Children devise and selectively use tools to offload cognition

Adam Bulley[1,2*], Thomas McCarthy[3], Sam J. Gilbert[4], Thomas Suddendorf[3], and Jonathan Redshaw[3]

[1] Department of Psychology, Harvard University, Cambridge, MA, USA, 02138

[2] The University of Sydney, School of Psychology and Brain and Mind Centre, Australia, 2006

[3] Early Cognitive Development Centre, School of Psychology, The University of Queensland, St. Lucia, Australia, 4072

[4] Institute of Cognitive Neuroscience, University College London, London, UK, WC1N 3AZ

*Corresponding author: adam_bulley@fas.harvard.edu

Summary

From maps sketched in sand to supercomputing software, humans ubiquitously enhance cognitive performance by creating and using artifacts that bear mental load [1–5]. This extension of information processing into the environment has taken center-stage in debates about the nature of cognition in humans and other animals [6–9]. How does the human mind acquire such strategies? In two experiments we investigated the developmental origins of cognitive offloading in 150 children aged between 4 and 11 years. We created a memory task in which children were required to recall the location of hidden targets. In one experiment, participants were provided with a pre-specified cognitive offloading opportunity: an option to mark the target locations with tokens during the hiding period. Even 4-year-old children quickly adopted this external strategy and, in line with a metacognitive account, across ages children offloaded more often when the task was more difficult. In a second experiment, we provided children with the means to devise their own cognitive offloading strategy. Very few younger children spontaneously devised a solution, but by ages 10 and 11 nearly all did so. In a follow-up test phase, a simple prompt greatly increased the rate at which the younger children devised an offloading strategy. These findings suggest that sensitivity to the difficulties of thinking arises early in development and improves throughout the early school years, with children learning to modify the world around them to compensate for their cognitive limits.

Results and discussion

Experiment 1. When and how do children begin to offload cognition with tools?

To investigate cognitive offloading in young children, we created a simple hiding game paradigm (Fig. 1) inspired by Piagetian object permanence tasks [10,11] but with drastically reduced chance performance levels. A circular array of 25 cups was arranged in front of participants (experiment 1: n = 80, mean age = 7.39 years, range = 4.01 years to 11.95 years). Targets (stickers) were hidden underneath the cups, and participants were instructed to try to find the rewards by choosing the correct cups after a 5-second time delay. Children first completed the game without any opportunity for cognitive offloading (*phase 1*), and then again after being introduced to an optional strategy with which they could offload cognition by placing tokens atop the target cups while the stickers were being hidden (*phase 2*). While previous studies have provided children with external cues that act as reminders for delayed intentions [e.g. 12–17], here we were interested in how young children would set their own reminders as a means of offloading cognition.

We varied the number of hidden stickers (1 vs 5) to investigate whether children would set proportionately more reminders in a more difficult version of the task. Previous work has shown that adults tend to be selective in their use of cognitive offloading, taking into account the costs of time and effort that these strategies require and thus offloading more frequently in situations of higher internal demand [18–20]. Although the cost of setting reminders in the current experiments is minimal, the design nonetheless allowed us to track changes in children's propensity to account for these costs across development.

INSERT FIGURE ONE HERE

In both phases, we scored search accuracy as the proportion of target cups searched after the delay (out of all cups searched), such that scores could range from 0 (no target cups searched) to 1 (all target cups searched), and we analyzed these data with General Linear Models. Children were more accurate in easy trials than in hard trials, confirming the efficacy of the difficulty manipulation (M_{easy} = 0.91, SE = .02, M_{hard} = 0.70, SE = .02; $F(1,78)$ = 113.96, $p < .001$, η_p^2 = .59). In line with our predictions, search accuracy was significantly higher in phase 2 (when the cognitive offloading strategy was allowed) than phase 1 (when it was not; M_{phase1} = 0.75, SE = .02, M_{phase2} = 0.87, SE = .02; $F(1,78)$ = 34.63, $p < .001$, η_p^2 = .31).

Furthermore, older children were more accurate than younger ones ($F(1,78) = 31.52$, $p < .001$, $\eta_p^2 = .29$), with age correlating with overall search accuracy at r (78) = .54.

The difference in children's accuracy between easy and hard trials was larger when children could not offload cognition than when they could ($M_{diff\_phase1} = 0.32$, $SE = .03$, $p < .001$, $M_{diff\_phase2} = 0.10$, $SE = .03$, $p < .001$; 2-way interaction $F(1,78) = 40.05$, $p < .001$, $\eta_p^2 = .34$), suggesting the cognitive offloading strategy supplemented performance in the harder trials. This narrowing of the easy/hard performance gap between phases was greater in the younger children than the older children (3-way interaction $F(1,78) = 5.98$, $p = .017$, $\eta_p^2 = .07$), likely because older children had a higher baseline proficiency on the hard trials.

Children's cognitive offloading scores were operationalized as the difference between the proportion of target cups marked and the proportion of non-target cups marked (as has been done in conceptually similar computer-based studies of cognitive offloading in adults [19,21]). Participants' cognitive offloading scores could therefore range from -1 (i.e., in the unlikely case that a child marked every non-target cup but no target cups) to 1 (i.e., in the case that a child marked every target cup but no non-target cups); and a participant who did not mark any cups on a given trial received a score of 0.

Out of all the tokens that participants placed on cups, 95.18% were placed on target cups. Overall, children had larger cognitive offloading scores in the hard trials than the easy trials ($M_{hard} = .57$, $SE = .05$, $M_{easy} = .19$, $SE = .04$; $F(1,78) = 63.64$, $p < .001$, $\eta_p^2 = .45$), indicating they were *selective* in their use of the tokens as a function of task difficulty. Average cognitive offloading scores did not significantly vary with age, $F(1, 78) = 3.75$, $p = .057$, $\eta_p^2 = .05$, and nor did the degree of selective offloading (2-way interaction $F(1,78) = 3.21$, $p = .077$, $\eta_p^2 = .04$; see Fig. 2). Post-hoc analyses revealed that, controlling for age, offloading scores in the easy condition did not significantly predict search accuracy in phase two, r (77) = .18, $p = .121$, whereas offloading scores in the hard condition *did* significantly predict search accuracy in phase two, r (77) = .86, $p < .001$.

In order to visualize the results (Fig. 2), participants were divided into 4 age groups: 4/5-year-olds ($n = 25$), 6/7-year-olds ($n = 27$), 8/9-year-olds ($n = 17$), and 10/11-year-olds ($n = 11$). Post-hoc testing indicated even the 4- and 5-year-old children offloaded cognition selectively, with a significantly higher offloading score in hard trials than easy trials, $p = .004$, $\eta_p^2 = .11$; as did the older children (all other p values < .003, η_p^2 values ranged from .12 to .25; Fig 2.C). Further post-hoc testing indicated that cognitive offloading scores on easy trials significantly decreased with increasing age, r (78) = -.32, $p = .003$, whereas scores on hard trials did not significantly correlate with age, r (78) = -.05, $p = .669$.

To ascertain the proportions of children who were offloading cognition selectively and non-selectively, each child was assigned to one of four mutually exclusive response categories (Fig 2.E; see STAR methods for details). Consistent with the GLM findings, at least 30% of children from all age groups offloaded cognition selectively. Many of the younger children offloaded cognition non-selectively (in both hard and easy conditions), whereas few children aged 8 years and older did so.

INSERT FIGURE TWO HERE

Experiment 2. How and when do children devise their own cognitive offloading strategies?

In the second experiment, we provided children with the opportunity to devise their own cognitive offloading strategy (n = 70, mean age = 7.95 years, range = 4.06 years to 11.83 years). This enabled us to (i) explore whether young children intuit cognitive offloading as a means to bolster cognition by themselves, and (ii) determine whether patterns observed in experiment 1 could merely reflect that children were following experimenter requests.

We used the same paradigm as experiment 1, with a few modifications. Instead of tokens, participants were given a non-permanent marker pen, which could be used to write on the target cups to offload cognition. No instructions were provided about using the pen to perform cognitive offloading. However, we took steps to pre-empt permission problems by implementing an initial drawing-on-cups training phase, and by introducing children to an observing puppet who "loves to watch people draw on the cups" (with no mention of drawing on *target cups*; see STAR methods). To incentivize children to devise a cognitive offloading strategy if they could, we heightened the difficulty of the unaided memory task by increasing the delay between target hiding and retrieval to 30 seconds. The pen was available throughout all 8 trials, but after the first 4 trials (*phase 1*) the experimenter gave participants a simple prompt by pointing to the pen and asking "Have you thought of a way to use that to help you remember where the stickers are?" (*phase 2*).

Once again, children were more accurate in easy trials than in hard trials, confirming the efficacy of the difficulty manipulation (M_{easy} = .84, SE = .02, M_{hard} = .68, SE = .02; $F(1,68)$ = 49.40, p < .001, η_p^2 = .42). This difference in accuracy between the easy and hard trials significantly decreased with age (see Fig. 3), potentially because older children were more

likely to devise the offloading strategy (2-way interaction $F(1,68) = 8.04$, $p = .006$, $\eta_p^2 = .11$). Search accuracy was significantly improved after the prompt was provided at the start of phase 2 ($M_{phase1} = .71$, SE = .02, $M_{phase2} = .82$, SE = .02; $F(1,68) = 17.39$, $p < .001$, $\eta_p^2 = .20$). Furthermore, older children were more accurate than younger ones, $F(1,68) = 72.46$, $p < .001$, $\eta_p^2 = .52$), with age correlating with accuracy at r (68) = .72.

Out of all the cups marked during phase 1, 89.22% were target cups. Out of all the cups marked during phase 2, 90.07% were target cups. Unlike in experiment 1, there was no significant difference in cognitive offloading scores between easy and hard trials ($M_{hard} = .39$, SE = .03, $M_{easy} = .37$, SE = .03; $F(1,68) = 1.75$, $p = .191$, $\eta_p^2 = .03$). However, and again in contrast to experiment 1, older children were significantly more likely to devise and employ the cognitive offloading strategy across conditions, $F(1, 68) = 51.14$, $p < .001$, $\eta_p^2 = .43$. These results indicate an age-related improvement in devising the cognitive offloading strategy, but also little difficulty-related selectivity in the deployment of this strategy once devised. Levels of cognitive offloading were significantly increased after the prompt was provided at the start of phase 2 ($M_{phase1} = .21$, SE = .03, $M_{phase2} = .55$, SE = .05; $F(1,68) = 68.57$, $p < .001$, $\eta_p^2 = .50$). Post-hoc analyses revealed that, controlling for age, offloading scores in the easy condition did not significantly predicted search accuracy in phase 1, $r (67) = .04$, $p = .773$, but did in phase 2, $r (67) = .38$, $p = .001$). Offloading scores in the hard condition also significantly predicted search accuracy in phase 1, $r (67) = .56$, $p < .001$, and in phase 2, $r (67) = .91$, $p < .001$).

As in experiment 1, each child was assigned to one of four mutually exclusive response categories for each phase (Fig 3.E, see STAR methods for details). None of the 4- and 5-year-olds devised the offloading strategy prior to receiving the prompt, although the rate of doing so increased with age. By 10 and 11 years nearly all children conceived of the strategy even without having been prompted (16 out of 18). After the prompt, the proportion of offloaders substantially increased among children aged 4 to 7 years, from approximately 9% (3 out of 35) to approximately 43% (15 out of 35). However, consistent with the GLM findings, and in contrast to experiment 1, very few children of any age offloaded selectively in either phase (9 out of 70 in phase 1; 5 out of 70 in phase 2).

INSERT FIGURE THREE HERE

Many theorists posit that the mind acquires much of its power from its ability to create and interface with external objects, props, aids, scaffolding, and tools [1,2,28–30,4,7,22–27]. While other species act upon their environment in ways that may feed back to influence cognition, such as when ants leave pheromone trails that guide themselves and other ants [31–33], humans appear to flexibly and intentionally exploit the environment for cognitive ends in unparalleled fashion [34,35]. Complementing a long tradition of research into the development of meta-memory and memory strategies [36,37,46,47,38–45], our findings empirically chart how human children first begin to augment their cognitive processing with external support to achieve their goals. Children used the tokens and pens to create external representations of hidden stickers [23], thereby transforming otherwise inert objects into cognitive artifacts: tools to aid in cognitive processes [2,48,49]. By setting reminders in such fashion, people establish "retrieval cues for communicating with themselves in the future" [50]. Our results reveal that this propensity for both devising and selectively using tools for self-communication between present and future arises during the preschool years, thereby laying the foundations for an "extended mind" [1,9].

Our findings demonstrate that when provided with a cognitive offloading strategy, around a third of even 4- and 5-year old children choose to use it selectively, in line with task demands. Such calibration implies a degree of metacognitive insight and control [4,18,51–53], given that children must differentiate between situations where offloading will and will not benefit performance (a previous study with more a more complicated task only found selective offloading in children aged around 9 years and older [52]). Nonetheless, when participants had to devise their own cognitive offloading strategy in experiment 2, we found very little evidence for selective offloading. Adults, on the other hand, have been shown to selectively offload cognition both when strategies are instructed and when they are spontaneously generated [20]. One likely explanation is that, because the delay time between target hiding and target retrieval was longer in experiment 2, children who devised the offloading strategy were more likely to use it globally "just in case". This explanation is consistent with post-hoc analyses showing that children benefited from offloading in phase 2 of the easy condition of experiment 2 (after pen use had been prompted), but not in the easy condition of experiment 1 (after token use had been introduced).

In experiment 2, children's rates of devising the offloading strategy increased linearly with age, with none of the 4- and 5-year-olds but nearly all 10- and 11-year-olds using the pen to mark target cups in phase 1. Interestingly, this pattern for devising thinking tools parallels some influential findings on children's ability to devise mechanical tools. One study, for

instance, found that fewer than 10% of 4- to 5-year-olds spontaneously bent a wire into a hook in order to retrieve a reward from a tube, although a large majority of 10- to 11-year-olds did so [54]. Notably, rates of offloading in our second experiment considerably increased in children aged 4 to 7 after a simple prompt. Again, this finding parallels patterns in the mechanical tool innovation literature showing that rates of tool use or modification in younger children substantially increase when affordances are made particularly obvious or explicitly prompted [55–58]. It is perhaps not surprising that children's cognitive offloading propensity strongly benefits from prompting, given that many memory enhancement strategies are cultural practices inherited from others [29,59]. An important avenue for future research will be to test the hypothesis that the developmental processes are similar whether tools are devised for performing physical work or cognitive work.

Humans create environmental scaffolding that overcomes the limits of raw biological computation: we write lists, keep records, program computers, sketch maps, use calendars, and set reminders. This cognitive offloading of mental labor into the external world presents somewhat of a paradox: to fully understand cognition, we must look outside the brain, to account for how mere objects get transformed into thinking tools. The present findings suggest that the basic capacities to learn, flexibly deploy, and spontaneously devise offloading strategies using tools emerges during the preschool years and continues to develop throughout childhood.

Author contributions

Conceptualization, A.B. T.M, and J.R.; Methodology, A.B. T.M, S.G., T.S., and J.R. Formal analysis, A.B., T.M., and J.R.; Investigation, A.B., and T.M.; Data Curation, A.B., T.M., and J.R., Writing – Original Draft, A.B., and J.R.; Writing – Reviewing & Editing, A.B. T.M, S.G., T.S., and J.R. Visualization, A.B., T.M., and J.R.; Supervision, T.S., and J.R., Project Administration, A.B., and J.R.

Declaration of interests

The authors declare no competing interests.

Figure titles and legends

Fig. 1. The cognitive offloading task.

(Ai- Aii) Children sat opposite an experimenter in front of an array of hiding locations. (Aiii) The experimenter hid either 1 or 5 target stickers. (Bi – Bii) The stickers were hidden in a pre-determined order that began and ended with distractor cups that were simply lifted and put back in place, here in order: 23 (distractor), 1, 7, 11, 14, 19, 20 (distractor). (Biii) Children were introduced to a cognitive offloading strategy (*tokens* in Exp. 1) or were given the means to devise their own (a *pen* in Exp. 2) and could use these strategies during target-hiding. After a delay (5s in Exp. 1, 30s in Exp. 2), participants were invited to search under the same number of cups as the number of hidden stickers.

Fig. 2. Experiment 1 results.

(A, B) Children's search accuracy grouped by age, across difficulty level and phase (no offloading vs offloading). Each datapoint is one child's performance, with some minor jitter added for discriminability. (C) Cognitive offloading scores in phase 2 grouped by age and across difficulty level. Each datapoint is one child's cognitive offloading score, with some minor jitter added for discriminability. (D) Each child was assigned to one of four mutually exclusive categories: non-markers, who did not mark any cups on any trial; indiscriminate markers, who marked at least one cup but did not mark target cups significantly more than non-

target cups; non-selective offloaders, who showed clear evidence of cognitive offloading in both easy and hard conditions; and selective offloaders, who showed clear evidence of cognitive offloading in the hard condition but not the easy condition. N's throughout: 4-5 years = 25; 6-7 years = 27; 8-9 years = 17, 10-11 years = 11.

Fig 3. Experiment 2 results.

(A, B) Children's search accuracy grouped by age, across difficulty level and phase (no prompt vs prompt). Each datapoint is one child's performance, with some minor jitter added for discriminability. (C, D) Cognitive offloading scores grouped by age, across difficulty level, and phase. Each datapoint is one child's cognitive offloading score, with some minor jitter added for discriminability. (E) Cumulative percentage of children from each age group who showed each category of cognitive offloading response in each phase (see Fig. 2.E. caption for definitions). N's throughout: 4-5 years = 17; 6-7 years = 18; 8-9 years = 17, 10-11 years = 18.

STAR Methods

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Adam Bulley (adam_bulley@fas.harvard.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The datasets generated during this study are available at the Open Science Framework. Experiment 1: https://osf.io/k48dx/. Experiment 2: https://osf.io/ywbge/.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human subjects

Across both experiments, 150 children were recruited and participated at the Queensland Museum in Brisbane, Australia. Experiment 1 included 80 children (41 female, 39 male, mean age = 7.39 years, range = 4.01 years to 11.95 years), and experiment 2 included 70 children (34 female, 36 male, mean age = 7.95 years, range = 4.06 years to 11.83 years). Parents or guardians provided verbal or written consent prior to testing, and children received a small gift for participating as well as any stickers they retrieved during the task. All study protocols were approved by the relevant ethics boards of the University of Queensland, Australia.

In experiment one, 7 additional participants were removed from analysis for falling outside the age range of interest ($n = 2$), experimenter error ($n = 2$), parental interference ($n = 2$), and for unwillingness to complete the task ($n = 1$). In experiment two, 17 additional participants were removed from analysis because they had observed another participant completing the study (and may have thus seen the pen used as a reminder-setting strategy; n = 9), unwillingness to complete the task ($n = 4$) or because they had clinical diagnoses ($n = 4$). For 13 participants in experiment 1, information regarding clinical diagnoses was not available due to data loss. However, these participants were included in analyses based on the low likelihood of having a diagnosis, with only 4 of 161 other total original participants (~2.5%) excluded for this reason.

METHOD DETAILS

Open science practices

Hypotheses, measures, and our analytical plan were pre-registered with the Open Science Framework for both experiments (experiment 1: https://osf.io/k48dx/; experiment 2: www.osf.io/qkm9j/).

Apparatus, materials, and procedure

Children sat opposite the experimenter in front of a 50 cm task board comprising a field of 25 opaque cups (7 cm in diameter). The cups were placed in 2 concentric rings (8 cups inner, 17 cups outer). The cups in the inner ring were placed at an angle of 45° to one another from the centre of the board, with their centres 11cm from the centre of the board. The cups in the outer ring were placed at an angle of approximately 21° to one another from the centre of the board, with their centres 20.5cm from the centre of the board. Adobe Illustrator was used to create a virtual diagram of the apparatus to these specifications, which was then exported to a CNC 1290 CO2 Laser Cutter to cut holes into a 50cm diameter, circular sheet of polypropylene, which was used as the top layer of the apparatus. Another 50cm circular sheet of polypropylene, without holes, comprised the base of the apparatus and was separated from the top of the apparatus by 2cm. The apparatus was constructed in this way to ensure that the cups would remain stable during and between trials.

Experiment 1 had a 2 (difficulty: easy/ hard) x 2 (phase: no cognitive offloading phase 1/ cognitive offloading phase 2) within-participants design, with age in days measured as a between-participants continuous variable (mean-centered for analyses). The design of experiment 2 was identical other than the fact that phase 1 included no instruction about how to use the offloading strategy (but it was available), and phase 2 included a simple prompt. In both experiments, participants completed 8 trials of the main task (4 trials in each phase). For each trial, the task difficulty was either easy (1 target hidden) or hard (5 targets hidden), with 4 trials of each difficulty in total. The order of trial difficulty was counterbalanced between participants (either easy first or hard first), and then alternated within-participants (i.e. 1-5-1-5 or 5-1-5-1). Targets (stickers) were hidden under the target cups with the cup being picked up by the experimenter in one hand and the sticker being placed underneath it with the other hand before the cup was returned to the board. After all stickers were hidden, a delay commenced (5 seconds counted down verbally by the experimenter in experiment 1; 30 seconds measured by a sand timer in experiment 2), before participants were instructed to choose which cup (or cups) they thought had the reward inside. Participants were given the same number of guesses

as the number of hidden stickers. Before the start of each trial, children were told how many stickers were going to be hidden in that trial.

The cups under which the stickers were hidden were predetermined pseudo-randomly using a random number generator, under the constraint that no two stickers were hidden under adjacent cups of the same circle. One distractor cup, under which no sticker was hidden, was lifted before hiding the stickers, and another distractor cup was lifted after hiding the stickers. During the delay after hiding, all actions by participants were permissible with the exception of attempting to lift a cup, when the experimenter would tell the participant to wait until the delay had finished.

In experiment 1, participants were provided a bucket of tokens at the start of phase 2 (25 tokens in the bucket; 1 for each cup). Participants were told that they could place tokens on the target cups while the stickers were being hidden to help them remember where the stickers were and were shown how to use this option. They were told that the token strategy was completely optional, and they could use it if they wanted to, but that they did not have to. Six participants had the cognitive offloading strategy explained again after the first trial of phase 2 because they appeared not to understand how to use the tokens or asked for clarification. In experiment 2, participants were provided with a non-permanent marker pen before the start of phase 1. The pen ink could be readily erased from the top of the cups between trials. It was important that children knew this, and also knew that they were allowed to draw on the cups without being told how this could be used to offload cognition. To resolve this problem, we therefore included a warm-up phase to make it clear that participants could mark the cups and that these markings were not permanent. Prior to the testing phases, the experimenter used the pen to draw on two cups in view of the participant, before handing the children the pen and inviting them to draw on the cups themselves. After participants had drawn on a number of cups, the experimenter erased the markings with a cloth to demonstrate that the markings were not permanent.

Participants were then introduced to a toy monkey named "Cup-Monkey". They were informed that Cup-Monkey "loves to watch people draw on the cups" and that they were allowed to draw on the cups whenever they liked when Cup-Monkey was around. Participants were then informed that Cup-Monkey would be seated next to them for the entire experiment, before being invited to draw on more cups for Cup-Monkey to watch. If the child had not done so already, they were asked to draw on some of the cups on the half of the board closest to the experimenter, to help them feel comfortable reaching over the board towards the experimenter. All markings were erased from the cups before children proceeded to the next phase.

Throughout the two test phases (at the start of the third, fifth and seventh trials), participants were given 'Cup-Monkey' reminders. Participants were asked "Who is this?" (as the experimenter pointed at Cup-Monkey, and "What does it mean when Cup-Monkey is around?". Regardless of the response to the participant's answer to the second question, the experimenter replied, "It means we are allowed to draw on the cups whenever we want." In both experiments, participants could use their cognitive offloading strategy while the hiding phase was in progress (throughout the time the experimenter was hiding the stickers).

In both experiments, and before the first phase commenced, children were asked a question assessing their metacognitive knowledge into task difficulty across conditions. The question was: "Would it be easier to remember where one sticker is hidden, where five stickers are hidden, or would it be the same?" The first two options were counterbalanced, but the "same" option was always presented last. However, children (especially younger children) frequently asked for clarification and appeared unsure how to answer the question. Furthermore, due to an experimenter error the question was asked differently in experiments 1 and 2. In experiment 1, the following answer prompt was included: "What would be easier?", and this inadvertently revealed that the response to the question was easily biased by such extraneous factors (i.e., there were far fewer responses to the "same" answer when the prompt was included). Given the complexity of the question, its revealed sensitivity to framing, dependence on linguistic competence, and different administration between experiments, we lacked confidence in the validity of the measure. We thus opted not to analyze the metacognitive self-report data, and focused on the behavioral data instead. For previous self-report studies of children's insight into reminder setting strategies see [50,60].

QUANTIFICATION AND STATISTICAL ANALYSIS

For each trial, we recorded the number of correctly retrieved targets. For easy trials this number was out of 1, and for hard trials it was out of 5. A derived accuracy score was then calculated by averaging across the 2 trials of each difficulty (easy vs. hard) within each phase (no cognitive offloading vs. cognitive offloading for experiment 1; pre-prompt vs. post-prompt for experiment 2). Accuracy was analyzed with a General Linear Model including within-subjects main effects of difficulty and phase, a continuous mean-centered main effect of age, and the full factorial set of interactions between these 3 variables. Similar to previous studies that used a conceptually related paradigm in adults [19,21,61], the focal cognitive offloading measure was operationalised as the proportion of target cups marked minus the proportion of non-target cups marked. For example, if a participant marked 4 out of 5 target cups and 1 out

of 20 non-target cups on a hard trial, then their cognitive offloading score for that trial would be (4/5) – (1/20) = 0.75. For the GLM analyses, each participant's cognitive offloading score was averaged across the two trials of each difficulty within each phase. The rationale for this measure is that participants occasionally marked non-target cups, which does not constitute cognitive offloading. The measure was therefore selectively directed toward target cups, corrected for any general tendency to mark cups. For experiment 1, this dependent variable was entered into a General Linear Model including the within-subject main effect of difficulty (easy vs. hard), a continuous, mean-centered main effect of age, and the interaction between these 2 variables. For experiment 2, the General Linear Model additional included a phase effect (pre-prompt vs. post-prompt) and the full factorial set of two-way and three-way interactions.

To distinguish genuine cognitive offloading from indiscriminate marking of cups at the individual level (see stacked bar charts in main text), each child's marking behaviour across both trials of each condition was analysed using Fisher's exact test. For these analyses, the independent data points were the 50 unique cups that children could mark across the two trials of each condition (within each phase). For example, if a participant marked 7 of the 10 target cups and 12 of the 40 non-target cups across the two hard trials of a given phase, the participant would be classed as a cognitive offloader for hard trials in that phase, $p = .030$. However, if the participant had instead marked 7 of the 10 target cups and 13 of the 40 non-target cups, they would be classed as an indiscriminate marker in that phase, $p = .067$. Using this approach, participants who were categorised as cognitive offloaders on both easy and hard trials (or on easy trials only) within a phase were ultimately categorised as *non-selective offloaders* in Figs 2 and 3. Participants who were categorised as cognitive offloaders for hard trials but not easy trials (i.e., they were either indiscriminate markers or did not mark at all on easy trials) were categorised as *selective offloaders* in Figs 2 and 3. Participants who did not meet either of these criteria were categorised as *indiscriminate markers* if they marked at least one cup on either type of trial, or as *non-markers* if they did not mark any cups at all.

All statistical analyses were performed in SPSS 24. Figures were created using Microsoft Excel and R-Studio (version 3.6.1) [62], using the ggplot2 package [63].

Supplemental information

Supplemental information includes 1 figure and additional descriptive statistics.

References

1.   Clark, A., and Chalmers, D. (1998). The extended mind. Analysis *58*, 7–19.

2.   Sterelny, K. (2010). Minds: Extended or scaffolded? Phenomenol. Cogn. Sci. *9*, 465–481.

3.   Sutton, J. (2006). Exaograms and Interdisciplinarity: history, the extended mind, and the civilizing process.

4.   Risko, E.F., and Gilbert, S.J. (2016). Cognitive offloading. Trends Cogn. Sci. *20*, 676–688.

5.   Storm, B.C., Stone, S.M., and Benjamin, A.S. (2017). Using the Internet to access information inflates future use of the Internet to access other information. Memory *25*, 717–723.

6.   Kaplan, D.M. (2012). How to demarcate the boundaries of cognition. Biol. Philos. *27*, 545–570.

7.   Japyassú, H.F., and Laland, K.N. (2017). Extended spider cognition. Anim. Cogn. *20*, 375–395.

8.   Adams, F., and Aizawa, K. (2001). The bounds of cognition. Philos. Psychol. *14*, 43–64.

9.   Bocanegra, B.R., Poletiek, F.H., Ftitache, B., and Clark, A. (2019). Intelligent problem-solvers externalize cognitive operations. Nat. Hum. Behav. *3*. Available at: http://dx.doi.org/10.1038/s41562-018-0509-y.

10.   Piaget, J. (1954). The construction of reality in the child (London: Routledge).

11.   Barth, J., and Call, J. (2006). Tracking the displacement of objects: A series of tasks with great apes (Pan troglodytes, Pan paniscus, Gorilla gorilla, and Pongo pygmaeus) and young children (Homo sapiens). J. Exp. Psychol. Anim. Behav. Process. *32*, 239–252.

12.   Guajardo, N.R., and Best, D.L. (2000). Do preschoolers remember what to do? Incentive and external cues in prospective memory. Cogn. Dev. *15*, 75–97.

13.   Kliegel, M., and Jäger, T. (2007). The effects of age and cue-action reminders on event-based prospective memory performance in preschoolers. Cogn. Dev. *22*, 33–46.

14.   Redshaw, J., Henry, J.D., and Suddendorf, T. (2016). Disentangling the effect of event-based cues on children's time-based prospective memory performance. J. Exp. Child Psychol. *150*, 130–140. Available at: http://dx.doi.org/10.1016/j.jecp.2016.05.008.

15.   Kvavilashvili, L., and Ford, R.M. (2014). Metamemory prediction accuracy for simple

prospective and retrospective memory tasks in 5-year-old children. J. Exp. Child Psychol. *127*, 65–81. Available at: http://dx.doi.org/10.1016/j.jecp.2014.01.014.

16.   Meacham, J.A., and Colombo, J.A. (1980). External retrieval cues facilitate prospective remembering in children. J. Educ. Res. *73*, 299–301.

17.   Mahy, C.E.V., Mazachowsky, T.R., and Pagobo, J.R. (2018). Do verbal reminders improve preschoolers' prospective memory performance? It depends on age and individual differences. Cogn. Dev. *47*, 158–167. Available at: https://doi.org/10.1016/j.cogdev.2018.06.004.

18.   Gilbert, S.J., Bird, A., Carpenter, J.M., Fleming, S.M., Sachdeva, C., and Tsai, P. (2020). Optimal Use of Reminders: Metacognition, Effort, and Cognitive Offloading. J. Exp. Psychol. Gen. *44*, 1–39.

19.   Gilbert, S.J. (2015). Strategic use of reminders: Influence of both domain-general and task-specific metacognitive confidence, independent of objective memory ability. Conscious. Cogn. *33*, 245–260. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1053810015000070.

20.   Boldt, A., and Gilbert, S.. (2019). Confidence Guides Spontaneous Cognitive Offloading. Cogn. Res. Princ. Implic. *4*.

21.   Gilbert, S.J. (2015). Strategic offloading of delayed intentions into the external environment. Q. J. Exp. Psychol., 1–22. Available at: http://www.tandfonline.com/doi/abs/10.1080/17470218.2014.972963.

22.   Clark, A. (2008). Supersizing the mind: Embodiment, action, and cognitive extension (Oxford University Press).

23.   Heersmink, R. (2013). A Taxonomy of Cognitive Artifacts: Function, Information, and Categories. Rev. Philos. Psychol. *4*, 465–481.

24.   Zhang, J. (1997). The Nature of External Representations in Problem Solving. Cogn. Sci. *21*, 179–217.

25.   Dunn, T.L., and Risko, E.F. (2016). Toward a Metacognitive Account of Cognitive Offloading. Cogn. Sci. *40*, 1080–1127.

26.   Hutchins, E. (1995). Cognition in the Wild (MIT press).

27.   Rowlands, M.J. (2010). The new science of the mind: From extended mind to embodied phenomenology (MIT Press).

28.   Mithen, S. (2000). Mind, brain and material culture: An archaeological perspective. In Evolution and the human mind: Modularity, language and meta-cognition, P. Carruthers and A. Chamberlain, eds. (Cambridge: Cambridge University Press), pp.

207–217.

29.   Donald, M. (1991). Origins of the modern mind: Three stages in the evolution of
       culture and cognition (Harvard University Press).

30.   Kirsh, D. (2010). Thinking with external representations. Artif. Intell. Soc. *25*, 441–
       454. Available at: http://gunpowder.ucsd.edu/kirsh/articles/interaction/Kirsh-
       interaction.pdf%5Cnhttp://adrenaline.ucsd.edu/kirsh/articles/Interaction/Kirsh-
       Interaction.pdf.

31.   Sterelny, K. (2003). Thought in a hostile world: The evolution of human cognition.

32.   Czaczkes, T.J., Grüter, C., and Ratnieks, F.L.W. (2015). Trail Pheromones: An
       Integrative View of Their Role in Social Insect Colony Organization. Annu. Rev.
       Entomol. *60*, 581–599. Available at:
       http://www.annualreviews.org/doi/10.1146/annurev-ento-010814-020627.

33.   Odling-Smee, F.J., Laland, K.N., and Feldman, M.W. (2013). Niche construction: the
       neglected process in evolution (Princeton university press).

34.   Redshaw, J., and Bulley, A. (2018). Future-thinking in animals: Capacities and Limits.
       In The Psychology of Thinking about the Future, G. Oettingen, A. T. Sevincer, and P.
       M. Gollwitzer, eds. (Guilford Press), pp. 31–51.

35.   Bulley, A., Redshaw, J., and Suddendorf, T. (2019). The future-directed functions of
       the imagination: From prediction to metaforesight. In The Cambridge Handbook of the
       Imagination, A. Abraham, ed. (Cambridge: Cambridge University Press), pp. 1–50.

36.   Neldner, K., Collier-Baker, E., and Nielsen, M. (2015). Chimpanzees (Pan troglodytes)
       and human children (Homo sapiens) know when they are ignorant about the location
       of food. Anim. Cogn. *18*, 683–699.

37.   Balcomb, F.K., and Gerken, L.A. (2008). Three-year-old children can access their own
       memory to guide responses on a visual matching task. Dev. Sci. *11*, 750–760.

38.   Dunlosky, J., and Metcalfe, J. (2008). Metacognition (Sage Publications).

39.   Flavell, J.H., and Wellman, H.M. (1975). Metamemory.

40.   Dufresne, A., and Kobasigawa, A. (1989). Children's spontaneous allocation of study
       time: Differential and sufficient aspects. J. Exp. Child Psychol. *47*, 274–296.

41.   Lockl, K., and Schneider, W. (2004). The effects of incentives and instructions on
       children's allocation of study time. Eur. J. Dev. Psychol. *1*, 153–169.

42.   Metcalfe, J., and Finn, B. (2013). Metacognition and control of study choice in
       children. Metacognition Learn. *8*, 19–46.

43.   Masur, E.F., McIntyre, C.W., and Flavell, J.H. (1973). Developmental changes in

apportionment of study time among items in a multitrial free recall task. J. Exp. Child Psychol. *15*, 237–246.

44.   Heisel, B.E. (1981). Young Children ' s Storage a Memory-for-Location Behavior Task. J. Exp. Child Psychol. *31*, 350–364.

45.   Miller, P.H., Seier, W.L., Probert, J.S., and Aloise, P.A. (1991). Age differences in the capacity demands of a strategy among spontaneously strategic children. J. Exp. Child Psychol. *52*, 149–165.

46.   Bjorklund, D.F., Dukes, C., and Brown, R.D. (2009). The development of memory strategies. In Studies in Developmental Psychology. The Development of Memory in Infancy and Childhood., M. L. Courage and N. Cowan, eds. (Hove, East Sussex: Psychology Press), pp. 145–175.

47.   Roebers, C.M. (2014). Children's Deliberate Memory Development: The Contribution of Strategies and Metacognitive Processes. In The Wiley Handbook on the Development of Children's Memory, Vol II, P. J. Bauer and R. Fivush, eds. (New York: Wiley), pp. 865–894.

48.   Hutchins, E. (1999). Cognitive artifacts. MIT Encycl. Cogn. Sci. *126*, 127.

49.   Norman, D.A. (1991). Cognitive artifacts. Des. Interact. Psychol. human-computer interface *1*, 17–38.

50.   Beal, C.R. (1985). Development of Knowledge about the Use of Cues to Aid Prospective Retrieval. Child Dev. *56*, 631–642.

51.   Risko, E.F., and Dunn, T.L. (2015). Storing information in-the-world: Metacognition and cognitive offloading in a short-term memory task. Conscious. Cogn. *36*, 61–74. Available at: http://dx.doi.org/10.1016/j.concog.2015.05.014.

52.   Redshaw, J., Vandersee, J., Bulley, A., and Gilbert, S.J. (2018). Development of children's use of external reminders for hard-to-remember intentions. Child Dev. *89*, 2099–2108. Available at: http://doi.wiley.com/10.1111/cdev.13040.

53.   Dunn, T.L., Lutes, D.J.C., and Risko, E.F. (2016). Metacognitive evaluation in the avoidance of demand. J. Exp. Psychol. Hum. Percept. Perform. *42*, 1372–1387.

54.   Beck, S.R., Apperly, I.A., Chappell, J., Guthrie, C., and Cutting, N. (2011). Making tools isn't child's play. Cognition *119*, 301–306.

55.   Neldner, K., Mushin, I., and Nielsen, M. (2017). Young children's tool innovation across culture: Affordance visibility matters. Cognition *168*, 335–343. Available at: http://dx.doi.org/10.1016/j.cognition.2017.07.015.

56.   Neldner, K., Redshaw, J., Murphy, S., Tomaselli, K., Davis, J., Dixson, B., and

Nielsen, M. (2019). Creation across culture: Children's tool innovation is influenced by cultural and developmental factors. Dev. Psychol. *55*, 877–889.

57.   Chen, Z., and Siegler, R.S. (2000). Across the great divide: bridging the gap between understanding of toddlers' and older children's thinking. Monogr. Soc. Res. Child Dev. *65*.

58.   Gredlein, J.M., and Bjorklund, D.F. (2005). Sex differences in young children's use of tools in a problem-solving task: The role of object-oriented play. Hum. Nat. *16*, 211–232.

59.   Vygotsky, L.S. (1978). Mind in Society: The Development of Higher Psychological Processes M. Cole, V. John-Steiner, S. Scribner, and E. Souberman, eds. (Cambridge, MA: Harvard University Press).

60.   Kreutzer, M.A., Leonard, C., and Flavell, J.H. (1975). An Interview Study of Childrens Knowledge About Memory. Monogr. Soc. Res. Child Dev. *40*, 1–60.

61.   Cherkaoui, M., and Gilbert, S.J. (2017). Strategic use of reminders in an 'intention offloading' task: Do individuals with autism spectrum conditions compensate for memory difficulties? Neuropsychologia *97*, 140–151.

62.   Team, Rs. (2018). RStudio: Integrated Development for R. Available at: http://www.rstudio.com/.

63.   Wickham, H. (2009). ggplot2 Elegant Graphics for Data Analysis Available at: http://had.co.nz/ggplot2/book.