Title
Discrepancy detection in the retrieval-enhanced suggestibility paradigm

Journal
Memory, 26(4)

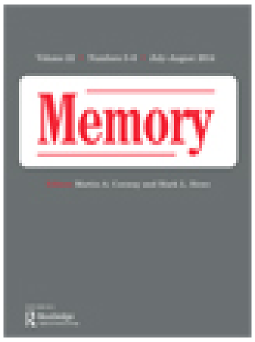
ISSN
0965-8211

Authors
Butler, Brendon Jerome
Loftus, Elizabeth F

Publication Date
2018-04-21

DOI
10.1080/09658211.2017.1371193

Peer reviewed



Memory

ISSN: 0965-8211 (Print) 1464-0686 (Online) Journal homepage: http://www.tandfonline.com/loi/pmem20

Discrepancy detection in the retrieval-enhanced suggestibility paradigm

Brendon Jerome Butler & Elizabeth F. Loftus

To cite this article: Brendon Jerome Butler & Elizabeth F. Loftus (2017): Discrepancy detection in the retrieval-enhanced suggestibility paradigm, Memory, DOI: 10.1080/09658211.2017.1371193

To link to this article: http://dx.doi.org/10.1080/09658211.2017.1371193

Published online: 12 Sep 2017.

MEMORY, 2017
https://doi.org/10.1080/09658211.2017.1371193

Routledge
Taylor & Francis Group

Discrepancy detection in the retrieval-enhanced suggestibility paradigm

Brendon Jerome Butler and Elizabeth F. Loftus

Department of Psychology and Social Behavior, University of California, Irvine, CA, USA

ABSTRACT

Retrieval-enhanced suggestibility (RES) refers to the finding that immediately recalling the details of a witnessed event can increase susceptibility to later misinformation. In three experiments, we sought to gain a deeper understanding of the role that retrieval plays in the RES paradigm. Consistent with past research, initial testing did increase susceptibility to misinformation – but only for those who failed to detect discrepancies between the original event and the post-event misinformation. In all three experiments, subjects who retrospectively detected discrepancies in the post-event narratives were more resistant to misinformation than those who did not. In Experiments 2 and 3, having subjects concurrently assess the consistency of the misinformation narratives negated the RES effect. Interestingly, in Experiments 2 and 3, subjects who had retrieval practice and detected discrepancies were more likely to endorse misinformation than control subjects who detected discrepancies. These results call attention to limiting conditions of the RES effect and highlight the complex relationship between retrieval practice, discrepancy detection, and misinformation endorsement.

ARTICLE HISTORY
Received 10 April 2017
Accepted 15 August 2017

KEYWORDS
Discrepancy detection; retrieval-enhanced suggestibility; misinformation; retrieval practice; testing

Nearly a century ago, Arthur Irving Gates (1922) showed that immediate testing improves later recall of studied or learned material. Since then, countless studies have shown this same *testing effect*. Researchers have examined the memorisation of word lists (Brewer, Marsh, Meeks, Clark-Foos, & Hicks, 2010; Pastötter, Schicker, Niedernhuber, & Bäuml, 2011; Tulving & Watkins, 1974), picture lists (Wheeler & Roediger, 1992), face-name patterns (Weinstein, McDermott, & Szpunar, 2011), and the learning of written narratives (Glover, 1989; Roediger & Karpicke, 2006). Despite changes in the type of tested material, the results remain consistent: testing enhances learning and memory for practiced material.

Testing is also known to help protect against retroactive interference, newly learned information that interferes with memories for previously learned material (Brewer et al., 2010; Chan & McDermott, 2007; Jang & Huber, 2008; Szpunar, McDermott, & Roedigger, 2009; Weinstein et al., 2011). Some researchers believe testing protects against proactive interference by improving source discrimination (Weinstein et al., 2011). According to this view, testing helps individuals isolate discrete sets of information (e.g. specific word lists) from one another, which improves speed and accuracy during retrieval. Alternatively, Pastötter et al. (2011) proposed that initial testing improves the encoding process of the learned material. Because of the enhanced encoding, an individual's memory is more resistant to potential interference from subsequent information. Testing improves later recall of studied material.

Despite the extant literature showing that initial testing buffers against proactive interference, there is a growing line of research surrounding an effect known as retrieval-enhanced suggestibility (RES), which in some ways can be thought of as a reverse testing effect. RES, or RES, refers to this finding: immediately recalling the details of a witnessed event can increase an individual's susceptibility to later misinformation (Chan, Thomas, & Bulevich, 2009). Some researchers believe that the RES effect is due to initial test questions serving as cues that guide attention to the misinformation (e.g., Gordon, Thomas, & Bulevich, 2015). For example, if on the initial test a witness is asked, "What color was the robber's hat?", the witness effectively gets a cue that the hat color is important. When the witness is later presented with post-event information, they will pay more attention to information concerning the hat color, which increases the likelihood of them learning the misinformation. Researchers studying RES have measured increased attention by recording how long participants take to read the misinformation narrative, and what is typically found is that subjects who took an initial test spend more time reading sentences that contain misinformation. Of interest to the current study are the other consequences of initial testing, particularly discrepancy detection. If a witness is paying more attention to the misinformation, would she also be more likely to notice that something is wrong with it?

CONTACT Brendon Jerome Butler butlerbj@uci.edu Department of Psychology and Social Behavior, University of California, Irvine, CA 92697-7085, USA

Discrepancy detection

Being able to detect discrepancies between something you have seen and something you are being told plays an important role in the acceptance or rejection of information. A recent study by Cochran, Greenspan, Bogart, and Loftus (2016) examined the role of discrepancy detection in a choice blindness/misinformation paradigm. In Experiment 2, subjects watched a slideshow that depicted a crime taking place. Following memory tasks and a retention interval, subjects were required to make a suspect identification from a six-person, target-absent lineup. Later, subjects were told, "Earlier in the study, you picked the photo of the man you saw in the slideshow. On the next page, you will briefly see the photo of this person." Subjects in the manipulated (misinformation) condition were not shown the suspect they selected. Instead, the suspect shown was randomly selected from the remaining five suspects that were not chosen. Additionally, they were asked to describe their reason(s) for making that identification. Concurrent detection was determined by coding the reasons participants gave for making their identification. After another retention interval, subjects were again presented with the six-person target-absent lineup and were required to make another suspect identification, which allowed the authors to determine whether the misinformation led to memory change from the first lineup to the second. The authors found that only 13.3% of subjects who detected the discrepancy between the two suspects they chose and the suspects they were shown changed from Lineup 1 to Lineup 2. In contrast, over 53% of subjects who failed to detect the discrepancy changed from Lineup 1 to Lineup 2.

Despite being an important determinant of misinformation acceptance, measures of discrepancy detection are used infrequently in misinformation research. Loftus (1979) found that when subjects are presented with blatantly contradictory information, they were more likely to reject it and were more resistant to other pieces of misinformation. Researchers have also found that reading misinformation-containing post-event narratives more slowly is associated with increased scrutiny, which leads to a greater likelihood of detecting discrepancies and resisting misinformation (Tousignant, Hall, & Loftus, 1986). This finding seems relevant to – and perhaps at odds with – some RES findings, particularly to the results found by Gordon et al. (2015). Tousignant et al. found slower reading times to be associated with discrepancy detection and increased *resistance* to misinformation, while Gordon and colleagues found slower reading times lead to increased *learning* of misinformation. Previous research has examined the role of source memory in the RES paradigm, but they never asked whether there was conflicting information in the two sources (e.g. Chan, Wilford, & Hughes, 2012). After answering a question on the final test, subjects were asked to indicate the source of the answer they provided. Subjects could respond: (a) video only, (b) audio only, (c) both video and audio, (d) neither (i.e., the answer was new). Although this provides an understanding of source memory, it is not sufficient for determining whether a subject detected a discrepancy between the sources of information. For example, a subject can respond "video only" and also remember hearing the answer in the audio narrative. Further, subjects could remember hearing the answer in both and remember them being different, or remember the answer in both and remember them being the same. A recent RES study by Gordon and Thomas (2017) measured discrepancy recollection. After taking the final test, subjects were presented with each test question and were asked to respond "YES" if they noticed discrepancies between the original event and the misinformation narrative, and "NO" if they did not. However, the authors did not investigate how recollecting discrepancies affected misinformation endorsement.

The present study

The present study was performed under the typical RES paradigm. In the first experiment,[1] all subjects watched two slideshows, both of which depicted a crime taking place. Immediately after viewing the slideshows, subjects in the *retrieval practice* condition took a cued recall test that pertained to details from the two slideshows, while subjects in the *control* condition performed an alternate task. After a retention interval, all subjects read the same post-event narratives that contained misinformation. Finally, all subjects were tested on their memory for details from the slideshows. We designed a funnelled source memory task that allowed us to determine which subjects detected discrepancies between details in the slideshows and details in the narratives.

Experiment 1

Method

Subjects

A total of 98 undergraduate students from the University of California, Irvine participated in this experiment in exchange for course credit. Of these, 16 subjects were excluded from the data analysis due to incomplete data resulting from computer problems. Previous RES research has shown medium-to-large effects with samples as small as 30 per condition. A power analysis based on a more conservative effect size of 0.55 suggested that we collect approximately 52 subjects per condition.

Design and materials

This experiment had a 2×2 mixed design, with condition (*retrieval practice vs. control*) being manipulated between subjects, and item type (*consistent vs. misinformation*) manipulated within subjects. Subjects were randomly assigned to one of the two conditions: *control* ($n = 42$) or *retrieval practice* ($n = 40$).

The materials used in this experiment are modified versions of those used in previous misinformation studies (e.g. Okado & Stark, 2005). The "witnessed events" consisted of two sets of slideshows, each of which depicted a crime taking place. In the first slideshow, a man is shown stealing a woman's wallet, and in the second, a man is shown burglarising a car. Each slideshow consisted of 50 slides that were shown at a rate of 3.5 seconds per slide.

The post-event information was presented as a written narrative. Each narrative was 50 sentences long, with each sentence summarising one of the 50 slides from the corresponding slideshow. The sentences were presented on the screen one at a time and subjects pressed a button after reading one sentence to advance to the next. Three of the sentences in each narrative (six total) were altered to include misinformation. For example, if the slideshow showed a man using a *credit card* to open a car door, the altered sentence read, "The man used a *clothes hanger* to open the car door."

There were two versions of the test, both relating to 18 details from the slideshows (nine questions per slideshow). The first version was cued recall, where subjects typed in their responses to the questions. The second version of the test was three-alternative forced choice (3 AFC). On the cued recall test, 12 of the 18 questions pertained to details from the slideshows that were unchanged in post-event narrative. On these questions, subjects could choose between three options: the correct response (consistent item) or one of two neutral lures. For the six questions that pertained to details that were later altered in the post-event narrative, subjects could choose between the consistent item (correct response), the misinformation item, or a neutral lure.

A funnelled source memory task was used to determine whether subjects detected change between the details in the slideshows and the narratives (Appendix A). After each question, subjects were asked how they knew the answer they selected. They could respond by selecting: (a) I saw it in the slideshow, (b) I read it in the narrative, (c) It was in both, and (d) I do not know. Subjects were then asked additional questions based on their initial source memory response (except those that selected "I do not know", who were not further-questioned on the source of their memory). Our funnelled source memory task is much improved on previous source memory measures, which typically give subjects the opportunity to indicate where they heard/saw/read something, but does not allow them to indicate the source (and differences between the sources) of information with granularity.

Procedure

A diagram of the experimental procedure can be found in Figure 1. Subjects were told that they would be watching a series of slideshows and that their memory for the slideshows would be tested later. Immediately after viewing the slideshows, subjects in the retrieval practice condition completed the cued recall test, while those in the control condition completed a series of health/life surveys as an alternate task. After either the immediate test or the health/life survey, all subjects filled out a demographics questionnaire and watched a distractor video to fill a retention interval. Subjects were then presented with post-event

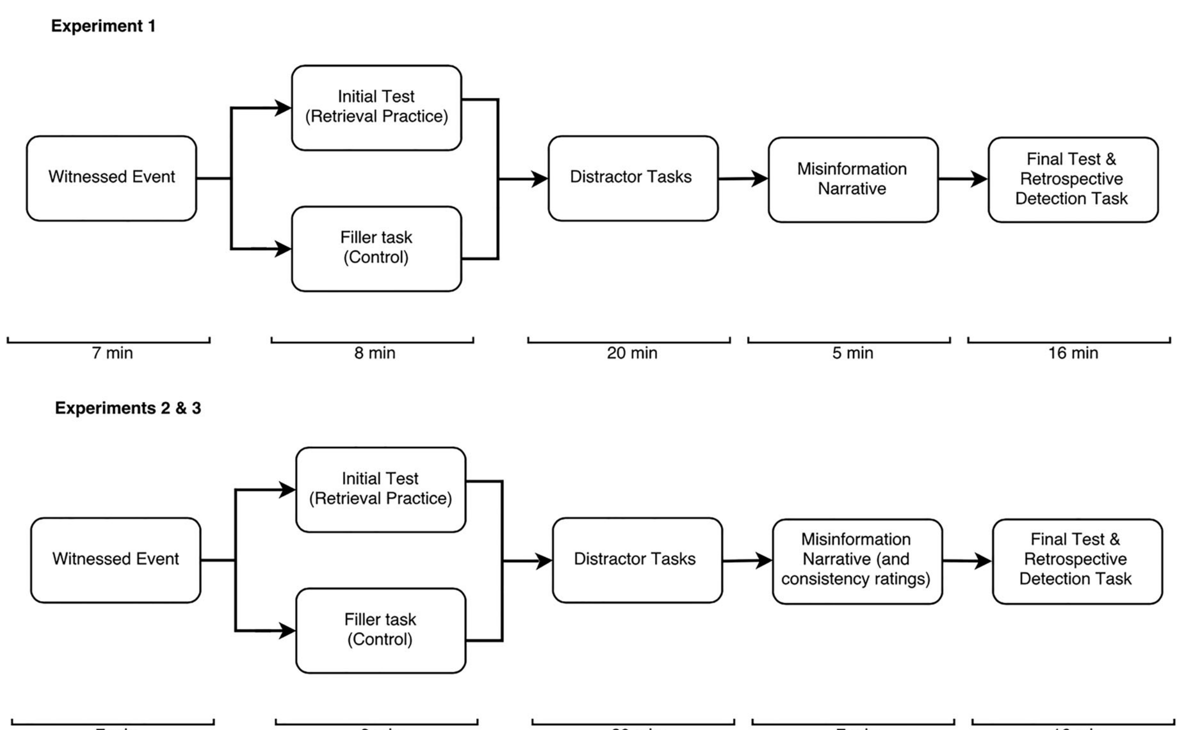


Figure 1. General experimental procedure.

information, which consisted of narratives that summarised the two slideshows. Once subjects finished reading the post-event narratives, they completed the final test and funnelled source memory task.

Results

Misinformation endorsement contingent on condition

Initial testing led to an RES effect – those in the retrieval practice condition ($M = 0.52$, 95% CI [0.44, 0.59]) were more likely to endorse misinformation on the final test than those in the control condition ($M = 0.41$, 95% CI [0.33, 0.47]), $t(80) = 2.16$, $p = .034$, $d = .476$. Consistent with past RES literature, there was no difference between conditions in the endorsement of consistent items, $t(80) = 1.57$, $p = 0.114$.

Misinformation endorsement contingent on retrospective detection status

Of importance to the present study was investigating how detecting discrepancies affects suggestibility. We used subjects' responses on the source memory task to determine detection status; when subjects noticed the discrepancy between the slideshows and the narratives for a particular item, that was considered a "detected" discrepancy, and when subjects failed to notice a discrepancy, that was considered an "undetected" discrepancy. As there were six pieces of misinformation in the narratives, each subject could endorse up to six misinformation items, and could detect a discrepancy for each of the six items.

Overall, subjects detected discrepancies 29.6% of the time. The odds of detecting discrepancies were 1.94 higher for subjects in the retrieval practice condition than for those in the control condition, $p = .031$, 95% CI [1.06, 3.55]. Collapsed across conditions, the odds of endorsing misinformation were 0.33 lower for subjects who detected discrepancies than for subjects who failed to detect, OR = 0.33, $p = .002$, 95% CI [0.16, 0.66].[2] As can be seen in Figure 2, there was no difference between conditions in misinformation endorsement rates for detectors. However, subjects in the retrieval practice condition who failed to detect discrepancies were more likely to endorse misinformation than control subjects who also failed to detect discrepancies, $z = 3.52$, $p < .001$.

Misinformation endorsement contingent on initial test performance

Overall, subjects who answered correctly on the initial test were much more likely to answer the same question correctly on the final test, OR = 6.65, $p < .001$, 95% CI [3.64, 12.14]. When subjects answered incorrectly on the initial test and later failed to detect the discrepancy, they endorsed misinformation at much higher rates ($M = 0.82$, 95% CI [0.74, 0.90]) than subjects who answered incorrectly

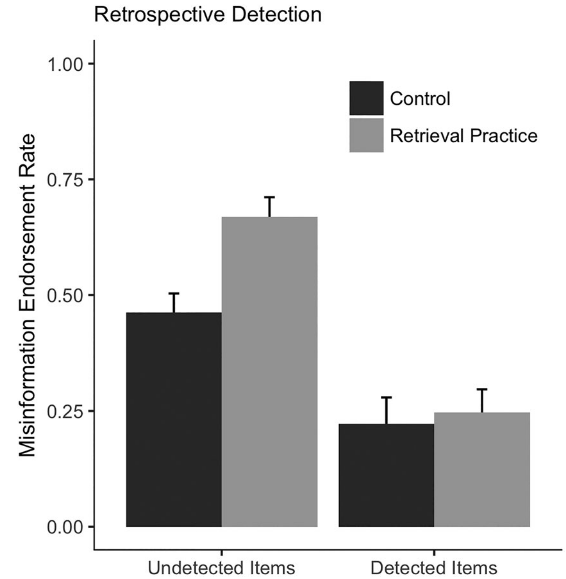


Figure 2. Misinformation endorsement rates for Experiment 1, broken down by item detection and condition. Error bars represent +1 SEM.

and detected the discrepancy ($M = 0.28$, 95% CI [0.06, 0.51]), $z = 3.98$, $p < .001$.

Reading times

Subjects in the retrieval practice condition ($M = 2.45$, 95% CI [2.22, 2.68]) took nearly the same amount of time to read misinformation sentences as control subjects ($M = 2.35$, 95% CI [2.16, 2.53]), $t(490) = 0.66$, $p > .250$. Contrary to previous findings (e.g. Gordon et al., 2015), subjects who answered questions incorrectly ($M = 2.56$, 95% CI [2.15, 2.97]) on the initial test did not spend more time reading sentences that contained misinformation than subjects who answered questions correctly ($M = 2.35$, 95% CI [1.95, 2.75]), $\beta = 0.21$, 95% CI [−0.21, 0.63]. We did not find reading times to be associated with retrospective detection, OR = 0.89, $p = .173$, 95% CI [0.77, 1.04].

Although initial test response was not associated with reading times, we did find initial test response to be associated with discrepancy detection; when subjects answered correctly on the initial test, the odds of them detecting a discrepancy for the corresponding item were over nine times higher than for subjects who answered incorrectly, OR = 9.32, $p < .001$, 95% CI [4.39, 19.77] (Table 1).

Experiment 2

Failing to retrospectively detect discrepancies predicted how likely a subject was to endorse misinformation, and this finding was most noticeable for those in the retrieval practice condition. We were not able to replicate previous findings showing that answering a question incorrectly on the initial test guides attention (as measured by reading time) towards corresponding sentences in the misinformation narrative. However, we did find that initial test response predicted whether subjects detected discrepancies.

Table 1 Average reading times for sentences containing misinformation for subjects in the retrieval practice condition, contingent on initial test accuracy and discrepancy detection (in seconds, confidence intervals in brackets).

	Detection type			
	Concurrent	Retrospective	Both	No detection
Experiment 1				
Correct	–	2.23 [1.69, 2.76]	–	2.50 [1.96, 3.04]
Incorrect	–	2.43 [1.59, 3.27]	–	2.58 [2.14, 3.02]
Experiment 2				
Correct	4.07 [3.53, 4.61]	3.60 [2.51, 4.70]	3.60 [2.51, 4.70]	4.19 [3.66, 4.72]
Incorrect	4.37 [3.70, 5.04]	4.38 [3.18, 5.57]	3.91 [2.03, 5.79]	3.36 [2.76, 3.96]

One possible concern with the findings is that subjects were asked retrospectively whether they detected discrepancies between the events and misinformation. Imperfect subject memory could mislead us into thinking detection occurred when it did not (or vice versa). Ideally, a method that measured detection concurrently, and did not depend so heavily on memory, would provide more useful information about detection. Thus, in Experiment 2, we introduced a method to measure detection concurrently.

Method

Subjects

A total of 121 undergraduate students from the University of California, Irvine participated in this experiment in exchange for course credit.

Design, materials, and procedure

Experiment 2 used the same design and materials as those in Experiment 1. Subjects were randomly assigned to one of the two conditions: *retrieval practice* ($n = 57$) or *control* ($n = 64$). The procedure for Experiment 2 was the same as Experiment 1, with the addition of a concurrent detection task, which was modelled after concurrent detection tasks used in other research (e.g. Wahlheim & Jacoby, 2013; Putnam, Sungkhasettee, & Roediger, 2017). As in Experiment 1, the post-event narrative was presented to subjects one sentence at a time. However, instead of simply pressing a button to advance to the next sentence, subjects were instructed to press a button on the screen indicating whether the sentence they just read was consistent or inconsistent with what they saw in the slideshows. As they read sentences containing misinformation, when subjects pressed the button labelled "Inconsistent" we classified that as a detected item, and when subjects pressed the button labelled "Consistent" we classified that as an undetected item.

Results

Misinformation endorsement contingent on condition

Subjects in the retrieval practice condition ($M = 53.5$, 95% CI [47.1, 59.9]) did not differ from control subjects ($M = 51.1$, 95% CI [44.9, 57.1] in misinformation endorsement on the final test. In other words, we did not observe an RES effect.

Misinformation endorsement contingent on concurrent detection status

Misinformation endorsement rates broken down by detection status can be found in Figure 3. There were no differences in concurrent detection rate between the two conditions, $t(119) = 0.57$, $p = .54$. Consistent with the discrepancy detection principle, subjects (in both conditions) who detected a discrepancy between the misinformation and the post-event narrative were less likely to endorse misinformation on the final test, $z = 12.78$, $p < .001$, 95% CI [0.36, 0.50].

Misinformation endorsement contingent on retrospective detection status

Subjects retrospectively detected discrepancies 21% of the time, and there were no differences in detection rate between condition, $t(122) = 0.56$, $p = .52$. Like in Experiment 1, subjects who failed to detect discrepancies were more likely to endorse misinformation ($M = 0.53$, 95%

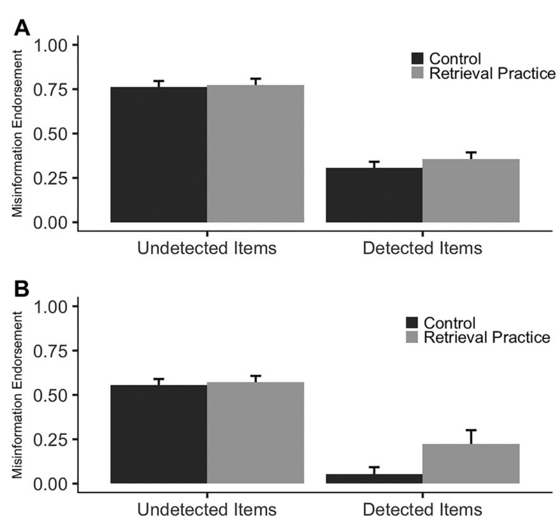


Figure 3. Misinformation endorsement rates for Experiment 2, broken down by condition and item detection. Panels A and B represent concurrent and retrospecive detection, repspectively. Error bars represent +1 SEM.

CI [0.49, 0.63]) than subjects who did detect discrepancies ($M = 0.15$, 95% CI [0.02, 0.25], $z = 5.27$, $p < .001$, 95% CI [0.24, 0.52]. Unlike in Experiment 1, however, subjects who failed to detect discrepancies in the retrieval practice condition ($M = 0.57$, 95% CI [0.50, 0.64]) were not more likely to endorse misinformation than subjects who failed to detect in the control condition ($M = 0.55$, 95% CI [0.49, 0.62]), $z = 0.53$, $p = .59$, 95% CI [-0.06, 0.12]. In fact, a completely different finding emerged – subjects in the retrieval practice who detected discrepancies ($M = 0.22$, 95% CI [0.07, 0.37] were *more likely* to endorse misinformation than subjects in the control condition who detected discrepancies ($M = 0.05$, 95% CI [−0.02, 0.12]), $z = 1.99$, $p = .047$, 95% CI [0.02, 0.57].

Misinformation endorsement contingent on initial test performance

Subjects who answered correctly on the initial test were much more likely to answer the same question correctly on the final test, OR = 7.58, $p < .001$, 95% CI [4.63, 12.42]. There was also an effect of detection and initial test performance on misinformation endorsement – when subjects answered incorrectly on the initial test and later failed to concurrently detect the discrepancy, they endorsed misinformation at much higher rates ($M = 0.90$, 95% CI [0.84, 0.97]) than subjects who answered incorrectly and detected the discrepancy ($M = 0.69$, 95% CI [0.57, 0.82]), $z = 3.14$, $p = .002$, 95% CI [0.07, 0.34]. The results for retrospective detection follow the same pattern, but they were not statistically significant; when subjects answered incorrectly on the initial test and later failed to retrospectively detect the discrepancy, they endorsed misinformation at higher rates ($M = 0.83$, 95% CI [0.76, 0.91]) than subjects who answered incorrectly and detected the discrepancy ($M = 0.59$, 95% CI [0.26, 0.93]), $z = 1.40$, $p = .161$, 95% CI [-0.09, 0.57].

As in Experiment 1, we found initial test response to be associated with discrepancy detection. When subjects answered correctly on the initial test, the odds of them concurrently detecting a discrepancy for the corresponding item were double than for subjects that answered incorrectly, OR = 2.04, $p = .002$, 95% CI [1.30, 3.22]. Although not statistically significant, the same pattern emerged for retrospective detection; when subjects answered correctly on the initial test, the odds of them retrospectively detecting a discrepancy for the corresponding item were higher than for subjects who answered incorrectly, OR = 1.68, $p = .197$, 95% CI [0.76, 3.74].

Reading times

Subjects in the retrieval practice condition ($M = 4.03$, 95% CI [3.77, 4.29]) took nearly the same amount of time to read misinformation sentences as control subjects ($M =$ 4.00, 95% CI [3.73, 4.27]), $t(712) = 0.12$, $p > .250$. As in Experiment 1, subjects who answered questions incorrectly ($M =$ 4.24, 95% CI [3.84, 4.64]) on the initial test did not spend more time reading sentences that contained misinformation than subjects who answered questions correctly ($M = 3.89$, 95% CI [3.51, 4.26]), $\beta = -0.35$, 95% CI [−0.88, 0.18]. We did not find reading times to be associated with retrospective (OR = 1.01, $p = .706$, 95% CI [0.92, 1.12]) nor concurrent (OR = 1.01, $p = .671$, 95% CI [0.95, 1.07]) detection.

Experiment 3

The results in Experiment 2 hinted that the presence of the concurrent detection task was affecting how subjects were engaging with the misinformation during the post-event narratives. We speculated that concurrent detection task used in Experiment 2 was causing subjects to pay much more attention to each sentence than they ordinarily would. We hypothesised that because the post-event narrative was presented one sentence at a time, and subjects had to make discrepancy decisions for each one, they were hyper-vigilant, resulting in more resistance to misinformation. However, our results indicated that neither condition nor initial test response was associated with misinformation sentence reading times. Thus, we sought after an alternative measure of attention/discrepancy detection.

We partnered with the market research company Dialsmith® and used their Perception Analyzer Online® (PAO) as a replacement to the concurrent detection task used in Experiment 2. The PAO is a state-of-the-art solution for moment-moment evaluation of recorded media (Dialsmith, www.dialsmith.com). As it is used in market research, respondents continuously rate – in real time – how positively or negatively feel about the media they are currently viewing. For our purposes, we used the PAO to track how subjects evaluated the consistency/inconsistency of information in the post-event narrative.

Method

Subjects

A total of 124 undergraduate students from the University of California, Irvine participated in this study in exchange for course credit.

Design, materials, and procedure

The design and materials of the experiment was the same as in Experiments 1 and 2. Subjects were randomly assigned to either the *retrieval practice* ($n = 62$) or *control* ($n = 62$) conditions.

The procedure for Experiment 3 was the same as Experiment 2, except for the post-event information and concurrent detection task. Instead of reading sentences one at a time, subjects listened to an audio recording of the post-event narrative.

As subjects listened to the audio narrative, they rated the consistency of the information they were hearing by continuously moving an on-screen slider from 0 (inconsistent) to 100 (consistent). We view the PAO as an improvement over the methods used in Experiment 2 and other studies for several reasons: First, subjects were not forced to read the post-event narrative sentence-by-sentence and make a binary consistent/inconsistent decision for each individual sentence; by using the PAO, subjects rated narrative consistency on a continuous scale. Second, using this new tool allows subjects to listen to the post-event narrative as opposed to reading it sentence-by-sentence, which mirrors a more realistic scenario where a person might hear post-event information on the news or from a fellow witness.

Results

Misinformation endorsement contingent on condition

We found no evidence for RES in this experiment; subjects in the retrieval practice condition ($M = 0.56$, 95% CI [0.50, 0.61]) were not more likely to endorse misinformation than control subjects ($M = 0.49$, 95% CI [0.44, 0.55]), OR = 1.29, 95% CI [0.93, 1.79].

Misinformation endorsement contingent on retrospective detection status

There were no differences in detection rate between the retrieval practice ($M = 0.07$, 95% CI [0.04, 0.10]) and control conditions ($M = 0.05$, 95% CI [0.02, 0.08]), $t(122) = 0.65$, $p = .51$ (Figure 4).

When collapsed across conditions, undetected items were more likely to be endorsed ($M = 0.56$, 95% CI [0.52, 0.60]) than detected items ($M = 0.19$, 95% CI [0.10, 0.31]), $z = 6.48$, $p < .001$, 95% CI [0.25, 0.48]. The same pattern from Experiment 2 emerged; when items were detected in the retrieval practice condition ($M = 0.35$, 95% CI [0.21, 0.49]), they were more likely to be endorsed than detected items in the control condition ($M = 0.09$, 95% CI [−0.01, 0.19]), $z = 2.97$, $p = .003$, 95% CI [0.08, 0.43].

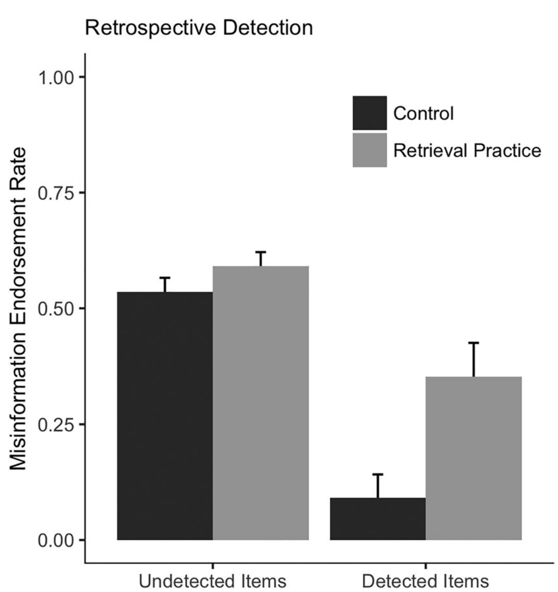


Figure 4. Misinformation endorsement rates for Experiment 3, broken down by condition and item detection. Error bars represent +1 SEM.

Misinformation endorsement contingent on initial test performance

Subjects who answered correctly on the initial test were much more likely to answer the same question correctly on the final test, OR = 7.58, $p < .001$, 95% CI [4.75, 12.68]. When subjects answered incorrectly on the initial test and later failed to detect the discrepancy, they endorsed misinformation at much higher rates ($M = 0.73$, 95% CI [0.67, 0.90]) than subjects who answered incorrectly and detected the discrepancy ($M = 0.42$, 95% CI [0.19, 0.68]), $z = 2.23$, $p = .022$.

Perception analyser online

Consistency ratings for misinformation items can be found in Table 2 and a visual representation of the PAO can be seen in Figure 5. We measured the consistency rating two and a half seconds after subjects heard the misinformation, giving them time to process and adjust their ratings accordingly. There were no significant differences between conditions in narrative consistency ratings for the six misinformation items. In addition, there was no relationship between initial test response and consistency rating – subjects who answered correctly on the initial test ($M = 66.47$, 95% CI [60.68, 72.25]) had similar ratings as subjects who answered incorrectly ($M = 68.17$, 95% CI [62.63, 73.70]), $p > .250$.

We were interested in seeing whether subjects' changes in their consistency ratings predicted retrospective detection. Overall, subjects whose consistency ratings decreased after hearing the misinformation were more likely to retrospectively detect that discrepancy, OR = 1.91, $p = .020$, 95% CI [1.09, 3.30]. Although, the interaction between conditions was not significant, (OR = 2.01, $p = .224$, 95% CI [0.65, 6.45]), there were differences within the retrieval practice condition; subjects who decreased their consistency ratings were more likely to endorse misinformation ($M = 0.16$, 95% CI [0.08, 0.27]) than subjects who increased or stayed the same ($M = 0.06$, 95% CI [0.04, 0.11]), $z = 2.57$, $p = .010$. There were no differences for subjects in the control condition, $z = 0.54$, $p > .250$.

Discussion

The goal of the present work was to investigate how retrieval practice affects discrepancy detection and susceptibility to misinformation. In all three experiments, subjects who retrospectively detected discrepancies in the post-event narratives were more resistant to misinformation than those who did not. In Experiments 2 and 3, even when detecting discrepancies, subjects who had retrieval practice were more likely to endorse misinformation than control subjects. These results highlight the complex relationship between discrepancy detection and retrieval practice. Surprisingly, the role of discrepancy detection

Table 2. Mean consistency ratings for each of the six misinformation items.

	Misinformation item				
	Item 1	Item 2	Item 3	Item 4	Item 5
Condition					
Control	74.04 [67.12, 80.96]	57.98 [49.12, 66.84]	60.82 [53.13, 68.51]	57.95 [49.44, 66.45]	76.08 [69.48, 82.67]
Retrieval practice	74.50 [68.11, 80.88]	58.04 [50.50, 65.59]	67.25 [60.44, 74.07]	56.46 [47.79, 65.14]	74.41 [67.34, 81.49]

has been mostly overlooked in the RES paradigm. Previous RES studies have used source monitoring tasks before (e.g. Chan et al., 2012), but the alternative forced choice methods used were inadequate at determining whether subjects detected a discrepancy between the two sources of information, or simply have source monitoring errors. Gordon and Thomas (2017) measured discrepancy recollection retroactively, but they did not examine how recollecting those discrepancies affected susceptibility to misinformation. Compared to previous research, our funnelled source memory task, combined with the use of binary (Experiment 2) and continuous (Experiment 3) consistency ratings, allowed us to draw a much more complete picture of how subjects engage with misinformation after retrieval practice.

We only found a general RES effect[3] in Experiment 1, where subjects read the post-event narratives without additional demands. When subjects were required to assess the consistency of the narratives, there were no longer overall differences in misinformation endorsement between the two conditions, i.e., no RES effect. This finding was present when subjects made explicit, binary consistency judgments in Experiment 2, as well as when they rated consistency freely on a continuous scale in

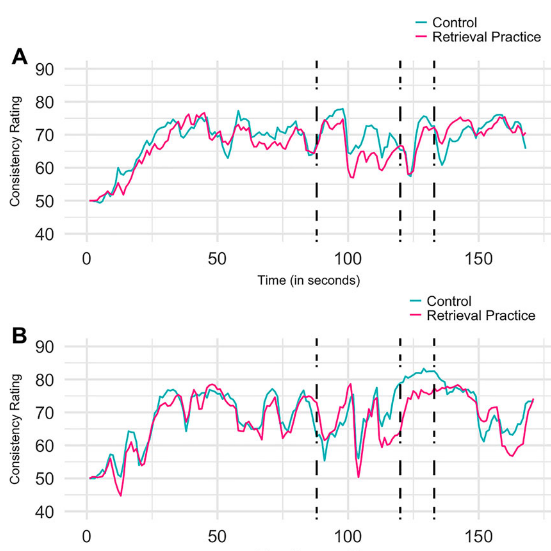


Figure 5. Visual representation of the PAO data. Mean consistency ratings are plotted over the time in the audio narrative (x-axis). Vertical dashed lines represent the location in the narrative of each misinformation item. Panels A and B represent the first and second misinformation narrative, respectively.

Experiment 3. It is possible that the concurrent tasks in Experiments 2 and 3 served as warnings, signalling to subjects that there were issues with the post-event narrative. Past research has shown warnings to reduce an individual's susceptibility to misinformation in both the traditional misinformation (e.g., Greene, Flynn, & Loftus, 1982) and RES (e.g., Thomas, Bulevich, & Chan, 2010) paradigms. Consistent with the discrepancy detection hypothesis (Tousignant et al., 1986), subjects who detected discrepancies were more resistant to misinformation than subjects who failed to detect discrepancies. Further, those who failed to detect discrepancies endorsed misinformation at much higher rates, regardless of whether they were in the retrieval practice or control condition.

In Experiments 2 and 3, even when detecting discrepancies, subjects who had retrieval practice were more likely to endorse misinformation than control subjects. These findings are consistent with prior work that shows noticing change leading to proactive facilitation. Wahlheim and Jacoby (2013) had subjects study word lists in an A–B, A–D paradigm. In the first phase of the experiment, subjects were instructed to read List 1 pairs (A–B) as quickly as possible. During the second phase, subjects were required to study and learn List 2 pairs (A–D), as well as indicate when they noticed that the word on the right (D) was different than the one presented earlier (B). The authors found a facilitative effect of memory such that memory for List 2 was improved when subjects noticed that a change had occurred. Further, increasing the number of presentations of the first event increased detection of change, which led to improved memory for the second event. In the present study, those who had an additional presentation of the first event (retrieval practice condition) and detected discrepancies were more likely to endorse misinformation than control subjects who detected discrepancies.

When using reading times as a measure of attention, we failed to replicate previous findings showing the association between initial testing and misinformation endorsement. For example, Gordon et al. (2015) found that subjects spent more time reading sentences that contained misinformation, and as a result, were more likely to endorse the misinformation on the final test. In our study, however, there were no differences in reading times between the two conditions, nor was initial test response associated with reading times. Admittedly, we were unsurprised to find no association between reading times and misinformation endorsement. On one side, there is a line of research showing that slower reading times are associated with enhanced scrutiny and increased misinformation resistance (e.g. Tousignant et al., 1986), while on the other side, recent RES studies have shown slower reading times to be associated with enhanced attention and increased misinformation endorsement. What was surprising, however, was that we also failed to see an association between reading times and discrepancy detection. As such, our findings failed to support both accounts of the effects of reading times on misinformation endorsement. Because we recognised that merely measuring how long subjects read each sentence may be a poor proxy for both scrutiny and attention, we sought an alternative, more nuanced measure that could either converge or diverge from the reading time & concurrent detection results. When using this alternative measure – the PAO – in Experiment 3, we found that when subjects had retrieval practice and decreased their ratings of event-narrative consistency, they were more likely to endorse misinformation. This pattern of results is like those found for retrospective detection in Experiments 2 and 3, where subjects who had retrieval practice and detected discrepancies were more likely to endorse misinformation than control subjects. It is possible that these findings can be accounted for by retrieval fluency – the ease with which information can be retrieved from memory (Benjamin, Bjork, & Schwartz, 1998). Under this view, detecting a discrepancy between the original memory and the misinformation narrative makes the latter information more salient, and thus, more likely to be endorsed. This hypothesis is consistent with previous RES research that has addressed the idea of retrieval fluency (e.g. Chan et al., 2012; Thomas et al., 2010).

In conclusion, our work adds to the RES literature in three important ways. First, it suggests that the RES effect is not as robust and generalisable as initially expected. Our data show that the effect is most prevalent in pristine situations that are free of additional task demands. Second, our work highlights the prominent role that discrepancy detection plays in the RES paradigm. When subjects detect discrepancies, they are much more resistant to misinformation, even after retrieval practice.[4] Third, our use of a non-traditional measurement tool – the PAO – sheds more light on how subjects engage and react to misinformation than methods typically used in this field. Future research should continue to explore the probative value of using various types of measurement tools when examining the complex relationship between retrieval practice, discrepancy detection, and misinformation endorsement.

Notes

1. Experiments 1–3 followed the same general procedure, which can be found in Figure 1.
2. When conditionalised on detection status and initial test response, misinformation endorsement was analysed as a dichotomous outcome. To analyse these data, we used a multilevel generalized linear model within the binomial family.
3. We refer to subjects in the *retrieval practice* condition endorsing misinformation at higher rates than *control* subjects as a "general" RES effect. The general RES effect does not take into account detection status.
4. Although *retrieval practice* subjects who detected discrepancies endorsed misinformation at higher rates than control subjects in Experiments 2 and 3, their overall rate of misinformation endorsement was much smaller than subjects who failed to detect discrepancies.

Acknowledgements

We would like to thank Dialsmith for allowing us to use their Perception Analyzer Online.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This material is based upon work supported by the National Science Foundation [Grant No. DGE-1321846].

ORCID

Brendon Jerome Butler http://orcid.org/0000-0003-3447-782X

References

Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory?: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General*, *127*(1), 55–68.

Brewer, G. A., Marsh, R. L., Meeks, J. T., Clark-Foos, A., & Hicks, J. L. (2010). The effects of free recall testing on subsequent source memory. *Memory (Hove, England)*, *18*(4), 385–393.

Chan, J. C. K. & McDermott, K. B. (2007). The testing effect in recognition memory: A dual process account. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *33*(2), 431–437.

Chan, J. C. K., Thomas, A. K., & Bulevich, J. B. (2009). Recalling a witnessed event increases eyewitness suggestibility. *Psychological Science*, *20*(1), 66–73.

Chan, J. C. K., Wilford, M. M., & Hughes, K. L. (2012). Retrieval can increase or decrease suggestibility depending on how memory is tested: The importance of source complexity. *Journal of Memory and Language*, *67*, 78–85.

Cochran, K. J., Greenspan, R. L., Bogart, D. F., & Loftus, E. F. (2016). Memory blindness: Altered memory reports lead to distortion in eyewitness memory. *Memory & Cognition*, *44*, 717–726.

Glover, J. A. (1989). The "testing" phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology*, *81*(3), 392–399.

Gordon, L. T., Thomas, A. K., & Bulevich, J. B. (2015). Looking for answers in all the wrong places: How testing facilitates learning of misinformation. *Journal of Memory and Language*, *83*, 140–151. doi:10.1016/j.jml.2015.03.007

Gordon, L. T. & Thomas, A. K. (2017). The forward effects of testing on eyewitness memory: The tension between suggestibility and learning. *Journal of Memory and Language*, *95*, 190–199.

Greene, E., Flynn, M. S., & Loftus, E. F. (1982). Inducing resistance to misleading information. *Journal of Verbal Learning and Verbal Behavior*, *21*(2), 207–219.

Jang, Y. & Huber, D. E. (2008). Context retrieval and context change in free recall: Recalling from long-term memory drives list isolation. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *34*(1), 112–127.

Loftus, E. F. (1979). Reactions to blatantly contradictory information. *Memory & Cognition*, *7*, 368–374.
Okado, Y. & Stark, C. E. L. (2005). Neural activity during encoding predicts false memories created by misinformation. *Learning & Memory*, *12*(1), 3–11.
Pastötter, B., Schicker, S., Niedernhuber, J., & Bäuml, K.-H. T. (2011). Retrieval during learning facilitates subsequent memory encoding. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *37*(2), 287–297.
Putnam, A. L., Sungkhasettee, V. W., & Roediger, H. L. (2017). When misinformation improves memory: The effects of recollecting change. *Psychological Science*, *28*(1), 36–46.
Roediger, H. L. & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Association for Psycological Science*, *17*, 249–255.
Szpunar, K. K., McDermott, K. B., & Roedigger, H. L. (2009). "Testing during study insulates against the buildup of proactive interference": Correction. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *35*(6), 156.
Thomas, A. K., Bulevich, J. B., & Chan, J. C. K. (2010). Testing promotes eyewitness accuracy with a warning: Implications for retrieval enhanced suggestibility. *Journal of Memory and Language*, *63*, 149–157. doi:10.1016/j.jml.2010.04.004
Tousignant, J. P., Hall, D., & Loftus, E. F. (1986). Discrepancy detection and vulnerability to misleading postevent information. *Memory & Cognition*, *14*(4), 329–338. doi:10.3758/BF03202511
Tulving, E. & Watkins, M. J. (1974). On negative transfer: Effects of testing one list on the recall of another. *Journal of Verbal Learning and Verbal Behavior*, *13*, 181–193.
Wahlheim, C. N. & Jacoby, L. L. (2013). Remembering change: The critical role of recursive remindings in proactive effects of memory. *Memory and Cognition*, *41*, 1–15. doi:10.3758/s13421-012-0246-9
Weinstein, Y., McDermott, K. B., & Szpunar, K. K. (2011). Testing protects against proactive interference in face-name learning. *Psychonomic Bulletin & Review*, *18*, 518–523.
Wheeler, M. A. & Roediger, H. L. (1992). Disparate effects of repeated testing: Reconciling Ballard's (1913) and Bartlett's (1932) results. *Psychological Science*, *3*(4), 240–246.

Appendix