Notice: This is the author's version of a work that was accepted for publication in *Developmental Psychology*. A definitive version was subsequently published in *Developmental Psychology*, *51*, 1704-1716. DOI: 10.1037/dev0000050

The Magic Shrinking Machine revisited: The presence of props at recall facilitates memory in 3-year-olds

Jonna J. Dahl

Osman S. Kingo

Peter Krøjgaard

Jonna J. Dahl, Pedagogical Psychological Counseling, Municipality of Hedensted, Juelsminde Rådhus,Denmark

Osman S. Kingo, Center on Autobiographical Memory Research, Department of Psychology and Behavioral Sciences, Aarhus University

Peter Krøjgaard, Center on Autobiographical Memory Research, Department of Psychology and Behavioral Sciences, Bartholins Allé 9, Aarhus University, DK-8000 Aarhus C. Mail: peter@psy.au.dk, Phone: 0045 87 16 58 61.

Corresponding Author: **Peter Krøjgaard**

Acknowledgement

This research was supported by the Danish National Research Foundation (Grant DNRF93) and MIND*Lab*. First of all we would like to thank Gabrielle Simcock and Harlene Hayne for providing us with detailed information about the Magic Shrinking Machine in order for us to replicate the device. Second, we thank Patricia Bauer and Kim Plunkett for constructive discussions of the present study. Third, we would like to thank Line R. Kristiansen and Trine Sonne for their help as assistants, and Line R. Kristiansen and Bettina Hald for scoring the data. Finally, we thank the children and parents who participated in this study.

Notice: This is the author's version of a work that was accepted for publication in *Developmental Psychology.* A definitive version was subsequently published in *Developmental Psychology*, *51*, 1704-1716. DOI: 10.1037/dev0000050

Abstract

In a seminal study Simcock and Hayne (2002) showed that 3-year-olds were unable to use newly acquired words to describe a "magic" event experienced six or twelve months earlier. In the reference study the children's verbal recall was tested *without* props being present. Inspired by recent evidence, the original design was replicated, testing 33-and 39-month-olds (n = 180), but *with* props present at recall while controlling for potential on-line reasoning. The results revealed that the children did use newly acquired words to describe their preverbal memory. Thus, the present study shows that nonverbal memories can be verbalized if the recall setting provides high level of contextual support, a finding relevant to researchers investigating the offset of childhood amnesia.

Notice: This is the author's version of a work that was accepted for publication in *Developmental Psychology*. A definitive version was subsequently published in *Developmental Psychology, 51*, 1704-1716. DOI: 10.1037/dev0000050

The Magic Shrinking Machine Revisited: The Presence of Props at Recall Facilitates Memory in 3-year-olds

Introduction

Early life experiences are by many developmental psychologists assumed to have a crucial impact on a person's life (Bowlby, 1951; Rutter, 2002). Meanwhile, most adults remember only few, if any, specific events from the first three years of their lives (e.g., Pillemer & White, 1989). The paucity of memories from early in life was described by Freud (1905/1953), who coined this "black hole" in our autobiographical memory *childhood amnesia*.[1] Thus, researchers studying childhood amnesia seem to be faced with a paradox: If early experiences are vital and affect our lives in general, then why is it that we as adults have severe difficulties recalling these events (Rovee-Collier & Cuevas, 2009)?

When older children and adults report memories from specific events or episodes, the report is by default *verbal*. However, events experienced in infancy and young childhood are encoded and stored at a time when the children's language abilities are under development. Thus, when attempting to understand childhood amnesia as it comes into play in childhood, the following two questions seem important: First, to what extent are young children capable of making verbal reports of events experienced when their language abilities were less mature? Second, as language matures, can newly acquired productive language be used to describe these early event memories? The second question is not only crucial regarding our understanding of childhood amnesia, but has also important practical relevance – for instance when children appear as witnesses in courtrooms (Goodman, Batterman-Faunce, Schaff, & Kenny, 2002). In this paper we focus especially on the second question. We begin by presenting a brief outline of the literature concerning young children's verbally produced memories of events that occurred when the children's productive language abilities were less mature, because this research serves as the foundation on which the studies on the second question are based.

Young children's ability to provide verbal reports of early life experiences

Until the beginning of the 1980's research on children's ability to verbally recall past events was sparse (Bauer, 2007) and evidence suggesting that children were able to verbally

[1] Some scholars define autobiographical memories as memories for personally experienced specific events having personal significance (e.g., Nelson, 1993). Following Berntsen and Rubin (2012) we prefer a broader definition concerning memory for experienced events, which in principle allows for considering autobiographical memories in both young children as well as in other species.

Notice: This is the author's version of a work that was accepted for publication in *Developmental Psychology*. A definitive version was subsequently published in *Developmental Psychology*, *51*, 1704-1716. DOI: 10.1037/dev0000050

recall events, experienced at a time when their language abilities were less mature, were mainly anecdotal (e.g., Nelson & Ross, 1980; Todd & Perlmutter, 1980). Since then a number of studies have shown that children around two years of age and older have verbal memories for naturally occurring events (e.g., Cleveland & Reese, 2008; Fivush, Gray, & Fromhoff, 1987; Peterson & Rideout, 1998) which in some instances are retained over several years (e.g., Peterson & Parsons, 2005; Peterson & Whalen, 2001). Meanwhile, the results from a number of other naturalistic studies indicate that once a long enough delay has ensued, events that were formerly retrievable may no longer be accessible to verbal recall (e.g., 29-35 month-olds: Fivush et al., 1987; 3-6 year olds: Goodman, Hirschman, Hepps, & Rudy, 1991; 3.5-4.5 year olds: Pillemer, Picariello, & Pruett, 1994).

In a different line of research a number of studies have investigated children's ability to talk about previously experienced events under more controlled conditions. Some lab-based studies have shown little evidence of verbal recall of memories encoded at a time when the children's language abilities were less mature (e.g., Boyer, Barron, & Farrar, 1994; Myers, Clifton, & Clarkson, 1987; Myers, Perris, & Speaker, 1994, study 3). However, these studies all share the same design characteristic that the children were asked to verbally recall the target event *without* the target props from the event being visibly present at recall.

In another series of lab-based studies Bauer and her colleagues have shown that the presence of the to-be-remembered props at recall may facilitate the children's ability to talk about their previous experiences from participating in elicited imitation studies (e.g., Bauer, Kroupina, Schwade, Dropnik, & Wewerka, 1998). In order to examine whether the children were able to verbally recall memories encoded at a time when the children's language skills were less developed, the children's language skills were assessed employing the MacArther Bates Communicative Developmental Inventory (CDI) both at the time of encoding as well as at the time of the verbal recall test. The results revealed that compared to naïve controls of equal age, children 20 months of age or older were able to verbally describe previously experienced events – even after substantial delays (e.g., Bauer et al. 1998; Bauer, Van Abbema, Wiebe, Cary, Phill, & Burch, 2004). However, successful verbal retrieval turned out to be highly dependent on the degree of contextual support: Whereas the 20-month-olds succeeded when the props were present at recall, they failed if the props were absent or if only pictures of the props were present (Bauer et al., 2004). The findings from Bauer and her colleagues thus indicate that memories encoded at a time when the productive vocabulary is less advanced, can be accessible to later verbal report, if the children's recall is supported by

Notice: This is the author's version of a work that was accepted for publication in *Developmental Psychology*. A definitive version was subsequently published in *Developmental Psychology, 51*, 1704-1716. DOI: 10.1037/dev0000050

a high degree of contextual support (e.g., having props present during the verbal recall test). However, although Bauer and colleagues assessed the children's productive vocabulary at both encoding and test, they did not explicitly investigate whether the children at the recall test used *newly* acquired words to describe the events.

Can children use newly acquired words to describe earlier memories?

The first study to directly address this question was Simcock and Hayne (2002). In an elegant design they taught 27-, 33- and 39-month-olds how to operate a Magic Shrinking Machine, a 'machine' which (apparently) could shrink toy objects into smaller, but identical looking replicas. Like Bauer and her collaborators, Simcock and Hayne (2002) assessed the children's language abilities at the time of encoding and at the time of test. However, in addition, Simcock and Hayne (2002) asked the parents to fill out an extra list of words specifically associated with the target event which allowed them to *directly* evaluate whether the children had acquired the necessary productive vocabulary to describe the event across the two delays *and* to assess if they used any of these new words when verbalizing their memories. The children's verbal and nonverbal recall was assessed six (or twelve) months later. The results showed that although the children nonverbally recalled the event, indexed by counting the number of correctly identified target photographs and the number of target actions re-enacted respectively, there was not a single instance in which a child used newly acquired words to describe the event (Simcock & Hayne, 2002). As stated by Simcock and Hayne (2002, p. 229): "…children's verbal reports of the event were frozen in time, reflecting their verbal skill at the time of encoding, rather than at the time of test." In contrast, in a follow-up study six years later, 46 children were re-interviewed about the Magic Shrinking Machine event (Jack, Simcock, & Hayne, 2012). The results revealed that nine (20%) of the children were able to recall the event verbally, and two of the children did use newly acquired words to describe the event. However, the children were now 8-10 years of age, and thus were not young children as the ones tested by Simcock and Hayne (2002).

On the face of it, the results obtained by Simcock and Hayne (2002) seem at odds with the results cited earlier from Bauer and colleagues. Although Bauer and colleagues did not directly assess whether the children used newly acquired words to describe the previously experienced event, the children in their study were able to talk about the event (Bauer et al., 2004) whereas the verbal recall task was found to be quite difficult for the children in the Simcock and Hayne (2002) study. However, there was at least one important design

Notice: This is the author's version of a work that was accepted for publication in *Developmental Psychology.* A definitive version was subsequently published in *Developmental Psychology, 51,* 1704-1716. DOI: 10.1037/dev0000050

difference between the two studies: Whereas the to-be-remembered props were *absent* at recall in the study by Simcock and Hayne (2002) where the children failed to use newly acquired words to describe the previously experienced event, the props were *present* in the study by Bauer et al. (2004) where the children succeeded in verbally recalling the to-be-remembered event. As suggested by Bauer et al. (2004) as well as Morris and Baker-Ward (2007), young children may need strong contextual support at recall in order to verbalize memories encoded at a time with limited language skills, and the absence of props at recall may have affected the findings reported by Simcock and Hayne (2002).

In a more recent lab study Morris and Baker-Ward (2007) investigated two-year-olds verbal recall of a unique event. In this study the children's verbal recall was facilitated by a high degree of contextual support during the verbal recall test by reinstating the encoding context including a to-be-remembered "Magic Bubble Machine". Like Simcock and Hayne (2002), the aim of Morris and Baker-Ward's (2007) study was to examine whether young children could use newly acquired words to describe the event. The Magic Bubble Machine could only be activated by each child's target color. The children were randomized into two conditions: those who already knew the label for their target color (known condition) and those who did not (unknown condition). During a two-month retention interval all children interacted with the experimenter in color-label and color-learning games. The results revealed that, although fragile, a significant portion of the children in both conditions who had acquired new verbal color labels were able to use these new words to describe critical components of the event. Thus, these findings indicate that some early life memories can be translated into language across time (Morris & Baker-Ward, 2007).

Summary

A substantial number of studies have investigated young children's ability to verbally recall events experienced at a time when their language abilities were less mature. However, to date only two studies (Morris & Baker-Ward, 2007; Simcock & Hayne, 2002) have examined young children's ability to use newly acquired words from their productive vocabulary to describe previously experienced events under controlled conditions. Given the fact that this question is highly relevant when examining the childhood amnesia enigma, it may seem surprising that only two studies have been conducted. Further, the results from these two studies are inconclusive. Whereas the children in Morris and Baker-Ward's (2007) experiment did use newly acquired words, the children in the study by Simcock and Hayne

Notice: This is the author's version of a work that was accepted for publication in *Developmental Psychology*. A definitive version was subsequently published in *Developmental Psychology, 51*, 1704-1716. DOI: 10.1037/dev0000050

(2002) did not. This divergence in results may be due to differences in the designs. In Simcock and Hayne's (2002) study the children's verbal recall was only supported by the experimenter's verbal prompts. In contrast, in the study by Morris and Baker-Ward (2007) the to-be-remembered props were present at recall. For children younger than three years of age, verbal prompts alone may not be effective retrieval cues (Hudson & Fivush, 1991; Hudson, 1993) and reinstating high degrees of contextual support tend to have a greater effect on young children's recall than more abstract cues such as verbal labels (Pipe, Gee, & Wilson, 1993). The previously cited evidence from Bauer's lab (Bauer et al., 2004) suggests that the presence of props at recall can be crucial. This reasoning follows the assumption that a careful match of distinct cues between the original to-be-remembered event and the context for retrieval is known to facilitate recall (Hayne, 2004; Newcombe, Lloyd, & Ratliff, 2007).

Whether the presence of props at recall would have made a difference in the Simcock and Hayne (2002) study can, however, not be resolved by speculation but requires replication. Based on the argument that the presence of props at recall "makes it impossible to differentiate between children's verbal recall of a preverbal memory and their on-line descriptions of objects or actions at the time of the test" (Simcock & Hayne, 2002, pp. 225-226), Simcock and Hayne (2002) deliberately chose to conduct the verbal recall test *without* the props being present. However, such a potential confound may be controlled for (Bauer et al., 2004). Here we replicate the Simcock and Hayne (2002) study, but this time with the props present at recall while controlling for the above mentioned potential confound. The basic idea was to use *two* equivalent but different 'magic' machines, a close replica of the Magic Shrinking Machine (MSM), and a novel device, the Crazy Duplicator (CDC). Half of the children were taught how to run the MSM, whereas the other half was shown how to operate the CDC. At recall both machines were visible, but out of reach. Thus, some children had the MSM as their target device at encoding whereas the CDC was a novel device at the delayed recall test and thus acted as a foil machine for this group of children. In contrast, for children having the CDC as their target machine at encoding, the MSM was novel at the delayed recall test and acted as their foil machine. Our reasoning was that if the children at the recall session were equally good at providing verbal reports of how to run the two machines, then their verbalizations could be attributed to on-line reasoning. If, on the other hand, the children's verbal performance was superior concerning the machine they had been taught how to operate (the target machine) relative to the new machine (the foil), then their verbalizations could be ascribed to memory. Based on the previously mentioned studies as

Notice: This is the author's version of a work that was accepted for publication in *Developmental Psychology.* A definitive version was subsequently published in *Developmental Psychology, 51,* 1704-1716. DOI: 10.1037/dev0000050

well as the generally accepted assumption that a careful match of distinct cues between the to-be-remembered event and the context for retrieval is known to facilitate recall (Hayne, 2004; Newcombe et al., 2007), we hypothesized that the children in the present study would be able to verbally recall the event and would use newly acquired words from their productive vocabulary to describe their target event.

The experiment

In the present study we replicated Simcock and Hayne's (2002) basic design but with the following changes: First, in order to allow for the props to be present at recall while controlling for possible online reasoning, we employed a mixed 2 (*Between*-subjects: MSM vs. CDC) x 2 (*Within*-subjects: MSM vs. CDC) design, where each child was taught *either* the MSM *or* the CDC at the time of encoding (T_1) as well as at test of encoding (T_2). However, at the delayed recall test (T_3) the child was tested on *both* machines (see Fig. 1).

Second, for methodological and pragmatic reasons all children were tested in the same laboratory on all visits and not in their homes as in the reference study. Our reasoning was that in the lab we were able to control for a number of potential confounding factors and saved time not visiting the children in their homes. Third, since the verbal recall task in the original experiment was very difficult for all three age groups (27-, 33- and 39-month-olds), we decided only to test the two oldest groups of children, that is, the 33-month-olds and the 39-month-olds. Fourth, whereas Simcock and Hayne (2002) employed two different retention intervals (six and twelve months), we only tested the children after a six month delay. Our reasoning was that the basic hypothesis we tested – that the presence of props may make a difference – would be testable regardless of whether we used one or two retention intervals. Fifth, contrary to Simcock and Hayne (2002) we employed a baseline measure at encoding (T_1) in order to assess whether the children *without* instructions would produce any of the target actions for the given target machine. This baseline measure allowed us to assess the impact of the instructions provided by the experimenter. Sixth, in order to assess the children's language skills Simcock and Hayne (2002) used the PPVT and the EVT, whereas in their sequel MSM study (Simcock and Hayne, 2003) they used the CDI to assess productive language skills in children 24-48 months of age. The CDI is a validated and reliable instrument when assessing young children's language skills (Fenson, Dale, Reznick, Bates, Thal, & Pethick, 1994; Fenson, Bates, Dale, Goodman, Reznick, & Thal, 2000) and

Notice: This is the author's version of a work that was accepted for publication in *Developmental Psychology.* A definitive version was subsequently published in *Developmental Psychology, 51,* 1704-1716. DOI: 10.1037/dev0000050

the CDI is easier to administer than the EVT when assessing children's productive language abilities. Consequently, in line with Simcock and Hayne (2003) we for pragmatic reasons choose to use the CDI as well. Seventh, we deliberately decided not to measure the children's ability to talk about the machines prior to instructions. During piloting we initially asked the children about what they thought their target machine was for, and the children said absolutely nothing and generally found the situation quite unpleasant. Note, since we were able to use the checklist of words (see below) specifically associated with the target event completed by the parents at both encoding and at the delayed recall test, we retained the possibility to assess whether the children did use newly acquired words from their productive vocabulary to describe the to-be-remembered event at the recall test. Hence, a verbal memory test at the test of encoding was not necessary to test the main hypothesis.

Method

Participants

A total of 180 children (87 female) participated in the present study. Ninety-four of the children were 33 months of age (45, female, M_{age} = 33.00 months, SD = 0.25, range: 32.50-33.50 months) and 86 of the children were 39 months of age (42 female; M_{age} = 39.08 months, SD = 0.27, range: 38.60-40.10 months). Within each age group, the children were randomly assigned to either the MSM Condition or the CDC Condition. Additionally six children (three 33-month-olds and three 39-month-olds) were tested but excluded from the final sample due to missing data from the CDI (5) and experimental error (1). The children were predominantly Scandinavian Caucasian, living in families with middle to higher SES. All the children were healthy and full-term and were recruited from the [Blinded] via registers from the National Board of Health. The families were contacted by letter and follow-up phone calls. The children received a small gift for their participation.

Apparatus

All three sessions (see below) took place in the same 30 m^2 lab room. At the first session, the encoding session (T_1) and the second session, the test of encoding (T_2), only one of the two custom-made "machines" was present (see Fig. 1), that is, *either* the MSM, 80 cm x 43 cm x 60 cm [L x W x H], in brown, yellow, and orange, a replica of the device developed and used by Simcock and Hayne (2002, 2003); *or* the CDC, 60 cm x 43 cm x 80 cm [L x W x H], in grey, red, yellow, blue, and green, which, as mentioned above, acted as a second 'magic' device enabling us to control for on-line reasoning. Both machines were designed specifically for the present experiment and not commercially available. Two JVC Everio digital camcorders recorded all sessions for later off-line scoring (see below).

Procedure

All children came to the laboratory three times: two visits separated by 24hr (T_1 and T_2) and a third visit (T_3) after a six month delay. On the first two visits half of the children in each age group were taught how the MSM worked and the other half of children were taught how to operate the CDC. On the third visit all the children were asked first to *tell* everything they could about both of the machines and subsequently to *demonstrate* by behavioral re-enactment how to operate them.

The children had to (re)enact five different target actions in the correct order in order to operate each machine. If they succeeded in operating the machines correctly, the MSM would 'shrink' inserted objects into smaller identically looking replicas whereas the CDC could 'copy' objects. The target actions related to the MSM were identical to those described by Simcock and Hayne (2002, 2003): (1) turn on the machine by pulling down a yellow lever activating a light on top of the machine, (2) take a toy from a bag, (3) place it inside the machine, (4) turn a green handle on the left side of the machine, and finally (5) retrieve a smaller identical looking toy by opening an orange door on the machine's front. Six objects were used in relation to the MSM: a giraffe, play dough, a scarf, a can, a T-shirt and a crayon. The objects used in the present study were not completely identical to the ones used by Simcock and Hayne (2002, 2003) because we were unable to find both large and small versions of the original objects.

In order to minimize carry-over effects between the two devices, all the *target actions* and *target objects* related to the CDC were chosen, so the key words to-be-remembered differed from the ones relevant for making the MSM work. The five target actions for the CDC were: (1) take a toy from a basket, (2) place the object on a shelf inside the machine and close a small door, (3) pick up a cymbal from a red pocket on the machine's left side and to put it on a stick on the right side of the machine, (4) take a club, placed on the machine's front, and hit the cymbal, and finally, (5) push a sliding door aside to retrieve the now two identical toys, 'copied ' by the CDC. The six target objects used with the CDC were: a squirrel, a camera, a belt, a cup, a rooster, and a stove.

Notice: This is the author's version of a work that was accepted for publication in *Developmental Psychology.* A definitive version was subsequently published in *Developmental Psychology, 51,* 1704-1716. DOI: 10.1037/dev0000050

Encoding (T_1).

Baseline. An assistant sat behind the machine in order to make the machine shrink or copy the objects provided that the correct actions were produced in the correct order. Following a brief warm-up, the experimenter first encouraged the child to show her what he or she could do with his or her target device ("Look at this exciting toy [Pointing to the bag (MSM) or the basket (CDC)] and this interesting machine I´ve got here [touching the machine]. What do you think you can do with it? Show me what you can do!"). The baseline phase provided us with a measure regarding whether the child *without* instructions would produce any of the target actions for the given machine. The trial was "child controlled" (Bauer & Mandler, 1992), meaning that the period ended when the child stopped playing, abandoned the props, or began engaging in repetitive behavior (e.g., pulling the lever up and down again and again).

Demonstration. The experimenter began by saying: "This machine is actually a Magic Shrinking Machine" or "This machine is actually a Crazy Duplicator" [depending on the child's target device]. "The machine only works if you operate it in a specific order. Let me show you how it works". Then the experimenter demonstrated the five target actions always starting with the giraffe (MSM) or the squirrel (CDC). During the demonstration the experimenter narrated what she was doing with the target machine and thus described all the to-be-remembered target actions. Each child was encouraged to participate in the game at any time during the first demonstration. The shrinking or copying sequence was repeated for all six target objects, one object at a time. When the last object had been shrunken or copied the child received a medal as a reward for his or her participation at the end of the session of encoding. The medal was used as a retrieval cue 24hr's later in relation to the nonverbal recall test of encoding (T_2) (cf. Simcock & Hayne, 2002, 2003).

Language development. Both at the time of encoding (T_1) and at the time of the final recall test (T_3), we asked the parents to fill out a Danish version of the MacArthur Communicative Development Inventory: Words and Sentences (CDI) in order to assess the children's general language and the development of their productive language skills. In the reference study Simcock and Hayne (2002, 2003) asked the parents to fill out an extra checklist of words regarding the children's productive vocabulary specifically associated with

Notice: This is the author's version of a work that was accepted for publication in *Developmental Psychology.* A definitive version was subsequently published in *Developmental Psychology, 51,* 1704-1716. DOI: 10.1037/dev0000050

the to-be-remembered event. Thus, in the reference study only the children's *productive* vocabulary was used in order to check for the possible use of newly acquired words. This strategy makes sense in order to assess whether possible verbal productions related to the to-be-remembered event by the children were new or not. In order to replicate the study, we employed the same strategy, but asked the parents to fill out a checklist involving words for *both* machines (see below).

Additionally, in line with Simcock and Hayne (2003) each child's parent was instructed not to initiate any conversation about the target event with their child during the 24hr retention interval. In case the child mentioned anything related to the target event, the parents were instructed only to respond briefly and in confirmatory fashion, but not to elaborate on any statements provided by the child.

Test of encoding (T_2).

Behavioral re-enactment. Replicating the procedure by Simcock and Hayne (2002) all the children returned to the laboratory for a booster session 24hr's after the first visit. In front of their target machine the experimenter asked each child: "Yesterday, I gave you a medal like this [holding the medal in her hand], do you remember that?" and "Before I gave you the medal, we played an exciting game with this machine [touching the machine] and the stuff in the bag or the basket [pointing at it]. Can you show me, what we did?" Similar to the procedure at T_1 the shrinking and copying sequences were repeated for all six objects, one object at a time. When the re-enactment stopped the experimenter asked "What else did we do?" in order for the child to re-enact as many target actions as possible by him or herself. In case a child failed to re-enact one or more actions, these missing actions were demonstrated again and the child was encouraged to participate in order to maximize encoding.

Recall test (T_3).

Six months (± 14 days; M_{days} =181.15, *SD* =7.66) after the test of encoding, the children returned to the lab for the recall test. The test involved three steps: First, we assessed whether the child could *recognize* his or her target machine based on two A4 sized photos (left and right balanced) (recognition test). The children replied verbally or by pointing. Second, with the machine in question fully visible, but firmly out of the children's reach, we tested how much each child could *tell* about each machine (verbal recall). Finally, we assessed whether the child was able to *operate* the two machines (nonverbal recall). The verbal recall test and the nonverbal recall test were run successively for each machine before turning to the other machine. The verbal recall test was always conducted *before* the nonverbal recall test in order to rule out the possibility that the latter would prompt or boost the former. The order of the machines was balanced across subjects. In the following the verbal recall test and the nonverbal recall test are outlined in more detail, one at a time.

Verbal recall. One of the machines was made visible by removing a blanket. In order to assess the child's verbal recall of the machine the experimenter asked each child general, open-ended questions about the event. Whereas the children could see the machine, they were *not* allowed to touch the machine in order to rule out a potential carry–over effect from present nonverbal (i.e., behavioral re-enactment) experiences to verbal recall. The target objects were hidden in the bag (MSM) and the basket (CDC) except for the giraffe and the squirrel as they were placed on top of their target machine, respectively. Because one of the machines was familiar to the child, whereas the other was new, the questions differed slightly. With the target machine the experimenter started by asking (e.g., "This is the machine you and I played with the last time you were here, please tell me everything you can remember about the machine", "Can you tell me more about the machine?"; "What else did we do?"). When the child stopped and could not report or add new information regarding the target machine the experimenter asked more direct questions (e.g. "What kinds of toys did we use?", "How did we make the machine work?").

When the child was asked about the foil machine the experimenter said "This is the new machine you haven't seen before; please tell me everything you can about this machine", "Can you tell me more about the machine?" and "What do you think the machine can do"). Regarding the foil machine the experimenter only asked about the functions of the machine (e.g., "How do you think the machine works"?). The experimenter kept asking the child questions about the to-be-remembered event or the foil device until the child could not report any more information about the event.

Nonverbal recall. After the verbal recall test the experimenter encouraged the child to show whether he or she was able to operate the machine. The experimenter only used general verbal prompts (e.g., Target Machine: "Show me what we did?", "What did we do next?"; Foil Machine: "Show me, what you can do?"; What else can you do?"). If the child was able to remember the five target actions in their specific order, the child was allowed to shrink or

Notice: This is the author's version of a work that was accepted for publication in *Developmental Psychology.* A definitive version was subsequently published in *Developmental Psychology, 51,* 1704-1716. DOI: 10.1037/dev0000050

copy all six objects, one at a time. However, if the child failed to operate the machine the experimenter encouraged the child to try again and mentioned that the correct order was necessary in order to make the machine work but without specifically telling the child what to do. The nonverbal recall test ended when the child had shrunken or copied all six objects or when the child was unable to produce any new target actions related to the machine.

Scoring and Data Reduction

All three sessions were video recorded and the content of the children's reports at the verbal recall test were transcribed verbatim. The primary scorer had to know about the event in detail in order to decide whether the verbal reports provided by the child were an indicator of memory of the to-be-remembered event or not. In order to objectify the scoring process a secondary scorer re-scored 20% of the recordings after having received training by the primary scorer based on a manual and training based on some of the children not selected for re-scoring. The rescoring involved all aspects of the scoring, that is, both verbal and behavioral assessments. The interrater agreement was very high 99.3%; Pearson r =.998.

Scoring: Nonverbal recall

Baseline at encoding (T_1). The number of correct target *actions* produced was scored following the scoring strategy employed by Bauer (1992). Hence, the child received a "1" for each target action produced correctly. Each target action could only be credited once (range: 0-5). All the children were awarded with their overall best score. For instance, if a child produced two target actions with one target object but later managed to produce three target actions then the latter performance constituted the child's final score.

Test of encoding (T_2). To compare the child's behavioral enactment at baseline to their re-enactment during demonstration the child's behavioral re-enactment was scored in the same manner as just described. For each of the six target objects the number of target actions (max. 5) produced by the child were registered. Only the actions produced by the child before the experimenter eventually demonstrated the missing actions again were scored.

Scoring: Recognition test at T_3.

Each child received a "1" if he or she was able to identify the target device and "0" if he or she could not.

Scoring: Verbal recall at T_3.

Notice: This is the author's version of a work that was accepted for publication in *Developmental Psychology.* A definitive version was subsequently published in *Developmental Psychology, 51,* 1704-1716. DOI: 10.1037/dev0000050

The primary scorer scored the content of the children's verbal accounts based on the recordings. In the reference experiment Simcock and Hayne (2002) used a checklist of words associated with the target event to assess whether the children used newly acquired words from their productive vocabulary to describe the machines or not. However, because the target objects used in the present study differed from the ones used in the original study, the wordlist we adopted was not completely identical to the one used in the reference experiment (see Appendix A). In addition, the checklist developed by Simcock and Hayne (2002) did not include words related to the CDC. Consequently, we developed a checklist consisting of 22 words related to this device (see Appendix A). Thus, our wordlist associated with the two machines consisted of 44 words (22 for each device) and the children received "1" point for each word he or she mentioned on this additional checklist. Hence, it was possible for each child to get a maximum score of 22 (range: 0-22) for each machine based on the wordlist. Note that the parents filled out a single wordlist consisting of the 44 words associated with the two target events (listed in alphabetic order), and thus did not know that we during the scoring operated with two different wordlists, one for each target machine. Last, but not least, based on the CDI reports on each child's vocabulary assessed at encoding as well as at test, we checked whether any of the newly acquired words from the children's productive vocabulary of relevance for the machines, was used by the child during the test (T_3) to describe the two machines.

Scoring: Nonverbal recall at T_3.

At the delayed recall test the child's behavioral re-enactment regarding the number of target actions produced was scored. Analogous with the scorings conducted at T_1 and T_2 sessions a total score for each child in relation to their target machine was calculated. Furthermore, the child also received one point for each target action (range: 0-5) produced in relation to their foil machine in order to compare the child's performance related to the target machine vs. the foil machine, respectively.

Results

Initially, we tested whether the children differed with regard to productive vocabulary across age groups and conditions. To address this question, we conducted a two-way ANOVA with Age (33 months vs. 39 months) and Condition (MSM vs. CDC) as between-subjects factors and CDI score at T_2 (encoding test) as the dependent variable. The analysis only revealed a main effect of Age (F[1, 176] = 21.47, $p < .001$, $\eta_p^2 = .11$; $M_{33\text{m}} = 512.04$; *SD*

Notice: This is the author's version of a work that was accepted for publication in *Developmental Psychology.* A definitive version was subsequently published in *Developmental Psychology, 51,* 1704-1716. DOI: 10.1037/dev0000050

= 102.51; M_{39m} = 580.21, *SD* = 93.21) whereas Condition had no effect ($F < .90$). Thus, across Conditions the children were considered to be equivalent with regard to their productive vocabulary. Second, as the core question was to investigate the children's possible use of newly acquired words at the verbal recall test, we analyzed by means of paired samples *t*-tests (two-tailed) whether the 33-month-olds' and 39-month-olds' productive vocabulary had developed during the six months retention interval. The results revealed that the 33-month-olds productive vocabulary expanded significantly from the time of encoding (M = 512.04, SD = 102.51) to the time of the verbal recall test (M = 609.64, SD = 78.90), $t(93) = -13.273$, $p < .001$, $r = .81$. A similar pattern of results was found for the 39-month-olds (Encoding: M = 580.21, SD = 93.21 vs. Delayed test: M = 634.24, SD = 71.60), $t(85) = -8.276$, $p < .001$, $r = .67$). Thus, in both age groups the children increased their productive vocabulary significantly across the six months retention interval. Note that even though the children learned new words, the number of productive words according to the CDI recordings (obtained for all children at T_1 and T_3) did not reach a ceiling effect (Bleses, Vach, Wehberg, Faber, & Madsen, 2007).

In the following we first present the results from the recognition test. Subsequently, the results from the verbal recall test are presented, and finally we briefly present the results from the nonverbal recall test.

Recognition test

By means of binominal tests (against 0.5 test prob., two-tailed) we analyzed whether the children were able to identify their target machine. The results revealed that 82% of the 33-month-olds, $p < .001$ and 93% of the 39-month-olds, $p < .001$ were capable of doing so. Thus, the children in both age groups were able to recognize their target machine. Meanwhile, a Chi-square test revealed that the older children performed significantly better than their younger peers, $\chi^2(1) = 4.97$, $p = .026$, $V = .166$.

Verbal recall test

First, we analyzed whether the number of words from the wordlists used for each machine differed depending on which machine the children had as target machine.

The children's production of words listed on the *MSM wordlist.*

The analysis was conducted by means of an ANOVA with Target Machine (MSM vs. CDC) and Age Group (33 months vs. 39 months) as between-subjects factors and number of

Notice: This is the author's version of a work that was accepted for publication in *Developmental Psychology.* A definitive version was subsequently published in *Developmental Psychology, 51*, 1704-1716. DOI: 10.1037/dev0000050

produced words from the wordlist related to the MSM as the dependent variable. This analysis revealed a main effect of the factor Target Machine, $F(1, 176) = 27.80$, $p < .001$, $\eta_p^2 = .136$. The results showed that the children who had the MSM as target machine produced more words from the wordlist related to the MSM, ($M_{\text{words\_MSM}} = 2.64$, $SD = 1.75$) than the peers who had the CDC as target machine ($M_{\text{words\_MSM}} = 1.36$, $SD = 1.43$). The ANOVA also revealed a main effect of Age Group, $F(1, 176) = 7.91$, $p = .005$, $\eta_p^2 = .043$. The 39-month-old children generally produced more words ($M_{\text{words\_MSM}} = 2.44$, $SD = 1.87$) from the wordlist related to the MSM than their younger peers ($M_{\text{words\_MSM}} = 1.67$, $SD = 1.51$). The analysis revealed no other main effect or interactions.

The children's production of words listed on the CDC wordlist.

This analysis was repeated for the wordlist related to the CDC by means of an ANOVA with Target Machine (MSM vs. CDC) and Age Group (33 months vs. 39 months) as between-subjects factors and number of produced words from the wordlist related to the CDC as dependent variable. The analysis resulted in a main effect of the factor Target Machine, $F(1, 176) = 31.71$, $p < .001$, $\eta_p^2 = .150$. The children who had the CDC as target machine produced more words from the wordlist related to the CDC, ($M_{\text{words\_CDC}} = 1.26$, $SD = 1.24$) than the peers who had the MSM as target machine ($M_{\text{words\_CDC}} = 0.42$, $SD = 0.74$). No other main effect or interactions were obtained.

To summarize, results for the children's production of words from the two target wordlists were clear and systematic. Regardless of which machine the children had as their target machine (MSM or CDC) they systematically produced more words from the wordlist related to their target machine compared to the wordlist related to the foil machine.

Did the children use newly acquired words to describe the original event?

We now turn to the crucial question of whether the children used newly acquired words to talk about their target events. Based on the CDI records regarding each child's productive vocabulary at encoding and at recall together with transcripts of each child's factual utterances at the delayed recall test, we were able to register any incidents of use of newly acquired words from the wordlists produced at the time of recall. Although the total number of new words used was sparse, the children in both age groups and across target machines did indeed use newly acquired words from their productive vocabulary to describe the original event (see Table 1a and 1b for details).

Notice: This is the author's version of a work that was accepted for publication in *Developmental Psychology.* A definitive version was subsequently published in *Developmental Psychology, 51,* 1704-1716. DOI: 10.1037/dev0000050

Overall, the children who had the MSM as target machine produced at the verbal recall test at T_3 altogether 39 newly acquired words out of a total of 253 words from the wordlist, whereas the children who had the CDC as target machine produced a sum of 18 newly acquired words form the wordlist out of a total of 106 words produced at the recall test.

Recall that in the Simcock and Hayne (2002) study there was not a single incident in which a child used a newly acquired word to talk about the previously experienced event. In order to test for the significance of the use of newly acquired words, and since the number of new words used at recall was registered for each child, we used one-sample t-test with the test value "0" (i.e., no new words). Across age groups and machines, the number of newly acquired words from the wordlists used to describe the machines did indeed differ from "0" (see Table 1b). Hence, the results presented in the present study indicate that, provided that the props are present at recall, it is possible for young children to use newly learned words to describe an event which occurred six months earlier at a time when the child's productive vocabulary was less developed.

The analyses above are restricted in the sense that, although they show that the children did indeed use newly acquired words from their productive vocabulary to describe the two machines, they do not explicitly test whether the use of new words depended on whether the children had the MSM or the CDC as target machine. The results from these explicit tests are presented in the following. We first analyzed the new words used to describe the MSM. By means of an ANOVA with Target Machine (MSM vs. CDC) and Age Group (33 months vs. 39 months) as between-subjects factors and with the number of new words used for the MSM as the dependent variable we analyzed whether the number of new words used to describe the MSM differed depending on which machine the children had had as target machine. This analysis revealed a main effect of Target Machine, $F(1, 176) = 7.22$, $p = .008$, $\eta_p^2 = .039$. No other main effects or interactions were obtained. Thus, the results show that the children who had the MSM as target machine produced reliably more newly acquired words when describing the MSM at the verbal recall test ($M_{\text{new\_words\_MSM\_w/MSM as Target}} = 0.41$, $SD = 0.82$) compared to the children who had the CDC as target machine ($M_{\text{new\_words\_MSM\_w/CDC as Target}} = 0.14$, $SD = 0.39$). Consequently, the difference in results must be attributed to memory and *not* to on-line reasoning.

This analysis was repeated for the number of new words used for the CDC by means of an ANOVA with Target Machine (MSM vs. CDC) and Age Group (33 months vs. 39

Notice: This is the author's version of a work that was accepted for publication in *Developmental Psychology*. A definitive version was subsequently published in *Developmental Psychology, 51*, 1704-1716. DOI: 10.1037/dev0000050

months) as between-subjects factors but this time with new words used for the CDC as dependent variable. This analysis again revealed a main effect of Target Machine, $F(1, 176) = 9.33$, $p = .003$, $\eta_p^2 = .050$. Again no other main effects or interactions were obtained. Thus, replicating the pattern of results obtained above, the results revealed that the children who had the CDC as target machine produced reliably more newly acquired words when describing the CDC at the verbal recall test ($M_{\text{new\_words\_CDC\_w/CDC as Target}} = 0.21$, $SD = 0.47$) compared to the children who had the MSM as target machine ($M_{\text{new\_words\_CDC\_w/MSM as Target}} = 0.05$, $SD = 0.22$). Again, these results show that the difference in results must be attributed to memory and *not* to on-line reasoning.

Nonverbal recall test

In line with Simcock and Hayne (2002, 2003) we analyzed the children's nonverbal recall as these data provide the basis for showing that the children did indeed remember the to-be-remembered event experienced six months prior to the test. However, since the primary focus of the present study is on the children's *verbal memory* for the to-be-remembered event, we only present the main effects in order to avoid digression. The children's ability to re-enact the five target actions were analyzed by means of a mixed-model ANOVA with Target Machine (MSM vs. CDC) and Age Group (33 months vs. 39 months) as between-subjects factors and Sessions_Machine as within-subjects factors (baseline-targetT_1 vs. encoding-targetT_2 vs. test-targetT_3 vs. test-foilT_3) and with the number of target actions as the dependent variable. This analysis revealed a strong main effect of the factor Sessions_Machine, $F(3, 528) = 599.39$, $p < .001$, $\eta_p^2 = .773$. The analysis also revealed a main effect of Target_Machine, $F(1, 176) = 4.83$, $p = .03$, $\eta_p^2 = .027$. The means and standard deviations for all measures regarding nonverbal recall across age groups and conditions are displayed in Table 2 and a graphic representation of these results is displayed in Figure 2.

Subsequent pairwise post hoc tests using the Bonferroni correction revealed that as hypothesized the children overall produced significantly more target actions immediately after the encoding session compared to baseline ($p < .001$); that the children produced significantly more target actions at T_3 compared to baseline ($p < .001$), but also that the children had forgotten a significant amount of target actions across the retention interval ($p <$

Notice: This is the author's version of a work that was accepted for publication in *Developmental Psychology.* A definitive version was subsequently published in *Developmental Psychology, 51,* 1704-1716. DOI: 10.1037/dev0000050

.001). Finally, the children produced significantly more target actions at T_3 with their target machine compared to with the foil machine ($p < .001$), indicating that their performance after the six months' delay could not be explained by maturation but should be attributed to memory. In short, the basic design worked as planned.

Discussion

In the present study we replicated the MSM study used by Simcock and Hayne (2002). In contrast to the reference experiment we allowed the props to be present at the verbal recall test while controlling for potential on-line reasoning. The results were clear and systematic. We found that the children in both age groups and conditions used newly acquired words from their productive vocabulary to describe the previously experienced event, and that the use of newly acquired words depended reliably on which machine they had had as target machine, indicating that the difference in results could not be explained by on-line reasoning, but should be attributed to memory.

Why were the children in the present study able to use newly acquired words to describe the event whereas in the Simcock and Hayne (2002) study none of the children employed newly acquired words from their productive vocabulary to describe the to-be-remembered event at recall? We hypothesized that the presence of props at recall would lead to the obtained differences in results. However, before we can make any conclusions in this regard, we need to carefully discuss alternative explanations: Even though our design differed with regard to other factors, for example *place of testing* (home vs. lab), *age groups* tested (we did not test the 27-month-olds), *retention interval* (we only used a six-month but not a twelve-month retention interval), and *target objects* (we used different ones), we do not believe that the differences in the results obtained can be explained by these differences in the designs. First, existing evidence from studies where the context for test has been manipulated (in the lab vs. at home) has not shown any difference in results (Bauer et al., 2004). Second, even though the exclusion of the 27-month-olds in our study may overall have improved the average results, this cannot explain the difference in results as we found that *both* the 33-month-olds and 39-month-olds used new words, whereas Simcock and Hayne (2002) found that no children used newly acquired words regardless of age group. Third, although we only used a six-month retention interval (and not a six- *and* a twelve-month retention interval as in the reference study) the obtained results are in our view still comparable and convincing. Even with the shortest retention interval (six months) Simcock and Hayne (2002) found that

Notice: This is the author's version of a work that was accepted for publication in *Developmental Psychology*. A definitive version was subsequently published in *Developmental Psychology, 51,* 1704-1716. DOI: 10.1037/dev0000050

none of the children used newly acquired words, whereas the children in the present study did after the six-month retention interval. Fourth, although we did not use the exact same target objects for the MSM as used in the reference study, we consider it unlikely that this difference could explain the results obtained. Our reasoning is mainly based on the fact that the majority (26 out of 39) of the times the children who had the MSM as target used newly acquired words at recall, the children used words that *also* appeared on the Simcock and Hayne (2002, 2003) wordlist. Therefore, it could not have been the partial use of different objects for the MSM that caused the effect.

Can we rule out that on-line reasoning contributed to the children's use of newly acquired words from their productive vocabulary at the verbal recall test? The fact that the children used newly acquired words to describe not only their target machine, but *also* their foil machine (although only very rarely) suggests that the presence of props actually contributed marginally to on-line reasoning as well. However, the crucial point is that the contribution from on-line reasoning was under experimental control. Remember, that the verbal report (1) was obtained while the machines were in sight but out of reach, and (2) was invariably obtained *before* the nonverbal recall measures were collected. The magnitude of the number of new words used to describe each machine depended systematically on the machine each child had had as target machine. Thus, the children who had had the MSM as target machine produced approximately three times as many new words from their productive vocabulary related to the MSM ($M_{\text{new\_words\_MSM\_w/MSM as Target}} = 0.41$) compared to the children who had had CDC as target machine ($M_{\text{new\_words\_MSM\_w/CDC as Target}} = 0.14$). Similarly, the children who had had the CDC as target machine produced about four times as many new words related to the CDC ($M_{\text{new\_words\_CDC\_w/CDC as Target}} = 0.21$) compared children who were exposed to the MSM as target machine ($M_{\text{new\_words\_CDC\_w/MSM as Target}} = 0.05$). The fact that the children's use of newly acquired words from their productive vocabulary to describe the machines depended reliably on which machine they had had as target machine provide compelling evidence that the use of newly acquired words related to the machines should be attributed to memory, and not to on-line reasoning.

Why should the presence of props make a difference? In both the Simcock and Hayne (2002) study as well as in the present study the investigation of the young children's verbal recall was elicited during free and, subsequently, prompted recall. This may have initiated a top-down retrieval process and thus forced a difficult task upon the children in order for them to retrieve the memory of the event due to less mature frontal lobe structures in this age range (e.g., Johnson, 2005). However, in the present study we also provided the children with a careful match of distinct cues between the original event and the context for retrieval (e.g., exact same lab, same female experimenter, props were present) which are known to facilitate recall in general (Hayne, 2004; Newcombe et al., 2007) and consistent with the notion of cue-item discriminability (Rubin, 1995). We speculate that this massive cuing using highly distinct props may have triggered bottom-up retrieval processes and thus facilitated the children's ability to verbalize their memory and to use newly acquired words while doing so.

By having the distinctive props present at recall our study share a crucial design characteristic with the previously mentioned study by Morris and Baker-Ward (2007), which so far had been the only controlled study demonstrating that young children could use newly acquired words to describe an event experienced at a time when the their vocabulary was less evolved. As argued above we have reason to believe that the presence of distinct props at recall was central to the obtained results in both the experiment by Morris and Baker-Ward (2007) as well as in the study reported here. However, one might argue that the presence of props at recall was only *partial* in the Morris and Baker-Ward (2007) study. Whereas the bubble machine and the experimenter were present, the specific color pegs used were actually *not* present at recall (Morris & Baker-Ward, 2007, p. 453). Does this design feature undermine our argument? We do not think so. In accordance with the notion of cue-item discriminability (Rubin, 1995) the crucial facilitating factor may not be the amount of overlapping cues *per se*, but rather whether the available cues have a distinct and unique reference to exactly the to-be-remembered situation (and not to any other incidents). Note, that in both the study by Morris and Baker-Ward (2007) and in the present study, the presence of the 'magic machine' fulfilled exactly this criterion of being a distinct and unique cue providing an unambiguous reference to the to-be-remembered event.

Can we rule out the possibility that some of the 'new' words relevant for describing the target event could have entered the children's receptive vocabulary during the encoding sessions?[2] We acknowledge that we cannot rule out this possibility. However, first of all, we do not think this possibility restricts the scope of the obtained results substantially, as this possibility was also present in the Simcock and Hayne (2002) study. Although Simcock and Hayne (2002) tested both the children's receptive and productive vocabulary, all aspects related to assessing the children's free verbal recall (including the parental reports concerning words specifically related to the events) were exclusively based on the children's

[2] We would like to thank one of the anonymous reviewers for suggesting this possibility.

Notice: This is the author's version of a work that was accepted for publication in *Developmental Psychology*. A definitive version was subsequently published in *Developmental Psychology, 51,* 1704-1716. DOI: 10.1037/dev0000050

productive vocabulary. In order to be able to make direct comparisons with the results obtained in the Simcock and Hayne (2002) study, we chose to employ the same strategy in this regard. Since the possibility of learning 'new' and relevant words for describing the events during encoding must be considered the same in the present study and in the Simcock and Hayne (2002) study as the encoding sessions were equivalent, we cannot see how this possibility should constrain the documented effect of having the props present at recall. Second, although we acknowledge that it would have been advantageous to possess information regarding the children's receptive vocabulary just after the encoding sessions, collecting such data may have been problematic. For instance, it would be difficult to collect such data without at the same time running the risk of teaching the children the 'new' words *during* such a test, which in turn would warrant an additional control test etc.. Note that the Morris and Baker-Ward (2007) study deserves credit for conducting this control (i.e., while returning from the encoding session the children were asked which color could turn on the machine). In the present design, such a control measure would be very difficult to entertain as each wordlist involved not one, but 22 words.

General Discussion

The circumstances by which young children can and cannot use newly acquired words to describe events experienced substantially earlier is crucial when attempting to understand the enigma of childhood amnesia. In this light, surprisingly few studies have addressed this question. The results from the present study together with the results from the Simcock and Hayne (2002) study provides a more complete picture of the puzzle: there is no doubt that young children have difficulties employing newly acquired words to describe an event that took place at a time when the children's productive vocabulary was less advanced. When the props are absent at recall, they cannot; when the props are present at recall, they can, although only to a limited extent.

We believe that the results of the present study have important implications for our current understanding of the mechanisms responsible for, and involved in, childhood amnesia. Over the years a number of different explanations of childhood amnesia have been put forward (see Bauer, 2007, 2014; Hayne & Jack, 2011, for reviews) and several important components have been mentioned. For example, components such as brain development related to declarative memory in general (Bauer, 2007; Richmond & Nelson, 2007), and to episodic memory specifically (Newcombe et al., 2007), the cognitive self (Howe & Courage, 1993), socio-cultural factors – in particular supported by evidence on maternal reminiscences

Notice: This is the author's version of a work that was accepted for publication in *Developmental Psychology.* A definitive version was subsequently published in *Developmental Psychology, 51,* 1704-1716. DOI: 10.1037/dev0000050

style (Nelson & Fivush, 2004; Reese, 2009), and language acquisition (Morrison & Conway, 2010; Pillemer & White, 1989; Simcock & Hayne, 2002).

In this context, the present study can be seen as a test of the *language acquisition hypothesis* (Jack et al. 2012) stating that "…language development may be the rate-limiting step in the offset of childhood amnesia" (Simcock & Hayne; 2002, p. 230). Our results do indeed challenge this hypothesis in its absolute form: Although only a few newly acquired words were used to describe the to-be-remembered event, it actually did happen, and the frequency of newly acquired words used differed reliably from zero. Thus, the results provide evidence that the ability to use newly acquired words to describe previously experienced events is not a barrier in the absolute sense.

As already argued a careful match between the distinct cues present in the to-be-remembered event and in the recall setting seem to facilitate young children's ability to recall previously experienced events. This is probably partly due to the principle of cue-item discriminability (Rubin, 1995). However, when considering young children as is the case here, other factors may contribute to the facilitating effect than cue-item discriminability *per se*. According to Berntsen (2009) external cues often initiate *involuntary* recall, an automatic associative bottom-up process, requiring little executive control and relying less on frontal lobe structures compared to voluntary recall which typically involves a self-initiated, top-down search process requiring high degree of executive control. The voluntary top-down retrieval processes are therefore also considered by many theorists (e.g., Suddendorf & Corballis, 2007; Tulving, 1984) to be a highly advanced process requiring sophisticated search and control processes compared to the bottom-up retrieval process which relies on associative processes. Consistent with this idea Hall, Gjedde and Kupers (2008) suggest that the bottom-up retrieval process is an earlier ontogenetic developed achievement compared to the voluntary top-down process since the former is less dependent on the development of frontal lobe structures than the latter.

Applied to the present study the activation of these two different retrieval processes may help explain why the children in the present study were able to verbalize and use newly acquired words to describe their memory of the target event whereas their peers in the Simcock and Hayne (2002) study were not. We do not argue that the children's recall was involuntary; clearly it was not as the children were explicitly asked to remember the previously experienced event. What we *do* suggest is that the strong prominence in the recall setting of distinct cues referring unambiguously to the to-be-remembered event may have

Notice: This is the author's version of a work that was accepted for publication in *Developmental Psychology.* A definitive version was subsequently published in *Developmental Psychology, 51,* 1704-1716. DOI: 10.1037/dev0000050

facilitated bringing additional material to the children's mind by simple, associative means. Thus, we speculate that, *because* of the presence of props in the recall situation, at least some of the mnemonic material may have been brought to the children's mind without the need for employing the kind of strategic, voluntary retrieval that seems to be highly demanding for young children, granted their immature frontal lobes. Such an explanation is tentative, however, since systematic research on spontaneous recall in early childhood is almost non-existent in the literature (Krøjgaard, Kingo, Dahl, & Berntsen, 2014).

In conclusion, and in accordance with the results obtained by Morris and Baker-Ward (2002) we provide evidence that young children, provided that the distinct props are present at recall, can use newly acquired words from their productive vocabulary to verbally describe events that were experienced six months earlier when the children's vocabulary was less advanced.

Notice: This is the author's version of a work that was accepted for publication in *Developmental Psychology.* A definitive version was subsequently published in *Developmental Psychology, 51,* 1704-1716. DOI: 10.1037/dev0000050

References

Bauer, P.J. (1992). Holding it all together: How enabling relations facilitate young children's event recall. *Cognitive Development, 7,* 1-28.Bauer, P. J. (2007). *Remembering the times of our lives: Memory in infancy and beyond.* Mahwah, NJ: Laurence Erlbaum Associates.

Bauer, P. J. (2014). The development of forgetting: Childhood amnesia. In P. J. Bauer & R. Fivush (Ed.), *The Wiley handbook of the development of children's memory, Vol II*, 519-544. Wiley & Sons, Ltd. UK.

Bauer, P.J., Kroupina, M.G., Schade, J.A., Dropnik, P.L., & Wewerka, S.S. (1998). If memory serves, will language? Later verbal accessibility of early memories. *Development and Psychopathology, 10,* 655-679.

Bauer, P. J., & Mandler, J. (1992). Putting the horse before the cart: The use of temporal order in recall of events by one-year-old children. *Developmental Psychology, 28,* 441-452.

Bauer, P.J., Van Abbema, D.L., Wiebe, S.A., Cary, M.S., Phill, C., & Burch, M.M. (2004). Props, not pictures, are worth a thousand words: Verbal accessibility of early memories under different conditions of contextual support. *Applied Cognitive Psychology, 18,* 373-392.

Berntsen, D. (2009). *Involuntary autobiographical memories: An introduction to the unbidden past.* Cambridge University Press.

Berntsen, D., & Rubin, D. C. (2012). Understanding autobiographical memory: An ecological theory. In D. Berntsen & D.C. Rubin (Eds.), *Understanding autobiographical memory: Theories and approaches* (pp. 333-355). Cambridge: Cambridge University Press.

Bleses, D., Vach, W., Wehberg, S., Faber, K. & Madsen, T.O. (2007) *Tidlig Kommunikativ Udvikling*. [Early communicative development.] Syddansk Universitetsforlag.

Bowlby, J. (1951). *Maternal care and mental health.* Geneva. World Health Organization.

Boyer, M. E., Barron, K. L., & Farrar, M. J. (1994). Three-year-olds remember novel event from 20 months: Evidence for long-term memory in children. *Memory, 2,* 417-445.

Cleveland, & Reese, E. (2008). Children remember early childhood: Long-term recall across the offset of childhood amnesia. *Applied Cognitive Psychology, 22,* 127-142.

Fenson, L., Bates, E., Dale, P., Goodman, J., Reznick, J. S., & Thal, D. (2000). Measuring Variability in Early Child Language: Don't Shoot the Messenger. *Child Development, 71*, 323-328.

Fenson, L., Dale, P., Reznick, J. S., Bates, E.,Thal, D., & Pethick, S. J. (1994). Variability in Early Communicative Development. *Monographs of the Society for Research in Child Development, 59* (5, Serial No. 242).

Fivush, R., Gray, J. T., & Fromhoff, F. A. (1987). Two-year-olds talk about the past. *Cognitive Development, 2*, 393-409.

Freud, S. (1905/1953). Childhood and concealing memories. In *The basic writings of Sigmund Freud.* A. A. Brill (Trans. and Ed.). New York: Modern Library.

Goodman, G. S., Batterman-Faunce, J. M., Schaaf, J. M., & Kenney, R. (2002). Nearly 4 years after an event: children's eyewitness memory and adults' perceptions of children's accuracy. *Child Abuse & Neglect, 26,* 849-884.

Goodman, G. S., Hirschman, J. E., Hepps, D., & Rudy, L. (1991). Children's memory for stressful events. *Merrill–Palmer Quarterly, 37,* 109–157.

Hall, N. M., Gjedde, A., & Kupers, R. (2008). Neural mechanism of voluntary and involuntary recall. *Behavioural Brain Research, 186*, 261-272.

Hayne, H. (2004). Infant memory development: Implications for childhood amnesia. *Developmental review, 24,* 33-73.

Hayne, H., & Jack, F. (2011). Childhood amnesia. *Wiley Interdisciplinary Reviews: Cognitive Science, 2*, 136-145.

Howe, M. L., & Courage, M. L. (1993). On resolving the enigma of infantile amnesia. *Psychological Bulletin, 113,* 305-326.

Hudson, J. A. (1993). Reminiscing with mothers and others: Autoibographical memory in young children two-year-olds. *Journal of Narrative & Life History, 3,* 1-32).

Hudson, J.A., & Fivush, R. (1991). As times goes by: Sixth graders remember a kindergarten experience. *Applied Cognitive Psychology, 5,* 347-360.

Jack, F., Simcock, G., & Hayne, H. (2012). Magic memories: Young children's verbal recall after a 6-year delay. *Child Development, 83,* 159-172.

Johnson, M.H. (2005). *Developmental cognitive neuroscience.* 2nd edition. Oxford: Blackwell Publishing.

Krøjgaard, P., Kingo, O. S., Dahl, J. J., & Berntsen, D. (2014). "That one makes things small": Experimentally induced spontaneous memories in 3.5-year-olds. *Consciousness and Cognition, 30,* 24-35.

Morris, G., & Baker-Ward, L. (2007). Fragile but real: Children's capacity to use newly acquired words to convey preverbal memories. *Child Development, 78,* 448-458.

Notice: This is the author's version of a work that was accepted for publication in *Developmental Psychology.* A definitive version was subsequently published in *Developmental Psychology, 51,* 1704-1716. DOI: 10.1037/dev0000050

Morrison, C. M., & Conway, M. A. (2010). First words and first memories. *Cognition, 116,* 23-32.

Myers, N. A., Clifton, R.K., & Clarkson, M.G. (1987). When they were very young: Almost-threes remember two years ago. *Infant Behavior and Development, 10,* 123-132.

Myers, N. A., Perris, E. E., & Speaker, C. J. (1994). Fifty months of memory: a longitudinal study in early childhood. *Memory, 2,* 383-415.

Nelson, K. (1993). The psychological and social origins of autobiographical memory. *Psychological Science, 4,* 7-13.

Nelson, K., & Fivush, R. (2004). The emergence of autobiographical memory: A social cultural developmental theory. *Psychological Review, 111,* 486-511.

Nelson, K., & Ross, G. (1980). The generalities and specifics of long-term memory in infants and young children. *New Directions for Child and Adolescent Development, 10,* 87-101.

Newcome, N. S., Lloyd, M. E., & Ratliff, K. R. (2007). Development of episodic and autobiographical memory: A cognitive neuroscience perspective. In R. V. Kail (Ed.), *Advances in child development and behavior*, Vol. 35. San Diego, CA: Elsevier.

Peterson, C., & Parsons, B. (2005). Interviewing former 1- and 2-years olds about medical emergencies 5 years later. *Law and Human Behavior, 29,* 743-754.

Peterson, C., & Whalen, N. (2001). Five years later: Children's memory for medical emergencies. *Applied Cognitive Psychology, 15,* S7-S24.

Peterson, C., & Rideout, R. (1998). Memory for medical emergencies experienced by 1- and 2-year-olds. *Developmental Psychology, 34,* 1059-1072.

Pillemer, D. B., Picariello, M. L., & Pruett, J. C. (1994). Very long-term memories of a salient preschool event. *Applied Cognitive Psychology, 8,* 95-106.

Pillemer, D. B., & White, S. H. (1989). Childhood events recalled by children and adults. *Advanced in Child Development and Behavior, 21,* 297-340.

Pipe, M-E., Gee, S., & Wilson, C. (1993). Cues, props, and context: Do they facilitate children's event reports? In G. S. Goodman & B. L. Bottoms (Ed.), *Child victims, child witnesses: Understanding and improving testimony.* The Guilford Press, NY.

Reese, E. (2009). The development of autobiographical memory: Origins and consequences. In P. Bauer (Eds.), *Advances in child development and behavior, 37,* 145-200. The Netherlands, NL: Elsevier.

Notice: This is the author's version of a work that was accepted for publication in *Developmental Psychology.* A definitive version was subsequently published in *Developmental Psychology, 51,* 1704-1716. DOI: 10.1037/dev0000050

Richmond, & Nelson (2007). Accounting for change in declarative memory: A cognitive neuroscience perspective. *Developmental Cognitive Neuroscience, 27,* 349-373.

Rovee-Collier, C., & Cuevas, N. (2009). The development of infant memory. In M.L. Courage & N. Cowan, (Eds.), *The development of memory in infancy and childhood* (pp. 11-41). Hove & New York: Psychology Press.

Rubin, D. (1995). *Memory in oral traditions: the cognitive psychology of epic, ballads, and counting-out rhymes.* New York: Oxford University Press.

Rutter, M. (2002). Maternal deprivation. In M. H. Bornstein (Ed.), *Handbook of parenting Vol. 4 Applied parenting* (pp. 181-202). Mahwah, NJ: Lawrence Erlbaum Associates.

Simcock, G., & Hayne, H. (2002). Breaking the barrier? Children fail to translate their preverbal memories into language. *Psychological Science, 13,* 225-231.

Simcock, G., & Hayne, H. (2003). Age-related changes in verbal and nonverbal memory during early childhood. *Developmental Psychology, 39,* 805-814.

Suddendorf, T. & Corballis, M. C. (2007). The evolution of foresight: What is mental time travel and is it unique to humans? *Behavioral and Brain Sciences, 30,* 299-313.

Todd, C.M., & Perlmutter, M. (1980). Reality recalled by preschool children. *New Directions for Child and Adolescent Development, 10,* 69-85.

Tulving, E. (1984). Précis of elements of episodic memory. *The Behavioral and Brain Sciences, 7,* 223-268.

Notice: This is the author's version of a work that was accepted for publication in *Developmental Psychology*. A definitive version was subsequently published in *Developmental Psychology*, *51*, 1704-1716. DOI: 10.1037/dev0000050

Table 1a

Displays the frequencies of newly acquired words used at T_3 across age groups and target machines. The line "Number of new words said" serves as a header for presenting from 0-4 newly acquired words for each target machine. The figures displayed in the lines headed with either "33-" or "39-months-olds" represent the number of children who reported either 0, 1, 2, 3, or 4 newly acquired words at T_3 relative to their target machine.

	The Magic Shrinking Machine						The Crazy Duplicator					
Number of new words used	0"	1"	2"	3"	4"	Total # of new words used	0"	1"	2"	3"	4"	Total # of new words used
33-month-olds (MSM: N=47; CDC: N=47)	6					18	8					9
39-month-olds (MSM: N=49; CDC: N=37)	4	2				21	0					9

Table 1b

Displays the of prevalence in M's and SD's of words said from wordlists as well as newly acquired words from the wordlist used by the children across age groups and target machines to describe the original event including results from one-sample t-tests.

	The Magic Shrinking Machine								The Crazy Duplicator							
	All words said from wordlist		Newly acquired words said from wordlist		Newly acquired words? One sample t-test (test value = "0")				All words said from wordlist		Newly acquired words said from wordlist		Newly acquired words? One sample t-test (test value = "0")			
	M	*SD*	*M*	*SD*	*t*	*df*	*p*	*r*		*SD*		*SD*	*t*	*df*	*p*	*r*
33-month-olds	2.15	(1.55)	.38	(.82)	.193	6	.003	.43	.26	(1.17)	.19	(.40)	.301	6	.002	.44
39-month-olds	3.10	(1.83)	.43	(.82)	.674	8	.001	.47	.27	(1.35)	.24	(.55)	.700	6	.010	.41

Notice: This is the author's version of a work that was accepted for publication in *Developmental Psychology*. A definitive version was subsequently published in *Developmental Psychology*, *51*, 1704-1716. DOI: 10.1037/dev0000050

Table 2

Means and Standard Deviations for the number of different Target Actions.

Sex	Age group and	The Magic Shrinking Machine				The Crazy Duplicator			
		T_{1-} Baseline	T_{2-} Encoding	$T_{3\text{-Target}}$ Machine	$T_{3\text{-Foil}}$ Machine	T_{1-} Baseline	T_{2-} Encoding	$T_{3\text{-Target}}$ Machine	$T_{3\text{-Foil}}$ Machine
		(*SD)*	(*SD)*	(*SD)*	(*SD)*	(*SD)*	(*SD)*	(*SD)*	(*SD)*
= 5.0)	Target Actions (Max								
	33-month-olds:	96 (.96)	.81 (.83)	.53 (1.64)	1.26 (.85)	.26 (.53)	.98 (.15)	.49 (1.25)	.00 (1.23)
	39-month-olds:	.12 (.99)	.96 (.29)	.82 (1.59)	.33 (.85)	.05 (.82)	.97 (.16)	.05 (1.62)	.73 (1.26)

Figure 1

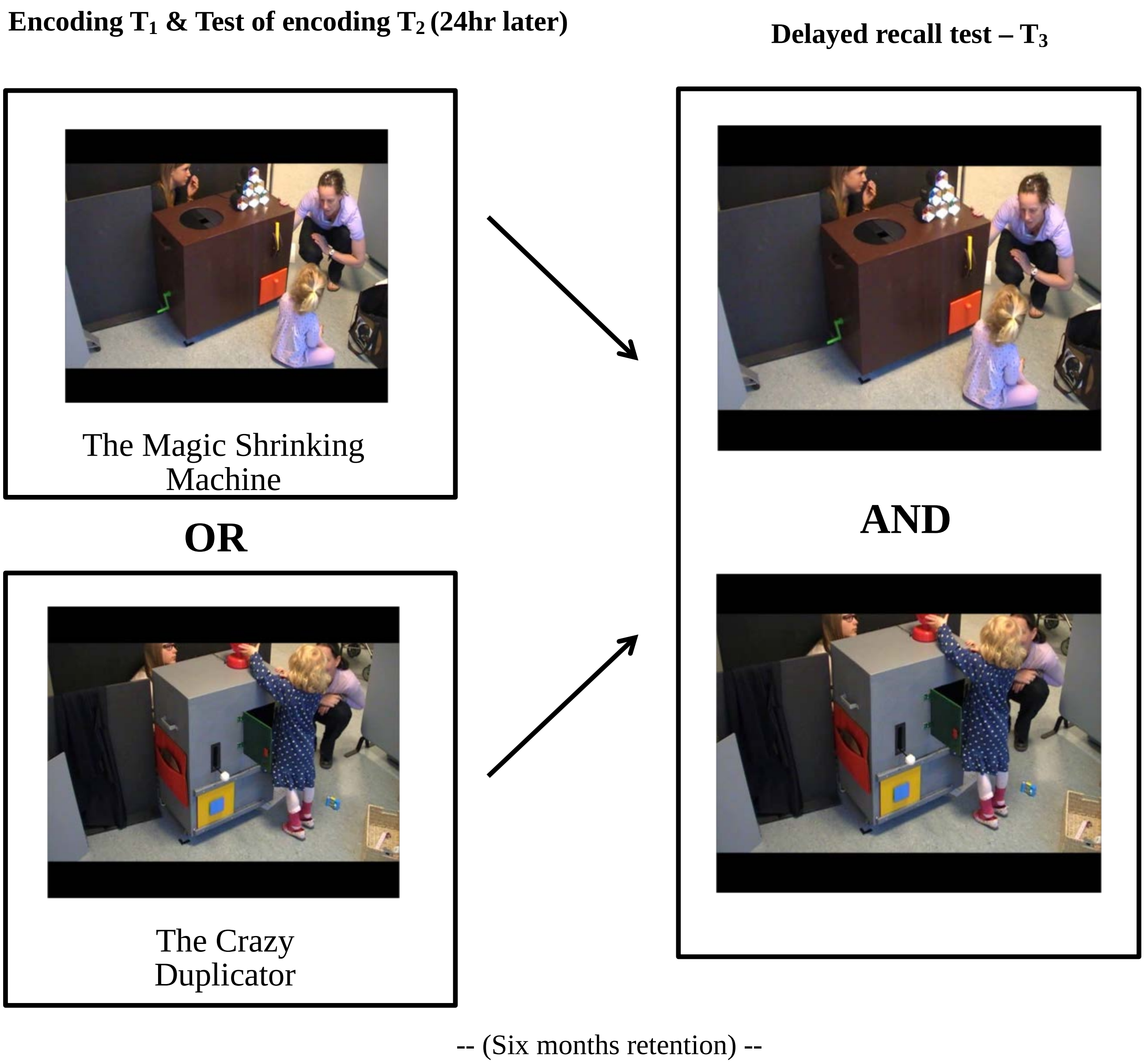
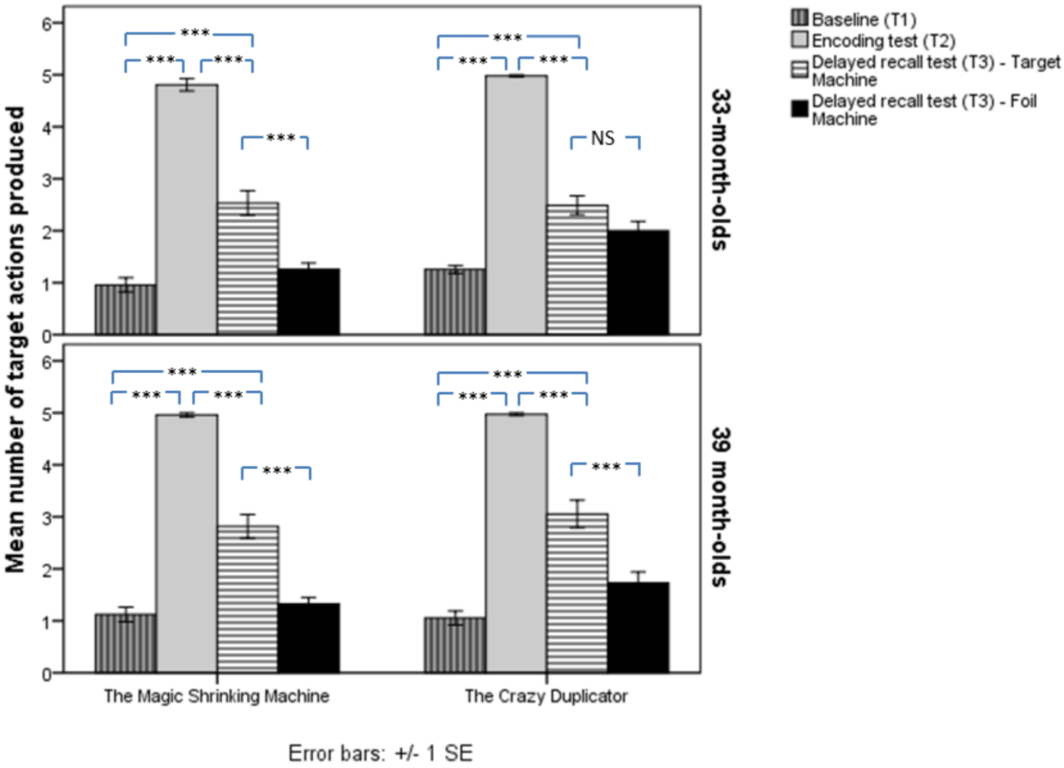


Figure 1. A schematic representation of the employed a mixed 2 (*Between*-subjects: MSM vs. CDC) x 2 (*Within*-subjects: MSM vs. CDC) design allowing for the to-be-remembered props to be present at test while controlling for potential online reasoning.

Figure 2



*** p < .001

Figure 2. Graphical presentation of the mean number of different target actions at $T_{1\text{Baseline}}$, $T_{2\text{Encoding}}$, $T_{3\text{Recall test-Target}}$, $T_{3\text{Recall test- Foil}}$ for both conditions and age groups.

Appendix A

The wordlists related to the target machines (in alphabetical order)

Simcock and Hayne's (2002, 2003) wordlist	Our MSM-wordlist	Our CDC-wordlist
Around	Bag*	Basket
Bag	Box	Belt
Ball	Can	Camera
Bell	Crayon	Close
Big	Door*	Club
Door	Giraffe	Copy Machine
Handle	Handle*	Crazy
Jacket/coat	Light*	Cup
Jar	Little*	Cymbal
Light	Magic Machine*	Door
Little	Open*	Hit
Magic Machine	Play dough	Out
Off	Pull	Push
On	Put*	Put*
Open	Scarf	Ready
Put	Shrink*	Rooster
Raisins	Smaller	Shelf
Shrink	Thing	Sliding door
Teddy bear	T-shirt	Squirrel
Turn	Turn*	Start
Flashlight	Turn off**	Stove
Toys	Turn on**	Toy**

* Identical words used by Simcock and Hayne (2002, 2003).

** Almost identical to words used by Simcock and Hayne (2002, 2003).