

## The Impact of Testing on the Formation of Children's and Adults' False Memories

NATHALIE BRACKMANN<sup>1,2,\*</sup>, HENRY OTGAAR<sup>1</sup>, MELANIE SAUERLAND<sup>1</sup> and MARK L. HOWE<sup>1,3</sup>

<sup>1</sup>Section Forensic Psychology, Department of Clinical Psychological Science, Faculty of Psychology and Neuroscience, Maastricht University, the Netherlands

<sup>2</sup>Department of Psychology, Gothenburg University, Sweden

<sup>3</sup>Department of Psychology, City University London, UK

*Summary:* Witnesses are frequently questioned immediately following a crime. The effects of such testing on false recall are inconclusive: Testing may inoculate against subsequent misinformation or enhance false memory formation. We examined whether different types of processing can account for these discrepancies. Drawing from Fuzzy-trace and Associative-activation theories, immediate questions that trigger the processing of the global understanding of the event can heighten false memory rates. However, questions that trigger the processing of specific details can inoculate memories against subsequent misinformation. These effects were hypothesized to be more pronounced in children than in adults. Seven/eight-, 11/12-, 14/15-year-olds, and adults ( $N = 220$ ) saw a mock-theft film and were tested immediately with meaning or item-specific questions. Test results on the succeeding day replicated classic misinformation and testing effects, although our processing hypothesis was not supported. Only adults who received meaning questions benefited from immediate testing and, across all ages, testing led to retrieval-enhanced suggestibility. © 2016 The Authors. Applied Cognitive Psychology Published by John Wiley & Sons, Ltd.

Witnesses to a crime frequently recall their experiences numerous times. They may immediately talk about the event to other witnesses or to the emergency staff arriving at the crime scene, may be interviewed by investigators later, and may be asked to tell their story in court several years later. To assess the reliability of witnesses' recollections, it is essential to examine whether such repeated interviews strengthen or distort memory traces. The current experiment addressed the issue of immediate testing and its influence on the development of erroneous recollections.

What we know about repeated interviewing is based largely on empirical studies that investigate how testing affects long-term memory performance. Immediate testing has been found to enhance memory retrieval on subsequent memory tasks compared to mere repeated learning of the to-be-remembered material (see Roediger & Karpicke, 2006, for a review). For example, children who answered multiple-choice and free recall questions after a lesson achieved higher correct recall scores on a later test than children who used the same time to restudy the material (Roediger & Butler, 2011). Furthermore, in the field of eyewitness interviewing, the immediate completion of the structured Self-Administered Interview (SAI) yielded more correct information than free recall or no initial retrieval on an interview one week later (Hope, Gabbert, Fisher, & Jamieson, 2014; Krix, Sauerland, Gabbert, & Hope, 2014). Thus, immediate testing seems to enhance the retention of correct information, and, consequently, yields a beneficial *testing effect*.

Intriguingly, the impact of testing on incorrect (as opposed to correct) recall has only recently become a subject of interest, and the results so far are inconclusive (e.g., LaPaglia, 2013). Recollections may be tainted by memory distortion

processes that lead to erroneous statements. Much of the research on such *false memories*<sup>1</sup> has concentrated on the formation of memory illusions triggered by the presentation of erroneous or misleading information (also referred to as misinformation) without directly examining the effect of testing as a moderating variable. An increase in incorrect recall or recognition for items that were misled compared to non-misled control items has been dubbed the *misinformation effect*. A standard procedure used to induce such erroneous recollections is the misinformation paradigm. Here, participants first see a stimulus event (e.g., a video of an armed robbery; Van Bergen, Horselenberg, Merckelbach, Jelicic, & Beckers, 2010), are then misinformed about items depicted in the event (e.g., an incorrect number of people present in the video), and are subsequently asked what they remember (see Loftus, 2005, for a review). Furthermore, the misinformation effect is more pronounced in younger children compared to older children and adults (see Bruck & Ceci, 1999, for a review).

Memory research has suggested that testing might not only positively affect correct recall but also protect against the production of false memories. Researchers argue that taking a test strengthens the memory trace for a specific event, which in turn reduces the impact of forgetting and external misinformation (e.g., Chan, Thomas, & Bulevich, 2009). In support of the protective value of testing, studies have frequently found an *inoculation effect* against misinformation (e.g., Pansky & Tenenboim, 2011; Wang, Paterson, & Kemp, 2014). In one of these studies, participants filled out the SAI immediately after seeing a video of a robbery (Gabbert, Hope, Fisher, & Jamieson, 2012, Experiment 1). One week later, misinformation was introduced via a mock report on a news homepage. On a final free-recall test, the SAI group demonstrated significantly higher correct recall rates and were less affected by the negative effects of

\* Correspondence to: Nathalie Brackmann, Section Forensic Psychology, Department of Clinical Psychological Science, Faculty of Psychology and Neuroscience, Maastricht University, PO Box 616, 6200 MD, Maastricht, the Netherlands.  
E-mail: n.brackmann@maastrichtuniversity.nl

<sup>1</sup> Please note that we use the term *false memories* to refer to false recollections of entire events as well as recollections of erroneous details.

misinformation compared to a control group that did not fill out the SAI prior to the misinformation.

Although these findings seem to show that testing may guard against the production of false memories, other studies have challenged this conclusion. That is, recent studies have demonstrated elevated false memory rates after taking a test (e.g., Chan & Langley, 2011; Chan & LaPaglia, 2011; Gordon & Thomas, 2014; Wilford, Chan, & Tuhn, 2014). This effect, labeled *retrieval-enhanced suggestibility* (Chan *et al.*, 2009), may be related to two underlying mechanisms: a focus on the misinformed items or the overwriting of the original memory trace because of testing. The former mechanism represents a preference for the newly stored (erroneous) information, while the latter causes the previously stored (correct) information to become inaccessible. Both processes disturb the reconsolidation of the original event and lead to higher false memory rates following the introduction of misinformation (Chan & Langley, 2011). In one of the studies that demonstrated retrieval-enhanced suggestibility, even the use of the Cognitive Interview (CI) increased the susceptibility to misinformation considerably (LaPaglia, Wilford, Rivard, Chan, & Fisher, 2014). Specifically, LaPaglia and colleagues' participants completed either the CI, a free-recall test, or a distractor task after seeing the video of a theft. After all participants had seen a distractor video, they listened to a summary of the theft that included misinformation. Correct recall rates were highest in the CI group, but so too were false memory rates. The authors suggest that the initial retrieval of information using the CI might have heightened participants' awareness of memory strengths and weaknesses. While listening to the post-event summary, this awareness has supposedly led to a focus on the conflicting information or information that filled in knowledge gaps and therefore promoted the acceptance of erroneous post-event information. These results are important because the CI is a generally accepted structured interview instrument that enhances correct recall without undermining overall accuracy (Memon, Meissner, & Fraser, 2010). Although it is known that the CI is accompanied by small increases in incorrect recall (Memon *et al.*, 2010), its aggravating effect on the proneness to subsequent misinformation is new.

One way to resolve these contrasting findings (inoculation vs. retrieval-enhanced suggestibility) is to consider the methodological details in those studies. These include delay between the event and testing (Pansky, 2012) or the introduction of misinformation (Chan & LaPaglia, 2011), recall type (cued recall vs. free recall; Wang *et al.*, 2014), form of misinformation (questions vs. narratives; LaPaglia & Chan, 2013), and contextual embedding of misinformation (LaPaglia, 2013). One factor that has rarely been examined is how information (i.e., the type of processing) is retrieved in the testing phase (but see Pansky & Tenenboim, 2011).

According to *Fuzzy-Trace Theory* (FTT, Brainerd, Reyna, & Ceci, 2008) and *Associative-Activation Theory* (AAT, Howe, Wimmer, Gagnon, & Plumpton, 2009; Otgaar, Howe, Peters, Smeets, & Moritz, 2014), different types of processing are directly linked to the formation of false memories. FTT postulates that when witnessing an event, at least two memory traces are stored in parallel. One trace captures the

underlying meaning and understanding of an event, the *gist* information, and the other trace stores item-specific details, the *verbatim* information. For instance, when witnessing a robbery, the gist trace captures that a weapon was involved while the verbatim trace stores what kind of weapon (a black object out of which bullets were shot, accompanied with noise). According to FTT, false memories are caused by a reliance on gist traces when retrieval of verbatim traces is not possible. Similarly, in AAT, both item-specific and relational (meaning) information is stored following an event, although both are stored in a single integrated memory trace. AAT links the production of false memories to an automatic activation of neighboring information in one's knowledge base. Thus, being confronted with a pistol simultaneously activates knowledge about other weapons. If relational information swamps item-specific information in memory and during retrieval, then memory illusions can prevail on tests of recall and recognition. According to both theories, the witness might erroneously infer that a knife was the weapon used in the robbery. Although earlier research has mainly focused on the formation of spontaneous false memories (see Brainerd *et al.*, 2008 for an overview), newer studies also link AAT and FTT to misinformation induced false memories (Otgaar, Howe, Brackmann, & Smeets, 2016).

Pansky and Tenenboim (2011) reasoned that only immediate testing on the item-specific (verbatim) and not on the relational or meaning (gist) level protects against misinformation while simultaneously heightening correct recall. In their study, adult participants first viewed a slide show depicting a normal day of a student. Subsequently, half of the participants had to answer one cued question that needed a one-word response (e.g., What did Inbal drink at the neighborhood pub?), while the other half of participants received a follow-up question to further specify the previous answer in a two-word response (i.e., What kind of wine?). The one-word response was supposed to reflect meaning processing and the two-word response item-specific processing.<sup>2</sup> After a delay of two days, participants received misinformation embedded in questions and then completed a final cued recall test on the item-specific processing level. False memory rates for the misinformed items that were tested with the two-word response approach were comparable to those obtained for the non-misinformed items. For the misinformed items that were tested with the one-word response approach, the false memory rates were significantly higher than for the non-misinformed items. Accordingly, only immediate item-specific questioning (two-word response approach) and not meaning questioning (one-word response approach) yielded an inoculation effect against misinformation.

Drawing from both practical and theoretical considerations, we extended Pansky and Tenenboim's (2011) approach in the current experiment. More specifically, we examined the effect of meaning versus item-specific testing on the susceptibility to misinformation from a developmental angle.

<sup>2</sup> Note that Pansky and Tenenboim (2011) used the terms verbatim and gist level to differentiate between their groups. To be consistent throughout the document, we refer to item-specific and meaning level, respectively. This is in line with the definition put forward by Hunt and Einstein (1981).

Children and adolescents may witness a crime and, like adults, be confronted with immediate questions about the witnessed event. Moreover, both AAT and FTT predict developmental changes in false memory rates. In FTT (see Brainerd & Reyna, 2004, for a review), the ability to extract the underlying meaning of an event (gist) develops more slowly with age compared to the ability to capture the item-specific (verbatim) details of an experience. In AAT (see Howe et al., 2009; Otgaar et al., 2014), children make fewer meaning-based inferences because their knowledge base (e.g., associative networks, script knowledge) is less developed than that of older children or adults. Thus, younger children have a less developed understanding of typical connections than adults and false memories that are because of reliance on understanding the meaning (gist) of an event increase with age.

Because of the importance of these developmental changes in memory illusions, we included children, adolescents, and adults in the current study. We triggered meaning or item-specific processing via immediate testing on several items after participants saw a video of a theft. Although all participants answered the same general questions, these questions were either followed by a question focussing on the underlying meaning (meaning group) or by a question focussing on specific details (item-specific group). On a subsequent testing day, we introduced misinformation for some items using a fictitious eyewitness account that summarized the video. Finally, all participants answered memory questions to assess memory for the original event.

We expected to replicate the classic misinformation effect with higher false memory rates for misinformed than for non-misinformed items and that this effect would be more pronounced in children than in adults. Drawing from findings concerning test effects in the absence of misinformation, we expected better memory performance for items that were tested compared to those that were untested. This effect should be present across all age groups and for both types of processing (meaning vs. item-specific). Given that findings of the inoculation effect stand in stark contrast to the findings of retrieval-enhanced suggestibility, the anticipated effect of immediate testing for misinformation items was non-directional and analyses were conducted in a purely exploratory manner.

Further, we hypothesized that meaning versus item-specific processing might account for the divergent findings in the extant literature such that participants who received meaning questions should be more vulnerable to subsequent misinformation than participants who received item-specific questions. Thus, retrieval-enhanced suggestibility is more likely to occur after immediate questioning at the meaning level while an inoculation effect is more likely after immediate questioning on an item-specific level. Although the ability to extract meaning and item-specific information undergoes considerable developmental improvements, the former development is more protracted than the latter (Bjorklund, 1987; Brainerd et al., 2008; Howe et al., 2009). Thus, we hypothesized that whereas item-specific testing should be beneficial for all ages (inoculation effect), meaning testing should affect younger children more than older children, adolescents, and adults increasing their

retrieval-enhanced suggestibility. This is based on both AAT and FTT predicting that false memories are more likely to occur when meaning-based information is relied upon instead of item-specific information. Furthermore, according to the tenets of AAT and FTT, an age-related increase exists in the ability to spontaneously extract the meaning (gist) information of an event. Forced processing at the meaning level through testing may enable younger children to better extract relational information and this may have the negative consequence of heightening false memory rates.

## METHOD

### Participants

Two-hundred-and-twenty participants took part in the experiment. Specifically, 59 7/8-year-olds ( $M_{age}=7.78$ ,  $SD=0.62$ ), 57 11/12-year-olds ( $M_{age}=11.94$ ,  $SD=0.84$ ), 53 14/15-year-olds ( $M_{age}=14.96$ ,  $SD=0.62$ ), and 51 adults ( $M_{age}=22.32$ ,  $SD=1.99$ ) were tested in their native language. Child and adolescent participants were recruited from elementary schools and high schools in the Netherlands and Germany. Consent of the school principals and parents were obtained prior to testing. Adult participants were university students, who received course credit or a 7.50€ voucher for their participation and gave their informed consent prior to participation. The study was approved by the standing ethical committee of the Faculty.

### Materials

#### *Video*

The 6:36-min version A of the video developed by Takarangi, Parker, and Garry (2006), which depicts a theft, was used as the stimulus film. It shows an electrician wandering around a house repairing electrical objects and using or stealing belongings from the owner. The video includes eight critical items which have been shown to reliably elicit a misinformation effect when erroneous information was introduced after the video.

#### *Immediate questions*

The immediate testing phase consisted of paired questions on nine items shown in the film. All participants answered the same general *what*-questions (e.g., 'What did Eric wear?'). The follow-up questions were either on a meaning or item-specific level. To elicit a meaning response, participants received a subsequent question asking for what the item mentioned in the first question is used for ('Why do people wear trousers?'; see Appendix A online<sup>3</sup> for a complete list of questions for the meaning testing group). Alternatively, to elicit an item-specific response, participants received a subsequent question asking *what kind of* item was mentioned in the first question ('What kind of trousers did Eric wear?'; see Appendix B online for a complete list of questions for the item-specific testing group). A two-alternative forced choice answer format was offered for the general *what*-questions as well as the follow-up

<sup>3</sup> All appendices can be viewed online in the Supporting Information.

questions to ensure processing comparability across conditions. Therefore, all participants received the same questions and answers for the general question and a follow-up question that was tailored to the previously given answer. Whether option *a* or *b* was correct was selected randomly and was the same for every participant (see Appendix C online for the answer key for the immediate questions). The order of the paired questions was randomized per participant.

The questions were developed based on the questionnaire used by Takarangi *et al.* (2006). All participants received the same general questions. A two-alternative answer format was chosen to avoid floor-effects and to ensure processing on the group-specific level. We used a theory-driven approach that relied on assumptions derived from AAT and FTT to construct the group-specific follow-up questions. To assess the semantic knowledge and to trigger processing on the meaning level, we added a follow-up question that requested a meaning-based explanation to the first answer. The item-specific questions were posed in a comparable manner to the study of Pansky and Tenenboim (2011) in the sense that the subsequent follow-up question prompted a more specific answer to the first question. To assess concrete event-specific knowledge and to trigger processing on the item-specific level, we added a follow-up question that requested an explanation in more detail. All participants received paired questions—the same general question and a follow-up question on either a meaning or an item-specific level—independent of group allocation.

### Misinformation

A 3:55-min eyewitness account of a female person describing the thief's activities was played aloud to introduce misinformation. The written narrative used by Takarangi *et al.*

(2006) was translated into German and Dutch and minor details were adjusted to resemble an eyewitness statement. A full verbatim transcript can be obtained from the first author upon request. The auditory summary contained misinformation about eight items (see Appendix C online for a specification of misinformed items). Half of these items had been previously tested in the immediate question phase on either the meaning or item-specific level (see Figure 1 for a visualization of the design and procedure). The column labeled misinformation gives an overview of the composition of the items, which depends on their former reference (column immediate testing and type of processing).

### Final memory test

The final memory test consisted of the same 18 questions for all participants (see Appendix D online). The questions were posed in the more specific *what kind of* item-specific format with two-alternative forced choice answer options. Eight questions referred to the misled items from the eyewitness account (see Figure 1). Four of these *misinformed* items had been previously tested in the immediate testing phase. The other 10 items were neutral and served as control items. Five of these *non-misinformed* items were previously tested in the immediate testing phase. Appendix C (online) gives an overview of the answer key for the final memory test, whether the item has been tested during immediate questions or not, and if it was erroneously depicted within the eyewitness account (misinformation manipulation).

### Design

The experiment used a 4 (Age: 7/8 vs. 11/12 vs. 14/15 years vs. adults)  $\times$  2 (Type of Processing: meaning vs. item-specific testing)  $\times$  2 (Immediate Testing: tested vs. untested)  $\times$  2

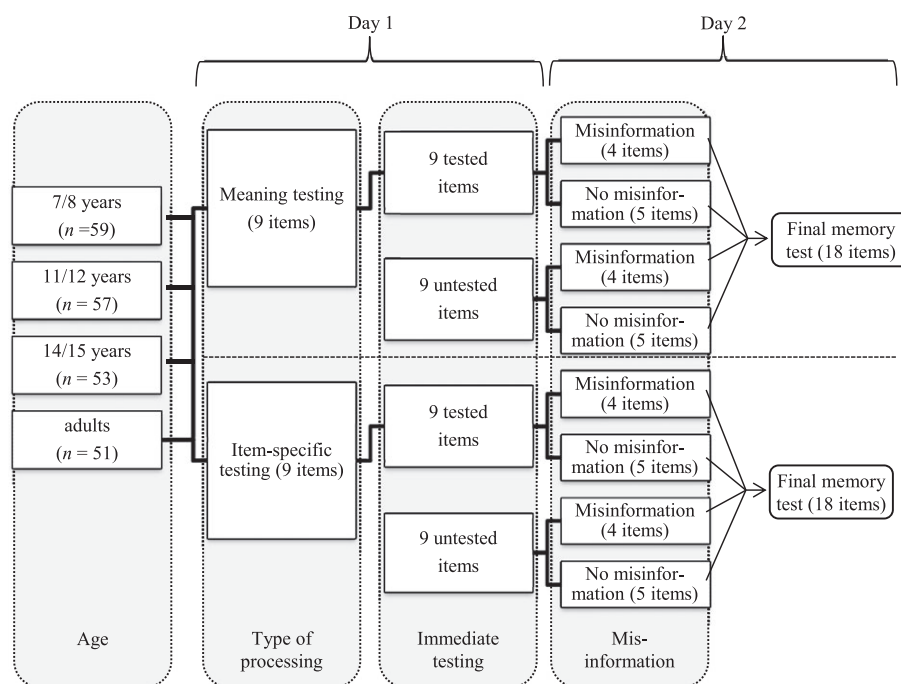


Figure 1. Overview of the design and procedure. Age and type of processing were between-subjects variables. Half of the participants of each age group were tested with meaning questions while the other half received item-specific questions after viewing the target video on day 1. The factors immediate testing and misinformation were manipulated within subjects and refer to the immediate questions on day 1 and the misinformation via the eyewitness account on day 2

Table 1. Mean proportion of incorrect answers in the final memory test as a function of age, type of processing, immediate testing, and misinformation

Age	Type of processing	Tested items				Untested items			
		Misinformation		No misinformation		Misinformation		No misinformation	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
7/8	Meaning	0.43	0.20	0.19	0.16	0.56	0.22	0.39	0.24
	Item	0.59	0.29	0.14	0.18	0.53	0.17	0.37	0.20
11/12	Meaning	0.54	0.23	0.09	0.13	0.54	0.19	0.23	0.20
	Item	0.53	0.31	0.15	0.17	0.50	0.23	0.30	0.17
14/15	Meaning	0.49	0.33	0.03	0.07	0.52	0.20	0.25	0.19
	Item	0.43	0.30	0.04	0.08	0.38	0.20	0.26	0.18
Adults	Meaning	0.20	0.22	0.05	0.12	0.26	0.23	0.23	0.20
	Item	0.20	0.18	0.07	0.11	0.36	0.20	0.18	0.16

Note. Mean scores were calculated by dividing the number of incorrect answers by the total number of items per condition and as such can range between 0 (all items were answered correct) and 1 (all items were answered incorrect).

(Misinformation: misinformation vs. no misinformation) mixed factorial design. Age and type of processing served as between-subjects variables while immediate testing and misinformation were manipulated within-subject.

## PROCEDURE

The experiment took place in quiet testing rooms at the university or elementary/high school on two consecutive days. Individual test sessions lasted approximately 15 min. Random group assignment was accomplished by providing scoring sheets that either affiliated the participants with the meaning or the item-specific group. Experimenters had no influence on the group assignment. Roughly half of the participants of each age group were allocated to the different question groups.

On the first day of testing, participants saw the video of the theft. After a short filler-task (Tetris, approximately 3 min), participants were tested on a meaning or item-specific level. One day later, participants listened to the eyewitness account that contained misinformation. Following a short filler task (Tetris, approximately 3 min), the final memory questions were asked in the same order as piloted by Takarangi et al. (2006). These questions served as a final memory test. They included items that were already questioned on day one as well as new items. On half of these untested and tested items, misinformation was introduced on the second day (see Appendix C online for a detailed description of these items).

## RESULTS

We wanted to examine the effects of different types of testing on the vulnerability to misinformation in different age groups. To this end, we computed a 4 (Age: 7/8 vs. 11/12 vs. 14/15 years vs. adults)  $\times$  2 (Type of Processing: meaning vs. item-specific testing)  $\times$  2 (Immediate Testing: tested vs. untested)  $\times$  2 (Misinformation: misinformation vs. no misinformation) mixed measures ANOVA. Four dependent scores computed from the final memory test served as the outcome variables. These concern the proportion of incorrect answers on the tested and untested

misinformation and no misinformation items which was used as a proxy measure of false memories.<sup>4</sup> Table 1 gives an overview of the descriptives of the final memory test scores.

## Preliminary analyses

To determine whether our outcomes are because of item characteristics rather than our specific manipulations, we examined whether test scores on the first day affected our final test outcomes independent of our misinformation manipulation. To do this, we first performed an analysis on item difficulty to assess whether items that were misinformed on the second day differed from those items that were non-misinformed prior to the introduction of the misleading eyewitness account during immediate testing on day one. To this end, we conducted a paired sample *t*-test to compare the proportion of correct answers to the immediate questions on day one in misinformed and non-misinformed items. There was no statistically significant difference between misinformed ( $M_{correct}=0.71$ ,  $SD=0.19$ ) and non-misinformed items ( $M_{correct}=0.68$ ,  $SD=0.15$ ) prior to the introduction of misinformation,  $t(219)=1.86$ ,  $p=0.064$ ,  $r=0.12$ . Thus, item difficulty during immediate testing on day one did not differ. Second, we performed a linear regression analysis on the item level. This was performed solely for the misinformed items. The combined final memory scores served as dependent variables. These combined variables differentiate between misinformed and non-misinformed items, but not between tested or untested ones. The mean score of misinformed items on day one served as predictor. No significant effect was found,  $F(1, 218)=1.85$ ,  $p=0.176$ ,  $R^2=0.008$ . The score on the misinformed items prior to misinformation on day one did not influence the test score after misinformation on day two. Thus, item difficulty on day one did not account for the final test outcome on day two. Therefore, the misinformation effects we report next were not driven by item difficulty.

<sup>4</sup> Note that slight violations in normality and homoscedasticity were observed. Nevertheless, because of split samples bigger than 20 and a predominance of low scores for the 14/15-year-olds, the model was not found to be invalid.

Table 2. Results of the four-way mixed measures ANOVA with age and type of processing as between-subject factors and immediate testing and misinformation as within-subject factors

	<i>F</i>	<i>df</i>	<i>p</i>	$\eta_p^2$
Age × type of processing × immediate testing × misinformation	2.23	3	0.085	0.03
Age × type of processing × immediate testing	0.79	3	0.500	0.01
Age × type of processing × misinformation	2.98	3	0.032	0.04
Age × immediate testing × misinformation	2.47	3	0.063	0.03
Type of processing × immediate testing × misinformation	0.67	1	0.415	0.00
Age × type of processing	1.22	3	0.305	0.02
Age × immediate testing	1.62	3	0.186	0.02
Age × misinformation	8.80	3	<0.001	0.11
Type of processing × immediate testing	1.57	1	0.211	0.01
Type of processing × misinformation	0.12	1	0.732	0.00
Immediate testing × misinformation	32.03	1	<0.001	0.13
Age	37.78	3	<0.001	0.35
Type of processing	0.05	1	0.825	0.00
Immediate testing	71.29	1	<0.001	0.25
Misinformation	270.77	1	<0.001	0.56

Note. Significant higher-order interactions are discussed in the text using simple interactions and simple main effects.

### Main analyses

The predicted interaction between age, type of processing, immediate testing, and misinformation was not statistically significant,  $F(3, 212) = 2.23$ ,  $p = 0.085$ ,  $\eta_p^2 = 0.03$ . However, there were statistically significant three- and two-way interactions which qualified the main effects (Table 2). Using simple interactions and simple main effects, we specify the interactive influences on the final memory score below.

#### Misinformation effect

The three-way interaction between age, type of processing, and misinformation was statistically significant,  $F(3, 212) = 2.98$ ,  $p = 0.032$ ,  $\eta_p^2 = 0.04$ . All three-way interactions including immediate testing were non-significant. Accordingly, we combined the final memory scores across tested and untested items. Figure 2 shows that the significant Age × Type of processing × Misinformation interaction may stem from differences in vulnerability to misinformation as a factor of age and type of processing. Specifically, the effect of misinformation versus no misinformation appears to be smaller in adults who were tested in the meaning group compared to all other ages and types of processing. To

confirm this impression, we first analyzed the simple Age × Misinformation interaction on each level of type of processing (meaning vs. item-specific testing). The results showed that the simple Age × Misinformation interaction was significant in the meaning testing group,  $F(3, 212) = 9.17$ ,  $p < 0.001$ ,  $\eta^2 = 0.12$ , but not in the item-specific testing group,  $F(3, 212) = 2.32$ ,  $p = 0.077$ ,  $\eta^2 = 0.03$ .

To gain a deeper understanding of the significant Age × Misinformation interaction in the meaning testing group, we subsequently looked at the second-order simple effect of misinformation (misinformation vs. no misinformation) on each level of age (7/8 vs. 11/12 vs. 14/15 years vs. adults) in the meaning testing group. This analysis revealed that the contaminating effect of misinformation differed as a function of age. Although the misinformation effect was comparably high in 11/12-year-olds (misinformation:  $M = 0.54$ ,  $SD = 0.16$ , no misinformation:  $M = 0.16$ ,  $SD = 0.12$ ),  $F(1, 216) = 46.71$ ,  $p < 0.001$ ,  $\eta^2 = 0.15$ , and 14/15-year-olds (misinformation:  $M = 0.51$ ,  $SD = 0.21$ , no misinformation:  $M = 0.22$ ,  $SD = 0.14$ ),  $F(1, 216) = 38.59$ ,  $p < 0.001$ ,  $\eta^2 = 0.12$ , it was smaller in 7/8-year-olds (misinformation:  $M = 0.49$ ,  $SD = 0.17$ , no misinformation:  $M = 0.29$ ,  $SD = 0.16$ ),  $F(1, 216) = 14.61$ ,  $p < 0.001$ ,  $\eta^2 = 0.05$ , and non-significant in

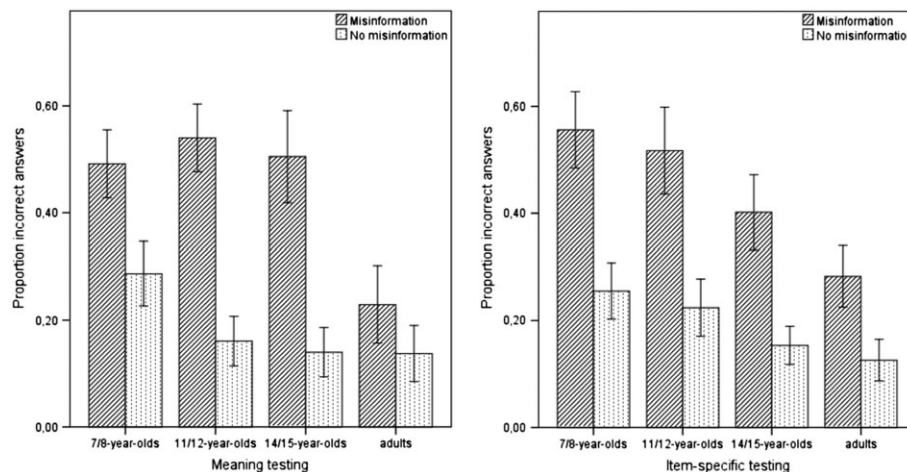


Figure 2. Proportion of incorrect answers as a function of age and misinformation for meaning (left graph) and item-specific testing (right graph). Error bars represent 95% confidence intervals

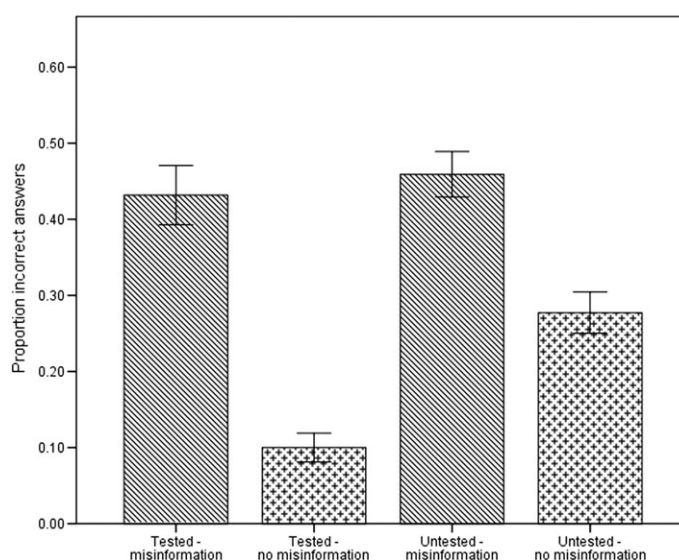


Figure 3. Proportion of incorrect answers as a function of immediate testing and misinformation. Error bars represent 95% confidence intervals

adults (misinformation:  $M=0.23$ ,  $SD=0.17$ , no misinformation:  $M=0.14$ ,  $SD=0.12$ ),  $F(1, 216)=2.34$ ,  $p=0.051$ ,  $\eta^2=0.01$ . Consequently, for the meaning tested sample, we found a misinformation effect in children and adolescents, but adults were not statistically affected by misinformation. Note that adults had low rates of incorrect recognition in misinformed as well as non-misinformed items. Even with a more lenient  $p$ -value criterion, considering  $p=0.051$  as marginally significant, the effect of the difference between these items would still be considered very low  $\eta^2=0.01$ . That is, we only found little support for a misinformation effect in adults. As predicted, the misinformation effect was much stronger in younger than older children and adolescents.

#### Testing effect

Interestingly, we did find a statistically significant two-way interaction in which immediate testing played a role. The two-way interaction between immediate testing and misinformation was significant,  $F(1, 212)=32.03$ ,  $p<0.001$ ,  $\eta_p^2=0.13$ . Simple main effects showed that the misinformation effect was almost double the size after testing (misinformation:  $M=0.43$ ,  $SD=0.29$ ; no misinformation:  $M=0.10$ ,  $SD=0.14$ ;  $F(1, 215)=241.93$ ,  $p<0.001$ ,  $\eta^2=0.50$ ) compared to no testing (misinformation:  $M=0.46$ ,  $SD=0.23$ ; no misinformation:  $M=0.28$ ,  $SD=0.20$ ;  $F(1, 215)=81.64$ ,  $p<0.001$ ,  $\eta^2=0.27$ ).<sup>5</sup> This is illustrated in Figure 3 in which the magnitude of the difference between the first and second bar (difference of 0.33), reflecting the misinformation effect for tested items, is considerably bigger than in the third and fourth bar (difference of 0.18), reflecting the misinformation effect for untested items respectively. Thus, ignoring type of processing and age, immediate testing almost doubled the size of the misinformation effect. Our data therefore support retrieval-enhanced suggestibility. All other main effects and interactions were non-significant.

<sup>5</sup> We thank an anonymous reviewer for suggesting this approach to analyzing our data.

## DISCUSSION

The present study was designed to determine the effect of different types of processing (meaning vs. item-specific) on testing effects in relation to vulnerability to misinformation. Extending previous studies, not only adults, but also children and adolescents were tested immediately after having seen a target event (i.e., theft). On a second testing day, misinformation was introduced using an eyewitness account. We replicated the beneficial effect of immediate testing (Roediger & Karpicke, 2006) and the harmful effect of misinformation (Loftus, 2005) on memory performance. Contrary to our expectation, our processing manipulation in immediate testing did not play a pivotal role in participants' subsequent vulnerability to misinformation. However, immediate testing increased the misinformation effect resulting in a clear retrieval-enhanced suggestibility effect. This is, to our knowledge, the first evidence of a retrieval-enhanced suggestibility effect in children. The results as well as their implications are discussed in more detail in the following paragraphs.

First, we found the predicted testing effect as described by Roediger and Karpicke (2006), which manifested itself in better memory performance for tested than for untested items. Second, in addition to the testing effect, we replicated the classic misinformation effect (e.g., Loftus, 2005). Here, we found higher false memory rates for the items that were misled than for those that were not misled across all ages. Importantly, the misinformation effect was larger for tested items than for untested items, supporting the notion of retrieval-enhanced suggestibility. We are the first to show this effect in a sample that not only included adults, but also children and adolescents. Furthermore, consistent with previous developmental findings (see Bruck & Ceci, 1999, for a review), the harmful effects of misinformation were more pronounced in children than in adults. Adding to previous studies that rarely included an adolescent age group (as pointed out by Jack, Leov, & Zajac, 2013 and by McGuire, London, & Wright, 2015), we showed that the adolescents were nearly as susceptible to misinformation as

the children. This suggests that adolescents' eyewitness reports should also be assessed cautiously and may not be put on the same level as statements by adults. Consequently, for adolescents as well as for children, a detailed analysis of the circumstances under which statements were obtained is important as is whether or not post-event misinformation was possibly involved.

We were most interested in the effect of different types of testing on the vulnerability to misinformation. We predicted that item-specific testing could account for the inoculation effect, while testing meaning might promote test-enhanced suggestibility. We based our assumptions on theoretical considerations and on a previous study investigating the effect of different types of testing in adults (Pansky & Tenenboim, 2011). In an attempt to determine the source of discrepancies with other studies, we conducted exploratory analyses into the level of processing during testing. Because of the age decrease in the vulnerability to misinformation (without previous testing), we expected that the effects of testing would be more pronounced in younger than older children and adults.

Our results showed that the type of processing (meaning vs. item-specific) apparently cannot account for the diverging findings regarding the impact of testing on the vulnerability to misinformation. That is, surprisingly, there were no overall differences as a function of meaning versus item-specific testing. The only exception in which type of processing had an influence was the adult group who was tested on meaning (see the result section on the misinformation effect). Adults tested in the item-specific group, and child and adolescent participants in both processing groups, showed increases in incorrect recognition after misinformation versus no misinformation irrespective of type of processing. However, adults tested in the meaning group showed no higher incorrect recognition rates after misinformation compared to no misinformation. Consequently, only for the adults tested in the meaning group were there no harmful effects of misinformation, providing some support for the inoculation effect.

Pansky and Tenenboim (2011) found that item-specific (verbatim) testing and not meaning (gist) testing yielded an inoculation effect of testing against misinformation in adults. However, we found the opposite pattern, namely, that meaning testing and not item-specific testing yielded some kind of inoculation effect inasmuch as meaning testing in the adult group prevented participants from falling prey to misinformation. This inconsistency may be because of differences in how misinformation was introduced. An inoculation effect of testing against misinformation was more frequently found when misinformation was introduced via questions (LaPaglia & Chan, 2013). Likewise, Pansky and Tenenboim used such an approach whereas we embedded the misinformation into a narrative. This approach has been found to favor test-enhanced suggestibility (LaPaglia & Chan, 2013), a pattern that was supported by our finding. Another possible reason for the inconsistencies with Pansky and Tenenboim is the diverging operationalization of the types of processing. According to Pansky and Tenenboim, item-specific (verbatim) processing, operationalized with a two-word response approach, yielded an inoculation effect

whereas meaning (gist) processing, operationalized with a one-word response approach, had no beneficial influence on false memory rates. An alternative explanation for their lower false memory rates in the two-word response group is that repetitive testing of the critical item occurred, compared to single testing in the one-word response group that only received one question. Likewise, only the two-word response condition focused on item-specific information whereas the one-word response did not necessarily trigger information retrieval at a meaning level as stipulated by AAT or FTT. In the current study, we addressed these limitations by employing a follow-up question not only for the item-specific, but also the meaning group. This question stimulated participants to think about why a depicted action, for example the theft of jewelry, had been performed. If repetitive questioning about an item accounted for the inoculation effect, this effect should have been detectable in the item-specific as well as the meaning testing group. However, this pattern of results was not found in the present experiment.

The question remains as to why we found a small inoculation effect against misinformation in our adult meaning testing group. The inoculation effect is thought to occur because immediate testing strengthens the memory trace for the original event (e.g., Wang *et al.*, 2014). This advantage preserves it from the subsequent, negative influences of misinformation. Apparently, the immediately retrieved information does not need to correspond entirely with the later requested information. Although our immediate meaning question led participants to process the underlying meaning of the details witnessed, our final memory test asked for item-specific information.

Besides the protective effect of meaning testing in adults, we found no effects of testing on the susceptibility to misinformation in other age groups. This may have been because of either no influence of type of processing or a failure in the induction of the manipulation. Even though our operationalization of meaning and item-specific processing was thought to emphasize the effect of testing on the susceptibility to misinformation and to be more forensically relevant, we did not find such an effect regardless of age. We expected immediate meaning testing compared with item-specific testing to elevate suggestibility to misinformation across all ages but, in particular, in younger children. However, no age effect was found. This does give us insight into the effect of different types of testing on subsequent vulnerability to misleading information. Whereas younger children are most prone to misinformation irrespective of type of testing, adults might somewhat benefit from meaning testing.

We tried to match our experiment as much as possible to naturally occurring circumstances during a crime (i.e., witnessing a theft), without losing control over possible extraneous factors. Naturally, because of ethical constraints and the age of our youngest participants, we were not able to show a highly upsetting video. It is known that stress, which is likely experienced during a real crime, can negatively influence memory (Deffenbacher, Bornstein, Penrod, & Kiemann, 2004). However, others argue that the impact might be neutral (e.g., Krix *et al.*, 2016; Sauerland *et al.*, 2016) or



even positive (e.g., Roozendaal & McGaugh, 2011). Thus, future research might also take the stress component into account when trying to disentangle the inoculating effect from test-enhanced suggestibility.

Another point to be considered is the answer format which should be as free as possible to match real-life interview techniques. The present study employed a forced-choice answer format to ensure processing on either the meaning or item-specific level. The advantage of such an answer format is that even very young children understand what is expected from them and no floor-effects emerge caused by difficulties in understanding the instructions. However, no differentiation between guessing and actually knowing can be drawn in a forced-choice answer format. It is unknown whether incorrect recognition rates were based on the formation of a false memory or on the wrong choice when no memory could be retrieved. Thus, we did not base our conclusions on absolute incorrect recognition rates but on differences in incorrect recognition rates between the conditions.

To conclude, this study sheds light on whether immediate interviewing has a beneficial or harmful effect on delayed memory performance. Although it did not provide evidence for the idea that type of processing is a crucial factor concerning inoculation versus retrieval-enhanced suggestibility, it did address questions about the nature of these opposing findings. One implication of our findings is that immediate conversations with other witnesses or bystanders, emergency personnel, or police officers at the crime scene do not necessarily impair memory consolidation processes. On the contrary, immediate testing can have beneficial effects when witnesses are not subsequently influenced by erroneous information. However, avoiding such an influence is nearly impossible. Indeed, our findings support the conclusion that testing may increase the magnitude of the misinformation effect at all ages. Therefore, it is crucial to better understand when immediate testing does (or does not) make participants vulnerable to subsequently incorporating misinformation into their memory of the witnessed (target) event.

## ACKNOWLEDGEMENTS

This paper has been supported by a fellowship from the Erasmus Mundus Joint Doctorate Programme in Legal Psychology (EMJD-LP; 2013-1438/001-001-EMII-EMJD) to Nathalie Brackmann and by a grant from the Edmund Hustinx Foundation and from the Netherlands Organization for Scientific Research (NWO 415-12-003) to Henry Otgaar. The authors acknowledge the help of Mike van Asseldonk, Dominique Nijpels, and Lisa Wilbers in data collection.

## REFERENCES

- Bjorklund, D. F. (1987). How age changes in knowledge base contribute to the development of children's memory: An interpretive review. *Developmental Review*, 7, 93–130. DOI:10.1016/0273-2297(87)90007-4.
- Brainerd, C. J., & Reyna, V. F. (2004). Fuzzy-trace theory and memory development. *Developmental Review*, 24, 396–439. DOI:10.1016/j.dr.2004.08.005.
- Brainerd, C. J., Reyna, V. F., & Ceci, S. J. (2008). Developmental reversals in false memory: A review of data and theory. *Psychological Bulletin*, 134, 343–382. DOI:10.1037/0033-2909.134.3.343.
- Brock, M., & Ceci, S. J. (1999). The suggestibility of children's memory. *Annual Review of Psychology*, 50, 419–439. DOI:10.1146/annurev.psych.50.1.419.
- Chan, J. C. K., & Langley, M. M. (2011). Paradoxical effects of testing: Retrieval enhances both accurate recall and suggestibility in eyewitnesses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37, 248–255. DOI:10.1037/a0021204.
- Chan, J. C. K., & LaPaglia, J. A. (2011). The dark side of testing memory: Repeated retrieval can enhance eyewitness suggestibility. *Journal of Experimental Psychology: Applied*, 17, 418–432. DOI:10.1037/a0025147.
- Chan, J. C. K., Thomas, A. K., & Bulevich, J. B. (2009). Recalling a witnessed event increases eyewitness suggestibility: The reversed testing effect. *Psychological Science*, 20, 66–73. DOI:10.1111/j.1467-9280.2008.02245.x.
- Deffenbacher, K. A., Bornstein, B. H., Penrod, S. D., & Kiernan, E. (2004). A meta-analytic review of the effects of high stress on eyewitness memory. *Law and Human Behavior*, 28, 687–706. DOI:10.1007/s10979-004-0565-x.
- Gabbert, F., Hope, L., Fisher, R. P., & Jamieson, K. (2012). Protecting against misleading post-event information with a Self-Administered Interview. *Applied Cognitive Psychology*, 26, 568–575. DOI:10.1002/acp.2828.
- Gordon, L. T., & Thomas, A. K. (2014). Testing potentiates new learning in the misinformation paradigm. *Memory & Cognition*, 42, 186–197. DOI:10.3758/s13421-013-0361-2.
- Hope, L., Gabbert, F., Fisher, R. P., & Jamieson, K. (2014). Protecting and enhancing eyewitness memory: The impact of an initial recall attempt on performance in an investigative interview. *Applied Cognitive Psychology*, 28, 304–313. DOI:10.1002/acp.2984.
- 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, 229–251. DOI:10.1016/j.jml.2008.10.002.
- Hunt, R. R., & Einstein, G. O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning and Verbal Behavior*, 20, 497–514. DOI:10.1016/S0022-5371(81)90138-9.
- Jack, F., Leov, J., & Zajac, R. (2013). Age-related differences in the free-recall accounts of child, adolescent, and adult witnesses. *Applied Cognitive Psychology*, 28, 30–38. DOI:10.1002/acp.2951.
- Krix, A. C., Sauerland, M., Gabbert, F., & Hope, L. (2014). Providing eyewitnesses with initial retrieval support: What works at immediate and subsequent recall? *Psychology, Crime & Law*, 20, 1005–1027. DOI:10.1080/1068316X.2014.902456.
- Krix, A. C., Sauerland, M., Raymaekers, L. H. C., Memon, A., Quaedflieg, C. W. E. M., & Smeets, T. (2016). Eyewitness evidence obtained with the Self-Administered Interview© is unaffected by stress. *Applied Cognitive Psychology*, 30, 103–112. DOI:10.1002/acp.3173.
- LaPaglia, J. A. (2013). Testing can enhance or reduce suggestibility: The importance of contextual detail during misinformation exposure. Unpublished doctoral dissertation. Iowa State University, Ames, IA.
- LaPaglia, J. A., & Chan, J. C. K. (2013). Testing increases suggestibility for narrative-based misinformation but reduces suggestibility for question-based misinformation. *Behavioral Sciences & the Law*, 31, 593–606. DOI:10.1002/bsl.2090.
- LaPaglia, J. A., Wilford, M. M., Rivard, J. R., Chan, J. C. K., & Fisher, R. P. (2014). Misleading suggestions can alter later memory reports even following a cognitive interview. *Applied Cognitive Psychology*, 28, 1–9. DOI:10.1002/acp.2950.
- Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory*, 12, 361–366. DOI:10.1101/lm.94705.
- McGuire, K., London, K., & Wright, D. B. (2015). Developmental trends in false memory across adolescence and young adulthood: A comparison of DRM and memory conformity paradigms. *Applied Cognitive Psychology*, 29, 334–344. DOI:10.1002/acp.3114.
- Memon, A., Meissner, C. A., & Fraser, J. (2010). The Cognitive Interview: A meta-analytic review and study space analysis of the past 25 years. *Psychology, Public Policy, and Law*, 16, 340–372. DOI:10.1037/a0020518.
- Otgaar, H., Howe, M. L., Brackmann, N., & Smeets, T. (2016). The malleability of developmental trends in neutral and negative memory

- illusions. *Journal of Experimental Psychology: General*, *145*, 31–55. DOI:10.1037/xge0000127.
- 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.
- Pansky, A. (2012). Inoculation against forgetting: Advantages of immediate versus delayed initial testing due to superior verbatim accessibility. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 1792–1800. DOI:10.1037/a0028460.
- Pansky, A., & Tenenboim, E. (2011). Inoculating against eyewitness suggestibility via interpolated verbatim vs. gist testing. *Memory & Cognition*, *39*, 155–170. DOI:10.3758/s13421-010-0005-8.
- Roediger, H. L. III, & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15*, 20–27. DOI:10.1016/j.tics.2010.09.003.
- Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, *1*, 181–210. DOI:10.1111/j.1745-6916.2006.00012.x.
- Roosendaal, B., & McGaugh, J. L. (2011). Memory modulation. *Behavioral Neuroscience*, *125*, 797–824. DOI:10.1037/a0026187.
- Sauerland, M., Raymaekers, L. H. C., Otgaar, H., Memon, A., Waltjen, T. T., Nivo, M., ... Smeets, T. (2016). Stress, stress-induced cortisol responses, and eyewitness identification performance. *Behavioral Sciences & the Law*. Advance online publication. DOI:10.1002/bsl.2249.
- Takarangi, M. K. T., Parker, S., & Garry, M. (2006). Modernising the misinformation effect: The development of a new stimulus set. *Applied Cognitive Psychology*, *20*, 583–590. DOI:10.1002/acp.1209.
- Van Bergen, S., Horselenberg, R., Merckelbach, H., Jelicic, M., & Beckers, R. (2010). Memory distrust and acceptance of misinformation. *Applied Cognitive Psychology*, *24*, 885–896. DOI:10.1002/acp.1595.
- Wang, E., Paterson, H., & Kemp, R. (2014). The effects of immediate recall on eyewitness accuracy and susceptibility to misinformation. *Psychology Crime and Law*, *20*, 619–634. DOI:10.1080/1068316X.2013.854788.
- Wilford, M. M., Chan, J. C. K., & Tuhn, S. J. (2014). Retrieval enhances eyewitness suggestibility to misinformation in free and cued recall. *Journal of Experimental Psychology: Applied*, *20*, 81–93. DOI:10.1037/xap0000001.