Cognitive offloading by children in perceptual discrimination tasks☆

Audrey E. Parrish[a,*], Maisy D. Englund[b], Andrew J. Kelly[c], Bonnie M. Perdue[d], Alexandra H. Daley[a], Charlotte K. Welsh[a], Michael J. Beran[c]

[a] Department of Psychology, The Citadel, USA

[b] Animals in Laboratories Program, Animal Welfare Institute, USA

[c] Department of Psychology, Georgia State University, USA

[d] Department of Psychology, Agnes Scott College, USA

Author Note: This research was supported by Grant BCS2043667 from the National Science Foundation and a research grant via the School of Humanities and Social Science at The Citadel. The authors thank the directors and staff at The Children's Museum of the Low country in Charleston, SC, The Charleston Catholic School in Charleston, SC, and Suburban Nursery School and Pre-K in Decatur, GA, as well as the parents of the children for their assistance and support.

* Correspondence to: 171 Moultrie St., Charleston, SC 29409, USA.

E-mail address: audrey.parrish1@gmail.com (A.E. Parrish).

Abstract

Cognitive offloading occurs when an individual modifies a current decision scenario in a way that reduces the cognitive load or difficulty of a task. Children begin to engage in such offloading even before formal schooling begins. Using a manual rotation paradigm, preschool and elementary school children (3- to 9-years-old) were given perceptual discrimination tasks in which they had to compare two visual stimuli (either vertical and horizontal lines that intersected and they had to determine which was longer, or rectangular shapes or clip art animals that they had to compare to determine if the stimuli were the same or different). On some trials, offloading to the environment via rotation of one stimulus was beneficial to make the discrimination easier from the perspective of those stimuli aligning. Children in all age groups showed rotation of the various stimuli to make the task easier, although there was a developmental trend such that likelihood of accuracy and rotation increased with age. Additionally, children were more likely to rotate objects on difficult trials than easier ones and this often resulted in increases in accuracy. This tendency to rotate for the more difficult trials was associated with age. These results confirm that children can manipulate stimuli in ways that make comparing those stimuli easier, reflecting a form of (meta)cognitive offloading using the external environment to resolve internal uncertainty.

Keywords: cognitive offloading, metacognition, cognitive development, rotation tasks, same-different judgments

Cognitive Offloading by Children in Perceptual Discrimination Tasks

Cognitive offloading refers to the use of external aids or physical actions to alleviate internal cognitive demands or to resolve uncertainty associated with a given task (Risko & Gilbert, 2016). For example, the use of memory aids (e.g., to-do lists, calendar reminders), internet search engines, maps, calculators, and even tilting one's head to interpret a rotated image are all examples of offloading routinely seen in adult humans. This ability is tied to metacognitive monitoring and control, such that use of offloading strategies is influenced by evaluations of one's own mental capabilities and limitations (e.g., Boldt & Gilbert, 2019; Dunn & Risko, 2016; Jolicoeur, 1988; Risko et al., 2014; Risko & Dunn, 2015). For example, engagement of a GPS navigational system is influenced by awareness of one's limited spatial memory for an intended route (e.g., Gardony et al., 2015; Risko & Gilbert, 2016). Cognitive offloading, such as setting reminders for to-be-recalled information, varies as a function of how valuable the information is perceived to be (e.g., Dupont et. al., 2023; Murphy, 2023) as well as how much information must be remembered and anticipation of future distraction (e.g., Gilbert, 2015), suggesting that such behavior is selective and flexible.

Children engage in help-seeking behavior that may reflect cognitive offloading. Infants will seek help from adults or refuse to respond when they are unsure where an item is hidden (e.g., Goupil et al., 2016). Coughlin et al. (2015) reported that children as young as 3-years-old monitored their uncertainty when asked to identify targets in a perceptual identification task. When targets were degraded images, the children increased their rates of asking for help, particularly for items marked incorrectly or that they verbally noted as uncertain, and they responded more quickly following assistance from a knowledgeable helper versus a less accurate helper. Similarly, Beran et al. (2019) reported that preschool children assigned the most difficult part of task to a symbolic helper in an offloading task and this strategic offloading generalized to different kinds of tasks (e.g., memory, counting, object identification, and word reading), although older children were more proficient in doing so for some stimuli. Thompson et al. (2012) reported that preschool children (mean age 4.22 years) increased requests for help as a function of how long they had been engaged in solving a puzzle, and that girls were more likely to seek help than boys despite similar performance levels. This effect was most evident among older preschool girls. In another test, children as young as four years old were presented with a spatial memory task where stickers were hidden in an array of cups (Bulley et al., 2020). They were more likely to pre-mark locations with tokens placed on top of the cups when the task was objectively harder (i.e., there were more locations to remember). In a second experiment, Bulley and colleagues modified the memory task to include a delay (30 s) between when the target was hidden and the retrieval phase. Children were provided with a pen to mark the cups (versus tokens) to explore whether children would spontaneously offload by marking/drawing on target cups. Children under the age of 10 rarely self-generated such an offloading strategy unless they were given a prompt. These results collectively highlight children's ability to offload difficult discriminations or stimuli to be processed and the effect of age on spontaneous generation of such offloading strategies.

Armitage et al. (2020) expanded on Shepard and Metzler's classic 1971 mental rotation task to assess developmental trends in offloading among children aged 4- to 11-years-old using a same-different discrimination that included two images of a human figure with either their right or left arm facing upwards. One figure was presented upright while the other was rotated at varying degrees (0° to 180°), affecting the objective difficulty of the discrimination. It was predicted that the same-different discrimination would be more difficult for increasingly larger differences in starting orientation between stimuli, and such trials should generate more offloading. For the offloading component, the rotated image was positioned on a turntable so that it could be physically rotated to align with the upright figure, and this was demonstrated to children in pre-trials. For all children, there was a linear relationship between starting angle and offloading via rotation as predicted. Six- and 7-year-olds were more selective in offloading via manual rotation than 4- and 5-year-olds, yet rotation decreased in the older age groups, likely due to ceiling effects in performance. In a follow-up experiment, children were more likely to rotate the turntable when it was useful for the discrimination (i.e., in an experimental counting condition versus a control color discrimination condition) and when the sheets were inverted versus upright. As age increased, children were more selective in their rotation behavior (i.e., they were more likely to rotate the images in the inverted experimental condition versus the inverted control condition). These results demonstrated a developmental increase in the use of selective offloading for children across the age range that was tested.

In a follow-up study, Armitage and Redshaw (2022) reported a developmental trend in which children's ability to spontaneously implement the rotation offloading strategy (i.e., with no experimenter instruction or demonstration) increased with age (from 4- to 11-years-old) when presented with misaligned maps. Specifically, from the age of 6 years and older, children's use of the manual rotation strategy improved task performance, whereas 4- and 5-year-olds showed little spontaneous use of the offloading rotation strategy. These results aligned with previous studies demonstrating limitations in 4- and 5-year-olds offloading performance (e.g., Armitage et al., 2020; Bremner & Andreasen, 1998; Bulley et al., 2020; Vosmik & Presson, 2004). Similarly, a developmental trend in the ability to recreate a rotated image was observed across 5-, 7-, and 9-year-olds if physical rotation of the image was prevented, suggesting that mental rotation is a high-level cognitive ability related to spatial processing that emerges late in development (Lange-Küttner & Green, 2007).

Current Study

Based in part on the design of Armitage et al. (2020), we assessed the ability of children to physically transform a test situation to aid performance. The current study serves as a conceptual replication and extension of their results with children to novel perceptual discrimination tasks, and we included younger children (3-year-olds) to determine whether this ability to offload will transfer to novel tasks and extend to younger ages. This is important to see if the ability to offload, and the age trends seen previously will extend to different tasks. In addition, our broader research program takes a comparative perspective, comparing the performance of humans across development with that of nonhuman primates, and the designed task used here will be adapted for use with nonhuman primates. Children 3 to 9 years of age (total *N* across experiments = 227) were given the opportunity to manually rotate stimuli to decrease cognitive demands of a line-length discrimination task (Experiment 1) as well as two same-different discrimination tasks (Experiment 2 and Experiment 3).

In Experiment 1, children were tasked with reporting which of two lines (horizontal or vertical) was longer in the test condition. Lines were positioned perpendicularly so that they formed an inverted letter ‘T,’ reflective of the vertical-horizontal illusion in which individuals, including children, overestimate the length of a vertical line that intersects a horizontal line of equal length (Brosvic et al., 1993, 2002; Fick, 1851). Children could offload by rotating the vertical line so that it was parallel to the horizontal line, assisting with line-length discrimination performance. Test performance was compared to a control condition, in which children were presented with the same set-up (inverted T formed by a horizonal and vertical line of varying lengths); however, they were asked to report which line was darker, and thus the offloading rotation behavior would not assist with discrimination performance. We predicted that children would be more likely to rotate the lines in test versus control trials, and more so for test trials in which the lines were closer in *subjectively experienced* length (i.e., the objectively more difficult discriminations because of the illusory error). We predicted an age effect such that older children would be more likely to rotate lines in the test trials, particularly for the more difficult discriminations.

In Experiment 2 and Experiment 3, children were asked to report whether two cards contained the same image (of either black squares on a grid – Experiment 2 or clip-art images of animals – Experiment 3) in the test condition. In these test conditions, rotation of the cards so that they aligned would assist with same-different discrimination performance, reflective of offloading. However, rotation of the cards for the control condition in Experiment 2, in which participants were asked to report which of two cards contained more black squares would not assist with discrimination performance because counting or estimating quantity is not aided by any particular orientation of the stimuli. We predicted that children would be more likely to rotate stimuli in test versus control trials, and more so for test trials with objectively more difficult discriminations (i.e., for more minute differences in animal images for Experiment 2). We also predicted an age effect such that older children would be more likely to engage in cognitive offloading (rotation) than younger children, strategically rotating more so in test trials (Experiment 2) and more so for the most difficult discriminations (Experiment 3). We expected that rotation of control trials would be low, in general, and not related to age.

Experiment 1

Method

Participants

Participants were recruited from an early childhood education center, an elementary school, and a children's museum. A total of 100 children participated in Experiment 1, including 56 females and 44 males. Age was calculated based on date of birth and date of testing. Some parents/guardians provided only the birth month and year for their child, and, in these cases (*N* = 30) age was calculated based on the first day of the birth month. The average age of the sample was 75 months (*SD* = 20), including 27 children 3- to 4-years-old, 31 children 5- to 6-years-old, and 35 children 7- to 9-years-old. A total of seven parents/guardians did not include birth date information on the optional demographics form, but these participants were within the age range tested (3- to 9-years-old). These seven participants were excluded from all analyses that included age as a variable. The sample was largely Caucasian (*N* = 87), whereas remaining children were identified as African American (*N* = 4), Asian/Asian American (N = 3), Hispanic/Latino (*N* = 1), Pacific Islander (*N* = 1), and multi-racial (*N* = 3). One consent form did not include this information. For all experiments, study protocols were approved via the institutions' Review Boards as well as through the directorship of the schools and museum. Parental permission was obtained, and the children provided assent for participation prior to the study.

Materials

The manual rotation task consisted of two sets of 8.5” x 11” laminated sheets depicting a printed horizonal line and a vertical line attached to the sheet via a brad fastener, so that it intersected with the horizontal line in an inverted letter T formation. The horizontal line was always 7” long, and the vertical line differed in length for the Test Lines condition. The lines differed in color (grayscale) for the Control Lines condition. Children were instructed that they could rotate the vertical line so that it was parallel to the horizontal line if it made the task easier (i.e., to decide which line was darker/longer). Each condition is described in detail below and depicted in Figure 1. Children were provided a small reward for their participation following study completion (e.g., small plush animals, fidget spinners, pop-it keychains).

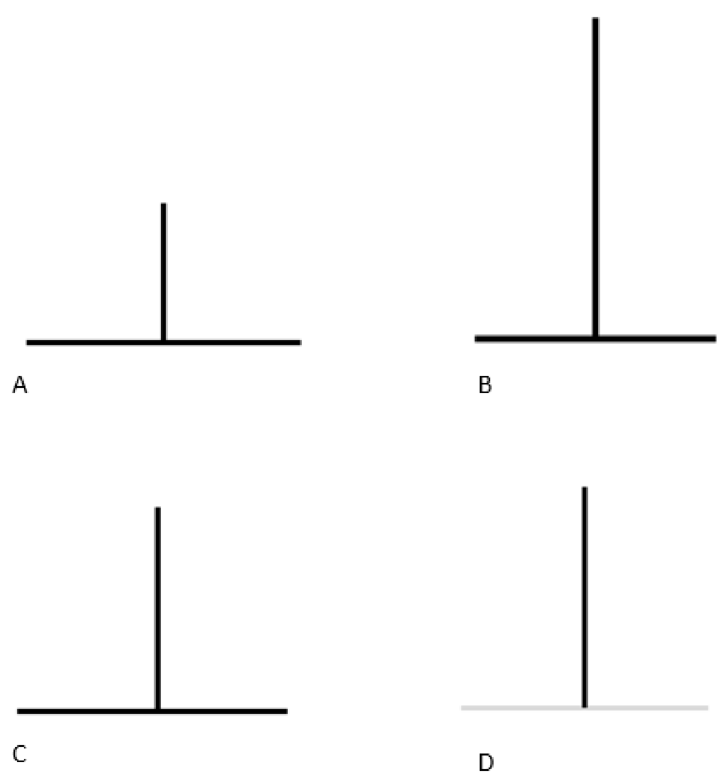

Figure 1. Experiment 1 sample trials. (A) shows a trial with a longer horizontal line. (B) shows a trial with a longer vertical line. (C) shows an illusory trial, in which the vertical line is shorter, but typically appears to most people to be longer. (D) shows a control trial, where the darker line has to be indicated.

Design and Procedure

Using a within-subjects design, participants completed both conditions – Control Lines and Test Lines – randomized for order. For both conditions, participants were shown a laminated sheet of paper with a printed horizontal line and the vertical line attached to the paper so that it formed an inverted letter T. Children were instructed to identify the darker line (Control Lines) or the longer line (Test Lines). In both conditions, children were instructed that they could rotate the vertical line so that it was parallel to the horizontal line if it made the task easier (i.e., to decide which line was darker/longer). If children responded that the lines were the same, the experimenter prompted the child to choose one of the two lines. Despite this prompt, some children maintained the response of "same." This happened rarely (in 0.23% of trials), and these trials were excluded from the analyses.

In both conditions, participants were given a practice trial in which they had to perform the discrimination (Control Lines – indicate which line is darker; Test Lines – indicate which line is longer). Participants were shown how to rotate the vertical line so that it was parallel to the horizontal line, and they were required to practice the rotation. For both conditions, children were given a verbal reminder that they could rotate the vertical line after 10 trials of no rotation (if that occurred). Children were given general feedback regarding effort but no performance-based feedback (e.g., "Good, keep going."). After completion of the first condition, children immediately began the next condition (randomized for condition order), and the instructions were repeated for the new condition (Control Lines or Test Lines).

For the Control Lines condition, children completed 12 trials. The objective was to determine which line was darker. One line was printed in black, and the alternative line varied in color opacity so that it appeared light to dark gray. The opacity of the alternative line varied from 20% (dark gray – most difficult discrimination) to 80% (light gray – easiest discrimination). The horizontal line was always 7” in length and the vertical line varied in length from 3” to 8”, including one trial of equal-length (7” vs. 7”). The width of the lines was held constant (.35”). We randomized trial order so that sometimes the horizontal line was black and sometimes the vertical line was black, and the alternative line varied in opacity. We also varied the length of the lines across trials so that sometimes the horizontal line was longer and sometimes the vertical line was longer. Children were shown a demonstration trial to practice the perceptual discrimination prompt and rotation instruction. This trial was not included in any of the analyses.

For the Test Lines condition, children completed 17 trials. The objective was to determine which line was longer, and thus both lines were black. Similar to the Control Lines condition, the horizontal line was always 7” in length and the vertical line varied in length from 3” to 8”, including two trials of equal length (7” vs. 7”). Note that these two illusory test trials (7” vs. 7”) were analyzed to determine whether the participants exhibited the classic illusion (i.e., identify the vertical line as longer despite its equal length with the horizontal line). The width of the lines was held constant (.35”). We randomized trial order so that sometimes the horizontal line was longer and sometimes the vertical line was longer. Children were shown a demonstration trial to practice the perceptual discrimination prompt and rotation instruction, which was not included in any of the analyses.

For both conditions, we recorded the children's response to the primary task (Control Lines – color discrimination; Test Lines – line-length discrimination). We also recorded whether the vertical line was rotated or not.

Data Analysis

Data analyses were conducted in SPSS (Version 28.0.1.0) or in R version 4.3.2 (R Core Team 2023) using the lme4 function (Bates et al., 2015). Adjusted predicted probabilities were computed using ggeffects (Lüdecke, 2018) and data visualizations were created using ggplot2 (Wickham, 2016) and ggeffects (Lüdecke, 2018). Data were analyzed via a series of generalized linear mixed-effects models (GLMMs) with binomial distributions (i.e., logistic regression), which are described in detail in the Results sections. For each experiment, we explored two outcome variables of interest: children's choice to rotate (did not rotate = 0, rotated = 1), and children's accuracy (incorrect = 0, correct = 1) on any given trial. Subject ID was included as a random effect in all models, except in two cases where including Subject ID led to a singular fit of the model. In those cases, Subject ID's variance was estimated at exactly 0, and therefore its effect was negligible on outcome measures. This was unsurprising considering our small sample size and our inclusion of age as a predictor. Therefore, in the cases of singular fit, Subject ID was removed from the model and a simple generalized linear model (GLM) was run in place of a generalized linear mixed-effects model (GLMM; Barr et al., 2013) – outcome estimates were unchanged by this exclusion (e.g., Pasch et al., 2013).

Specific fixed effects are described for each experiment. To reduce multicollinearity and improve interpretability of coefficients for lower order effects, binary categorical fixed effects were contrast coded and centered around zero (-1, 1), categorical fixed effects with three levels were also coded using effect contrast coding and continuous variables were centered and scaled using Z-score standardization (Aiken & West, 1991; UCLA: Statistical Consulting Group, n.d.). This means that the "choice to rotate" variable was coded differently depending on if it was a predictor variable (did not rotate = -1, rotated = 1) or if it was the outcome variable (did not rotate = 0, rotated = 1). In instances in which the Levene's test indicated unequal variances, we adjusted the degrees of freedom accordingly; these instances are described below.

Results

Five children did not pass the trial demonstration phase in which they were required to correctly identify the longer/darker line. These children were not included in the following analyses, leaving a total sample of 95 children. We first examined whether condition order impacted the likelihood to rotate in the experimental condition for lines with a true difference in length (Test Lines) excluding the two illusory test trials (7" vs 7") for each participant, which are analyzed separately below. A total of five trials (or 0.2% of these test trials with a true difference in line length) across all participants were excluded due to a response of "same" as the primary dependent variable was accuracy in reporting the longer line, thus a response of horizontal or vertical was necessary.

Because we had not recorded the condition that was completed first (test or control) due to experimenter error for some children (N = 22), this analysis was important to determine whether order affected rotation likelihood. A one-way between-subjects ANOVA was conducted to compare the overall percentage of trials in which participants rotated the line for those children in which condition order was not recorded (*N* = 22; *M* = 46.97%, *SD* = 37.91%), those who first completed the Control Lines condition (*N* = 31; *M* = 46.88%, *SD* = 33.71%), and those who first completed the Test Lines condition (*N* = 42; *M* = 56.63%, *SD* = 34.05%). There was not a significant effect of condition order on rotation behavior, *F*(2, 92) = .91, *p* = .407, η_p^2= .019.

We next assessed performance in the Control Lines condition. A total of 28 children (29.47%) rotated the line in this condition, and as predicted, accuracy in making the darkness judgment did not differ between children who did not rotate the line (*M* = 91.67%, *SD* = 18.58%) and those that rotated the line (*M* = 88.39%, *SD* = 19.42%), independent samples *t*(93) = .77, *p* = .442, Cohen's *d* = .17. In contrast, 88 children (92.63%) rotated the line in the Test condition. Accuracy in making the line length judgment in Test trials was significantly higher for children who rotated the line (*M* = 92.94%, *SD* = 7.17%) than for those who did not rotate the line (*M* = 79.05%, *SD* = 14.62%), independent samples *t*(6.23) = 2.5, *p* = .046, Hedges' *g* = 1.75. Note that Levene's test indicated unequal variances (*F* = 7.65, *p* = .007), so degrees of freedom were adjusted from 93 to 6.23, and Hedges' g was reported due to different sample sizes across children who rotated versus those who did not rotate.

For the illusory Test trials, (7" vertical versus 7" horizontal), we again excluded any trials (*N* = 4, or 2.11% of equal-length trials) in which children reported "same" for line length as the primary dependent variable to determine illusory susceptibility is which line is perceived as longer – horizontal or vertical. On average, children reported the vertical line as the longer line in 73.66% of those trials. A binomial test confirmed that choice for the vertical line over the horizontal line was significantly higher than 50%, $p < .001$, providing evidence for perception of the horizontal-vertical illusion in the current sample.

GLMM Results

Choice to Rotate. First, we explored the effect of age and trial difficulty on children's choice to rotate in the experimental Test Lines condition. For these analyses, we categorized trial difficulty into three levels: "*Easy Short*" (the vertical line was less than or equal to 5 inches in length), "*Easy Long*" (the vertical line was greater than 7.5 inches in length), and "*Hard*" (the vertical line was greater than 5 inches in length and shorter than 7.5 inches in length which encompasses the range for which the illusory experience makes these lines seem similar in length). We excluded the two illusory test trials for each participant (7" vs. 7") as there was no correct response for these equal-length trials. Additionally, we again excluded the test trials in which participants responded that the lines were the "same" (0.23%). Finally, participants who were missing birthdate/age information ($N = 7$) were excluded.

We conducted a GLMM that explored the interaction of trial difficulty (EasyShort, EasyLong, and Hard) and age on choice to rotate. The model revealed a significant two-way interaction between difficulty level and age ($\beta = 0.267$; $p = .020$; see Table 1, Supplementary Material). In general, the likelihood to rotate increased with both increasing age and higher difficulty level; the interaction effect can be explained by the S-shape curve representing hard trials (Figure 2). That is, in hard trials, the likelihood to rotate increased with increasing age and then begins to level off, representing a ceiling effect at older ages for hard trials that is absent or less pronounced for the easier trial types (see also Table 2, Supplementary Material).

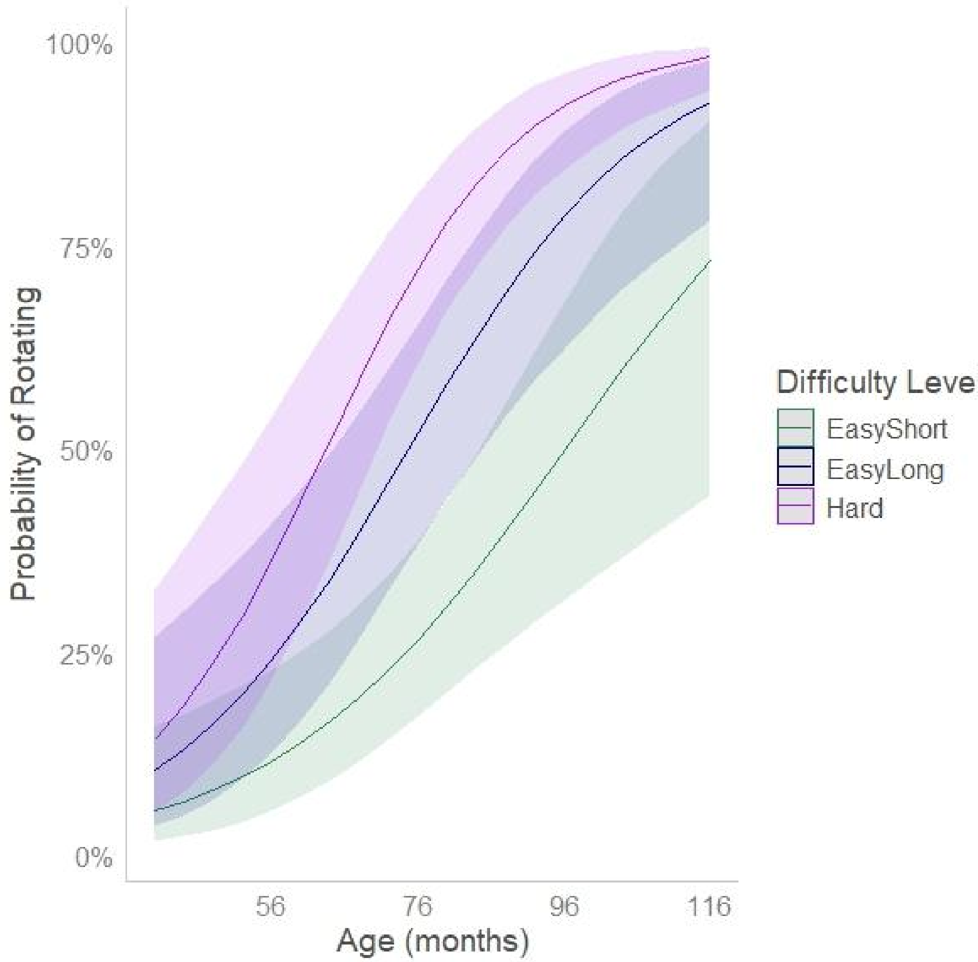

Figure 2. Probability of rotating as a function of age and difficulty level in Experiment 1 Test Lines condition.

Accuracy. Next, we explored the effect of age, difficulty level, and choice to rotate on children's accuracy. Subject ID was initially included as a random effect; however, the variance within Subject ID was negligible, causing a singular fit of the model. Consequently, we removed Subject ID, and changed the model from a GLMM to a GLM. This model revealed a significant three-way interaction ($\beta = 0.764$, $p = .035$ see Table 3, Supplementary Material). Figure 3 depicts the nature of this interaction: when children did not rotate, age differentially affected accuracy. All children were generally likely to be accurate for easy trials, but for hard trials, likelihood of accuracy increased with increasing age. However, when children did choose to rotate, this age effect decreases, and all children show higher overall probabilities of being accurate regardless of difficulty level (see Table 4, Supplementary Material).

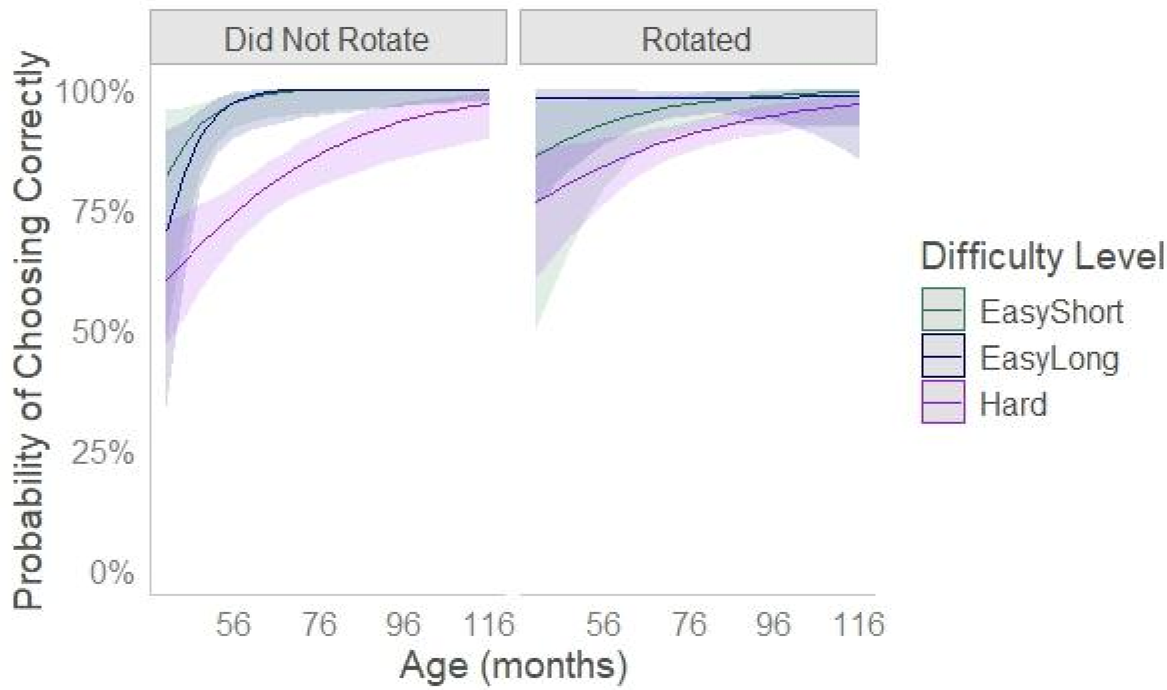

Figure 3. Probability of choosing correctly as a function of age, difficulty level, and the choice to rotate in Experiment 1 Test Lines condition.

Experiment 2

In Experiment 1, when children chose to rotate, the relative influence of age and difficulty level on accuracy was reduced. Experiment 2 sought to test the generalizability of these findings by using a new discrimination (same/different task) to differentiate novel stimuli (4 x 4 cards with gridded squares that were partially shaded) and that varied in their degree of starting orientation. Rotation of the cards so that they aligned would assist with same-different discrimination performance, reflective of offloading. We again compared performance to a control condition in which participants were asked to report which of two cards contained more black squares. Rotation would not assist with discrimination performance because counting or estimating quantity is not aided by any particular orientation of the stimuli. To accommodate COVID-19 restrictions, we switched testing venues from schools to a local museum, and subsequently employed a between-subjects design in order to limit testing sessions to approximately 10 minutes per participant.

Method

Participants

A new set of participants were recruited from a local children's museum. A total of 84 children participated in Experiment 2, including 44 males and 40 females. Age was calculated based on date of birth and date of testing. Some parents/guardians provided only the birth month and year for their child, and, in these cases (*N* = 26) age was calculated based on the first day of the birth month. The average age of the sample was 66 months (*SD* = 17), including 32 children 3- to 4-years-old, 35 children 5- to 6-years-old, and 12 children 7- to 8-years-old. A total of five parents/guardians did not include birth date information on the optional demographics form, but these participants were within the age range tested (3- to 9-years-old). These five participants were excluded from all analyses that included age as a variable. The sample was largely Caucasian (*N* = 72), followed by multi-racial (*N* = 8), African American (*N* = 3), and Hispanic/Latino (*N* = 1).

Materials

The manual rotation task consisted of two sets of approximately 4"x 4" laminated cards depicting images corresponding to each condition (described in detail below and depicted in

Figure 4). Children were provided a small reward for their participation following study completion.

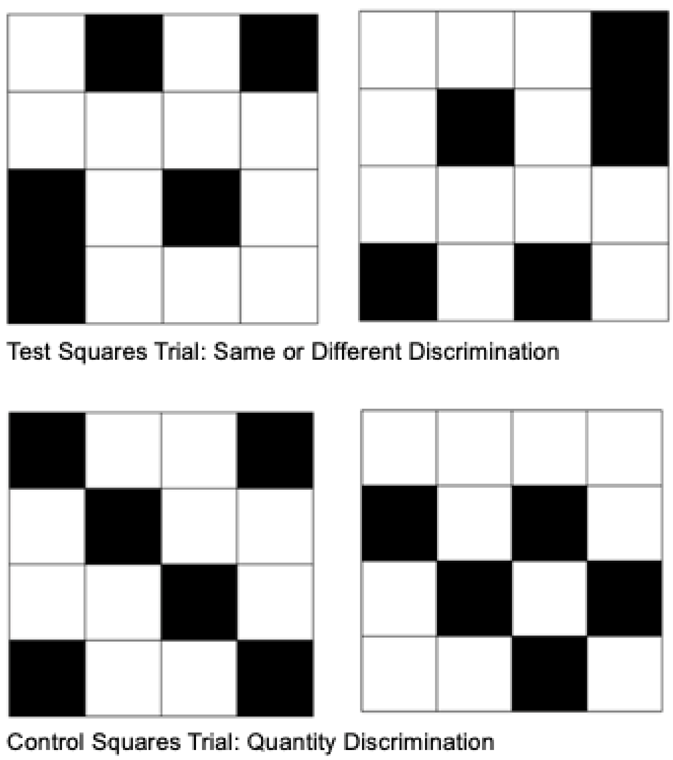

Figure 4. Experiment 2 sample trials. The top row shows a same or different discrimination trial for the Test Squares condition. In this case, the images are the same, but rotated 180 degrees. The bottom row shows a quantity discrimination trial for the Control Squares condition where the goal was to indicate which image had more black squares.

Design and Procedure

Using a between-subjects design, children participated in the Control Squares condition (*N* = 42) or the Test Squares condition (*N* = 42). Children were shown a pair of cards simultaneously that they were instructed to evaluate based on a particular prompt: Control Squares (‘which card has more black squares?’) or Test Squares (‘are these two cards the same or different?'). In each condition, children also were instructed that they could rotate a card if it made the task easier (i.e., to decide which card had more black squares or to decide if the cards were the same or different). Note that we asked about having more black squares to avoid any prompting of a specific way of answering that, such as by using counting, subitization, or estimating the total area of black squares.

In both conditions, participants were given a practice trial in which they had to perform the discrimination (Control Squares – indicate which card had more black squares; Test Squares – indicate whether the cards were the same or different). During this practice demonstration, participants were shown how to rotate the card positioned on the right so that it aligned with the card positioned on the left for the Test Squares or so that it was fully rotated in the Control Condition. In all conditions, children were given a verbal reminder that they could rotate the card on the right after every five trials if no rotation was performed by a child during those five trials. Children were given general feedback regarding effort but no performance-based feedback.

For the Control Squares condition, children completed 16 trials. Each trial presented a pair of cards with a 4x4 grid of white squares. Each card depicted between 2 to 7 of these 16 squares colored black, with the objective of determining which card in the pair contained more black squares. The quantitative difference between the left and right cards (in terms of the number of black squares) ranged from 1 to 3 (e.g., 5 vs. 4 black squares; 2 vs. 4 black squares; 3 vs. 6 black squares). We randomized left-right placement of the larger quantity card across trials. We did not vary the starting orientation of the cards, as the cards always differed in the number of black squares and their placement within the 4x4 grid, thus the starting orientation would inherently never align. Children were shown a demonstration trial to practice the perceptual discrimination prompt and rotation instruction. All children accurately indicated the card with a greater quantity of black squares during this practice trial. This trial was not included in the analyses.

For the Test Squares condition, children completed 23 trials. Each trial again presented a pair of cards with a 4x4 grid of white squares. Each card depicted between 2 to 7 of these 16 squares colored black, with the objective of determining whether the cards depicted the same or different shape placements. Eleven of the trials were “Same Pattern” and twelve were “Different Pattern” The starting orientation of the “Same Pattern” trials varied from 0° to 270° across trials, such that the card placed on the right was presented in the same orientation as the left card (0°), one rotation to the right (90°), opposite vertical orientation (180°), or one rotation to the left (270°). There were two to three trials of each starting rotation for the Same Pattern trials (0°: 3 trials, 90°: 3 trials, 180°: 3 trials, 270°: 2 trials). As with the Control Squares condition, we did not vary the starting orientation of the Different Pattern trials, as the cards always differed in their patterns, thus the starting orientation would inherently never align. We randomized trials for same-different pattern and starting orientation. Children were shown a demonstration trial to practice the perceptual discrimination prompt and rotation instruction. All children accurately indicated the cards were the same on this practice trial. This trial was not included in the analyses.

For both conditions, we recorded the children’s response to the primary task (quantity discrimination or same-different discrimination). We also recorded whether the righthand card was rotated.

Results

In the Control Squares condition, only 9 of 42 children ever rotated an image, and five of those children only did so on one trial. Overall, an image was rotated in only 2.23% of the trials in this condition. Children chose the correct square on 89.58% of the Control trials. When the difference in quantity between images was three items, performance was 100%, for the difference of two items, performance was 94.22%, and when the difference was only one item, performance was 81.97%. In contrast, 36 of 42 children rotated images in the Test Squares condition with squares. Overall, children rotated an image on 18.84% of trials. This was a significantly higher percentage of rotations than in the Control condition, $t(51.72) = 7.19$, $p < .001$, Cohen's $d = 1.57$. Note that Levene's test indicated unequal variances ($F = 62.27$, $p < .001$), so degrees of freedom were adjusted from 82 to 51.72. Accuracy in making the same/different discrimination was higher for children who rotated the card ($M = 86.96\%$, $SD = 12.98\%$) than for those who did not rotate ($M = 77.54\%$, $SD = 11.48\%$), but this difference was not significant (independent samples $t(40) = 1.67$, $p = .103$, Cohen's $d = .77$. A closer examination of specific trial types showed the importance of rotating stimuli.

GLMM Results

In the Test Squares condition, children compared images that either consisted of the same pattern presented in the same starting orientation, the same pattern presented in different starting orientations, or two different patterns. We investigated children's likelihood to rotate across these conditions as well as their likelihood to be accurate in identifying whether the images were the same pattern or different patterns. As before, we did not include participants with missing birthdate/age information ($N = 5$).

Choice to Rotate. First, we explored children's likelihood to rotate as a function of trial type (Same Pattern Same Orientation: hereafter SPSO; Same Pattern Different Orientation: hereafter SPDO; Different Pattern: hereafter DP) and age. This model revealed a significant interaction between age and trial type (see Table 5 and Table 6, Supplementary Material). Children of all ages were exhibiting floor effects for likelihood to rotate in trials where the image was the same pattern and the same orientation (SPSO trials) and in trials where the image was an entirely different pattern (DP trials), indicating that these trials were sufficiently easy that rotations were not needed. Children were significantly more likely to rotate in trials where the image was the same pattern but presented at a different starting orientation (SPDO trials; β = 2.362; p <.001), and as age increased, so did the likelihood of rotating in this condition (β = 0.596, p = .027). Across the 468 total DP trials, there were only 33 instances of rotating (7%); across the 117 SPSO trials, there were only 3 instances of rotating (2%). In contrast, out of 312 total SPDO trials, children rotated 136 times (44%).

Accuracy. A ceiling effect was observed for the accuracy data in the DP and SPSO conditions. Only 27 incorrect choices were made out of 468 DP trials (6%), and only 9 incorrect choices were made out of 117 SPSO trials (8%). Again, there was greater outcome variability for SPDO trials; in those trials, 88 incorrect choices were made out of 312 trials (28%).

The lack of outcome variability in two of the three trial types caused overfitting of the model when attempting to look at 3-way interaction effects between trial type, age, and choice-to-rotate on accuracy. Because children were overall unlikely to be inaccurate in DP and SPSO conditions, we decided to explore age and choice-to-rotate effects in only the SPDO condition. We conducted a logistic GLMM with age and choice to rotate on accuracy. The model revealed that the main effects of age (β = 1.278, p <.001) and choice to rotate (β = 1.147, p < .001) were significant, but no significant interaction effect was observed (β = -0.175, p = .466; see Table 7 and Table 8, Supplementary Material). Children were more likely to be correct in trials where they rotated the images than when they did not rotate the images, and likelihood of accuracy increased with age (Figure 5).

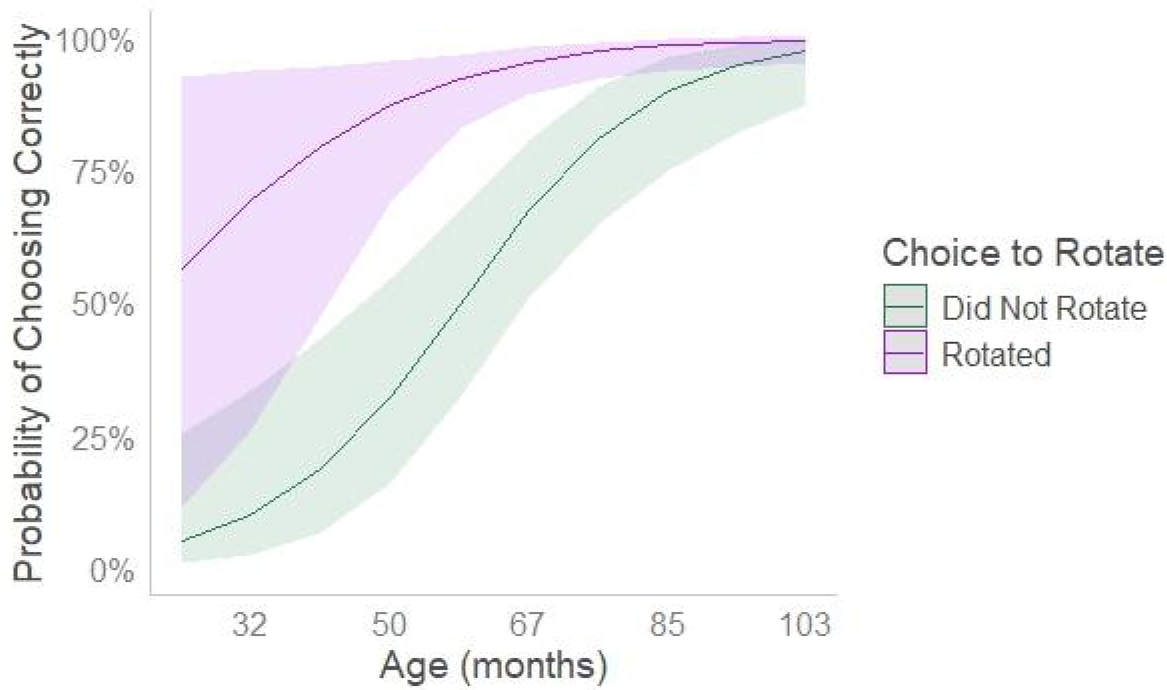

Figure 5. Probability of choosing correctly as a function of age and choice to rotate in Experiment 2 Test Squares, Same Pattern Different Orientation condition.

Experiment 3

The results of Experiment 2 once again demonstrated that children were more likely to rotate objects on difficult trials than easier ones and this resulted in increases in accuracy. In the final experiment, we tested whether this would hold true with a new sample of children for less abstract images by using pictures of animals. Instead of employing a test versus control condition, we focused on strategic use of the rotation offloading behavior by introducing objectively easier animal pairs (i.e., different species with obvious featural differences) versus more difficult discriminations (i.e., same species with more minute featural differences). Consistent with Experiment 2, data was collected in a local children's museum, and this experiment was pre-designed for data collection before any analyses of data from the earlier experiments to provide another test of offloading capacity in children at this age. Pilot testing indicated that for the short duration of testing, we could again give participants trials in all conditions, and hence the return to a within-subjects design.

Method

Participants

A new set of participants were recruited from a local children's museum. A total of 43 children participated in Experiment 3, including 22 males and 21 females. Age was calculated as before, including for parents who only gave their child's birth month ($N = 11$). The average age of the sample was 64 months ($SD = 16$), including 19 children aged 3- to 4-years-old, 15 children aged 5- to 6-years-old, and 8 children aged 7-years-old. One parent/guardian did not include birth date information on the optional demographics form. This participant was excluded from all analyses that included age as a variable. The sample was largely Caucasian ($N = 35$), followed by multi-racial ($N = 5$), Hispanic/Latino ($N = 2$), and Asian/Asian American ($N = 1$).

Materials

The manual rotation task consisted of a set of approximately 4 inch x 4 inch laminated cards depicting images corresponding to each condition (described in detail below and depicted in Figure 6). Children were provided a small reward for their participation following study completion.

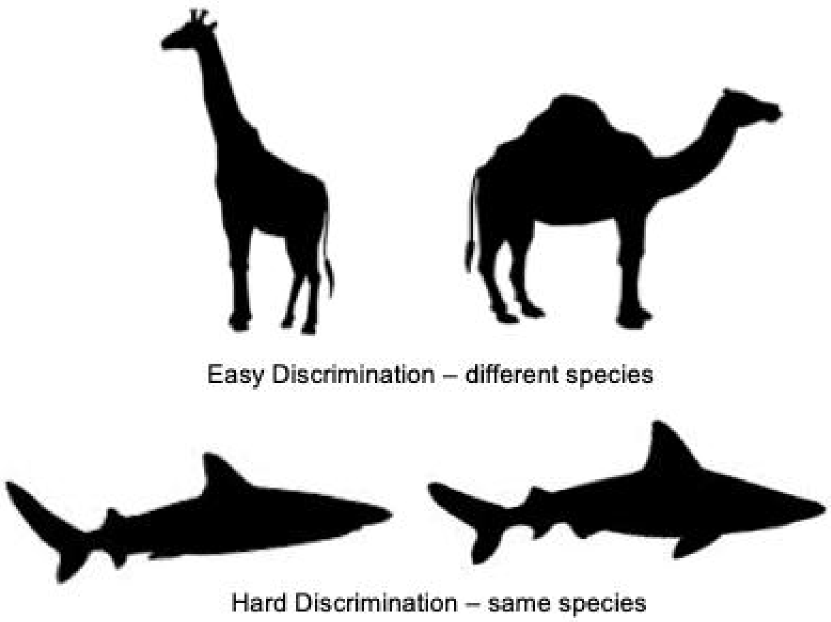

Figure 6. Experiment 3 sample Different trials. The goal was to indicate whether the animals were the same or different. For the Different trials, we varied difficulty by presenting easy discriminations (top row – two different species, which were visibly unique from one another, e.g., camel vs. giraffe) and hard discriminations (bottom row – two of the same species with slight differences in their features, e.g., two sharks with minor differences in their fins).

Design and Procedure

Children were shown a pair of cards simultaneously that they were instructed to evaluate based on a particular prompt ("are these two animals the same or different?"). Consistent with Experiment 2, children also were instructed that they could rotate the righthand card if it made the task easier (i.e., to decide if the cards were the same or different).

Participants were given two practice trials in which they had to perform the discrimination (indicate whether the cards were the same or different). During these practice demonstrations, participants were shown how to rotate the righthand card so that it aligned with the other card. Children were given a verbal reminder that they could rotate the card on the right after every five trials if no rotation was performed by a child during those five trials. They were given general feedback regarding effort but no performance-based feedback.

Children completed 22 trials. Each trial presented a pair of cards that depicted an animal on each card, and those images were either the same or different. Half of the trials were Same (N = 11 trials), and half of the trials were Different (N = 11 trials). In Same animal trials, two identical images of the same animal were presented, and these were categorized as "Easy" discriminations. In Different animal trials, children were presented with either animals of two different species that were visibly unique from each other (e.g., camel vs. giraffe), which were also categorized as Easy discriminations (N = 4 trials); or they were presented with two animals that were the same species but had slight differences in their features (e.g., two sharks with minor differences in their fins), and these were categorized as "Hard" discriminations (N = 7 trials). The starting orientation for Same and Different trials varied from 0° to 270° across all trials, such that the card placed on the right was presented in the same orientation as the left card (0°), one rotation to the right (90°), opposite vertical orientations (180°), or one rotation to the left (270°). We randomized trials for same-different animals and starting orientation. For the Same and Different animal trials, children were presented with two cards in 0° rotation, three cards in 90° rotation, three cards in 180° rotation, and three cards in 270° rotation. Children were shown two demonstration trials to practice the perceptual discrimination prompt and rotation instruction. All children accurately indicated whether the cards contained the same or different animals on the practice trials. These trials were not used in the analyses.

For all conditions, we recorded the children's response to the primary task (determining whether the two presented stimuli were the same or different). We also recorded whether the righthand card was rotated.

Results

We first examined performance as a function of trial difficulty (Easy vs. Hard). Accuracy was significantly higher in the Easy trials (*M* = 82.33%, *SD* = 17.87) as compared to the Hard trials (*M* = 52.49%, *SD* = 21.31), paired samples $t(42) = 7.12$, $p < .001$, Cohen's $d = 1.52$. Average percent rotation did not differ across Easy (*M* = 40.31%, *SD* = 29.95) and Hard trials (*M* = 40.86%, *SD* = 30.0), paired samples $t(42) = -.25$, $p = .807$, Cohen's $d = .018$. A closer examination of specific trial types showed the importance of rotating stimuli.

GLMM Results

Choice to Rotate. We explored the effect of starting orientation, trial type, and age on children's choice to rotate, once again excluding participants who were missing birthdate/age information (*N* = 1). Starting orientation was categorized as *Same* (i.e., 0° different) or *Different* (including 90°, 180°, or 270° rotated). Trial type was categorized as *Easy* (the two animal images were exactly the same or the two animals were entirely different species) or *Hard* (the two animal images were of the same species but had slight differences in the images). We could not consider the interaction of all three of these factors (starting orientation, trial type, and age) on children's choice to rotate given the complexity of the resulting model in relation to our data set. Therefore, we compared simpler candidate models that included combinations of lower-level main effects and interactions to determine which simplified model would best fit the data. These comparisons were made using the package AICmodavg (Mazerolle, 2023) and were based on the models'

Akaike information criterion with correction (AICc) values. The complete list of candidate models that were compared can be found in Table 9 of the Supplementary Material.

The best fitting model was one that included the interaction between starting orientation and age and their main effects. This model revealed only a main effect of starting orientation ($\beta = 2.483$, $p < .001$; see Table 10, Supplementary Material). Children were significantly more likely to rotate an image if the images started at Different orientations compared to if they started at the Same orientation (for which they had a nearly 0% likelihood of rotation; see Table 11, Supplementary Material).

Accuracy. Next, we explored the effects of starting orientation, difficulty level, age, and choice to rotate on children's accuracy. We did not consider a four-way interaction due to the complexity of the model in relationship to the dataset and feasibility of interpretation. Candidate models can be found in Table 12 of the Supplementary Material. The best fitting model was one which included the three-way interaction between difficulty level, age, and choice to rotate and all lower-level interactions and main effects. Including Subject ID resulted in a singular fit of the model, so Subject ID was removed, and the model was run as a GLM. This model revealed two significant two-way interactions, and no significant three-way interaction (see Table 13, Supplementary Material). The first significant two-way interaction was between age and difficulty level ($\beta = -0.354$, $p = .003$). That is, for easy trials, the likelihood of accuracy increased with increasing age, whereas in hard trials, age had a minimal effect on accuracy likelihood (Figure 7; Table 14, Supplementary Materials). The second significant two-way interaction was between choice to rotate and difficulty level ($\beta = -0.318$, $p = .001$). For easy trials, children were more likely to be accurate if they rotated, but for hard trials, children were about equally likely to be accurate whether they chose to rotate or not (Figure 7; Table 14, Supplementary Materials). In other words, for easy trials, both age and the choice to rotate significantly impacted children's likelihood of being accurate, but for hard trials neither age nor choice to rotate significantly impacted children's likelihood of being accurate. These interactions confirmed the relative difficulty of these trial types.

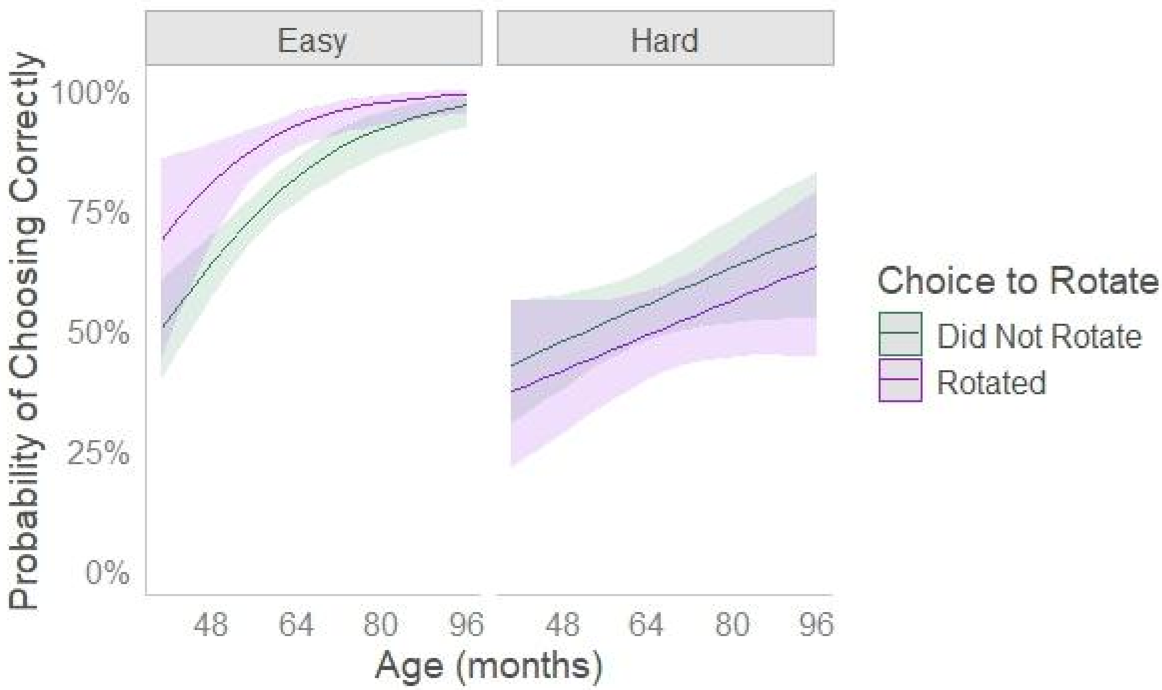

Figure 7. Probability of choosing correctly as a function of age, choice to rotate, and trial type for Experiment 3.

General Discussion

The current study extended previous work in the developmental literature suggesting that children can effectively engage their metacognitive abilities through spatially rotating objects when making perceptual discriminations and that older children were consistently more likely to do so than younger children (e.g., Armitage et al., 2020). Across three experiments, we explored this ability with different stimuli and with younger children (3- to 9-years-old) than had been previously studied. Children in all age groups showed rotation of the various stimuli to make the task easier, although there was a developmental trend such that likelihood of accuracy and rotation increased with age. Additionally, children were more likely to rotate objects on difficult trials than easier ones and this resulted in increases in accuracy. This tendency to rotate for the more difficult trials was associated with age. These results confirm that children can manipulate stimuli in ways that make comparing those stimuli easier, reflecting a form of (meta)cognitive offloading using the external environment to resolve internal uncertainty.

Rotation Findings

Children consistently chose to rotate more when it would be helpful for making a relevant discrimination compared to when it was not needed. Within the various test conditions, children were selective about when to offload. In particular, children were more likely to rotate when the discrimination was more difficult as compared to when it was easier. For instance, in Experiment 2, when the grid patterns were identical and presented in the same orientation (SPSO) or when the patterns were different (DP), children, no matter their age, demonstrated less than a 5% probability of rotating the stimuli. However, when the same object was presented in a different orientation (SPDO), children were more likely to rotate. They also were more likely to be accurate when they rotated compared to when they did not rotate. Importantly, this tendency to rotate for the more difficult trials in the experimental conditions was associated with age, such that as age increased, so did the tendency to rotate these kinds of trials. Similar patterns were observed in Experiment 1 with horizontal and vertical lines.

For the horizontal-vertical illusion in Experiment 1, children overwhelmingly reported the vertical line as longer in the equal-length trials (in 73.66% of trials), reflecting the standard line-length illusion (Brosvic et al., 1993, 2002). Moreover, children were more likely to rotate the vertical line when it was close in length with the horizontal line and less likely to rotate the vertical line when it was considerably shorter than the horizontal line, and this also reflects the illusory experience. These findings are consistent with a recent study in which preschool children showed a clear metacognitive error in judging vertical and horizontal lines. In this task, children confidently moved towards a separate location to collect a reward in trials that they experienced the illusion, which gave them false confidence that the vertical line was longer and thus should have generated a reward for that trial (James et al., 2021).

Our data are consistent with the work of Armitage et al. (2020) who demonstrated that rotation probability increased with the degree of misalignment. Armitage and colleagues also found that the younger children in their sample (4- and 5-year-olds) were less likely to rotate objects. We observed the same pattern in all three experiments, particularly when considering the most difficult test trials in Experiments 1 and 2. In a different paradigm, Redshaw et al. (2018) demonstrated that children aged between 6.9- and 13-years-old predicted a task that required future responding to three targets would be more difficult to remember than a task that involved responding to one target. These predictions were also fairly well calibrated with their task performance. Despite correctly predicting that responding in the 3-target condition would be difficult, younger children did not set as many reminders (i.e., offload) during these trials as older children within the sample age range tested. In other words, younger elementary-aged children showed the ability to metacognitively monitor, but not to enact that knowledge in the task context, which is consistent with the larger developmental metacognition literature (see Lockl & Schneider, 2004).

One possibility that helps explain the age-related changes comes from the relation of age and working memory span in children (e.g., Gathercole et al., 2004; Isaacs & Vargha-Khadem, 1989). It has been shown that low-span children struggle more to engage in effective cognitive offloading. Berry et al. (2019) presented children with a color block sequencing task, and they were required to arrange colored blocks in the same order as they heard them announced (the working memory task). Low span children showed task facilitation when the blocks available to arrange were already sorted by color (offloading the need to search among random colors). However, low span children did not pre-sort blocks by color (i.e., engage in offloading) when given the chance, suggesting a lack of metacognitive awareness of the value of this effort.

Based on the design of the current experiment, it is unclear whether children would have picked up on this rotation strategy spontaneously. The instructional manipulation in our experiment required explicit practice with rotation as well as in-task reminders. A natural next step would be to see if children spontaneously use the rotation response for newer stimuli without explicit instruction or prompts. Nevertheless, the present data along with previous research (e.g., Armitage et al., 2023) suggest that as children get older, they seem better able to identify when offloading is needed. We next consider how the choice to rotate and age affected the accuracy of the decisions.

Accuracy Findings

As anticipated by the manipulation, easier discriminations resulted in higher accuracy than difficult discriminations. In Experiment 1, all children performed near ceiling on the easy trials, but older children (i.e., +1 SD above mean age) were more likely to be accurate than younger children (i.e., -1 SD below the mean age) for hard trials, especially when the children did not rotate the stimuli. This age effect decreased when children rotated the line, such that offloading resulted in higher levels of accuracy for all children in hard trials. In Experiment 2, all children showed high levels of accuracy and low levels of rotation when the patterns were the same and in the same starting orientation (SPSO) as well as when the patterns were different (DP) (i.e., the easier conditions). In the more difficult condition in Experiment 2 (SPDO), both increasing age and choosing to rotate provided an advantage for accuracy. In Experiment 3, all children were more likely to be accurate for easy trials compared to hard trials. Increasing age and choosing to rotate the stimulus provided an advantage for accuracy on easy trials but not hard trials. The positive effect of choice to rotate on accuracy in these tasks suggests that offloading ability is an important determinant of success. This is consistent with Armitage and Redshaw (2022) who found that when children were allowed to rotate a map and bring it into alignment, they were better able to find the hidden rewards in the task.

Limitations

One key difference between the current study and previous work is that our study did not include a condition where participants were not allowed to rotate. Thus, we were not able to directly assess performance changes from an unaided to an aided condition. Therefore, the possibility exists that the children who rotated would have performed better even when not allowed to rotate. We believe this is an unlikely possibility, given that rotating was consistently associated with better performance, but our design was not able to unequivocally rule out this possibility.

When considering the hard trials, rotation was associated with higher accuracy across experiments, consistent with previous rotation work (Armitage et al., 2020). The one exception to this pattern was in Experiment 3 for the Hard discriminations, where there was a numerical advantage for those who did not rotate, but it is important to note that probability of accuracy for these difficult trials hovered around 50%, even for older children. This speaks to the difficulty of the discrimination children had to make. Specifically, in this experiment, the hard trials were the same animal with slight featural differences (e.g., two pictures of a shark but one was missing a pectoral fin). Previous developmental research demonstrated increases in perceptual sensitivity to small differences in objects across 5-, 6-, and 11-year-olds, providing support to the idea that identifying minor perceptual violations in stimuli may prove difficult for early elementary-aged children (Lange-Küttner, 2000). In a future experiment using these images, stimuli with larger perceptual differences may be employed or children might be warned during the instructions that some of the differences may be very small and therefore they should be very careful. Under such a set of instructions, we might anticipate that rotations would increase on any trials where the animals were of the same type. We would also predict that choosing to rotate on these difficult trials would have a large impact on accuracy. Additionally, because of the nature of convenience sampling, the number of children who were 7 years old and older in our sample decreased across experiments, which may have impacted potential age effects, particularly in the final experiment. Future studies that include a larger number of older children are needed to fully map the trajectory of off-loading across the elementary years.

Conclusions

We employed a manual rotation task to assess cognitive offloading in children as young as 3 years of age using perceptual discrimination tasks, including a visual line-length illusion and same-different discrimination. Children were more likely to rotate on trials in which they would benefit from such a decision (test versus control) as well on trials in which the discriminations were more difficult relative to easier trials. Older children outperformed younger children in this task, including a greater likelihood to rotate, which benefited their overall performance. These results are indicative of metacognitive monitoring, as measured using a procedural task which required little verbal explanation or prompting.

Data Availability Statement

All raw data are freely available online at https://osf.io/msx4u/.

References

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions.* Sage Publications, Inc.

Armitage, K. L., Bulley, A., & Redshaw, J. (2020). Developmental origins of cognitive offloading. *Proceedings of the Royal Society B, 287,* 20192927. https://doi.org/10.1098/rspb.2019.2927

Armitage, K. L., & Redshaw, J. (2022). Children boost their cognitive performance with a novel offloading technique. *Child Development*, *93*, 25-38. https://doi.org/10.1111/cdev.13664

Armitage, K. L., Suddendorf, T., Bulley, A., Bastos, A. P. M., Taylor, A. H., & Redshaw, J. (2023). Creativity and flexibility in young children's use of external cognitive strategies. *Developmental Psychology, 59*, 995–1005. https://doi.org/10.1037/dev0001562

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J., (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68,* 255-278. https://doi.org/10.1016/j.jml.2012.11.001

Beran, M. J., Kelly, A. J., Perdue, B. M., Whitham, W., Love, M., Kelly, V., & Parrish, A. E. (2019). Divide and conquer: Preschool children can assign the hardest tasks to a symbolic helper. *Experimental Psychology, 66,* 296-309. https://doi.org/10.1027/1618-3169/a000454

Berry, E. D., Allen, R. J., Mon-Williams, M., & Waterman, A. H. (2019). Cognitive offloading:

structuring the environment to improve children's working memory task performance. *Cognitive Science*, *43*, e12770. https://doi.org/10.1111/cogs.12770

Boldt, A., & Gilbert, S. J. (2019). Confidence guides spontaneous cognitive offloading. *Cognitive Research: Principles and Implications, 4,* 1-16. https://doi.org/10.1186/s41235-019-0195-y

Bremner, J. G., & Andreasen, G. (1998). Young children's ability to use maps and models to find ways in novel spaces. *British Journal of Developmental Psychology*, *16*, 197-218. https://doi.org/10.1111/j.2044-835X.1998.tb00919.x

Brosvic G. M., Bailey S., Baer A., Engel J., Dihoff R. E., Carpenter L., Baker S., & Cook, M. (1993). Developmental susceptibility to the horizontal-vertical illusion. *Bulletin of the Psychonomic Society, 31,* 609–612. https://doi.org/10.3758/BF03337369

Brosvic, G. M., Dihoff, R. E., & Fama, J. (2002). Age-related susceptibility to the Müller-Lyer and the horizontal-vertical illusions. *Perceptual and Motor Skills, 94,* 229–234. https://doi.org/10.2466/pms.2002.94.1.229

Bulley, A., McCarthy, T., Gilbert, S. J., Suddendorf, T., & Redshaw, J. (2020). Children devise and selectively use tools to offload cognition. *Current Biology*, *30*, 3457-3464. https://doi.org/10.1016/j.cub.2020.06.035

Coughlin, C., Hembacher, E., Lyons, K. E., & Ghetti, S. (2015). Introspection on uncertainty and judicious help-seeking during the preschool years. *Developmental Science, 18,* 957–971. https://doi.org/10.1111/desc.12271

Dunn, T. L., & Risko, E. F. (2016). Toward a metacognitive account of cognitive offloading. *Cognitive Science*, *40*, 1080-1127. https://doi.org/10.1111/cogs.12273

Dupont, D., Zhu, Q., & Gilbert, S. J. (2023). Value-based routing of delayed intentions into brain-based versus external memory stores. *Journal of Experimental Psychology: General*, *152*, 175-187. https://doi.org/10.1037/xge0001261

Fick, A. (1851). *De errone quodam optic asymmetria bulbi effecto* [Errors due to optical asymmetries]. J. A. Kochii.

Gardony, A. L., Brunyé, T. T., & Taylor, H. A. (2015). Navigational aids and spatial memory impairment: The role of divided attention. *Spatial Cognition & Computation*, *15*, 246-284. https://doi.org/10.1080/13875868.2015.1059432

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*, 177–190. https://doi.org/10.1037/0012-1649.40.2.177

Gilbert, S. J. (2015). Strategic offloading of delayed intentions into the external environment. *Quarterly Journal of Experimental Psychology, 68*, 971-992. https://doi.org/10.1080/17470218.2014.972963

Goupil, L., Romand-Monnier, M., & Kouider, S. (2016). Infants ask for help when they know they don't know. *Proceedings of the National Academy of Sciences, 113*, 3492–3496. https://doi.org/10.1073/pnas.1515129113

Isaacs, E. B., & Vargha-Khadem, F. (1989). Differential course of development of spatial and verbal memory span: A normative study. *British Journal of Developmental Psychology, 7*, 377-380. https://doi.org/10.1111/j.2044-835X.1989.tb00814.x

James, B. T., Parrish, A. E., Guild, A. S., Creamer, C., Kelly, V., Perdue, B., ... & Beran, M. J.

(2021). Go if you know: Preschool children's movements reflect their metacognitive monitoring. *Cognitive Development*, *57*, Article 101001. https://doi.org/10.1016/j.cogdev.2020.101001

Jolicoeur, P. (1988). Mental rotation and the identification of disoriented objects. *Canadian Journal of Psychology*, *42*, 461-478. https://doi.org/10.1037/h0084200

Lange-Küttner, C. (2000). The role of object violation in the development of visual analysis. *Perceptual and Motor Skills*, *90*, 3-24. https://doi.org/10.2466/pms.2000.90.1.

Lange-Kuttner, C., & Green, H. (2007, July). What is the age of mental rotation?. In *2007 IEEE 6th International Conference on Development and Learning* (pp. 259-263). IEEE. https://doi.org/10.1109/DEVLRN.2007.4354043

Lockl, K., & Schneider, W. (2004). The effects of incentives and instructions on children's allocation of study time. *European Journal of Developmental Psychology*, *1*, 153–169. https://doi.org/10.1080/17405620444000085

Lüdecke, D. (2018). ggeffects: Tidy data frames of marginal effects from regression models. *Journal of Open Source Software, 3*(26), 772. https://doi.org/10.21105/joss.00772

Mazerolle, M. J. (2023). AICcmodavg: Model selection and multimodel inference based on (Q)AIC(c)_. R package version 2.3.3, <https://cran.r-project.org/package=AICcmodavg>.

Murphy, D. H. (2023). Strategic offloading: How the value of to-be-remembered information influences offloading decision-making. *Applied Cognitive Psychology*, 37, 749–767. https://doi.org/10.1002/acp.4051

Pasch, B., Bolker, B. M., & Phelps, S. M. (2015). Interspecific dominance via vocal interactions mediates altiduinal zonation in neotropical singing mice. *The American Naturalist, 182,* E161-E173. https://doi.org/10.1086/673263

R Core Team (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>

Redshaw, J., Vandersee, J., Bulley, A., & Gilbert, S. J. (2018). Development of children's use of external reminders for hard-to remember intentions. *Child Development*, *89*, 2099–2108. https://doi.org/10.1111/cdev.13040

Risko, E. F., & Dunn, T. L. (2015). Storing information in-the-world: Metacognition and cognitive offloading in a short-term memory task. *Consciousness and Cognition, 36,* 61-74. https://doi.org/10.1016/j.concog.2015.05.014

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences, 20,* 676-688. https://doi.org/10.1016/j.tics.2016.07.002

Risko, E. F., Medimorec, S., Chisholm, J., & Kingstone, A. (2014). Rotating with rotated text: A natural behavior approach to investigating cognitive offloading. *Cognitive Science, 38*, 537-564. https://doi.org/10.1111/cogs.12087

Schneider, W. (2008). The development of metacognitive knowledge in children and adolescents: Major trends and implications for education. *Mind Brain and Education*, *2*, 114-121. https://doi.org/10.1111/j.1751-228X.2008.00041.x

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*, 701-703. https://doi.org/10.1126/science.171.3972.701

Thompson, R. B., Cothran, T., & McCall, D. (2012). Gender and age effects interact in preschoolers' help-seeking: Evidence for differential responses to changes in task difficulty. *Journal of Child Language*, *39*, 1107-1120. https://doi.org/10.1017/S030500091100047X

UCLA: Statistical Consulting Group. (n.d.) How do I interpret the coefficients of an effect-coded variable involved in an interaction in a regression model? Retrieved from https://stats.oarc.ucla.edu/other/mult-pkg/faq/general/faq-how-do-i-interpret-the-coefficients-of-an-effect-coded-variable-involved-in-an-interaction-in-a-regression-model/#effect-coding-for-categorical-predictors-with-3-or-more-levels.

Vosmik, J. R., & Presson, C. C. (2004). Children's response to natural map misalignment during wayfinding. *Journal of Cognition and Development*, *5*, 317-336. https://doi.org/10.1207/s15327647jcd0503_2

Wickham, H. (2016). Ggplot2: Elegant graphics for data analysis. Springer-Verlag, New York. https://ggplot2.tidyverse.org https://doi.org/10.1007/978-3-319-24277-4_9