Offloading information to an external store increases false recall.

Xinyi Lu

Megan O. Kelly

Evan F. Risko

University of Waterloo

Corresponding Author: Xinyi Lu (xinyi.lu@uwaterloo.ca)

Abstract

Offloading to-be-remembered information is a ubiquitous memory strategy, yet in relying on external memory stores, our ability to recall from internal memory is often diminished. In the present investigation, we examine how offloading impacts true and false recall. Across three experiments, participants studied and wrote down word lists that were each strongly associated with an unstudied critical word. Recall in the Offloading condition (i.e., when they were told that they would have access to their written lists during recall) was contrasted with a No-Offloading condition (i.e., when they were told that they would *not* have access to their written lists during recall). We found that offloading decreased true recall of presented words while increasing false recall for unpresented critical words. Results are discussed in terms of offloading's differential effects on the formation of gist and verbatim traces during encoding.

Keywords: memory; cognitive offloading; false memory; recall

1. Introduction

In order to cope with the limitations of our internal memory, we have long turned to *offloading* to-be-remembered information to external memory stores (Clark, 2008; Clark & Chalmers, 1998; Nestojko, Finley, & Roediger, 2013; Risko & Gilbert, 2016). For example, one might write down a shopping list, or make a note of an important future appointment. While offloading to an external store can aid us in remembering, it reduces successful retrieval (relative to relying on internal memory) when the external store is unavailable (Eskritt & Ma, 2014; Kelly & Risko, 2019a, 2019b; Sparrow, Liu, & Wegner, 2011). Recent attempts to better understand the nature of this offloading effect (Eskritt & Ma, 2014; Kelly & Risko, 2019a, 2019b; Sparrow et al., 2011) have drawn parallels to directed forgetting and, in particular, have suggested a role for reduced top-down maintenance or rehearsal (i.e., selective rehearsal; Sheard & MacLeod, 2005) during encoding when participants expect to have access to external memory aids. In the present investigation we examine a prediction based on this idea in the context of offloading's influence on false recall.

One of the defining aspects of memory is our penchant for connecting related information. The inherently associative nature of our memories means that false memories for semantically related yet unencountered information can be common. For example, in the DeeseRoediger-McDermott (DRM) paradigm (Deese, 1959; Roediger & McDermott, 1995), participants study a list of words (e.g., *bed, rest, awake*) that are semantically related to an unpresented word (e.g., *sleep*), known as the critical lure. During a subsequent memory test, participants show substantial rates of false recall for the unstudied critical lure (Payne, Elie, Blackwell & Neuschatz, 1996; Roediger & McDermott, 1995). Researchers generally agree that the items on a DRM list activate semantically related concepts during encoding that are consistent with the theme or gist of the list (Reyna & Brainerd, 1995; Roediger, Watson, McDermott & Gallo, 2001). For example, according to fuzzy-trace theory (Reyna & Brainerd, 1995), two different types of traces are created in parallel during the encoding process: gist traces and verbatim traces. While the gist trace contains the general meaning or theme of the information, the verbatim trace contains item-specific, surface information (Brainerd, Wright, Reyna & Mojardin, 2001). Although the critical lure in the DRM paradigm is never actually presented, it is nonetheless repeatedly cued during encoding, and is consistent with the general theme of the list. Thus, fuzzy-trace theory considers false recall in the DRM paradigm to be based predominantly on gist traces, while true recall is based on both gist and verbatim traces (e.g., Payne et al., 1996; Toglia, Neuschatz, & Goodwin, 1999). In addition, verbatim traces can also help suppress false recall in a process known as *recollection rejection* (Brainerd, Reyna, Wright & Mojardin, 2003): while false-yet-gist-consistent items (such as the critical lure) may come to mind during recall, these can be rejected on the basis of lacking verbatim detail.

How might offloading affect the likelihood of falsely recalling related information? As noted above, researchers have suggested that offloading might serve as a kind of self-imposed forget cue, analogous to instructions to forget in the directed forgetting paradigm (Eskritt & Ma, 2014). Researchers who have studied directed forgetting effects on verbatim and gist information have found that forget cues disrupt verbatim traces more than gist traces (Ahmad, Tan & Hockley, 2019; Fawcett, Taylor & Nadel, 2013; Montagliani & Hockley, 2019). When directed forgetting manipulations have been used with item sets that encourage false recall, the results have been somewhat mixed depending on whether participants are cued to forget/remember after each word (item-method), or whether they are given the forget cue only after a whole list is studied (list-method). While item-method studies have tended to find that forget cues lead to a decrease in false recall (Marche, Brainerd, Lane & Loehr, 2005; Lee, 2008; Montagliani &

Hockley, 2019), list-method studies have found either no effect (e.g., Seamon, Luo, Shulman, Toner & Caglar, 2002) or a slight *increase* in false recall (e.g., Kimball & Bjork, 2002; Pitarque et al., 2018). In explaining this pattern, Marche et al. (2005) argued that verbatim traces (thought to be more susceptible to interference; Reyna & Brainerd, 1995) are more impacted by forget cues than gist traces. In list-method directed forgetting, participants receive the forget cue only after encoding the entire list and presumably forming a strong gist trace. As such, the list-level forget cue would primarily influence the strength of verbatim traces (e.g., by reducing post-list rehearsal), thus leading to increased false recall, given the established gist trace and impaired support for recollection rejection (Kimball & Bjork, 2002). On the other hand, Marche et al. (2005) suggested that item-level forget cues, in addition to disrupting verbatim traces, also disrupt the formation of the gist traces during encoding. Since the remember/forget cue appears soon after every item, participants can adopt a strategy of waiting for the appropriate cue before engaging in elaborative semantic processing, thereby compromising the formation of the gist trace. Thus, in item-method directed forgetting, one observes a decrease in false memory (though to a lesser extent than the decrease in true recall; Marche et al., 2005; Montagliani & Hockley, 2019).

Offloading memory demands to an external store bears more similarity to item-method directed forgetting than to list-method directed forgetting. For example, when we write down reminders for future events, the putative forget cue is present at the time of encoding (i.e., we know at the time of writing the reminder that we are creating an external record of events). Indeed, offloading in this manner could be argued to provide an even more potent forget cue as individuals know they can rely on the external store earlier in the encoding episode than a typical item-level directed forgetting cue. From this perspective, offloading externally should decrease false recall compared to maintaining information in internal memory.

1.1 Present Investigation

In the present investigation, we report three preregistered experiments (Experiment 1a: https://osf.io/pcdtg; Experiment 1b: https://osf.io/gdjc2; Experiment 2: https://osf.io/x56hg) using an adaptation of a paradigm developed by Kelly and Risko (2019a; 2019b). Specifically, we manipulated the opportunity to offload in a free recall task to examine how this would affect false memory formation.

In Experiments 1a and 1b, participants performed a series of trials on which they were presented with two lists of to-be-remembered words on every trial. Each list was composed of words designed to produce the false recall of a particular critical lure. On each trial, participants were told to write down words presented in one color (e.g., red) on one list that would be accessible during the recall task (i.e., individuals could rely on the external memory store), and words in the other color (e.g., blue) on another list that would be inaccessible during the recall task. As such, the manipulation of offloading consists of an instruction that the participant can expect to have access to the external store during the forthcoming memory test, contrasted with an instruction that they would not have access to the external store. While participants write the words down in both conditions, the former condition (where they have been told that the external store will be accessible) mimics a typical scenario in which an individual would engage in offloading memory demands (henceforth referred to as the "Offloading" condition). On the other hand, the latter condition (where they have been told that the external store will be inaccessible) is akin to a typical scenario where individuals would not offload (henceforth referred to as the "No-Offloading" condition). On the first three trials, the participant's expectations about whether they would or would not have access to their list was accurate. That is, when they were told they would have access to the words they wrote down, they were indeed allowed access to that list (and vice versa when they were told they would not have access to their list). We used this procedure in order for participants to develop trust/familiarity with using and relying on the external store. Critically, before the final trial's recall phase (but after encoding), participants were notified that they would not be able to refer to their external store during recall for either list. Thus, recall on the final trial contrasts recall of the words for which participants thought they could rely on an external memory store (i.e., the "Offloading" condition where access to the external store was expected) with recall of the words for which they knew they had to rely on their internal memory (i.e., the "No-Offloading" condition where access to the external store was not expected).

2. Experiments 1a and 1b

2.1 Method

Experiments 1a and 1b are identical (1b was intended as a replication of 1a) and are described together.

2.1.1 Participants

In both experiments, data from 40 participants were analyzed based on an a priori power analysis with a desired power of .80 ($\alpha = .05$, two-tailed) to detect a medium-sized effect when comparing the false recall between offloading conditions (Offloading vs No-Offloading). Participants were undergraduate students at the University of Waterloo participating for course credit.

2.1.2 Apparatus

Participants sat at individual workstations separated by occlusion screens. Each workstation had a computer, a monitor, a pen, a folder, and an envelope.

2.1.3 Stimuli

We created four 20-item word lists (see Appendix) adapted from Stadler, Roediger and McDermott (1999) by taking the first ten items from eight lists that produced the highest rates of recall for the critical lure. Each list was formed by combining two sets of ten DRM items, resulting in four lists of 20 items each. The two sets of DRM items within each list were randomly interleaved with one another. Within each set of ten DRM items, word order was fixed in order of decreasing backwards association strength to the critical lure (as is typical in studies using the DRM paradigm; Roediger et al., 2001). Each DRM item set would appear in either blue or red, with the other set in each list appearing as the other color. Lists were counterbalanced across trial position (i.e., 1 to 4) such that each list appeared in each trial position equally across participants. Moreover, we also counterbalanced the assignment of word color to DRM word set (i.e., A or B) as well as color to offloading condition (i.e., Offloading vs No-Offloading) across participants.

2.1.4 Procedure

Participants followed the instructions given on-screen and by the researcher throughout the session (four trials total). Each trial had three components: encoding, a brief retention interval with the external store inaccessible, and recall. A researcher in the room monitored participants for compliance (e.g., that no participants used the list on the final trial, and that lists were put away promptly), and replaced participants' used lists with blank ones after each trial.

At the start of the experiment, participants were told that they would be learning two intermixed lists of words on each trial: words from one list would be presented in red font with words from the other in blue font. They were told to write each word as it appeared onto one of two pieces of paper that had been pre-labelled as "red list" and "blue list". They were told which of the two lists they would have access to during the recall phase of each trial (i.e., Offloading), and which they would not have access to (i.e., No-Offloading).

Encoding. Before each trial began, each participant was given two pieces of paper labelled as "red list" and "blue list", which were otherwise blank. At the beginning of each trial, the participant was presented with a list of words on the screen, presented one at a time. Each word was presented alone for 3000 ms at the center of the screen in either red or blue font, followed by a blank screen for 3000 ms. Participants were told to write each word as it appeared onto their corresponding “red list” or “blue list”.

Retention interval. Once all list words had been presented, twenty-two seconds were provided to participants to place one list in the provided envelope, where it would be inaccessible for the recall task, and the other list in the provided folder, where it would be accessible. This was followed by a five-second screen instructing them to take their lists out of the folder for the impending recall task.

Recall. Participants were instructed to recall all the words that they could (both from the accessible list and the inaccessible list) into an onscreen text field (all entered words were visible throughout recall). Specifically, on the first three trials, they were instructed that they may use their external stores to aid recall by opening their folders to access the list; they were not able to access the list that was discarded into the envelope. Thus, participants were told that they would always have access to any items written on the list that would be placed into the folder (Offloading) but they would not have access to any items that were written on the list that would be placed into the envelope (No-Offloading). Critically, on the fourth (final) trial, participants were instructed not to take the list out of the folder, unlike in the previous three trials. Thus, participants had to recall the words without use of their external stores. Participants were given 180 seconds to complete each free recall phase and were debriefed and excused when finished.

2.2 Results

Data from five participants in Experiment 1a and seven participants in Experiment 1b were not analyzed because they did not follow instructions. Additional participants who took part after the stopping rule of 40 (8 in Experiment 1a, 8 in Experiment 1b) were also excluded. As reported in the pre-registrations, we elected to analyze the recall of the critical lures and presented items separately, then followed up with a combined ANOVA and mixed-effects logistic regression analyses. All mixed-effects logistic regression analyses revealed a similar pattern of results and are reported in the Supplementary Materials. Data and analysis code are available at https://osf.io/bqwyc/files/ (1a) and https://osf.io/rn5wq/files/ (1b). Table 1 shows the recall of presented items and critical lures across the four trials of the experiment. As expected, when participants had access to their written lists (in the first three trials), true recall for the words on that list was near ceiling, while the false recall rate was close to zero. Across the first three trials, participants often falsely recalled the critical lure in the No-Offloading condition (i.e., the unpresented critical word associated with the list that they knew they would not have access to). As our interest was not in these first three trials, we provide these means for descriptive purposes.

Table 1. Mean (SD) recall for presented items and related critical lures in Experiments 1a and 1b as a function of offloading condition across trials 1 to 4.

	Experiment 1a				Experiment 1b			
	Trial 1	Trial 2	Trial 3	Trial 4	Trial 1	Trial 2	Trial 3	Trial 4
Offloading								
Presented	.98 (.09)	.96 (.10)	.98 (.04)	.42 (.18)	.97 (.13)	.97 (.10)	.95 (.10)	.46 (.21)

Critical Lures	.05	.02	00	.38	.08	.03	.03	.35
	(.22)	(.16)	(00)	(.49)	(.27)	(.16)	(.16)	(.48)
No-Offloading								
Presented	.65	.72	.77	.78	.59	.71	.74	.73
	(.21)	(.18)	(.13)	(.19)	(.21)	(.20)	(.24)	(.19)
Critical Lures	.38	.22	.18	.10	.28	.28	.20	.13
	(.49)	(.42)	(.38)	(.30)	(.45)	(.45)	(.41)	(.34)

Recall of Critical Lures on the Final Trial. In Experiment 1a, an exact McNemar's test revealed a statistically significant difference in the recall of the critical lure (CL) across offloading conditions, OR = 6.50, p = .007; in the Offloading condition 38% of participants recalled the CL while in the No-Offloading condition 10% recalled the CL. A similar result emerged in Experiment 1b, OR = 3.25, p = .049, with 35% of participants recalling the CL in the Offloading condition while 13% recalled the CL in the No-Offloading condition. Thus, participants were more likely to recall the critical lure associated with the words in the Offloading condition (where they expected they would have access to their written list) compared the words in the No-Offloading condition (where they knew they would not have access to their written list).

Recall of Presented Items on the Final Trial. In Experiment 1a, a within-subjects t-test, $t(39) = 7.99$, $p < .001$, $d = 1.90$, revealed that participants recalled fewer items in the Offloading condition (M = .42, bootstrapped 95% CI: .38 – .47) than in the No-Offloading condition (M = .78, bootstrapped 95% CI: .73 – .82). A similar result emerged in Experiment 1b, $t(39) = 6.88$, $p < .001$, $d = 1.40$ (Offloading: .46, bootstrapped 95% CI: .41 – .52; No-Offloading: .73, bootstrapped 95% CI: .69 - .78). Thus, the word lists in the Offloading condition were more poorly recalled compared to those in the No-Offloading condition.

Omnibus Analyses. In Experiment 1a, a 2 x 2 within-subjects ANOVA with Offloading condition (Offloading vs No-Offloading) and recalled Word Type (Presented vs Critical Lure) revealed a significant main effect of Word Type, $F(1, 39) = 40.72$, $p < .001$, $\eta_{G^2} = .25$, demonstrating that there was a higher rate of recall for presented items compared to the critical lures, and no significant main effect of Offloading, $F(1, 39) = 0.68$, $p = .414$, $\eta_{G^2} < .01$. There was, however, a significant interaction between Offloading and Word Type, $F(1, 39) = 40.09$, $p < .001$, $\eta_{G^2} = .20$. A similar result emerged in Experiment 1b; again there was a significant main effect of Word Type, $F(1, 39) = 52.64$, $p < .001$, $\eta_{G^2} = .24$, such that there was a higher rate of recall for presented items compared to the critical lures, and no significant main effect of Offloading, $F(1, 39) = 0.25$, $p = .620$, $\eta_{G^2} < .01$. There was also a significant interaction between Offloading and Word Type, $F(1, 39) = 20.19$, $p < .001$, $\eta_{G^2} = .13$, such that there was lower recall of the presented items but higher recall of the critical lures in the Offloading condition (relative to No-Offloading). Figure 1 shows the mean proportion of recall by Offloading and Word Type in Experiments 1a and 1b.

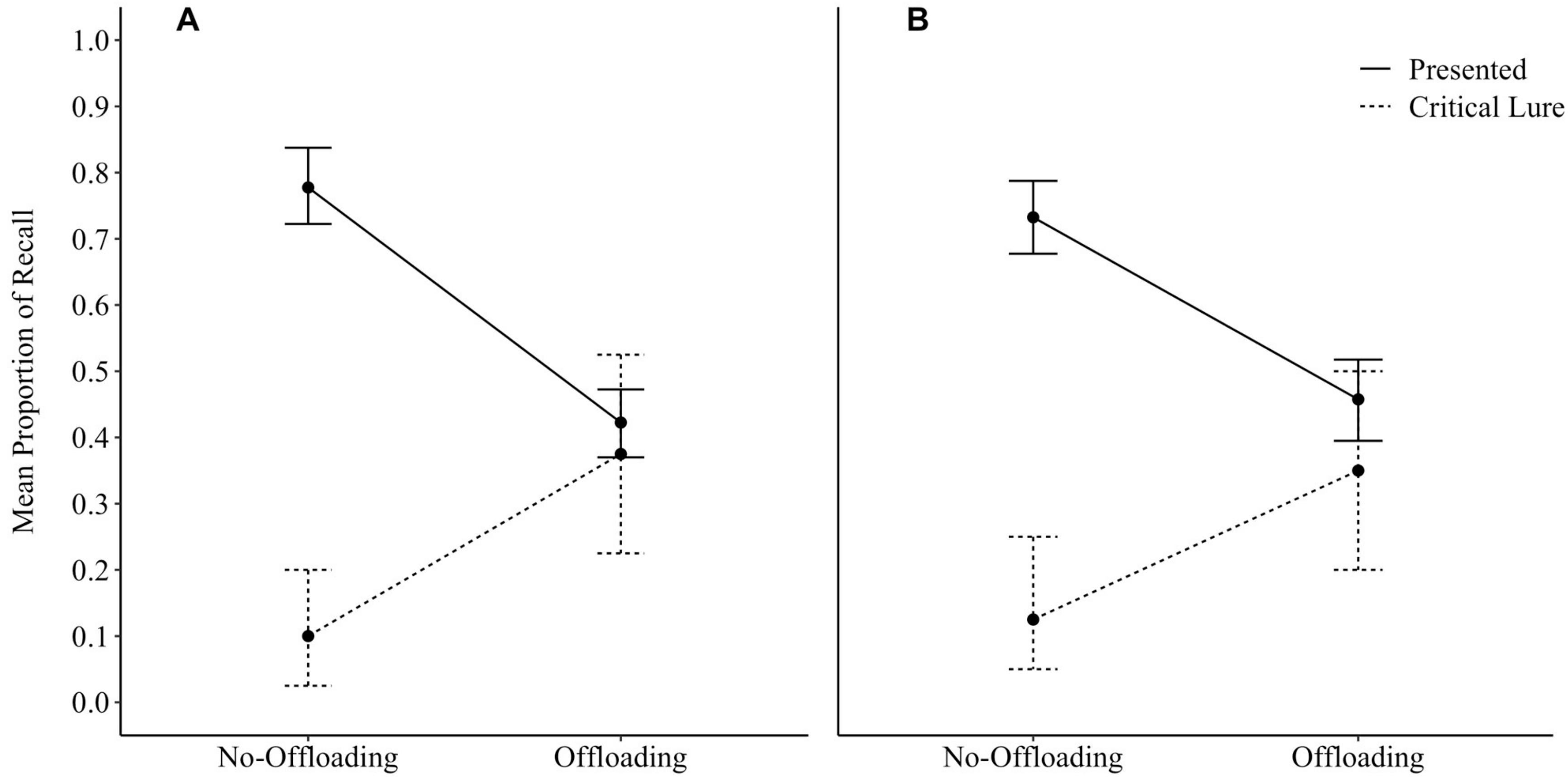

Figure 1. Mean proportion of recall by Offloading condition and Word Type in Experiment 1a (A) and Experiment 1b (B). Error bars represent bootstrapped 95% confidence intervals.

2.3 Exploratory

Recall of other intrusions. We examined recall intrusions (other than the critical lure) across the four trials. For each participant, the proportion of recall intrusions was calculated by dividing the number of intrusions by the number of recalled words for that trial. We also classified intrusions under the following scheme: (1) semantically related to presented words from the list in the Offloading condition, or (2) No-Offloading condition; (3) semantically unrelated intrusions. (A more detailed breakdown of unrelated intrusions can be found in the Supplementary Materials.) Since the number of participants who had intrusions was very small, we provide the raw intrusion counts for both related and unrelated intrusions across participants. Table 2 shows the mean proportion of intrusions recalled across participants (calculated by dividing each participant's number of intrusions by their total number of recalled items) as well as the related intrusion counts (for Offloading and No-Offloading) and unrelated intrusion counts.

Table 2. Mean proportion of non-critical lure intrusions recalled in Experiments 1a and 1b.

	Experiment 1a				Experiment 1b			
	Trial 1	Trial 2	Trial 3	Trial 4	Trial 1	Trial 2	Trial 3	Trial 4
Proportion of Intrusions	< .01	< .01	< .01	.02	< .01	< .01	< .01	.01
Related Intrusion Count								
Offloading	0	0	0	6	0	0	0	2
No-Offloading	3	0	0	0	0	1	0	2
Unrelated Intrusion Count								
All	3	2	4	4	2	2	4	1

Output order of Offloading and No-Offloading lists. As our free recall task allowed participants to output the words from both the Offloading and No-Offloading conditions in any order, we explored whether participants exhibited any patterns in their output order. (A few participants recalled no words in the Offloading condition and are not included in this analysis.) Participants tended to recall the words from the lists in the Offloading condition later than the words from the lists in the No-Offloading condition. Table 3 shows the mean output positions of recalled words for the Offloading and No-Offloading conditions across all four trials. Note that on average participants recalled more items in total on trials 1, 2 and 3 (respectively, Experiment 1a: 16.20, 16.80, 17.40; Experiment 1b: 15.50, 16.80, 16.90) than trial 4 (Experiment 1a: 12.00; Experiment 1b: 11.90).

Table 3. Mean output positions (SD) of recalled words across trials in the Offloading and NoOffloading conditions in Experiments 1a and 1b.

	Experiment 1a				Experim nt 1b			
	Trial 1	Trial 2	Trial 3	Trial 4	Trial 1	Trial 2	Trial 3	Trial 4
Offloading	10.30	11.30	13.00	9.53	8.88	12.50	12.90	9.38
	(3.72)	(3.53)	(2.17)	(3.35)	(3.56)	(2.78)	(2.65)	(3.53)
No-Offloading	7.49	6.51	5.07	5.50	8.42	5.36	4.82	5.16
	(3.99)	(3.38)	(1.96)	(1.58)	(4.65)	(2.26)	(1.67)	(1.76)

Further exploratory analyses on the relative output position of the critical lures and serial position effects are available in the Supplementary Materials.

3. Experiment 2

In the final trial of both experiments, expecting to have access to an external store during encoding lead to both lower true recall of studied words, and higher false recall of unstudied critical lures. Experiment 2 addresses three issues that arose in Experiments 1a and 1b. First, we observed an unexpected trend: the false recall rate in the No-Offloading condition (i.e., the list for which individuals knew they would not have access during recall) decreased across the four trials (Table 1). This could be indicative of a kind of "practice" effect across the four trials, such that participants selectively increase their criterion for rejecting the critical lures associated with the lists stored internally. Therefore, the observation that offloading lead to increased false memory in the Offloading condition might reflect returning to a baseline (i.e., unpracticed) level of false memory recall (without the experience of the previous three trials). Second, provided the lists in the Offloading and No-Offloading conditions have to be recalled at the same time, there is also the potential for competition between lists during output. Indeed, our exploratory analyses showed individuals tended to report the items from the lists in the Offloading condition after the items from the lists in the No-Offloading condition. In order to better understand why individuals are more prone to false recall in the Offloading condition, understanding the extent to which the increase in false recall in the Offloading condition reflects inter-list competition would be valuable. Lastly, the mixed list design restricted the analysis of intrusions in the sense that only related intrusions could be unambiguously assigned to a condition (i.e., either Offloading or NoOffloading). The analysis of non-related intrusion rates could provide critical clues as to the mechanism underlying the increase in false recall in the Offloading condition. For example, attempting to recall items from a previous list for which you thought you could rely on an external store might encourage a more liberal criteria for recall.

Experiment 2 addresses the issues outlined above. In particular, we moved to a single-list per trial design (similar to previous work; Kelly & Risko, 2019b), rather than the two-lists per trial design in Experiments 1a and 1b. Again, the Offloading manipulation occurred withinsubjects, but in Experiment 2 this manipulation was between-lists, such that participants studied a single list of words on each trial, with the key Offloading manipulation occurring across the final two trials. That is, on each trial, words from a single (critical lure associated) list were presented, and participants were told to write all of the words down. At the beginning of the study, participants were told that they would have access to their external stores on all trials except one, but that they would be informed before that trial began. On the first three trials, participants had access to their external stores. The critical Offloading manipulation happened on the final, fourth and fifth, trials. On one trial, participants were told before encoding that they would not have access to their external stores (No-Offloading condition; i.e., they knew they had to rely on internal memory). On the other trial, participants were not told this before encoding, but right before the recall phase (Offloading condition; i.e., they thought they would be able to rely on their external stores).

By using a single-list design instead of two lists per trial, the problem of a potential "practice" effect in the No-Offloading condition across the four trials is eliminated; because the Offloading vs No-Offloading manipulation occurs only on the final two trials, no list-specific experience can be accumulated. Secondly, as participants recall only one list per trial, the problem of inter-list competition at retrieval is also eliminated. Lastly, the single-list design allows us to better compare related and unrelated intrusion rates across conditions, as both related and unrelated intrusions can be clearly assigned to each condition.

3.1 Methods

3.1.1 Participants

Data from 80 participants were analyzed based a power analysis (desired power of .80, $\alpha = .05$, two-tailed) using the McNemar test odds ratios obtained from Experiments 1a and 1b.

3.1.2 Stimuli

We created five 15-item word lists (see Appendix) adapted from Stadler, Roediger and McDermott (1999) by taking the first fifteen items from each list that produced the highest rates of recall for the critical lure. Lists were counterbalanced across trial position (i.e., 1 to 5) such that each list appeared in each trial position equally across participants.

3.1.3 Procedure

Participants followed the instructions given on-screen and by the researcher throughout the session (five trials total). Each trial had three components: encoding, a brief retention interval with the external store inaccessible, and recall. A researcher in the room monitored participants for compliance.

Encoding. At the beginning of each trial, the participant was presented with a list of words on the screen, presented one at a time. Each word was presented for 3000 ms in black font, followed by a blank screen for 3000 ms before the next word appeared. Participants were instructed to write down each word as they saw it onto a piece of paper that was provided to them. They were told that they would have access to this written list for recall except during one of the trials.

Retention interval. Once all list words had been presented, twenty-two seconds were provided to participants to place the list in the provided folder. This was followed by a screen instructing them to take their lists out of the folder for the impending recall task that was displayed for five seconds.

Recall. Participants were instructed to recall all the words that they could into an onscreen text field (all entered words were visible throughout recall). Specifically, on the first three trials, they were instructed that they may use their external stores to aid recall by opening their folders to access their lists. The critical Offloading manipulation occurred on the final two trials, where participants were instructed not to take their lists out of the folders. Thus, on the final two trials, participants had to recall the words without the use of their external stores. On one of the trials, participants were warned before encoding that they would not have access to their lists (No-Offloading condition); on the other trial, participants were not warned beforehand and would have expected to have access (Offloading condition). The order of the Offloading and No-Offloading conditions was counterbalanced across participants. Participants were given 180 seconds to complete each free recall phase and were debriefed and excused when finished.

3.2 Results

Data from six participants were not analyzed because they did not follow instructions during the experiment. One participant's data also had to be excluded due to technical errors. One additional participant who took part after the stopping rule of 80 was also excluded. As reported in the pre-registration, we elected to analyze the recall of the critical lures and presented items separately, then followed up with a combined ANOVA and mixed-effects logistic regression analyses. When we included the order of the Offloading and No-Offloading trials as a factor, there was no main effect of order nor were there interactions with order. All mixed-effects logistic regression analyses revealed a similar pattern of results and are reported in the Supplementary Materials. Data and analysis code are available at https://osf.io/tserw/files/.

Table 4 shows the recall of presented items and critical lures across the five trials of the experiment. As expected, when participants had access to their written lists (in the first three trials), true recall for those words was near ceiling, while the false recall rate was close to zero. As our interest was not in these first three trials, we provide these means for descriptive purposes.

Table 4. Mean recall (SD) for presented items and related critical lures in Experiment 2 as a function of offloading across trials 1 to 5.

	Trial 1	Trial 2	Trial 3	No-Offloading	Offloading
Presented	.98 (.07)	.99 (.05)	.98 (.05)	.71 (.14)	.48 (.18)
Critical Lures	0 (0)	0 (0)	.01 (.11)	.21 (.41)	.40 (.49)

Recall of Critical Lures on the Final Two Trials. An exact McNemar's test revealed a statistically significant difference in the recall of the critical lure (CL) when comparing the Offloading trial to the No-Offloading trial, OR = 2.50, p = .017; in the Offloading condition 40% of participants recalled the CL while in the No-Offloading condition 21% recalled the CL. Thus, participants were more likely to recall the critical lure associated with the words in the Offloading condition (when they expected to have access to their external store) compared to words in the No-Offloading condition (that they expected to maintain internally).

Recall of Presented Items on the Final Two Trials. A within-subjects t-test, $t(79) = 10.88$, $p < .001$, $d = 1.38$, revealed that recall in the Offloading condition (M = .48, bootstrapped 95% CI: .45 – .51) was significantly worse than recall in the No-Offloading condition (M = .71, bootstrapped 95% CI: .69 – .73). Thus, the word lists in the Offloading condition were more poorly recalled compared to those in the No-Offloading condition.

Omnibus Analyses. A 2 x 2 within-subjects ANOVA with Offloading (No-Offloading vs

Offloading) and recalled Word Type (Presented vs Critical Lure) showed that there was a significant main effect of Word Type, $F(1, 79) = 54.11$, $p < .001$, $\eta_{G^2} = .15$, demonstrating that there was a higher rate of recall for presented items compared to the critical lures, and no significant main effect of Offloading, $F(1, 79) = 0.39$, $p = .537$, $\eta_{G^2} < .01$. There was, however, a significant interaction between Offloading and Word Type, $F(1, 79) = 27.28$, $p < .001$, $\eta_{G^2} = .09$, such that Offloading (relative to No-Offloading) lead to lower recall of the presented items but higher recall of the critical lures. Figure 2 shows the mean proportion of recall by offloading and word type in Experiment 2.

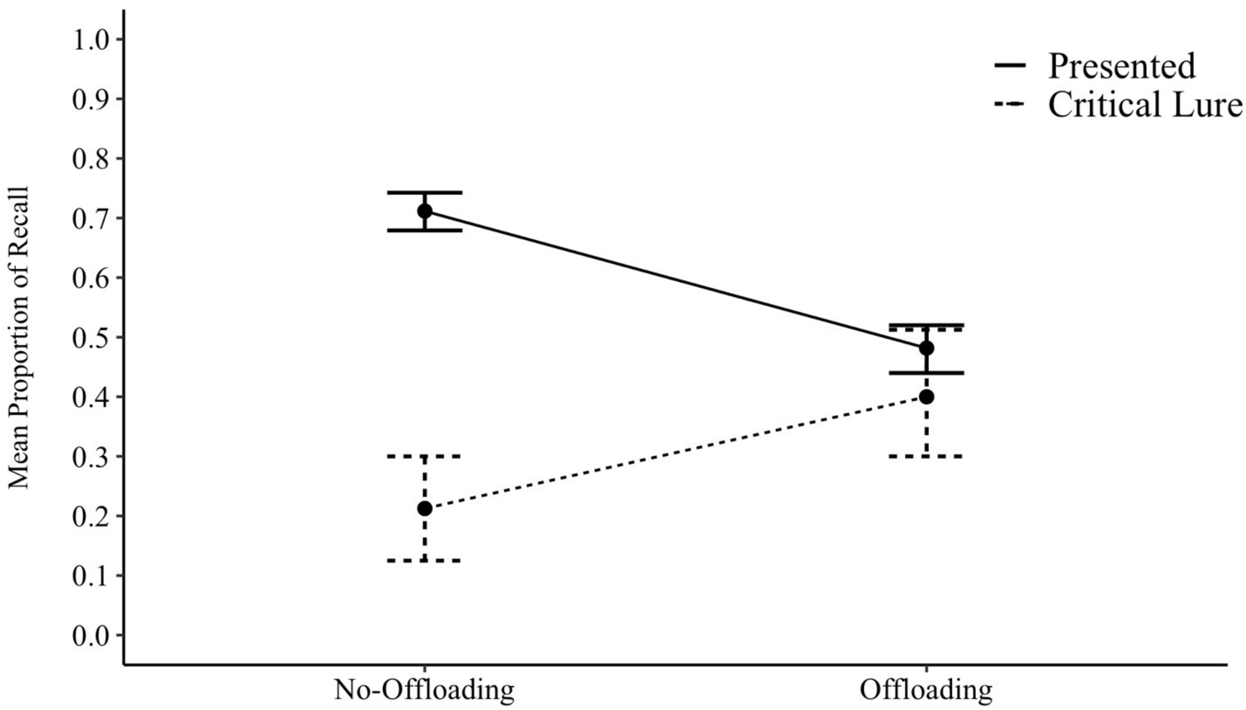

Figure 2. Mean proportion of recall by Offloading condition and Word Type in Experiment 2. Error bars represent bootstrapped 95% confidence intervals.

Recall of Related and Unrelated Intrusions. We calculated intrusion rates (other than the critical lure) across the five trials. Table 5 shows the mean proportion of intrusions recalled across participants. Since the number of participants who had an intrusion (and the number of intrusions) was small, we provide the raw intrusion counts for both related and unrelated intrusions across participants.

Table 5. Mean proportion of non-critical lure intrusions recalled and raw intrusion counts in Experiment 2.

	Trial 1	Trial 2	Trial 3	No-Offloading	Offloading
Proportion of Intrusions	< .01	< .01	< .01	.02	.04
Total Intrusion Count by Type					
Related	0	0	0	3	16
Unrelated	3	6	4	11	12

In our pre-registration, we had planned to classify the intrusions in a three-category (semantically related, orthographically similar, and unrelated) scheme. However, the intrusion rate for each category was very small, and some intrusions did not fall neatly under this classification; for example, an intrusion could be phonologically similar to a presented word. Therefore, we chose to aggregate all intrusions that were not semantically related intrusions into the unrelated intrusions category. (A more detailed breakdown of these intrusions can be found in the Supplementary Material.) Also, we had planned to compare intrusion rates across the Offloading conditions with *t*-tests, but this was problematic as the outcome data were not normally distributed. Therefore, we used Wilcoxson signed rank tests to compare the effect of Offloading condition on related and unrelated intrusion counts in the final two trials. For related intrusions, the Offloading condition (relative to No-Offloading) was associated with higher intrusion counts, $Z = 2.29$, $p = .022$, while for unrelated intrusions there was no effect of offloading, $Z = 0.11$, $p = .909$.

Further exploratory analyses on the relative output position of the critical lures and serial position effects are available in the Supplementary Materials.

3.3 Discussion

Experiment 2 replicated the findings of Experiments 1a and 1b. In the Offloading condition, where participants believed that they could rely on their external stores, there was lower true recall of studied words, and higher false recall of unstudied critical lures and other semantically-related words (compared to the No-Offloading condition, where participants knew that the external store would be unavailable). Provided the Offloading manipulation occurred only on the final two trials, we can rule out the differential accumulation of experience across trials in Experiments 1a and 1b as an explanation for the difference between conditions. Also, provided each participant studied a single list per trial, it does not appear that inter-list competition plays a major role in the observed results. Lastly, the exploratory analysis of intrusions suggests that while there was evidence of a greater number of non-critical but related intrusions in the Offloading condition, this was not true for unrelated intrusions. Each of these observations offers useful constraints for understanding offloading's effect on false recall.

4. General Discussion

Across all three experiments, offloading (when participants believed they would have access to their external stores during encoding) led to reduced recall of the studied words, consistent with previous research (e.g., Eskritt & Ma, 2014; Kelly & Risko, 2019a, 2019b; Sparrow et al., 2011). Critically, giving participants the opportunity to offload also increased the likelihood of false recall of related information when the external store was unavailable. The current results are in the opposite direction from that predicted by previous work on directed forgetting. If offloading represents a form of self-imposed item-level directed forgetting (Eskritt & Ma, 2014), then offloading should have decreased false recall, as in previous studies that had used item-level forget cues with DRM word lists (Marche et al., 2005; Lee, 2008). This was clearly not the case.

How do we reconcile these disparate results? One interesting possibility is that the increase in false recall observed here emerges from aspects of the act of offloading itself. That is, while our offloading paradigm may be similar to item-level directed forgetting in the sense that the putative remember/forget cues are present at the time of encoding, it is different in that participants needed to actively store the to-be-remembered items in the external store (i.e., they write the words onto a list). In other words, while Marche et al. (2005) suggested that participants in item-level directed forgetting studies can bypass the processing of 'forget' words by intentionally waiting for the remember/forget cue and thereby compromising the formation of the gist trace, avoiding encoding associative information might be more difficult in typical offloading scenarios, given that participants typically have to actively transfer information to an external store. In the current experiments, this occurred in the act of writing the words down onto a list, similar to creating a to-do-list in real life. This externalized form of encoding, and possibly subsequent exposure to the full written list (i.e., participants' lists were visible to them during the study phase), might be sufficient to promote the extraction of gist information. By this account, the disparate results between offloading and item-level directed forgetting lie in differences in the depth of encoding afforded/required by each act. This account makes the straightforward prediction that having participants write words (and/or view the to-be-remembered words in an accumulating list) in a standard item-level directed forgetting paradigm would lead to forget instructions increasing, rather than decreasing, false recall. It would also predict that eliminating the need to participate in the act of externalizing the to-be-remembered information in the Offloading condition here would lead to false memory performance more similar to that found in past item-level directed forgetting work (Marche et al., 2005). This also raises an interesting question about how central the act of *externalization* (e.g., writing down to-be-remembered information; Risko & Gilbert, 2016) should be to definitions of offloading. For example, it might be fruitful to distinguish between more active and more passive forms of offloading (e.g., putting information into an external store vs. leaving information in an external store). It will be important for future work to consider how various aspects of the act of offloading can impact memory (and other domains; see Risko & Gilbert, 2016 for a review).

Whether the foregoing account of the disparate results is correct or not, the current results support an interpretation of typical cases of offloading (i.e., actively storing to-be-remembered information in an external store in order to reduce memory demands) as disrupting verbatim traces more than gist traces. According to fuzzy-trace theory, the formation of gist traces occurs automatically and unconsciously (Brainerd et al., 2001; Brainerd & Reyna, 2005), while verbatim traces become inaccessible more rapidly and require repetition to maintain (Brainerd, Reyna, & Kneer, 1995; Brainerd, Reyna & Mojardin, 1999). From this perspective, our results are consistent with the idea that typical cases of offloading result in reduced top-down mnemonic activities intended to improve memory during encoding (i.e., reduced engagement in those processes more under the control of the participant; Kelly & Risko, 2019a, 2019b).

While the reliance on gist memory when one can depend on an external store yielded a “negative” outcome in the present context (i.e., higher rates of false recall), the storage of such information might play an important role in a transactive memory system (Wegner, 1987). In such a system, memories are thought to be distributed across multiple "locations", such as a significant other (Wegner, Erber & Raymond, 1991) or even a computer (Sparrow et al., 2011). Individuals are able to rely on these transactive memory partners to store to-be-remembered information for them, thereby relieving themselves of the burden of remembering all of it. Yet relying on these external forms of memory does not entirely eliminate the memory requirements of the task. When information has been offloaded to such a transactive memory partner, rather than needing to store the "what" information, we need to store the "where" information for future retrieval (Sparrow et al., 2011). The existence of a gist representation might be sufficient to perform this function, at least in many contexts. In addition, a gist-based representation of offloaded information might help individuals solve the endorsement problem present in distributed memory systems. That is, when individuals store information externally, one has to decide at retrieval whether to endorse or accept the information contained within/retrieved from that external store as accurately reflecting what was originally encoded into it (Arango-Muñoz, 2013; Clark & Chalmers, 1998; Risko, Kelly, Patel & Gaspar, 2019). While individuals tend to have difficulty detecting manipulations of their external memory stores (Risko et al., 2019), relying on gist memory could allow them to detect flagrant thematic inconsistencies.

5. Conclusions

Although offloading to-be-remembered information allows us to escape the limitations of our internal memory (Clark, 2008; Clark & Chalmers, 1998; Nestojko et al., 2013; Risko & Gilbert, 2016), relying on an external store can compromise our ability to remember in the absence of that store (relative to relying on our internal/biological memory). In addition, in the present investigation, we have demonstrated that storing to-be-remembered information in an external store also elevates false memory for related information.

Acknowledgements

We acknowledge Batul Karimjee, Sarah Schywiola, and Hyacinth Raymundo for their assistance in data collection and data entry.

Funding

This work was supported by a Discovery Grant (#04091) from the Natural Sciences and Engineering Research Council of Canada (NSERC), an Early Researcher Award from the Province of Ontario (#ER14-10-258), funding from the Canada Foundation for Innovation and Ontario Research Fund (#37872) and from the Canada Research Chairs (#950-232147) program to E.F.R.

References

Ahmad, F. N., Tan, P., & Hockley, W. E. (2019). Directed forgetting for categorised pictures: Recognition memory for perceptual details versus gist. *Memory*, *27*(7), 894–903.

Arango-Muñoz, S. (2013). Scaffolded memory and metacognitive feelings. *Review of Philosophy and Psychology, 4*(1), 135–152.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

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

Brainerd, C. J., Reyna, V. F., & Kneer, R. (1995). False-recognition reversal: When similarity is distinctive. *Journal of Memory and Language*, *34*(2), 157-185.

Brainerd, C. J., Reyna, V. F., & Mojardin, A. H. (1999). Conjoint recognition. *Psychological Review*, *106*(1), 160.

Brainerd, C. J., Reyna, V. F., Wright, R., & Mojardin, A. H. (2003). Recollection rejection: False-memory editing in children and adults. *Psychological Review*, *110*(4), 762.

Brainerd, C. J., Wright, R., Reyna, V. F., & Mojardin, A. H. (2001). Conjoint recognition and phantom recollection. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(2), 307-327.

Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension*. New York, NY: Oxford University Press.

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, *58*(1), 7-19.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, *58*(1), 17.

Eskritt, M., & Ma, S. (2014). Intentional-forgetting: Note-taking as a naturalistic example. *Memory & Cognition*, *42*, 237-246.

Fawcett, J. M., Taylor, T. L., & Nadel, L. (2013). Intentional forgetting diminishes memory for continuous events. *Memory*, *21*(6), 675–694.

Kelly, M. O., & Risko, E. F. (2019a). Offloading memory: Serial position effects. *Psychonomic Bulletin & Review*, *26*(4), 1347-1353.

Kelly, M. O., & Risko, E. F. (2019b). The isolation effect when offloading memory. *Journal of Applied Research in Memory and Cognition, 8*(4), 471-480.

Kimball, D. R., & Bjork, R. A. (2002). Influences of intentional and unintentional forgetting on false memories. *Journal of Experimental Psychology: General*, *131*(1), 116.

Lee, Y. S. (2008). Can intentional forgetting reduce false memory? Effects of list-level and itemlevel forgetting. *Acta Psychologica*, *127*(1), 146-153.

Marche, T. A., Brainerd, C. J., Lane, D. G., & Loehr, J. D. (2005). Item method directed forgetting diminishes false memory. *Memory*, *13*(7), 749-758.

Montagliani, A., & Hockley, W. E. (2019). Item-based directed forgetting for categorized lists: Forgetting of words that were not presented. *Canadian Journal of Experimental Psychology, 73*(3), 135–143.

Nestojko, J. F., Finley, J. R., & Roediger, H. L. (2013). Extending cognition to external agents. *Psychological Inquiry*, 24, 321–325.

Payne, D. G., Elie, C. J., Blackwell, J. M., & Neuschatz, J. S. (1996). Memory illusions: Recalling, recognizing, and recollecting events that never occurred. *Journal of Memory and Language*, *35*(2), 261-285.

Pitarque, A., Satorres, E., Escudero, J., Algarabel, S., Bekkers, O., & Meléndez, J. C. (2018). Motivated forgetting reduces veridical memories but slightly increases false memories in both young and healthy older people. *Consciousness and Cognition*, *59*, 26-31.

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences*, *7*(1), 1-75.

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

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 803.

Roediger, H. L., Watson, J. M., McDermott, K. B., & Gallo, D. A. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review*, *8*(3), 385407.

Seamon, J. G., Luo, C. R., Shulman, E. P., Toner, S. K., & Caglar, S. (2002). False memories are hard to inhibit: Differential effects of directed forgetting on accurate and false recall in the DRM procedure. *Memory*, *10*(4), 225-237.

Sheard, E. D., & MacLeod, C. M. (2005). List method directed forgetting: Return of the selective rehearsal account. In N. Ohta, C. M. MacLeod, & B. Uttl (Eds.), *Dynamic Cognitive Processes* (pp. 219-248). Tokyo: Springer.

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.

Stadler, M. A., Roediger, H. L., & McDermott, K. B. (1999). Norms for word lists that create false memories. *Memory & Cognition*, *27*(3), 494-500.

Toglia, M. P., Neuschatz, J. S., & Goodwin, K. A. (1999). Recall accuracy and illusory memories: When more is less. *Memory, 7*(2), 233-256.

Wegner, D. M. (1987). Transactive memory: A contemporary analysis of the group mind. In *Theories of group behavior* (pp. 185-208). Springer, New York, NY.

Wegner, D. M., Erber, R., & Raymond, P. (1991). Transactive memory in close relationships. *Journal of Personality and Social Psychology*, *61*(6), 923-929.

Appendix

Word lists presented during study (unpresented critical lures marked with *)

Experiments 1a and 1b

List 1	List 2	List 3	List 4
Set A			
door	nose	sour	cigarette
glass	breathe	candy	puff
pane	sniff	sugar	blaze
shade	aroma	bitter	billows
ledge	hear	good	pollution
sill	see	taste	ashes
house	nostril	tooth	cigar
open	whiff	nice	chimney
curtain	scent	honey	fire
frame	reek	soda	tobacco
window*	smell*	sweet*	smoke*
Set B			
bed	nurse	table	smooth
rest	sick	sit	bumpy
awake	lawyer	legs	road
tired	medicine	seat	tough
dream	health	couch	sandpaper
wake	hospital	desk	jagged
snooze	dentist	recliner	ready
blanket	physician	sofa	coarse

doze	ill	wood	uneven
slumber	patient	cushion	rugged
sleep*	doctor*	chair*	rough*

Experiment 2

List 1	List 2	List 3	List 4	List 5
door	nose	sour	bed	nurse
glass	breathe	candy	rest	sick
pane	sniff	sugar	awake	lawyer
shade	aroma	bitter	tired	medicine
ledge	hear	good	dream	health
sill	see	taste	wake	hospital
house	nostril	tooth	snooze	dentist
open	whiff	nice	blanket	physician
curtain	scent	honey	doze	ill
frame	reek	soda	slumber	patient
view	stench	chocolate	snore	office
breeze	fragrance	heart	nap	stethoscope
sash	perfume	cake	peace	surgeon
screen	salts	tart	yawn	clinic cure
shutter	rose	pie	drowsy	
window*	smell*	sweet*	bed*	doctor*