Study effort and the memory cost of external store availability

Megan O. Kelly

Evan F. Risko

University of Waterloo

This is the peer reviewed version of the following article: **Kelly, M. O., & Risko, E. F. (2022). Study effort and the memory cost of external store availability. *Cognition*, *228, 105228*.**, which has been published in final form at https://doi.org/10.1016/j.cognition.2022.105228.

Abstract

Previous work demonstrates that individuals often recall less information if, at study, there is expectation that an external memory store will be available at test. One explanation for this effect is that when individuals can expect access to an external memory store, they forgo intentional, controlled efforts at encoding. The present work offers a novel test of this account by examining study effort, indexed by study time and self-reported strategy use, as a function of instructed external store availability. In two preregistered experiments, participants studied lists of to-be-remembered items for a free recall test and were either instructed that they could use their study list to support them at test or that they could not. Critically, participants controlled their own study time, and no participant had their study list at test, regardless of instruction. Consistent with the effort at encoding account, external store availability influenced both study time and strategy use, and there was evidence that these effects mediated the influence of external store availability on recall performance. Interestingly, much of the memory cost remained when controlling for study effort, thus, suggesting that the cost is potentially multiply determined.

We often incorporate the support of external memory stores (or *external memory*, e.g., a notebook or a computer file, anything or anyone that can contain information) to accomplish our goals. These memory aids allow us to store information that we intend to remember without having to memorize that information ourselves, letting us skirt the limitations of our internal memory. This behavior allows us to offload the associated memory demands to the external store (Clark, 2010; Clark & Chalmers, 1998; Risko & Gilbert, 2016). While this approach to "remembering" has long been common and helpful, we are only beginning to make concerted efforts towards understanding this memory strategy at a basic, mechanistic level.

Much work has begun to explore the relation between internal memory and external supports, the various ways individuals use external memory supports to support internal memory, and the effects of doing so (e.g., Bulley et al., 2020; Eskritt & Ma, 2014; Finley et al., 2018; Gilbert et al., 2022; Grinschgl et al., 2021; Henkel. 2014; Kelly & Risko, 2019a, 2019b; Lu et al., 2020, 2021; Marsh & Rajaram, 2019; Meyerhoff et al., 2021; Park et al., 2022; Pereira et al., 2021; Risko et al., 2019; Risko et al., in press; Sachdeva & Gilbert, 2020; Scarampi & Gilbert, 2020; Sparrow, Liu, & Wegner, 2011; Storm & Stone, 2015). Of recent interest is how storing information externally influences the internal memory for that information (Eskritt & Ma, 2014; Grinschgl et al., 2021; Kelly & Risko, 2019a, 2019b; Lu et al., 2020, 2021; Park et al., 2022; Sparrow, Liu, & Wegner, 2011). One consistent finding to emerge from this research is that when individuals can rely on an external memory store, they demonstrate significantly poorer subsequent unaided recall of the stored information compared to if no external memory store was presumed to be available (Eskritt & Ma, 2014; Kelly & Risko, 2019a, 2019b; Lu et al., 2020; Sparrow, Liu, & Wegner, 2011). One explanation for this relative cost is that when individuals expect access to an external memory store, they forgo top-down, controlled efforts at study. For example, in cost/benefit models of effort allocation (e.g., the Expected Value of Control theory;

Shenhav et al., 2021), these kinds of control demanding strategies (e.g., rehearsal, narrative generation) would be viewed as inherently costly and deployed only when the payoff (e.g., in increased accuracy) exceeded that cost (Shenhav et al., 2021). With an external store capable of ensuring (presumably) perfect accuracy available, the expected value of control (i.e., the payoff due to engaging strategies aimed at improving memory accuracy) would be correspondingly low. Thus, the anticipation of having access to an external store effectively acts as a cue to *not* invest such mnemonic effort. We refer to this explanation as the *study-effort hypothesis*. In the present investigation, we test this account directly by examining both study time allocation and selfreported strategy use as a function of the availability of an external memory store.

The cost of external store reliance

A variety of different memory paradigms have been used to investigate the influence of relying on an external memory device. Sparrow et al. (2011) examined the influence of saving information on its subsequent recall during a test wherein the saved information was not available. Individuals who were told that their saved information had been erased before test demonstrated significantly better recall than individuals who were given the impression their information saved. Similarly, Eskritt and Ma (2014) found that when individuals were given the opportunity to use study notes to assist memory for an array of images, they demonstrated poorer unaided memory for image location than those who were unable to make study notes. Recent work has focused on explaining the cost of external store availability using the study-effort hypothesis. Kelly and Risko (2019a; 2019b) and Lu et al. (2020) found that the instructed availability of an external store led to *reduced* recall for information thought to be *more* reliant on top-down, controlled memory efforts at study and *no* reduced recall for information thought to be *less* reliant on such effort. Specifically, while memory for salient information (Kelly & Risko, 2019b), and the general theme of a list (e.g., when categorized; Lu et al., 2020) appears preserved when individuals have an external memory store available to rely on, recall is often more substantially reduced for beginning-of-list than end-of-list information (Kelly & Risko 2019a; 2019b) and for more item-specific verbatim information (Lu et al., 2020). That said, these studies did not directly attempt to index top-down, controlled efforts made at study. The present work does so through the examination of study time allocation and selfreported study strategy use.

Study time allocation and strategy use

One means of indexing study effort in the context of a memory task is to measure study time allocation, that is, the amount of time individuals spend "studying" a to-be-remembered stimulus when that time is controlled by the individual (i.e., when the task is self-paced; e.g., Mazzoni & Cornoldi, 1993; Metcalfe & Kornell, 2005; Son & Metcalfe, 2000). A number of studies have used study time allocation in this manner. For example, Ariel and colleagues (2015) examined whether the influence of information value on memory (i.e., memory for high-value versus low-value information) reflected differences in study time allocation as a function of value. They found that high value items received significantly more study and were remembered significantly better than lower value items. Similarly, Dunlosky and Thiede (1998) manipulated the probability that a specific study item would be tested later, finding that items very likely to be tested (i.e., 90% chance) were studied longer and remembered better than items less likely to be tested (i.e., 50% or 10% chance). Thus, under conditions wherein participants might modulate their effort and individuals are given control over study time, study time varies in relation to those putative modulations in effort.

While study time can provide a window into the allocation of study effort, it provides little in the way of insight into what individuals may be doing (or not doing) during that time.

For example, increased study time might be associated with the use of time intensive (Beaudoin & Desrichard, 2017; Froger et al., 2012), top-down, controlled memory strategies or mnemonics aimed at increasing later memory performance (e.g., repeated reading of items, imagery, sentence generation, grouping items in some meaningful way; based on Dunlosky & Kane, 2007; Unsworth et al., 2019; Unsworth, 2016). Moreover, study time alone will not distinguish between study strategies if those study strategies do not differ in the time spent engaging in them (Dunlosky & Kane, 2007). Previous findings suggest that engaging in effective controlled memory strategies (e.g., imagery or sentence generation vs. repetition or reading) can effectively increase the probability of recall success (e.g., Bailey et al., 2008; Bower, 1972; Dunlosky & Kane, 2007; Richardson, 1998; Rundus, 1971; Unsworth et al., 2019). For example, Unsworth et al. (2019) found evidence that individual differences in self-reported study strategies predicted recall performance (see also Dunlosky & Kane, 2007). Just as individuals can differ in their controlled study strategies, aspects of the study situation, such as the current manipulation of external memory store availability, could influence the use of study strategies. Thus, it might be useful to complement investigations of variation in study time with the concurrent examination of strategy use. Asking participants to report on whether they used a study strategy and/or to provide a description of what strategy they employed (if they did so) could provide additional insight into what individuals were doing at study across conditions. Thus, we examine both study time allocation and self-reported strategy use here as a function of whether individuals are told during study that they will be able to use an external memory store at test.

The current investigation

In the present work, we report two preregistered experiments (Experiment 1a and Experiment 1b; the second as a replication of the first) wherein the basic procedure of each follows that of previous work (Kelly & Risko, 2019a), such that participants complete a series of trials wherein they store to-be-remembered information on which they are subsequently tested in a free recall test. On the first three trials, participants are instructed that their memory test will be supported by their externally stored list of study items, and this is, indeed the case. Critically, however, on the fourth trial, they are not supported by their external store at test. Half of the participants are given notice of this (instructed-no-store), and half of the participants are not (instructed-store). Unlike previous studies, participants were given control of their study time, advancing study items as they saw fit. In addition, participants provided post-task reports of whether they used a memory strategy at study. The study-effort account predicts that participants who anticipate relying on the external memory store at test (instructed-store condition) should allocate less time *and* be less likely to report the use of a mnemonic strategy than those who are instructed that they cannot rely on the external memory store at test (instructed-no-store condition). Furthermore, study time allocation and strategy use should each mediate the effect of instruction on recall performance (i.e., explain the cost of relying on an external store). We index the former both in terms of a difference score across Trial 3 (i.e., the last trial wherein participants expect access to the external store) and Trial 4 (i.e., wherein half the participants expected access to the external store and the other half did not) and absolute study time allocation on Trial 4. The former allows us to control for differences in study time allocation across participants (i.e., Trial 3 is identical for all participants).

Experiments 1a & 1b

Provided that Experiment 1b is a replication of Experiment 1a, they are described together as such from hereon.

Method

Data was collected online (e.g., without a researcher “present”), as this investigation took place during the COVID-19 pandemic. Both experiments followed the same procedure unless otherwise noted. Experiments 1a and 1b were preregistered at osf.io/smk8n and osf.io/84cgk respectively, however, improvements were made to the analytical procedure in Experiment 1b, ergo, we follow that preregistered analytical procedure for both experiments. Deviations from this procedure are clarified (as is the case particularly for Experiment 1a).

Participants

Data was collected and analyzed from participants (E1a: *N* = 90, *n* = 45 per external store condition; E1b: *N* = 110, *n* = 55 per external store condition) recruited on PROLIFIC for GBP 5.00 (~USD 6.40) based on our estimation that the study would take no longer than 40 minutes to complete[1]. This was based on the *N* needed to have 80% power (alpha = .05, two-tailed) to detect a medium-large effect (E1a: *d* = 0.60; E1b: *d* = 0.55) of external store instructions (manipulated between-participants). The countries of residence of participants included Canada, the United Kingdom, and the United States. The exclusion criteria for participant data in analyses were the following (preregistered): (1) did not type at least 80% of words for Trials 2-4 during study, (2) did not reach at least 80% recall rate for the words across Trials 1–3 wherein they had access to their external store, (3) indicated that they were not paying attention or did not give reasonable effort during the task (e.g., self-reported that they were actually doing something else during the study), (4) answered *yes* to storing any words to aid memory outside of the means offered within the experiment, and (5) indicated during post-debriefing that they would not like their data to be used. Excluded participants were to be replaced to preserve the final *Ns* of each experiment and no participants from Experiment 1a were recruited in Experiment 1b.

Stimuli

Stimuli were four 20-item word lists (available at osf.io/8s75w) derived randomly from a stimulus set of 80 items. Word lengths ranged from four to 10 letters and word frequencies ranged

[1] Prolific computes a maximum study time based on the estimated time of completion and for our 40minute study, this is 106 minutes. Two participants in E1a and 3 in E1b reached this maximum and thus, were excluded and replaced.

from 26 to 82060 using *FreqCount* from *SUBTLEX-UK* (Van Heuven et al., 2014) and from two to 21384 using *FreqCount* from *SUBTLEX-US* (Brysbaert & New, 2009)[2]. Items were presented randomly within each list and lists were counterbalanced across trial position (i.e., first through fourth) such that each list appeared in each trial position.

Procedure

The experiments began by directing participants to follow instructions provided on the screen and to remain attentive to the screen throughout the study. Each of the four trials comprised three phases: a study phase, a 15-s period without access to their external store (as implemented in previous work, e.g., Kelly & Risko 2021), and then a test phase. Participants were not told how many trials they would be partaking in.

Study phase. At the start of each trial, participants were presented visually with the list of to-be-remembered words onscreen. Words were presented one at a time and participants selected "next word" whenever they were ready to advance to the next to-be-remembered word and there were no lower- or upper-time limits set for the self-paced study. A single line text field also appeared right below the word simultaneously and participants were instructed to type each item as they saw them into the textbox. Their typed word list appeared and accumulated on the left or right side of the screen (counterbalanced across participants) during the study phase. After the final item, a 15-s countdown elapsed wherein participants were without the list visible before moving on to the test phase.

Test phase. In the test phase, participants were instructed to type the items that were presented in the study phase into a large text field onscreen with their saved list as a resource (counterbalanced across participants to appear on the right or left side of the screen). An important feature of our

[2] Not including *ladybug, recombine, layover, deform,* and *perspire* which were not in the SUBTLEX-UK database and not including *foundation* and *uncle* which were not in the SUBTLEX-US database.

procedure is that on the first three trials, participants had access to their saved list during the test phases. This was to help participants establish a sense of trust in the external store akin to how they might trust their own external stores (e.g., their phones). Critically, participants do not get their external memory support during the test phase of Trial 4. Half of the participants were told of this after they completed Trial 3 (instructed-no-store condition); the other half of participants were not given notice that their external store would not be available during recall (instructed-store condition).

Post-task questionnaire. The final portion of the study was a short questionnaire wherein we asked participants about (1) any potential memory strategies they engaged in during the final trial wherein their list was unavailable (although this procedure differed slightly between experiments, as described shortly) and (2) whether they expected access to their list on the final trial.

In Experiment 1a, Question 1 asked, "In performing tasks like the ones we asked you to do in the current study, individuals sometimes report engaging in strategies they believe might help them remember the presented words. On the final trial, did you engage in any such strategies during the study phase (i.e., when you were typing the to-be-remembered words)? Please elaborate on your strategy or why you did not choose to use one (it is OK if you did not)." Question 1 in Experiment 1b had the same preamble but contained slightly different wording at the end: "In performing tasks like the ones we asked you to do in the current study, individuals sometimes report engaging in strategies they believe might help them remember the presented words. On the final trial, did you engage in any such strategies during the study phase (i.e., when typing the tobe-remembered word list)? It is completely OK if you did not. Please respond only based on the final trial." If participants responded "Yes" to this question, they were led to a follow-up free response question before advancing: "You responded that you did engage in such memory strategies on the final trial. Please briefly state what you did. Please only comment with respect to the final trial only." If participants responded to the initial question with "No", they advanced straight to Question 2.

In both experiments, the subsequent question read "Please rate how much you agree with the following statement: On the final trial, I had expected to get access to my list." Participants responded by selecting one option from the following scale: "Strongly agree", "Agree", "Neither agree nor disagree", "Disagree", "Strongly disagree", "I prefer not to answer". At the very end of the study, participants were asked if they used any written notes, etc. to aid them in the task, as well as whether there was any reason their data should not be used.

Results

Data from 19 participants in Experiment 1a and 22 participants in Experiment 1b were excluded and replaced because they did not follow instructions and/or did not meet the preregistered inclusion criteria. In Experiment 1a, two participants requested that their data not be used, three participants reported doing something else during their completion of the study, one participant did not encode at least 80% of items during study across Trials 2–4, and 13 participants did not recall at least 80% of items during test across trust trials (Trials 1–3). In Experiment 1b, two participants reported doing something else during their completion of the study, four participants did not encode at least 80% of items during study across Trials 2–4, and 16 participants did not recall at least 80% of items during test across trust trials. In the final samples (E1a: $N = 90$; E1b: $N = 110$), the mean reported age was 35.9 years-old in Experiment 1a (information unavailable for two participants) and 38.7 in Experiment 1b (information unavailable for one participant), with 53 participants identifying as female and 36 as male in Experiment 1a (information unavailable for one participant), and with 65 identifying as female and 44 as male in Experiment 1b.

Across all participants and lists, there were 28 instances in Experiment 1a and 55 in Experiment 1b (< 1% of trials in both experiments) wherein participants failed to correctly type an item into their store and these trials were removed before analysis. Across all trials of each participant in each experiment, there were 109 instances in Experiment 1a (1.49% of total trials) and 111 in Experiment 1b (1.25% of total trials) wherein participants recalled an item not on their list. In Experiment 1a, 38% of these instances involved participants recalling items from other lists within the study and this rate was 47% in Experiment 1b. The mean proportion of items recalled during the initial trials wherein participants had access to the external store ranged from 96.9% to 98.4% in Experiment 1a and 97.5% to 97.9% in Experiment 1b, thus performance was near ceiling, as to be expected when participants are, indeed, making use of their external memory stores.

For the free-response strategy question (applicable to all participants in Experiment 1a but only applicable to participants responding that they used a strategy in Experiment 1b), two naïve coders independently coded responses based on categories of strategies adapted from Unsworth et al. (2019) and Dunlosky and Kane (2007; see Table 1 for adapted categories). Note that while the plan was to follow exactly the outlined categories of Dunlosky and Kane (2007) and Unsworth et al. (2019), the responses of participants warranted some minor alterations to the original six categories. Initial agreement for responses in Experiment 1a was 69% and was 67% in Experiment 1b. Any disagreements between coders were reconciled by having them reach agreement on the classifications without researcher involvement.

As preregistered (and again, for both experiments, we are reporting results based on the preregistered analyses for Experiment 1b, noting deviations), we report mixed effects models (wherein continuous variables are standardized), as well as ANOVAs and t-tests over participant means to provide an assessment of the statistical claims across different statistical approaches (see Steegen et al., 2016). Note that the analyses often yield the same statistical conclusions, and that, more importantly, the same qualitative conclusions can be drawn. If applicable in the reported mixed effects models, by-participant and by-item slopes and intercepts are included whenever model fit is improved significantly by their inclusion (i.e., we deviate from the preregistration which did not mention using a model-fitting approach) unless convergence issues were encountered. In preregistering these analyses, we specified the use of by-participant random effects for external store condition which was manipulated between-participants, ergo we deviate from this preregistered specification by forgoing the inclusion of by-participant slopes varying for the external store factor (e.g., Brown, 2021). The model fitting approach was done using the *stats* package in R (R Core Team, 2020). The mixed models used dummy coding for categorical variables and were conducted using the *lme4* package in R (Bates et al., 2015). For mixed linear models (e.g., as used for study time data), the package *lmerTest* (Kuznetsova et al., 2017) was used to extract *p* values. While their utility is debated, we have included them here for ease of reading and consistency.

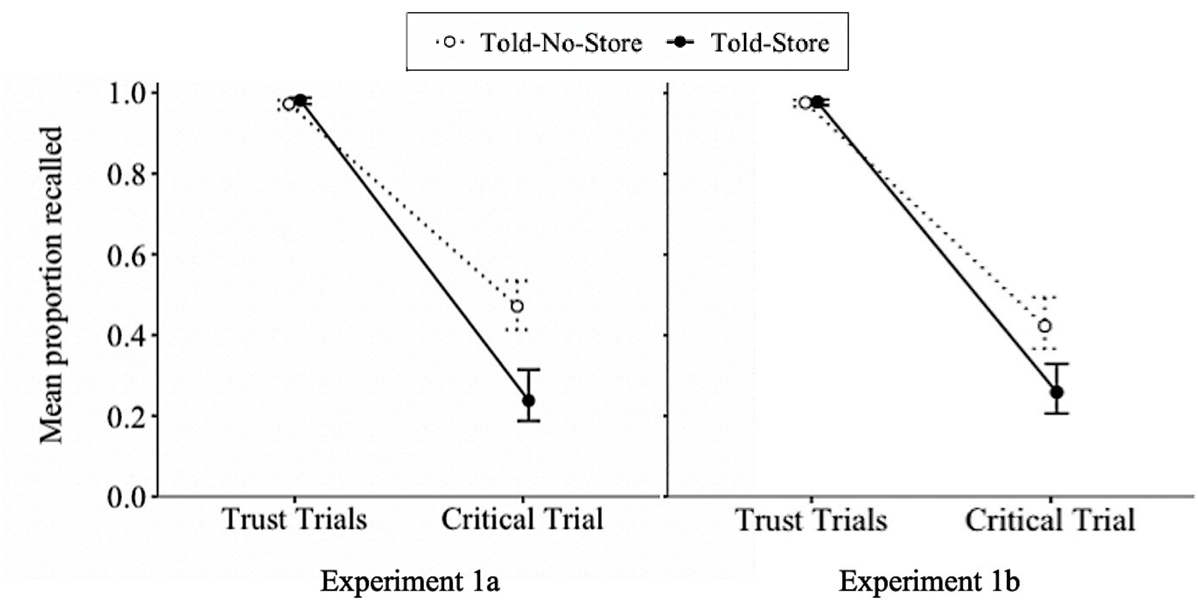
For the analyses involving the dependent variable of study time, outliers (3.6% of all observations in Experiment 1a and 3.7% of all observations in Experiment 1b) were defined using a non-recursive procedure within-participant for each block wherein any trials producing z-scores beyond a criterion based on sample size were removed (Van Selst & Jolicoeur, 1994). Thus, for analyses including the study time dependent variable, trimmed data (i.e., without outliers) were reported in addition to untrimmed data (i.e., not excluding outliers) [with the latter presented in square parentheses]. The random effects structures of the mixed effects models of untrimmed analyses follow that of the trimmed analyses unless convergence issues arose. Data and analyses codes for both experiments are available at osf.io/8s75w/.

The effect of external store availability

Recall. In both experiments, mixed effects logistic regression with external store condition as a predictor of final trial (Trial 4) recall performance (with by-participant and by-item intercepts as random effects) found that those in the told-store condition were significantly less likely to accurately recall items (**E1a:** told-store: .24; told-no-store: .47, $b = -1.31$, $SE = 0.25$, $z = -5.34$, $p < 001$; **E1b:** told-store: .26; told-no-store: .42, $b = -1.04$, $SE = 0.26$, $z = -4.00$, $p < 001$). An analogous analysis using Welch's independent samples t-test found qualitatively similar results in (**E1a:** $t(88.00) = 5.23$, $p < .001$, $d = 1.10$; **E1b:** $t(107.77) = 3.63$ $p < .001$, $d = 0.69$). Figure 1 presents the mean proportion of study items recalled as a function of external store condition for both trial types (trust and critical) in both Experiments 1a and 1b.

Figure 1

Mean proportion of study items recalled by external store condition, trial type, and experiment.



Note. Error bars are bias corrected accelerated bootstrap 95% confidence intervals using 10,000 replications.

Study time. Mixed effects linear regression with external store condition (told-store vs.

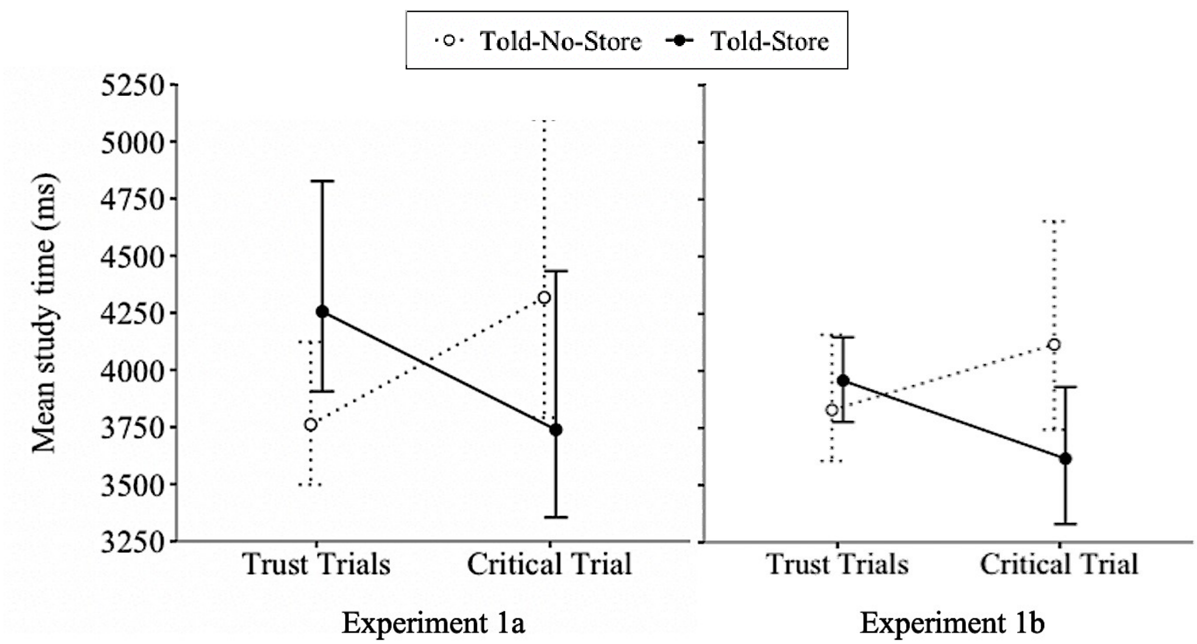
told-no-store) and trial (Trial 3 vs. Trial 4) as fixed effects on study time was conducted.[3] This analysis revealed a significant interaction between external store condition and trial such that the difference in study time between Trial 3 and Trial 4 was larger for the told-no-store condition (**E1a:** told-store: -11 ms, told-no-store: 760 ms, *b* = -0.29, *SE* = 0.11, *t* = -2.77, *p* = .007, [toldstore: -30 ms, told-no-store: 947 ms, *b* = -0.29, *SE* = 0.09, *t* = -3.29, *p* = .001]; **E1b:** told-store: 108 ms, told-no-store: 527 ms, *b* = -0.28, *SE* = 0.09, *t* = -3.12, *p* = .002, [told-store: 63 ms, toldno-store: 697 ms, *b* = -0.16, *SE* = 0.08, *t* = -2.08, *p* = .040]). Simple effects tests revealed that the difference in study time between Trials 3 and 4 was not significant in the told-store condition (**E1a:** *b* = -0.01, *SE* = 0.05, *t* = 0.10, *p* = .919 [*b* = -0.01, *SE* = 0.04, *t* = -0.23, *p* = .823]; **E1b:** *b* = -0.06, *SE* = 0.03, *t* = -1.70, *p* = .089, [*b* = 0.02, *SE* = 0.04, *t* = 0.50, *p* = .617]) but was significant in the told-no-store condition (**E1a:** *b* = 0.30, *SE* = 0.09, *t* = 3.30, *p* = .002, [*b* = 0.30, *SE* = 0.08, *t* = 3.74, *p* < .001]; **E1b:** *b* = 0.20, *SE* = 0.07, *t* = 2.86, *p* = .006, [*b* = 0.15, *SE* = 0.06, *t* = 2.43, *p* = .019]). The same interaction was found when using mixed analysis of variance (**E1a:** *F*(1, 88) = 8.15, *p* = .005, η_G^2 = .01, [*F*(1, 88) = 11.16, *p* = .001, η_G^2 = .02]; **E1b:** *F*(1, 108) = 10.62, *p* = .001, η_G^2 = .01 [*F*(1, 108) = 4.24, *p* = .042, η_G^2 = .01]). Analogous paired-samples t-test analyses also found no significant difference in study time in the told-store condition (**E1a:** *t*(44) = 0.09, *p* = .932, *d* = 0.01, [*t*(44) = 0.21, *p* = .838, *d* = 0.03] ; **E1b:** *t*(54) = 1.58, *p* = .119, *d* = 0.21, [*t*(54) = 0.57, *p* = .571, *d* = 0.08]) and a significant difference in the told-no-store condition (**E1a:** *t*(44) = 3.31, *p* = .002, *d* = 0.49, [*t*(44) = 3.73, *p* = .001, *d* = 0.56]; **E1b:** *t*(54) = 2.89, *p* = .006, *d* = 0.39, [*t*(54) = 2.42, *p* = .019, *d* = 0.33]).

[3] In Experiment 1a, the analysis of study time was not preregistered to include trial as a fixed effect and instead, only specified examining the effect of external store condition on Trial 4 recall, thus we deviate from this specification in following the preregistration for Experiment 1b.

Using mixed effects linear regression with by-participant and by-item random intercepts, we examined the fixed effect of external store condition on study time on Trial 4 (these analyses were not preregistered). Those in the told-store condition did not allocate significantly less study time (**E1a:** told-store: 3738 ms, told-no-store: 4317 ms, $b = -0.21$, $SE = 0.15$, $t = -1.36$, $p = .178$ [told-store: 3968 ms, told-no-store: 4675 ms, $b = -0.19$, $SE = 0.12$, $t = -1.59$, $p = .115$]; **E1b:** told-store: 3616 ms, told-no-store: 4115 ms, $b = -0.21$, $SE = 0.12$, $t = -1.79$, $p = .076$ [told-store: 4000 ms, told-no-store: 4632 ms, $b = -0.14$, $SE = 0.08$, $t = 1.71$, $p = .091$]), although the results are in the predicted direction. Figure 2 presents the mean study time per item as a function of external store condition for both trial types (trust and critical) in both Experiments 1a and 1b.

Figure 2

Mean study time per study item by external store condition, trial type, and experiment.



Note. Error bars are bias corrected accelerated bootstrap 95% confidence intervals using 10,000 replications.

Self-reported strategy—yes/no. A chi-square test found that those in the told-store condition were significantly less likely to report using a memory strategy on Trial 4 (**E1a:** toldstore: .40, told-no-store: .82, $c^2(1) = 15.15$, $p < .001$; **E1b**: told-store: .27, told-no-store: .53, $c^2(1) = 6.40$, $p = .011$). See Table 1 for a percentage breakdown of reported strategy type by external store condition and experiment.

Self-reported strategy—free response. In each experiment, we compared the selfreported strategies by external store condition for participants specifically indicating clear use of a strategy using a chi-square test of independence (i.e., excluding those who reported no strategy use).[4] There was no significant difference in types of strategy by external store condition (**E1a:** $c^2(5) = 4.88$ $p = .431$; **E1b:** $c^2(5) = 7.29$, $p = .200$).

Table 1.

Percentage of self-reported strategy (or no strategy) by external store condition and experiment.

	Experiment 1a		Experiment 1b	
	Told-no-store	Told-store	Told-no-store	Told-store
1: Read/Type each word	6.7%	4.4%	3.6%	0%
2: Repeated word often or as much as possible	4.4%	6.7%	0%	1.8%
3: Sentence/story generation, including to link multiple words	28.9%	6.7%	18.2%	16.4%
4: Mental imagery/visuals of words	4.4%	0%	5.5%	0%

[4] In Experiment 1a, all participants were presented with a free response field to respond with whether they used a strategy or not during final trial (Trial 4) recall, and to describe their strategy—thus all responses

5: Group/cluster words in some way	20.0%	8.9%	16.4%	3.6%
6: Other (e.g., unlisted strategy and/or multiple strategies)	17.8%	13.3%	9.0%	5.5%
Total proportion reporting (clear) strategy	82.2%	40%	52.7%	27.3%
0: No (clear) strategy reported	17.8%	60.0%	47.3%	72.7%

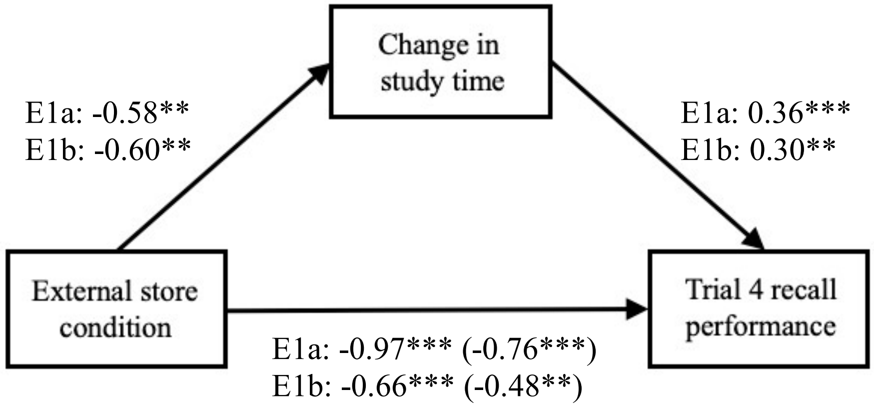
Mediation analyses

As outlined in our preregistration, we conducted three sets of mediation analyses testing whether the relation between external store condition and Trial 4 recall performance was mediated by (1) Change in study time from Trial 3 to Trial 4, (2) Trial 4 study time, and (3) Selfreported strategy use. We used of the *mediation* package in R (Tingley et al., 2014) which, through the use of non-parametric bootstrapping procedures, provides estimates of the indirect effect, its 95% confidence intervals, and the significance. We follow the same convention as before in reporting the analyses involving study time or change in study time, with outliers removed [and without outliers removed]. Mediation model figures depict the results with outliers removed.

Figure 3

The effect of External store condition on Trial 4 recall performance is partially mediated by the Change in study time between Trials 3 and 4

in Experiment 1a were qualitative and coded. We streamlined this procedure in Experiment 1b by having participants first respond with either *yes* or *no* options to using a strategy during final trial recall and only if participant responded yes were they given opportunity to describe their strategy. Thus, only data from participants responding *yes* to using a strategy were qualitative and coded in Experiment 1b.

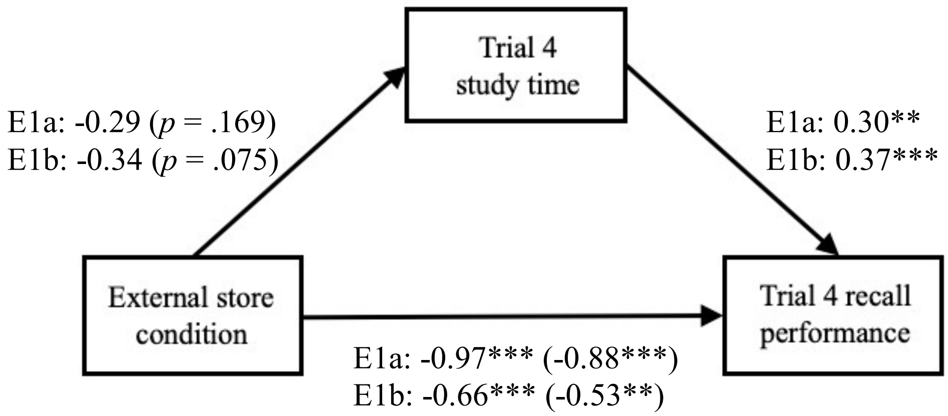


Note. Standardized regression coefficients for the relation between external store condition and Trial 4 recall performance as mediated by change in study time for each experiment. The standardized regression coefficient between external store condition and Trial 4 recall performance controlling for change in study time is in parentheses (outlier data of study time not included). *$p < .05$, **$p < .01$, ***$p < .001$.

As depicted in Figure 3, change in study time (Trial 4 – Trial 3) significantly mediated the relation between external store condition and Trial 4 recall performance. The total effect of external store condition on Trial 4 recall performance was significant (**E1a:** $b = -0.97$, $SE = 0.19$, $t = -5.23$, $p < .001$, [$b = -0.97$, $SE = 0.19$, $t = -5.23$, $p < .001$]; **E1b:** $b = -0.66$, $SE = 0.18$, $t = -3.63$, $p < .001$, [$b = -0.67$, $SE = 0.18$, $t = -3.70$, $p < .001$]). This total effect is the same for all mediations reported, therefore, we do not report the result in full again in the mediation analyses to follow. The indirect effect was significant (**E1a:** -0.21, $p = .004$ [-0.26, $p < .001$]; **E1b:** -0.18, $p = .002$ [-0.10, $p = .032$]). The direct effect was also significant (**E1a:** -0.76, p < .001 [-0.71, p < .001]; **E1b**: -0.48, $p = .006$ [-0.56, $p = .004$]), demonstrating that change in study time did not fully mediate the effect of external store. The proportion of the effect of external store condition on recall performance mediated by change in study time was 0.21 [0.27] in Experiment 1a and 0.27 [0.16] in Experiment 1b.

Figure 4

The effect of External store condition on Trial 4 recall performance is partially mediated by Trial 4 study time

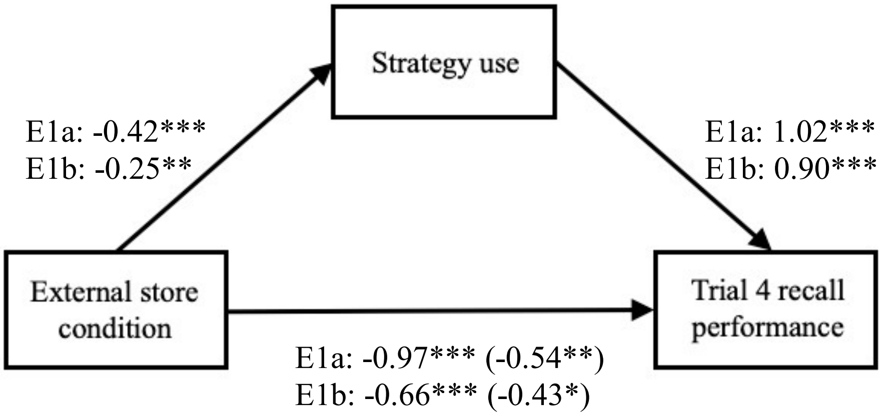


Note. Standardized regression coefficients for the relation between external store condition and Trial 4 recall performance as mediated by Trial 4 study time for each experiment. The standardized regression coefficient between external store condition and Trial 4 recall performance controlling for Trial 4 study time is in parentheses (outlier data of study time not included). $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

As depicted in Figure 4, in testing study time as a mediator through the effect of external store condition on Trial 4 recall, the indirect effect was not significant despite being in the predicted direction (**E1a:** -0.09, $p = .210$ [-0.13, $p = .110$]; **E1b:** -0.12, $p = .056$, [-0.12, $p = .080$]). Provided the total effect (**E1a:** -0.97; **E1b:** -0.66) reported earlier, the direct effect was significant (**E1a:** -0.88, $p < .001$ [-0.84, $p < .001$]; **E1b**: -0.53, $p = .006$ [-0.55, $p = .002$]).

Figure 5

The effect of External store condition on Trial 4 recall performance is partially mediated by Strategy use



Note. Standardized regression coefficients for the relation between external store condition and Trial 4 recall performance as mediated by self-reported strategy use for each experiment. The standardized regression coefficient between external store condition and Trial 4 recall performance controlling for strategy use is in parentheses (outlier data of study time not included). $^*p < .05$, $^{**}p < .01$, $^{***}p < .001$.

Finally, as depicted in Figure 5, the effect of strategy use mediated the relation between external store condition and Trial 4 recall performance. The indirect effect was significant (**E1a:** -0.43, $p < .001$; **E1b:** -0.23, $p = .006$). The direct effect was also significant (**E1a:** -0.54, $p = .006$; **E1b:** -0.43, $p = .010$). The proportion of the effect of external store condition on recall mediated by strategy was .45 in Experiment 1a and .35 in Experiment 1b.

Exploratory

The following analyses were not preregistered.

Correlational analyses. We conducted three exploratory Pearson correlational analyses to investigate the relation between change in study time, Trial 4 study time, and reported strategy use. In both experiments, as expected, Trial 4 study time correlated significantly with change in study time (**E1a:** $r = .56$, $t(88) = 6.28$), $p < .001$, [$r = .55$, $t(88) = 6.17$), $p < .001$]; **E1b:** $r = .47$, $t(108) = 5.52$, $p < .001$ [$r = .59$, $t(108) = 7.50$, $p < .001$]). Trial 4 study time did not correlate significantly with reported strategy in Experiment 1a, but did in Experiment 1b when excluding outliers (**E1a:** $r = .19$, $t(88) = 1.82$, $p = .072$; **E1b:** $r = .21$, $t(108) = 2.28$, $p = .024$). When including outliers, the correlation was significant in Experiment 1a and not in Experiment 1b (**E1a:** $r = .25$, $t(88) = 2.39$, $p = .019$, **E1b:** $r = .18$, $t(108) = 1.89$, $p = .062$) but results were qualitatively consistent with those when excluding outliers. Change in study time correlated significantly with reported strategy Experiment 1a but not in Experiment 1b (**E1a:** $r = .21$, $t(88) = 2.00$, $p = .048$ [$r = .22$, $t(88) = 2.09$, $p = .039$]; **E1b:** $r = .17$, $t(108) = 1.80$, $p = .074$ [$r = .07$, $t(108) = 0.73$, $p = .468$]).

General Discussion

Memory for to-be-remembered information is often worse if, at study, we anticipate the support of an external memory store. The present work provided a direct test of the study-effort hypothesis of this cost (Kelly & Risko 2019a; 2019b; Lu et al., 2020; Park et al. 2021) by examining study time allocation and self-reports of study strategies. Not surprisingly, there was a significant effect of external store condition on memory performance, like that found in related studies (Eskritt & Ma, 2014; Kelly & Risko 2019a; 2019b; Lu et al., 2020; Park et al. 2021; Sparrow et al., 2011). Novel to the present effort, we also found direct evidence for the studyeffort hypothesis, such that when individuals are provided the impression that they can rely on an external memory support for future memory testing, there is less effort devoted to studying the stored information. That is, when individuals thought that they could rely on an external store they allocated less study time per item and were less likely to report using a memory strategy to aid recall. Across both investigations, we also examined the extent to which these effects mediated the relation between external store condition and Trial 4 recall performance. There was clear evidence that change in study time from Trial 3 to Trial 4 (between which individuals found out whether or not they could continue to rely on their external store), and self-reported strategy use partially mediated the relation between external store condition and recall performance. This was not the case with Trial 4 absolute study time as a mediator, but the estimate was in the predicted direction. In general, these results are consistent with the study effort hypothesis of the cost associated with relying on an external memory store.

While the mediation analyses across both experiments revealed that study time allocation and strategy mediated the effect of external store condition on free recall performance, this effect was clearly only partial. Indeed, a sizable portion of the effect appears to *not* be due to these mediators alone. This implies two non-mutually exclusive possibilities: the current indices did not fully capture effort devoted to encoding and/or the cost of external memory store reliance is caused by more than just this withdrawal of study effort. With respect to the former possibility, the current measures of effort during study are (of course) imperfect. For example, in the present experiments, participants were always expected to type out the study items to have them stored in the external memory store. This requirement could obscure the relation between study effort and recall performance as it imposes a minimal time allocated at study when it is plausible that individuals would allocate even less time without such a requirement. From this perspective, the present work may be a conservative test of the influence of external store reliance on study effort and subsequent memory performance. In addition, self-reported strategy use is limited to strategies for which individuals can report and remember deploying and may be subject to demand effects (i.e., participants reporting engaging in a strategy that they did not). Given that the strategy question came after recall, individual's recall performance may well have impacted their strategy reports. It is also worth considering whether participants understand what is meant by "strategies". We opted to avoid feeding participants ideas and found that their responses were still easily coded into the strategy types adapted from previous work (Dunlosky & Kane, 2007; Unsworth et al., 2019). Moreover, those who reported strategy use tended to perform higher. Taken together, it appears participants understood the strategy prompt. Future research employing different measures of effort during study (e.g., psychophysiological measures such as pupil dilation; Papesh et al., 2012; van der Wel & Steenbergen, 2018) would be valuable to further examine the study-effort hypothesis.

Beyond limitations in measuring effort at study, the partial mediations might also indicate that the cost of external store reliance cannot be explained fully by the withdrawal of such effort. Thus, the cost might be multiply determined. This raises exciting new questions about additional potential contributions to the cost of expecting access to an external store. While the current investigation examined a study-based mechanism, the cost of external store reliance may also be driven by processes at retrieval. For example, the memory cost could partly result from feelings of surprise or alarm once individuals realize that the external store is not available when they expected it to be. This momentary experience of surprise or alarm (e.g., similar to realizing you lost a note containing important information) may be enough to compromise one's ability to recover the studied information. For example, Wessel and colleagues (2016) found that accuracy on a working memory task was significantly reduced when the task followed a surprising tone compared to a standard tone. They suggested that this might reflect inhibitory mechanisms engaged by surprise. Relatedly, the concomitant worry about one's performance could also, itself, interfere with performance. Task-unrelated processes, such as worries or anxiety about performance, could occupy working memory, impairing memory performance (e.g., Eysenck et al., 2007; Moran, 2016). Thus, the present investigation both confirms an effort-based contribution while also opening the door to the future examination of contributions heretofore not considered.

Finally, we would be remiss if we did not address the possibility that the scenario that we constructed in the current empirical work using list memory does not translate perfectly to how individuals use external memory stores in their day-to-day lives. This limitation of the current work could pose some difficulty in interpreting the current findings in the context of our day to day lives. That said, we believe the paradigm used captures key features about how stores are used in the “real world”. For example, many of us regularly trust and rely on various external stores to complete our memory goals. Our trust trials are intended to capture this aspect of regularly using external devices. Many of us also know the feeling of being unable to retrieve some key information kept on those stores in the face of misplacing them. Thus, although we are constructing a lost-phone scenario in the laboratory, we believe the findings from implementing this paradigm could provide deeper understanding of the mechanisms involved (or not) when relying on external memory supports.

Previous related work has proposed that the reliable cost of external memory store reliance is driven by a reduced intentional engagement in remembering strategies at study—the *study-effort hypothesis*. To date, this hypothesis has acted as little more than an inference and thus, in the current investigation, we sought to directly test it. We have found evidence not only confirms this hypothesis as an explanation for the robust memory cost under offloading conditions, but also evidence that suggests it explains a modest portion of this memory cost.

Conclusion

The current work provides further insights into the mechanism underlying the memory cost associated with external store reliance. The present work is has made two important theoretical contributions (1) it has provided the first direct evidence for the study-effort hypothesis of the cost of external store reliance and (2) it has provided the first evidence consistent with the idea that the cost of external store reliance may be caused by factors other than variations in study effort (or that the measures of study effort used here fail to capture it completely). The latter discovery opens the door to the consideration of novel contributions to the cost of external store reliance and to a deeper understanding of the costs and benefits of external memory stores.

Acknowledgments

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) from the Canada Research Chairs (#950-232147) program to E.F.R., and an Alexander Graham Bell Canada Graduate Scholarship from NSERC to M.O.K.

References

Ariel, R., Price, J., & Hertzog, C. (2015). Age-related associative memory deficits in value-based remembering: The contribution of agenda-based regulation and strategy use. *Psychology and Aging, 30*(4), 795 – 808.

Bailey, H., Dunlosky, J., & Kane, M. J. (2008). Why does working memory span predict complex cognition? Testing the strategy affordance hypothesis. *Memory & Cognition, 36*, 1383–1390.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Beaudoin, M., & Desrichard, O. (2017). Memory self-efficacy and memory performance in older adults: The mediating role of task persistence. *Swiss Journal of Psychology, 76*(1), 23-33.

Bower, G. H. (1972). Mental imagery and associative learning. In L. W. Gregg (Ed.), Cognition in learning and memory (pp. 51–88). New York, NY: Wiley.

Brown, V. A. (2021). An introduction to linear mixed-effects modeling in R. *Advances in Methods and Practices in Psychological Science, 4*(1), 1-19.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41,* 977-990.

Bulley, A., McCarthy, T., Gilbert, S. J., Suddendorf, T.,& Redshaw, J. (2020). Children devise and selectively use tools to offload cognition. *Current Biology, 30*(17), 3457 – 3464.

Clark, A. (2010). Supersizing the Mind. Oxford, UK: Oxford University Press.

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

Dunlosky, J., & Kane, M. J. (2007). The contributions of strategy use to working memory span: A comparison of strategy assessment methods. *The Quarterly Journal of Experimental Psychology, 60*(9), 1227-1245.

Dunlosky, J., & Thiede, K. W. (1998). What makes people study more? An evaluation of factors that affect self-paced study. *Acta Psychologica, 98*(1), 37-56.

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

Eysenck, M. W., Santos, R., Derakshan, N., & Calvo, N. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion*, 7(2), 336-353.

Finley, J. R., Naaz. F., & Goh, F. W. (2018). Memory and technology: How we use information in the brain and the world. Springer.

Froger, C., & Bouazzaoui, B., Isingrini, M., & Taconnat, L. (2012). Study time allocation deficit of older adults: The role of environmental support at encoding? *Psychology and Aging, 27*(3)*,* 577-588.

Gilbert, S., Boldt, A., Sachdeva, C., Scarampi, C., & Tsai, P. (2022). Outsourcing memory to external tools: A review of 'intention offloading'. https://doi.org/10.31234/osf.io/ahqtz

Grinschgl, S., & Papenmeier, F., & Meyerhoff, H. S. (2021). Consequences of cognitive offloading: Boosting performance but diminishing memory. *Quarterly Journal of Experimental Psychology, 74*(9), 1477-1496.

Henkel, L. A. (2014). Point-and-shoot memories: The influence of taking photos on memory for a museum tour. Psychological Science, 25(2), 396–402

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

Kelly, M. O., & Risko, E. F. (2021). Revisiting the influence of offloading memory on free recall. *Memory & Cognition*, 1-12.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1-26.

Lu, X., Kelly, M. O., & Risko, E. F. (2020). Offloading information to an external store increases false recall. *Cognition, 205,* 104428.

Lu, X., Kelly, M. O., & Risko, E. F. (2021). The gist of it: ofoading memory does not reduce the benefit of list categorisation. *Memory, 30*(4), 1-16. https://doi.org/10.1080/ 09658211.2021.1989465

Marsh, E. J., & Rajaram, S. (2019). The Digital Expansion of the Mind: Implications of Internet Usage for Memory and Cognition. *Journal of Applied Research in Memory and Cognition*. https://doi.org/10.1016/j.jarmac.2018.11.001

Mazzoni, G., & Cornoldi, C. (1993). Strategies in study time allocation: Why is study time sometimes not effective? *Journal of Experimental Psychology: General, 122*(1), 47-60.

Metcalfe, J., & Kornell, N. (2005). A region of proximal learning model of study time allocation. *Journal of Memory and Language, 52*, 463-477.

Meyerhoff, H. S., Grinschgl, S., Papenmeier, F., & Gilbert, S. J. (2021). Individual differences in cognitive offloading: A comparison of intention offloading, pattern copy, and short-term memory capacity. *Cognitive Research: Principles and Implications, 6*(1), 34. https://doi.org/10.1186/s41235-021-00298-x

Moran, T. P., (2016). Anxiety and working memory capacity: A meta-analysis and narrative review. *Psychological Bulletin, 142*(8), 831-864.

Goldinger S. D., & Papesh, M. H. (2012). Pupil dilation reflects the creation and retrieval of memories. *Current Directions in Psychological Science, 21*(2), 90-95.

Park, J., & Kelly, M. O., & Risko, E. F. (2021). Park, J., Kelly, M. O., Hargis, M., & Risko, E. F. (under review). Influence of external store reliance on predicted and actual value-directed remembering. Manuscript submitted for publication.

Pereira, A. E., Kelly, M. O., Lu, X., & Risko, E. F. (2021). 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. https://doi. org/10.1080/09658211.2021.1990347

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R project.org/

Richardson, J. T. E. (1998). The availability and effectiveness of reported mediators in associative learning: A historical review and an experimental investigation. *Psychonomic Bulletin & Review, 5,* 597–614.

Risko, E. F. & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Science*, 20(9), 676–688.

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

Risko, E. F., Kelly, M. O., Lu, X., & Pereira, A. (in press). Varieties of Offloading Memory: A Framework. In Q. Wang & A. Hoskins, The Remaking of Memory in the Age of Social Media and the Internet. Oxford University Press.

Rundus, D. (1971). Analysis of rehearsal processes in free recall. Journal of Experimental Psychology, 89(1), 63–77. https://doi.org/10.1037/h0031185

Shenhav, A., Fahey, M. P., & Grahek, I. (2021). Decomposing the motivation to exert mental effort. *Current Directions in Psychological Science, 30*(4), 307-314.

Sachdeva, C., & Gilbert, S. J. (2020). Excessive use of reminders: Metacognition and effortminimisation in cognitive offloading. *Consciousness and Cognition, 85*, 103024. https://doi.org/10.1016/j.concog.2020.103024

Scarampi, C., & Gilbert, S. J. (2020). The effect of recent reminder setting on subsequent strategy and performance in a prospective memory task. *Memory, 28*(5), 677–691. https://doi.org/10.1080/09658211.2020.1764974

Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. Science. 333, 776-778

Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(1), 204-221.

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science, 11*(5), 702-712.

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.

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. Mediation: R package for causal mediation analysis. *Journal of Statistical Software, 59*(5).

Unsworth, N. (2016). Working memory capacity and recall from long-term memory: Examining the influences of encoding strategies, study time allocation, search efficiency, and monitoring abilities. Journal of Experimental Psychology: Learning, Memory, and Cognition, 42, 50–61.

Unsworth, N., Miller, A. L., & Robison, M. K. (2019). Individual differences in encoding strategies and free recall dynamics. *Quarterly Journal of Experimental Psychology, 72*(10), 2495-2508.

Van der Wel, P., & van Steengergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic Bulletin & Review, 25,* 2005-2015.

Van Heuven, W.J.B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, 67, 1176-1190.

Van Selst, M., Jolicoeur, P. (1994). A solution to the effect of sample size on outlier elimination. *The Quarterly Journal of Experimental Psychology, 47A*(3), 631-650.

Wessel, J. R., Jenkinson, N., Brittain, J-S., Voets, S. H. E. M., Aziz, T. P., & Aron, A. R. (2016). Surprise disrupts cognition via a fronto-basal ganglia suppressive mechanism. *Nature Communications, 7*, 11195.