Research Article

ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Psychological Science
1–16

DOI: 10.1177/09567976241306424
www.psychologicalscience.org/PS
Sage

The Extended Mind in Young Children: Cost-Dependent Trade-Off Between External and Internal Memory

Yibiao Liang[1,3], Erik Blaser[1], Jia Ying Yi[2], Liyang Sai[2], and Zsuzsa Kaldy[1]
[1]Department of Psychology, University of Massachusetts Boston; [2]Department of Psychology, Hangzhou Normal University; and [3]Institute of Psychology, Goethe University Frankfurt

Abstract
Most work on working memory development has children remember a set of items as well as they can. However, this approach sidesteps the *extended mind*, the integration of external information with memory. Indeed, adults prefer to use external resources (e.g., lists, models) but will remember more as the cost to access them increases. Here, in our shopping game, we investigated this trade-off in 5- to 8-year-olds. Using a touchscreen, children shopped in a virtual store. Their shopping list and the store were not visible simultaneously but could be toggled. We manipulated access cost by varying a delay (0–4 s) before the list's reappearance. Across three preregistered experiments at two sites (the United States and China, $N = 141$), a pattern emerged: When it was costlier to do so, children revisited the list less often, studied it longer, and selected more correct items. Also, children recognized the costs, identifying the no-delay condition as easier. Young children showed a cost-dependent trade-off of external-resource use versus working memory.

Keywords
extended mind, working memory, cognitive effort, cognitive development, preregistered, cross-cultural samples

Received 3/6/24; Revision accepted 11/21/24

Introduction

Most people cannot imagine living without their phones. Indeed, access to the internet has inflated people's confidence in their knowledge (Ward, 2021). Of course, people have long used external resources to support problem-solving, from the strings and knots of ancient *quipu* (Ascher & Ascher, 1981) to writing itself (despite concerns that this might corrupt memory itself; Király et al., 2017; Plato, 1997). This exploitation of external resources to support internal cognitive processes is a reflection of the *extended mind* (Clark & Chalmers, 1998). For instance, if the information in a notebook is accessible (easily reachable), reliable (available when needed), and trustworthy (accurate), then it extends the mind; the external resources do not just support cognitive processes, but they become part of them (Clark, 2008; Gallagher, 2018). This is advantageous, allowing one to overcome internal limitations and solve otherwise intractable tasks (Hayhoe & Ballard, 2005; Risko & Gilbert, 2016). In the current study, we will focus on the effect of accessibility on the extended minds of young children.

The extended mind in adults

Solving a task requires integrating the processing of goal-related information from the environment with the processing of internal representations (e.g., those in working memory). One line of work on this interplay measured participants' ability to switch processing between (versus within) these domains (for a review, see Verschooren et al. (2019). Alongside this was the

Corresponding Author:
Yibiao Liang, Goethe University Frankfurt, Institute of Psychology
Email: yikbiu@outlook.com

related but distinct line of research on the factors that influenced participants' choice to use external versus internal resources (Ballard et al., 1995; Draschkow et al., 2021; Hayhoe et al., 1998; Inamdar & Pomplun, 2003; Somai et al., 2020). For instance, Ballard et al. (1995) asked participants to copy a puzzle-like model. Participants most frequently used a strategy in which they looked at the model both before picking up a puzzle piece and before placing the piece down. That is, instead of memorizing the identity and position of multiple pieces, participants referred to the model frequently, memorizing only the identity or the location of a single piece. Indeed, it has been suggested that we are "cognitive misers," minimizing cognitive effort when possible (e.g., the effort to encode, maintain, and retrieve information in working memory; Kool et al., 2010; Taylor, 1981)—for instance, physically rotating a page of text, or tilting one's head, instead of performing mental rotation (Risko et al., 2014).

To manipulate the accessibility of external resources, Ballard et al. (1995) increased the distance between the model and the participant's puzzle, thereby increasing the locomotive cost (eye and head movements) to shift between them. This led participants to increase working memory use. Time can also be a cost (Grinschgl et al., 2021; Sahakian et al., 2023; Somai et al., 2020). In Somai et al. (2020), participants used more working memory when they had to wait to see the model. (Our paradigm similarly uses a delay-to-access manipulation.) In general, it seems people adaptively trade off the use of external and internal resources to strike a balance for a given set of task demands between the subjective effort of using external resources and the subjective effort of using internal processes like working memory (Ballard et al., 1995; Draschkow et al., 2021; Grinschgl et al., 2021; Kenderla & Kibbe, 2023; Somai et al., 2020).

The mechanisms that underlie the trade-off are those of *cognitive control*. Cognitive control is commonly understood to underlie flexible, goal-directed, adaptive processing (Badre & Nee, 2018; Botvinick et al., 2001; Egner, 2017). The decision processes that lead to resource distribution are based on cost-benefit computations (see, e.g., Chater & Oaksford, 1999; Lieder & Griffiths, 2019). For example, gating policies that govern which representations to encode into working memory (and when to discard them) are specified in terms of cost-benefit computations (Chatham & Badre, 2015; O'Reilly & Frank, 2006); such computations may themselves be influenced by how confident one is in one's own memory (Gilbert et al., 2020). Recent models of cognitive control have put cognitive effort at the center of these computations (Agrawal et al., 2022; Inzlicht et al., 2018; Kool et al., 2017; Kool & Botvinick, 2018; Shenhav et al., 2017; Westbrook & Braver, 2015). Cognitive-effort exertion is in many ways analogous to physical-effort exertion: It is the straining of the "cognitive musculature" (Shenhav et al., 2021). According to the highly influential *expected value of control* model (Shenhav et al., 2013, 2017), people determine how much cognitive effort to exert by weighing the costs and benefits of allocating cognitive control to various processes, including memory maintenance (Westbrook et al., 2013); they then choose the one with the highest expected value. Here we study the cost-dependent trade-off of external resource use versus working memory in the extended minds of young children.

The extended mind in children

In the development of the extended mind, 5 to 8 years of age is an important period, when children's developing working memory abilities (Ahmed et al., 2022) begin to be guided explicitly as they begin formal schooling. As with adults, if children take advantage of external resources, this can help them mitigate demands on internal processing (Armitage et al., 2020; Bulley et al., 2020). This is also the age at which the ability to spontaneously monitor cognitive demands (i.e., *metacognition*) begins to emerge (Niebaum & Munakata, 2020). One of the reasons to use external resources is to avoid exerting cognitive effort (Risko & Gilbert, 2016), and this likely requires recognizing the demands of different strategies or conditions. Consequently, some researchers have suggested that young children (5-year-olds) are not yet cognitive misers, indicating that they do not yet spontaneously minimize their cognitive effort adaptively (O'Leary & Sloutsky, 2017) but that this attribute begins to emerge at around 6 years of age. For instance, 6-year-olds, but not 4- to 5-year-olds, will (as in the Risko et al., 2014, study with adults) physically rotate a paper map rather than engage in mental rotation (Armitage & Redshaw, 2022). The present study will test whether children show a cost-dependent trade-off and whether they recognize the relative costs.

There are only a few studies that have investigated the extended mind in children. Haselen et al. (2000) and Hoffman et al. (2003) used a model-copying paradigm similar to Ballard et al.'s (1995) and found that 5- to 11-year-old children showed similar eye-movement patterns to adults', indicating that children will also spontaneously use external resources (however, the sample sizes were relatively small). Studies on cognitive offloading have shown that even young children (4-year-olds) can create external resources to help solve problems (e.g., physically marking the location of hidden targets for later search; Armitage & Redshaw, 2022;

Bulley et al., 2020), but cost-dependent trade-offs were not tested. In a recent study (Kenderla & Kibbe, 2023), 8- to 10-year-olds were asked to find a set of cards that satisfied a rule. Cards were face down, but had faces that differed in features (color, shape, pattern, etc.) that could be exposed with a click. Access cost was manipulated by varying the delay before a card was revealed. Children relied more on external resources—that is, they viewed cards more often before making their final selections when cards were more accessible.

Our work builds on these results, using a streamlined task that targets working memory and can be tailored for younger children. Being able to test younger children is critical in this context. As we discussed above, our reading of the literature identifies 5+ years of age as a plausible age of emergence for this trade-off. In Experiment 1, we cast a wide net testing children from 5 to 8 years of age, and then in Experiment 2, we specifically targeted these younger children (5- to 6-year-olds). By testing children at this particular age, then, our results are certain to have an impact on our understanding of developmental trajectories: We will either corroborate that 5 to 6 years of age is, indeed, the age of emergence of the cost-dependent trade-off, or, alternatively, we will help establish that the age of emergence is yet younger.

The current study

Here, we use a novel naturalistic working memory task—an innovative, tablet-based shopping game—to study the extended mind in 5- to 8-year-old children. In this game, children shop for items with a shopping list. The list and the store are not visible simultaneously, but children can toggle between them, and we can manipulate the time delay before the list reappears. We predicted (a) that children would take advantage of the available external resources (here, the shopping list) to conserve cognitive effort and (b) that they would adaptively trade off the use of external versus internal resources as the accessibility of the external resources changed. That is, we expected that when access cost was increased children would reduce the number of times they used the list, would study it longer (consistent with attempting to remember more), and would select more correct items in the store (consistent with having remembered more).

Research Transparency Statement

General disclosures

Conflicts of interest: All authors declare no conflicts of interest. **Funding:** This research was supported by the National Institutes of Health (Grant No. R15HD086658 awarded to Z. Kaldy and E. Blaser). **Artificial intelligence:** No artificial-intelligence-assisted technologies were used in this research or the creation of this article. **Ethics:** This research complies with the Declaration of Helsinki (2023) and received approval from a local ethics board (IDs 3143 for University of Massachusetts Boston and 2022031 for Hangzhou Normal University).

Experiment 1 disclosures

Preregistration: The hypotheses, methods, and analysis plan were preregistered (https://doi.org/10.17605/OSF.IO/9QRC8) on 06-18-2022, after data collection, which began on 05-21-2022 (only 6 participants were run during this period, and their data was not accessed until all data collection was completed). There were major and minor deviations from the preregistration (for details, see Table S1 in the Supplemental Material available online). **Materials:** All study materials are publicly available (https://doi.org/10.17605/OSF.IO/A2U48). **Data:** All primary data are publicly available (https://doi.org/10.17605/OSF.IO/A2U48). **Analysis scripts:** All analysis scripts are publicly available (https://doi.org/10.17605/OSF.IO/A2U48). **Computational reproducibility:** The computational reproducibility of the results has been independently confirmed by the journal's STAR team.

Experiment 2a disclosures

Preregistration: The hypotheses, methods, and analysis plan were preregistered (https://doi.org/10.17605/OSF.IO/DBKJS) on 02-28-2023, prior to collection, which began on 03-01-2023. There were major and minor deviations from the preregistration (for details, see Table S1 in the Supplemental Material). **Materials:** All study materials are publicly available (https://doi.org/10.17605/OSF.IO/A2U48). **Data:** All primary data are publicly available (https://doi.org/10.17605/OSF.IO/A2U48). **Analysis scripts:** All analysis scripts are publicly available (https://doi.org/10.17605/OSF.IO/A2U48). **Computational reproducibility:** The computational reproducibility of the results has been independently confirmed by the journal's STAR team.

Experiment 2b disclosures

Preregistration: The hypotheses, methods, and analysis plan were preregistered (https://doi.org/10.17605/OSF.IO/7CVS8) on 03-11-2023, prior to data collection, which began on 03-11-2023. There were major and minor deviations from the preregistration (for details, see Table S1 in the Supplemental Material). **Materials:** All study materials are publicly available (https://doi.org/10.17605/OSF.IO/A2U48). **Data:** All primary data are publicly available (https://doi.org/10.17605/OSF.IO/A2U48). **Analysis scripts:** All analysis scripts are publicly available (https://doi.org/10.17605/OSF.IO/A2U48). **Computational reproducibility:** The computational reproducibility of the results has been independently confirmed by the journal's STAR team.

Experiment 1

Method

The research aims, hypotheses, methods, and analysis plan were preregistered on June 18, 2022, after data collection, which began on May 21, 2022 (only 6 participants were run during this period, and their data was not accessed until all data collection was completed). There were major and minor deviations from the preregistration (for details, see below and see Table S1 in the Supplemental Material).

Participants. Sample size was determined a priori using G*Power 3.2 (Faul et al., 2009). On the basis of a pilot study and theoretical considerations, we expected a medium effect size ($d = 0.5$). Given a paired-sample t-test focusing on our main contrast (the time spent on the list in the two conditions) with power of 0.8 and an alpha of .05, the required minimum sample size was 34. We recruited and tested 37 children in all: 27 were tested at the Children's Museum of New Hampshire, and 10 were recruited via mailings and recruitment events from the Greater Boston area and tested in the laboratory. Three children did not complete the experiment. The final sample included 34 children aged 5 to 8 ($M = 6.98$ years, $SD = 1.23$, age range = 5.09–8.97 years; 22 girls). Parents reported that their children were Asian ($n = 4$), Black or African American ($n = 1$), White ($n = 26$), more than one race ($n = 1$), other ($n = 1$), or did not report their children's race ($n = 1$). Of the 34 participants, 5 were reported as being Hispanic/Latino. All procedures received approval from the University of Massachusetts Boston's Institutional Review Board.

Procedure. We developed a naturalistic memory paradigm that was child-friendly but extensible for a wide age range—the shopping game. It was created using PsychoPy (Peirce et al., 2019) and played on a tablet (Fig. 1). Participants were asked to select several items from a store using a shopping list. The store and the list were not visible on the screen simultaneously, instead, participants could toggle between them by tapping the store icon (an image of an arrow) or the list icon (an image of an adult with a cart), respectively. There was no time limit on the task, and crucially, children were told that they were allowed to go back to the list as many times as they wanted. Correctly selected items were crossed off on the 10-item list (to make it clear that they did not need to be selected again). Because we wanted to focus on measuring the memory component, we minimized the search demands of the game by keeping the same 20 items in the same location in the store across all test trials. To further minimize search, the items on the shelves in the store were grouped into four different, consistent categories: fruits, vegetables, proteins (e.g., meat, egg, milk), and baked goods (e.g., cake, cupcakes). In addition, before testing began, the experimenter guided the child through naming all the items in the store.

Children only needed to find six out of the 10 items on the list in each trial, but this information was not disclosed to the child explicitly. The rationale for this was that we wanted the children, throughout the trial, to always have a relatively large set of yet-to-be-selected items on the list. As motivation, every correct selection was rewarded with a star and accompanied by a pleasant sound, whereas every incorrect selection resulted in the loss of a star and an unpleasant sound. (Children were told that the stars indicated how many stickers they would get at the end of the experiment, but in reality they could get as many stickers as they wanted at the end of the session.) Once the child picked six correct items in the store, the trial was over and feedback appeared on the screen, reminding children how many stars they had earned.

The key manipulation in the shopping game was the accessibility of the external resource, which was determined by the delay between tapping on the list icon and the appearance of the list. There were two conditions: a low-cost, 0-s delay (no-delay condition) and a high-cost, 4-s delay (long-delay condition). The difference between the two conditions was explained to the children with a plausible framing story—finding the items in a small convenience store, where their parent is always close by (i.e., no delay), versus finding the items in a large supermarket where the parents have to take a long walk to the children to check the list (i.e., a long delay). In addition, as a visual-reminder cue, a picture of a small convenience store or a large supermarket was presented before each trial. In addition, the distance between the adult and the child icons in the store view also served as a cue (reminder) of the condition (and, during the 4-s time delay, there was an accompanying animation of the list icon—including an adult with a cart holding the list—moving toward the icon of the child). Each condition was organized into a block with one initial practice trial followed by three test trials. (Practice trials were identical to test trials, except that

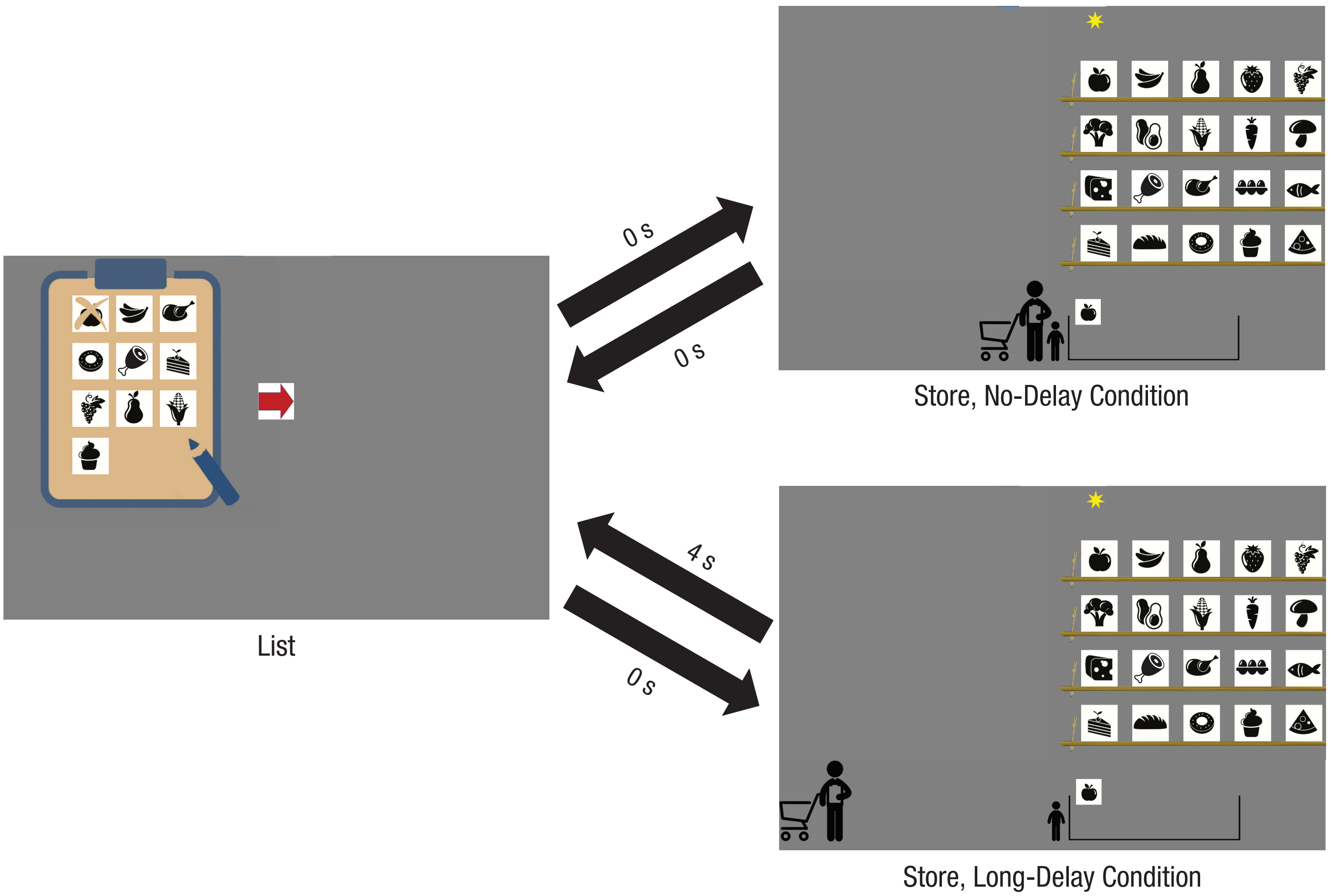

Fig. 1. Schematic overview of the shopping game. On each trial, the store was presented first. Participants could then tap on the list icon (an adult with a cart) to make a trip to the list, and then tap on the store icon (arrow) to go back. In the no-delay condition, the list appeared immediately after the list icon was tapped. In the long-delay condition, the list appeared 4 s after the list icon was tapped. There was never any delay from the list to the store in either condition. Importantly, children were allowed to revisit the list as many times as they wanted during a trial. Children won a star for each correctly chosen item and lost a star for each incorrectly chosen one. After the children correctly chose six (of the 10) list items, the trial ended.

the experimenter explained the game as the trial played out, including the meaning of different icons, the reward system, and how to operate the game.) Children participated in both conditions, and the order of conditions was counterbalanced between participants. The entire session lasted, on average, 9.66 min (*SD* = 2.10).

Measures and analyses. In each trial, we measured three variables: *Dwell time* represents the time children spent studying the shopping list of to-be-remembered items during each trip, which captures the effort they made to encode items in working memory (i.e., use of internal resources); *trips* refers to the number of times participants looked back at the list during each trial, which measures their use of external resources; and *streak correct* refers to the length of the run of correct choices on each trip until the first incorrect choice or until the child opted to return to the list, whichever came first (Sahakian et al., 2023); this estimates working memory use.[1] Following our general hypotheses, we expected that children would have more trips, shorter dwell time, and shorter streak correct in the no-delay condition compared with the long-delay condition. To have a fair comparison with the one-shot condition in Experiments 2a and 2b, dwell time and streak correct were based just on the first trips from each trial (the preregistered analyses based on all trips, which show the same overall pattern of results, are presented alongside first-trip analyses in Tables S2–S10 in the Supplemental Material).

We used R (R Core Team, 2023) and RStudio (Rstudio Team, 2015) for data processing, analyses, and data visualization. We used the *lattice* package (Sarkar, 2008) and the *tidyverse* package (Wickham et al., 2019) for data processing and visualization. We ran our mixed-model analyses with the *lme4* (Bates et al., 2015) and *lmerTest* (Kuznetsova et al., 2017) packages. The *MASS*

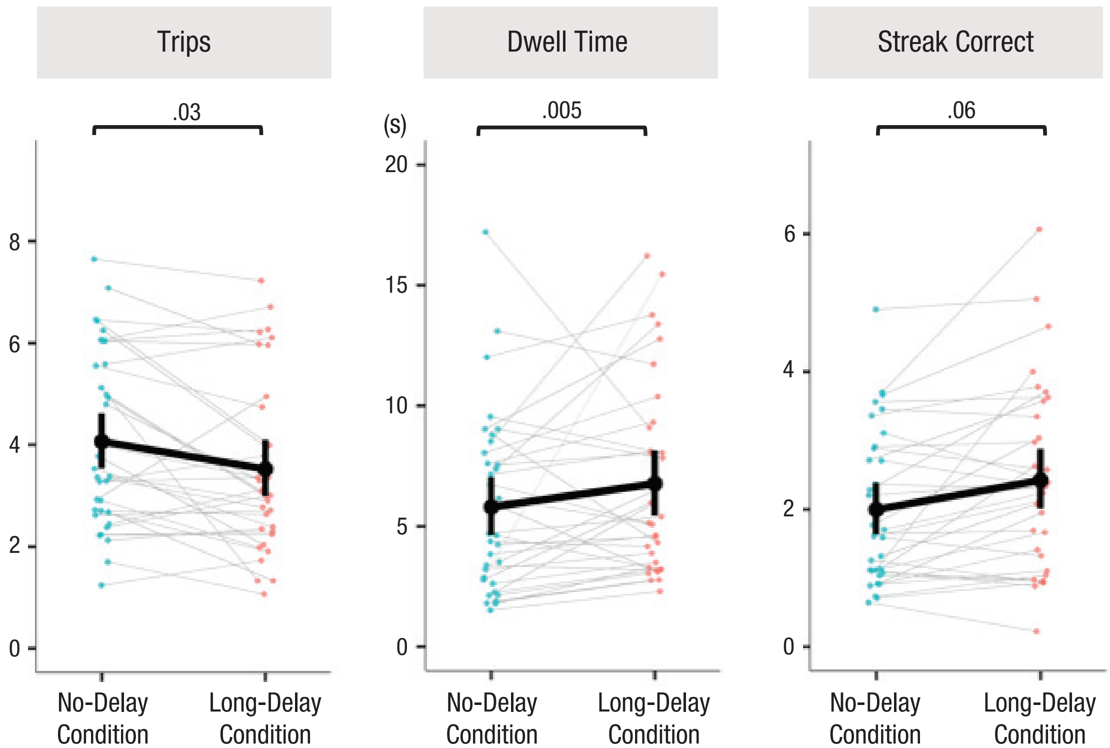

Fig. 2. Individual and group means for the three measures in the no-delay and long-delay conditions. Values above horizontal bars represent the *p* values. Error bars reflect 95% confidence intervals.

package (Venables & Ripley, 2013) was used for the Box-Cox procedure. We conducted our model comparison and selection with the *car* package (Fox & Weisberg, 2018) and the *MuMIn* package (Bartoń, 2023). We used the *emmeans* package (Lenth, 2023) for posthoc tests and the *effectsize* package (Ben-Shachar et al., 2020) for calculating effect sizes. Confidence intervals (CIs) of the graph were calculated with the *Hmisc* package (Harrell & Dupont, 2023).

In our data-analysis procedure, we first fitted the data with a linear mixed model. For continuous data, we visually inspected the distribution of the residuals and used the Box-Cox procedure to calculate the 95% CIs for lambda to transform the response variable to approximate a normal distribution. We chose a lambda value in the CI or close to the CI that was easy to interpret (e.g., −1, 0, 1, 2) when possible. We then refitted the model with the transformed dwell time and visually inspected the distribution of the residuals of the new model. We chose the initial model or the new model on the basis of the distribution of residuals. For count data (trips and streak correct), we fitted the data with generalized linear mixed models with Poisson distribution. For our main analyses, we fitted each response variable with condition (long delay or no delay) as the fixed effect. In our exploratory analyses of the age effect, we fitted each response variable with conditions, age, and the interaction between conditions and age as fixed effects. In all these models, we included a by-participant random slope and a by-participant random intercept to account for the individual variances in the baseline response and in the response difference toward conditions.[2]

Results.

Main effects. (Fig. 2) Trips showed a significant effect of condition (χ^2 = 4.50, p = .034, d = 0.36), revealing more trips to the list in the no-delay condition (M = 4.07 times, SD = 1.73) than in the long-delay condition (M = 3.52 times, SD = 1.73). For dwell time, there was also a significant effect of condition (χ^2 = 7.96, p = .005, d = 0.49), with children studying the list longer in the long-delay condition (M = 6.77 s, SD = 4.00) compared with the no-delay condition (M = 5.81 s, SD = 3.70). Although there was a longer streak correct in the long-delay condition (M = 2.42, SD = 1.32) compared with the no-delay condition (M = 2.00, SD = 1.11), this was not significant (χ^2 = 3.41, p = .065, d = 0.32; but see Tables S4 and S5 in the Supplemental Material for additional analyses that corroborate the trend).

A note regarding children's performance in our task. False alarm (error) rates per trip were generally very low (~3%), indicating that children adopted a highly conservative, risk-averse response bias, likely to avoid

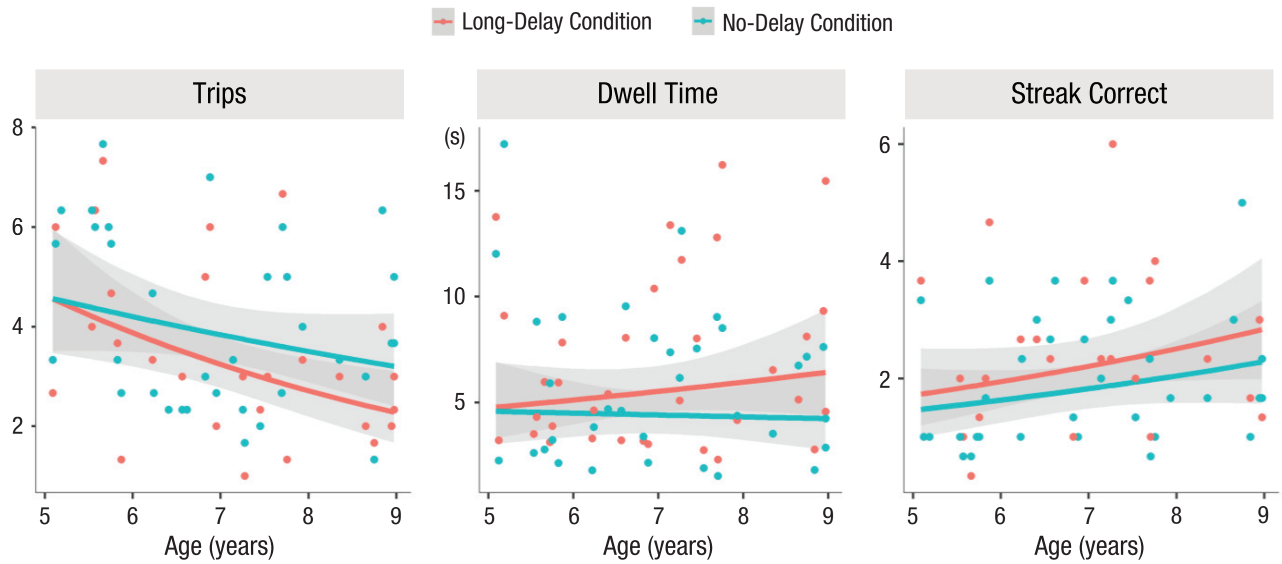

Fig. 3. Age trends for the three measures for both conditions. Dots represent data from the two conditions (long delay and no delay) for each participant. Gray bands represent 95% confidence intervals (CIs).

losing a reward (star). Future studies should manipulate rewards (similar to Gilbert et al., 2020, in adults) to explore the cost-benefit computations in children's trade-offs.

Age effects. We also looked for age effects across our 4-year age range. For trips, we found a main effect of age ($\chi^2 = 5.44$, $p = .020$, $\eta_p^2 = .15$) with older children making fewer trips, but no interaction with condition ($\chi^2 = 2.18$, $p = .139$, $\eta_p^2 = .06$). This main effect of age on trips shows an age-related increase in working memory usage. For dwell time, we found no significant main effect of age ($\chi^2 = 0.38$, $p = .537$, $\eta_p^2 = .004$) nor an interaction ($\chi^2 = 2.26$, $p = .133$, $\eta_p^2 = .07$). Finally, for streak correct, we also found no significant main effect of age ($\chi^2 = 2.77$, $p = .096$, $\eta_p^2 = .08$) and no interaction ($\chi^2 = 0.03$, $p = .860$, $\eta_p^2 = .0003$).

A parallel set of exploratory preregistered analyses, where the per-trial measures of dwell time and streak correct were based on all trips (as opposed to just the first trips, as reported here), showed a similar pattern with one notable exception. For dwell time, we did find a significant interaction effect of age and condition, with younger children less affected by the manipulation of delay time (see Table S6 in the Supplemental Material). This result, taken together with the suggestive trends visible in Fig. 3 here (especially for trips and dwell time) and the backdrop of evidence, as discussed above, that related cognitive processes are only beginning to emerge in 5-year-olds, led us to specifically target this younger age group in Experiment 2.

Results summary. Overall, we found that 5- to 8-year-old children were able to take advantage of external resources (here, a shopping list) and trade off their use against internal resources (working memory) on the basis of accessibility: When we introduced an annoying delay in accessing the shopping list, children reduced the number of trips they made to refer back to it, attempted to remember more by studying longer, and tended to select more items correctly with each visit to the virtual store. Although we did not find compelling evidence for an interaction of age and trade-off in our exploratory analyses, we did see trends that warranted a targeted follow-up with the younger children in our age range.

Experiments 2a (China Site) and 2b (U.S. Site)

Here we investigate the trade-off in younger children (5- to 6-year-olds) and how it may be related to their emerging metacognition. We introduced an additional one-shot condition designed to maximize internal resource use and metacognition questions designed to investigate children's impressions of the subjective effort required for the no-delay versus long-delay conditions. It was not obvious what to predict for children in this age range, which is the reason that we have targeted this issue here. Shifts in the use of external resources likely require monitoring the effort needed to use working memory (Kelly & Risko, 2022), but children's ability to identify, or opt for, an easier task (Niebaum & Munakata, 2020; Niebaum et al., 2019, 2021; O'Leary & Sloutsky, 2017) or to use external resources selectively on the basis of the difficulty of a task (Armitage et al., 2020; Bulley et al., 2020) is only beginning to emerge at 5 years of age. Although we did not find compelling age interactions in Experiment 1, there were trends that, combined with this literature, led us to the conservative prediction that children in this 5- to 6-year age range would not be able to trade off external versus internal resources in response to changing access costs. We found, however, that the opposite was true.

Experiments 2a and 2b had identical procedures and only differed in the testing site, counterbalancing, and sample size. Most published studies in cognitive developmental psychology are based on samples from Western, educated, industrialized, rich, and democratic (WEIRD) societies (Singh et al., 2023). This creates a considerable gap in our understanding, given that culture can influence, for instance, memory (Leger & Gutchess, 2021) and response bias (Freire & Pammer, 2020). Given this, researchers have been advocating for the consideration of cultural differences in cognition and developmental research (Qu et al., 2021). Here, alongside a sample from the United States (Experiment 2b), we also tested a sample of children in another geographic area and culture: eastern China (Experiment 2a).

Method

The research aims, hypotheses, methods, and analysis plan in Experiment 2a were preregistered on February 28, 2023, prior to collection, which began on March 1, 2023. The research aims, hypotheses, methods, and analysis plan in Experiment 2b were preregistered on March 11, 2023, prior to data collection, which began on March 11, 2023. There were major and minor deviations from the preregistration (for details, see below and see also Table S1 in the Supplemental Material).

Participants. In Experiment 2a, the order of the long-delay and no-delay conditions was counterbalanced as in Experiment 1. Now, however, these conditions were either preceded, or followed by, a new one-shot condition (see below). We therefore doubled our planned sample size to 68 to test for any potential effect of position on the trade-off (i.e., an interaction between position and condition). Eighty-seven 5- to 6-year-old children participated in the study, but 15 children were excluded because of technical difficulties, leaving 72 in the final sample (M = 6.06 years, SD = 0.31, age range = 5.52–6.51 years; 36 girls). (One additional participant's data could not be included for the metacognition questions, leaving a sample of 71 for that particular analysis.) All participants were recruited from the same kindergarten in Shengzhou City in eastern China. They were tested in a separate quiet room in the kindergarten by a trained local experimenter. All children were from the Han ethnic group.

Because we did not find an interaction between the position of the one-shot condition (either first or last in the condition order) and condition in Experiment 2a (see Table S10 in the Supplemental Material), in Experiment 2b we tested just the one positional order, with the one-shot condition at the end. Because of this, we no longer needed the double sample size of Experiment 2a. Forty-two 5- to 6-year-old children were recruited and tested at the Children's Museum of New Hampshire in Dover, New Hampshire (n = 17), or at the Discovery Museum in Acton, Massachusetts (n = 7), or tested in the laboratory (n = 18; recruited from the greater Boston area via mailings). Five children in the museums and two in the lab quit the experiment before finishing. The final sample in Experiment 2b included 35 children (M = 6.01 years, SD = 0.60, age range = 5.08–6.93; 13 girls). Caregivers reported that their children were Asian (n = 3), Black (n = 1), White (n = 23), more than one race (n = 3), other (n = 3), or did not report their children's race (n = 2). Of the participants, 3 were reported as being Hispanic/Latino. All procedures received approval from the University of Massachusetts Boston's and Hangzhou Normal University's Institutional Review Board.

Procedure. The procedures of Experiments 2a and 2b were largely identical, and similar to Experiment 1's, with the following changes. A new one-shot condition was added, in which children had only one chance to look at the list, meaning they could not make multiple trips back to the list. This condition served as a frame of reference, assessing children's performance when they are encouraged to maximize their effort to use internal resources. As in Experiment 1, along with instruction and practice, we used visual cues to help children keep track of conditions. For the one-shot condition, we showed a grocery-truck image at the beginning of each trial and a small shopping-list icon (instead of an adult icon) when the store was shown, with the framing story that the truck would drive away. As with Experiment 1, each condition was organized into a block with one initial practice trial followed by three test trials (in one minor change, all 20 items on the store shelves were kept in the same locations for all trials within a blocked condition, as in Experiment 1, but they were rearranged between conditions). Most importantly, we now introduced two metacognitive questions. After the long-delay and no-delay conditions, children were asked, "Which of the two games did you think was easier?" (easier question) and "Which game would you play again to earn more stars?" (preference question). The entire session lasted on average 16.31 min (SD = 2.71) for Experiment 2a and 14.81 min (SD = 5.11) for Experiment 2b.

Measures and analyses. Measures were the same as in Experiment 1, with some additional analyses. The first was a comparison of the three conditions (no delay, long delay, and one shot) using a linear mixed model. We conducted post hoc comparisons with Holm correction for the *p* value (Holm, 1979) if there was a significant interaction. The second new analysis was a comparison of the ratio of choosing the no-delay condition in the metacognition questions (the "which one is easier" and the preference questions) versus chance level (50%) with a binomial test. Then, to compare the two different sites (China and the United States), we ran the linear mixed model with condition, site, and the interaction between these two factors as fixed effects and a random effect structure as the other linear model (a by-participant random intercept and a by-participant random slope). We conducted post hoc comparisons with Holm correction for the *p* value (Holm, 1979) if there was a significant interaction. Experiment 2a showed a main effect of the position of the one-shot condition, with one-shot at the beginning having generally higher dwell time than when one-shot was at the end (χ^2 = 8.20, p = .004, η_p^2 = .09), but there was no interaction effect between order and condition (χ^2 = 0.29, p = .866, η_p^2 = .004; for all results, see Table S10). To match the position and sample size for the site comparison, we compared the data in Experiment 2a (where the one-shot position was last) to data in Experiment 2b (where the one-shot position was always last). We also analyzed the correlation between dwell time and performance.

Result.

Comparison between conditions. (Fig. 4) Our results showed that in both Experiments 2a and 2b, children made more trips in the no-delay condition compared with the long-delay condition (Experiment 2a, no delay: M = 3.88 times, SD = 1.40, long delay: M = 3.32 times, SD = 0.14, χ^2 = 5.95, p = .015, d = 0.29; Experiment 2b, no delay: M = 4.69 times, SD = 1.53, long delay: M = 3.59 times, SD = 1.35, χ^2 = 10.81, p = .001, d = 0.56).

Dwell time[3] also showed a significant effect of condition in both experiments (Experiment 2a, χ^2 = 79.62, p < .001, η_p^2 = .53; Experiment 2b, χ^2 = 37.54, p < .001, η_p^2 = .52), with dwell time significantly longer in the long-delay condition compared with the no-delay condition (Experiment 2a, no delay: M = 8.18 s, SD = 6.01, long delay: M = 8.76 s, SD = 5.13, t = 3.36, p = .001, d = 0.40; Experiment 2b, no delay: M = 4.54 s, SD = 3.32, long delay: M = 5.74 s, SD = 2.37, t = 4.34, p < .001, d = 0.74). In addition, dwell time was significantly higher in the one-shot condition (Experiment 2a, M = 14.36 s, SD = 7.86; Experiment 2b, M = 9.46 s, SD = 6.49) compared to the no-delay and the long-delay conditions (Experiment 2a, one shot versus no delay: t = 8.88, p < .001, d = 1.05; one shot versus long delay: t = 7.27, p < .001, d = 0.86; Experiment 2b, one shot versus no delay: t = 6.03, p < .001, d = 1.03; one shot versus long delay: t = 3.66, p < .001, d = 0.63).

Correspondingly, there was a significant effect on streak correct (Experiment 2a, χ^2 = 65.97, p < .001, η_p^2 = .36; Experiment 2b, χ^2 = 24.60, p < .001, η_p^2 = .43). Post hoc analyses show that streak correct was higher in the long-delay condition compared with the no-delay condition, but it was only significant in Experiment 2b (Experiment 2a, no delay: M = 1.97, SD = 0.90, long delay: M = 2.18, SD = 0.88, z = 1.75, p = .081, d = 0.21; Experiment 2b, no delay: M = 1.55, SD = 0.77, long delay: M = 2.16, SD = 0.91, z = 3.16, p = .003, d = 0.53). The one-shot condition (Experiment 2a, M = 3.23, SD = 1.27; Experiment 2b, M = 2.56, SD = 0.82) also had a higher streak correct than the no-delay condition (Experiment 2a, z = 7.65, p < .001, d = 0.90; Experiment 2b, t = 4.96, p < .001, d = 0.84). Compared to the long-delay condition, the one-shot condition had a higher streak correct numerically, but this was only significant in Experiment 2a (Experiment 2a, z = 5.43, p < .001, d = 0.64; Experiment 2b, t = 1.81, p = .070, d = 0.31). Longer dwell time should lead to better performance. To explore this, we determined the Kendall correlation between dwell time and streak correct. This confirmed that there was a significant positive correlation between dwell time and streak correct in both samples (Experiment 2a, τ = 0.41, z = 14.36, p < .001; Experiment 2b, τ = 0.34, z = 8.13, p < .001).

Metacognition of cognitive demands. For the easier question ("Which of the two games did you think was easier?"), most children chose the no-delay condition in both samples (Experiment 2a: 66.20%, Experiment 2b: 82.86%)—significantly higher than the 50% chance level (Experiment 2a: p = .009, Experiment 2b: p < .001). For the preference question ("Which game would you play again to earn more stars?") the pattern was similar, with significantly more of the children choosing the no-delay condition in Experiment 2a (63.38%, p = .032). The direction of results was consistent but not significant for Experiment 2b (60%, p = .311). There was no significant interaction between condition and response to either of the two metacognition questions in any of the measures (Experiment 2a, easier: trips, χ^2 = 0.42, p = .519, η_p^2 = .006; dwell time, χ^2 = 0.26, p = .613, η_p^2 = .004; streak correct, χ^2 = 1.65, p = .199, η_p^2 = .02; preference: trips, χ^2 = 0.09, p = .763, η_p^2 = .001; dwell time, χ^2 = 0.03, p = .873, η_p^2 = .0004; streak correct, χ^2 = 0.21, p = .646, η_p^2 = .003; Experiment 2b, easier: trips, χ^2 = 0.43, p = .512, η_p^2 = .01; dwell time, χ^2 = 0.61, p = .435, η_p^2 = .02; streak correct, χ^2 = 0.07, p = .788, η_p^2 = .002; preference: trips, χ^2 = 0.04, p = .846, η_p^2 = .001; dwell time, χ^2 = 0.48, p = .490, η_p^2 = .01;

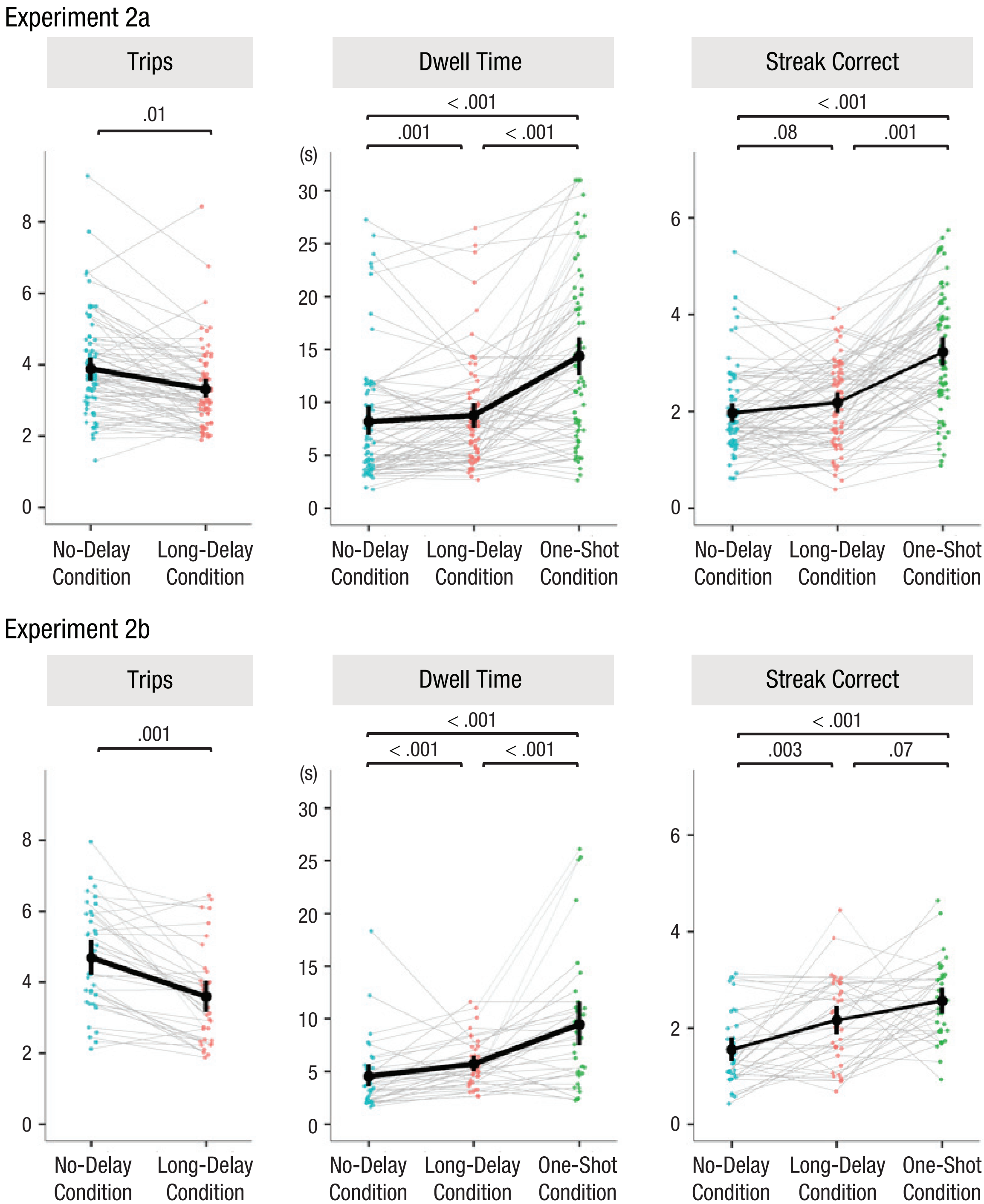

Fig. 4. Individual participant and group means for the three measures in the no-delay, long-delay, and one-shot conditions in Experiments 2a (China) and 2b (United States). Values above the horizontal bars represent *p* values of post hoc pairwise comparisons. Error bars reflect the 95% confidence intervals of the group means.

streak correct, $\chi^2 = 0.21$, $p = .643$, $\eta_p^2 = .01$. See results from all trips and other preregistered measures in Table S8 in the Supplemental Material). These results suggest that most children had an explicit and valid impression of the relative cognitive demand and cost of the two conditions.

Comparison between the Chinese and U.S. samples. For trips, there was neither an effect of site (China: M = 3.71 times, SD = 1.13, United States: M = 4.14 times, SD = 1.30; $\chi^2 = 3.45$, $p = .063$, $\eta_p^2 = .05$), nor an interaction between condition and site ($\chi^2 = 1.14$, $p = .286$, $\eta_p^2 = .02$). Similarly, for streak correct, there was neither an effect of *site (China: M = 2.34, SD = 0.89, United States: M = 2.09, SD = 0.59;* $\chi^2 = 1.74$, *p = .187,* $\eta_p^2 = .03$*) nor an interaction between condition and site (*$\chi^2 = 2.63$, *p = .268,* $\eta_p^2 = .04$*). Dwell time, however, showed a significant effect of site (*$\chi^2 = 3.88$, *p = .049,* $\eta_p^2 = .08$*) with an overall longer dwell time in the Chinese sample (China: M = 9.03 s, SD = 5.44, United States: M = 6.58 s, SD = 3.22;* $\chi^2 = 3.88$, *p = .049,* $\eta_p^2 = .08$*), but no significant interaction between condition and site (*$\chi^2 = 3.53$, *p = .172,* $\eta_p^2 = .05$*). The dwell-time result suggests that children in the Chinese sample may have devoted greater effort, in general, to memorizing list items. We did find a correlation between dwell time and streak correct (see above), but here we did not see a significant difference in streak correct between the two sites. Taken together, this suggests that more time devoted to studying list items, although generally an effective way to remember more, was not sufficiently different between the sites to yield a meaningful difference in performance.*

Results summary. Across two experiments, with independent samples from two testing sites (one in the United States and one in China), we found the same pattern: 5- to 6-year-old children (a) made more trips to visit the shopping list in the no-delay condition than in the long-delay condition, (b) spent a significantly shorter time studying the shopping list in the no-delay condition than in the long-delay and one-shot conditions, and (c) showed a consistent trend (that did not always reach significance) toward a shorter streak correct in the no-delay condition than in the long-delay and one-shot conditions. These results replicate those of Experiment 1, but now with a younger age range. (See also the results from a corroborative combined analysis of all 5- to 6-year-old children from all three experiments, $N = 125$, in Tables S13 and S14 in the Supplemental Material.) In addition, these results show that even if there is a substantial cost to access the external resource (as in the long-delay condition), this does not immediately elicit maximal effort toward internal resource use. Children gave yet more effort (longer dwell times) in the one-shot condition. In short, children balanced the costs (effort) associated with using internal resources with those of accessing external resources.

General Discussion

In our shopping-game paradigm, children were asked to shop for items in a virtual store using a shopping list. The list and the store were not visible simultaneously, but children could opt to toggle between the store and the list. When given the chance and when the accessibility cost was low, 5- to 8-year-olds revisited their shopping list relatively frequently, spent a relatively short time studying their shopping list, and chose relatively few items on each visit to the store. However, when we increased the accessibility cost (imposing a 4-s delay until the list appeared), children made fewer trips, studied longer, and chose more correct items per trip. These results show that 5- to 8-year-old children adaptively trade off the use of internal versus external resources according to the accessibility of the external resources.

In our second experiment, we focused on 5- to 6-year-olds and tested children in two different sites (in China and in the United States). In addition to confirming the same findings regarding the trade-off as in Experiment 1, we tested children's metacognition by asking them which of the games was easier and which one they would prefer to play again. We found that most 5- to 6-year-olds chose the no-delay condition as "easier" and "preferred" over the long-delay condition, indicating they had a subjective impression of the cognitive demands of the different conditions. However, it is also possible that children may have also found the long-delay condition more boring (because it actually took a longer time to complete it) and that this may affected their answers, especially with respect to the preference question. With that said, given that we know that our delay manipulation had an effect on how much cognitive effort children exerted, we believe it is most parsimonious to attribute the observed differences in the answers to the "which one was easier" question to the delay itself.

The extended mind and the trade-off between looking and remembering

Given that even 4- to 5-year-olds have been shown to choose to use external resources (Armitage et al., 2020; Bulley et al., 2020), it is not surprising that children in our study could take advantage of their shopping list to support playing the shopping game. However, it is surprising that our participants—especially the younger ones (5- to 6-year-olds)—could adaptively trade off external versus internal resource use in the face of changing task demands. In line with the literature discussed in the introduction, previous studies have shown that 4- to 7-year-olds are not good at using external resources selectively depending on the difficulty of a task (Armitage et al., 2020; Bulley et al., 2020). In addition, the trade-off we found requires the evaluation of the effort of different strategies, and studies have shown that spontaneous metacognition is only starting to emerge between 5 and 7 years of age (Niebaum & Munakata, 2020).

Our paradigm had a few design features that may have helped support this trade-off in younger children. First, visual cues can facilitate children's spontaneous metacognitive monitoring (Niebaum & Munakata, 2020). In fact, although 5-year-olds were not able to choose an easier task when given no visual cues or prompts (O'Leary & Sloutsky, 2017), they were able to choose when the visual cues were provided (Wang & Bonawitz, 2022). In our paradigm, children were informed of the differences between conditions with visual cues throughout the experiment. Second, more experience with a task can improve metacognitive monitoring of task difficulty (O'Leary, 2017) and one's own performance (Urban & Urban, 2021), even without feedback. Our paradigm simulated an everyday task—helping with shopping—so the child may have had experience from daily life that transferred advantageously.

Similarities and differences between the Chinese and U.S. samples

In Experiment 2, we found no interaction effect between site (United States and China) and condition, indicating that the groups were not significantly different in their ability to trade off the use of internal versus external resources. However, we did find a main effect of site on dwell time, showing that the Chinese sample may have devoted greater effort to encoding the shopping list (see a similar pattern for a drawing task in Long et al., 2023). It might be tempting to attribute this difference to the Asian culture's purported emphasis on hard work (see the criticism of this simple cultural essentialism in Kobakhidze et al., 2023), but we speculate that there is a more parsimonious factor. The Chinese sample was tested by researchers in their kindergarten, whereas the U.S. sample was tested by researchers primarily in a children's museum. We speculate that the more quiet setting in the kindergarten versus the more "free-exploration" museum environment might account for the greater overall effort (longer study times) in the Chinese sample (see Rance et al., 2023, for similar contextual effects). Again, this did not affect our main results, but it would be interesting to follow up to see how testing context affects overall task motivation.

Limits on the generalizability of our findings

Our convenience sampling yielded child participants largely from middle-income families and representing the racial and ethnic distribution of the northeastern United States and eastern China. These participants are likely familiar with tablet-based games. Although our sample sizes met the a priori requirements for our analyses, they were still relatively modest.

Conclusion

The ability to exploit external resources to extend one's mind is adaptive, allowing one to increase task performance and reduce cognitive effort. Here we show that even children as young as 5 to 6 years of age can extend their minds and do so flexibly in response to task demands. Given this, it will be interesting to see whether yet younger children—to whom very limited internal resources make external resources that much more valuable—show a similar trade-off. In addition, there is an element of practice and experience that may aid adaptive, cost-effective trade-offs, so future work on the influence of schooling, where the skilled use of external resources is explicitly taught, would be especially interesting.

Transparency

Action Editor: Evie Vergauwe
Editor: Simine Vazire
Author Contributions
Yibiao Liang: Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Software, Visualization, Writing – original draft; Writing – review & editing.
Erik Blaser: Conceptualization; Funding acquisition; Methodology; Resources; Supervision; Writing – original draft; Writing – review & editing.
Jia Ying Yi: Investigation; Writing –review & editing.
Liyang Sai: Investigation; Supervision; Resources; Writing – review & editing.
Zsuzsa Kaldy: Conceptualization; Funding acquisition; Methodology; Project administration; Resources; Supervision; Writing – original draft; Writing – review & editing.
Declaration of Conflicting Interests
The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.
Funding
This research was supported by the National Institutes of Health (Grant No. R15HD086658) awarded to Z. Kaldy and E. Blaser).
Artificial Intelligence
No artificial-intelligence-assisted technologies were used in this research or the creation of this article.
Open Practices
Experiment 1 disclosures. Preregistration: The hypotheses, methods, and analysis plan were preregistered (https://doi.org/10.17605/OSF.IO/9QRC8) on 06-18-2022, after data collection, which began on 05-21-2022 (only 6 participants were run during this period, and their data was not

accessed until all data collection was completed). There were major and minor deviations from the preregistration (for details, see Table S1 in the Supplemental Material available online). Materials: All study materials are publicly available (https://doi.org/10.17605/OSF.IO/A2U48). Data: All primary data are publicly available (https://doi.org/10.17605/OSF.IO/A2U48). Analysis scripts: All analysis scripts are publicly available (https://doi.org/10.17605/OSF.IO/A2U48). Computational reproducibility: The computational reproducibility of the results has been independently confirmed by the journal's STAR team.

Experiment 2a disclosures. Preregistration: The hypotheses, methods, and analysis plan were preregistered (https://doi.org/10.17605/OSF.IO/DBKJS) on 02-28-2023, prior to collection, which began on 03-01-2023. There were major and minor deviations from the preregistration (for details, see Table S1 in the Supplemental Material). Materials: All study materials are publicly available (https://doi.org/10.17605/OSF.IO/A2U48). Data: All primary data are publicly available (https://doi.org/10.17605/OSF.IO/A2U48). Analysis scripts: All analysis scripts are publicly available (https://doi.org/10.17605/OSF.IO/A2U48). Computational reproducibility: The computational reproducibility of the results has been independently confirmed by the journal's STAR team.

Experiment 2b disclosures. Preregistration: The hypotheses, methods, and analysis plan were preregistered (https://doi.org/10.17605/OSF.IO/7CVS8) on 03-11-2023, prior to data collection, which began on 03-11-2023. There were major and minor deviations from the preregistration (for details, see Table S1 in the Supplemental Material). Materials: All study materials are publicly available (https://doi.org/10.17605/OSF.IO/A2U48). Data: All primary data are publicly available (https://doi.org/10.17605/OSF.IO/A2U48). Analysis scripts: All analysis scripts are publicly available (https://doi.org/10.17605/OSF.IO/A2U48). Computational reproducibility: The computational reproducibility of the results has been independently confirmed by the journal's STAR team.

ORCID iDs

Yibiao Liang https://orcid.org/0000-0003-3049-4307
Erik Blaser https://orcid.org/0000-0002-9968-8946
Zsuzsa Kaldy https://orcid.org/0000-0002-0679-4505

Acknowledgments

We would like to thank Tessyia Roper for her help with participant recruitment, the staff at the children's museums for their help and support, and the families who participated.

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/09567976241306424

Notes

1. We preregistered the measures *d′* and *C*, but because of the high number of trips without errors (~90%), they could not be properly calculated. Instead, we opted for streak correct. For completeness, we include *d′* and *C* analyses in all relevant tables in the Supplemental Material, which show the same overall pattern of results.
2. We preregistered an exploratory analysis of learning (trial) effects. There were no significant effects. Results are shared in Table S7 in the Supplemental Material. In addition, we preregistered that we would collect additional data (e.g., children's education level and their screen time) for exploratory analyses, but we did not analyze this data because we realized during data collection that our survey was not well designed.
3. We had preregistered that we would exclude trials with dwell times > 30 s on the basis of data from Experiment 1, but we failed to anticipate that some children might dwell longer in the one-shot condition. Therefore, instead of excluding these values, we winsorized them by setting all dwell times > 30 s to 31 s (Shete et al., 2004) for all experiments. Results using the strict exclusion (see Tables S11–S12 in the Supplemental Material) show similar patterns.

References

Agrawal, M., Mattar, M. G., Cohen, J. D., & Daw, N. D. (2022). The temporal dynamics of opportunity costs: A normative account of cognitive fatigue and boredom. *Psychological Review*, *129*(3), 564–585. https://doi.org/10.1037/rev0000309

Ahmed, S. F., Ellis, A., Ward, K. P., Chaku, N., & Davis-Kean, P. E. (2022). Working memory development from early childhood to adolescence using two nationally representative samples. *Developmental Psychology*, *58*(10), 1962–1973. https://doi.org/10.1037/dev0001396

Armitage, K. L., Bulley, A., & Redshaw, J. (2020). Developmental origins of cognitive offloading. *Proceedings of the Royal Society B: Biological Sciences/The Royal Society*, *287*(1928), Article 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*(1), 25–38. https://doi.org/10.1111/cdev.13664

Ascher, M., & Ascher, R. (1981). *Code of the Quipu* (pp. 56–74). University of Michigan Press. https://courses.cit.cornell.edu/quipu/data/title-2.pdf

Badre, D., & Nee, D. E. (2018). Frontal cortex and the hierarchical control of behavior. *Trends in Cognitive Sciences*, *22*(2), 170–188. https://doi.org/10.1016/j.tics.2017.11.005

Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience*, *7*(1), 66–80. https://doi.org/10.1162/jocn.1995.7.1.66

Bartoń, K. (2023). *MuMIn: Multi-model inference* (Version 1.47.5). https://cran.r-project.org/web/packages/MuMIn/index.html

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

Ben-Shachar, M., Lüdecke, D., & Makowski, D. (2020). Effectsize: Estimation of effect size indices and standardized parameters. *Journal of Open Source Software*, *5*(56), Article 2815. https://doi.org/10.21105/joss.02815

Botvinick, M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624–652. https://doi.org/10.1037/0033-295x.108.3.624

Bulley, A., McCarthy, T., Gilbert, S. J., Suddendorf, T., & Redshaw, J. (2020). Children devise and selectively use tools to offload cognition. *Current Biology*, *30*(17), 3457–3464.e3. https://doi.org/10.1016/j.cub.2020.06.035

Chater, N., & Oaksford, M. (1999). Ten years of the rational analysis of cognition. *Trends in Cognitive Sciences*, *3*(2), 57–65. https://doi.org/10.1016/s1364-6613(98)01273-x

Chatham, C. H., & Badre, D. (2015). Multiple gates on working memory. *Current Opinion in Behavioral Sciences*, *1*, 23–31. https://doi.org/10.1016/j.cobeha.2014.08.001

Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension*. Oxford University Press. https://play.google.com/store/books/details?id=n5wRDAAAQBAJ

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, *58*(1), 7–19. https://doi.org/10.1093/analys/58.1.7

Draschkow, D., Kallmayer, M., & Nobre, A. C. (2021). When natural behavior engages working memory. *Current Biology*, *31*(4), 869–874.e5. https://doi.org/10.1016/j.cub.2020.11.013

Egner, T. (2017). *The Wiley handbook of cognitive control*. John Wiley & Sons. https://play.google.com/store/books/details?id=y75KDwAAQBAJ

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Fox, J., & Weisberg, S. (2018). *An R companion to applied regression*. Sage. https://play.google.com/store/books/details?id=uPNrDwAAQBAJ

Freire, M. R., & Pammer, K. (2020). Influence of culture on visual working memory: Evidence of a cultural response bias for remote Australian Indigenous children. *Journal of Cultural Cognitive Science*, *4*(3), 323–341. https://doi.org/10.1007/s41809-020-00063-4

Gallagher, S. (2018). The extended mind: State of the question. *The Southern Journal of Philosophy*, *56*(4), 421–447. https://doi.org/10.1111/sjp.12308

Gilbert, S. J., Bird, A., Carpenter, J. M., Fleming, S. M., Sachdeva, C., & Tsai, P.-C. (2020). Optimal use of reminders: Metacognition, effort, and cognitive offloading. *Journal of Experimental Psychology: General*, *149*(3), 501–517. https://doi.org/10.1037/xge0000652

Grinschgl, S., Papenmeier, F., & Meyerhoff, H. S. (2021). Consequences of cognitive offloading: Boosting performance but diminishing memory. *Quarterly Journal of Experimental Psychology*, *74*(9), 1477–1496. https://doi.org/10.1177/17470218211008060

Harrell, F. E., & Dupont, C. (2023). *Hmisc: Harrell miscellaneous* (Version 5.1-1). https://hbiostat.org/r/hmisc/

Haselen, G. C. L.-V., Van Der Steen, J., & Frens, M. A. (2000). Copying strategies for patterns by children and adults. *Perceptual and Motor Skills*, *91*(2), 603–615. https://doi.org/10.2466/pms.2000.91.2.603

Hayhoe, M. M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences*, *9*(4), 188–194. https://doi.org/10.1016/j.tics.2005.02.009

Hayhoe, M. M., Bensinger, D. G., & Ballard, D. H. (1998). Task constraints in visual working memory. *Vision Research*, *38*(1), 125–137. https://doi.org/10.1016/s0042-6989(97)00116-8

Hoffman, J. E., Landau, B., & Pagani, B. (2003). Spatial breakdown in spatial construction: Evidence from eye fixations in children with Williams syndrome. *Cognitive Psychology*, *46*(3), 260–301. https://doi.org/10.1016/s0010-0285(02)00518-2

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, Theory and Applications*, *6*(2), 65–70. http://www.jstor.org/stable/4615733

Inamdar, S., & Pomplun, M. (2003). *Comparative search reveals the tradeoff between eye movements and working memory use in visual tasks* [Paper presentation]. Proceedings of the Annual Meeting of the Cognitive Science Society, 25, Boston, MA. https://escholarship.org/content/qt4k01x26b

Inzlicht, M., Shenhav, A., & Olivola, C. Y. (2018). The effort paradox: Effort is both costly and valued. *Trends in Cognitive Sciences*, *22*(4), 337–349. https://doi.org/10.1016/j.tics.2018.01.007

Kelly, M. O., & Risko, E. F. (2022). Study effort and the memory cost of external store availability. *Cognition*, *228*, Article 105228. https://doi.org/10.1016/j.cognition.2022.105228

Kenderla, P., & Kibbe, M. M. (2023). Explore versus store: Children strategically trade off reliance on exploration versus working memory during a complex task. *Journal of Experimental Child Psychology*, *225*, Article 105535. https://doi.org/10.1016/j.jecp.2022.105535

Király, I., Takács, S., Kaldy, Z., & Blaser, E. (2017). Preschoolers have better long-term memory for rhyming text than adults. *Developmental Science*, *20*(3), Article 12398. https://doi.org/10.1111/desc.12398

Kobakhidze, M. N., Ma, Y., & Tsaloukidis, A. (2023). Tiger parenting beyond cultural essentialism: Discourses of class, culture, and competition in Hong Kong. *ECNU Review of Education*, *7*(3), Article 20965311231198252. https://doi.org/10.1177/20965311231198252

Kool, W., & Botvinick, M. (2018). Mental labour. *Nature Human Behaviour*, *2*(12), 899–908. https://doi.org/10.1038/s41562-018-0401-9

Kool, W., Gershman, S. J., & Cushman, F. A. (2017). Cost-benefit arbitration between multiple reinforcement-learning systems. *Psychological Science*, *28*(9), 1321–1333. https://doi.org/10.1177/0956797617708288

Kool, W., McGuire, J. T., Rosen, Z. B., & Botvinick, M. M. (2010). Decision making and the avoidance of cognitive demand. *Journal of Experimental Psychology: General*, *139*(4), 665–682. https://doi.org/10.1037/a0020198

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*, 1–26. https://doi.org/10.18637/jss.v082.i13

Leger, K. R., & Gutchess, A. (2021). Cross-cultural differences in memory specificity: Investigation of candidate mechanisms. *Journal of Applied Research in Memory and Cognition*, *10*(1), 33–43. https://doi.org/10.1016/j.jarmac.2020.08.016

Lenth, R. V. (2023). *Emmeans: Estimated marginal means, aka least-squares means*. https://CRAN.R-project.org/package=emmeans

Lieder, F., & Griffiths, T. L. (2019). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, *43*, Article e1. https://doi.org/10.1017/S0140525X1900061X

Long, B., Wang, Y., Christie, S., Frank, M. C., & Fan, J. E. (2023). Developmental changes in drawing production under different memory demands in a U.S. and Chinese sample. *Developmental Psychology*, *59*(10), 1784–1793. https://doi.org/10.1037/dev0001600

Niebaum, J. C., Chevalier, N., Guild, R. M., & Munakata, Y. (2019). Adaptive control and the avoidance of cognitive control demands across development. *Neuropsychologia*, *123*, 152–158. https://doi.org/10.1016/j.neuropsychologia.2018.04.029

Niebaum, J. C., Chevalier, N., Guild, R. M., & Munakata, Y. (2021). Developing adaptive control: Age-related differences in task choices and awareness of proactive and reactive control demands. *Cognitive, Affective & Behavioral Neuroscience*, *21*(3), 561–572. https://doi.org/10.3758/s13415-020-00832-2

Niebaum, J. C., & Munakata, Y. (2020). Deciding what to do: Developments in children's spontaneous monitoring of cognitive demands. *Child Development Perspectives*, *14*(4), 202–207. https://doi.org/10.1111/cdep.12383

O'Leary, A. P. (2017). *Using scaffolding to examine the development of metacognitive monitoring and control* (Doctoral dissertation, The Ohio State University). Retrieved from https://etd.ohiolink.edu/acprod/odb_etd/ws/send_file/send?accession=osu1492172107513459&disposition=inline

O'Leary, A. P., & Sloutsky, V. M. (2017). Carving metacognition at its joints: Protracted development of component processes. *Child Development*, *88*(3), 1015–1032. https://doi.org/10.1111/cdev.12644

O'Reilly, R. C., & Frank, M. J. (2006). Making working memory work: A computational model of learning in the prefrontal cortex and basal ganglia. *Neural Computation*, *18*(2), 283–328. https://doi.org/10.1162/089976606775093909

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Plato. (1997). *Plato: complete works* (J. M. Cooper & D. S. Hutchinson, Eds.). Hackett Publishing Co. https://play.google.com/store/books/details?id=fMhgDwAAQBAJ

Qu, Y., Jorgensen, N. A., & Telzer, E. H. (2021). A call for greater attention to culture in the study of brain and development. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *16*(2), 275–293. https://doi.org/10.1177/1745691620931461

Rance, G., Dowell, R. C., & Tomlin, D. (2023). The effect of classroom environment on literacy development. *NPJ Science of Learning*, *8*(1), 9. https://doi.org/10.1038/s41539-023-00157-y

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

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences*, *20*(9), 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*(3), 537–564. https://doi.org/10.1111/cogs.12087

RStudio Team. (2015). *RStudio: Integrated development environment for R. RStudio*. http://www.rstudio.com/

Sahakian, A., Gayet, S., Paffen, C. L. E., & Van der Stigchel, S. (2023). Mountains of memory in a sea of uncertainty: Sampling the external world despite useful information in visual working memory. *Cognition*, *234*, Article 105381. https://doi.org/10.1016/j.cognition.2023.105381

Sarkar, D. (2008). *Lattice: Multivariate data visualization with R* (1st ed.). Springer. https://doi.org/10.1007/978-0-387-75969-2

Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The expected value of control: An integrative theory of anterior cingulate cortex function. *Neuron*, *79*(2), 217–240. https://doi.org/10.1016/j.neuron.2013.07.007

Shenhav, A., Fahey, M. P., & Grahek, I. (2021). Decomposing the motivation to exert mental effort. *Current Directions in Psychological Science*, *30*(4), 307–314. https://doi.org/10.1177/09637214211009510

Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., & Botvinick, M. M. (2017). Toward a rational and mechanistic account of mental effort. *Annual Review of Neuroscience*, *40*, 99–124. https://doi.org/10.1146/annurev-neuro-072116-031526

Shete, S., Beasley, T. M., Etzel, C. J., Fernández, J. R., Chen, J., Allison, D. B., & Amos, C. I. (2004). Effect of winsorization on power and type 1 error of variance components and related methods of QTL detection. *Behavior Genetics*, *34*(2), 153–159. https://doi.org/10.1023/B:BEGE.0000013729.26354.da

Singh, L., Cristia, A., Karasik, L. B., Rajendra, S. J., & Oakes, L. M. (2023). Diversity and representation in infant research: Barriers and bridges toward a globalized science of infant development. *Infancy: The Official Journal of the International Society on Infant Studies*, *28*(4), 708–737. https://doi.org/10.1111/infa.12545

Somai, R. S., Schut, M. J., & Van der Stigchel, S. (2020). Evidence for the world as an external memory: A tradeoff between internal and external visual memory storage. *Cortex*, *122*, 108–114. https://doi.org/10.1016/j.cortex.2018.12.017

Taylor, S. E. (1981). The interface of cognitive and social psychology. In J. H. Harvey (Ed.), *Cognition, social* *behavior, and the environment* (pp. 189–211). Lawrence Erlbaum.

Urban, K., & Urban, M. (2021). Effects of performance feedback and repeated experience on self-evaluation accuracy in high- and low-performing preschool children. *European Journal of Psychology of Education*, *36*(1), 109–124. https://doi.org/10.1007/s10212-019-00460-6

Venables, W. N., & Ripley, B. D. (2013). *Modern applied statistics with S*. Springer. https://doi.org/10.1007/978-0-387-21706-2

Verschooren, S., Schindler, S., De Raedt, R., & Pourtois, G. (2019). Switching attention from internal to external information processing: A review of the literature and empirical support of the resource sharing account. *Psychonomic Bulletin & Review*, *26*(2), 468–490. https://doi.org/10.3758/s13423-019-01568-y

Wang, J., & Bonawitz, E. (2022). Children's sensitivity to difficulty and reward probability when deciding to take on a task. *Journal of Cognition and Development*, *4*(3), 341–353. https://doi.org/10.1080/15248372.2022.2152032

Ward, A. F. (2021). People mistake the internet's knowledge for their own. *Proceedings of the National Academy of Sciences*, *118*(43), Article e2105061118. https://doi.org/10.1073/pnas.2105061118

Westbrook, A., & Braver, T. S. (2015). Cognitive effort: A neuroeconomic approach. *Cognitive, Affective & Behavioral Neuroscience*, *15*(2), 395–415. https://doi.org/10.3758/s13415-015-0334-y

Westbrook, A., Kester, D., & Braver, T. S. (2013). What is the subjective cost of cognitive effort? Load, trait, and aging effects revealed by economic preference. *PLOS ONE*, *8*(7), Article e68210. https://doi.org/10.1371/journal.pone.0068210

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., . . . Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), Article 1686. https://doi.org/10.21105/joss.01686