Memory

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

On our susceptibility to external memory store manipulation: examining the influence of perceived reliability and expected access to an external store

April E. Pereira, Megan O. Kelly, Xinyi Lu & Evan F. Risko

To cite this article: April E. Pereira, Megan O. Kelly, Xinyi Lu & Evan F. Risko (2022) On our susceptibility to external memory store manipulation: examining the influence of perceived reliability and expected access to an external store, Memory, 30:4, 412-428, DOI: 10.1080/09658211.2021.1990347

To link to this article: https://doi.org/10.1080/09658211.2021.1990347

MEMORY
2022, VOL. 30, NO. 4, 412–428
https://doi.org/10.1080/09658211.2021.1990347

On our susceptibility to external memory store manipulation: examining the influence of perceived reliability and expected access to an external store

April E. Pereira, Megan O. Kelly, Xinyi Lu and Evan F. Risko

Department of Psychology, University of Waterloo, Waterloo, Ontario, Canada

ABSTRACT
Offloading memory to external stores (e.g., a saved file) allows us to evade the limitations of our internal memory. One cost of this strategy is that the external memory store used may be accessible to others and, thus, may be manipulated. Here we examine how reducing the perceived reliability of an external memory store and manipulating one's expectation for future access to such a store can influence participants' susceptibility to its manipulation (i.e., endorsing manipulated information as authentic). Across three pre-registered experiments, participants were able to store to-be-remembered information in an external store. On a critical trial, we surreptitiously manipulated the information in that store. Results demonstrated that an explicit notification of a previous manipulation of that store and the warning that the store will be inaccessible in the future can decrease susceptibility to manipulation of that store. Results are discussed in the context of the metacognitive monitoring and control of memory reports in situations that involve the distribution of memory demands across both internal and external spaces.

ARTICLE HISTORY
Received 1 November 2020
Accepted 1 October 2021

KEYWORDS
Memory; recognition; recall; offloading; reliability

Individuals are often presented with to-be-remembered information that is critical for accomplishing their future goals. Given that the ability to store and retrieve accurate information from internal/biological memory is limited, this can cause problems when what we wish to remember exceeds what we may be capable of accurately remembering (Cowan, 2010). As such, it is often easier and/or more beneficial to rely on external storage devices rather than to rely on our internal/biological memory (Eskritt & Ma, 2014; Hutchins, 1995; Kelly & Risko, 2019a; Lu et al., 2020; Lu et al., 2021; Sparrow et al., 2011). In this digital age, the amount of information that can be stored externally (e.g., in cyberspace) is virtually limitless and is usually readily accessible.

The use of external memory storage in place of internal memory storage can be thought of as a form of *cognitive offloading* (Risko & Gilbert, 2016). While offloading memory demands in this manner grants individuals the benefit of having an extended memory system with a vast capacity, there are risks associated with taking such an approach to "remembering" (Ferguson et al., 2015; Kelly & Risko, 2019a; Kelly & Risko, 2021; Lu et al., 2020; Sparrow et al., 2011). One risk is that the external store can be manipulated by others (Clark, 2010b; Sterelny, 2004) unbeknownst to us. This is particularly problematic when our external memory stores are in places accessible via the Internet (e.g., personal information stored "in the cloud") and thus, in principle, accessible by others.

Endorsement of information

When retrieving from an external memory store (e.g., a file stored in the cloud, a notebook), one must decide whether to endorse the information in the external store as that which had originally been stored there (i.e., the *endorsement problem*; Arango-Muñoz, 2013). In a series of experiments examining this general problem, Risko et al. (2019) focused on the individual's susceptibility to manipulation of their external memory stores. They presented participants with to-be-remembered words and instructed them to save the presented information to a computer file that they could access during a subsequent recall test. Doing so provided the participants the opportunity to offload the memory demands to the external store. Unsurprisingly, this allowed near-perfect "recall" of the stored information at test. After repeating this procedure across multiple trials, on the final trial (of critical interest), the researchers manipulated the information in the participant's external memory store by inserting a novel word into it. Individuals often failed to notice this manipulation, with most recalling the inserted information as if it had been initially presented. Importantly, endorsement was not absolute; that is, individuals did not appear to merely trust their external store uncritically. How, then, do individuals decide whether to endorse the information in their external memory stores?

CONTACT April E. Pereira april.pereira@uwaterloo.ca

A useful means of framing this question theoretically is to think of the endorsement problem from a metacognitive perspective (Arango-Muñoz, 2013). For example, Goldsmith and Koriat's (1999) discuss a metacognitive framework for understanding memory reports in the context of situations where individuals must decide whether to volunteer an answer to a query (e.g., the answer to a trivia question). According to Goldsmith and Koriat (1999), this decision combines information from (1) a monitoring process that provides a subjective sense of the likely correctness of a retrieved answer with (2) a control mechanism that is sensitive to situational demands and ultimately decides whether the answer will be reported. The endorsement problem can be seen as a similar kind of problem, as an individual must decide whether to endorse (i.e., report) information from their external store as being the information stored there initially. Thus, we can imagine that similar monitoring and control mechanisms are at play here. For example, when we encounter information in our external memory stores, it likely comes with a feeling of familiarity. In addition, we also have a history of external memory store use, both in general and with the particular external store in question, and face various demands associated with that retrieval (e.g., a need for accuracy versus speed). From this theoretical perspective, what seems clear is the need to better understand what factors are considered in the face of such an endorsement problem and how they come to influence the endorsement of information in external memory stores.

In the present investigation, we pursue this broad question through an examination of the influence of two manipulations on the endorsement of information inserted into an individual's external memory store—the perceived reliability of the external memory store and the expected access to that external store during a future test of memory. How reliable an individual considers a given external memory store to be is likely to play an important role in whether an individual endorses information stored within it (Lewandowsky et al., 2000; Muir & Moray, 1996; Storm & Stone, 2015; Weis & Wiese, 2019). Research consistent with this idea has demonstrated that reliability is related to the individual's reliance on external aids to perform cognitive tasks.

Weis and Wiese (2019) examined the effect of actual and believed reliability on an individual's decision to offload task demands in a mental rotation task. In this task, participants had the option to rotate the stimuli either internally (mentally) or externally, with a rotation knob that rotated the object on a computer screen. The knob's actual reliability and an instruction altering participants' beliefs about the knob's reliability (believed reliability) were manipulated, and the frequency of cognitive offloading (i.e., the use of the knob) and perceived knob utility were measured. They found that participants adjusted their offloading based on the actual and believed reliability of the knob. When participants experienced a decrease in the knob's actual reliability or were led to believe that the knob's reliability was lower than it actually was, participants reduced their use of the external rotation option.

In the context of offloading memory demands, Storm and Stone (2015) provided evidence that the reliability of an external memory store modulated the benefit of offloading. Across three experiments, Storm and Stone (2015) demonstrated that when participants believed (at study) that a file containing a list of to-be-remembered words would be saved and accessible at the time of test, there was a benefit to the recall of an intervening list that was not saved. The authors proposed that offloading the initial list reduced proactive interference on the subsequent list. Particularly relevant to the present effort, this benefit of offloading was not observed when the external memory store was considered unreliable. Unreliability in this case was manipulated by participants experiencing an ineffective saving process. Storm and Stone (2015) suggested that when the external memory store was perceived as unreliable, individuals were less likely to offload their memory to that store (despite it being available), thus reducing the benefit to the subsequent list.

The Storm and Stone (2015) explanation highlights two ideas central to the present investigation. First, reducing the perceived reliability of an external store can reduce reliance on it. If we view accepting information inserted into an external store as an issue related to too much reliance on that store, then reducing the external store's perceived reliability should reduce susceptibility to manipulation of that store. The second idea is that in the context of storing information in an external store, if an individual does not believe that external memory store to be reliable, then they might not offload memory to that store, instead opting to store that information internally. The notion that offloading during study might influence later susceptibility to external store manipulation was raised in the original work investigating this issue. Risko and colleagues (2019) argued that one potential reason that participants often accepted an item inserted into their external memory stores was that their expectation of having access to that external store during initial study led to poor encoding of the actually presented information. For example, in research investigating cognitive offloading, participants who expect to have access to an external memory store at recall (i.e., those that can offload the memory demands), recall less than those who do not expect to have access to an external store (i.e., those that cannot offload memory demands; Kelly & Risko, 2019a; 2019b; Lu et al., 2020). This might reflect individuals forgoing efforts to internally store information when they can rely on it being available externally. Returning to the individual's susceptibility to external store manipulation, a poor internal representation for information that was actually presented (i.e., legitimate information) would presumably make it more difficult to differentiate it from inserted items (i.e., illegitimate information). This can be thought of as a basic signal detection problem wherein poor encoding, due to an expectation of future access to the external memory store, leads to greater overlap in the distributions of memory strength/familiarity and as a result, a reduced ability to distinguish actually presented from unpresented items.

Another route through which encoding activities might influence susceptibility to external store manipulation is that it can influence one's expectations with respect to their own memory. For example, Scoboria and colleagues (2007) found that they could enhance people's belief in a childhood event (i.e., a belief that an event occurred regardless of an accompanying memory), by providing participants with both high prevalence information (e.g., "this event is common") and a rationale for the common experience of forgetting past events. That is, when participants are instructed that the likelihood of an event is high and that forgetting often occurs, they are more likely to increase their belief that an event happened to them. In the context of external memory stores, if participants encoded the information poorly, they may not expect items in their external memory store to be associated with an experience of remembering (e.g., a feeling of familiarity). Consequently, the lack of such experience when they encounter an inserted item in their external store would not itself set off any proverbial alarm bells. This might make it difficult, again, to tell legitimate from illegitimate information.

Present investigation

In the present investigation, we examined both the perceived reliability of the external memory store and encoding conditions as two possible factors influencing endorsement of information inserted into an external memory store. In Experiment 1, we extended previous work which examined individuals' susceptibility to endorsing information that has been surreptitiously inserted into their external memory store (Risko et al., 2019). In Experiment 2, we compared this susceptibility in a condition wherein individuals were made "naïve" to the insertion, as in the work by Risko et al. (2019), to a condition in which individuals were informed that we had previously manipulated their external memory store. Lastly, in Experiment 3, we sought to investigate how differences in expected future access to an external memory store (i.e., the opportunity to offload) influenced susceptibility to insertion (Kelly & Risko, 2019a; 2019b; Lu et al., 2020).

The reported experiments followed the same general procedure as that of Risko et al. (2019). On each trial, participants were shown a list of to-be-remembered words, one a time, and had to type them into a computer file that they were instructed would be available during test (which was always the case). Participants then completed a simple arithmetic distractor task. During the recognition test in Experiment 1 or the recall test in Experiments 2 and 3, participants were given access to their saved file to consult if desired. The procedure was the same for the first three trials, to develop a sense of trust in and familiarity with the external memory store. On the fourth trial, the researcher surreptitiously inserted a word in the middle position of the participant's saved list in the time between the encoding task and retrieval (i.e., while participants completed the distractor task). Participants then completed their recognition/recall test on the fourth trial. In Experiments 1 and 2, diverging from Risko and colleagues (2019) in which the task ended after this fourth trial, we explicitly notified participants that this manipulation of their external memory store had taken place. Critically, participants then completed a fifth trial, similar to the fourth trial. That is, we again inserted an item in the middle position of the participant's external memory store while they performed the distractor task between the encoding task and recognition/recall test. Thus, this fifth trial took place when participants knew that the reliability of their external store was compromised. In Experiment 3, after the third trial, we warned half of the participants that their external store would not be available at test, although they were still to type the words at encoding.

The critical question is whether individuals endorse the inserted item as having been presented during encoding and, further, whether the likelihood of this endorsement differs following being apprised of the external memory store's vulnerability to manipulation (Experiments 1 & 2) or future inaccessibility (Experiment 3). Based on previous research (Storm & Stone, 2015; Weis & Wiese, 2019), we predicted that when participants are told that their external memory store could be manipulated or that it will be inaccessible, they should be less susceptible to a manipulation of their external store. Also of interest is the form that this putative decrease in susceptibility might take. For example, this reduced susceptibility might emerge as a decrease in endorsement for all items (e.g., a kind of general skepticism or bias against the external store) or a more specific increase in the likelihood that the inserted item is detected as such (e.g., increased sensitivity). In addition to endorsement, we also assessed participants' ability to pick out the inserted item on the last trial, and self-reports of strategies employed (from internal memory reliance to external store reliance).

Experiment 1

In Experiment 1 (preregistered at https://osf.io/cm9fq), participants performed the tasks described above. The retrieval test was a modified recognition test wherein participants were presented with each study word (i.e., originally presented during encoding), and, on Trials 4 and 5, the inserted item as well—the only foil. After Trial 4, participants were told about the insertion of the item into their external store (i.e., their typed list) and asked whether they noticed. After Trial 5, participants were first asked about the offloading strategy that they employed wherein they rated on a scale from 1–5 the extent to which they relied on their typed list (i.e., the external store) versus their internal memory. Participants were then asked whether they noticed if a word was inserted on Trial 5 and finally, asked to select a word from their external memory store (i.e., list) that they thought was most likely to have been inserted. Data and materials for Experiment 1 are available at https://osf.io/xzw4t/.

Method

Participants

Data from 32 participants were collected based on an a priori power analysis with the desired power of .80 ($\alpha = .05$, two-tailed) to detect a medium sized effect in participants' confidence in the inserted word from Trial 4 to Trial 5 (see *Confidence* below for details). Participants were undergraduate psychology students at the University of Waterloo participating for course credit. Data from two participants were replaced due to incomplete data.

Apparatus

Both the participant and researcher were seated in the same room with a divider separating their workstations. At the participant's workstation were two computers and two monitors, one to display the instructions and task (display monitor), and the other used to create and save their typed lists (workspace monitor). These monitors were connected to the computers and monitors at the researcher's workstation to remotely control them and to observe the participant's progression through the experiment (this was not made explicit to the participants, however). At the researcher's workstation were three computers with three corresponding monitors displaying each of the two monitors from the participant's workstation; and one was used to covertly access the participant's list and to insert a word when required.

Stimuli

Five lists were created using the SenticNet 4 word corpus (Cambria et al., 2016). The lists were counterbalanced across trial position. The word lists varied in lengths (i.e., 15, 17, 19, 21, 23) so that when participants progressed to the insertion trials (Trials 4 and 5), a one-word insertion would not be easily detectable by counting. The inserted words were yoked to specific word lists, such that each list had the same designated word as the inserted word (see Appendix for the word lists). Whenever a list was presented on Trials 4 or 5, the designated inserted word for that list was inserted in the middle position of the list. The word lists presented for the non-manipulated trials (Trials 1-3) did not include its yoked inserted word, and thus, list lengths were 14, 16, 18, 20, and 22. In the analysis, the inserted item was compared to a control item, which was the word presented directly before it, or in rare cases where that item was not stored, its preceding item was used as the control. The control item was chosen to be the immediately preceding item to approximately control for serial position and to avoid the item following the inserted item. The latter could be problematic if participants notice the inserted item. Within and between each word list (including the inserted words), words were not meaningfully different in length or frequency, with median word lengths of 6 to 7 letters and median list frequencies ranging from 258–764 (using frequency count from SUBTLEX-US; Brysbaert & New, 2009). At encoding, words were presented visually in the centre of the screen in Arial font and each word was presented for 5 s with a 1-s interstimulus interval.

Post-Trial 4 notification question

After completing the recognition test for the first word-insertion trial (Trial 4), participants responded to a question which asked, "During the arithmetic task, we typed *"[inserted word]"* into your text file. Did you notice?"

Post-task questionnaire

Upon the completion of the second word-insertion trial, Trial 5, participants were asked three questions specific to that final trial. Question 1 asked, "Please select the option that best describes your recognition strategy during the final (fifth) trial of this study." Participants had six options to choose from, including: (a) I relied exclusively on my typed list during the recognition test, (b) I relied mostly on my typed list during the recognition test, (c) I relied about equally on both my list and my internal memory during the recognition test, (d) I relied mostly on my internal memory during the recognition test, (e) I relied exclusively on my internal memory during the recognition test, and (f) None of the above. Question 2 asked, "On the last trial we may have added a word to your typed list that was not presented originally. Please respond yes or no as to whether you believe we inserted a word into your list on your final trial." Question 3 stated, "Please open up your last list. Please review this list and type out a word you think was inserted. Even if you don't think something was added, please guess." Participants were shown their final manipulated list to refer to for Question 3.

Procedure

Participants were seated at their workstation, approximately 50 cm in front of two adjacent monitors (display and workspace monitors). Each trial began with an encoding task, in which one word at a time was presented in white on a grey background. As each word was presented on the right display monitor, participants simultaneously typed each word into a text file on the left workspace monitor. On the rare occasion that a participant missed writing a word, they would not have the opportunity for it to be presented again. After the encoding task was complete, participants were asked to save their ".txt" file on the left workspace monitor. With their saved list now closed, participants completed a 30-s arithmetic distractor task

Table 1. Experiment 1: Means (SDs) of all dependent variables.

	Trial 1	Trial 2	Trial 3	Trial 4 (pre-notification)	Trial 5 (post-notification)
Control confidence	3.97 (0.26)	3.95 (0.38)	3.99 (0.16)	4.00 (0.00)	3.94 (0.25)
Inserted confidence	–	–	–	3.78 (0.75)	3.16 (1.27)
Control endorsement	0.99 (0.08)	0.98 (0.13)	0.99 (0.06)	1.00 (0.00)	1.00 (0.00)
Inserted endorsement	–	–	–	0.94 (0.25)	0.72 (0.46)
Notification question	–	–	–	0.19 (0.40)	–
Strategy	–	–	–	–	3.16 (1.02)
Think inserted	–	–	–	–	0.66 (0.48)
Guess accuracy	–	–	–	–	0.34 (0.48)

Note: Dependent variables (Confidence, Endorsement, Post-Trial 4 notification question, Post-task questions 1-3; Strategy, Think inserted, Guess accuracy) are reported across the various conditions in Experiment 1. For Trials 1-3, the control confidence and endorsement are mean values for all encoded items. For Trials 4-5, the control confidence and endorsement are means of the one control item.

on the right display monitor, which asked them to answer "true" or "false" to simple arithmetic equations.

After the distractor task, participants were instructed to open their ".txt" file on the workspace monitor and to complete a recognition test on the display monitor, using the list as an aid if they chose to. During each recognition test, participants were asked to provide a confidence rating for each word one at a time, corresponding to whether they believed each word in the recognition test was presented during the encoding task. For each word, participants provided a confidence rating of (1) definitely not presented during encoding, (2) probably not presented during encoding, (3) probably presented during encoding, or (4) definitely presented during encoding. There was no time limit. Three trials were completed in this manner. No items were inserted on Trials 1-3, thus, all the words presented in the recognition test were targets.

Critically, on the fourth trial, while participants were completing the arithmetic distractor task, the researcher used one of the monitors at their workstations to covertly access the participant's saved, closed list, and to insert a word into the middle position of that list. This took place undisclosed to participants, and their display monitor did not change while the researcher altered the contents of the file it held. When opening their saved list for the recognition test, participants unknowingly accessed this now manipulated list. Participants then performed the recognition test for Trial 4, on which the inserted item was presented as a foil. After the recognition test, participants answered the Post-Trial 4 notification question, the wording of which informed them that their external memory store was vulnerable to manipulation. Participants then completed the fifth (final) trial which included the same insertion manipulation as Trial 4. Participants subsequently answered Questions 1, 2, and 3 from the Post-task questionnaire. To conclude, the researchers debriefed the participants about the true purpose of the study and reason for deception.

Results

Descriptive data from Experiment 1 are available in Table 1. All mixed-effects models reported throughout were conducted using the *lme4* package (Bates et al., 2015). Interactions among the fixed factors were also included in the model when appropriate—as indicated in the preregistered analyses. We included intercepts for participant as a random effect unless otherwise specified. In the case that models resulted in singular fits, this factor was removed. When an interaction term is not significant, we report results with and without it in the model. As described earlier, responses to the inserted item were compared to a control item (an actually presented item), which was the word presented directly before the inserted item was placed, or in rare cases where that item was not encoded, the control item was directly preceding that item. Lastly, when a non-pre-registered analysis is conducted, we refer to it in text as exploratory.

Endorsement

Endorsement was calculated by dichotomizing confidence responses. If participants responded "1" or "2" (i.e., *definitely* or *probably not presented during encoding*), this was considered a "no" response (i.e., not endorsed), whereas if they responded "3" or "4" (*probably* or *definitely presented during encoding*), this was considered a "yes" response (i.e., endorsed).

As can be seen in Figure 1, mean endorsement on Trials 4 and 5 were both 1.00 for the control items; for inserted items, they were .94 and .72 respectively. We analyzed the effect of notifying participants of the unreliability of their external memory store by comparing responses on Trial 4 and Trial 5 on endorsement for each item type (inserted vs. control) using separate McNemar's Chi-squared tests with a continuity correction. There was a statistically significant difference in the endorsement of the inserted item across Trials 4 and 5, $\chi2(1) = 4.00$, $p = .046$, such that the inserted item was endorsed more on Trial 4 than Trial 5. Because participants endorsed the control item 100% of the time on both Trials 4 and 5, no statistical analysis is reported.

We also analyzed the effect of item type (inserted vs. control) on endorsement separately for each trial (Trial 4 vs. Trial 5) using the same statistical test. There was no statistically significant difference in the endorsement of control and inserted items on Trial 4, $\chi2(1) = 0.50$, $p = .480$, but there was on Trial 5, such that inserted

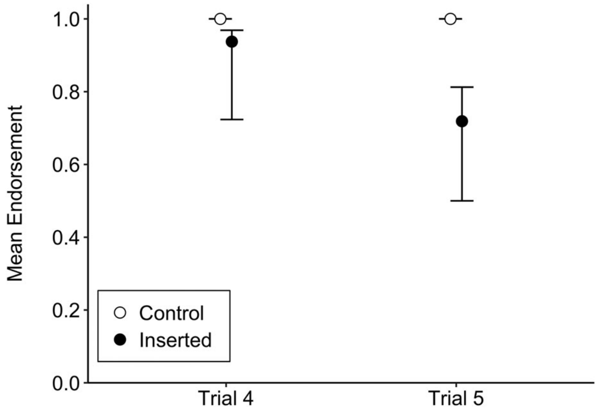

Figure 1. Mean percentage of endorsement for control and inserted items across Trial 4 and 5. Error bars are bias-corrected accelerated bootstrap 95% confidence intervals using 10,000 replications. There are no error bars for the control items, as they were 100% endorsed.

items were endorsed significantly less often than control items, χ2(1) = 7.00, *p* = .008. We preregistered a mixed-effects logistic regression to test the interaction between the effects of item type (inserted vs. control) and trial (Trial 4 vs. Trial 5) on endorsement with random intercepts for participant, however, this model failed to converge and as such no results are reported.

Confidence

We also analyzed confidence ratings as a continuous variable (using the entire 1–4 scale). In Table 2, the confidence scale is presented with the proportion of participants reporting each rating (1, 2, 3, or 4) for the inserted and control items in Trials 4 and 5. An exploratory within-subject Analysis of Variance (ANOVA) was conducted to examine the effects of trial (Trial 4 vs. Trial 5) and item type (inserted vs. control) on confidence ratings. The results revealed main effects of trial, $F(1, 31) = 9.59$, $p = .004$, $\eta^2_G = .052$, and of item type, $F(1, 31) = 12.40$, $p = .001$, $\eta^2_G = .103$, and a significant interaction between trial and item type, $F(1, 31) = 5.07$, $p = .003$, $\eta^2_G = .035$. Using pre-registered paired-samples t-tests, confidence ratings for inserted items were significantly higher on Trial 4 ($M = 3.78$, $SD = 0.75$) than on Trial 5 ($M = 3.16$, $SD = 1.27$), $t(31) = 2.69$, $p = .011$, $d = 0.48$. For control items, there was no significant difference in the confidence ratings between Trial 4 ($M = 4.00$, $SD = 0$) and Trial 5 ($M = 3.94$, $SD = .25$), $t(31) = 1.44$, $p = .161$, $d = 0.25$. When analyzing the effect of item type separately for Trials 4 and 5, there was no significant difference in the confidence ratings for control ($M = 4.00$, $SD = 0$) and inserted ($M = 3.78$, $SD = 0.75$) items on Trial 4; $t(31) = 1.65$, $p = .109$, $d = 0.29$. On Trial 5, confidence was significantly lower for inserted items ($M = 3.16$, $SD = 1.27$) than for control items ($M = 3.94$, $SD = 0.25$), $t(31) = 3.37$, $p = .002$, $d = 0.59$. A mixed effects regression with random intercepts for participant was conducted to examine the interaction between effects of trial (Trial 4 vs. Trial 5) and item type (inserted vs. control) on confidence ratings and revealed the interaction to be significant, $b = -0.56$, $SE = 0.25$, $t = -2.22$, $p = .029$.

Table 2. Experiment 1: Proportions of confidence ratings (1, 2, 3, or 4) for control and inserted items on trials 4 and 5.

		Rating			
Trial	Item	1	2	3	4
4	Control	0	0	0	1
	Inserted	0.06	0	0.03	0.91
5	Control	0	0	0.06	0.94
	Inserted	0.22	0.06	0.06	0.66

Post-task questionnaire

For means of responses to the Post-Trial 4 notification question and Post-task Questions 1 (strategy; 0: completely external – 5: completely internal), 2 (think inserted; 0: no; 1: yes), and 3 (guess accuracy; 0: incorrect guess; 1: correct guess), see Table 1. In a series of regressions, we used individuals' reported strategy at retrieval on Trial 5 as a predictor of whether they endorsed the inserted item on Trial 5 (logistic regression), their confidence (1-4) for the inserted item on Trial 5 (linear regression), whether they thought a word had been inserted on Trial 5 (logistic regression), and whether they correctly selected the inserted word on Trial 5 when asked (logistic regression). The overall mean self-reported strategy at retrieval was rated 3.16 ($SD = 1.02$) on a scale from 1 (exclusive reliance on the external list) to 5 (exclusive reliance on internal memory). Strategy was not a significant predictor of endorsement of the inserted item, $b = -0.92$, $SE = 0.54$, $z = -1.68$, $p = .092$, but did predict confidence, $b = -0.46$, $SE = 0.21$, $t = -2.16$, $p = .039$, such that the more external the recognition strategy reported, the higher the confidence rating for the inserted item. Strategy did not predict whether participants thought a word was inserted on Trial 5, $b = -0.51$, $SE = 0.38$, $z = 1.34$, $p = .180$, or whether they accurately guessed the inserted word, $b = 0.69$, $SE = 0.46$, $z = 1.52$, $p = .129$. These relations should be considered with caution in light of the small sample size in Experiment 1.

Discussion

Consistent with previous research, participants often failed to notice a word inserted into their external memory stores. Indeed, on Trial 4, 94% of participants responded "yes" (i.e., a 3 or 4 on the confidence scale) that the inserted item had been presented and they were highly confident in their endorsement (3.78 on a 4-point scale). Critically, both endorsement and confidence decreased on Trial 5, after participants were told that we had previously manipulated their external memory store, though both endorsement rate (72%) and confidence rating (3.16) remained high. The notice in between Trials 4 and 5 appeared to have no substantive effect on the control item (i.e., the item that was actually presented). That said, it is important to note that the control item on Trial

4 is close to ceiling, thus an increase, but not a decrease, in control endorsement would be impossible to detect. Nonetheless, the overall pattern is consistent with the notion that the effect of the notice was to increase individual's ability to discern the inserted item (i.e., foil) from actual target (control) items, rather than to a general skepticism about the contents of the external store. The strategy report results were mixed, but there was some limited evidence that a self-reported reliance on the external memory store was related to a higher confidence rating for the inserted item.

Experiment 2

In Experiment 2 (pre-registered at https://osf.io/3v7j2), we sought a conceptual replication of Experiment 1 using a modified recall test rather than a recognition test. One potential issue with using a recognition test is that participants can respond "yes" to the inserted item for reasons other than the presence of the inserted item in the external memory store. For example, individuals might have simply got into the habit of responding with a confidence rating of "4" (definitely presented during encoding) to all of the items, provided that almost all of the items (except the inserted item) were presented during the encoding phase. A free recall test does not suffer from this limitation. For this recall test, participants were provided with a text box in which they typed all of the words that had been presented on that trial. As in Experiment 1, during the recall test, participants could consult their saved lists (i.e., their external memory stores). Thus, the act of "recalling" the inserted word (i.e., typing it into the response box) would be unlikely, unless participants were actively endorsing the information in the external memory store.

In Experiment 2, we continued to collect confidence ratings, but given the change in memory test, these ratings took on a different meaning. That is, participants were asked to provide confidence ratings for all of the items they recalled. We again used a four-point scale but here each point corresponded to a percentage range of confidence that the item had been presented starting at above 50% (1: 51-60%; 2: 61-75%; 3: 76-94%; 4:95-100%). In addition to switching to a recall test, we also included a no-notice condition wherein participants did *not* receive notice of the insertion after Trial 4. Lastly, we collected a much larger sample than in Experiment 1 to increase power, and participants completed the study online, thus minor procedural changes from Experiment 1 were made to accommodate the online platform. Data and materials for Experiment 2 are available at https://osf.io/xzw4t/.

Method

Participants

160 participants were included in the study and recruited online (during the Covid-19 pandemic) using Amazon's Mechanical Turk and completed the study within one hour for $9.00 USD. All participants were over the age of eighteen. One participant was replaced due to incomplete data and sixty participants were replaced based on preregistered exclusion criteria (see below for details). The number of usable participants collected was based on increasing power from an unpublished recall experiment (https://osf.io/wk62f) to better detect critical interactions between notice and item type. Compared to Experiment 1, we roughly doubled our sample size for each condition present in Experiment 2.

Materials

The *Stimuli* and *Post-Trial 4 notification question* used were the same as in Experiment 1.

Confidence measure

Beside each word that they typed ("recalled"), participants were asked to provide a confidence rating corresponding to how much they believed it was presented to them in the encoding task. For each word that they recalled, participants provided a confidence rating of (1) possibly presented originally (i.e., between 51% and 60% chance it was presented), (2) moderately likely presented originally (i.e., between 61% and 75% chance it was presented), (3) very likely presented originally (i.e., between 76% and 94% chance it was presented), or (4) definitely presented originally (i.e., between 95% and 100% chance it was presented). There was a 5-min time limit for the recall test before the program automatically proceeded to the next task.

Post-task questionnaire

Upon completion of the second word-insertion trial, Trial 5, participants were asked three questions specific to that final trial. Question 1 asked, "Please select the option that best describes your recall strategy during the final (fifth) trial of this study." Participants had six options to choose from, including: (a) I relied exclusively on my typed list during the recall test, (b) I relied mostly on my typed list during the recall test, (c) I relied about equally on both my list and my internal memory during the recall test, (d) I relied mostly on my internal memory during the recall test, (e) I relied exclusively on my internal memory during the recall test, and (f) None of the above. Question 2 asked, "On the last trial we may have added a word to your typed list that was not presented originally. Please respond yes or no as to whether you believe we inserted a word into your list on your final trial." Question 3 stated, "Please open up your last list. Please review this list and type out a word you think was inserted. Even if you don't think something was added, please guess." Participants were shown their final manipulated list to refer to for Question 3.

Debriefing questionnaire

Not to be confused with the *Post-task questionnaire*, we administered a debriefing questionnaire to help ensure data quality from online collection. Specifically, at the end of the experiment, participants were asked three questions that we used to exclude participants. Question 1 asked, "Did you take any notes or write anything down while completing the task?" Question 2 asked, "Were you doing anything else while completing this task? (e.g., Netflix)." For Questions 1 and 2, the options of *yes* or *no* were provided in multiple-choice format. Question 3 asked, "Is there any reason we should or should not use your data? (It's okay if you think you weren't able to give it your best, just let us know)." The options of "feel free to use my data" and "don't use my data" were provided in multiple-choice format.

Procedure

Each trial began with an encoding task, in which one word at a time was presented in blue on a white background. Participants were told to type each word as it appeared in exactly the way it was presented. As each word was presented in the middle of the screen, participants had 6 s to type the word in a text box below it to "save it" on a list (counterbalanced to populate on the left or right side of the screen, at the participant level). After 6 s, participants were presented with the next word and their previously typed word was added to the list. This list was presented on the right or left side of the screen under the title "saved list." No special characters, numbers, or capitalizations that the participant typed would be translated to their saved list. If on the rare occasion participants missed writing a word, then they would not have the opportunity for it to be presented again and it would not be added to their saved list. After the encoding task was complete, participants had an opportunity to view their list for 10 s before it disappeared, and they moved on to the 30-s arithmetic distractor task, which had a time limit of 10 s per question.

After the distractor task, participants completed the recall test, during which they were presented with their saved list on the same side of the screen as it had been presented during encoding. In the middle of the screen, there was a text box and participants were instructed to only type ("recall") words that they thought were presented during the encoding task along with a confidence rating, using their saved list as an aid if they chose to. Participants were advised that if they thought a word had not been presented to them, they should not type ("recall") it in the text box. Three trials were completed in this way. On the fourth trial, when presented with their saved list at recall, it was presented with the inserted word halfway into their typed list, undisclosed to participants. Once participants completed the recall test for Trial 4, those in the notice condition were asked the Post-Trial 4 notification question, the wording of which informed them that their external memory store was vulnerable to manipulation. Those in the no-notice condition moved on to Trial 5 without any notice. Afterward, all participants completed the fifth and final trial, including the same manipulation as Trial 4. Participants subsequently answered Questions 1, 2, and 3 from the Post-task questionnaire, completed the debriefing questionnaire, and were debriefed on the true purpose of the study and the reason for deception.

Results

Descriptive data from Experiment 2 are presented in Table 3. Average confidence ratings reported are based only on items that were recalled. The single control item to be compared to the single inserted item was decided in the same manner as Experiment 1. Sixty participants were replaced based on not meeting any of the following preregistered criteria: (1) typing the word before the

Table 3. Experiment 2: Means (SDs) of all dependent variables.

Condition		Trial 1	Trial 2	Trial 3	Trial 4 (pre-notification)	Trial 5 (post-notification)
No-notice	Control confidence	3.57 (0.91)	3.80 (0.65)	3.81 (0.63)	3.81 (0.63)	3.77 (0.71)
	Inserted confidence	–	–	–	3.46 (1.06)	3.65 (0.85)
	Control recall	0.82 (0.38)	0.90 (0.30)	0.93 (0.25)	0.94 (0.24)	0.88 (0.33)
	Inserted recall	–	–	–	0.65 (0.48)	0.58 (0.50)
	Notification question	–	–	–	–	–
	Strategy	–	–	–	–	3.84 (1.12)
	Think inserted	–	–	–	–	0.55 (0.50)
	Guess accuracy	–	–	–	–	0.41 (0.50)
Notice	Control confidence	3.58 (0.88)	3.90 (0.39)	3.88 (0.40)	3.85 (0.46)	3.78 (0.56)
	Inserted confidence	–	–	–	3.52 (0.97)	3.03 (1.24)
	Control recall	0.85 (0.36)	0.93 (0.25)	0.92 (0.28)	0.93 (0.27)	0.90 (0.30)
	Inserted recall	–	–	–	0.78 (0.42)	0.41 (0.50)
	Notification question	–	–	–	0.40 (0.50)	–
	Strategy	–	–	–	–	3.78 (0.83)
	Think inserted	–	–	–	–	0.78 (0.42)
	Guess accuracy	–	–	–	–	0.58 (0.50)

Note: Experiment 2: Dependent variables (Confidence, Recall, Post-Trial 4 notification, Post-task question answers 1-3; Strategy, Think inserted, Guess accuracy) are reported across the various conditions. For Trials 1-3, the control confidence and recall are mean values for all encoded items. For Trials 4-5, the control confidence and recall are means of the one control item.

inserted word (used as the control) or the word before that, (2) typing at least 90% of the words they were supposed to on Trials 4 and 5 (the instruction was to write down 100% of the words), (3) accurately answering over 70% on the simple math problems during the arithmetic distractor task, (4) providing a confidence rating of 1-4, as instructed, to any recalled word (since our DVs include the confidence of that recalled item, but we are not able to infer it from no confidence rating or a rating outside of the range). In the debriefing questionnaire at the end of the experiment, participants were excluded from all analyses if they answered yes to any of the following: (1) doing something other than the task, (2) writing/screenshotting any words down during the encoding task, or (3) responding that we should not use their data. All mixed-effects models reported throughout were conducted in the same manner as outlined in Experiment 1.

Recall

The mean proportions of items recalled as a function of condition (no-notice and notice) and item type (control vs. inserted) are presented in Figure 2. A mixed effects logistic regression with the predictors notice condition (no-notice vs. notice), trial (Trial 4 and 5), and item type (inserted vs. control) revealed a three-way interaction, $b = -3.03$, $SE = 1.22$, $z = -2.49$, $p = .013$. Two separate regressions revealed a significant interaction between trial and item type in the notice condition, $b = -2.25$, $SE = 0.86$, $z = -2.62$, $p = .009$, but not in the no-notice condition, $b = 0.55$, $SE = 0.86$, $z = 0.64$, $p = .520$. When the interaction term in the latter model was removed, participants were significantly more likely to recall items on Trial 4 than on Trial 5, $b = -0.82$, $SE = 0.40$, $z = -2.06$, $p = .039$, and significantly more likely to recall the control item than the inserted item, $b = -3.50$, $SE = 0.60$, $z = -5.78$, $p < .001$. We next performed separate regressions on the inserted and control items in the notice condition. In the notice condition, for inserted items, recall was significantly higher on Trial 4 compared with Trial 5, $b = -2.17$, $SE = 0.55$, $z = -3.92$, $p < .001$. No significant difference was revealed for control items, $b = -1.17$, $SE = 1.16$, $z = -1.01$, $p = .310$.

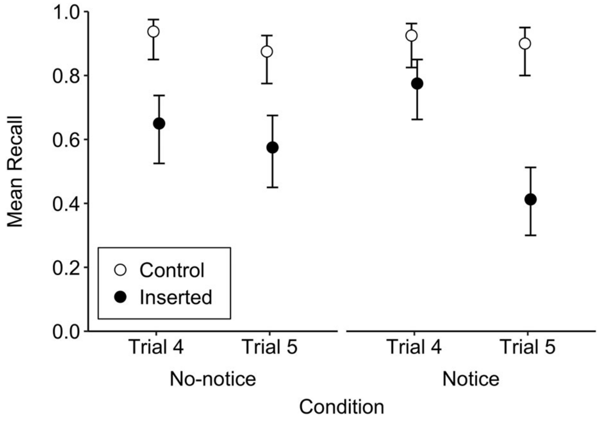

Figure 2. Mean percentage of recall for control and inserted items across Trial 4 and 5 for both notice conditions. Error bars are bias-corrected accelerated bootstrap 95% confidence intervals using 10,000 replications.

As is clear in Figure 2, on Trial 4 there seems to be a difference in individual's recollection of the inserted item, which is unexpected since these conditions do not differ until after Trial 4, when the notification takes place. To assess whether the recollection of inserted items on Trial 4 differed by condition, we performed an exploratory regression (i.e., not preregistered), which was a standard logistic regression since there was no within-subject factor. For inserted items in Trial 4, no significant difference was revealed for condition, $b = 0.62$, $SE = 0.36$, $z = 1.74$, $p = .083$. To assess whether the recollection of inserted items on Trial 5 differed by condition, we again performed an exploratory regression, which revealed that for inserted items on Trial 5, items in the notification condition were recalled significantly less than those in the no-notification condition, $b = -0.66$, $SE = 0.32$, $z = -2.05$, $p = .041$. Notably, while recall of the inserted item did reduce (in Trial 5) after the explicit notification that we had inserted an item (after Trial 4), 41% of participants in the notified condition still recalled the inserted word.

Confidence

In Table 4, the confidence scale is presented with the proportion of participants reporting each rating (1, 2, 3, or 4) for the inserted and control items in the last trial (Trial 5). For the following analyses, confidence was treated as a continuous variable. A mixed effects regression including notice condition (no-notice vs. notice), trial (Trial 4 and 5), and item type (inserted vs. control), revealed a three-way interaction, $b = -0.67$, $SE = 0.23$, $t = -2.93$, $p = .004$. Two separate regressions for the notice and no-notice conditions revealed a significant two-way interaction between trial and item type in the notice condition, $b = -0.44$, $SE = .18$, $t = -2.46$, $p = .015$, but no such interaction in the no-notice condition, $b = 0.21$, $SE = 0.14$, $t = 1.50$, $p = .135$. When the interaction term in the model was removed for the no-notice condition, participants did not differ in their confidence between Trials 4 and 5, $b = 0.01$, $SE = 0.07$, $t = 0.11$, $p = .911$, but reported significantly lower confidence in the inserted item than in the control item, $b = -0.25$, $SE = 0.07$, $t = -3.39$, $p < .001$. We next performed separate regressions for the effect of trial on inserted and control items in the notice condition. For inserted items, confidence was significantly lower on Trial 5 than Trial 4,

Table 4. Experiment 2: Proportions of confidence ratings (1, 2, 3, or 4) for control and inserted items on the last trial (Trial 5) across the notice and no-notice reliability conditions.

		Rating			
Condition	Item	1	2	3	4
No-notice	Control	0.04	0.03	0.04	0.89
	Inserted	0.06	0.05	0.06	0.83
Notice	Control	0.00	0.07	0.08	0.85
	Inserted	0.18	0.18	0.06	0.58

$b = -0.56$, $SE = 0.18$, $t = -3.18$, $p = .003$. No significant difference was revealed for control items, $b = -0.07$, $SE = 0.08$, $t = -0.94$, $p = .350$.

Post-task questionnaire

For means across the Post-Trial 4 notification question and Post-task Questions 1 (strategy; 0: completely external – 5: completely internal), 2 (think inserted; 0: no; 1: yes), and 3 (guess accuracy; 0: incorrect guess; 1: correct guess), see Table 3.

To assess whether the notification manipulation influenced self-reported strategy at retrieval, an exploratory Welch t-test was conducted. Recognition strategy at retrieval did not differ across the no-notification condition ($M = 3.84$, $SD = 1.12$) and the notification condition ($M = 3.78$, $SD = .0.83$), $t(145.43) = 0.40$, $p = .688$. Because both the normality and the homogeneity of variance assumptions were violated (p's < .05), a non-parametric test (Mann–Whitney-Wilcoxon Test) was also conducted and revealed similar results, $W = 3450$, $p = .400$. We also compared (again exploratory) responses across the no-notice and notice conditions for whether participants believed we had inserted an item on Trial 5, and their accuracy at guessing the inserted item on Trial 5, using separate Chi-squared tests with a continuity correction. There was a statistically significant difference in the belief of insertion across conditions, $\chi2(1) = 8.08$, $p = .004$, such that those in the notice condition more often reported believing that a word was inserted on Trial 5. There was no difference in the accuracy of guessing the inserted item on Trial 5 across conditions, $\chi2(1) = 3.60$, $p = .058$.

Next, in a series of regressions, we used individuals' reported strategy at retrieval on Trial 5 as a predictor of whether they recalled the inserted item on Trial 5 (using logistic regression), their confidence for the inserted item on Trial 5 (using linear regression), whether they thought a word had been inserted on Trial 5 (using logistic regression), and whether they correctly selected the inserted word for Trial 5 when asked (using logistic regression).

Participants reporting a more external strategy were (i) more likely to recall the inserted item (foil) than were those reporting a more internal strategy, $b = 0.75$, $SE = 0.19$, $z = 3.89$, $p < .001$, (ii) more likely to have a higher confidence rating for the inserted item when recalled, $b = 0.35$, $SE = 0.12$, $t = 2.88$, $p = .005$, (iii) less likely to report that a word was inserted, $b = -0.57$, $SE = 0.20$, $z = -2.91$, $p = .003$, and (iv) lower in their accuracy at guessing the inserted word, $b = -0.43$, $SE = 0.17$, $z = -2.49$, $p = .013$. It is important to note that within Trial 5, only those that recalled the inserted item *and* had a subsequent confidence rating (46/80 participants in the no-notice condition, and 33/80 participants in the notice condition) were included in the linear regression for confidence rating. Exploratory (not preregistered) regression analyses analogous to the four regressions listed above, but with both condition and a condition by recall strategy interaction as additional predictors (with recall strategy), revealed no interaction for any of the four regressions listed above (recall of inserted item, confidence in inserted item, reporting a word was inserted, accurately guessing the inserted word).

Discussion

Experiment 2 extends the main result of Experiment 1 to a modified recall test. That is, receiving notice that an external memory store was potentially unreliable reduced individuals' susceptibility to the acceptance of manipulated information in their external store. Consistent with Experiment 1, participants often failed to notice a word inserted into their external memory stores. Indeed, across conditions on Trial 4, a majority of participants (65% in the no-notice condition, 78% in the notice condition) recalled the inserted word and were confident that it had been previously presented (3.46/4 in the no-notice condition, 3.52/4 in notice condition). Critically, for those given notice of the previous manipulation, recall and confidence decreased significantly on Trial 5, such that 41% of participants recalled the inserted item and with reduced confidence when they did (3.03/4). While recall of the inserted item decreased after the notification of the insertion, still almost half of the participants failed to detect the insertion (41% recall the inserted word).

Consistent with Experiment 1, the notice between Trials 4 and 5 appeared to have no substantive effect on the endorsement and confidence rating of the control item (i.e., an actually presented item). This suggests that any effect of the notice was primarily to increase individuals' abilities to discern actually presented target/control items from the inserted item (i.e., the foil). Again, as in Experiment 1, control endorsement on Trial 4 was high and thus an increase in control endorsement might be difficult to detect. In the no-notice condition, there was a small general reduction in items recalled from Trial 4 to Trial 5 and participants were generally more likely to recall and have higher confidence in control than inserted items. This suggests that when given no notice of the manipulation, participants do not subsequently show an increased ability to discriminate between control and inserted items, but instead show evidence of overall reduced trust (or general skepticism) in the store. It is unclear, at this point, what the cause of that effect might be, though it is important to note that while individuals in the no-notice condition were never told of the manipulation on Trial 4, a word was nevertheless inserted.

The strategy report results demonstrated that self-reported reliance on the external memory store at retrieval was related to more recall of the inserted item, higher confidence in the inserted item, lower likelihood of thinking that a word was inserted, and lower accuracy in guessing the inserted item. Each of these results is consistent with the notion that higher self-reported reliance on one's internal/biological memory during retrieval leads to less susceptibility to manipulation of their external store. Interestingly, while those that were notified of the previous insertion were less likely to endorse the inserted item, there was no evidence via the self-reports that they relied less on the external store during retrieval.

Experiment 3

As suggested in the introduction (see also Risko et al., 2019), one reason that individuals might be susceptible to the manipulation of their external memory stores is that, when using such a store, they initially encode information poorly (Kelly & Risko, 2019a; 2019b; Lu et al., 2020) and/or believe that they did. This poor encoding might lead to a memory experience when retrieving information from the external store that is insufficient to detect the manipulation (i.e., one cannot tell the poorly encoded information that was actually presented from the inserted information). This kind of mechanism may provide one route through which the reliability manipulation in Experiments 1 and 2 has its effect. Specifically, if one comes to believe that their external store is unreliable, they may not offload the memories (i.e., forego storing them internally) to their external store (see Storm & Stone, 2015) and, instead, might encode the information more strongly into internal memory, leading to an increase in their ability to detect the manipulation.

To examine the link between encoding and susceptibility to external memory store manipulation, in Experiment 3 (pre-registered at https://osf.io/5uayq/) we manipulated individual’s expectation that they could rely on an external store during a future recall test. When participants are told not to expect access to their external memory store at recall, they recall more than when they are told to expect to have access to that store. Kelly and Risko (2019a; 2019b) argued that this offloading cost was due to a disengagement of effortful memorization of the list of to-be-remembered words when individuals believe they can rely on their external store. In Experiment 3, for the first three trials of the task, participants were given to-be-remembered words to type into a saved list and had access to this saved list to aid in recall. On the last trial, half of the participants were told that they would not have access to their saved list at recall and the other half were told to expect access to their saved list. Critically, everyone received access to their list at recall which included an inserted item (as in Experiments 1 and 2). Again, the main dependent variable of interest is the extent to which individuals recall the inserted item across these two conditions. If devoting more effort to encoding can decrease susceptibility to manipulation of the external store, then participants who do not expect access to their list at recall should be better able to detect an inserted item. Data and materials for Experiment 3 are available at https://osf.io/xzw4t/.

Method

Participants

The 160 participants included in the study were recruited online (during the Covid-19 pandemic) using Prolific and completed the study within one hour for £3.75. All participants were over the age of eighteen. 22 participants were replaced based on the same exclusion criteria used in Experiment 2.

Materials

The *Stimuli, Post-task questionnaire,* and *Debriefing questionnaire* used were the same as in Experiments 1 and 2. The *Confidence measure* used was the same as in Experiment 2.

Procedure

The first three trials of the experiment were the same as Experiment 2. Once participants completed the recall test for Trial 3, participants in the warned condition were told that they would not receive their next typed list at recall. Participants in the not-warned condition were told that, like the other trials, they would receive their next typed list at recall. On the fourth trial, everyone was presented with their typed list at recall, and it was presented with the inserted word halfway into their list, undisclosed to participants. Participants subsequently answered Questions 1, 2, and 3 from the Post-task questionnaire, completed the debriefing questionnaire, and were debriefed on the true purpose of the study and the reason for deception.

Results

Descriptive data from Experiment 3 are available in Table 5. Average confidence ratings reported are only for items that were recalled. The single control item compared to the single inserted item was the item preceding the inserted item, as in Experiments 1 and 2.

Recall

The mean proportions of items recalled as a function of warning condition (no-warning vs. warning) and item type (control vs. inserted) are presented in Figure 3. A mixed effects logistic regression with the predictors condition (no-warning and warning) and item type (inserted vs. control) revealed a two-way interaction between condition and item type, $b = -7.91$, $SE = 1.35$, $z = -5.85$, $p < .001$. We next performed separate regressions on the inserted and control items, which were both logistic regressions since there was no within-subject factor. For inserted items, recall was significantly higher in the no-warning condition compared with the warning condition, $b = -0.93$, $SE = 0.33$, $z = -2.83$, $p = .005$. No significant difference was revealed for control items, $b = -0.36$, $SE = 0.61$, $z = -0.60$, $p = .550$.

Table 5. Experiment 3: Means (SDs) of all dependent variables.

Condition		Trial 1	Trial 2	Trial 3	Trial 4 (post-warning)
No-warning	Control confidence	3.51 (0.97)	3.88 (0.52)	3.91 (0.45)	3.89 (0.42)
	Inserted confidence	–	–	–	3.42 (1.06)
	Control recall	0.83 (0.37)	0.91 (0.28)	0.91 (0.28)	0.94 (0.24)
	Inserted recall	–	–	–	0.66 (0.48)
	Strategy	–	–	–	2.19 (1.09)
	Think inserted	–	–	–	0.58 (0.50)
	Guess accuracy	–	–	–	0.54 (0.50)
Warning	Control confidence	3.51 (1.00)	3.83 (0.60)	3.87 (0.47)	3.92 (0.36)
	Inserted confidence	–	–	–	3.06 (1.19)
	Control recall	0.81 (0.39)	0.87 (0.33)	0.87 (0.34)	0.91 (0.28)
	Inserted recall	–	–	–	0.44 (0.50)
	Strategy	–	–	–	2.73 (1.11)
	Think inserted	–	–	–	0.66 (0.48)
	Guess accuracy	–	–	–	0.75 (0.44)

Note. Dependent variables (Confidence, Recall, Post-task question answers 1-3; Strategy, Think inserted, Guess accuracy) are reported across the conditions in Experiment 3. For Trials 1-3, the control confidence and recall are mean values for all encoded items. For Trial 4, the control confidence and recall are means of the one control item.

Confidence

In Table 6, the confidence scale is presented with the proportion of participants reporting each rating (1, 2, 3, or 4) for the inserted and control items in the last trial (Trial 4). For the following analyses, confidence was treated as a continuous variable. A mixed effects regression including warning condition (no-warning vs. warning) and item type (inserted vs. control), did not reveal a two-way interaction between condition and item type, $b = -0.37$, $SE = 0.19$, $t = -1.92$, $p = .055$. When the interaction term in the model was removed, participants were not significantly more likely to report higher confidence in a given condition, $b = 0.10$, $SE = 0.10$, $t = -1.02$, $p = .309$, but were significantly more likely to report lower confidence in the inserted than control item, $b = -0.64$, $SE = 0.10$, $t = -6.73$, $p < .001$.

Post-task questionnaire

For means across the notification question and Post-task Questions 1 (strategy; 0: completely external – 5: completely internal), 2 (think inserted; 0: no; 1: yes), and 3 (guess accuracy; 0: incorrect guess; 1: correct guess), see Table 5.

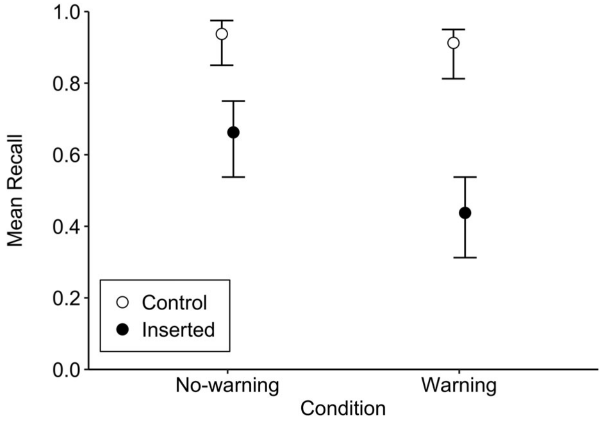

Figure 3. Mean percentage of recall for control and inserted items across both conditions. Error bars are bias-corrected accelerated bootstrap 95% confidence intervals using 10,000 replications.

To assess whether the warning manipulation influenced self-reported strategy at retrieval, an exploratory Welch t-test was conducted (i.e., not preregistered). Recall strategy significantly differed across the warning condition ($M = 2.73$, $SD = 1.09$) and the no-warning condition ($M = 2.19$, $SD = 1.11$), $t(157.94) = 3.08$, $p = .002$. Because the Shapiro–Wilk normality assumption was violated ($p < .05$), a non-parametric test (Mann–Whitney-Wilcoxon Test) was also conducted and revealed the same result, $W = 2317.5$, $p = .002$. We also compared (again not preregistered) responses across the no-warning and warning conditions for whether participants believed we had inserted an item on Trial 4, and their accuracy at guessing the inserted item on Trial 4, using separate Chi-squared tests with a continuity correction. There was no difference in the belief of insertion across conditions, $\chi2(1) = .95$, $p = .329$, on Trial 4. There was a statistically significant difference in the accuracy of guessing the inserted item on Trial 4 across conditions, $\chi2(1) = 6.98$, $p = .008$, such that those in the warning condition more often accurately guessed the inserted word on Trial 4.

Next, in a series of pre-registered regressions, we used individuals' reported strategy at retrieval on Trial 4 as a predictor of whether they recalled the inserted item on Trial 4 (using logistic regression), their confidence for the inserted item on Trial 4 (using linear regression), whether they thought a word had been inserted on Trial 4 (using logistic regression), and whether they correctly selected the inserted word for Trial 4 when asked (using logistic regression).

Table 6. Experiment 3: proportions of confidence ratings (1, 2, 3, or 4) for control and inserted items on the Last Trial (Trial 4) across the no-warning and warning conditions.

		Rating			
Condition	Item	1	2	3	4
No-warning	Control	0	0.04	0.03	0.93
	Inserted	0.11	0.09	0.06	0.74
Warning	Control	0	0.03	0.03	0.94
	Inserted	0.17	0.14	0.14	0.54

Table 7. Experiment 2: endorsement of control items across Trial 1–5 as a function of whether the participant noticed the inserted item on Trial 4.

	Overall Control Endorsement				
	1	2	3	4	5
Non-noticer	0.88 (0.25)	0.95 (0.17)	0.96 (0.14)	0.97 (0.08)	0.95 (0.12)
Noticer	0.74 (0.32)	0.84 (0.23)	0.82 (0.22)	0.82 (0.24)	0.83 (0.21)

Note. The status of noticer and non-noticer was based on Trial 4.

Participants who reported a more external strategy were more likely to recall the inserted item than were those reporting a more internal strategy, $b = -0.63$, $SE = 0.16$, $z = -3.91$, $p < .001$, and had lower accuracy in guessing the inserted word, $b = 0.36$, $SE = 0.16$, $z = -2.31$, $p = .021$. Recall strategy was not a significant predictor of confidence rating for the inserted item, $b = -0.20$, $SE = 0.12$, $t = -1.67$, $p = .098$, or for reporting that a word was inserted, $b = 0.28$, $SE = 0.15$, $z = 1.83$, $p = .067$. It is important to note that within Trial 4, only those that recalled the inserted item *and* had a subsequent confidence rating (53/80 participants in the no-warning condition, and 35/80 participants in the warning condition) were included in the linear regression for confidence rating. Exploratory (not preregistered) regression analyses analogous to the four regressions listed above, but with both condition and a condition by recall strategy interaction as additional predictors (with recall strategy), revealed no interaction for any of the four regressions listed above with the following dependent variables: recall of inserted item, confidence in inserted item, reporting a word was inserted, and accurately guessing the inserted word.

Discussion

Experiment 3 assessed whether expecting access to one's external memory store influences susceptibility to a manipulation of that store. Consistent with Experiments 1 and 2, participants often failed to notice a word inserted into their external memory stores. Indeed, a large percentage of participants (66% in the no-warning condition) recalled the inserted word and were confident that it had been previously presented (3.42/4 in the no-warning condition). The novel observation in Experiment 3 was that when participants were told that they would not have access to their external store, recall and confidence in the inserted word lessened, such that only 44% recalled the inserted item and with less confidence (3.06/4) than in the no-warning condition. This result is consistent with the idea that investing more effort during encoding (because individuals believed they could not rely on an external store) can protect one against manipulation of one's external memory store (when, in this case, it becomes unexpectedly available). This might be because better encoded items are more easily discriminated from the inserted item and/or more effort at encoding leads to a greater expectation that items feel familiar at retrieval. Another interesting possibility is that the surprise availability of their list made individuals who were warned that they would not have their list more skeptical and thus more willing to accept that the experimenter might have manipulated their list. The encoding manipulation appeared to have no substantive effect on the control item, suggesting that any effect of the manipulation was primarily to increase individuals' ability to discern actually presented target/control items from the inserted item (i.e., the foil). Again, similar to Experiments 1 and 2, control endorsement on Trial 4 was high and an increase might be difficult to detect. The participants' self-reports demonstrated that on average, individuals in the no-warning condition reported relying more heavily on their saved list than on their internal memory during the final recall test and that self-reported reliance on the external memory store was related to more recall of the inserted item and lower accuracy in guessing the inserted item.

The relation between endorsement of control and inserted items (exploratory analysis)

In the following exploratory analysis, we examine the relation between control performance and the detection of inserted items. In particular, we focus on control performance as a function of whether a given participant detected the inserted item and focus on Experiments 2 and 3 provided the larger sample sizes.

In Table 7, overall control performance (i.e., the proportion of control items endorsed by the participant) is presented as a function of whether the participant noticed the inserted item (i.e., identified it as having not been presented) on Trial 4 in Experiment 2. As can be seen in Table 7, participants that noticed the inserted item also appear less likely to endorse control items. This was true even prior to Trial 4. To confirm, we conducted a two-sample t-test comparing control performance on trials 1, 2, and 3 across non-noticers ($M = .93, SD = .14$) and noticers ($M = .80, SD = .22$). The difference was significant, $t(61.07) = 3.82$, $p < .001$. A similar pattern was present in Experiment 3. As can be seen in Table 8, participants that noticed the inserted item also appear less likely to endorse control items. To confirm, we conducted a two-sample t-test comparing control performance on trials 1, 2, and 3 across non-noticers ($M = .92$, $SD = .14$) than noticers ($M = .78$, $SD = .25$). The difference was

Table 8. Experiment 3: all control items endorsement per trial.

	Trial			
	1	2	3	4
Non-noticer	0.87 (0.26)	0.94 (0.17)	0.96 (0.10)	0.97 (0.08)
Noticer	0.73 (0.35)	0.80 (0.28)	0.80 (0.28)	0.81 (0.26)

significant, $t(105.84) = 4.33$, $p < .001$. This was also true if we considered only those participants in the no-warning condition (non-noticer: $M = .92$, $SD = .16$; noticers: $M = .81$, $SD = .23$; $t(38.63) = 2.17$, $p = .036$). Overall, it appears as though participants most likely to detect an item inserted into their external store are those less likely to endorse items that were actually presented. This might suggest a more conservative criterion for "yes" responses in these participants. Interestingly, while this approach improves detection of the inserted item (i.e., leads to its correct rejection) it also impairs performance on control items (i.e., leads to more misses).

General discussion

Using external aids to offload cognitive demands has long been a memorial strategy allowing us to evade the limitations of our internal/biological memory (Clark, 2010a; Donald, 1991; Nestojko et al., 2013; Risko & Gilbert, 2016). There are costs, however, to allocating memory demands to external locations. Here we focused on one such cost, originally reported by Risko and colleagues (2019), that individuals are susceptible to manipulation of their external memory stores. In the present investigation, we again found that a large percentage of participants did not notice a manipulation of their external memory store. This basic result replicates and extends Risko and colleagues (2019). We also found two manipulations that reliably influenced one's susceptibility to such manipulation: first, when individuals were given explicit notification that we had previously manipulated their external memory store, and second, when we told participants not to expect access to their external store at recall. In both situations, individuals were better able to detect a manipulation of their external memory store. In addition, neither of these manipulations appeared to compromise how participants endorsed the original (i.e., legitimate, not inserted) contents. This pattern of results puts an important constraint on understanding how these manipulations influence the decision to endorse information in one's external memory store. That is, neither manipulation appeared to lead to a general unwillingness to accept items in the external store, an influence that would have decreased endorsement of control items as well. Still, even with an explicit notification of an insertion or a presumably better encoded list of words (because they did not expect access to their external memory store), many participants (i.e., > 40%) remained unable to discriminate target words from words inserted into their external memory stores.

Each participant also provided a self-report rating of their reliance on their internal memory versus external memory store at retrieval. If participants were to rely on their internal memory, then one could imagine that they would be better equipped to not endorse the inserted item. Overall, the relations between self-reported strategy at retrieval and the various measures of one's susceptibility to the manipulation of their external store reported here seems consistent with this idea. That is, in Experiments 2 and 3, reported strategy at retrieval was a significant predictor of endorsement of the inserted item and accurately guessing the inserted word (these effects were in the same direction but not significant in Experiment 1, which had a smaller sample and used recognition instead of recall as the test). Thus, those who self-report being more reliant on their internal stores were less susceptible to manipulation of their external stores. Interestingly, strategy at retrieval did not differ across the reliability conditions in Experiment 2 but did differ across the expected access conditions in Experiment 3. This result might suggest that the two manipulations are reducing susceptibility to external storage manipulation via different mechanisms.

While the self-reported strategy at retrieval data are interesting, it is important to note that individuals may not be able to accurately assess the extent to which they relied on their internal versus external stores. In addition, given that the self-report questions followed the retrieval phase, participants' retrieval performance (e.g., last trial) could have influenced their answers to these questions. For example, participants may have successfully detected the inserted item and because of this, reported relying on their internal memory or vice versa. An alternative approach to indexing individual differences in reliance on an external store by self-report could involve more indirect methods (e.g., pupil dilation during encoding). The exploratory analysis of the relation between control and inserted item endorsement also provides some preliminary insight into individual differences in detecting an item inserted into one's external store. Namely, participants that were less likely to endorse control items (i.e., items that were actually presented) were more likely to correctly reject the inserted item.

Routes to reliability

The manipulation of reliability in the reported experiments is one of the few ways in which reliability has been manipulated in the literature thus far (Storm & Stone, 2015; Weis & Wiese, 2019). Despite the differences in how the manipulations were implemented, the effects on behaviour were similar (e.g., decreased offloading with reduced reliability; decreased susceptibility to external store manipulation). Nonetheless, it is clear that one can come to not trust an external store to perform a cognitive task in different ways. It will be interesting in future research to compare these different types of violations of trust or reliability manipulations directly.

Understanding endorsement when memory is distributed: a metacognitive approach

Regardless of whether information is stored internally or externally, upon retrieval of that information, participants must decide whether to endorse it or not (Arango-Muñoz, 2013). One approach to understanding this problem is in the context of metacognitive monitoring and control. For example, as noted in the Introduction, Goldsmith and Koriat's (1999) schematic model of free report memory performance captures a similar problem to that facing participants here. In their model, monitoring processes produce a subjective sense of the correctness of a retrieved candidate answer (i.e., the assessed probability). This output is then compared to a response threshold that is influenced by information regarding situational demands and incentives, in order to come to a decision about whether to report the candidate answer.

How might one extend this kind of model to the endorsement problem and, more generally, to contexts wherein individuals' "memory" is distributed across internal and external spaces? As a first pass, an item inserted into an external store could be thought to yield little evidence from a monitoring process that the item was previously presented. For example, there is little reason to expect that the inserted items would feel familiar. Nevertheless, the inserted items are often endorsed. In Goldsmith and Koriat's (1999) framework, this might reflect the control mechanism enacting a low threshold, given the situational demands are such that the inserted item is present in what has proven to be a reliable external store. From this perspective, a decrease in the perceived reliability of the external store (Experiments 1 and 2) might raise this threshold. As long as this threshold is not raised too high and/or the assessed probability output from the monitoring process is high for most control items, this should lead to a decrease in reports of the inserted item without much of an effect on control items. The influence of the expectation of access to the external store, putatively encoding effort, could arguably have a similar effect. That is, following encoding items deeply, one might adopt a higher threshold for accepting items as legitimate.

A different perspective is that the reliability manipulation and/or expectation of access manipulation influences the output of the monitoring process as opposed to the response threshold. As detailed previously, better encoding would serve to improve memory for the control (actually presented) items. This could enhance the experienced difference between the inserted item and the control items, thus improving detection performance. Returning to Goldsmith and Koriat's (1999) framework, this would require that the monitoring process be able to consider information that would capture this difference (e.g., relative familiarity). A related idea would be that improved encoding of control items leads to the retrieval of information that decreases the assessed probability that the inserted item had been presented. For example, the inserted item here is placed in between items that would have originally been presented in sequence. Thus, the assessed probability that an item was legitimate might be exceptionally low if participants recollect studying two items in sequence that now have an additional unfamiliar item between them. Placing the influence of the reliability and expectation of access manipulations in the monitoring mechanism leaves open the question of how an item's presence in an external store influences the decision to endorse the inserted item. From this perspective, all items that appear in an external store might receive a kind of "boost" to the assessed probability that is outputted to the control mechanism and compared to some threshold. Of course, these described means of grounding the present findings within a monitoring and control framework are not mutually exclusive. Furthermore, the influence of each of the manipulations used here might well affect different parts of such a model. Tentative evidence that this might be the case is available in the different effects that the manipulations had on strategy reports (i.e., the reliability manipulation did not influence retrieval strategy whereas the expected access manipulation did). Future work aimed at further refining our understanding of how individuals approach solving the endorsement problem from this perspective would be valuable.

Conclusion

Offloading memory to external stores is a critical strategy allowing us to evade the limitations of our internal memory. One cost of this approach is that it potentially exposes our "memories" to manipulation, provided that they reside out in the proverbial open. The present research reinforces this idea (replicating and extending the original work), as most participants failed to notice a manipulation of their external store, and also demonstrates that an explicit notification of either a previous manipulation or the future inaccessibility of our external memory store can decrease this susceptibility. These manipulations did not appear to influence the likelihood that individuals endorsed legitimate information in their external stores, thus placing a potential constraint on explanations of their influence. Lastly, individual variation in reliance on one's external store (self-reported) and willingness to endorse control items (i.e., items that were actually presented) were both demonstrated to be related to susceptibility to accepting illegitimate information inserted into one's external store. The latter two results suggest that investigating individual differences in managing memory in distributed contexts would be a fruitful future direction. In a technologically advanced age, in which a large amount of to-be-remembered information is externally stored, understanding the associated risks (in addition to associated benefits) is crucial to using our distributed memory systems efficiently. The present investigation has added to our understand of one such risk.

Disclosure statement

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

Funding

This work was supported by the Canada Research Chairs program: [Grant Number 950-232147]; a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (NSERC): [Grant Number 04091]; Canada Foundation for Innovation and

Ontario Research Fund: [Grant Number 37872]; an Early Researcher Award from the Province of Ontario: [Grant Number 37872].

References

Arango-Muñoz, S. (2013). Scaffolded memory and metacognitive feelings. *Review of Philosophy and Psychology*, *4*(1), 135–152. https://doi.org/10.1007/s13164-012-0124-1

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

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990. https://doi.org/10.3758/BRM.41.4.977

Cambria, E., Poria, S., Bajpai, R., & Schuller, B. (2016). *Senticnet 4: A semantic resource for sentiment analyses based on conceptual primitives*. Proceedings of COLING 2016, the 26th international conference on computational linguistics: Technical papers, pp. December. 2666–2677.

Clark, A. (2010a). *Supersizing the mind*. Oxford University Press.

Clark, A. (2010b). Memento's revenge: The extended mind, extended. In R. Menary (Ed.), *The extended mind* (pp. 43–66). MIT press.

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science*, *19*(1), 51–579. https://doi.org/10.1177/0963721409359277

Donald, M. (1991). *Origins of the modern mind: Three stages in the evolution of culture and cognition*. Harvard University Press.

Eskritt, M., & Ma, S. (2014). Intentional forgetting: Note-taking as a naturalistic example. *Memory and Cognition*, *42*(2), 237–246. https://doi.org/10.3758/s13421-013-0362-1

Ferguson, A. M., McLean, D., & Risko, E. F. (2015). Answers at your fingertips: Access to the Internet influences willingness to answer questions. *Consciousness and Cognition*, *37*, 91–102. https://doi.org/10.1016/j.concog.2015.08.008

Goldsmith, M., & Koriat, A. (1999). The strategic regulation of memory reporting: Mechanisms and performance consequences. In D. Gopher, & A. Koriat (Eds.), *Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application* (pp. 373–400). MIT Press.

Hutchins, E. (1995). How a cockpit remembers its speeds. *Cognitive Science*, *19*(3), 265–288. https://doi.org/10.1207/s15516709cog1903_1

Kelly, M. O., & Risko, E. F. (2019a). The isolation effect when offloading memory. *Journal of Applied Research in Memory and Cognition*, *8*(4), 4. https://doi.org/10.1016/j.jarmac.2019.10.001

Kelly, M. O., & Risko, E. F. (2019b). Offloading memory: Serial position effects. *Psychonomic Bulletin and Review*, *26*(4), 1347–1353. https://doi.org/10.3758/s13423-019-01615-8

Kelly, M. O., & Risko, E. F. (2021). Revisiting the influence of offloading memory on free recall. *Memory & Cognition*. https://doi.org/10.3758/s13421-021-01237-3

Lewandowsky, S., Mundy, M., & Tan, G. P. A. (2000). The dynamics of trust: Comparing humans to automation. *Journal of Experimental Psychology: Applied*, *6*(2), 104–123. https://doi.org/10.1037//1076-898x.6.2.104

Lu, X., Kelly, M. O., & Risko, E. F. (2020). Offloading information to an external store increases false recall. *Cognition*, 104428. Advance online publication. https://doi.org/10.1016/j.cognition.2020.104428

Lu, X., Kelly, M. O., & Risko, E. F. (2021). The gist of it: Offloading memory does not reduce the benefit of list categorisation. *Memory*, 1–16. https://doi.org/10.1080/09658211.2021.1989465

Muir, B. M., & Moray, N. (1996). Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics*, *39*(3), 429–460. https://doi.org/10.1080/00140139608964474

Nestojko, J. F., Finley, J. R., & Roediger, H. L. (2013). Extending cognition to external agents. *Psychological Inquiry*, *24*(4), 321–325. https://doi.org/10.1080/1047840X.2013.844056

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences*, *20*(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002

Risko, E. F., Kelly, M. O., Patel, P., & Gaspar, C. (2019). Offloading memory leaves us vulnerable to memory manipulation. *Cognition*, *191*, https://doi.org/10.1016/j.cognition.2019.04.023

Scoboria, A., Lynn, S. J., Hessen, J., & Fisico, S. (2007). So that's why I don't remember: Normalising forgetting of childhood events influences false autobiographical beliefs but not memories. *Memory*, *15*(8), 801–813. https://doi.org/10.1080/09658210701685266

Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. *Science*, *333*(6043), 776–778. https://doi.org/10.1126/science.1207745

Sterelny, K. (2004). Externalism, epistemic artefacts, and the extended mind. In R. Schantz (Ed.), *The externalist challenge* (pp. 239–254). Walter de Gruyter.

Storm, B. C., & Stone, S. M. (2015). Saving-enhanced memory: The benefits of saving on the learning and remembering of new information. *Psychological Science*, *26*(2), 182–188. https://doi.org/10.1177/0956797614559285

Weis, P. P., & Wiese, E. (2019). Using tools to help us think: Actual but also believed reliability modulates cognitive offloading. *Human Factors*, *61*(2), 243–254. https://doi.org/10.1177/0018720818797553

Appendix

	List				
	1	2	3	4	5
1	shoulder	carpet	sweep	highway	exercise
2	lunchtime	seat	gardener	gasoline	article
3	tree	venue	river	offer	walkway
4	colour	territory	trailer	paperwork	schoolyard
5	judgment	slush	train	amphibian	reptile
6	trashcan	recombine	percussion	early	point
7	kale	picnic	rainfall	gambling	engaged
8	home	alligator	seashore	cabbage	**theatre**
9	freshness	doors	unseen	**rabbit**	sculpture
10	stereo	body	**clock**	chase	beverage
11	nerves	**campground**	frozen	stone	pencil
12	**carpenter**	neighbour	rumour	timing	matrix
13	computer	dolphin	grain	sushi	horse
14	kidney	desk	drawing	store	mechanic
15	vein	camera	sidewalk	teen	broccoli
16	pickle	liquid	bean	patio	
17	liver	lawn	filter	squirrel	
18	keyboard	obstacle	stapler		
19	figure	windy	hidden		
20	couch	stint			
21	soon	centre			
22	atrium				
23	second				

Word lists of varying lengths (yoked inserted item bolded)