Memory

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/pmem20

"I can't skip it": does free report improve accuracy in false memories?

Marina C. Wimmer, Ben Whalley & Timothy J. Hollins

To cite this article: Marina C. Wimmer, Ben Whalley & Timothy J. Hollins (2021) "I can't skip it": does free report improve accuracy in false memories?, Memory, 29:3, 353-361, DOI: 10.1080/09658211.2021.1895223

To link to this article: https://doi.org/10.1080/09658211.2021.1895223

Published online: 12 Mar 2021.

MEMORY
2021, VOL. 29, NO. 3, 353–361
https://doi.org/10.1080/09658211.2021.1895223

"I can't skip it": does free report improve accuracy in false memories?

Marina C. Wimmer[a], Ben Whalley[b] and Timothy J. Hollins[b]

[a]School of Applied Sciences, Psychology, Edinburgh Napier University, Edinburgh, UK; [b]School of Psychology, University of Plymouth, Plymouth, UK

ABSTRACT
Strategic monitoring of recognition memory by children and adults was examined using a semantic DRM procedure. Children (7- and 10-year-olds) and adults (overall $N = 393$) studied lists of semantically related words either incidentally or intentionally and were tested with old items, new items and critical lures to judge as old or new. Participants either made a decision about every item they saw (forced report), or they had the opportunity to withhold answers they were uncertain about (free report). Children were less likely to withhold an answer than adults. However, 7-year-olds were more able to resist false memories when given the opportunity to withhold an answer compared to 10-year-olds or adults. In contrast, adults were unable to improve false memory accuracy. These data suggest that once semantically induced false memories have been encoded they are amenable to strategic monitoring at retrieval in children but not adults.

ARTICLE HISTORY
Received 10 March 2020
Accepted 21 February 2021

KEYWORDS
False memories; DRM paradigm; strategic monitoring; withholding answers

Remembering well can be assessed by measuring how much is remembered (*quantity),* or by looking at the proportion of output that is correct (*accuracy*). These two measures often dissociate. For instance, developmental research shows a clear memory quantity increase with age (Ghetti & Lee, 2011), but a mixed pattern for memory accuracy (Brainerd et al., 2008; Koriat et al., 2001; Krebs & Roebers, 2010) and even evidence of a developmental decline in memory accuracy for semantically related non-presented information ("critical lures" in so-called Deese-Roediger-McDermott (DRM) materials). That is, although adults and older children (10-year-olds) remember more than younger children (7- and 5-year-olds), they are also more susceptible to false memories (Brainerd et al., 2008).

Central to the understanding of memory accuracy is the concept of memory control. In order to increase memory accuracy, it is necessary to excise control over memory decisions about what to report, or how to report it (Koriat & Goldsmith, 1996). There is good evidence that adults and children older than 6 years can strategically control their memory accuracy by withholding information when unsure, across a range of memory settings (Hollins & Weber, 2017; Koriat et al., 2001; Koriat & Goldsmith, 1994; Krebs & Roebers, 2010; Perfect & Weber, 2012; Roebers et al., 2009). However, developmental change in memory control for DRM lists has yet to be studied and so this is an aim of the present work.

The DRM paradigm (Deese, 1959; Roediger & McDermott, 1995) involves the study of semantically associated word lists (e.g., hot, snow, warm, winter, ice, etc) that relate to a non-presented item (i.e., *cold*, the critical lure). As well as correctly recalling or recognising items from the studied lists (presented items, e.g., *hot*), participants also frequently falsely remember unpresented information such as the critical lure (e.g., *cold*). Participants remember less frequently unpresented information that is semantically related items relating to the individual list (e.g., *water*) than in turn unpresented semantically unrelated items (e.g., *beautiful*) (Brainerd et al., 2008; Brainerd & Reyna, 2005; Gallo, 2010; Howe et al., 2009; Roediger et al., 2001). That is, false memories for critical lures are not a result of a generalised memory deficit but are specific to the semantic association between the critical lure and the presented information. These false memories for critical lures increase from the age of 5 to adulthood (Dewhurst & Robinson, 2004; Howe et al., 2009; Otgaar et al., 2014). An overall false memories increase from the age of 5 onward may appear counterintuitive, but is compatible with theories of memory development that posit increasing gist extraction abilities through childhood (Fuzzy-Trace-Theory, Brainerd & Reyna, 2005) or greater associative activation (Associative-Activation-Theory, Howe et al., 2009). That is, Associative-Activation theory suggests that a single process of association formation in a semantic network via spreading activation between nodes and concepts accounts for true and false memories.

CONTACT Marina C. Wimmer m.wimmer@napier.ac.uk School of Applied Sciences, Psychology, Edinburgh Napier University, Edinburgh EH11 4BN, UK

Children's false memories for critical lures increase with age because association formation becomes more refined and automatic with increasing age (Howe et al., 2009). In contrast, Fuzzy-Trace-Theory posits memories are formed via a verbatim memory trace processing surface features of presented information and a gist trace processing the overall semantic meaning (Brainerd & Reyna, 2005). False memories underlie gist extraction abilities which increase with increasing age (Brainerd et al., 2008). Therefore, in the DRM paradigm, false memories for critical lures increase with increasing age.

A large body of literature has manipulated the study of DRM lists to increase monitoring of the studied material eliminating or attenuating the effect of age increase for critical lures. For instance, when information gives rise to distinctive or verbatim processing, retrieval monitoring strategies such as "recall-to reject" can be employed by older children and adults to avoid errors but there is debate of underlying metacognitive strategy versus automatic retrieval process (Brainerd et al., 2003; Carneiro et al., 2009; Roediger et al., 2001). Forewarning about the false memory effect, having an easily identifiable common theme, presenting distinctive pictures, and repeating lists during encoding have reduced false memories in older children and adults but not in younger children (e.g., Brainerd & Reyna, 2005; Carneiro et al., 2009; Del Prete et al., 2014; Ghetti, 2003; 2008; Tempel et al., 2015). However, these manipulations are implemented at encoding and mostly of implicit nature. It remains unclear whether these manipulations have increased gist extraction in younger children (e.g., Lampinen et al., 2006; Odegard et al., 2008) or activation of association with the critical lure (De Petro et al., 2014) in addition to memory editing versus automatic retrieval (Brainerd & Reyna, 2005; Ghetti, 2003) to avoid errors at retrieval. Consequently, in the current work, we adopt an explicit monitoring manipulation that only applies to the retrieval phase.

Explicit retrieval manipulations such as directed forgetting ("forget the just studied list") reveals that children are able to reduce the occurrence of false memories for critical lures whereas adults are unable to do so (Howe, 2005; Kimball & Bjork, 2002; Knott et al., 2011). Specifically, following a standard DRM study episode, Kimball and Bjork (2002) found that false memories for critical lures were unaffected by the instruction to recall, or forget, the previous list, indicating lack of retrieval monitoring at output. In children, on the other hand, directed forgetting reduces false recall (Howe, 2005; Knott et al., 2011), suggesting they may be able to distinguish true from false memories at retrieval.

However, drawing firm conclusions about retrieval monitoring from this work is difficult for the following reason. The effects of directed forgetting may not uniquely reflect output monitoring. Previous work has argued that list-based directed forgetting effects may reflect shifts in mental context (e.g., Sahakyan & Kelley, 2002) or inhibition (e.g., Bjork, 1989). Therefore, in the present work, we use withholding individual responses as a more direct measure of the ability to monitor and control memory for specific items not yet researched to date.

Explicit retrieval instructions to withhold answers to individual items when uncertain can increase children's memory accuracy in non-misleading contexts such as remembering a particular event rather than semantically related information (Koriat et al., 2001; Krebs & Roebers, 2010; Roebers et al., 2009). Specifically, 7- and, particularly, 11-year-olds' output accuracy was higher when they were free to withhold answers about a previously presented story (*free-report*), relative to the case when an answer is required (*forced-report*: Koriat et al., 2001). Likewise, 8–12-year-olds can increase their accuracy by withholding difficult test answers they feel uncertain about, (Krebs & Roebers, 2010; Roebers et al., 2009). Additionally, children aged 6 and above show higher accuracy for the items they would want their memory to be judged on ("golden items") (Hembacher & Ghetti, 2013). Thus, children from at least 6 years onwards can explicitly use monitoring strategies at retrieval to increase their memory accuracy in many different contexts. However, it remains unclear whether or not this extends to semantically induce false memories for critical lures in the DRM paradigm.

The specificity of the DRM procedure is that it differentiates between memories for semantically presented information (presented items, e.g., hot), semantically related non-presented information (critical lures, cold), semantically related list items that were non-presented (related distractors, e.g., water) and semantically unrelated information (unrelated distractors, e.g., beautiful). Examining the ability to improve memory accuracy through withholding a response across the different DRM item types will differentiate between an ability to monitor memory accuracy *per se* versus an ability to monitor memory accuracy for semantically induced false memories as in critical lures. That is, if both children and adults have a generalised ability to monitor memory accuracy then given the opportunity to withhold an answer should have similar effects across all item types. In contrast, if children but not adults distinguish true from false memories at retrieval (Howe, 2005; Kimball & Bjork, 2002; Knott et al., 2011) then children but not adults may be able to monitor memory accuracy for critical lures.

The latter pattern would be predicted by developmental theories explaining occurrence of false memories for critical lures. According to AAT (Howe et al., 2009; Wimmer & Howe, 2009) adults should have difficulty in monitoring false memories for critical lures compared to children, due to automatic activation outside of conscious control than in turn semantically related non-presented items that do not repeatedly and directly activate the critical lure. FTT would also predict that adults should have more difficulty to monitor false memories for critical lures than semantically related items compared to children because the former underlie more gist extraction processes for false memories that increase with age and increasingly underlie recollective processes. As a result, adults are not able to accurately monitor them at retrieval (Brainerd et al., 2004; Brainerd & Reyna, 2005).

Additional empirical reason to suspect that adults may not be able to monitor false memories for critical lures at retrieval is that encoding manipulations of processing levels have little effect on adult's false memories as opposed to children's. Encoding information in standard form (memorise everything) versus shallow form (focus on surface features e.g., font, divided attention) does not affect adults' false memories for critical lures (Dodds & MacLeod, 2004). Dividing attention during encoding reduces children's false memories for critical lures but not adults' (Otgaar, Peters, & Howe, 2012). Moreover, under shallow processing, children's but not adults' false memories are correlated with the overall amount remembered (Wimmer & Howe, 2010), suggesting false memory formation for critical lures may be automatic in adults but not children. If so, they should be amenable to accuracy monitoring at retrieval in children but not adults. Indeed, adults' monitoring accuracy at retrieval for memory of words and faces is unaffected by shallow processing (Arnold & Prike, 2015). However, shallow processing reduces the willingness to report an answer when given the opportunity to withhold them (Arnold & Prike, 2015). Thus, shallow processing affects adults' amount of withholding answers but not the ability to monitor accuracy for presented information, suggesting these processes are dissociated. It is unclear whether these processes are dissociated to begin with. Therefore, a secondary aim of the current research was to manipulate processing level at encoding to investigate its effect on accuracy monitoring at retrieval in both children and adults.

To address this, the current research examined the effects of retrieval monitoring strategies (skip an answer when unsure) in the DRM paradigm. We compared children at the age when monitoring develops (around 7 years) with both older children who are assumed to have developed monitoring abilities (10-year-olds; Ghetti, 2003; 2008; Koriat et al., 2001), and with adults.

We also explored whether strategic output monitoring is susceptible to level of processing manipulations (shallow versus standard) at encoding. It is expected that shallow processing increases adults' willingness to report an answer, but does not affect the accuracy of strategic monitoring (Arnold & Prike, 2015). If the two processes are dissociated to begin with, we should also find the same pattern in children. That is, shallow processing should affect response bias in terms of how much information is volunteered without affecting the accuracy of strategic monitoring.

Experiment

To manipulate the level of encoding, participants were either instructed to memorise the presented information (standard encoding) or instructed to press the colour button the word is written in without receiving memory instructions (shallow encoding). At retrieval, half of the participants were given the opportunity to skip an answer when unsure (free reporting) whereas the other half was required to provide an answer (forced reporting).

Method

Participants

Overall 393 participants (223 females) took part: 148 (75 females) 7-year-olds ($M = 7;8$, $SD = 8$ months), 147 (72 females) 10-year-olds ($M = 10;2$, $SD = 9$ months), and 98 (76 females) adults ($M = 24$, $SD = 11$ years). Children were recruited from local primary schools with a varied socio-economic intake and took part following parental consent and their own assent on the testing day. This was a convenience sample, with sample size limited by the number of children to take part in local schools which had parental consent and gave their own assent to participate in the research. Adults were recruited via a university participation pool and were seen in the university laboratory receiving course credit.

Design

A 3(age group: 7-year-olds vs. 10-year-olds vs. adults) × 2 (condition: standard vs. shallow processing) × 2(reporting: forced vs. free) × 4(item type: critical lures vs. presented items vs. related distractors vs. unrelated distractors) design was used where all variables were manipulated between participants apart from item type which was the within participant variable.

Materials and procedure

Children were tested individually in a quiet area outside their classroom and received verbal instructions from the experimenter that were also displayed on the computer screen. Adults were tested in groups of six, on individual computers and followed the instructions displayed on their screen wearing headphones. The experimental session consisted of three phases: study, distractor, and recognition phase.

The six, 15-item DRM lists (bread, cold, doctor, fruit, lion, sleep) from Roediger et al. (2001) were presented visually and orally at a 2-s rate on a computer screen, one word at a time, and each word was written in colour (either red, green, or blue) (see Appendix). The 15 items were presented in descending backwards associative strength (BAS = the probability that a list item activates the critical lure) to the critical lure. Lists were chosen according to their high word frequency values using the word frequency norms of British primary school aged children's printed vocabulary (Stuart et al., 1993–1996). After each DRM list, a filler picture of a star appeared for 3-s indicating the end of one list and the beginning of the next.

Participants were randomly allocated to the standard condition ($N = 196$; 72 7-year-olds, 75 10-year-olds, 49 adults) or the shallow condition ($N = 197$; 76 7-year-olds, 72 10-year-olds, 49 adults). In the standard condition, participants were instructed to remember as many words as possible. In the shallow condition, there was no intention to memorise but participants were instructed to press the colour button on the keyboard that corresponded to the colour of the presented item. Red, green, and blue stickers were affixed to the c, b, m keys, respectively. Participants received six practise items that were semantically and phonologically unrelated to the presented DRM lists. When participants reached criterion, that is, all practise trials correct, they proceeded with the DRM lists.

After all six lists have been presented, children received a 3-min "Where's Wally?" distractor task and adults computed mathematical calculations. Finally, they received a 36-item recognition test consisting of 18 presented words (the first three items from each list that were the strongest related to the critical lure, e.g., *hot*), 6 critical lures (the strongest associated non-presented item from each list, e.g., *cold*), 6 related distractors (non-presented items related to presented items, e.g., *water*), and 6 unrelated distractors (non-presented items not related to presented items, e.g., *beautiful*) that were presented orally and visually in random order. All items were presented in black font. Responding was self-paced.

Forced reporting

At the beginning of the recognition phase participants ($N = 206$; 68 7-year-olds, 82 10-year-olds, 56 adults) were instructed to say "yes" if the item was presented before and "no" if it was a new item. For child participants, the experimenter recorded the response whereas adults responded themselves by pressing the "y" and "n" keys.

Free reporting

The other half of the participants ($N = 187$; 80 7-year-olds, 65 10-year-olds, 42 adults) were told that they were playing a memory game where they could win one bonus point for each correct answer made. However, they would lose all their points for an incorrect answer (see Koriat et al., 2001). Therefore, if they were unsure they could "skip" the answer. If they skipped they did not gain or lose any points, so they were told that it was beneficial to skip when they were unsure. Before the recognition phase began, child participants were asked to repeat the instructions. The experimenter repeated instructions until children could correctly explain the rules themselves. During the recognition phase, participants could either respond, "yes", "no" or "skip" and the experimenter pressed the corresponding button on the keyboard for children. At the end of the experiment, all child participants were told that they did not lose any points and were praised for their performance. Adults received all instructions on the screen and recorded their responses themselves via button presses on the keyboard, y, n, or s.

Statistical analysis

Our primary analysis estimated the effect of free (skipping) vs. forced reporting on memory accuracy, and we modelled the proportion of correct responses within each item type using a mixed effects binomial regression. Our model was designed to address three potential challenges in analysing these data: First, average rates of skipping differ between age groups. Because we are interested in the effect of permitting skipping, rather than the rate of skipping *per se*, we accounted for skipping rates in our analyses. This is important because, to take an extreme example, if skipping is zero, then performance cannot improve relative to forced reporting. If skipping is 100%, then accuracy must be worse than in forced reporting. Second, the potential benefit of free reporting (skipping) varies according to the baseline performance of forced reporting: if performance under forced reporting is high then skipping has little potential to improve accuracy. As a result, we examined whether the option to skip improves accuracy, accounting for the baseline rate of forced reporting accuracy.

The model was fit with neutral, but weakly informative priors using rstanarm (Goodrich et al., 2018). Convergence of multiple chains was checked visually, and with reference to the \hat{R} statistic and the number of effective samples (Gelman & Shirley, 2011) for each parameter. Marginal odds ratios (for the probability of a correct response) and intervals (95% highest posterior density) for comparisons of interest were computed based on these models (Kruschke, 2011) using the emmeans package (Lenth, 2019). Equivalent models were used to estimate rates of skipping between age groups, item types, and encoding conditions within the skipping group. All data and analyses for this paper are provided in doi://10.5281/zenodo.3369209.

Results and discussion

Eleven participants (five adults, two 10-year-olds, four 7-year-olds) were excluded from the analyses because they responded "yes" to all items in the recognition test.

Skipping rates

Skipping rates were substantially above zero across all items for all age groups in both encoding conditions (mean number skipped 7.42 [5.89–9.06]) although variability in skipping rates between items and conditions was evident (see Figure 1). More related distractors were skipped than any other type (odds ratio (OR) for related

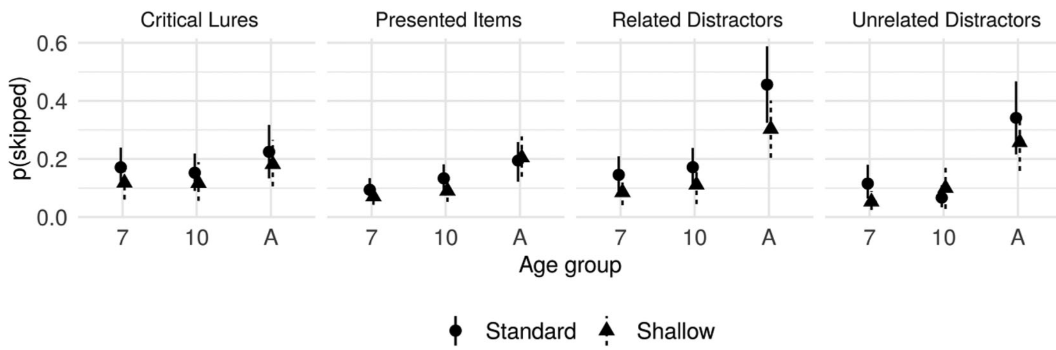

Figure 1. Rates of skipping by age group, item type and condition with 95% confidence interval.

distractors compared with the mean of other types = 1.35 [1.10–1.62]). In contrast, fewer presented items (OR = 0.81 [0.69–0.93]) and unrelated distractors (OR = 0.81 [0.64–0.99]) were skipped.

There was also an effect of age group whereby adults skipped more items than both 7-year-olds (OR = 3.25 [1.79–5.26]) and 10-year-olds (OR = 2.92 [1.54–4.84]). The 7- and 10-year-olds did not substantially differ (OR = 1.12 [0.63–1.74]). Additionally, more items were skipped in the standard than shallow encoding condition (OR = 0.65 [0.39–0.95]).

Effect of reporting on accuracy

To explore the effects of skipping on accuracy, we estimated the benefit of free reporting (skipping) based on the forced choice responding baseline by age groups. We first estimated whether the marginal benefit of free reporting (skipping) vs. forced choice reporting was different in the two encoding conditions (standard versus shallow) for the different age groups, across the different item types. This was not the case (Figure 2): The Bayes factor in favour of the null hypothesis (no effect of or interactions with encoding condition) was very large (BF = 57,624). In other words, there was an equal benefit of skipping on accuracy in the shallow and standard condition, across age groups and item types. Consequently, in our subsequent analyses, we report the results collapsed across encoding conditions. Our data supplement provides additional plots and tables split by encoding condition.

Figure 3(A) indicates that when participants could skip responses then accuracy for presented items increased for both adults and 10-year-olds. In contrast, 7-year-olds did not experience a benefit for presented items. As can be seen in Figure 3(B), both 10-year-olds and adults increased in accuracy relative to 7-year-olds (OR = 1.15 [1.03–1.28] and 1.21 [1.08–1.36]), but adults and 10-year-olds did not differ substantially (OR = 1.05 [0.94–1.17]).

For related and unrelated distracters, all age groups experienced greater accuracy when allowed to skip (OR from 1.09 [1.03–1.15] to OR = 1.42 [1.22–1.67]). Adults gained more benefit than children (OR = 1.23 [1.02–1.45] and OR = 1.27 [1.06–1.51]), but the two groups of children did not differ (OR = 0.96 [0.87–1.07] and OR = 0.97 [0.89–1.05])

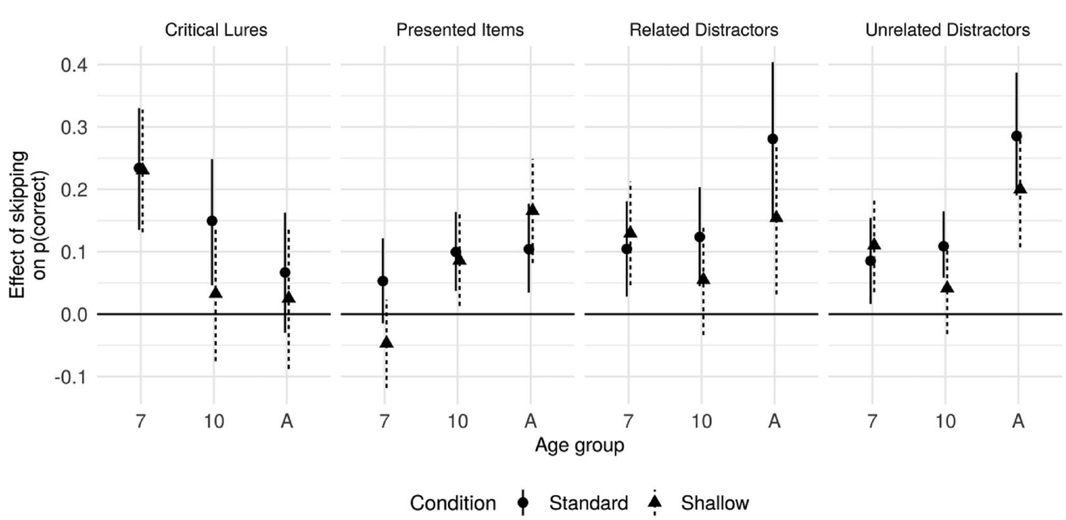

Figure 2. Effect of skipping on accuracy by age group, item type, and condition with 95% confidence interval.

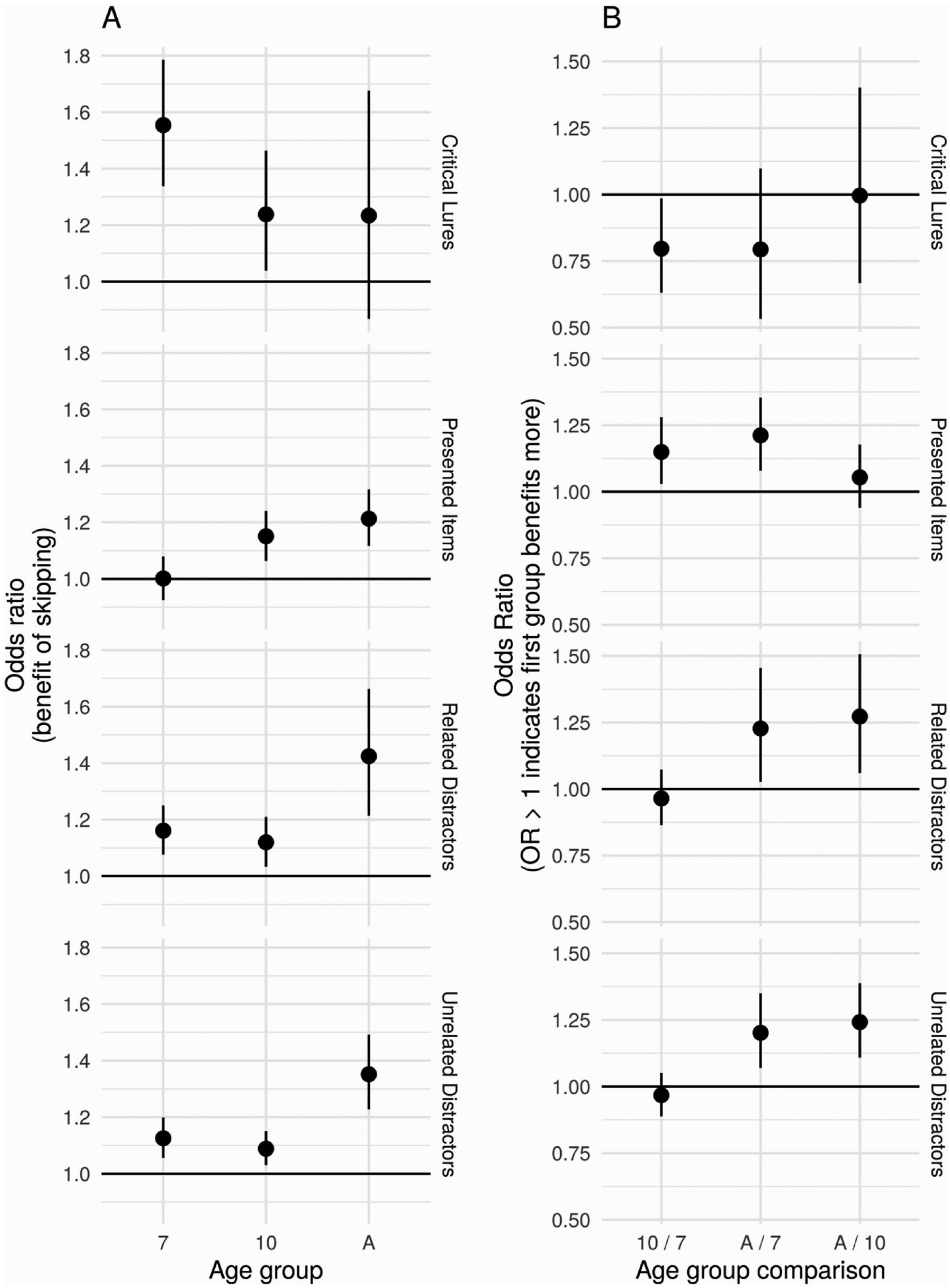

Figure 3. Effect of skipping on response accuracy, within and between age groups. Intervals are 95% highest posterior density. Panel A shows the effect of skipping for each group as an odds ratio, so an OR of 1 means there was no effect. Panel B shows the differences in this effect for each pair of groups. Here an OR of 1 means there is no difference in the effect. A/7 and A/10 mean adults compared with 7-year-olds or 10-year-olds.

For critical lures, the pattern was reversed. When allowed to skip responses both 7-year-olds (OR = 1.56 [1.34–1.79]) and 10-year-olds (OR = 1.24 [1.03–1.46]) showed increased accuracy. Adults' accuracy did increase somewhat, but this value was estimated with less precision and the credible interval for the OR did not exclude zero (OR = 1.23 [0.86–1.68]). Seven-year-olds gained greater benefit in accuracy for critical lures than 10-year-olds when allowed to skip items (0.80 [0.63–0.98]). A similar trend was evident for the comparison between 7-year-olds and adults, but again the credible interval for this difference did not exclude zero (OR = 0.79 [0.54–1.11]). Adults and 10-year-olds did not differ (OR = 1.00 [0.65–1.40]).

Discussion

Our study was the first to examine developmental changes in strategic monitoring at retrieval on accuracy in the DRM paradigm. We found that false memories are reduced (accuracy for critical lures increased) in children when the option to withhold an answer is available. This was the case for both for 7- and 10-year-olds who have developed monitoring abilities (Koriat et al., 2001). In contrast, adults were unable to reduce their rate of false memories when given the opportunity to skip responses. Contrary to the established finding that withholding answers increases accuracy for episodic memory tests (Hollins & Weber, 2017; Koriat & Goldsmith, 1994; Perfect & Weber, 2012) adults failed to do so in a misleading context as in the DRM paradigm. Thus, when false memories are semantically induced adults not only have more false memories than children (Brainerd et al., 2008; Howe et al., 2009) but are also unable to monitor and withhold them at output (current study). Our current DRM findings bear similarity to a pattern found in Koriat and Goldsmith (1996, Experiment 2), where adults showed little benefit in using free report to improve the accuracy for deceptive general knowledge questions. These are questions for which people think they know the answer but are often wrong (such as thinking that Sydney is the capital of Australia). Confidence is high for these types of questions and participants do not skip when they should.

Previous work has shown, using directed forgetting instructions, that children are able to reduce false memories when given an instruction to forget the just studied information, but that this is not true of adults (Howe, 2005; Kimball & Bjork, 2002; Knott et al., 2011). However, directed forgetting studies are list based ("forget all you have just studied") rather than item based, and it is unclear whether list-based directed forgetting reflects output monitoring (Kimball & Bjork, 2002; Knott et al., 2011) or shifts in mental context (e.g., Sahakyan & Kelley, 2002) or inhibition (e.g., Bjork, 1989). In contrast, the current results provide direct evidence for the role of output monitoring: Children, but not adults, could reduce rates of false memories by withholding an answer to individual items. This finding is not a generic effect of lack of output monitoring for non-presented information: we found different patterns of accuracy for related and unrelated distractors. For distractors, all participants were able to increase their accuracy given the opportunity to skip, but adults particularly-so. Thus, the inability of adults to increase accuracy by withholding an answer was unique to semantically induced false memories.

This pattern of results can be accounted for theoretically by either AAT or FTT and does not allow distinguishing between the two. Critical lures are directly semantically related to the list items, whereas related distractors have semantic similarity to the list items but are not directly semantically related to all list items nor have list items backward associate strength (probability to activate) to related distractors (Brainerd et al., 2003; Odegard et al., 2008). Therefore, AAT predicts critical lures not to be amenable to retrieval monitoring as opposed to related distractors particularly in adults whose associative activation formation is automatic compared to children's (Howe et al., 2009). FTT posits that younger children are less able to extract the gist from presented information and because related distractors are not related to all list items nor directly activated, they do not allow gist extraction as opposed to critical lures. As a result, they should be easier to avoid at output (Brainerd et al., 2004; Brainerd & Reyna, 2005; Odegard et al., 2008; see also Lampinen et al., 2006). Support for this notion comes from conjoint recognition paradigms, where children and adults are either asked to recognise presented items only (verbatim), items preserving meaning (gist), or both presented items and items preserving meaning (verbatim + gist) from which familiarity (feeling of oldness) and recollection (remembering defining details about an event) can be derived from the different DRM item types. For both critical lures and related distractors recollection increased between 7 and 14 years (also termed "phantom recollection") (Brainerd et al., 2004). For critical lures, familiarity did not increase between 7 and 14 years whereas it did for related distractors (Brainerd et al., 2004, Experiment 1; see also Odegard et al., 2008). Thus, whereas critical lures are increasingly recollected with increasing age they are not increasingly familiar. Consequently, they are more difficult to monitor at output in adults compared to children. It may also speak to the notion that these types of adults' false memories may be generated during encoding rather than during retrieval (Jou, 2008; Jou & Flores, 2013) and are therefore hard to avoid at output.

For true memories, free reporting did increase accuracy for adults and 10-year-olds consistent with previous episodic tasks that used a non-semantic context (Hollins & Weber, 2017; Koriat et al., 2001; Krebs & Roebers, 2010; Perfect & Weber, 2012; Roebers et al., 2009). In contrast, 7-year-olds did not increase their accuracy for presented items when given the opportunity to skip a response. As they were able to increase their accuracy across all other item types it is not a result of an inability to monitor *per se*. Rather, this finding may reflect that monitoring abilities are still rudimentary across contexts as this is the age range when they develop.

It is important to note that in our version children and adults were given the "highest" possible incentive to skip a response (win all, lose all) and the level of incentive given has impact on memory accuracy as previously shown, that is, higher incentive leading to more memory accuracy (Koriat et al., 2001). Therefore, in our paradigm, motivation to skip and accuracy monitoring are conflated and we are not directly able to disentangle the two. What is crucial though, is that participants had the same incentive to skip an item across all four item types, and yet, there were differences in the ability to monitor them at retrieval. This indicates that our results are not purely an artefact of motivation to skip but do indeed provide insight into accuracy monitoring across age in DRM items at retrieval.

Additionally, there were differences in the proportion of items that was skipped. Both 7- and 10-year-olds were less likely to withhold an answer than adults. It is unlikely that their reduced skipping is due to a lack of understanding the instructions: First, they only proceeded with the recognition phase once they successfully explained their instructions. Second, 4- and 5-year-olds can already make adequate decisions to disclose an answer after giving confidence judgements about their memory (Hembacher & Ghetti, 2014) and are more likely to withhold an answer when uncertain and are more accurate (Lyons & Ghetti, 2013). Moreover, 5-year-olds have metacognitive insight into what constitutes a false memory ("I think he really thought he saw it") and a guess (Jaswal & Dodson, 2009). Third, and most importantly, they were able to withhold answers when given the opportunity to skip to increase their memory accuracy. Thus, even though children and adults differ in how much they withheld, children are able to withhold information at output to increase accuracy. Moreover, adults who skip more are not universally more successful in monitoring as shown by the reverse age pattern for critical lures.

Finally, it is worth noting that the level of encoding (shallow versus standard) affected the probability of information being skipped, that is, more skipping took place in the standard condition, whilst having no effect on accuracy. This is in line with previous research with adults showing that encoding level affects adults' willingness to report an answer but leave accuracy unaffected (Arnold & Prike, 2015), suggesting the two are dissociated in both adults and children.

In sum, we report novel findings on children's and adults' metacognitive monitoring of DRM false memories at retrieval. Children, but not adults, show benefits of withholding information to improve memory accuracy for DRM false memories at retrieval. Thus, once semantic false memories have been encoded children but not adults can strategically monitor them at retrieval by withholding an answer.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Arnold, M. M., & Prike, T. (2015). Comparative difficulty and the strategic regulation of accuracy: The impact of test-list context on monitoring and meta-cognition. *Acta Psychologica*, *157*, 155–163. https://doi.org/10.1016/j.actpsy.2015.02.018

Bjork, R. A. (1989). Retrieval inhibition as an adaptive mechanism in human memory. In H. L. Roediger, & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honor of endel tulving* (pp. 309–330). Erlbaum.

Brainerd, C. J., Holliday, R. E., & Reyna, V. F. (2004). Behavioral measurement of remembering phenomenologies: So simple a child can do it. *Child Development*, *75*(2), 505–522. https://doi.org/10.1111/j.1467-8624.2004.00689.x

Brainerd, C. J., & Reyna, V. F. (2005). *The science of false memory*. Oxford University Press.

Brainerd, C. J., Reyna, V. F., & Ceci, S. J. (2008). Developmental reversals in false memory: A review of data and theory. *Psychological Bulletin*, *134*(3), 343–382. https://doi.org/10.1037/0033-2909.134.3.343

Brainerd, C. J., Reyna, V. F., Wright, R., & Mojardin, A. H. (2003). Recollection rejection: False-memory editing in children and adults. *Psychological Review*, *110*(4), 762–784. https://doi.org/10.1037/0033-295X.110.4.762

Carneiro, P., Fernandez, A., & Dias, A. R. (2009). The influence of theme identifiability. *Memory & Cognition*, *37*(2), 115–129. https://doi.org/10.3758/MC.37.2.115

Deese, J. (1959). On the prediction of occurrence of certain verbal intrusions in free recall. *Journal of Experimental Psychology*, *58*(1), 17–22. https://doi.org/10.1037/h0046671

Del Prete, F., Mirandola, C., Konishi, M., Cornoldi, C., & Ghetti, S. (2014). Paradoxial effects of warning in the production of children's false memories. *Journal of Cognition and Development*, *15*(1), 94–109. https://doi.org/10.1080/15248372.2012.721036

Dewhurst, S. A., & Robinson, C. A. (2004). False memories in children: Evidence for a shift from phonological to semantic associations. *Psychological Science*, *15*(11), 782–786. https://doi.org/10.1111/j.0956-7976.2004.00756.x

Dodd, M D, & Macleod, C M. (2004). False recognition without intentional learning. *Psychonomic Bulletin and Review*, *11*, 137–142.

Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Memory & Cognition*, *37*(7), 833–848. https://doi.org/10.3758/MC.38.7.833

Gelman, A., & Shirley, K. (2011). Inference from simulations and monitoring convergence. In S. P. Brooks, A. Gelman, G. Jones, & X.-L. Meng (Eds.), *Handbook of Markov chain Monte Carlo* (pp. 163–174). Chapman Hall.

Ghetti, S. (2003). Memory for nonoccurrences: The role of metacognition. *Journal of Memory and Language*, *48*(4), 722–739. https://doi.org/10.1016/S0749-596X(03)00005-6

Ghetti, S. (2008). Rejection of false events in childhood. A metamemory account. *Current Directions in Psychological Science*, *17*(1), 16–20. https://doi.org/10.1111/j.1467-8721.2008.00540.x

Ghetti, S., & Lee, J. (2011). Children's episodic memory. *Wiley Interdisciplinary Reviews*, *2*, 353–448. https://doi.org/10.1002/wcs.114

Goodrich, B., Gabry, J., Ali, I., & Brilleman, S. (2018). rstanarm: Bayesian applied regression modelling via Stan. R Package Version 2.17.4. http://mc-stan.org/

Hembacher, E., & Ghetti, S. (2013). How to bet on a memory: Developmental linkages between subjective recollection and decision making. *Journal of Experimental Child Psychology*, *115*(3), 436–452. https://doi.org/10.1016/j.jecp.2013.03.010

Hembacher, E., & Ghetti, S. (2014). Don't look at my answer: Subjective uncertainty underlies preschoolers' exclusion of their least accurate memories. *Psychological Science*, *9*(9), 1768–1776. https://doi.org/10.1177/0956797614542273

Hollins, T. J., & Weber, N. (2017). Evidence of metacognitive benefit to memory. *Memory*, *25*(3), 317–325. https://doi.org/10.1080/09658211.2016.1171363

Howe, M. L. (2005). Children (but not adults) can inhibit false memories. *Psychological Science*, *16*(12), 927–931. https://doi.org/10.1111/j.1467-9280.2005.01638.x

Howe, M. L., Wimmer, M. C., Gagnon, N., & Plumpton, S. (2009). An associative activation theory of children's and adults' memory illusions. *Journal of Memory and Language*, *60*(2), 229–251. https://doi.org/10.1016/j.jml.2008.10.002

Jaswal, V. K., & Dodson, C. S. (2009). Metamemory development: Understanding the role of similarity in false memories. *Child Development*, *80*(3), 629–635. https://doi.org/10.1111/j.1467-8624.2009.01286.x

Jou, J. (2008). Recall latencies, confidence, and output positions of true and false memories: Implications for recall and metamemory theories. *Journal of Memory and Language*, *58*(4), 1049–1064. https://doi.org/10.1016/j.jml.2007.12.003

Jou, J., & Flores, S. (2013). How are false memories distinguishable from true memories in the Deese-Roediger-McDermott paradigm? A review of the findings. *Psychological Research*, *77*(6), 671–686. https://doi.org/10.1007/s00426-012-0472-6

Kimball, D. R., & Bjork, R. A. (2002). Influences of intentional and unintentional forgetting on false memories. *Journal of Experimental Psychology: General*, *131*(1), 116–130. https://doi.org/10.1037/0096-3445.131.1.116

Knott, L. M., Howe, M. L., Wimmer, M. C., & Dewhurst, S. A. (2011). The development of automatic and controlled inhibitory retrieval processes in true and false recall. *Journal of Experimental Child Psychology*, *109*(1), 91–108. https://doi.org/10.1016/j.jecp.2011.01.001

Koriat, A., & Goldsmith, M. (1994). Memory in naturalistic and laboratory contexts: Distinguishing the accuracy-oriented and quantity-oriented approaches to memory assessment. *Journal of Experimental Psychology: General*, *123*(3), 297–315. https://doi.org/10.1037/0096-3445.123.3.297

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*, *103*(3), 490–517. https://doi.org/10.1037/0033-295X.103.3.490

Koriat, A., Goldsmith, M., Schneider, W., & Nakash-Dura, M. 2001. The credibility of children's testimony: Can children control the accuracy of their memory reports? *Journal of Experimental Child Psychology*, *79*(4), 405–437. https://doi.org/10.1006/jecp.2000.2612

Krebs, S. S., & Roebers, C. M. (2010). Children's strategic regulation, metacognitive monitoring, and control processes during test taking. *Educational Psychology*, *80*, 325–340. https://doi.org/10.1348/000709910X485719

Kruschke, J. K. (2011). Bayesian assessment of null values via parameter estimation and model comparison. *Perspectives on Psychological Science*, *6*(3), 299–312. https://doi.org/10.1177/1745691611406925

Lampinen, J M, Leding, J K, Reed, K B, & Odegard, T N. (2006). Global gist extraction in children and adults. *Memory*, *14*, 952–964.

Lenth, R. (2019). emmeans: Estimated marginal means, aka least-squares means. R Package Version 1.4.0.009996. https://github.com/rvlenth/emmeans.

Lyons, K. E., & Ghetti, S. (2013). I don't want to pick! introspection on uncertainty supports early strategic behaviour. *Child Development*, *84*(2), 726–736. https://doi.org/10.1111/cdev.12004

Odegard, T J, Holliday, R B, & Brainerd, C J. (2008). Attention to global gist processing eliminates age effects in false memories. *Journal of Experimental Child Psychology*, *99*, 96–113.

Otgaar, H., Howe, M. L., Peters, M., Smeets, T., & Moritz, S. (2014). The production of spontaneous false memories across childhood. *Journal of Experimental Child Psychology*, *121*, 28–41. https://doi.org/10.1016/j.jecp.2013.11.019

Otgaar, H, Peters, M, & Howe, M L. (2012). Dividing attention lowers children's but increases adults' false memories. *Journal of Experimental Psychology: Learning, Memory*, *38*, 204–210.

Perfect, T. J., & Weber, N. (2012). How should witnesses regulate the accuracy of their identification decisions: One step forward, two steps back? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(6), 1810–1818. https://doi.org/10.1037/a0028461

Roebers, C. M., Schmid, C., & Roderer, T. (2009). Metacognitive monitoring and control processes involved in primary school children's test performance. *British Journal of Educational Psychology*, *79*(4), 749–767. https://doi.org/10.1348/978185409X429842

Roediger, H. L., & McDermott, B. K. (1995). Creating false memories: Remembering words not presented on lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 803–814. https://doi.org/10.1037/0278-7393.21.4.803

Roediger, H. L., Watson, J. M., McDermott, K. B., & Gallo, A. D. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review*, *8*(3), 385–407. https://doi.org/10.3758/BF03196177

Sahakyan, L., & Kelley, C. M. (2002). A contextual change account of the directed forgetting effect. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *28*(6), 1064–1072. https://doi.org/10.1037/0278-7393.28.6.1064

Stuart, M., Masterson, J., Dixon, M., & Quinlan, P. (1993–1996). *Children's Printed Word Database*. http://www.essex.ac.uk/psychology/cpwd/

Tempel, T., Frings, C., & Mecklenbräuker, S. (2015). Dual processes of false recognition in kindergarten children and elementary school pupils. *Journal of Experimental Child Psychology*, *138*, 135–142. https://doi.org/10.1016/j.jecp.2015.05.004

Wimmer, M. C., & Howe, M. L. (2009). The development of automatic associative processes and children's false memories. *Journal of Experimental Child Psychology*, *104*(4), 447–465. https://doi.org/10.1016/j.jecp.2009.07.006

Wimmer, M. C., & Howe, M. L. (2010). Are children's memory illusions created differently from those of adults? Evidence from levels-of-processing and divided attention paradigms. *Journal of Experimental Child Psychology*, *107*(1), 31–49. https://doi.org/10.1016/j.jecp.2010.03.003

Appendix

All DRM study lists/items were obtained from Roediger et al. (2001) and were chosen according to their high backward associative strength (BAS) to the critical lure and their high word frequency values using the word frequency norms of British primary school aged children's printed vocabulary (Stuart et al., 1993–1996).

Critical lure: **Bread** Critical lure: **Doctor** Critical lure: **Lion**
1. Rye 1. Physician 1. Roar
2. Loaf 2. Nurse 3. Tiger
4. Toast 4. Surgeon 4. Mane
5. Dough 5. Patient 5. Fierce
6. Crust 6. Clinic 6. Den
7. Flour 7. Dentist 7. Cub
8. Sandwich 8. Medicine 8. Cage
9. Jam 9. Lawyer 9. Jungle
10. Jelly 10. Health 10. Bears
11. Slice 11. Sick 11. Pride
12. Milk 12. Cure 12. Africa
13. Food 13. Hospital 13. Circus
14. Eat 14. Office 14. Hunt
15. Wine 15. Ill 15. Feline

Critical lure: **Sleep** Critical lure: **Fruit** Critical lure: **Hot**
1. Nap 1. Kiwi 1. Hot
2. Doze 2. Citrus 2. Shiver
3. Bed 3. Pear 3. Arctic
4. Awake 4. Berry 4. Frigid
5. Drowsy 5. Vegetable 5. Freeze
6. Snooze 6. Banana 6. Chilly
7. Slumber 7. Orange 7. Frost
8. Tired 8. Cherry 8. Warm
9. Rest 9. Apple 9. Ice
10. Snore 10. Ripe 10. Winter
11. Wake 11. Basket 11. Snow
12. Dream 12. Juice 12. Heat
13. Yawn 13. Bowl 13. Wet
14. Blanket 14. Salad 14. Weather
15. Peace 15. Cocktail 15. Air

Recognition Words:

Critical lures:
Bread, Cold, Doctor, Fruit, Lion, Sleep,

Presented items:
Rye, Dough, Milk, Hot, Freeze, Heat, Physician, Patient, Cure, Kiwi, Vegetable, Juice, Roar, Fierce, Africa, Nap, Drowsy, Dream,

Related distractors:
Meat, Water, Help, Nut, Safari, Relax,

Non-related distractors
Sport, Beautiful, Black, Car, High, King