procedure

This experiment was identical to Experiment 2a, except that all participants completed all three phases in a single session. In addition, adults received the same visual and auditory feedback for their responses that children received in Experiment 2a, but they were not given verbal feedback by the experimenter.

4.2. Results and discussion

No adults were excluded from analyses based on the performance criteria introduced in Experiment 2a. **Fig. 3C** shows adults' accuracy before and after each stimulus-response shift. We first analyzed adults' performance using a 2 (Order: Shift 1 vs. Shift 2) \times 2 (Shift: Pre-Shift vs. Post-Shift) \times 2 (Pair Type: Overlapping vs. Unique) repeated-measures ANOVA. Similar to children's results in Experiment 2, there was a significant three-way interaction among these factors, $F(1,28) = 12.21$, $p = .002$, $\eta_p^2 = 0.30$. To more closely examine these effects, we conducted separate analyses for unique and overlapping pairs.

Unique Pairs: We first analyzed adults' performance for unique pairs using a 2 (Order: Shift 1 vs. Shift 2) \times 2 (Shift: Pre-Shift vs. Post-Shift) repeated-measures ANOVA. There was a significant interaction, $F(1,28) = 7.60$, $p = .01$, $\eta_p^2 = 0.21$, pointing to somewhat lower accuracy on trials after Shift 1 ($M_{\text{decrease}} = 0.11$, $SD = 0.12$), $t(28) = 4.95$, $p < .001$, $CI_{\text{difference}} [0.06, 0.15]$, $d = 0.92$, $CI_d [0.48, 1.35]$, but not after Shift 2 ($M_{\text{decrease}} = 0.02$, $SD = 0.11$), $p = .31$. However, accuracy was above chance following both the first shift ($M_{\text{accuracy}} = 0.82$, $SD = 0.12$), $t(28) = 14.31$, $p < .001$, $CI_{\text{accuracy}} [0.78, 0.87]$, $d = 2.66$, $CI_d [1.87, 3.44]$, and the second shift ($M_{\text{accuracy}} = 0.90$, $SD = 0.08$), $t(28) = 25.95$, $p < .001$, $CI_{\text{accuracy}} [0.87, 0.93]$, $d = 4.82$, $CI_d [3.51, 6.12]$, indicating that adults adjusted their responses following both shifts, perhaps more so after the second shift.

Overlapping Pairs: Similar to unique pairs, there was a significant interaction, $F(1,28) = 39.86$, $p < .001$, $\eta_p^2 = 0.59$, which resulted from a significant decrease in accuracy between the pre- and post-shift trials for Shift 1 ($M_{\text{decrease}} = 0.23$, $SD = 0.02$), $t(28) = 10.03$, $p < .001$, $CI_{\text{difference}} [0.18, 0.27]$, $d = 1.86$, $CI_d [1.25, 2.46]$, but not for Shift 2 ($M_{\text{decrease}} = 0.01$, $SD = 0.14$), $p = .69$. Importantly, adults' accuracy was above chance after the first shift ($M_{\text{accuracy}} = 0.74$, $SD = 0.15$), $t(28) = 8.55$, $p < .001$, $CI_{\text{accuracy}} [0.68, 0.80]$, $d = 1.59$, $CI_d [1.03, 2.13]$, as well after the second shift ($M_{\text{accuracy}} = 0.89$, $SD = 0.13$), $t(28) = 16.32$, $p < .001$, $CI_{\text{accuracy}} [0.84, 0.94]$, $d = 3.03$, $CI_d [2.16, 3.90]$, indicating that they adjusted their responses following both shifts, perhaps more so after the second shift. These results indicate that adults experienced some perseveration for overlapping pairs following the first shift (likely due to memory for the previous phase), but they were able to overcome this interfering memory to quickly and successfully adjust their responses.

How did humans avoid severe perseveration with overlapping pairs, especially following the second stimulus-response shift? One possibility is that they used working memory to override responses learned in the previous phase. A second possibility is that they formed configural memory traces that included contextual information that uniquely predicted the correct response across phases. For example, it is possible that children and adults noticed the difference in background brightness between phases and bound that information in memory along with the correct response to overlapping pairs in each phase.

To investigate this possibility, we included a short testing phase at the end of the experiment for adults. In this testing phase, adults again indicated whether pairs belonged to Mickey Mouse or Winnie the Pooh, but this time without feedback. Participants were first tested on contingencies presented in Phase 2 (8 trials), and then (without interruption) they were tested on contingencies presented in Phases 1 and 3 (8 trials). Note that in contrast to the main experiment, no feedback was given to participants during this testing.

If participants formed configural memories that included contextual

background information, then they should perform well on both blocks, even for overlapping pairs. To test this possibility, we averaged performance between these two blocks for overlapping and unique pairs. Adults had high accuracy for unique pairs ($M_{accuracy} = 0.81$, $SD = 0.22$), $t(28) = 7.77$, $p < .001$, $CI_{accuracy} [0.73, 0.90]$, $d = 1.44$, $CI_d [0.091, 1.96]$, but they were not above chance for overlapping pairs ($M_{accuracy} = 0.54$, $SD = 0.21$), $p = .33$. Therefore, we did not find evidence that adults formed configural memory representations including contextual background cues that supported flexible response shifting in the absence of performance feedback. Given that adults did not exhibit any evidence of forming configural memory representations, it seems unlikely that young children did so. Although it is possible that configural memory representations were formed but not activated in the memory test, it seems more likely that both adults and young children used working memory, rather than configural long-term memory, to flexibly adjust their responses.

4.2.1. Developmental differences in experiments 2a and 2b

To examine the possibility of developmental differences between children and adults, we performed 2 (Order: Shift 1 vs. Shift 2) \times 2 (Shift: Pre-Shift vs. Post-Shift) \times 2 (Age: Children vs. Adults) mixed ANOVAs with Order and Shift as within-subject factors, and Age as a between-subject factor, separately for overlapping and unique pairs. For unique pairs, there was a significant main effect of age, $F(1,78) = 9.32$, $p = .003$, $\eta_p^2 = 0.11$, indicating overall higher accuracy in adults than children, but no interactions involving age were significant. For overlapping pairs, there was no main effect of age and no interactions involving age. Overall, these results suggest very similar performance between children and adults in these experiments.

4.2.2. Cross-species differences in Experiments 1, 2a, and 2b

An important consideration of this project concerns ways in which pigeons and humans performed similarly or differently in our tasks. Given some methodological difference between the species, as well as differences in sample size, it would be inappropriate to directly compare the species using inferential statistics. However, there are clear qualitative trends in the data that merit discussion.

First, whereas pigeons showed very strong perseveration following both response shifts for overlapping pairs, human children and adults showed relatively moderate perseveration following the first shift and no perseveration at all following the second shift. This clear qualitative difference in the pattern of results suggests substantially greater cognitive flexibility in the human participants.

Although the species differed dramatically in performance for overlapping pairs following the shifts, there was a clear similarity in the pattern of performance for unique pairs. Specifically, although both species showed a moderate drop in performance when introduced to new pairs following the first shift, both pigeon and human participants showed no performance decrement for unique pairs following the second shift. Furthermore, for the unique pairs, responding in all cases was above chance after every shift.

This similarity in responding to the unique pairs is important because it suggests that both pigeons and humans had no difficulty remembering information across phases of the task. This suggests that pigeons' low performance for overlapping pairs following the response shifts was due to a lack of cognitive flexibility rather than a general inability to learn or adapt to the shifting task demands. Crucially, the similarity in performance for the unique pairs also suggests that procedural differences between tasks for the different species did not greatly affect the results. If the species were performing the tasks in very different ways, then we would expect to see a different pattern of results for unique pairs as well as overlapping pairs.

Although these findings support the view that species differences in cognitive flexibility were responsible for the observed disparities in performance—and not the exact tasks that were performed—it is

nonetheless possible that differences in how the stimuli were processed could have played some role. Specifically, it is possible that the stimuli were familiar and easily nameable to human children and adults, but not to pigeons; this factor could have affected the results. To examine this possibility, we conducted an additional control experiment in which adults (using a new sample) were presented with the same task as in Experiment 2b, except that the stimuli were abstract images that were novel and not easily nameable (see Fig. S4 in Supplementary Materials). The results of this additional experiment suggest that making the task more abstract made learning more difficult for adults, but it had little impact on cognitive flexibility. Specifically, and similar to Experiment 2b, there was no performance decrement following the second response shift, with participants exhibiting above chance accuracy. These findings suggest little effect of novel and abstract stimuli on cognitive flexibility in adults (see Fig. S5 in Supplementary Materials). We return to the important topic of possible effects of task differences in the General Discussion.

5. General discussion

5.1. Summary of findings

The reported experiments examined cognitive flexibility in pigeons as well as in human children and adults. In our task, pairs of objects were associated with different characters across three phases. Overlapping pairs were identical across all phases, but the associated responses shifted between Phases 1 and 2 (Shift 1) and again, back to the first learned associations, between Phases 2 and 3 (Shift 2). Unique pairs involved different objects in Phase 2 compared to those in Phases 1 and 3.

For the unique pairs, we found similar patterns of responding in pigeons and humans, which suggests that participants in all groups were able to adjust to at least some of the changing contingencies in the task. At the same time, there were substantial differences with respect to overlapping pairs: pigeons exhibited stronger and more persistent perseveration than either children or adults.

5.2. Mechanisms of cognitive flexibility

To interpret these results, it is useful to consider extant theories of perseveration and cognitive flexibility. One such theory is the attentional inertia account. According to this theory, young children perseverate in the DCCS and other tasks measuring cognitive flexibility because they have difficulty inhibiting attention to a previously relevant stimulus dimension; as a result they find it difficult to shift attention to a newly relevant dimension (Kirkham et al., 2003). Although successful in explaining perseveration in some tasks, the attentional inertia theory may have difficulty accounting for perseveration in the current task (as well as cross-species differences in perseveration), as no attentional shifting was needed for task performance, because responses were not tied to specific dimensions of the experimental stimuli.

Although the current results support the memory account of perseveration and challenge the attentional inertia account, they cannot eliminate the possibility that attentional inertia underlies perseveration in dimension shifts (as in the DCCS), but not response shifts (as in the current task). For example, it is possible that different kinds of flexibility (i.e., response shifts, dimension shifts, and task shifts) are based on different mechanisms (Kehagia et al., 2010). Specifically, different flexibility tasks may activate different cortical-subcortical neural networks (McAlonan & Brown, 2003; Rogers, Andrews, Grasby, Brooks, & Robbins, 2000), suggesting the possibility of mechanistic differences between performance on cognitive flexibility tasks. Note, however, that the current findings are not the only challenge to the attentional inertia account; this account has been also challenged as an explanation of dimension shifts with the DCCS task (Jordan & Morton, 2012; Müller, Dick, Gela, Overton, & Zelazo, 2006).

Given the current findings, as well as others using variants of the DCCS task (Jordan & Morton, 2012; Müller et al., 2006), we suggest that the attentional inertia framework does not provide a compelling explanation for cognitive flexibility (or lack thereof). An alternative account of cognitive flexibility is based on competition between representations in long-term and working memory (Morton & Munakata, 2002). According to this theory, perseveration occurs when new goals, supported by working memory, are overwhelmed by established long-term memory representations. A memory-based interpretation of the current results is that, on the one hand, pigeons have strong long-term memory for learned contingencies, but they have difficulty updating their responses based on the current working memory representations; on the other hand, humans have the ability to quickly update their responses, perhaps using working memory to overcome established long-term memory representations.

5.3. Memory competition

One important issue is the nature of competition between long-term and working memory representations. We suggest that in our task, long-term memory representations were formed during training in each phase, especially given that each stimulus pair was learned over multiple trials and that there was some delay between phases. These representations likely included not only associations between objects and the associated characters, but also between these contingencies and a particular motor response. Following the introduction of a new phase, however, working memory was necessary to maintain new contingencies for the old stimuli. Given the old stimuli, the old response contingencies were likely to be retrieved, thus resulting in competition between re-activated old contingencies and the new contingencies maintained in working memory. Any failure to maintain the new contingencies in working memory and/or inhibit the old contingencies would result in perseveration.

5.4. Cognitive flexibility across species

Recall that, in contrast to the current results, pigeons have demonstrated notable cognitive flexibility in prior work (Castro & Wasserman, 2016; Cook & Rosen, 2010). However, it seems likely that such flexible behaviors were bootstrapped by extensive experience with the different tasks and, importantly, rapid shifting between these tasks had come under strong contextual stimulus control after extensive training. Using mechanisms such as associative learning, pigeons could slowly and gradually construct representations in memory, as well as the ability to flexibly shift between representations if those representations were firmly associated with salient contextual stimuli (see Castro & Wasserman, 2016). In contrast, the current work suggests that pigeons struggle to exhibit cognitive flexibility in the absence of such extensive training and contextual cues.

In contrast to pigeons, humans (both children and adults) exhibited clear evidence of flexible response shifting. Children and adults did exhibit a cost for overlapping pairs following the first shift, but this cost was expected given that no instructions indicated that a shift would occur and participants needed to learn to change responses based solely on feedback. Although this first shift resulted in a cost to accuracy, neither children nor adults exhibited a cost following the second shift.

Why did humans not exhibit a cost following the second shift? One possibility is that humans formed complex representations in long-term memory including contextual cues (such as the brightness of the background) and used memory cues to make the correct responses in Phase 3. To address this possibility we included a testing phase following Experiment 2b, in which trials from Phases 1 and 2 were both tested, and no feedback was provided. Because participants could not rely on feedback for an error signal indicating that the correct responses had shifted, they could only rely on contextual cues such as the color of the background to maintain high accuracy. Adults' low performance on

the testing phase suggests that they were likely not using configural memory to flexibly adjust their responses in the main experiment.

Another possibility is that humans were able to use working memory to quickly and flexibly adjust responses. It is possible that children and adults did so on the basis of feedback: once they saw that the correct response had just been changed, they could use working memory to promote the new response. It is also possible that working memory was boosted by strategy use: perhaps children and adults realized after the first transition between phases that responses to overlapping pairs had changed, and as a result were more likely to change their responses again following the transition between Phases 2 and 3. Although it is unclear what exact mechanisms were used by children and adults to avoid a cost of switching after the second response shift, it is clear that humans' responses were not dominated by memory for what had been learned in the previous phase, suggesting cognitive flexibility from working memory use.

One finding of note from our work is the similarity of performance between 5-year-old children and adults. Given clear developmental improvement in working memory capacity (Gathercole et al., 2004), such similarity may seem surprising. However, the demands on working memory were arguably low in this experiment compared to tasks in which 5-year-old children have performed poorly (Hongwanishkul et al., 2005).

Recall that in the advanced version of the DCCS task, for example, participants must shift between tasks on a trial-by-trial basis, with the sorting dimension being signaled by contextual cues. To succeed, participants had to keep in working memory the mappings between contextual cues and dimension rules, as well as between specific values of those dimensions and associated responses. In contrast, in the current task, participants needed to remember only the current stimulus-response mappings.

Given that working memory demands were minimized, the difference in the pattern of performance between pigeons and humans is especially striking. Although we did not observe differences in performance between 5-year-olds and adults, we expect that humans with more reduced working memory (such as younger children or patients with lesions to the frontal lobes), would perform similarly to pigeons, with large costs of response shifts and no reduction of this cost across multiple shifts.

Much recent work on avian cognition has emphasized that pigeons do not possess a cortex, and yet they demonstrate impressive cognitive abilities (Castro & Wasserman, 2016; Güntürkün & Bugnyar, 2016). Pigeons do possess a structure, the nidopallium caudolaterale (NCL), that many have suggested is analogous to prefrontal cortex (Güntürkün, 2005); however, it is unclear what consequences may arise from the different organization of avian and mammalian brains. One interesting possibility is that prefrontal cortex allows stronger working memory representations that in turn allow greater flexibility in overcoming long-term memory representations. Clearly, this speculative possibility requires future research, but the nature and consequences of neural differences between species is an exciting avenue of future work.

5.5. Limitations

A potential limitation of this work is the presence of task differences between the species. We attempted to make the variants of the task presented to pigeons and to humans as similar as possible, but some differences that could have affected the results remained. One difference between the tasks was that pigeons were extensively trained, with each phase being learned across hundreds of trials over several days, whereas children and adults completed the task in a single session with relatively few trials in each phase. Perhaps pigeons were more prone to perseveration simply because they formed very stable representations of the contingencies in each phase, whereas humans formed less stable memories and as a result showed less perseveration. Nonetheless, the differential amount of training that pigeons received (because each

pigeon was trained as many days as needed to reach the learning criterion) did not predict post-shift accuracy or the magnitude of perseveration in pigeons (see Fig. S2 in Supplementary Materials), and a subsample of children who demonstrated high accuracy in Phase 1 (i.e., they were effectively trained to the same criterion as the pigeons) showed the same pattern of results as the whole sample, with marked qualitative differences compared to pigeons' performance (see Fig. S3 in Supplementary Materials). It seems unlikely, therefore, that differences in perseveration between species were due to differences in the amount or effectiveness of training.

Another potentially important factor to consider is that the stimuli were familiar and easily nameable for human children and adults, but not pigeons. However, a control experiment conducted with adults using abstract and novel stimuli resulted in the same pattern of findings, suggesting that species differences in how the stimuli were processed did not affect differences in cognitive flexibility (see Fig. S5 in Supplementary Materials). In addition, other task differences could have also affected the results. For example, humans received different forms of feedback than pigeons, including verbal feedback, and it is possible that these differences may have affected perseveration. Although such feedback provided explicit rules to human children and adults, we think it is unlikely to have played a large role, given evidence from other studies that children struggle to benefit from explicit rules when learning complex contingencies (Napolitano & Sloutsky, 2004; Robinson & Sloutsky, 2004; Sloutsky & Fisher, 2008).

Of course, it would be impossible to fully equate procedures between species, but our methods allowed an important comparison that mitigates these concerns. Specifically, we included object pairs that were unique across phases and, as a result, presumably did not cause memory competition. Notably, performance for unique pairs was qualitatively similar across species, suggesting that the striking differences observed in performance for overlapping pairs were unlikely to stem from differences in general task demands.

5.6. Conclusions

In this study, we used a comparative approach to study the effect of long-term memory on cognitive flexibility. We found clear evidence of associative learning and long-term memory in pigeons as well as human children and adults. When responses to the same stimuli changed, pigeons greatly struggled, suggesting that their memory for previous contingencies overwhelmed their working memory representations for new contingencies. However, human participants readily learned to overcome the influence of long-term memory, rapidly and flexibly changing responses following a second shift. These findings are readily interpreted in light of memory-based theories of cognitive flexibility. They further suggest that, although pigeons can demonstrate highly flexible behavior when provided with extensive training involving rapid and repeated shifts in stimulus-response contingencies under contextual stimulus control, they struggle to do so without such training. Therefore, one important difference between avian species and humans may be the ability to use working memory to rapidly adjust responses despite the influence of long-term memory for previous behaviors.

Acknowledgements

This research is supported by NIH grant R01HD078545 to VMS and by NIH P01HD080679 grant to VMS and EAW.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.cognition.2018.03.015>.

References

- Berg, E. A. (1948). A simple objective technique for measuring flexibility in thinking. *The Journal of General Psychology*, 39, 15–22. <http://dx.doi.org/10.1080/00221309.1948.9918159>.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10(4), 433–436. <http://dx.doi.org/10.1163/156856897X00357>.
- Castro, L., & Wasserman, E. A. (2016). Executive control and task switching in pigeons. *Cognition*, 146, 121–135. <http://dx.doi.org/10.1016/j.cognition.2015.07.014>.
- Cook, R. G., Levison, D. G., Gillett, S. R., & Blaisdell, A. P. (2005). Capacity and limits of associative memory in pigeons. *Psychonomic Bulletin & Review*, 12(2), 350–358. <http://dx.doi.org/10.3758/BF03196384>.
- Cook, R. G., & Rosen, H. A. (2010). Temporal control of internal states in pigeons. *Psychonomic Bulletin & Review*, 17(6), 915–922. <http://dx.doi.org/10.3758/PBR.17.6.915>.
- Darby, K. P., & Sloutsky, V. M. (2015a). The cost of learning: Interference effects in memory development. *Journal of Experimental Psychology: General*, 144(2), 410–431. <http://dx.doi.org/10.1037/xge0000051>.
- Darby, K. P., & Sloutsky, V. M. (2015b). When delays improve memory: Stabilizing memory in children may require time. *Psychological Science*, 26(12), 1937–1946. <http://dx.doi.org/10.1177/0956797615607350>.
- Diekamp, B., Kalt, T., & Güntürkün, O. (2002). Working memory neurons in pigeons. *The Journal of Neuroscience*, 22(4), RC210 <https://doi.org/Artn Rc210>.
- Emery, N. J. (2006). Cognitive ornithology: The evolution of avian intelligence. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 361(1465), 23–43. <http://dx.doi.org/10.1098/rstb.2005.1736>.
- Fagot, J., & Cook, R. G. (2006). Evidence for large long-term memory capacities in baboons and pigeons and its implications for learning and the evolution of cognition. *Proceedings of the National Academy of Sciences of the United States of America*, 103(46), 17564–17567. <http://dx.doi.org/10.1073/pnas.0605184103>.
- Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology*, 40(2), 177–190. <http://dx.doi.org/10.1037/0012-1649.40.2.177>.
- 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, and Computers*, 36(2), 355–362. <http://dx.doi.org/10.3758/BF03195582>.
- Gibson, B. M., Wasserman, E. A., Gosselin, F., & Schyns, P. G. (2005). Applying bubbles to localize features that control pigeons' visual discrimination behavior. *Journal of Experimental Psychology: Animal Behavior Processes*, 31(3), 376–382. <http://dx.doi.org/10.1037/0097-7403.31.3.376>.
- Gibson, B. M., Wasserman, E., & Luck, S. J. (2011). Qualitative similarities in the visual short-term memory of pigeons and people. *Psychonomic Bulletin & Review*, 18(5), 979–984. <http://dx.doi.org/10.3758/s13423-011-0132-7>.
- Güntürkün, O. (2005). The avian "prefrontal cortex" and cognition. *Current Opinion in Neurobiology*, 15, 686–693. <http://dx.doi.org/10.1016/j.conb.2005.10.003>.
- Güntürkün, O., & Bugnyar, T. (2016). Cognition without cortex. *Trends in Cognitive Sciences*, 20(4), 291–303.
- Hongwanishkul, D., Happaney, K. R., Lee, W. S. C., & Zelazo, P. D. (2005). Assessment of hot and cool executive function in young children: Age-related changes and individual differences. *Developmental Neuropsychology*, 28, 617–644. <http://dx.doi.org/10.1207/s15326942dn2802.4>.
- Jordan, P. L., & Morton, J. B. (2012). Perseveration and the status of 3-year-olds' knowledge in a card-sorting task: Evidence from studies involving congruent flankers. *Journal of Experimental Child Psychology*, 111(1), 52–64. <http://dx.doi.org/10.1016/j.jecp.2011.06.011>.
- Kehagia, A. A., Murray, G. K., & Robbins, T. W. (2010). Learning and cognitive flexibility: Frontostriatal function and monoaminergic modulation. *Current Opinion in Neurobiology*, 20(2), 199–204. <http://dx.doi.org/10.1016/j.conb.2010.01.007>.
- Kendler, H. H., & Kendler, T. S. (1969). Reversal-shift behavior: Some basic issues. *Psychological Bulletin*, 72, 229–332.
- Kirkham, N. Z., Cruess, L., & Diamond, A. (2003). Helping children apply their knowledge to their behavior on a dimension-switching task. *Developmental Science*, 6(5), 449–467. <http://dx.doi.org/10.1111/1467-7687.00300>.
- Kirsch, J. A., Güntürkün, O., & Rose, J. (2008). Insight without cortex: Lessons from the avian brain. *Consciousness and Cognition*, 17(2), 475–483. <http://dx.doi.org/10.1016/j.concog.2008.03.018>.
- Lazareva, O. F., & Wasserman, E. A. (2016). No evidence for feature binding by pigeons in a change detection task. *Behavioural Processes*, 123, 90–106. <http://dx.doi.org/10.1016/j.beproc.2015.09.007>.
- Lazareva, O. F., & Wasserman, E. A. (2010). Category learning and concept learning in birds. In D. Mareschal, P. C. Quinn, & S. E. G. Lea (Eds.). *The making of human concepts*. Oxford: Oxford University Press.
- Levenson, R. M., Krupinski, E. A., Navarro, V. M., & Wasserman, E. A. (2015). Pigeons (*Columba livia*) as trainable observers of pathology and radiology breast cancer images. *PLoS One*, 10(11), 1–21. <http://dx.doi.org/10.1371/journal.pone.0141357>.
- Mathôt, S., Schrei, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, 44(2), 314–324. <http://dx.doi.org/10.3758/s13428-011-0168-7>.
- McAlonan, K., & Brown, V. J. (2003). Orbital prefrontal cortex mediates reversal learning and not attentional set shifting in the rat. *Behavioural Brain Research*, 146(1–2), 97–103. <http://dx.doi.org/10.1016/j.bbr.2003.09.019>.
- Morton, J., & Munakata, Y. (2002). Active versus latent representations: A neural network model of perseveration, dissociation, and decalage. *Developmental Psychobiology*, 40(3), 255–265. <http://dx.doi.org/10.1002/dev.10033>.

- Müller, U., Dick, A. S., Gela, K., Overton, W. F., & Zelazo, P. D. (2006). The role of negative priming in preschoolers' flexible rule use on the dimensional change card sort task. *Child Development*, 77(2), 395–412. <http://dx.doi.org/10.1111/j.1467-8624.2006.00878.x>.
- Napolitano, A. C., & Sloutsky, V. M. (2004). Is a picture worth a thousand words? The flexible nature of modality dominance in young children. *Child Development*, 75(6), 1850–1870. <http://dx.doi.org/10.1111/j.1467-8624.2004.00821.x>.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10, 437–442. <http://dx.doi.org/10.1163/156856897X00366>.
- Robinson, A. L., Heaton, R. K., Lehman, R. A., & Stilson, D. W. (1980). The utility of the Wisconsin Card Sorting Test in detecting and localizing frontal lobe lesions. *Journal of Consulting and Clinical Psychology*, 48(5), 605–614. <http://dx.doi.org/10.1037/0022-006X.48.5.605>.
- Robinson, C. W., & Sloutsky, V. M. (2004). Auditory dominance and its change in the course of development. *Child Development*, 75, 1387–1401. <http://dx.doi.org/10.1111/j.1467-8624.2004.00747.x>.
- Rogers, R. D., Andrews, T. C., Grasby, P. M., Brooks, D. J., & Robbins, T. W. (2000). Contrasting cortical and subcortical activations produced by attentional-set shifting and reversal learning in humans. *Journal of Cognitive Neuroscience*, 12(1), 142–162. <http://dx.doi.org/10.1162/089892900561931>.
- Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, 84(1), 1–66. <http://dx.doi.org/10.1037/0033-295X.84.1.1>.
- Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, 84, 127–190. <http://dx.doi.org/10.1037/0033-295X.84.2.127>.
- Sloutsky, V. M., & Fisher, A. V. (2008). Attentional learning and flexible induction: How mundane mechanisms give rise to smart behaviors. *Child Development*, 79(3), 639–651. <http://dx.doi.org/10.1111/j.1467-8624.2008.01148.x>.
- Soto, F. A., & Wasserman, E. A. (2014). Mechanisms of object recognition: What we have learned from pigeons. *Frontiers in Neural Circuits*, 8, 1–22. <http://dx.doi.org/10.3389/fncir.2014.00122>.
- Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 27(1), 92–114. <http://dx.doi.org/10.1037/0096-1523.27.1.92>.
- Wasserman, E. A. (2016). Conceptualization in pigeons: The evolution of a paradigm. *Behavioural Processes*, 123, 4–14. <http://dx.doi.org/10.1016/j.beproc.2015.09.010>.
- Wasserman, E. A., Brooks, D. I., & McMurray, B. (2015). Pigeons acquire multiple categories in parallel via associative learning: A parallel to human word learning? *Cognition*, 136, 99–122. <http://dx.doi.org/10.1016/j.cognition.2014.11.020>.
- Wasserman, E. A., & Young, M. E. (2010). Same-different discrimination: The keel and backbone of thought and reasoning. *Journal of Experimental Psychology: Animal Behavior Processes*, 36(1), 3–22.
- Watanabe, S., Sakamoto, J., & Wakita, M. (1995). Pigeons' discrimination of paintings by Monet and Picasso. *Journal of the Experimental Analysis of Behavior*, 63(2), 165–174. <http://dx.doi.org/10.1901/jeab.1995.63-165>.
- Wright, A., Santiago, H., Sands, S., Kendrick, D., & Cook, R. (1985). Memory processing of serial lists by pigeons, monkeys, and people. *Science*, 229(4710), 287–289. <http://dx.doi.org/10.1126/science.9304205>.
- Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): A method of assessing executive function in children. *Nature Protocols*, 1(1), 297–301.
- Zelazo, P. D., Müller, U., Frye, D., Marcovitch, S., Argitis, G., Boseovski, J., ... Sutherland, A. (2003). The development of executive function in early childhood. *Monographs of the Society for Research in Child Development*, 68(3) vii–137.
- Zentall, T. R., Clement, T. S., Bhatt, R. S., & Allen, J. (2001). Episodic-like memory in pigeons. *Psychonomic Bulletin & Review*, 8(4), 685–690. <http://dx.doi.org/10.3758/BF03196204>.