Cognitive Development 59 (2021) 101068

Contents lists available at ScienceDirect

Cognitive Development

journal homepage: www.elsevier.com/locate/cogdev

Thinking inside the box: Mental manipulation of working memory contents in 3- to 7-year-old children

Eva Reindl[a,*], Divya Parkash[a], Christoph Johannes Völter[a,b], Amanda Madeleine Seed[a]

[a] School of Psychology and Neuroscience, University of St Andrews, St Mary's Quad, South Street, St Andrews KY16 9JP, UK
[b] Comparative Cognition Messerli Research Institute, University of Veterinary Medicine Vienna, Medical University of Vienna, University of Vienna, Veterinärplatz 1, 1210 Vienna, Austria

ARTICLE INFO

Keywords:
Working memory
Mental manipulation
Mental rotation
Executive functions
Updating

ABSTRACT

We developed a non-verbal task assessing updating and manipulation of working memory contents. 80 3- to 7-year-olds (47 girls; predominantly European White) were tested with a 4 × 4 grid containing 8 boxes (in the 4 centre cells and 4 outer corners). A sticker was hidden and children searched for it after a delay phase. In the updating trials, the grid was rotated during delays, in the manipulation trials, the grid was both occluded and rotated. Rewards were hidden in either the inner or outer boxes (between-subjects design). Performance was affected by age, rotation degree and hiding condition. Performance was better in outer boxes trials, where visual tracking was easier. Occluded inner trials added a substantial cognitive load (which increased with degree of rotation), resulting in children performing at chance level, suggesting that manipulation involving mental rotation is a distinct skill from tracking invisible object displacement, with a more protracted development.

1. Introduction

Executive functioning lies at the heart of any cognitive processing and plays a crucial role in cognitive development throughout childhood and beyond (Cowan, 2016; Diamond, 2013; Garon, Bryson, & Smith, 2008; van der Sluis, de Jong, & van der Leij, 2007). Two core aspects of executive functions are the updating and manipulation of information in mind. *Updating* describes the capacity to revise information held in memory in light of new input and/or to replace old with new information (Morris & Jones, 1990; van der Sluis et al., 2007). *Manipulation* refers to the ability to re-order items in mind, to relate them, and to integrate new pieces of information with existing knowledge (Diamond, 2013; Garon et al., 2008). While researchers in developmental psychology often use the terms updating and working memory (henceforth WM) interchangeably (see e.g., Garon et al., 2008; but see Lee & Bull, 2015), adult research has usually used them to describe distinct concepts (Ecker, Lewandowsky, Oberauer, & Chee, 2010; Frischkorn, von Bastian, Souza, & Oberauer, 2020; Miyake et al., 2000; Morra, Panesi, Traverso, & Usai, 2018). In that line, WM is often described as the temporarily activated set of representations in mind which require attentional resources to maintain active (Barrouillet & Camos, 2010; Cowan, 2017; Engle, 2002; Morra et al., 2018). Here, we also aim to keep the two terms updating and WM separate, in order to avoid any potential misunderstandings. We focus on the two processes updating and manipulation as defined above, but assume that these

* Corresponding author.
E-mail address: eva.reindl@live.de (E. Reindl).

https://doi.org/10.1016/j.cogdev.2021.101068
Received 12 February 2021; Received in revised form 10 May 2021; Accepted 19 May 2021
Available online 27 May 2021

processes operate on representations that are currently held active in mind, i.e., in WM (Ecker et al., 2010; Miyake et al., 2000).

Research from the last three decades has shown that the mental processes that can be carried out on WM representations undergo rapid improvements during the pre-school period and beyond across a variety of tasks (Cowan, 2016; Garon et al., 2008; Hughes, 2011; Simmering & Perone, 2013). With increasing age, children are able to perform increasingly complex cognitive operations. Infants from 5 months of age are able to hold information in mind over a short delay (i.e., a few seconds) and to retrieve it when prompted (measured e.g., with the delayed response task; Diamond & Doar, 1989). This *storage and retrieval* capacity is often labelled short-term memory, with the term WM being reserved for a cognitively more complex phenomenon. However, storage and retrieval are often seen as indispensable parts of WM (Cowan, 2014). Over infancy and the pre-school years, the length of delay and the number of remembered items increase. Forward span tasks (e.g., digit span and word span tasks) are often used to assess storage and retrieval abilities in older children, with the number of items that can be remembered increasing across the pre-school years and beyond (Garon et al., 2008).

Updating operations are already present in the second year of life (Evrard et al., 2011; Wiebe, Lukowski, & Bauer, 2010) and undergo equally rapid changes in the pre-school and school years (Bull, Espy, & Wiebe, 2008; Espy et al., 2004; Jenkins & Berthier, 2014; Wiebe, Espy, & Charak, 2008). Compared to research with older children and adults (Miyake et al., 2000; van der Sluis et al., 2007), fewer tasks exist to assess updating in pre-schoolers, as many of the tasks developed for adults rely on an understanding of letters and digits. Usually, self-ordered pointing tasks such as the Scrambled Box or the Spin the Pots tasks are used in research with young children as they require only few instructions (Diamond, 1997; Hughes & Ensor, 2007). Note that the authors introducing these tasks often do not themselves label the tasks as updating tasks but refer to them generally as WM tasks. This is in line with the observation that developmental researchers often use the terms updating and WM interchangeably. Other, more recent updating tasks for young children are the Magic House task, in which children watch sequences of toy animals being placed into a house (the sequences are of different lengths) and need to remember the last two animals that were placed (Panesi & Morra, 2020), or adaptations of the Keep Track task, a task similar to the Magic House task but which seems to pose slightly higher demands on resistance to interference as participants are required to label all the items presented (Traverso, Viterbori, & Usai, 2015; van der Sluis et al., 2007; Ven, Kroesbergen, Boom, & Leseman, 2012). The Noisy Book Task (Hughes, 1998) has previously also been described as an updating task (Garon et al., 2008). In this task, children are presented with a 3 × 3 array of pictures that each produce a sound when pressed. In the test, the pictures are covered and children are presented verbally with a sequence of items whose pictures they need to press in order (again, sequences differ in length between trials). However, the Noisy Book Task might be better described as a simple forward span task, in which the updating demand does not stem from the design of the trial itself but results from the need to erase memory traces from the previous trial when starting a new sequence (for a similar critique, see below our comment on the updating trials of the task by Boudreau et al., 2018). From school age onward, updating is usually assessed by Letter/Digit Memory tasks, Keep track or n-back tasks (Kirchner, 1958; Miyake et al., 2000; St Clair-Thompson & Gathercole, 2006; van der Sluis et al., 2007). Note that there is another set of tasks – complex span tasks – assessing the simultaneous storage and processing of information, aiming to assess WM capacity, which we do not focus on here (for more information see e.g., Bayliss, Jarrold, Gunn, & Baddeley, 2003; Conway et al., 2005; Engle, 2002; Kane et al., 2004).

Regarding the manipulation of WM content, there is a relative shortage of tasks suitable for pre-schoolers (Boudreau, Dempsey, Smith, & Garon, 2018; Garon et al., 2008). Studies have attempted to investigate WM manipulation in children as young as 3 years of age using backward word and digit span tests or the backward Corsi span task (Garon et al., 2008). However, these tasks have been adapted from tasks for adults and due to their verbal and representational demands might underestimate children's capacities. Indeed, studies have shown floor effects for 3-year-olds on the backward digit span task (Bull et al., 2008; Carlson, 2005; Carlson, Moses, & Breton, 2002; Davis & Pratt, 1995) and Carlson et al. (2002) reported that the majority of children's errors resulted from children repeating the sequence forward, which suggests that at least the youngest children did not understand the task.

A similar issue arises with the backward Corsi span task – a spatial version of the backward span test – in which participants are presented with a board of nine blocks and observe the experimenter touch a sequence of blocks, after which they are asked to repeat the touches in the reverse order (Berch, Krikorian, & Huha, 1998; Milner, 1971). While this task poses somewhat fewer additional cognitive demands, it still requires an understanding of the concept "reverse" or "backward". Again, there is evidence of floor effects in young pre-schoolers in some studies (Bull et al., 2008), which could stem from a lack of task validity at this age. Nevertheless, evidence from span tasks suggests that even though performance within age groups is strongly dependent on task features and the studied domain (e.g., verbal or visuo-spatial WM), from 4 to 5 years of age onwards children's ability to manipulate WM content increases throughout the primary school years and beyond (Carlson, 2005; Carlson et al., 2002; Gathercole, Pickering, Ambridge, & Wearing, 2004; Luciana, Conklin, Hooper, & Yarger, 2005; Simmering & Perone, 2013).

Invisible displacement tasks have also been suggested to capture manipulation of information in mind (Garon et al., 2008). In these tasks, an individual observes an object being hidden underneath a cup. Then, the experimenter moves the cup underneath a bigger cup (single displacement) or two bigger cups in sequence (double displacement) before retrieving it again. Finally, the individual is shown the now empty small cup. In order to find the object, the individual has to represent, update, and finally infer the location of the reward. Single displacement tasks can be solved by infants from 15 months of age, double displacement tasks from 24 months. However, there is the possibility that invisible displacement task can be solved by visual tracking (supported by object permanence), without the requirement to mentally manipulate the representation.

Recently, Boudreau et al. (2018) presented a task with minimal verbal demands to measure retrieval, updating, and manipulation within a single task for pre-schoolers. One advantage of using a single task is that it is possible to establish the validity of the empirical approach, by establishing good baseline performance on the most basic task requirements. The retrieval trial was a delayed response task in which children were presented with a box with several compartments and observed the experimenter put 1–3 toys behind the doors. 3-year-olds showed good performance in retrieving items after a 10 s delay with performance further increasing in 4- and 5-year-olds, as could be expected based on the findings from previous research that had shown that even 12-month-olds can remember the location of a single item for 10 s (Garon et al., 2008).

Updating was measured using trials with the same procedure as above but in which the toy was hidden behind different doors. The demand to update information about the location of the object did not occur within each trial (as e.g., in the invisible displacement task) but between trials. Therefore, Boudreau et al.'s (2018) updating trials do not seem to be an updating task; rather, they look like a series of different items of a retrieval task.

In their manipulation task, a magic wand was used to indicate a change in the location of a hidden toy. In two additional trials, the box was rotated by 180° after the removal of the wand. However, as with the invisible displacement task, the question arises whether these manipulation trials could be solved by visual tracking – first by visually following the movement of the wand across the box until it stopped and disappeared and then by following the (occluded) location during the rotation. If so, no manipulation of the memory content would be required. Another difficulty for the interpretation of the experiment comes from the fact that the manipulation score represented averaged performance from all magic wand trials. Performance on the rotation trials alone was not reported. Therefore, children's manipulation performance could have been overestimated.

Given these issues and the fact that in general only few tasks measuring manipulation of WM content exist, the current study set out to develop a novel task, suitable from 3 years of age on, measuring the retrieval, updating, and – crucially – the manipulation of WM content. We also aimed to investigate the role played by sustained visual attention, by manipulating how easy it is to keep an eye on the reward location. Following Boudreau and colleagues, we also drew upon mental rotation rather than using a span task to study the manipulation of WM content.

Mental rotation describes the ability to imagine a rotational movement of one or several objects in 2D or 3D space (Frick, Möhring, & Newcombe, 2014; Shepard & Metzler, 1971). In the classic mental rotation task, participants are asked to compare an image of an object to an image of a target object and judge whether the object is a rotated version (or rather a mirror image) of the target (Shepard & Metzler, 1971). The greater the angular difference between the two images, the longer participants need to make a judgment. For example, in Shepard and Metzler (1971) participants needed 1 s for a 0° rotation, about 3 s for a 90° rotation, and 4–6 s for a 180° rotation, indicating that participants simulate the rotation in their mind and need longer time the larger the required rotation is. Developmental work has shown a gradual, but steady progression of mental rotation abilities between the ages of 3 and 5 (Frick, Hansen, & Newcombe, 2013). By 5 years, children are capable of solving the classic mental rotation task, albeit performing at a slower speed than adults (Frick, Ferrara, & Newcombe, 2013; Frick, Hansen et al., 2013; Kosslyn, Margolis, Barrett, Goldknopf, & Daly, 1990; Marmor, 1975, 1977). In contrast, 3-year-olds' accuracy rates have been found to be at chance level (Frick, Ferrara et al., 2013; Frick, Hansen et al., 2013). However, when tasks demands are lowered – e.g., when requiring children to perform a goal-directed action on an object rather than asking them to compare two objects and make an explicit decision – 3-year-olds are capable of mental rotation (Krüger, Kaiser, Mahler, Bartels, & Krist, 2014). In 4-year-olds, mental rotation abilities are still not robust and depend on task difficulty: while some studies conclude that 4-year-olds are already fully capable of mental rotation (Marmor, 1975, 1977), research using more abstract images (Dean & Harvey, 1979), a more difficult procedure (e.g., denying children the chance to receive feedback to verify their choices; Frick, Ferrara et al., 2013; Frick, Hansen et al., 2013), or tangible, 3D stimuli (Hawes, Lefevre, Xu, & Bruce, 2015) has found 4-, and sometimes even 5-year-olds to perform at chance level. Furthermore, substantial individual differences in the capacity of mental rotation have been found at 4 years of age, highlighting the importance of this age window for the development of mental rotation abilities (Estes, 1998; Noda, 2010). Thus, as with WM manipulation tasks, performance in mental rotation tests is highly dependent on task features.

WM plays an important role in mental rotation (some might also view mental rotation as just another sort of WM manipulation): in order to rotate a mental representation of an image and align it to the target image, intermediate mental representations have to be generated, stored and manipulated. Dual task studies have found that general attentional resources are required during the rotation process: parallel tasks requiring object memory or other aspects of executive functions interfere with performance on mental rotation tasks (Bruyer & Scailquin, 1998; Hyun & Luck, 2007). Another study found that 55 % of the performance variance of 3- to 6-year-olds in a mental rotation task could be explained by these children's success in WM tests (Lehmann, Quaiser-Pohl, & Jansen, 2014).

The current study took inspiration from the mental rotation literature to develop a task suitable for children as young as 3 years of age to measure retrieval, updating, and one form of manipulation of WM content within a single task (following Boudreau et al., 2018): the *Rotating Grid Task*. We also aimed to create a task with minimal verbal requirements, in order to reduce task demands not directly related to WM. We used a rectangular grid containing eight boxes (four in the outer corners, four in the centre) as potential hiding places for a reward. In each trial, children observed one box being baited and had to identify the location of the reward after a retention interval. To measure retrieval, children had to remember the location of the reward for 5 s. In the updating trials, the box was rotated either 90°, 180°, or 270°. Here, children needed to update their representation of the location of the reward, but performance could be aided by visually tracking the box containing the reward. The manipulation trials were less easily solvable by visual tracking and the demand on mental manipulation increased: here, the box was occluded before the rotation. All children received all three phases (retrieval, updating, manipulation), but children differed by whether the reward was always hidden in one of the inner boxes or in one of the outer boxes. We assumed that performance of children in the *Outer compartment* condition would be better than that of children in the *Inner compartment* condition, as the edges of the box might serve as an additional visual anchor facilitating visual tracking in the Outer compartment condition – even in the occluded (manipulation) trials. Therefore, the need for mental manipulation – while increased by the occlusion of the box – might be only slight if one of the outer boxes is baited, compared to when one of the inner boxes is baited. We assumed that the inner boxes, by being further away from the box edge and by being closer to each other, were harder to visually track than the outer boxes, and even more so during occluded trials. Indeed, there is evidence that visual anchors play a facilitating role in visual search (Boettcher, Draschkow, Dienhart, & Võ, 2018). Due to the spatial proximity of an anchor to the target object, the anchor provides additional spatial information about the target object. Thus, we suspect that visual anchors can not only facilitate visual search, but also reduce the cognitive efforts involved in mental rotation.

2. Methods

2.1. Participants

The final sample was 80 children (47 girls, 33 boys) between 3y1m and 7y11 m (M = 68.97 months, SD = 16.53) recruited and tested in a Science Centre in a medium-sized town in Scotland, UK, where opportunity sampling was employed. Data were collected between December 2018 and February 2019. There were 11 3-year-olds, 16 4-year-olds, 15 5-year-olds, 20 6-year-olds, and 18 7-year-olds (see Table S1 for the cell sample sizes split by age and condition). The sample size reflects the maximum sample size that we were able to gather given the time constraints for this student project. No a priori sample size analysis was conducted; instead, we aimed to test as many children as possible in the available time window. The ethnic background was predominantly White (n = 65, 81 %), with 3 children (4 %) being of "Mixed race", and one child (1 %) Turkish. The ethnic background was undisclosed for 11 children (14 %). We tested another two children but excluded them from the analysis because of unknown age. Participants were randomly assigned to two conditions: *Outer compartment* (25 girls, 17 boys) and *Inner compartment* (22 girls and 16 boys). In addition to the Rotating Grid Task, children were administered a scrambled box task (Diamond, 1997). This task had been included for validation purposes, but was eventually not used for that purpose due to a ceiling effect in performance which prevented a meaningful correlational analysis between the scrambled box task and the Rotating Grid Task (see Supplementary Material for more information). Written informed consent was obtained by participants' parents or guardians prior to the study. Ethical approval was granted by the University of St Andrews, UK, School of Psychology and Neuroscience Ethical Review Committee.

2.2. Materials

In order to measure the WM operations retrieval, updating, and manipulation, we used a novel task, the Rotating Grid Task, consisting of a 4 × 4 grid (a plastic storage box, 22.5 × 28 × 6.5 cm, Fig. 1a). Eight blue mini storage boxes (6.5 × 8 × 4.5 cm) were inserted in the four corners and in the middle square of the grid. The remaining eight compartments stayed empty. For the manipulation phase of the experiment, the grid was occluded with a laminated white A4 sheet attached to the front with small pieces of Bluetac (Fig. 1b). Stickers were used as reward, which children put on an A4 coloured paper used as a sticker sheet. Video recordings were made using a handycam mounted on a tripod. A short video demonstrating the rotation of the box can be found as Supplementary Video 1.

2.3. Design and procedure

Children first played the Rotating Grid Task and were then administered the scrambled box task (not reported here). For the Rotating Grid Task, children were randomly assigned to one of two conditions: For children in the Outer compartment condition, stickers were hidden only in one of the four outer boxes (i.e., those in the corners). The four boxes in the middle of the grid stayed empty. For children in the Inner compartment condition, stickers were hidden only in one of the four inner boxes (i.e., those in the middle square). The four boxes in the four corners stayed empty. In both conditions, children were administered eight trials (Table 1, Fig. S1). The order of boxes used as hiding places was chosen semi-randomly so that in the first four trials each box was baited once (in

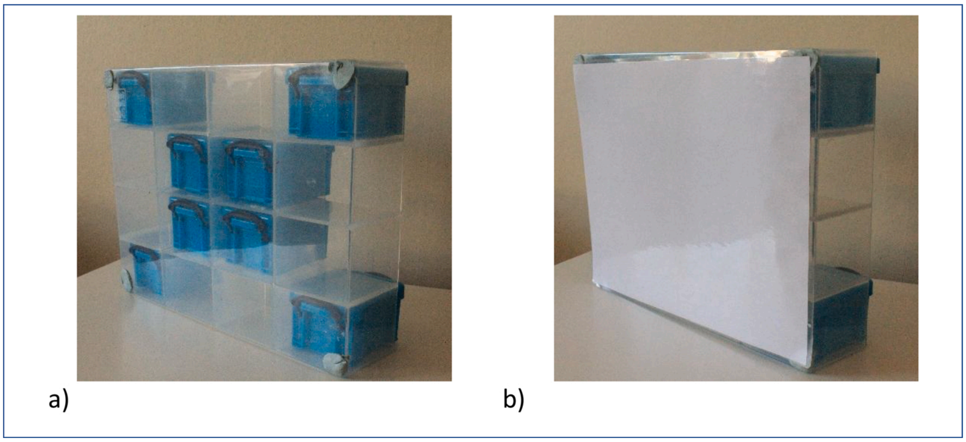

Fig. 1. (a) The Rotating Grid Task as seen by the participant in un-occluded trials. (b) In occluded trials, the front of the box was covered with a white sheet the size of the box.

random order) as well as in the second four trials. The order differed between children and was counterbalanced.

All children were tested by the same female experimenter. Children and their families were recruited in a Science Museum. The experimenter had a small stall set up in the museum, with signs informing about the study. Families interested in taking part could approach the stall and were given more information orally and through an information letter. Only those children took part who had given oral assent and whose parent or caregiver had given written informed consent. Children were informed that they could play a game in which they could win stickers, but that they could stop the game at any time. At the beginning of the task, children were encouraged to choose four stickers from a box which they wanted to win in the following game. If children had more than four trials correct, additional stickers were chosen by the experimenter. Once children had chosen their stickers and handed them over to the experimenter, the experimenter explained the game in a warm-up trial:

In this game our hands will always be flat on the table like this. I will take a sticker [experimenter takes sticker] *and put it in a box* [experimenter opens one of the boxes, randomly chosen (but always a box from the compartment that was relevant in the condition the child was in) puts the sticker inside, closes box, and puts it back into the grid]*, and then after a while I will ask you where the sticker is. Then you can point and tell me where you think the sticker is. Does this sound okay to you?*

The experimenter waited with the child for a few seconds and then asked: "So, where is the sticker?" The child was encouraged to find the sticker by pointing. The experimenter then opened the box the child pointed to and praised the child for finding the sticker. Then the experimenter explained that the child could stop the game at any time and they did a short practise on what the child could say if they wanted to stop the game. After this, the experimenter asked: "Are you ready to win some stickers?" and the first trial began. Trial 1 (Supplementary Video 2) measured retrieval; here, children had to keep the location of a sticker that the experimenter hid in one of the four boxes of the relevant compartment in mind for a short period of time (~ 5 s). During the retention interval, the experimenter broke any visual contact of the child with the location of the box by handing them a sheet of paper and saying "I will give you this sheet of paper, so you can stick your stickers on here and take the sheet home with you." Then children were encouraged to find the sticker by pointing to the box they thought the sticker was in. The experimenter retrieved and opened the box; if the correct box was chosen, the experimenter handed the sticker to the child. If an incorrect box was chosen, children were shown the empty box, the experimenter retrieved the sticker from the correct box and encouraged the child to try again in the next round.

In Trials 2–4 (Supplementary Video 3) we additionally measured children's ability to update information about the location of the sticker in their mind. Here, children watched a sticker being hidden in one of the boxes of the relevant compartment, then the experimenter said: "Now I'm going to turn the box, only point when I ask you where the sticker is.", after which the experimenter rotated the grid (90°, 180°, or 270°, order across Trials 2–4 randomised). After each 90° rotation, the experimenter put the grid back on the table for 1 s in order to emphasize each completed 90° rotation and to facilitate children's tracking of the rotation. The rotations were conducted in a standardized fashion in order to avoid any unintentional cueing by the experimenter's hand placement: The experimenter always picked the box up in the same way: the left hand on the top side of the box, the right hand on the right side of the box (from the experimenter's view). Then the experimenter turned the box 90° anti-clockwise (from their view), removed the hands from the box, and – in trials with more than one rotation – placed the hands on the top and side of the box again to do the next 90° rotation in the same fashion. After the final rotation, both hands were removed from the box. After the last rotation, the child was allowed to look for the sticker. In order to succeed in these trials, children had to remember where the sticker was hidden and to update the information about its location during the rotation (by tracking the location of the box).

The last four trials (Trials 5–8, Supplementary Video 4) measured manipulation of WM (in addition to retrieval). Here, children again watched a sticker being hidden, then the experimenter said: "Now I'm going to cover the box, only point when I ask you where the sticker is.", after which the experimenter attached an occluder at the front of the grid (same size as front panel of the grid) to cover the boxes and rotated the grid (with the occluder attached) 90°, 180°, or 270° (order across trials randomised); in one trial, there was no rotation (0°). In order to succeed, children had to hold in mind the location of the box and mentally manipulate its location (by mental rotation of the grid and its boxes). Total testing time was 10−15 min.

Table 1
Overview of the eight trials and three experimental phases that children in both conditions (Inner/Outer compartment) of the Rotating Grid Task received.

Trial	Phase	Occlusion	Rotation[a]	Description
1	Retrieval	No	0	E puts sticker in one box[b] within relevant compartment, breaks child's visual contact with the grid by asking to select a sticker sheet. Child then allowed to look for sticker.
2	Updating	No	180	E puts sticker in one box[b] within relevant compartment and rotates grid. Child then allowed to look for sticker.
3			270	
4			90	
5	Manipulation	Yes	180	E puts sticker in one box[b] within relevant compartment, occludes the front of the grid and rotates grid. Child then allowed to look for sticker.
6			90	
7			0	
8			270	

Notes. [a]In trials 2–4 and 5–8, the order of the rotation degrees was randomized. [b]The box was chosen in a pre-set, semi-random fashion so that for each child, each of the four boxes in the relevant compartment was baited twice within the 8 trials. E = Experimenter.

2.4. Hypotheses

We expected to find a three-way interaction between compartment, occlusion, and age. Firstly, we expected an interaction between compartment and occlusion in that occluded trials would be more difficult than non-occluded trials (as non-occluded trials could be aided by visual tracking and occluded trials would demand more mental manipulation) and that this effect of occlusion would be more pronounced in the Inner compartment condition. This is because in the Outer compartment condition the boxes are further apart, facilitating distinction, and children might be able to use the corners of the box as cues, which would still allow them to use visual tracking, whereas in the Inner compartment condition, the demand on mental manipulation coming from the occlusion is more pronounced. Secondly, we expected to find an increase in performance with age, and that this increase would be more pronounced in those conditions in which the demands on mental manipulation would be highest (e.g., in the occluded Inner compartment condition). Note that given the findings of previous studies on WM manipulation and mental rotation we could also have formulated more specific hypotheses regarding age (rather than just an increase in performance), for example that 3-year-olds would show a floor effect in the occluded Inner compartment condition. Yet, we only had a general and broad hypothesis with respect to age, as our task was completely novel and it was unclear how difficult it would be for children of different ages.

We also hypothesized that within each condition, performance would be negatively affected by the degree of rotation (which would replicate many previous findings, see e.g. Kaltner & Jansen, 2016; Shepard & Metzler, 1971). We also tested for an interaction effect between age and rotation, as several previous studies did so (but without finding any such effect; Frick, Ferrara et al., 2013; Frick, Hansen et al., 2013; Krüger et al., 2014; Noda, 2010). We also hypothesized that we would find an effect for the interaction between occlusion and rotation, in that the effect of rotation would be more pronounced in the occluded trials.

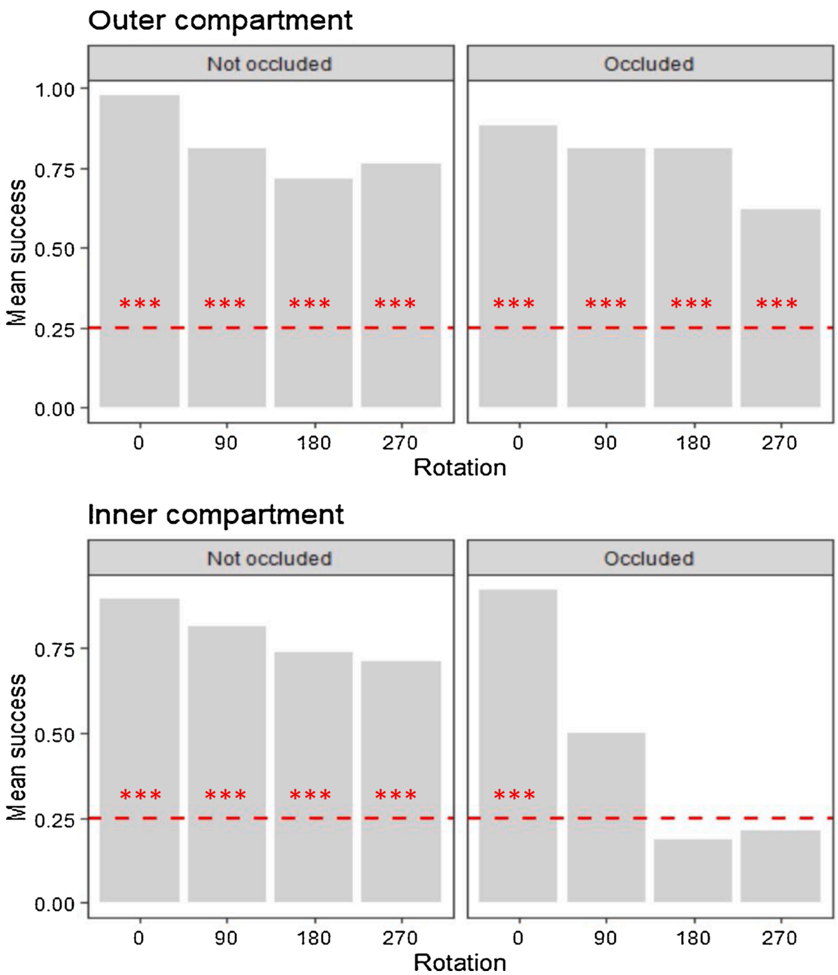

Fig. 2. Mean success rates in the Inner (n = 38) and Outer (n = 42) compartment conditions, split by occlusion and rotation degree.
Notes. The red horizontal line represents chance performance (25 %). The red asterisks indicate that the mean success rate was significantly different from the chance value .25 (Bonferroni-corrected α-level was .0031).

2.5. Coding and analysis

We scored for each child and trial whether or not the correct box was chosen (*Success*) as well as the total number of mistakes per child (0–8). To establish inter-rater reliability, a second coder blind to the hypotheses coded 25 % of the videos (i.e., $n = 20$). Inter-rater reliability was found to be perfect.

Analyses were carried out in R version 3.6.1 (R Core Team, 2019). We first checked for a possible difference in performance between boys and girls as some previous studies found gender differences in mental rotation tasks with children (Frick et al., 2014). We checked for differences between boys and girls in the number of mistakes made by running an unpaired two-samples Wilcoxon test (also known as Mann-Whitney test) using the R package *exactRankTests* (Hothorn & Hornik, 2017). A non-parametric test was used as the number of mistakes was not normally distributed within the gender groups. We used the same test to investigate if the number of mistakes differed between conditions. Effect sizes for the unpaired two-samples Wilcoxon tests were computed using the *wilcox_effsize* function from the *rstatix* package version 0.7.0 (Kassambara, 2021).

We also visualized the mean success rates for the Outer ($n = 42$) and Inner ($n = 38$) compartment conditions, split by occlusion, and compared performance against chance level. For this, we set the chance level to 25 %, which was a conservative estimate: despite there being eight boxes, most children picked a box within the same compartment in which the baited box was placed; therefore, we decided that chance performance would be to pick one of the four boxes within the relevant compartment. In order to determine whether children's performance differed significantly from chance, we performed two-sided one-sample Wilcoxon signed rank tests (i.e., the non-parametric alternative to a one-sample t-test) in R, comparing children's mean success rate against the value of .25. The α-level was Bonferroni-corrected to .05/16 = .0031. Effect sizes were again computed using the *wilcox_effsize* function from the *rstatix* package (Fig. 2; Table S2). Note that in Table S3 we also provide the mean success rates split by age groups (as well as condition and occlusion) and tests against chance level. As cell sizes within the age groups were very low, we did not further interpret the results from Table S3.

In order to examine children's performance with regard to compartment (Inner/Outer boxes), the occlusion of the grid, the degree of rotation as well as their age, we created a Generalized Linear Mixed Model (GLMM; Baayen, 2008) with binomial error structure and logit link function (McCullagh & Nelder, 1989) using the function *glmer* of the R-package *lme4* (Bates et al., 2015) with Success (yes/no) as Dependent Variable and age (z-transformed to a mean of 0 and a SD of 1), occlusion (yes/no), compartment (Inner/Outer), and rotation (as factor, 0, 90, 180, 270) as well as the three-way interaction between age, compartment and occlusion, and the two-way interactions between age and rotation, and occlusion and rotation as Independent Variables. Child ID was included as a random effect. To keep type I error rate at the nominal level of 5% (Barr, Levy, Scheepers, & Tily, 2013; Schielzeth & Forstmeier, 2009), we initially included random slopes for occlusion, rotation, and the interaction between occlusion and rotation on child ID, but did not include the correlation between the random effects. However, due to convergence issues (possibly because of the small sample size), we had to remove all random slopes from the model (for further details see Supplementary Results). The sample for this model consisted of 640 observations from 80 children. We decided to enter rotation as a factor as we assumed that the increases in rotation steps would not necessarily be qualitatively the same, i.e., a difference between 0° and 90° rotations could be qualitatively larger than a difference between 90° and 180° rotations. However, in response to the review process we also conducted the models with rotation entered as a numeric variable (see below and Supplementary Results).

Effect sizes for the entirety of the fixed effects and the entirety of the fixed and random effects were obtained using the function *r.squaredGLMM* of the package *MuMIn* (Barton, 2020). Model stability was assessed by comparing the estimates obtained from the model based on all data with those obtained from models with the levels of the random effects excluded one at a time (Nieuwenhuis, te Grotenhuis, & Pelzer, 2012). There were no issues with model stability (Tables S3 and S4). In order to determine the significance of the full model (Forstmeier & Schielzeth, 2011), we compared it against a null model only comprising the intercept and the random effects structure with a likelihood ratio test using the R function *anova* with the argument *test* set to "Chisq" (Dobson, 2002). To test the significance of the individual estimates we compared the full model against a reduced model not comprising the estimate using the same type of likelihood ratio test. Post-hoc multiple comparisons were conducted using the package *multcomp* (Hothorn, Bretz, & Westfall, 2008).The data and the analysis code are available at https://osf.io/2ncm3/?view_only=becdb69b1e92401f8d915610d4762ff9.

3. Results

As the total number of mistakes made by boys ($M = 2.54$, SD = 1.66, range 0–6) and girls ($M = 2.13$, SD = 1.75, range 0–5) did not differ significantly (unpaired two-samples Wilcoxon test, $U = 647$, $p = .317$, $r = .113$ (small effect size; 95 % CI [0.003; 0.33]), data from these two groups were collapsed for all further analyses. The number of mistakes was significantly larger in the Inner ($M = 3.05$, SD = 1.35, range 0–5) than in the Outer compartment condition ($M = 1.62$, SD = 1.74, range 0–6), unpaired two-samples Wilcoxon test, $U = 409.5$, $p < .001$, $r = .425$ (moderate effect size; 95 % CI [0.24; 0.61]).

The mean success rates against chance level split by condition, occlusion, and rotation degree are depicted in Fig. 2 (see also Table S2). Children performed above chance level in all sub-conditions, apart from the occluded trials in the Inner condition when rotations were conducted. However, it is ambiguous whether these non-significant findings represent true chance performance or are rather a result of low power. Indeed, for the Inner compartment condition, the one-sample Wilcoxon tests only had a power of 52 % to detect a moderate effect of $r = 0.50$. Therefore, a future study would need to replicate this investigation with a larger sample size.

Our full model, comprising the three-way interaction of age, occlusion and compartment, and the two-way interactions between age and rotation and occlusion and rotation, as well as a random intercept for child ID, explained the data significantly better than a null model only comprising an intercept (χ^2 (16) = 188.92, $p < .001$). Contrary to our hypothesis, the 3-way interaction between age, compartment and occlusion did not significantly contribute to the model fit (χ^2 (1) = 0.072, p = .789), so we reran the model without this interaction. We then tested which of the 2-way interactions contributed significantly to the model fit and found significant effects of the interaction between age and compartment (χ^2 (1) = 13.339, p < .001), occlusion and compartment (χ^2 (1) = 9.540, p = .002), but no effect of the interactions between age and occlusion (χ^2 (1) = 0.389, p = .553), age and rotation (χ^2 (3) = 3.803, p = .284) nor between occlusion and rotation (χ^2 (3) = 4.143, p = .246). To further investigate the significant interactions, we split the dataset by compartment (i.e., by the between-subject factor) into two subsets, and ran new GLMMs with success as dependent variable and age in months, occlusion, and rotation as independent variables (fixed effects) and a random effect for child ID.

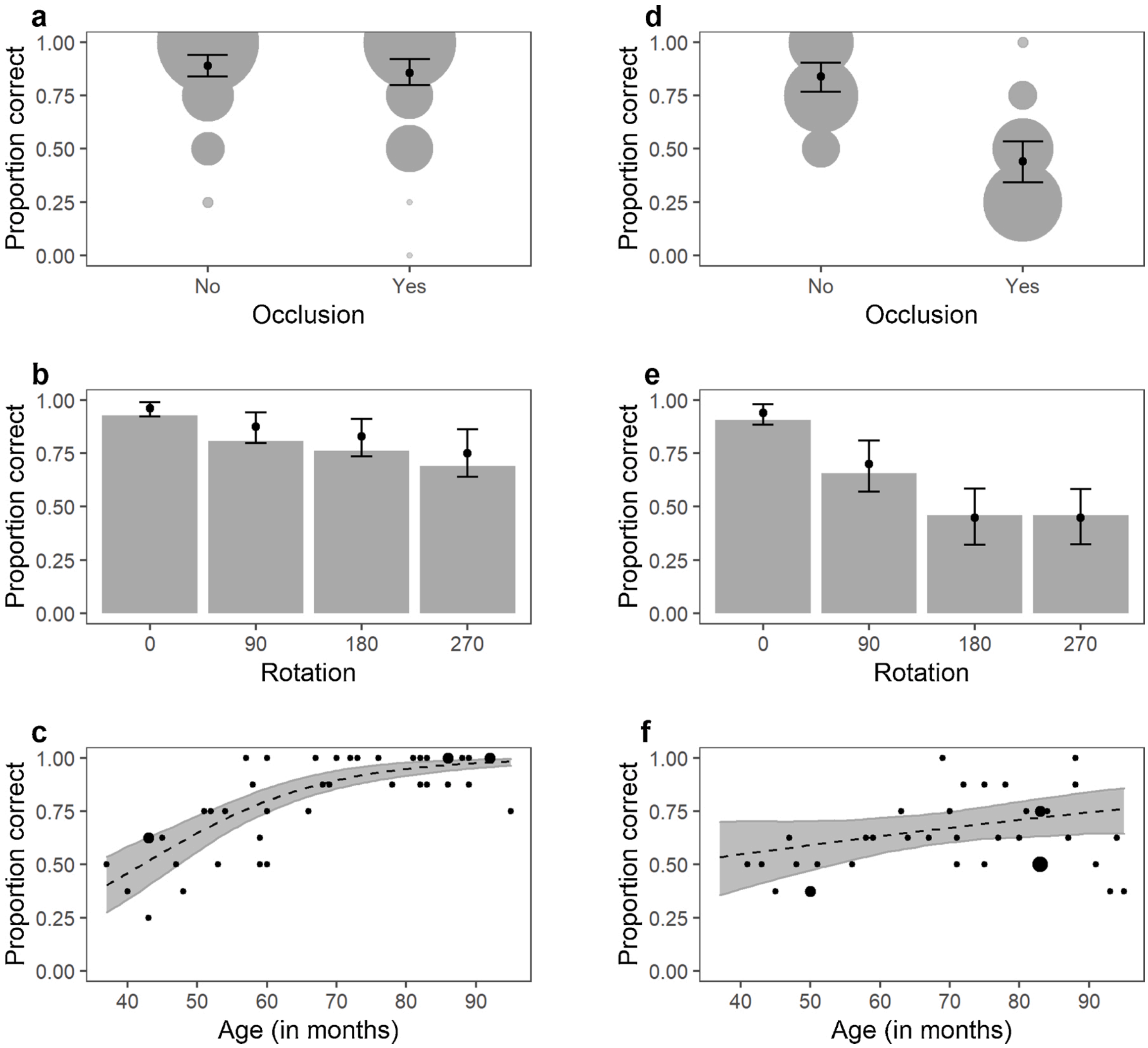

Fig. 3. Performance in the Outer (left column; a–c) and Inner (right column; d-f) compartment conditions.

Notes. Row 1 (a, d): Performance as a function of occlusion. The points show the model predictions (with all predictor variables centred except for occlusion), the error bars show the bootstrapped 95 % CI. The size of the points is proportional to the number of represented individuals (a – Outer compartment: ranging between 1 and 22; b – Inner compartment: ranging between 2 and 17). In the Outer compartment condition, performance did not differ significantly between the non-occluded and the occluded condition. In the Inner compartment condition, performance was significantly better in the non-occluded condition. Row 2 (b, d): Barplot showing the performance as a function of rotation. The points show the model prediction (with all predictor variables except for rotation centred) and the error bars the bootstrapped 95 % CI. Performance decreased significantly with increasing rotation in both conditions. Row 3 (c, f): Performance as a function of age. The dashed line shows the model predictions (with all predictor variables centred), the grey shaded area shows the bootstrapped 95 % CI. The points show the mean individual performance; the size of the points is proportional to the number of represented individuals (Outer compartment condition: ranging between 1 and 2; Inner compartment condition: ranging between 1 and 3). In the Outer compartment condition, performance improved significantly with age, but did not do so in the Inner compartment condition.

3.1. Outer compartment condition

The model explained the data significantly better than the null model (χ^2 (5) = 57.60, $p < .001$). The proportion of variance in the response explained by the entirety of the fixed effects was .27 and by the entirety of the fixed and random effects together was .28. We found a significant positive effect of age (χ^2 (1) = 35.321, $p < .001$; Fig. 3c) and a significant effect of rotation (χ^2 (3) = 20.974, $p < .001$; Fig. 3b). With each standard deviation increase in age (i.e., 16.8 months), there was a 267 % increase in the odds of succeeding in a trial. Posthoc multiple comparisons showed that trials with 180° rotation ($p = .010$) and trials with 270° rotation ($p < .001$) were significantly more difficult than no rotation trials. Compared to trials with no rotation, in trials with 180° rotation there was an 81 % decrease in the odds of being successful, and in trials with 270° rotation there was an 88 % decrease. No other comparisons were significant (but note that this might be due to insufficient power). There was no significant effect of occlusion (χ^2 (1) = 0.888, $p = .346$, Fig. 3a), indicating that occluded and non-occluded trials were of comparable difficulty. See Table S3 for the estimated coefficients.

3.2. Inner compartment condition

The model explained the data significantly better than the null model (χ^2 (5) = 99.34, $p < .001$). The proportion of variance in the response explained by the entirety of the fixed effects was .36 and by the entirety of the fixed and random effects together was .40. As in the Outer compartment condition, we found a significant effect of rotation (χ^2 (3) = 58.24, $p < .001$; Fig. 3e), with all rotation trials being significantly more difficult than no rotation trials (all $p < .001$) and 180° and 270° trials being significantly more difficult than 90° trials (both $p = .032$). Compared to trials with no rotation, in trials with 90° rotation there was an 85 % decrease in the odds of being successful, and in trials with 180° and 270° rotation there was a 95 % decrease. Compared to trials with a 90° rotation, in trials with 180° and 270° rotation there was an 65 % decrease in the odds of being successful. There was no difference between the 180° and 270° trials ($p = .100$). In contrast to the Outer compartment condition, there was a significant effect of occlusion (χ^2 (1) = 46.32, $p < .001$, Fig. 3d), with occluded trials being more difficult than non-occluded trials. Compared to non-occluded trials, in occluded trials there was a 86 % decrease in the odds of being successful. There was no effect of age (χ^2 (1) = 3.43, $p = .064$, Fig. 3f), indicating that for children across the age range the Inner trials were comparably difficult. See Table S4 for the estimated coefficients.

4. Discussion

The aim of this study was to develop a novel, non-verbal task measuring specific aspects of WM updating and manipulation in young children (3- to 7-year-olds). We designed the Rotating Grid Task, a task in which participants are asked to locate a reward after having observed the reward being hidden in a compartment within a box and the box being rotated (here by either 90°, 180° or 270°). The box is either not occluded during the rotation, allowing the participant to visually track and mentally update the location of the reward, or occluded, increasing the demand of mental manipulation in order to infer its new position, particularly when the reward is hidden in one of the Inner compartments.

As expected – given previous results that even infants can hold the location of a single object in mind for 10 s – we found that 3-year-olds performed near ceiling in the retrieval trial, in which they had to hold the location of a single reward in mind for 5 s. In terms of updating, our results mirror previous work by showing a developmental increase in performance in the updating trials, with children showing very good performance by 5 years of age (Boudreau et al., 2018; Garon et al., 2008; Hongwanishkul, Happaney, Lee, & Zelazo, 2005; Panesi & Morra, 2020). With regard to our trials designed to require mental manipulation (i.e., when the box was occluded), we showed that the easier it was to keep track of the location visually (i.e., when the reward was hidden in one of the outer boxes), the better children's performance tended to be. This suggests that tasks aiming to measure WM manipulation through invisible displacement (e.g., the double displacement task, the magic wand trials with rotation in Boudreau et al. (2018) and our manipulation trials) do not necessarily pose equal cognitive demands. For example, in our study, in our most difficult manipulation trials (Inner compartment condition, occluded, rotation trials) children's performance was at chance level; in contrast, in those "manipulation" trials in which performance could be aided by visual tracking and updating (i.e., Outer compartment condition, occluded) performance was better and increased with age. Thus, neither occlusion nor the Inner compartment condition alone were difficult and demanded mental manipulation, but it was the combination of occlusion and the Inner compartment condition that made the task especially challenging and unlikely to be solved through visual tracking. Similar to our Outer compartment non-occluded trials, tasks such as the double displacement task or the manipulation trials in Boudreau et al. (2018) likely also allow for visual tracking, thus scaffold children's performance and might not be as demanding on WM manipulation as previously thought. Our findings support the view that depending on the context, WM manipulation can consist of many different operations of varying difficulty. Here, we investigated one particularly challenging kind of manipulation.

In the following, we discuss our results in more detail with regard to our hypotheses. First, we found support for the hypothesis that performance generally decreases with increasing degree of rotation (regardless of the hiding condition), replicating many previous findings in adults and children (Frick, Ferrara et al., 2013; Frick, Hansen et al., 2013; Kaltner & Jansen, 2016; Shepard & Metzler, 1971). Second, we found no support for our hypothesis that there would be a three-way interaction between compartment (Inner/-Outer), occlusion, and age. Instead, there were interactions between compartment and occlusion as well as between compartment and age. That is, occlusion affected performance only in the Inner compartment condition; and an age effect was only found in the Outer compartment condition. As expected, the occluded trials of the Inner compartment condition were the most difficult trials and, unexpectedly, that was true for all age groups (3–7 years). In the Outer compartment condition, occluded trials were not more difficult than non-occluded trials. As mentioned above, this could have been due to the possibility that performance in the occluded trials of the Outer compartment condition could have been aided by visual tracking of the corner in which the baited compartment was located, thus not requiring the manipulation of a mental representation. This might have been further facilitated by the fact that the box was rectangular, and so the lengths of the sides of the box could have served as an additional cue for the location of the reward. Therefore, the occluded trials of the Outer compartment condition do not seem to measure manipulation (at least not with the box used in this study). Thus, future studies should use the "occluded inner boxes" trials as manipulation trials of this task. However, note that this sort of mental manipulation would then expected to be too difficult for children between 3 and 7 years and might be more appropriate for older age groups.

Our findings suggest that young children (from 4 years of age) are already proficient at solving those "manipulation" trials which can be aided by visual tracking. The occluded trials in the Inner compartment condition provide no such scaffolding for children and thus pose (higher) demands on WM manipulation, resulting in children in our study not performing above chance level in those trials. These results support previous evidence that manipulation of WM content has a protracted developmental trajectory (Gathercole et al., 2004; Luciana et al., 2005). Note, however, that as the current study was cross-sectional, it does itself not provide direct evidence for a protracted development and should therefore be replicated with a longitudinal design.

As much as children in our study were able to "keep an eye" on the baited box during the rotation in the non-occluded trials, children in Boudreau et al. (2018) just needed to track the location of the magic wand. Furthermore, our finding that the occluded version of the Outer compartment condition could probably also have been solved through visual tracking suggests that in Boudreau et al.'s (2018) additional trials in which the box was occluded and rotated, children might similarly have been able to use the edges of the box as visual cues. This suggestion fits with the relatively good performance that was achieved by 5-year-olds in the Boudreau et al. (2018) task.

In contrast, in the current study, 5-year-olds did not yet perform above chance levels in the occluded trials of the Inner compartment condition. We conclude that these trials pose a substantial demand on WM manipulation abilities which are not yet fully developed even in older children. These results add to previous findings from backward span tasks that also suggest that manipulation of WM content is a distinct skill set with protracted development (Carlson, 2005; Carlson et al., 2002; Davis & Pratt, 1995; Gathercole et al., 2004; Isaacs, 1989; Luciana et al., 2005; Simmering & Perone, 2013). Five-year-olds still perform rather poorly in backward span tasks (Carlson et al., 2002; Simmering & Perone, 2013). Nevertheless, older pre-schoolers still seem to perform better in the backward span tests than in the manipulation trials of the current study. Similarly, in contrast to the current study, age-related increases in performance between 4 and 8 years of age have been found in a tube rotation task in which children watched a tube rotate 180° and then had to draw, verbally describe, and reproduce the movement (Piaget & Inhelder, 1971). However, the tube rotation task might have posed slightly smaller memory demands (as there was no occlusion of objects) and also only involved a 180° rotation.

In addition, while there are performance increases within pre-school age in the backward digit span test (Gathercole et al., 2004), we found no such increase with age in the occluded trials of the Inner condition in the current study. However, before strong interpretations of this potential mismatch can be made, the current study should be replicated with a larger sample to investigate whether the lack of an increase in performance with age is a true null effect. Future studies could also carry out the task in different testing environments: it could be that the Science Museum environment in which we collected the data for this study was at times too distracting or noisy and resisting distraction from outside while engaging in our task might have been a greater demand for the younger children in our sample. However, if these results can be reproduced, this could potentially hint at a dissociation in WM manipulation abilities in young children based on task domain (spatial (the current study) vs verbal (backward digit span)). This would be in line with those suggesting that all WM and short-term memory tasks are separable to some degree due to their individual task demands, information content, required operations and the strategies they afford (Alloway, Gathercole, & Pickering, 2006; Doebel, 2020; Perone, Simmering, & Buss, 2021; Simmering & Perone, 2013). At the same time, there is strong evidence that performance in visuo-spatial and verbal WM tasks are dependent on the same general attentional resources (Barrouillet & Camos, 2010; Camos & Barrouillet, 2018; Swanson, 2017; Vergauwe, Barrouillet, & Camos, 2009).

In sum, we argue that verbal measures of WM manipulation are not well suited for very young children as they likely pose too heavy verbal demands. Visuo-spatial tasks seem to be better suited, however – as we have argued here – researchers should prevent participants from using visual tracking by removing visual anchors in order to ensure that their task poses sufficient demands on WM manipulation.

Our results also fit with the existing literature on the development of mental rotation in children. While 5-year-olds have been found to be able to solve most "classic" mental rotation tasks (see studies listed in Pedrett, Kaspar, & Frick, 2020), their mental rotation skills are – as our study also suggests – far from being fully developed (Dean & Harvey, 1979; Hawes et al., 2015). It seems that 5-year-olds' overall good performance in many mental rotation studies is often boosted by their good performance in trials with rotations below 90°, whereas their performance decreases in trials with rotations between 90° and 180°. For example, Frick, Hansen et al. (2013) presented 3- to 5-year-olds with rotated pairs of puzzle pieces, one of which would fit into a cut-out shape after rotating it upright while the other would not fit as it was a mirror image of the target shape. While 5-year-olds had an around 90 % accuracy rate for 90° rotations, they showed poorer performance for rotations between 90° and 180°, with the accuracy rate being at around 60 % for 180° rotations (note here also that these classic tasks do not require participants to mentally represent the images as they can see it all the time; therefore, they are not memory tasks unlike our task). Another study, comparing performance on a classic 2D and a novel 3D mental rotation test, found that 4- and 5-year-olds did not perform above chance in either task and that only 57 % of 7- and 8-year-olds were classified as "successful rotators" (i.e., children who scored the majority of the 3D test trials correct; Hawes et al., 2015).

Future work using the Rotating Grid Task should collect data from a larger sample and from a wider age range in order to investigate the developmental trajectory of performance in this task. In addition, in order to validate our new task, it is essential to explore how children's performance relates to their performance in other manipulation and updating tasks, such as backward span tests, The Magic House task (Panesi & Morra, 2020) or the Keep track task (van der Sluis et al., 2007). While we attempted to compare performance in the Rotating Grid Task with success rates in the Scrambled Box task, ceiling effects in the latter task prevented us from doing so (see Supplementary Material). Future studies could also employ eye tracking technology in order to investigate to what extent children really visually track the location of the hidden reward in the different conditions of the task. Finally, note that the sample in this study was very homogeneous, mainly consisting of children from Westernized, rich, and highly educated backgrounds. However, as the Rotating Grid Task poses only minimal verbal demands, it can be easily transferred to studies in different cultures, to gain a better understanding of potential cross-cultural differences in the development of WM. It can also be applied to atypical populations such as children with language impairments or to populations who suffer from cognitive decline such as patients with dementia. Lastly, the task can also be used with non-human animals, to address wider questions concerning the evolution of WM.

Children's ability to manipulate WM content has been found to be predictive of their academic success, especially in mathematics (Bull et al., 2008). Therefore, understanding the development of WM in childhood and its implications for the educational setting is crucial. Previous research has often focused on the developmental onset of WM and has highlighted that pre-school-aged children are already able to manipulate WM content (Boudreau et al., 2018; Garon et al., 2008). However, if this conclusion is taken at face value by researchers and practitioners in the educational context, there is a risk of overestimating the abilities of pre-school and even school-aged children. Our study – in line with closer scrutiny of previous research – supports the view that WM manipulation has a protracted development. Young children show still rather poor performance in backward span tests and more demanding mental rotation tasks, and the current study showed that even 7-year-olds find it difficult to truly manipulate mental representations if they have no concrete observable stimuli available to guide their understanding. This highlights the need for awareness of the cognitive limitations that children in the pre-school and early school years still face in the classroom as well as for the provision of tools and guidance to scaffold their learning.

Funding

This work was supported by a European Research Council (ERC) starting grant awarded to A.M.S. (INQMINDS, no. SEP-210159400). The funding source had no involvement in study design, collection, analysis, interpretation of the data, writing the paper, nor in the decision to submit the article for publication.

Author contributions

E.R.: conceptualization, methodology, formal analysis, investigation, data curation, writing original draft, visualization, supervision. D.P.: conceptualization, methodology, formal analysis, investigation, data curation, writing original draft, visualization. C.J.V.: conceptualization, writing original draft, visualization, supervision. A.M.S.: conceptualization, methodology, resources, writing original draft, supervision, project administration, funding acquisition.

Declaration of Competing Interest

Authors declare no competing interests.

Acknowledgments

General: We thank all children and parents for their participation and Dundee Science Centre for their support of our research. We thank Geraldine Brown for the help in coordinating the research and Abby Schwartz for reliability coding, and six anonymous reviewers for their very helpful comments and feedback on earlier versions of this manuscript.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.cogdev.2021.101068.

References

Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and visuospatial short-term and working memory in children: Are they separable? *Child Development, 77*(6), 1698–1716. https://doi.org/10.1111/j.1467-8624.2006.00968.x

Baayen, R. H. (2008). *Analyzing linguistic data. A practical introduction to statistics using R*. New York: Cambridge University Press.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255–278.

Barrouillet, P., & Camos, V. (2010). Working memory and executive control: A time-based resource-sharing account. *Psychologica Belgica, 50*(3&4), 353–382.

Barton, K. (2020). *MuMIn: Multi-model inference*. Retrieved from https://cran.r-project.org/package=MuMIn.

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

Bayliss, D. M., Jarrold, C., Gunn, D. M., & Baddeley, A. D. (2003). The complexities of complex span: Explaining individual differences in working memory in children and adults. *Journal of Experimental Psychology: General, 132*(1), 71–92. https://doi.org/10.1037/0096-3445.132.1.71

Berch, D. B., Krikorian, R., & Huha, E. M. (1998). The Corsi block-tapping task: Methodological and theoretical considerations. *Brain and Cognition, 38*, 317–338. Retrieved from http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed4&NEWS=N&AN=1999000559.

Boettcher, S. E. P., Draschkow, D., Dienhart, E., & Võ, M. L. H. (2018). Anchoring visual search in scenes: Assessing the role of anchor objects on eye movements during visual search. *Journal of Vision, 18*(13), 1–13. https://doi.org/10.1167/18.13.11

Boudreau, A. M., Dempsey, E. E., Smith, I. M., & Garon, N. (2018). A novel working memory task for preschoolers: Sensitivity to age differences from 3-5 years. *Child Neuropsychology, 24*(6), 799–822. https://doi.org/10.1080/09297049.2017.1333592

Bruyer, R., & Scailquin, J.-C. (1998). The visuospatial sketchpad for mental images: Testing the multicomponent model of working memory. *Acta Psychologica, 98*, 17–36. https://doi.org/10.1016/S0001-6918(97)00053-X

Bull, R., Espy, K. A., & Wiebe, S. A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. *Developmental Neuropsychology, 33*(3), 205–228. https://doi.org/10.1080/87565640801982312

Camos, V., & Barrouillet, P. (2018). *Working memory in development.* London: Routledge.

Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology, 28*(2), 595–616. https://doi.org/10.1207/s15326942dn2802_3

Carlson, S. M., Moses, L. J., & Breton, C. (2002). How specific is the relation between executive function and theory of mind? Contributions of inhibitory control and working memory. *Infant and Child Development, 11*, 73–92. https://doi.org/10.1002/icd

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12*(5), 769–786.

Cowan, N. (2014). Short-term and working memory in childhood. In P. J. Bauer, & R. Fivush (Eds.), *The Wiley handbook on the development of children's memory* (pp. 202–229). Chichester: Wiley-Blackwell.

Cowan, N. (2016). Working memory maturation: Can we get at the essence of cognitive growth? *Perspectives on Psychological Science, 11*(2), 239–264. https://doi.org/10.1177/1745691615621279

Cowan, N. (2017). The many faces of working memory and short-term storage. *Psychonomic Bulletin and Review, 24*(4), 1158–1170. https://doi.org/10.3758/s13423-016-1191-6

Davis, H. L., & Pratt, C. (1995). The development of children's theory of mind: The working memory explanation. *Australian Journal of Psychology, 47*(1), 25–31. https://doi.org/10.1080/00049539508258765

Dean, A. L., & Harvey, W. O. (1979). An information-processing analysis of a Piagetian imagery task. *Developmental Psychology, 15*(4), 474–476. https://doi.org/10.1037/0012-1649.15.4.474

Diamond, A. (1997). Prefrontal cortex cognitive deficits in children treated early and continuously for PKU. *Monographs of the Society for Research in Child Development, 62*, 1–206.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*(1), 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Diamond, A., & Doar, B. (1989). The performance of human infants on a measure of frontal cortex function, the delayed response task. *Developmental Psychobiology, 22*(3), 271–294. https://doi.org/10.1002/dev.420220307

Dobson, A. J. (2002). *An introduction to generalized linear models.* Boca Raton: Chapman & Hall/CRC press.

Doebel, S. (2020). Rethinking executive function and its development. *Perspectives on Psychological Science, 15*(4), 942–956. https://doi.org/10.1177/1745691620904771

Ecker, U. K. H., Lewandowsky, S., Oberauer, K., & Chee, A. E. H. (2010). The components of working memory updating: An experimental decomposition and individual differences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(1), 170–189. https://doi.org/10.1037/a0017891

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*(1), 19–23.

Espy, K. A., McDiarmid, M. D., Cwik, M. F., Stalets, M. M., Hamby, A., & Senn, T. E. (2004). The contribution of executive functions to emergent mathematic skills in preschool children. *Developmental Neuropsychology, 26*(1), 465–486. https://doi.org/10.1207/s15326942dn2601

Estes, D. (1998). Young children's awareness of their mental activity: The case of mental rotation. *Child Development, 69*(5), 1345–1360. https://doi.org/10.2307/1132270

Evrard, D., Charollais, A., Marret, S., Radi, S., Rezrazi, A., & Mellier, D. (2011). Cognitive and emotional regulation developmental issues in preterm infants 12 and 24 months after birth. *European Journal of Developmental Psychology, 8*(2), 171–184. https://doi.org/10.1080/17405620903504538

Forstmeier, W., & Schielzeth, H. (2011). Cryptic multiple hypotheses testing in linear models: Overestimated effect sizes and the winner's curse. *Behavioral Ecology and Sociobiology, 65*(1), 47–55. https://doi.org/10.1007/s00265-010-1038-5

Frick, A., Möhring, W., & Newcombe, N. S. (2014). Development of mental transformation abilities. *Trends in Cognitive Sciences, 18*(10), 536–542. https://doi.org/10.1016/j.tics.2014.05.011

Frick, A., Ferrara, K., & Newcombe, N. S. (2013). Using a touch screen paradigm to assess the development of mental rotation between 3 and 5 years of age. *Cognitive Processing, 14*, 117–127. https://doi.org/10.1007/s10339-012-0534-0

Frick, A., Hansen, M. A., & Newcombe, N. S. (2013). Development of mental rotation in 3- to 5-year-old children. *Cognitive Development, 28*(4), 386–399. https://doi.org/10.1016/j.cogdev.2013.06.002

Frischkorn, G. T., von Bastian, C. C., Souza, A. S., & Oberauer, K. (2020). Individual differences in updating are not related to reasoning ability and working memory capacity. *PsyArXiv*. https://doi.org/10.31234/osf.io/92uhb

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin, 134*(1), 31–60. https://doi.org/10.1037/0033-2909.134.1.31

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology, 40*(2), 177–190. https://doi.org/10.1037/0012-1649.40.2.177

Hawes, Z., Lefevre, J.-A., Xu, C., & Bruce, C. D. (2015). Mental rotation with tangible three-dimensional objects: A new measure sensitive to developmental differences in 4- to 8-year-old children. *Mind, Brain, and Education, 9*(1), 10–18. https://doi.org/10.1111/mbe.12051

Hongwanishkul, D., Happaney, K. R., Lee, W. S. C., & Zelazo, P. D. (2005). Assessment of hot and cool executive function in young children: Age-related changes and individual differences. *Developmental Neuropsychology, 28*(2), 617–644. https://doi.org/10.1207/s15326942dn2802

Hothorn, T., & Hornik, K. (2017). *exactRankTests: Exact distributions for rank and permutation tests (R package version 0.8-29)*.

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal, 50*(3), 346–363. https://doi.org/10.1002/bimj.200810425

Hughes, C. (1998). Finding your marbles: Does preschoolers' strategic behavior predict later understanding of mind? *Developmental Psychology, 34*(6), 1326–1339.

Hughes, C. (2011). Changes and challenges in 20 years of research into the development of executive functions. *Infant and Child Development, 20*, 251–271. https://doi.org/10.1002/icd

Hughes, C., & Ensor, R. (2007). Executive function and theory of mind: Predictive relations from ages 2 to 4. *Developmental Psychology, 43*(6), 1447–1459. https://doi.org/10.1037/0012-1649.43.6.1447

Hyun, J.-S., & Luck, S. J. (2007). Visual working memory as the substrate for mental rotation. *Psychonomic Bulletin & Review, 14*(1), 154–158. https://doi.org/10.3758/BF03194043

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

Jenkins, I. L., & Berthier, N. E. (2014). Working memory and inhibitory control in visually guided manual search in toddlers. *Developmental Psychobiology, 56*, 1252–1262. https://doi.org/10.1002/dev.21205

Kaltner, S., & Jansen, P. (2016). Developmental changes in mental rotation: A dissociation between object-based and egocentric transformations. *Advances in Cognitive Psychology, 12*(2), 67–78. https://doi.org/10.5709/acp-0218-6

Kane, M. J., Tuholski, S. W., Hambrick, D. Z., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General, 133*(2), 189–217. https://doi.org/10.1037/0096-3445.133.2.189

Kassambara, A. (2021). *Rstatix: Pipe-friendly framework for basic statistical tests*.

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology, 55*(4), 352–358.

Kosslyn, S. M., Margolis, J. A., Barrett, A. M., Goldknopf, E. J., & Daly, P. F. (1990). Age differences in imagery abilities. *Child Development, 61*(4), 995–1010. https://doi.org/10.2307/1130871

Krüger, M., Kaiser, M., Mahler, K., Bartels, W., & Krist, H. (2014). Analogue mental transformations in 3-year-olds: Introducing a new mental rotation paradigm suitable for young children. *Infant and Child Development, 23*, 123–138. https://doi.org/10.1002/icd

Lee, K., & Bull, R. (2015). Developmental changes in working memory, updating, and math achievement. *Journal of Educational Psychology, 108*(6), 869–882. https://doi.org/10.1037/edu0000090

Lehmann, J., Quaiser-Pohl, C., & Jansen, P. (2014). Correlation of motor skill, mental rotation, and working memory in 3- to 6-year-old children. *European Journal of Developmental Psychology, 11*(5), 560–573. https://doi.org/10.1080/17405629.2014.888995

Luciana, M., Conklin, H. M., Hooper, C. J., & Yarger, R. S. (2005). The development of nonverbal working memory and executive control processes in adolescents. *Child Development, 76*(3), 697–712. https://doi.org/10.1111/j.1467-8624.2005.00872.x

Marmor, G. S. (1975). Development of kinetic images: When does the child first represent movement in mental images? *Cognitive Psychology, 7*, 548–559. https://doi.org/10.1016/0010-0285(75)90022-5

Marmor, G. S. (1977). Mental rotation and number conservation: Are they related? *Developmental Psychology, 13*(4), 320–325. https://doi.org/10.1037/0012-1649.13.4.320

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd edition). Washington, DC: Chapman and Hall/CRC.

Milner, B. (1971). Interhemispheric differences in the localization of psychological processes in man. *British Medical Bulletin, 27*(3), 272–277. https://doi.org/10.1093/oxfordjournals.bmb.a070866

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology, 41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

Morra, S., Panesi, S., Traverso, L., & Usai, M. C. (2018). Which tasks measure what? Reflections on executive function development and a commentary on Podjarny, Kamawar, and Andrews (2017). *Journal of Experimental Child Psychology, 167*, 246–258. https://doi.org/10.1016/j.jecp.2017.11.004

Morris, N., & Jones, D. M. (1990). Memory updating in working memory: The role of the central executive. *British Journal of Psychology, 81*(2), 111–121. https://doi.org/10.1111/j.2044-8295.1990.tb02349.x

Nieuwenhuis, R., te Grotenhuis, M., & Pelzer, B. (2012). influence.ME: Tools for detecting influential data in mixed effects models. *The R Journal, 4*(2), 38–47.

Noda, M. (2010). Manipulative strategies prepare for mental rotation in young children. *European Journal of Developmental Psychology, 7*(6), 746–762. https://doi.org/10.1080/17405620903465771

Panesi, S., & Morra, S. (2020). Executive functions and mental attentional capacity in preschoolers. *Journal of Cognition and Development, 21*(1), 72–91. https://doi.org/10.1080/15248372.2019.1685525

Pedrett, S., Kaspar, L., & Frick, A. (2020). Understanding of object rotation between two and three years of age. *Developmental Psychology, 56*(2), 261–274. https://doi.org/10.1037/dev0000871

Perone, S., Simmering, V. R., & Buss, A. T. (2021). A dynamical reconceptualization of executive function development. *Perspectives on Psychological Science*, 1–33. https://doi.org/10.1177/1745691620966792

Piaget, J., & Inhelder, B. (1971). *Mental imagery in the child. A study of the development of imaginal representation.* London: Routledge.

R Core Team. (2019). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Schielzeth, H., & Forstmeier, W. (2009). Conclusions beyond support: Overconfident estimates in mixed models. *Behavioral Ecology, 20*(2), 416–420. https://doi.org/10.1093/beheco/arn145

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171*(3972), 701–703. https://doi.org/10.1126/science.171.3972.701

Simmering, V. R., & Perone, S. (2013). Working memory capacity as a dynamic process. *Frontiers in Psychology, 3*, 1–26. https://doi.org/10.3389/fpsyg.2012.00567

St Clair-Thompson, H. L., & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *The Quarterly Journal of Experimental Psychology, 59*(4), 745–759. https://doi.org/10.1080/17470210500162854

Swanson, H. L. (2017). Verbal and visual-spatial working memory: What develops over a life span? *Developmental Psychology, 53*(5), 971–995. https://doi.org/10.1037/dev0000291

Traverso, L., Viterbori, P., & Usai, M. C. (2015). Improving executive function in childhood: Evaluation of a training intervention for 5-year-old children. *Frontiers in Psychology, 6*(525), 1–14. https://doi.org/10.3389/fpsyg.2015.00525

van der Sluis, S., de Jong, P. F., & van der Leij, A. (2007). Executive functioning in children, and its relations with reasoning, reading, and arithmetic. *Intelligence, 35*, 427–449. https://doi.org/10.1016/j.intell.2006.09.001

Ven, S. H. G. Van Der, Kroesbergen, E. H., Boom, J., & Leseman, P. P. M. (2012). The development of executive functions and early mathematics: A dynamic relationship. *British Journal of Developmental Psychology, 82*, 100–119. https://doi.org/10.1111/j.2044-8279.2011.02035.x

Vergauwe, E., Barrouillet, P., & Camos, V. (2009). Visual and apatial working memory are not that dissociated after all: A time-based resource-sharing account. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(4), 1012–1028. https://doi.org/10.1037/a0015859

Wiebe, S. A., Espy, K. A., & Charak, D. (2008). Using confirmatory factor analysis to understand executive control in preschool children: I. Latent structure. *Developmental Psychology, 44*(2), 575–587. https://doi.org/10.1037/0012-1649.44.2.575

Wiebe, S. A., Lukowski, A. F., & Bauer, P. J. (2010). Sequence imitation and reaching measures of executive control: A longitudinal examination in the second year of life. *Developmental Neuropsychology, 35*(5), 522–538. https://doi.org/10.1080/875656412010494751