Memory & Cognition
https://doi.org/10.3758/s13421-025-01688-y

Does expecting external memory support cost recognition memory?

Megan O. Kelly[1] · Batul Karimjee[1] · April E. Pereira[1] · Xinyi Lu[1] · Evan F. Risko[1]

Accepted: 10 January 2025

Abstract
We often use tools and aids to help us achieve our cognitive goals – that is, we often *offload* to external supports. One such variety of offloading is the use of external memory stores (e.g., phones, computers, notebooks, calendars) to support memory. Recent work aimed at better understanding the consequences of offloading memory on aspects of *unaided* memory have revealed a clear cost to unaided memory performance when an external memory store is unexpectedly lost, but this work has focused on examining this cost in free-recall paradigms. Using key theoretical differences between recall and recognition, we sought to examine the influences of expecting external memory supports in a recognition memory context across five preregistered experiments, finding evidence for a small cost to unaided recognition memory. We found evidence for a specific cost in recollection (Experiments 2, 3a, and 3b). When testing the effects of expecting external memory support on indices of study effort, there was a reduction to study time which partially mediated the relation between expecting support and memory performance indices, consistent with earlier work using free recall. Individuals did not predict a cost to memory of losing expected support in recognition, contrasting earlier work using free recall.

Keywords Recognition · Memory · Offloading · Recollection · Study effort

Introduction

Every day, we encounter information that we hope to remember. One effective way that we manage these memory demands is to use external memory supports. For example, we may make to-do lists or record important dates into an agenda to reduce forgetting. This approach often affords us the information that we intend to remember without the same required cognitive demands otherwise involved in remembering on our own, internally (i.e., a form of *cognitive offloading;* Risko & Gilbert, 2016). Despite our long history of integrating such external support (e.g., Nestojko et al., 2013), we have only recently begun focusing on better understanding this approach to "remembering" and the underlying memory processes with which it is associated.

To better understand memory processes in the context of using external memory supports, researchers have examined the influence of expecting such support on memory performance. Here, individuals are often tasked to remember studied information for a memory test but are also promised access to an external memory support at the time of testing (e.g., often a computer file; Eskritt & Ma, 2014; Kelly & Risko, 2019a, 2019b, 2022a, 2022b; Lu et al., 2020, 2022; Murphy, 2023a; Richmond et al., 2023; Risko & Kelly, 2024; Risko et al., 2019, 2024; Sparrow et al., 2011; Storm & Stone, 2015). In one variety of these external support paradigms, prior to study, half of the participants are instructed that their externally stored (yet-to-be-remembered) information will be available at test to aid remembering whereas the other half are instructed that their stored information will be *unavailable*. Critically, *all* individuals perform the memory test without access to the stored information (i.e., unaided). Of primary interest is how unaided memory differs depending on whether participants were instructed to expect or *not* to expect their stored information to be available.

In these experiments, participants reliably demonstrate significantly poorer memory performance for the to-be-remembered information when told that their stored information will be available compared to when told it will be unavailable. This relative cost to unaided memory performance associated with those expecting external support has been shown to be quite robust, often appearing as a large or very large effect by various standards (i.e., Cohen's d of

✉ Megan O. Kelly
megan.kelly@princeton.edu

[1] Department of Psychology, University of Waterloo, Waterloo, ON, Canada

Published online: 31 January 2025

0.80 or higher; e.g., Cohen, 1988; Kelly & Risko, 2019a, 2019b, 2022a, 2022b).[1] Thus, one consequence of expecting external memory support is a clear cost to *unaided* internal memory performance (Kelly & Risko, 2019a, 2019b, 2022a, 2022b; Lu et al., 2020, 2022; Murphy, 2023a; Park et al., 2022; Risko et al., 2019; Sparrow et al., 2011).

Explaining the cost

Recent work has argued that the relative cost associated with expecting external memory support is attributable to expectations of support leading to reduced internal mnemonic efforts at study (e.g., Eskritt & Ma, 2014; Kelly & Risko, 2019a, 2019b, 2022a, 2022b; Park et al., 2022). This effort-based explanation draws from suggestions that expending effort is avoided unless necessary (Kool et al., 2010) and that expected payoffs of investing effort ought to outweigh or be worth the cost of investing effort (Shenhav et al., 2021). In the context of a memory task, retrieval success would typically constitute a desired payoff of investing costly internal study efforts to remember. When an external memory support containing the to-be-remembered information is presumably available come time to "remember" the to-be-remembered information, equivalent (or better) retrieval success can be expected *without* the investment of study effort simply through access to the external memory support. Hence, according to a study effort hypothesis, individuals are less likely to invest study effort when expecting the aid of an external memory support.

Kelly and Risko (2022b) reported support for the idea that the cost associated with expecting external support is driven by differential mnemonic efforts at study. Critically, Kelly and Risko also found that, although a study effort explanation could explain some of the cost, it could not explain *all* of the cost – that is, study effort indices *partially* mediated the relation between expecting external support (or not) and the resulting *unaided* memory performance. Further evidence for the idea that study effort represents only a partial explanation, Kelly and Risko (*submitted*) demonstrated that expecting *partial* support (participants received the beginning portions of to-be-remembered items) demonstrated a clear cost to *unaided* memory performance when compared to expecting *no* support, but, critically, individuals who expected to receive that partial support seemed *not* to significantly differ in their study effort from those expecting no external support. Therefore, the cost to memory performance demonstrated by those who lost *partial* support suggests a cost to memory that cannot easily be explained by reduced study efforts. Taken together, differential levels of study effort appear to explain *some* but not *all* of the cost to unaided memory associated with expecting external memory support.

The discussed findings have provided various insights into the evident cost to unaided memory associated with expecting external support; they have all done so in the context of a recall paradigm. Although free recall and recognition memory often do demonstrate highly similar effects (e.g., temporal manipulations such as study time, massing, and spacing, and retention interval; Anderson & Bower, 1972; Kintsch, 1966; Olson, 1969), there exist several contexts wherein recognition and recall are affected differently or to different extents by the same manipulation (e.g., effect of age-related decline – Danckert & Craik, 2013; word frequency – Schwarts & Rouse, 1961; Shepard, 1967; semantic association – Cofer, 1967; Kintsch, 1966; Anderson & Bower, 1972; incidental vs. intentional learning – Anderson & Bower, 1972; Dornbush & Winnick, 1967; Eagle & Leiter, 1964; Postman et al., 1955; although see Popov & Dames, 2023; list-method-directed forgetting – Sahakyan et al., 2009, 2013; value-directed remembering – Murphy, 2023b), thus suggesting that they differ in theoretically important ways.

One salient difference between free recall and recognition that is especially relevant to the current work is that recognition tests provide more environmental support during retrieval than recall tests do. That is, there is a higher degree of feature overlap between the environmental stimuli or context at study and at retrieval for recognition than for recall (Craik, 1994). Retrieval on a recognition test, hence, requires less effortful, self-initiated processing (Craik, 1994). Given the differences between recall and recognition, here we extend the investigation of the potential underlying mechanisms of external memory support use to recognition memory.

The idea that expecting external memory support would be associated with a cost in unaided recognition memory draws indirect support from the item-method-directed forgetting literature where directed forgetting is apparent in both free recall and recognition (item method; Basden et al., 1993; MacLeod, 1999). Directed forgetting has been likened to offloading memory (e.g., Eskritt & Ma, 2014; Kelly & Risko, 2019a; Lu et al., 2020) because participants are presented with items that are subsequently paired with an instruction to either remember or forget (*remember items* and *forget items*, respectively, hereon in). In both directed forgetting and the current offloading memory procedures, internal memory is tested for items that presumably need

[1] Although some have argued that there is ambiguity in interpreting even standardized effect sizes, we believe that our doing so in the present work is warranted given that (i) we outline the specific context in which we are examining the effects of interest and (ii) we are intending to compare the standardized effects found in free recall and recognition, rather than on their own without context. Indeed, this aligns with Laken's (2013) suggestion that "the best way to interpret Cohen's *d* is to relate it to other effects in the literature…".

not be committed to internal memory. That is, the forget items in directed forgetting could be likened to the externally stored items in offloading paradigms (with the remember items being analogous to internally stored items).

It is noteworthy that the study effort hypothesis – that expecting external support leads to reduced intentional study efforts, resulting in a cost to unaided memory – is consistent with a leading account of the intentional/directed forgetting effect, the *selective rehearsal* account (Bjork, 1972; MacLeod, 1975; Sheard & MacLeod, 2005; Tan et al., 2020; see also Fellner et al., 2020). According to this account, the difference between remember and forget items in memory performance is driven by "remember" items getting selectively more rehearsal (i.e., intentional study effort) than "forget" items, which can enhance both free recall and recognition (e.g., Rundus, 1971; Woodward et al., 1973).

Whereas we might expect an effect of offloading on recognition memory, there is also good reason to expect that it may be much less robust than in recall. Previous work comparing incidental and intentional memory suggests that investing intentional efforts at study may have *less* influence on recognition compared to free recall (Anderson & Bower, 1972; Dornbush & Winnick, 1967; Eagle & Leiter, 1964; Postman et al., 1955; though see Craik, 2023, Popov & Dames, 2023). For example, intentional learning instructions lead to better memory performance on free recall tests than do incidental learning instructions, whereas this has been far less clear for recognition memory. Instead, incidental learning instructions often lead to recognition memory performance that is not worse (and is sometimes better) than under intentional learning instructions (Anderson & Bower, 1972; Dornbush & Winnick, 1967; Eagle & Leiter, 1964; Postman et al., 1955). According to the study effort hypothesis, the cost of offloading is a product of the withdrawal of intentional effort at study, so based on this body of work such a withdrawal ought to have greater consequences for recall than for recognition.

Recognition memory tests also provide a unique opportunity to distinguish between levels of remembering (Basden & Basden, 1996; Jacoby et al., 1997; Mangels et al., 2001; Ochsner, 2000; Ozubko et al., 2012; Yonelinas, 2002; Yonelinas & Jacoby, 1995). Here, individuals respond to targets and foils by indicating whether they *recollect* the item as old (e.g., remember studying it episodically), or whether it is *familiar* (e.g., they know it to be studied although they cannot remember it episodically), or new (the item was not studied). The ability to recollect in this context is thought to require considerable attention toward study items. Manipulations at study have been found to affect recollection items (items reported as "remember") differently from familiarity items (reported as "know"). For example, Basden and Basden (1996) found an effect of directed forgetting on items reported as *Remember*[2] but not on items reported as *Know*. According to a study effort hypothesis, expecting external support should lead to lower rates of recollection than expecting no such support.

The relation between study effort and familiarity is arguably less clear (and as such so are the predictions for familiarity estimates based on the study effort hypothesis). Jacoby's (1991) interpretation of familiarity as largely a result of automatic, unintentional memory processes would lead the study effort hypothesis to predict little to no cost of expecting external support on familiarity estimates. However, expecting external support could also demonstrate costs in familiarity if expectations of support also reduce these more automatic, unintentional processes. That said, prior work has found that memory phenomena less dependent on intentional efforts to remember internally (e.g., recency, von Restorff/isolation effects, gist-based memory) tend to be less affected by expecting external support than memory phenomena more dependent on intentional efforts to remember (primacy, verbatim-based memory; Kelly & Risko, 2019a, 2019b, 2022a; Lu et al., 2022). The study effort hypothesis specifically in combination with Jacoby's (1991) interpretation of familiarity, along with prior findings, would suggest there should be little cost of expecting support on familiarity processes.

Present investigation

We investigated whether expecting external memory support influenced recognition performance across five preregistered experiments. Across experiments, we followed the same general procedure wherein each experiment consisted of five trials. At the beginning of each experiment, participants were instructed to write down (or type) lists of to-be-remembered items. During the initial three trials, participants were instructed that their lists of to-be-remembered words would be available to facilitate their memory at test, which was indeed the case. Thus, these first three trials served to establish trust in the external store. Participants were told at the beginning of the study that they would always have access to their external stores except on *one* trial, but that they would be given notice of this before this trial starts (before study/encoding). Critically, on the fourth *and* fifth trials, their external stores (the lists) were unavailable at test. Participants were only given advance notice of this on *one* of the two trials. Our main interest is in comparing recognition performance when participants were told that they could expect their external store (told-store) to performance when they were told *not* to expect their external store (told-no-store).

[2] They referred to these items as *recollect* items; however, they take on the role of *remember* items as described currently.

In addition to investigating the influence of expecting external memory support on unaided recognition memory generally, we investigated its effect on indices of recollection and familiarity (Experiments 2, 3a, and 3b; RKN paradigm; Basden & Basden, 1996; Jacoby et al., 1997; Mangels et al., 2001; Ochsner, 2000; Ozubko et al., 2012; Yonelinas, 2002; Yonelinas & Jacoby, 1995). According to the study effort hypothesis, expecting external memory support should lead to reduced recollection.

In the final experiment, we followed Kelly and Risko (2022b) by examining the influence of external store expectation on study time allocation as a direct test of the study effort hypothesis in the context of a recognition test. We also extended the investigation to follow earlier work (e.g., Lu et al., 2022; Park et al., 2022) in considering whether expectations of external memory support influenced expectations of recognition memory performance and metacognitive accuracy in the final experiment.

Experiment 1

Experiment 1 was preregistered via the Open Science Framework at osf.io/e2bxm.

Method

Participants

Data were collected from 40 undergraduate students from the University of Waterloo, who completed this study in exchange for course credit. The preregistered N of 40 was based on an a priori power analysis with a desired power of 0.80 ($\alpha = 0.05$, two-tailed) to detect a Cohen's d of 0.45 for the difference between hit rate in the told-no-store and told-store conditions using the dependent-samples t-test from G*Power. This effect size was chosen as a conservative approximation of *half* the effect size usually found in the proportion of freely recalled items between told-no-store and told-store conditions.

Stimuli

Stimuli comprised ten 30-item word lists generated randomly from a stimulus set of 300 (available at osf.io/y36p2). Word lengths ranged from four to ten letters and word frequencies ranged from three (*inartistic*) to 82,060 (*second*) using FREQCount from SUBTLEX-UK (Van Heuven et al., 2014) and from one (*inartistic*) to 21,384 (*friend*) using FREQCount from SUBTLEX-US (Brysbaert & New, 2009).[3] Lists were counterbalanced across trial position (i.e., first through fifth) and across target and foil designations.

Procedure

The experimental task was administered on desktop computers via E-Prime version 3.0 stimulus presentation software. Participants were invited in groups of three to four and sat at individual workstations separated by dividers. They were instructed that they would learn and save a set of to-be-remembered words and then engage in a recognition memory test for which they would be given the saved list of items at test to help them in the memory test. Participants were told that they would repeat this procedure multiple times but that for *one* of these trials, they would *not* have an external memory support available at test (be unaided) but that they would be given advance notice of when this would happen (i.e., before the start of the trial). Participants engaged in a total of five trials, three *trust* trials wherein the external support was available at test as instructed, and, afterward, the two critical trials wherein the external memory support expectation manipulation (told-store vs. told-no-store) occurred. The order of these external support conditions was counterbalanced. Each trial comprised two main phases: a study phase and a recognition test phase.

Study In the study phase, participants were presented with a list of 30 words one at a time; these were displayed in white on a black background at the center of the screen. Participants were provided pen and paper during each study phase to record the study words in list format. Each word was presented for 3 s with an interstimulus interval of 2.5 s, and participants were instructed to write down each word as it appeared on the screen so that it would be saved to their list, which they would be able to use at test.

Recognition After a brief retention interval of 15 s, participants completed a recognition test which consisted of a total of 60 words, 30 studied items (targets) and 30 new items (foils) that were randomly intermixed. Participants were to respond to each item with either OLD (to identify a target – studied prior) or NEW (to identify a foil – *not* studied prior). As mentioned, for the three *trust trials* (the initial three trials), all participants had access to their external supports (i.e., their saved list of to-be-remembered items) as they were instructed at study. This portion of the task was to help participants establish a sense of trust that the external

[3] The SUBTLEX-US word frequencies were not available for the following words: *midway, uncle, avenue, chapstick, foundation, carpenter, stitch, harbor, matrix, raven*; the SUBTLEX-UK word frequency was not available for the word *campground.*

supports would be available when indicated given that this might not have matched their expectations for what participation in a memory study would entail. Importantly, on the critical trials (the final two trials), participants were *not* provided with their external supports unlike in the trust trials. Half of the participants were told about this beforehand right before the study phase (told-no-store condition) while the other half were not given the promised advanced notice of this (told-store). On the second critical trial (the final trial), this was also the case. Because participants were told ahead of time that there would only be one recognition test wherein they would not have their support, the logic was that those who had the told-no-store trial *first* would understand that to be the *one* test that they would not be supported for (as indicated in the instructions) and that this would not matter for those in the told-store first condition as once they found out they would *not* get the support when told they would, any consequence of violating their trust would have little influence given that they only had the told-no-store trial left.

At the end of the study, participants were asked two final questions to gain further insight into how they might be using the external supports in the context of a recognition memory test (similar to post-task prompts of Kelly & Risko, 2022a, who used a free-recall procedure). The first one was: "*Question 1. On ONE of the trials, you were NOT told ahead of time that you would not have access to your saved list. On this particular trial, what was your EXPECTED strategy on the recognition test?*" to which participants would respond with [1: Rely only on saved list; 2: Rely mostly on saved list; 3: Rely equally on saved list and own memory; 4: Rely mostly on own memory; 5: Rely only on own memory]. The second question was: "*Question 2. On ONE of the trials, you WERE told ahead of time that you would not have access to your saved list. On this particular trial, what was your EXPECTED strategy on the recognition test?*" to which participants responded along the same scale as in Question 1. As reported by Kelly and Risko (2022a), we would expect that for the told-store condition, the rating (from external-to-internal) would be lower than that for the told-no-store condition – that is, closer to the exclusively internal end of the scale (closer to the "1: Rely only on saved list" endpoint) and hence further from the exclusively internal end of the scale (further from "5: Rely only on own memory" endpoint).

Results

We preregistered the replacement of participant data if they showed that they were unable to follow procedures or task instructions (e.g., they did not write down the right words during encoding/study), or if performance was at or below chance. No participants needed replacing for not following procedures. While performance for three participants was below chance, we could not replace them in time before work from home restrictions due to COVID-19. Moreover, upon further reflection, that criterion seemed potentially overly restrictive, and retaining those people also allowed more power as intended by the preregistered N of 40. All results were qualitatively the same regardless of their inclusion/exclusion. Data from one participant was lost due to technical difficulties with the E-prime program software bringing the final *N* to 39.

Statistical analyses were conducted using R version 4.3.1 with ANOVAs conducted using the *ez* package (Lawrence, 2016), mixed-effects models conducted using *lme4* (Bates et al., 2015) and *lmerTest* (Kuznetsova et al., 2017), and *t*-tests and mixed-model fitting for random effects structures conducted using the *stats* package (R Core Team, 2022). For relevant analyses on sensitivity (*d'*), extreme values of 0 or 1 for hit rate and false alarm rate were corrected by adding 0.5 to both hit and false alarm counts to both targets and foils to prevent *d'* values approaching negative or positive infinity (*loglinear* correction method based on Stanislaw & Todorov, 1999). Note that we deviate from the preregistration by reporting exploratory analyses comparing performance indices on the final trust trial with that of the critical trials. Any other deviations from the preregistration (osf.io/e2bxm) are also specified. Table 1 presents the means across key dependent variables during trust trials and as a function of external store condition for the critical trials. Data and analyses code for Experiment 1 are available via the Open Science Framework at osf.io/y36p2.

Hit rate

A paired-samples *t*-test revealed no difference between the told-no-store condition and the told-store condition [$t(38) = 1.38$, $p = 0.174$ $d = 0.22$]. Analogous mixed-effects logistic regression revealed that participants in the told-no-store condition were significantly more likely to respond correctly to a target [$b = 0.39$, $SE = 0.17$, $z = 2.37$, $p = 0.018$]. Comparing final trust trial performance (i.e., when the external support was available) with critical trial performances (i.e., when the external support was unavailable),

Table 1 Experiment 1: Means and confidence intervals (CIs) for hit rate, false alarm rate, and sensitivity across trials and conditions

	Hit rate	False alarm rate	Sensitivity
Trust Trial 1	0.85 [0.75, 0.88]	0.05 [0.02, 0.17]	2.85 [2.12, 3.11]
Trust Trial 2	0.86 [0.83, 0.89]	0.05 [0.03, 0.08]	2.86 [2.61, 3.10]
Trust Trial 3	0.85 [0.81, 0.88]	0.05 [0.03, 0.08]	2.91 [2.63, 3.19]
Told-no-store	0.85 [0.84, 0.91]	0.08 [0.05, 0.11]	2.66 [2.38, 2.94]
Told-store	0.82 [0.80, 0.87]	0.09 [0.07, 0.14]	2.43 [2.14, 2.74]

CIs are bias corrected, bootstrap 95% CIs using 10,000 samples

paired-samples t-tests revealed no statistically significant differences between the final trust trial hit rate and either external store condition (trust: 0.85; told-no-store: 0.84; told-store: 0.82) [trust vs. told-no-store: $t(38) = 0.24$, $p = 0.809$, $d = 0.04$; trust vs. told-store: $t(38) = 1.86$, $p = 0.071$, $d = 0.30$], although hit rate was numerically higher on the trust trial in both cases.

False alarm rate

A paired-samples t-test found no significant difference between external support conditions for false alarm rate [$t(38) = 1.11$, $p = 0.274$, $d = 0.18$]. Analogous mixed-effects logistic regression revealed the same result [$b = 0.01$, $SE = 0.30$, $z = 0.03$, $p = 0.974$]. The final trust trial false alarm rate was significantly lower than the false alarm rate for either external store condition (trust: 0.05; told-no-store: 0.08; told-store: 0.09) [trust vs. told-no-store: $t(38) = 2.70$, $p = 0.010$, $d = 0.43$; trust vs. told-store: $t(38) = 2.46$, $p = 0.019$, $d = 0.39$].

Sensitivity (d′)

There was no significant difference between external support conditions for discriminating between targets and foils [$t(38) = 1.64$, $p = 0.110$, $d = 0.26$]. The final trust trial sensitivity was significantly higher than sensitivity for either external store condition (trust: 2.91; told-no-store: 2.63; told-store: 2.43) [trust vs. told-no-store: $t(38) = 2.48$, $p = 0.018$, $d = 0.40$; trust vs. told -store: $t(38) = 3.65$, $p = 0.001$, $d = 0.59$].

Correct response time (ms)

The told-no-store condition had significantly quicker correct responses than those in the told-store condition (told-no-store: 829; told-store 890) [$t(38) = 2.23$, $p = 0.032$, $d = 0.36$].

Post-task prompts

As mentioned at the end of the *Procedure* section, participants were prompted with two questions meant to elicit the degree to which they intended to use external supports during the final two critical trials. Exploratory analyses (not preregistered) comparing the difference in rating between external support questions revealed that there was no significant difference in ratings [$t(37) = 0.47$, $p = 0.644$, $d = 0.08$]. Of the told-store condition, 72% responded that they would rely mostly or only on their own memory, whereas 69.2% said the same in the told-no-store condition. Table 2 below presents the breakdown for internal–external ratings for Experiment 1.

Discussion

Experiment 1 tested whether expectations of external memory support would influence recognition memory – that is, whether the cost of external memory store availability found consistently in free recall would also manifest in recognition memory. Although most of the analyses on the memory indices were not statistically significant, the results of Experiment 1 do suggest that recognition memory can be negatively affected by expecting external memory support. Those told that they could expect external support demonstrated a significantly lower hit rate (in the mixed-effects model but not in the t-test), and *numerically* lower sensitivity than those expecting no support. Those told that they could expect support also had significantly slower correct response times, suggesting that expecting external support influences, at some level, the ease with which one can remember the externally stored information. Thus, the results from Experiment 1 provide some preliminary support for a cost from expecting external memory support on unaided recognition memory. That said, the effects found would generally be considered “small” (e.g., $d < 0.40$) especially compared to the large “offloading” effects often reported in free-recall paradigms (where sometimes $d > 1.00$; e.g., Kelly & Risko, 2019a, 2019b, 2022a, 2022b).

Of note in the current work is the similarity between performance on trust trials (where the external support was available) and critical trials (where external support was not available). Exploratory (not preregistered) analyses demonstrated that, although hit rate was not significantly different on trust trials compared to critical trials, the false alarm rate and sensitivity showed significant differences in the predicted directions (i.e., higher false alarm rates and lower sensitivity on critical trials). However, this performance benefit associated with having the external store available was

Table 2 Experiment 1: Mean ratings and proportions for each internal–external rating options as a function of prompt type (i.e., Q1: told store available vs. Q2: told no store available)

	Mean rating (1: list to 5: own memory)	1: Rely only on saved list	2: Rely mostly on saved list	3: Rely equally on saved list and own memory	4: Rely mostly on own memory	5: Rely only on own memory
Told store	3.87	2.6%	10.3%	12.8%	43.6%	28.2%
Told no store	4.00	0.0%	23.1%	7.7%	15.4%	53.8%

negligible and much smaller than has been found in research using a recall task.

One explanation for the similarity in performance when the external store was available (trust trials) and when it was unavailable (critical trials) is that, arguably, participants were not relying on the external support during trust trials as much as they could have, given that they were not performing at ceiling with the supports on the trust trials (e.g., above 95%, as in the case with free recall in earlier work, e.g., Kelly & Risko, 2019a, 2019b; Kelly & Risko, 2022a, 2022b; Lu et al., 2022; Park et al., 2022). Ceiling performance is clearly possible given that the external store is available. This idea seems to garner further support from the ratings participants gave regarding their memory strategy. Whether told to expect the external store or not, they reported relying heavily on their own memory. Indeed, there was no difference in the mean ratings of reported reliance (from 1: external to 5: internal) across conditions.

Kelly and Risko (2022a) asked the same question in a recall memory context and found that the mean rating in the told-no-store condition was significantly greater compared to that of the told-store condition (4.39 vs. 2.02 whereas here, 4.00 vs. 3.87). Furthermore, they found that 77% of those in the told-store condition reported a reliance mostly or exclusively on the external support (i.e., their saved list; see Table 2 of Kelly & Risko, 2022a) compared to the 13% of those in the told-store condition in the current work. In interpreting participants' reported memory strategy, it is important to consider the possibility that the participants misunderstood the questions. For example, for the told-no-store condition, 31% of participants reported expecting to rely equally or mostly on their saved list, which they should have understood would not have been possible for that trial. Nonetheless, these results add some additional context to the potential differences in how people are using external supports in the context of recognition memory. In particular, these results point to potential differences in how people use external memory stores as a function of the demands of the memory task (e.g., they might rely less on an external support in a recognition context relative to a recall context) as well as potential differences in the consequences of that use (e.g., relying on an external support may have lesser costs in a recognition context than a recall context).

Experiment 2

Experiment 1 provided some (but certainly not overwhelming) support for an effect of expecting external support on unaided recognition memory performance. In Experiment 2, we aimed to further test the effect of external store condition as a function of two ways of remembering in the context of recognition: recollection and familiarity (Basden & Basden, 1996; Jacoby et al., 1997; Mangels et al., 2001; Ochsner, 2000; Ozubko et al., 2012; Yonelinas, 2002; Yonelinas & Jacoby, 1995).

Recollection and familiarity

From a dual-process perspective, recollection and familiarity represent two contributing processes to responses made during a recognition test (for a review, see Yonelinas, 2002; see also Donaldson, 1996; Dunn, 2004; Yonelinas et al., 2010). Recollection is thought to be a slower, more conscious retrieval of the encoding context whereas familiarity is thought to reflect a quicker, less detailed feeling. One method that has been used to index this distinction in recognition responses is the Remember/Know/New procedure (Jacoby et al., 1997; Mangels et al., 2001; Ochsner, 2000; Ozubko et al., 2012; Yonelinas & Jacoby, 1995). Here, participants learn the distinction between items and respond with either R (*remember*), K (*know*), or N (*new*) to each item at test, with these responses used to index recollection and familiarity processes during recognition.

The ability to recollect information is dependent on how richly one encodes the original encoding instance – for example, elaborative processing increases recollection but not familiarity (Rajaram, 1993; survival processing – Cho et al., 2018). If the offloading cost is due, at least in part, to a lack of controlled study efforts to encode and maintain the stored information as predicted by a study effort hypothesis, then this cost should be apparent in reports of recollection such that recollection estimates should be lower when expecting external support compared to when not expecting support. In contrast, the study effort hypothesis would predict that there is little cost to familiarity estimates, provided familiarity is less dependent on intentional, effortful processes (Jacoby, 1991). Indirect support for these predictions come from the findings of Basden and Basden (1996), who found an effect of directed forgetting on items reported as Remember[4] whereas items reported as Know showed no directed forgetting effect. Experiment 2 was preregistered via the Open Science Framework at osf.io/ap2tr.

Method

The method of Experiment 2 was the same as that of Experiment 1 apart from the implementation of the Remember/Know/New procedure instead of the standard Yes/No recognition test used in Experiment 1. Here, participants were asked to make a "remember," "know," or "new" response by pressing "R", "K," or "N", respectively, on the keyboard.

[4] They referred to these items as *recollect* items; however, they take on the role of *remember* items as described currently.

There was a 500-ms blank screen between successive words, with each word appearing on the screen until the participant pressed one of the above keys. Participants were trained to make "remember" responses when they had a conscious recollection of specific details/information related to the initial presentation of the word and to make a "know" response if they believed the word to have been presented before but did not remember specific details/information about its initial presentation. Lastly, they were asked to make a "new" response if they believed that they had not seen the word in the study before. If needed, clarification was provided.

Data were collected and analyzed from $N = 55$ undergraduate students from the University of Waterloo who completed this study in exchange for course credit. Note that we originally preregistered a stopping rule of $N = 60$ based on an a priori power analysis with a desired power of 0.80 ($\alpha = 0.05$, two-tailed) to detect a Cohen's d of ~ 0.35 for the difference between overall *sensitivity* (d') in the told-no-store and told-store conditions using the dependent-samples t-test from G*Power. However, due to the COVID-19 pandemic and work-from-home protocols, we had to stop our collection prior to reaching the final N. Despite our final sample size being less than the intended sample size, we deemed it reasonable to move forward with analyses given the prior Experiment 1 sample size and the intent to replicate the investigation online in future experiments.

As mentioned earlier, we implemented the independent Remember/Know procedure to estimate the processes of recollection ("remember" responses) and familiarity ("know" responses) independently (i.e., so that they do not sum to 1.00; e.g., Jacoby et al., 1997; Mangels et al., 2001; Ochsner, 2000; Ozubko et al., 2012; Yonelinas & Jacoby, 1995). Here, recollection estimates reflect the proportion of "remember" responses out of the total responses, thus, $R = [\text{p(“remember”)}]$; and familiarity estimates reflect the *proportion* of "know" responses divided by the *proportion* of *non*-remember responses: $F = \text{p(“know”)} / [1 - \text{p(“remember”)}]$. These estimates were each computed separately for targets and foils for each external support condition. Note that this means that if participants responded with all "remember" responses, familiarity estimates could not be derived given a denominator of 0, hence, any participants who did so were excluded from the analyses involving familiarity estimates, and we report when this was necessary.

Results

As in Experiment 1, our preregistered exclusion criteria included participants who (i) were unable to follow procedures or task instructions (e.g., they did not write down the right words during encoding/study), or (ii) performed at or below chance. Again, we relaxed the latter criterion, preserving power and avoiding a potentially overly restrictive criterion. All results were qualitatively the same regardless of their inclusion/exclusion. For analyses involving familiarity, six participants were excluded because they had 100% recollection responses, which would mean a divisor of 0 because the divisor for the familiarity estimate calculation is 1 – p("remember").

Note that to better evaluate the effect of offloading on items reported as "remember" and "know," we deviated from the preregistration by conducting additional simple-effects analyses. Specifically, the preregistered analyses indicated the use of a 2 (told-store vs. told-no-store) × 2 (recollection vs. familiarity) ANOVA; however, because we were interested in the potential effects of expecting support on each of recollection and familiarity *regardless* of whether these two effects differed from each other (i.e., regardless of the significance of the interaction in the 2 × 2 ANOVA), we followed up the ANOVA with separate paired-samples t-tests on recollection and familiarity estimates to examine the effect of expecting external support (we did this in Experiments 3a and 3b as well despite not preregistering these analyses initially). Table 3 presents the means across key dependent variables during trust trials and as a function of external store condition for the critical trials. Data and analyses code for Experiment 2 are available via the Open Science Framework at osf.io/y36p2.

Hit rate

A paired-samples t-test revealed that the told-no-store condition had a higher hit rate than the told-store condition [$t(54) = 2.15$, $p = 0.036$, $d = 0.30$] and analogous mixed-effects logistic regression confirmed the same result [$b = 0.40$, $SE = 0.17$, $z = 2.42$, $p = 0.016$]. Exploratory analyses revealed no difference between the final trust trial and told-no-store critical trial hit rate (trust: 0.85; told-no-store: 0.86; told-store: 0.82) [$t(54) = 0.50$, $p = 0.617$, $d = 0.07$], although the trust trial hit rate was significantly higher than that of the told-store condition [$t(54) = 2.16$, $p = 0.035$, $d = 0.29$].

Table 3 Experiment 2: Means and confidence intervals (CIs) for Hit rate, False alarm rate, and Sensitivity across trials and conditions

	Hit rate	False alarm rate	Sensitivity
Trust Trial 1	0.84 [0.80, 0.87]	0.10 [0.07, 0.16]	2.52 [2.32, 2.70]
Trust Trial 2	0.84 [0.81, 0.86]	0.08 [0.05, 0.14]	2.65 [2.46, 2.82]
Trust Trial 3	0.85 [0.82, 0.88]	0.07 [0.04, 0.14]	2.73 [2.51, 2.94]
Told-no-store	0.86 [0.85, 0.90]	0.14 [0.11, 0.24]	2.53 [2.26, 2.79]
Told-store	0.82 [0.81, 0.87]	0.14 [0.11, 0.24]	2.29 [2.05, 2.51]

CIs are bias corrected, bootstrap 95% CIs using 10,000 samples

False alarms

A paired-samples t-test revealed no significant difference between conditions for false alarm rate [$t(54) = 0.47$, $p = 0.642$, $d = 0.06$]. Analogous mixed-effects logistic regression confirmed the same result [$b = 0.12$, $SE = 0.18$, $z = 0.68$, $p = 0.498$]. The trust trial false alarm rate was significantly lower than the false alarm rate for either external store condition (trust: 0.07; told-no-store: 0.14; told-store: 0.14) [trust vs. told-no-store: $t(54) = 4.44$, $p < 0.001$, $d = 0.60$; trust vs. told-store: $t(54) = 4.04$, $p < 0.001$, $d = 0.54$].

Sensitivity (d′)

Those in the told-no-store condition demonstrated greater sensitivity than those in the told-store condition [$t(54) = 2.06$, $p = 0.045$, $d = 0.28$]. Trust trial sensitivity was significantly higher than sensitivity for either external store condition (trust: 2.73; told-no-store: 2.46; told-store: 2.23) [trust vs. Told-no-store: $t(54) = 2.36$, $p = 0.022$, $d = 0.32$; trust vs. told-store: $t(54) = 4.48$, $p < 0.001$, $d = 0.60$].

Correct response time (ms)

There was no significant difference in correct response times between external support conditions (told-no-store: 1,119; told-store: 1,162) [$t(55) = 1.19$, $p = 0.238$, $d = 0.16$].

Recollection and familiarity

Targets A 2 (told-no-store vs. told-store) × 2 (recollection vs. familiarity) ANOVA demonstrated no main effect of condition (told-no-store: 0.53; told-no-store: 0.53) [$F(1, 48) < 0.01$, $p = 0.984$, $\eta_G^2 < 0.01$], no significant main effect of process (recollection: 0.58; familiarity: 0.48) [$F(1, 48) = 2.12$, $p = 0.152$, $\eta_G^2 = 0.02$], and no significant interaction such that the effect of external store expectations was not significantly different between recollection and familiarity estimates (recollection: 0.04; familiarity: = −0.04) [$F(1, 48) = 3.43$, $p = 0.070$, $\eta_G^2 < 0.01$]. Paired-samples t-tests (not preregistered) on the effect of expecting support on recollection and familiarity estimates revealed no clear effect of expecting support on recollection [$t(48) = 1.68$, $p = 0.100$, $d = 0.24$] or familiarity [$t(48) = 1.05$, $p = 0.299$, $d = 0.15$] estimates.

Foils A 2 (told-no-store vs. told-store) × 2 (recollection vs. familiarity) ANOVA demonstrated no main effect of condition (told-no-store: 0.09; told-store: 0.09), $F(1, 48) < 0.01$, $p = 0.970$, $\eta_G^2 < 0.01$, a main effect of process such that the estimates for recollection were lower than for familiarity (recollection: 0.06; familiarity: 0.11) [$F(1, 48) = 10.43$, $p = 0.002$, $\eta_G^2 = 0.03$], and no interaction (recollection: < 0.01; familiarity: < 0.01) [$F(1, 48) = 0.23$, $p = 0.633$, $\eta_G^2 < 0.01$]. Paired-samples t-tests (not preregistered) on recollection and familiarity estimates revealed no effect of expecting support on recollection [$t(48) = 0.48$, $p = 0.634$, $d = 0.07$] or familiarity [$t(48) = 0.28$, $p = 0.781$, $d = 0.04$] estimates.

Post-task prompts

Exploratory analyses (not preregistered) comparing the difference in rating between external support questions revealed that there was no significant difference in mean rating in Experiment 2 [$t(51) = 0.47$, $p = 0.640$, $d = 0.07$]. Table 4 presents the breakdown for internal–external ratings for Experiment 2.

Discussion

In Experiment 2, we found results consistent with Experiment 1 that more clearly supported a cost of external memory support on recognition. We found a clear effect of condition on hit rate, and no clear effect on false alarm rate. Unlike in Experiment 1, there was a significant effect of condition on sensitivity, and no effect on correct response times. That is, in Experiment 2, we found additional evidence of a small cost ($d = \sim 0.30$) to recognition memory (through hit rate and sensitivity) when expecting external support, but no clear evidence that such expectations matter for correct response times.

Critically, we sought to extend the examination of Experiment 1 by breaking recognition memory down further into recollection and familiarity processes in Experiment 2. This was based on the idea that a study effort hypothesis would predict that recollection estimates would be lower for those

Table 4 Experiment 2: Mean ratings and proportions for each internal–external rating option as a function of prompt type (i.e., Q1: told store available vs. Q2: told no store available)

	Mean rating (1: list to 5: own memory)	1: Rely only on saved list	2: Rely mostly on saved list	3: Rely equally on saved list and own memory	4: Rely mostly on own memory	5: Rely only on own memory
Told store	3.87	3.6%	9.1%	14.5%	38.2%	30.9%
Told no store	3.96	5.5%	7.3%	20%	18.2%	47.3%

expecting support than for those expecting no such support. We did not find much evidence for this idea. When testing this for studied items ("targets"), there was no significant interaction between support condition and process type, although the differences (0.04 vs. −0.04) suggested that while *recollection* estimates were lower when *expecting* support (compared to expecting no support), *familiarity* estimates were lower when expecting *no* support (compared to expecting support). The effect of expecting support was not significant for recollection or familiarity in isolation.

As in Experiment 1, the similarity between performance on trust trials wherein the external support was available and the critical trials suggest that participants might not be relying heavily on the external store, especially compared to behavior in a recall context (e.g., Kelly & Risko, 2019a, 2019b, 2022a, 2022b; Lu et al., 2020, 2022; Morrison & Richmond, 2020; Park et al., 2022; Risko & Dunn, 2015). Specifically, exploratory analyses demonstrated that hit rate was not significantly different for the third trust trial compared to the told-no-store condition (i.e., performance was the same on a trial where the participants had the to-be-remembered words in front of them and a trial where they did not). Importantly, hit rate was significantly lower in the told-store condition than on the trust trial (i.e., when they did not have the words in front of them, but they thought they would). Furthermore, like in Experiment 1, false alarm rate and sensitivity showed significant differences in the predicted directions (i.e., higher false alarm rates and lower sensitivity on critical trials). Thus, performance was better with the external store, but not much better. Also like Experiment 1, participants did not report relying heavily on the external store and did not significantly differ across the told-store and told-no-store conditions in how much they reporting relying on the external store.

As mentioned, in recall, there are clear and large differences between these conditions (Kelly & Risko, 2022a). Again, there could have been some confusion in interpreting the questions, given that for the told-no-store condition, 33% of participants implied expecting to rely equally, mostly, or only on their saved list, which, again, they should have understood would not have been possible for that trial. Taken together it seems that individuals are, at least to some extent, using the support on the trust trials though clearly not as much as in free recall (Kelly & Risko, 2022a).

Experiments 3a and 3b

Experiment 2 demonstrated clearer evidence than Experiment 1 that expecting external support influences recognition memory. However, there was little support for a study effort hypothesis in terms of an effect of expecting support on recollection processes. In Experiments 3a and 3b, we aimed to put these results on a stronger footing via two replications with larger sample sizes. Experiment 3a was replicated in Experiment 3b, thus, the method and results are described together. Experiments 3a and 3b were preregistered via the Open Science Framework at osf.io/5p8m4 and osf.io/kx835, respectively.

Method

Participants

Participants were undergraduate psychology students taking part for course credit. For Experiment 3a, data from $N = 90$ participants (44.4% female, 13.3% male; $M_{age} = 21.92$ years, $SD_{age} = 4.71$ years)[5] were collected and analyzed based on an a priori power analysis with a desired power of 0.80 ($\alpha = 0.05$, two-tailed) to detect a Cohen's d of ~ 0.30 for the difference between sensitivity in the told-no-store and told-store conditions using a dependent-samples t-test in G*Power. For Experiment 3b, we collected and analyzed $N = 120$ participants (71.1% female, 23.9% male, 2.5% nonbinary/nonconforming; $M_{age} = 19.93$ years, $SD_{age} = 4.01$ years) based on an a priori power analysis with a desired power of 0.80 ($\alpha = 0.05$, two-tailed) to detect a Cohen's d of ~ 0.25 for the difference between sensitivity in the told-no-store and told-store conditions.

Adapting to online procedure

The stimuli in Experiments 3a were those of earlier experiments, just with five fewer words per list (randomly determined) such that lists were 25 items long. In Experiment 3b, however, we returned to using 30-item word lists using the same 300 words from Experiments 1 and 2 but redrawing items randomly to generate ten novel lists. Unlike Experiments 1 and 2, the remaining experiments were collected online due to work-from-home protocols given the status of COVID-19 pandemic. The procedure for Experiments 3a and 3b was largely the same as earlier experiments except for minor amendments to allow for online data collection. These amendments consisted of programming the experiments in JavaScript including with the *JSPsych* framework (de Leeuw et al., 2023) to allow for collection via internet browsers. Instead of writing to-be-remembered words in a pen and paper list, participants typed words into a textbox during one-at-a-time item presentation at study so that we could collect their saved information. Words typed within 6 s were saved to a list on the left or right of the screen (randomly

[5] Demographic information was unavailable for 42.2% of participants in Experiment 3a and 2.5% of participants in Experiment 3b; percentages may not add up[to 100% due to rounding.

determined at the level of the participant) to mimic an accumulating list of items much like the procedure of using pen and paper in earlier experiments. A final critical difference from earlier studies was additional exclusion criteria to help encourage data quality in the context of online collection (see Rodd, 2024, for a general guide). We elaborate on these exclusion criteria in the *Results* section below.

Results

All analyses followed the preregistration unless otherwise specified. As mentioned, due to moving to online data collection for Experiments 3a and 3b, we preregistered the following replacement of participants who met these exclusion criteria: (1) did not type at least 85% of words for trials 2–4 during encoding; (2) did not reach at least 80% hit rate for the offloading words during the first three trials when they had access to their typed lists; (3) indicated that they were not paying attention or did not make an effort during the task (e.g., doing something else during the experiment; via self-report); (4) performance during trials 4 and/or 5 was at or below chance; (5) answered yes to writing/screen-capturing any words to aid memory outside of the means offered within the experiment; (6) indicated that they would not like their data to be used. We opted to relax criterion (4) as in earlier experiments (results are qualitatively the same regardless of their inclusion/exclusion).

In Experiment 3a, 13 participants were replaced due to not encoding at least 85% of study words across trials 2–5, and eight participants were replaced due to not reaching at least an 80% hit rate on trials wherein external support was available (i.e., during the trust trials). In Experiment 3b, seven participants were removed due to not encoding at least 85% of study words across trials 2–5, and 22 participants were removed due to not meeting the minimum of 80% hit rate on trials wherein external support was available (i.e., trust trials). One extra participant was collected in Experiment 3b, and we include them in the results below (results do not change with their exclusion).

As mentioned earlier, we were interested in the potential effects of expecting external support on recollection and familiarity, regardless of whether these two effects differed, thus we deviated from the preregistration by also conducting paired-samples *t*-tests on recollection and familiarity to examine the effect of expecting external support as we had done in Experiment 2. Eight participants in Experiment 3a and five participants in Experiment 3b were not included in the analyses involving familiarity estimates because they had 100% recollection responses, which, as mentioned earlier, makes the computation for the familiarity estimate undefined given the divisor for the familiarity estimate calculation is 1 – p("remember"). Table 5 below presents the means across key dependent variables during trust trials and as a function of external store condition for the critical trials. Data and analyses code for Experiments 3a and 3b are available via the Open Science Framework at osf.io/y36p2.

Hits

There was no significant difference in hit rate between the told-store and told-no-store instruction conditions in Experiment 3a (told-no-store: 0.85; told-store: 0.83) [$t(89) = 1.69$, $p = 0.094$, $d = 0.18$] but there was in Experiment 3b (told-no-store: 0.86; told-store: 0.83) [$t(120) = 2.86$, $p = 0.005$, $d = 0.26$]. Analogous mixed-effects models revealed the same results [E3a: $b = 0.15$, $SE = 0.12$, $z = 1.23$, $p = 0.219$; E3b: $b = 0.29$, $SE = 0.10$, $z = 2.80$, $p = 0.005$]. In Experiment 3a, exploratory analyses revealed hit rate was higher in the trust trial than for either external store condition (trust: 0.92; told-no-store: 0.85; told-store: 0.83) [trust vs. told-no-store: $t(89) = 6.18$, $p < 0.001$, $d = 0.65$; trust vs. told-store: $t(89) = 7.24$, $p < 0.001$, $d = 0.76$. The same was found for Experiment 3b [trust vs. told-no-store: $t(120) = 3.42$, $p = 0.001$, $d = 0.31$; trust vs. told-store: $t(120) = 5.40$,

Table 5 Experiments 3a and 3b: Means and confidence intervals (CIs) for Hit rate, False alarm rate, and Sensitivity across trials and conditions

	Hit rate	False alarm rate	Sensitivity
Trust Trial 1	E3a: 0.92 [0.90, 0.93] E3b: 0.92 [0.91, 0.93]	E3a: 0.03 [0.02, 0.05] E3b: 0.04 [0.02, 0.06]	E3a: 3.28 [3.16, 3.41] E3b: 3.33 [3.19, 3.45]
Trust Trial 2	E3a: 0.92 [0.90, 0.93] E3b: 0.90 [0.89, 0.92]	E3a: 0.02 [0.01, 0.04] E3b: 0.03 [0.02, 0.05]	E3a: 3.29 [3.16, 3.41] E3b: 3.19 [3.07, 3.30]
Trust Trial 3	E3a: 0.92 [91, 0.94] E3b: 0.90 [0.88, 0.91]	E3a: 0.03 [0.02, 0.05] E3b: 0.04 [0.03, 0.07]	E3a: 3.26 [3.12, 3.39] E3b: 3.13 [2.99, 3.26]
Told-no-store	E3a: 0.85 [0.83, 0.88] E3b: 0.86 [0.84, 0.88]	E3a: 0.11 [0.08, 0.14] E3b: 0.13 [0.11, 0.17]	E3a: 2.49 [2.37, 2.79] E3b: 2.46 [2.25, 2.59]
Told-store	E3a: 0.83 [0.80, 0.86] E3b: 0.83 [0.82, 0.86]	E3a: 0.12 [0.08, 0.14] E3b: 0.15 [0.12, 0.19]	E3a: 2.47 [2.28, 2.65] E3b: 2.22 [2.08, 2.42]

CIs are bias corrected, bootstrap 95% CIs using 10,000 samples

$p<0.001$, $d=0.49$] (trust: 0.90; told-no-store: 0.86; told-store: 0.83).

False alarms

There was no significant difference in false alarm rate between the told-no-store and told-store instruction conditions in Experiment 3a [$t(89)=1.31$, $p=0.195$, $d=0.14$] or Experiment 3b [$t(120)=1.96$, $p=0.052$, $d=0.18$]. Analogous mixed-effects logistic regression revealed qualitatively the same result in Experiment 3a [$b=0.19$, $SE=0.11$, $z=1.73$, $p=0.083$], but for Experiment 3b revealed that those in the told-no-store were more likely to incorrectly identify a foil as "old" [$b=0.31$, $SE=0.11$, $z=2.83$, $p=0.005$]. For Experiment 3a, false alarm rate was lower in the trust trial than for either external store condition (trust: 0.03; told-no-store: 0.11; told-store: 0.12) [trust vs. told-no-store: $t(89)=6.74$, $p<0.001$, $d=0.71$; trust vs. told -store: $t(89)=6.65$, $p<0.001$, $d=0.70$]. The same was found for Experiment 3b (trust: 0.04; told-no-store: 0.13; told-store: 0.15) [trust vs. told-no-store: $t(120)=8.21$, $p<0.001$, $d=0.75$; trust vs. told-store: $t(120)=8.96$, $p<0.001$, $d=0.81$].

Sensitivity

There was no significant difference in sensitivity between the told-no-store and told-store instruction conditions in Experiment 3a [$t(89)=1.41$, $p=0.161$, $d=0.15$], but there was in Experiment 3b [$t(120)=3.04$, $p=0.003$, $d=0.28$]. For Experiment 3a, sensitivity was higher in the trust trial than for either external store condition (trust: 3.26; told-no-store: 2.49; told-store: 2.37) [trust vs. told-no-store: $t(89)=8.97$, $p<0.001$, $d=0.95$; trust vs. told -store: $t(89)=9.76$, $p<0.001$, $d=1.03$]. Again, the same was found for Experiment 3b (trust: 3.13; told-no-store: 2.46; told-store: 2.22) [trust vs. told-no-store: $t(120)=8.28$, $p<0.001$, $d=0.75$; trust vs. told-store: $t(120)=11.14$, $p<0.001$, $d=1.01$].

Correct response time (ms)

There was no significant difference in response times for correct responses between conditions in Experiment 3a (told-no-store: 1,512; told-store: 1,534) [$t(89)=1.42$, $p=0.159$, $d=0.15$] but there was a significant difference in Experiment 3b such that those in the told-no-store condition had slower correct response times (told-no-store: 1,603; told-store: 1,478) [$t(120)=2.17$, $p=0.032$. $d=0.20$].

Process estimates

Targets In Experiment 3a, a 2 (told-no-store vs. told-store) × 2 (recollection vs. familiarity) ANOVA demonstrated a main effect of condition, such that process estimates were higher in general for the told-no-store condition than for the told-store condition (told-no-store: 0.59; told-store: 0.55) [$F(1, 89)=4.29$, $p=0.041$, $\eta_G^2<0.01$]. There was also a main effect of process such that the estimates were lower for recollection than for familiarity (recollection: 0.51; familiarity: 0.62) [$F(1, 89)=5.76$, $p=0.019$, $\eta_G^2=0.04$]. The effect of external store did not differ between processes (recollection: 0.05; familiarity: 0.02) [$F(1, 89)=1.05$, $p=0.309$, $\eta_G^2<0.01$]. Paired-samples *t*-tests (not preregistered) on recollection and familiarity estimates revealed a significant effect of expecting support on recollection estimates [$t(84)=2.86$, $p=0.005$, $d=0.31$] but not on familiarity estimates [$t(84)=0.72$, $p=0.382$, $d=0.08$].

In Experiment 3b, a 2 (told-no-store vs. told-store) × 2 (recollection vs. familiarity) ANOVA demonstrated no main effect of condition on estimates (told-no-store: 0.57; told-store: 0.55) [$F(1, 115)=2.12$, $p=0.148$, $\eta_G^2<0.01$], and no main effect of process on estimates (recollection: 0.57; familiarity: 0.54) [$F(1, 115)=0.39$, $p=0.532$, $\eta_G^2<0.01$]. The effect of external store condition did not differ between processes (recollection: 0.03; familiarity: 0.01) [$F(1, 115)=0.35$, $p=0.555$, $\eta_G^2<0.01$]. Paired-samples *t*-tests (not preregistered) on estimates revealed a significant effect of expecting support on recollection [$t(115)=2.17$, $p=0.032$, $d=0.20$] but not on familiarity [$t(115)=0.24$, $p=0.662$, $d=0.04$].

Foils In Experiment 3a, a 2 (told-no-store vs. told-store) × 2 (recollection vs. familiarity) ANOVA demonstrated no main effect of condition (told-no-store: 0.05; told-store: 0.06) [$F(1, 84)=1.96$, $p=0.166$, $\eta_G^2<0.01$]. There was a main effect of process such that the estimates were lower for recollection than for familiarity (recollection: 0.03; familiarity: 0.09) [$F(1, 84)=27.45$, $p<0.001$, $\eta_G^2=0.09$]. The effect of external store did not differ between processes (recollection: −0.01; familiarity: −0.01;) [$F(1, 84)<0.01$, $p=0.957$, $\eta_G^2<0.01$]. Paired-samples *t*-tests (not preregistered) on estimates revealed no effect of expecting support on recollection or familiarity estimates in Experiment 3a [recollection: $t(84)=1.52$, $p=0.132$, $d=0.17$; familiarity: $t(84)=0.88$, $p=0.382$, $d=0.10$].

In Experiment 3b, a 2 (told-no-store vs. told-store) × 2 (recollection vs. familiarity) ANOVA demonstrated no main effect of condition (told-no-store: 0.07; told-store: 0.08) [$F(1, 115)=2.32$, $p=0.130$, $\eta_G^2<0.01$]. There was a main effect of process such that the estimates were lower for recollection than for familiarity (recollection: 0.06; familiarity: 0.10) [$F(1, 115)=8.33$, $p=0.004$, $\eta_G^2=0.02$]. External store condition did not differ between processes (recollection: −0.01; familiarity:—0.01) [$F(1, 115)=0.06$, $p=0.815$, $\eta_G^2<0.01$]. Paired-samples *t*-tests (not preregistered) on estimates revealed no effect of expecting support on recollection or familiarity estimates in Experiment 3b [recollection: $t(115) = 0.95$, $p = 0.343$, $d = 0.09$; familiarity: $t(115) = 1.20$, $p = 0.233$, $d = 0.11$].

Post-task prompts

As done in prior experiments, the exploratory analyses compared the difference in rating between external support questions. These analyses revealed that the mean rating was significantly lower in the told-store condition than in the told-no-store condition (i.e., towards the "1: Rely only on saved list" endpoint of the scale) in both Experiments 3a [$t(88) = 6.32$, $p < 0.001$, $d = 0.67$] and 3b [$t(119) = 9.61$, $p < 0.001$, $d = 0.88$]. Many participants still reported a reliance mostly or only on internal memory in both Experiments 3a (53.3%) and 3b (52.9%) for the told-store condition. Table 6 below presents the breakdown for internal–external ratings for Experiments 3a and 3b.

Discussion

In Experiments 3a and 3b, we aimed to elucidate the effects of external support on recognition memory performance replicating the method of Experiment 2 in an online setting with increased power. Overall, the evidence for an influence of expecting external memory support on recognition memory was mixed. There was no significant cost to hit rate or sensitivity in Experiment 3a but there was a significant cost to hit rate and sensitivity in Experiment 3b, the latter paralleling the results of Experiments 1 and 2. There was no effect of external store condition on false alarm rate across Experiments 3a or 3b, except for the mixed model analysis of Experiment 3b suggesting a higher likelihood of false alarming for those expecting external support. The effects of expecting external support on correct response times appear unclear given that Experiment 3a suggested no significant effect of expecting support on correct response time (like Experiment 2), which was not the case in Experiment 1 (wherein expecting no support demonstrated significantly quicker correct response times) or in Experiment 3b (wherein expecting no support demonstrated significantly slower response times).

Overall, a clear picture seems to be emerging despite some inconsistencies across analyses in terms of statistical significance. With respect to hit rate and sensitivity across all experiments so far, performance has been *consistently* numerically lower for those told that they can expect external support compared to those told to expect no support – that is, in the direction of a cost to expecting external support. The effect was statistically significant in some cases but not in others. Thus, it would seem reasonable to argue for a *small* cost of expecting external support on recognition memory performance.

In Experiments 3a and 3b, we continued our examination from Experiment 2 of the potential influence of expecting external support on recollection and familiarity more specifically and with increased power. Neither Experiment 3a nor Experiment 3b demonstrated significant interactions between external store condition and recognition process although, in both experiments, familiarity demonstrated no effects of external store condition while recollection estimates demonstrated a cost in the told-store condition compared to the told-no-store condition. Thus, there is some evidence that expectations of external support affect recollection. What is less clear from the study effort hypothesis is whether expectations of external support influence recollection significantly more than familiarity. That is, familiarity might increase with intentional study effort processes; however, it could also be driven by unintentional processes in the current context. Together, the results of Experiments 2, 3a, and 3b suggest that if there *is* a differential influence of external support expectations on recollection versus familiarity estimates, it is small and difficult to consistently detect. Indeed, this conclusion parallels the conclusions we can draw from the general influence of external support expectations on recognition memory performance. If the effect exists, it appears to be small and difficult to reliably detect, although the numerical patterns in the data are relatively consistent across the four experiments so far.

Table 6 Experiments 3a and 3b: Mean ratings and proportions for each internal–external rating options as a function of prompt type (i.e., Q1: told store available vs. Q2: told no store available)

	Mean Rating (1: list to 5: own memory)	1: Rely only on saved list	2: Rely mostly on saved list	3: Rely equally on saved list and own memory	4: Rely mostly on own memory	5: Rely only on own memory
E3a						
Told-store	3.42	5.6%	13.3%	27.8%	40.0%	13.3%
Told-no-store	4.42	0.0%	6.7%	7.8%	22.2%	62.2%
E3b						
Told-store	3.42	4.1%	18.2%	24.9%	38.0%	14.9%
Told-no-store	4.60	0.0%	2.5%	8.3%	16.5%	72.7%

As done in the earlier experiments, we also conducted exploratory analyses on whether performance differed on the final trust trial compared to the critical trials (wherein the told-no-store vs. told-store condition manipulations occurred). The results were much clearer here than in the earlier experiments, with clear and consistently superior performance on trust trials compared to critical trials. Thus, Experiments 3a and 3b corroborate that people are, indeed, using external supports when available on the trust trials. This was also clearer in participants' reported memory strategy in Experiment 3a and 3b. When rating the chosen strategy from external (relying on list only) to internal (relying on own memory only), ratings in Experiments 3a and 3b did differ significantly as a function of external store condition such that for the told-store condition, participants reported relying on the external memory store more. That said, most people even in this condition reported choosing to rely mostly or only on their own memory. This still contrasts with results found in recall (here, over 50% whereas it was 13% from Kelly & Risko, 2022a). The difference between Experiments 3a and 3b could reflect the different samples used (i.e., in person vs. online), the different implementations of external support (i.e., written on paper vs. appearing on screen), and/or the different exclusion criteria employed (e.g., participants were excluded if they were not using the store as instructed).

Experiment 4

In the earlier experiments, we derived a prediction from the study effort hypothesis that expectations of external support could incur a cost to unaided recognition memory performance, and in particular, to recollection. Given the evidence that such a cost seems "small," we take a step back in Experiment 4 to examine the extent to which study effort can explain this observed cost. Thus, in Experiment 4 we shift our focus away from dissecting recognition into recollection/familiarity and toward more directly testing the feasibility of a study effort hypothesis in the current context of recognition memory.

Previous work has used study time allocation during self-paced study as an indicator of study effort (e.g., Ariel et al., 2015; Dunlosky & Thiede, 1998; Kelly & Risko, 2022b; Mazzoni & Cornoldi, 1993; Metcalfe & Kornell, 2005). Hence, in Experiment 4, we test the study effort hypothesis by indexing the study time allocation of participants in a self-paced adaptation of our current general method. If there is a cost to recognition memory of expecting external support, and this is driven by differences in study effort, then we should see that this cost is mediated (at least in part) by our index of study effort: study time. Alternatively, the smaller/null effect we have found in recognition might reflect individuals not modulating study efforts in response to expecting external support in the context of recognition memory and/or any modulation of study effort not influencing recognition memory performance.

A secondary interest that we had in Experiment 4 was to explore the potential influence of expecting external support on *expectations* of recognition memory performance through eliciting performance predictions as a metacognitive index (e.g., Lu et al., 2022; Park et al., 2022). We collected predictions of memory performance and tested whether the external store manipulation influenced predictions and the accuracy of those predictions. Previous work in the context of free recall has found that people predict a cost to memory performance upon losing expected support (e.g., Park et al., 2022). Examining whether this general finding would extend to the current context of recognition memory should provide additional insight into the metacognitions associated with offloading memory. That is, how does the test type (i.e., recognition vs. recall) affect people's predictions about how the loss of an expected memory store influences their memory performance?

A final interest that we had in Experiment 4 is whether the expectation of external memory support at study could lead to changes in response bias upon testing. For example, people who expected to be supported externally but were actually unaided at test could have less internal memory evidence to call upon during the recognition memory test, which could lead to response bias. If so, they could respond conservatively (being less likely to accept an item as studied), or they could respond liberally (being more likely to accept an item as studied), trying to increase hit rate (at the cost of increased false alarms). Such response strategies provide further context to the influence of losing expected support upon decision making during retrieval. The preregistration for Experiment 4 is available via the Open Science Framework at osf.io/qjgct.

Method

Data from 118 participants (58.5% female, 35.3% male; $M_{Age} = 38.82$ years, $SD_{Age} = 13.20$ years),[6] each paid GBP 7.50 as compensation for their participation, were collected and analyzed from PROLIFIC based on an a priori power analysis with a desired power of 0.80 ($\alpha = 0.05$, two-tailed) to detect a Cohen's d of ~ 0.25. The procedure generally follows that of the prior experiments except that we moved away from the "Remember/Know/New" procedures of Experiments 2, 3a, and 3b and returned to the Old/New

[6] Demographic information was unavailable for 5.1% of participants in Experiment 4. Percentages may not add up to 100% due to rounding.

response options from Experiment 1. Again, participants completed five trials in total, three trust trials wherein the external store was available (as instructed) and two critical trials on which the store was not available at test but wherein notice of this was given prior to the study phase only on one trial (the order of these critical trials was counterbalanced).

On the final three trials (i.e., the final *trust* trial *with* list available, and the two *critical* trials *without* list available), performance predictions were elicited by asking "*How accurate do you think you will be in terms of % correct on the memory test [with/without] your list available? A correct response is when you correctly respond 'old' to items presented during study and 'new' to items not presented during study.*" Participants were then given the option to select 0–100% in 5% increments. Word items were identical to those of Experiments 1, 2, and 3b, with item *lists* identical to that of Experiment 3b. As is the case throughout these experiments, lists appeared in all trial positions (1–5) and as both target and foil lists. We did not include the two memory strategy question prompts for this fifth experiment, given the generally converging findings from earlier experiments that individuals do not tend to rely heavily on external supports in the current context of recognition memory and given the potential confusion participants were having with answering the prompts (particularly in the case of Experiments 1 and 2).

Results

We had the same exclusion criteria for participants as in Experiments 3a and 3b (and we also relaxed the fourth criterion of at least chance performance on critical trials, again, results did not differ regardless of their inclusion/exclusion). Six participants were excluded for not having the minimum of 80% encoding accuracy into the external support and 15 participants for not reaching the minimum 80% hit rate on the trust trials where external support was available. Note that we preregistered outlier analyses on study time and that we foreground the results with outliers removed at the trial level (i.e., for every participant). Analyses that include these outliers were also conducted and we note where there are deviations from the main foregrounded analyses. Two participants did not provide critical trial predictions of performance: They are excluded from analyses of predictions and metacognitive accuracy.

Given that the manipulation of external support is within-participant, we followed the recommendations of Montoya & Hayes (2017) and conducted the critical mediation analysis in SPSS with their macro MEMORE and the percentile bootstrap confidence interval method (10,000 samples). We deviated from the preregistration by analyzing whether study time mediated the relation between external store condition and recognition performance rather than whether *change in* study time mediated the relation. This is because the change in study time measure for each external support condition would be calculated by subtracting the prior trust trial study time, but the external support conditions share this trust trial study time, hence, the same value would be subtracted from the study times of both the told-no-store and told-store conditions, rendering it equivalent to comparing the study time measures for these conditions (see this same note when testing the effect of external store condition on study time). We focus the mediation analyses on our main measures of performance including hit rate (i.e., the performance of items that have an associated study time), sensitivity, and percentage correct.

In addition to study time, we introduced a general recognition performance index of percentage correct given that the predictions elicited from participants were for percentage correct (to facilitate ease of interpretation). We also added an examination of response bias or decision criterion, C, to determine whether expecting support affected how liberal (i.e., more likely to say "old"; $C<0$) or conservative (i.e., more likely to say "new"; $C>0$) responses were, with the formula for C as $[-1/2 * [z(H)+z(F)]$ (Macmillan & Creelman, 1990; Stanislaw & Todorov, 1999).

The preregistration for Experiment 4 differed from that of earlier experiments by foregrounding mixed-effects regression analyses over analogous paired t-tests where applicable. Dues to an oversight, we did not preregister analyses on metacognitive predictions, so we deviate from the preregistration by adding an analysis of the effect of external store condition on predicted performance. All other analyses followed the preregistration unless otherwise specified. Table 7 presents the means across key dependent variables during trust trials and as a function of external store condition for the critical trials. Data and analyses code for Experiment 4 are available via the Open Science Framework at osf.io/y36p2.

Hit rate

There was a significant effect of condition on hit rate such that those in the told-no-store condition were more likely

Table 7 Experiment 4: Means and confidence intervals (CIs) for Hit rate, False alarm rate, and Sensitivity across trials and conditions

	Hit rate	False alarm rate	Sensitivity
Trust Trial 1	0.93 [0.92, 0.94]	0.02 [0.02, 0.03]	3.33 [3.23, 3.43]
Trust Trial 2	0.89 [0.88, 0.90]	0.03 [0.02, 0.04]	3.14 [3.03, 3.25]
Trust Trial 3	0.91 [0.90, 0.93]	0.02 [0.01, 0.03]	3.32 [3.21, 3.42]
Told-no-store	0.81 [0.79, 0.84]	0.15 [0.13, 0.18]	2.13 [1.98, 2.29]
Told-store	0.75 [0.73, 0.78]	0.16 [0.14, 0.19]	1.83 [1.69, 1.97]

CIs are bias corrected, bootstrap 95% CIs using 10,000 samples

to correctly identify a target as "old" than those in the told-store condition [$b = 0.47$, $SE = 0.08$, $z = 5.61$, $p < 0.001$; $t(119) = 5.31$, $p < 0.001$, $d = 0.49$]. Exploratory analyses revealed that trust trial hit rate was significantly higher than hit rate for either external store condition (trust: 0.91; told-no-store: 0.81; told-store: 0.75) [trust vs. told-no-store: trust vs. told-no-store: $t(117) = 7.95$, $p < 0.001$, $d = 0.73$; trust vs. told-store: $t(117) = 11.14$, $p < 0.001$, $d = 1.03$].

False alarm rate

There was no significant effect of condition on likelihood of incorrectly identifying a foil as "old" [$b = 0.22$, $SE = 0.12$, $z = 1.84$, $p = 0.066$; $t(117) = 1.57$, $p = 0.120$, $d = 0.14$]. Trust trial false alarm rate was significantly lower than false alarm rate for either external store condition (trust: 0.02; told-no-store: 0.15; told-store: 0.16) [trust vs. told-no-store: $t(117) = 9.95$, $p < 0.001$, $d = 0.92$; trust vs. told-store: $t(117) = 12.71$, $p < 0.001$, $d = 1.17$].

Sensitivity (d′)

There was a clear effect of condition on sensitivity such that those in the told-no-store condition had higher sensitivity than those in the told-store condition [$t(117) = 4.23$, $p < 0.001$, $d = 0.40$]. Trust trial sensitivity was significantly higher than sensitivity for either external store condition (trust: 3.32; told-no-store: 2.13; told-store: 1.82) [trust vs. told-no-store: $t(117) = 14.05$, $p < 0.001$, $d = 1.29$; trust vs. told -store: $t(117) = 18.83$, $p < 0.001$, $d = 1.73$].

Correct response time (ms)

There was no significant effect of external store condition on correct response time (told-no-store: 1,319; told-store: 1,269) [$b = 54.03$, $SE = 47.80$, $t = 1.15$, $p = 0.251$].

Response bias (C)

There was a significant effect on response bias such that those in the told-no-store condition were significantly less liberal in their response bias than those in the told-store condition (told-no-store: 0.09; told-store: 0.18) [$t(117) = 2.85$, $p = 0.005$, $d = 0.26$]. Exploratory one-sided *t*-tests revealed that the bias of those in both conditions (not preregistered) was significantly different from zero, indicating a clear tendency to respond "yes" to items [told-no-store: $t(117) = 2.71$, $p = 0.008$, $d = 0.25$; told-store: $t(117) = 4.94$, $p < 0.001$, $d = 0.46$].

Percentage correct

There was a significant effect of condition on percentage correct: Those in the told-no-store condition performed better than those in the told-store condition [$b = 0.04$, $SE = 0.01$, $t = 4.59$, $p < 0.001$; $t(117) = 4.53$, $p < 0.001$, $d = 0.42$].

Study time

The effect of condition on change in study time was significant such that moving from the trust trial to the told-no-store trial had a significantly larger effect on study time than did moving from the trust trial to the told-store trial (told-no-store: + 222 ms; told-store: −101 ms) [$t(117) = 3.74$, $p < 0.001$, $d = 0.34$]. While we followed the preregistered analysis of conducting this test on change in study time (from trust trial to critical trial), this is equivalent to analyzing the actual study time on the critical trials as a function of external support condition (told-no-store: 3,342 ms vs. told-store: 3,020 ms) given that computing *change* in study time means subtracting the same amount of time on the trust trial for each external store condition (i.e., the external store conditions share the baseline comparison of the final trust trial) [$t(117) = 3.74$, $p < 0.001$, $d = 0.34$].

Correlations between study time and performance indices

To gain a better sense of the relation between study effort and recognition performance, exploratory (i.e., not preregistered) correlation analyses were conducted between study time and each key performance index (hit rate, false alarm rate, and sensitivity) separately for each external support condition. In the told-no-store condition, study time was significantly correlated with hit rate [$r(116) = 0.23$, $p = 0.011$] and sensitivity [$r(116) = 0.25$, $p = 0.007$], but was not significantly correlated with false alarm rate [$r(116) = -0.09$, $p = 0.347$]. For the told-store condition, study time was significantly correlated with false alarm rate [$r(116) = -0.19$, $p = 0.035$], and sensitivity [$r(116) = 0.25$, $p = 0.006$] but not with hit rate [$r(116) = 0.10$, $p = 0.304$].

Mediation analyses

Study time mediating hit rate Given that there was an effect of external store condition on unaided hit rate, and an effect of external store condition on study time, we conducted a mediation analysis examining whether the relation between external store condition and hit rate was mediated by study time. Here, there was no significant mediation of study time [*indirect effect*: 0.006, CI_{95} (−0.001, 0.017)] and the effect of external store condition on hit rate [*total effect*: $b = 0.06$,

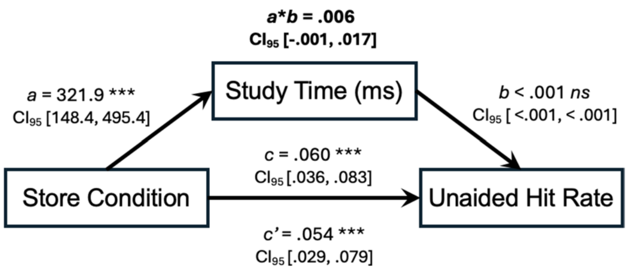

Fig. 1 Experiment 4: Study time as a mediator on the relation between external store condition and unaided hit rate. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, *ns* nonsignificant

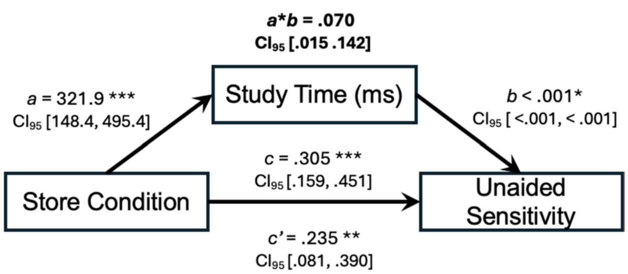

Fig. 2 Experiment 4: Study time as a mediator on the relation between external store condition and sensitivity. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, *ns* nonsignificant

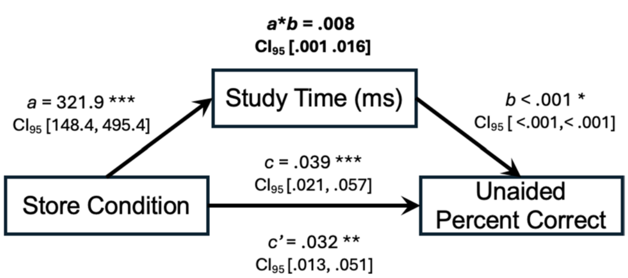

Fig. 3 Experiment 4: Study time as a mediator on the relation between external store condition and percentage correct. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, *ns* nonsignificant

$SE = 0.01$, $t = 5.04$, $p < 0.001$] remained robust when controlling for the mediator [*direct effect*: $b = 0.05$, $SE = 0.01$, $t = 4.24$, $p < 0.001$] (see Fig. 1).

Study time mediating sensitivity Given an effect of external store condition on unaided sensitivity, and an effect of external store condition on study time, we conducted a mediation analysis examining whether the relation between external store condition and sensitivity was mediated by study time. Here, we did find evidence that study time mediated the relation between external store condition and sensitivity [*indirect effect*: 0.070, CI_{95} (0.015, 0.142)] and the effect of external store condition on sensitivity [*total effect*: $b = 0.31$, CI_{95} (0.159, 0.451), $SE = 0.07$, $t = 4.13$, $p < 0.001$] remained robust when controlling for the mediator [*direct effect*: $b = 0.24$, CI_{95} (0.081, 0.390), $SE = 0.08$, $t = 3.02$, $p = 0.003$] (see Fig. 2).

Study time mediating percentage correct Finally, we conducted a mediation analysis examining whether the relation between external store condition and percentage correct was mediated by study time. Here, we also found evidence of study time as a mediator [*indirect effect*: 0.008, CI_{95} (0.001, 0.016)] and, again, the effect of external store condition on percentage correct [*total effect*: $b = 0.04$, CI_{95} (0.021, 0.057), $SE = 0.01$, $t = 4.31$, $p < 0.001$] remained robust when controlling for the mediator [*direct effect*: $b = 0.03$, CI_{95} (0.013, 0.051), $SE = 0.01$, $t = 3.27$, $p = 0.001$] (see Fig. 3).

Predicted percentage correct

There was no significant effect of condition on predictions, $t(115) = 1.08$, $p = 0.281$, $d = 0.10$ (told-no-store: 0.52; told store: 0.50; not preregistered). Exploratory analyses testing the influence of condition order (told-no-store-first vs. told-store-first) on this result revealed that for those in the told-no-store-first order condition predicted worse memory upon losing expected support compared to expecting no support, $t(57) = 4.16$, $p < 0.001$, $d = 0.55$ (told-no-store: 0.56; told store: 0.48). In contrast, those in the told-store-first order

Table 8 Experiment 4: Relation between predictions of memory performance and actual performance indices

	Index of memory performance			
	% Correct	Hit rate	False alarm rate	Sensitivity (*d'*)
Final trust trial	0.32 ***	0.18 ns	−0.34 ***	0.35 ***
Told store	0.39 ***	0.25 **	−0.28 **	0.37 ***
Told no store	0.33 ***	0.15 ns	−0.37 ***	0.32 ***
Statistical comparison between external store conditions	*ns*	*ns*	*ns*	*ns*

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, *ns* nonsignificant

condition predicted better memory upon losing expected support compared to expecting no support $t(57) = 3.58$, $p = 0.001$, $d = 0.47$ (told-no-store: 0.48; told store: 0.53).

Metacognitive accuracy Correlations of predicted and actual performance indices, with the correlations being compared statistically as a function of external store condition, are presented in Table 8. All correlations between predicted performance and actual memory performance were significant for each external support condition and index of memory performance except in the case of predicted performance and hit rate for the told-no-store condition. None of the correlations significantly differed as a function of external support condition.

Exploratory mini meta-analysis

Using the *metafor* package in R (Viechtbauer, 2010), we conducted an exploratory meta-analysis to get a better sense of the overall effect of expecting support on recognition performance across the $k = 5$ experiments. To reduce the number of analyses, we focused on the recognition performance index of sensitivity given that it incorporates performance on both targets and foils. The overall effect size was 0.28 ($SE = 0.05$, $CI_{95}[0.18, 0.38]$, $z = 5.68$, $p < 0.001$, see Fig. 4 for the forest plot of the individual study estimates and overall effect). While study estimates ranged from 0.15 to 0.40, heterogeneity across experiments was negligible ($\tau^2 < 0.001$, $I^2 < 0.01\%$, $Q(4) = 3.09$, $p = 0.543$).

Discussion

In Experiment 4, external support condition significantly affected hit rate, sensitivity, and overall percentage correct in the recognition test, such that all were significantly lower when told to expect external support (although no support was available). Thus, Experiment 4 provided clear evidence of a cost to unaided recognition memory from expecting external support. Indeed, these findings of a cost to recognition performance were further supported by an exploratory mini meta-analysis on sensitivity across the present five experiments. The mini meta-analysis provided robust evidence of a cost to recognition memory when expecting external support across the experiments which demonstrated minimal between-experiment variability.

Given this cost of external memory support expectations to recognition memory, one of the main aims in Experiment 4 was to test whether differences in study effort (i.e., study time as an index) could explain this effect. We found a clear effect of external support condition on study time such that those who were aware of the unavailable external support (told-no-store condition) increased study time significantly more than those not told of this ahead of time (told-store condition), thus, supporting the idea that expecting external support influences study time even in the context of recognition memory. In addition, the mediation analysis revealed a significant partial mediating influence of study time on the relation between expecting support and resulting memory for two of the three performance indices (i.e., sensitivity and % correct). Therefore, the cost to recognition associated with expecting external support can be partially explained by a study effort hypothesis. This is consistent with recent work suggesting that expecting support (i.e., expecting being able to offload) can lead to costs not completely explicable in terms of study effort (Kelly & Risko, 2022b).

In Experiment 4, we were also curious about how the loss of expected support might influence decision making during the recognition memory test. Thus, we examined whether expecting support affected response bias (conservative versus liberal) given that the lower memory performance of the told-store condition suggests that participants have less evidence in memory to call upon during testing. In general,

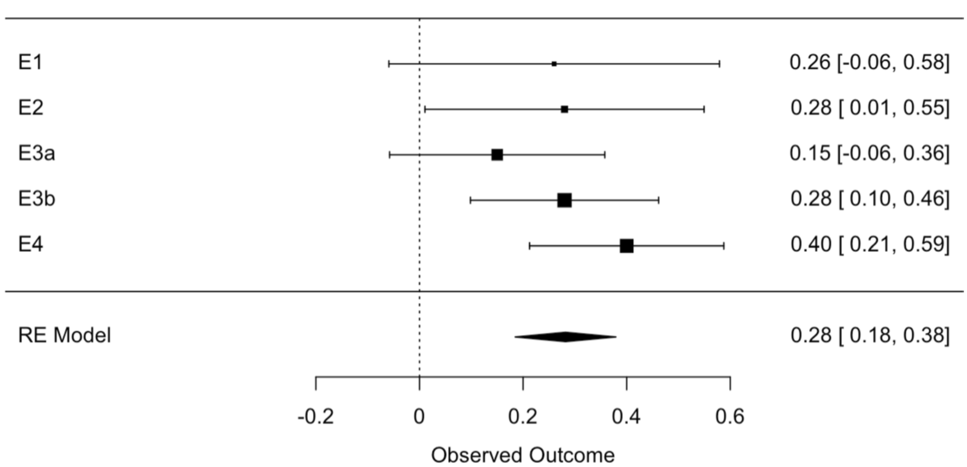

Fig. 4 Forest plot of the effect of expecting support on sensitivity across experiments

results suggest that participants tended to respond liberally (a tendency to respond "old") given that the positive c values in both external support conditions differed significantly from 0, a value which represents the absence of bias. In comparing external support conditions, those told to expect support were significantly more liberal in their responding on the recognition test, thus, the unexpected loss of external support led to a higher propensity to respond "old." The current findings are consistent with the idea that individuals try to compensate for the loss of expected support by maximizing hit rate (i.e., not missing a studied item) at the cost of increasing false alarm rate.

Our remaining interest in the current work was to explore the potential effect of expecting external memory support on metacognitive predictions of performance and metacognitive accuracy. Here, participants' performance predictions, in general, did not differ as a function of external support expectations, despite the cost observed in actual memory performance in the told-store condition. This was initially surprising and contrasted with the clear cost that individuals predict in unaided free recall when losing an expected support (Kelly & Risko*, submitted;* Lu et al., 2022; Park et al., 2022). This result initially suggested that individuals did not see the loss of an external support as necessarily leading to a reduction in performance in the case of recognition memory despite investing less effort during study when they expected the external store (as suggested by the study time data). However, upon further examination, the general lack of predicted cost appeared to be the aggregate result of two opposing effects as a function of condition order (told-*no*-store trial first vs. told-store trial first). Specifically, there was a relative *cost* to recognition memory predicted by those who experienced the told-no-store trial first (and told-store trial second), and a relative *benefit* predicted by those who experienced the told-store trial first (and told-no-store trial second). The initial critical trial yielded higher performance predictions than the second critical trial, regardless of condition order. The effect of expecting external supports on performance expectations appears more complex than predicting a general (lack of) cost. Participants were able to make predictions of performance that correlated positively with their performance. This was true regardless of external support condition (the exception being in the told-no-store condition when relating prediction to hit rate), and that there was no discernable influence of expecting support on the magnitude of these correlations. It appears from this current experiment that individuals have *some* metacognitive awareness of their general recognition memory ability in contexts that include external memory supports but that they are insensitive to the cost of losing an external support.

With respect to use of external supports, although we did not include the post-task prompts from earlier experiments in Experiment 4, we did find evidence that individuals are using the external support during the trust trials. That is, like in Experiments 3a and 3b, it was clear in Experiment 4 that performance indices were significantly better (i.e., higher hit rate and sensitivity, lower false alarm rate) when the external support was available compared to on the critical trials where it was not available.

General discussion

We have long used external memory aids to support our ability to remember. In the present work, we aimed to better understand this behavior and its potential consequences on recognition memory. Across five preregistered experiments (and an exploratory mini meta-analysis of these experiments), we found support for the idea that expecting external memory supports can negatively influence recognition memory performance. These findings support the broader idea that, while the support of external aids is helpful and often necessary, there are clear costs to unaided memory if an expected support has been lost. We also sought to understand the potential influence of external support expectation on recollection and familiarity. We found some evidence that recollection is affected by expecting external support which did not seem as clear for familiarity (Experiments 2, 3a, and 3b). Our results are consistent with the idea that expecting external supports affects effortful encoding processes that are particularly key for developing a sense of recollection for the to-be-remembered information. This is in line with a study effort hypothesis about the influence of expecting external support on the degree to which one expends top-down, intentional mnemonic efforts to commit to-be-remembered information to internal memory. Despite these findings, the present work also suggests that the general effect of expecting external memory support in recognition memory is far less of a cost than that seen in free recall.

To better understand the nature of the cost to recognition memory of expecting external support, the final experiment of the present work sought to test the feasibility of reduced study effort – indexed via study time – as an explanation for this cost. Along with clear effects of external store condition on study time, there was a positive influence of study time on the resulting memory performance indices (though not significant for hit rate) and, critically, a partial mediation of study time (though not significant for hit rate). All these patterns mirror those that have been found in recall. In addition, as mentioned earlier, these findings suggest that a portion of the cost to recognition memory performance is *not* explained easily with a study effort hypothesis and therefore align with the findings that some portion of the cost observed in free recall is also not easily explained by study effort reduction (Kelly & Risko, 2022b).

Given the clear effect of external support condition on study effort, the current results also suggest that the small effect of expecting external support on unaided recognition memory performance is not because participants are not varying study effort as a function of external store condition. Instead, the current results are consistent with the idea that participants *do* modulate study efforts in response to expecting external support (compared to expecting no such support), but that, critically, this modulation seems to show little cost to recognition memory performance, especially compared to the costs reported in previous in free recall (e.g., Kelly & Risko, 2019a, 2019b, 2022a, 2022b; Lu et al., 2020, 2022; Park et al., 2022). One potential explanation for this is that recognition testing is less sensitive to more shallowly encoded information, as it provides additional environmental support (the stimuli to endorse as “old” or “new”). Thus, study effort (needed for deeper encoding) would have less influence on memory in the context of recognition.

Some support for the idea that study effort has less influence on memory performance in the context of recognition comes from comparing the relation between study effort (study time) and performance between Experiment 4 of the current work and earlier work by Kelly and Risko (2022b). In Experiment 4 of the current work, correlations between study time and performance indices were $r = 0.23$ for hit rate, $r = -0.08$ false alarm rate (not significant) and $r = 0.25$ for sensitivity in the told-no-store condition. For the told-store condition, correlations between study time and performance indices were $r = 0.10$ for hit rate (not significant), $r = -0.19$ for false alarm rate and $r = 0.25$ for sensitivity. Although not reported, using the recall data of Kelly and Risko (2022b) posted to the Open Science Framework, the correlations between study time and performance were $r = 0.28$ (E1a) and $r = 0.38$ (E1b) for the told-no-store group and $r = 0.41$ (E1a) and $r = 0.40$ (E1b) for the told-store group. While it may not be advisable to statistically compare these given design differences, the correlations between study time and performance for the recall data appear consistently greater than those for the recognition data. Whether this is true in general may be a valuable direction for future research.

The trust trial performance (when external store is available), especially relative to critical trial performance (when external store is unavailable), and the memory strategy self-reports, together, suggest that individuals may not have been relying as much on external support in the current task as has been the case in previous research using a recall task. For trust trial performance, wherein the support was available and, therefore, virtually perfect performance was possible, actual performance was below 90% in the earlier experiments; in later experiments, it still did not exceed 95% unlike in prior work using free recall tests (e.g., Kelly & Risko, 2019a, 2019b; Kelly & Risko, 2022a, 2022b; Lu et al., 2020, 2022; Park et al., 2022). Furthermore, throughout Experiments 1–3b, the post-task prompts suggested that regardless of external support condition, individuals opted for a memory strategy more reliant on their own internal memory versus the saved list, even in the case of the told-store condition. Thus, the smaller cost to recognition memory performance relative to free recall could reflect, at least in part, less reliance on external supports for recognition memory. It remains an open question whether the cost of relying on an external memory support would remain different across recognition and recall if individuals were to rely on the stores to the same extent across tasks.

Less reliance on external memory supports during recognition could be due to a variety of reasons. For example, given that the trust trials allow participants to gain a sense of the task demands, on the critical trials participants may be content with the performance that they can achieve without relying on the external store. It might be the case that in the kind of recognition task used here, achieving “acceptable” performance might be deemed possible without relying much on the external support, whereas this might be less likely to be the case in free recall. If participants feel they can achieve “acceptable” performance without the external store, then electing not to use the external store seems rational, given that use itself could be associated with costs.

In the *Introduction*, we outlined a cost-payoff framework of investing study effort given its costs and the expected payoffs (e.g., memory benefits) of doing so. Again, when an external support containing the to-be-remembered information is expected, the same payoffs (or better) from study effort can be expected without investing that study effort. That is, retrieval success can be expected without study effort as the external support can provide this information upon retrieval. Hence, a study effort hypothesis suggests that there is less effort likely to be invested in this scenario compared to expecting no external support. Using this same cost-payoff framework, the payoff from consulting the external support (i.e., the perceived effect on performance) during recognition may not be worth the cost to consult it (i.e., the effort involved). Hence, using the store is also effortful in some way and must be balanced with the efforts required from internal memory to achieve “acceptable” levels of performance. This is consistent with a recent computational model proposed by Gilbert (2024) wherein storing information internally gives rise to an opportunity cost (given limited capacity) and wherein storing information externally gives rise to a small physical cost (though with unlimited capacity).

One final objective of this work was to explore the potential influence of expecting support on expected (predicted) recognition memory performance. We found that individuals provided predictions that generally correlated (and positively so) with their actual performance, suggesting some metacognitive awareness of their recognition performance. At first glance, individuals’ predictions did not show a general cost of expecting external support, which was unexpected as this has been repeatedly demonstrated in the context of recall. However, upon further analysis, this lack of general cost appeared to be due to some participants predicting a cost (those who experienced the told-no-store condition before the told-store condition) and some participants predicting a benefit (those who experienced the told-store condition before the told-no-store condition). The modulation of condition order (told-no-store first vs. told-store first) on these predictions suggest that expectations of performance are easily influenced by unexpected factors. These varied findings could also reflect the fact that the cost is smaller in the present recognition experiments than in the previous recall experiments and more work is needed to determine whether this reflects an area of genuine difficulty for metacognition. It might also be worth examining whether individuals predict a *benefit* to recognition of having external support.

The general costs of expecting support across both types of memory testing are clear. While we wish to avoid overspeculation and believe there to be nuanced differences between recognition and recall, this robustness seems to suggest that other forms of memory testing (e.g., cued recall, source memory) are likely to demonstrate similar findings. Exploring how current (and prior) results extend to other forms of remembering could provide support for more generalized insights.

Conclusion

The present work sought to extend our understanding of the effects of expecting external supports to a recognition memory test. Across five preregistered experiments, results evidenced a cost to recognition memory but one not nearly as pronounced as that found in the context of free recall. The current work also supports the idea that study effort reduction can explain, *in part*, the cost of expecting support on unaided memory and thus, critically, a portion of the cost is *not* easily explainable due to differences in study effort.

Acknowledgements We would like to thank Dahwi Ahn for her guidance on within-subjects mediation analysis.

Authors' contributions M.O.K.: Conceptualization, methodology, investigation, data curation, formal analysis, writing – original draft, writing – review and editing, Visualization; B.K.: Investigation (supporting), formal analysis (supporting); A.P.: Conceptualization (supporting); X.L.: Conceptualization (supporting); E.F.R.: Conceptualization (supporting), resources, writing – review and editing, supervision, funding acquisition.

Funding This work was supported by Alexander Graham Bell Canada Graduate Scholarships to MOK, AEP, and XL from the Natural Sciences and Engineering Research Council of Canada (NSERC), and by a Discovery Grant (#04091) from NSERC, by an Early Researcher Award from the Province of Ontario (#ER14-10–258), by funding from the Canada Foundation for Innovation and Ontario Research Fund (#37872), and by funding from the Canada Research Chairs (#950–232147) program to EFR.

Data Availability The data and materials of all the experiments of the current work are available via the Open Science Framework at osf.io/y36p2.

Code availability The analyses code for all experiments of the current work is available via the Open Science Framework at osf.io/y36p2.

Declarations

Conflicts of interest/Competing interests The authors declare no conflicts of interest or competing interests.

Ethics approval The conducting of these experiments was reviewed and approved by a University of Waterloo Research Review Committee.

Consent to participate Informed consent was obtained from all participants prior to their participation.

Consent for publication Post-debrief consent to use data was obtained from all participants included in this work.

Open Practices Statements As stated throughout, the data, materials, and analyses code for all experiments are available via the Open Science Framework at osf.io/y36p2. The preregistrations for the experiments are available at the following: E1: [osf.io/e2bxm]; E2: [osf.io/ap2tr]; E3a: [osf.io/5p8m4]; E3b: [osf.io/kx835]; E4: [osf.io/qjgct].

References

Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review, 79*(2), 97.

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.

Basden, B. H., Basden, D. R., & Gargano, G. J. (1993). Directed forgetting in implicit and explicit memory tests: A comparison of methods. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(3), 603.

Basden, B. H., & Basden, D. R. (1996). Directed forgetting: Further comparisons of the item and list methods. *Memory, 4*(6), 633–654.

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

Bjork, R. A. (1972). Theoretical implication of directed forgetting. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory* (pp. 217–235). Winston.

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.

Cho, K. W., Kazanas, S. A., & Altarriba, J. (2018). Survival processing in recognition memory: Separating recollection from familiarity. *The American Journal of Psychology, 131*(1), 19–32.

Cofer, C. N. (1967). Does conceptual organization influence the amount retained in immediate free recall? In B. Kleinmuntz (Ed.), Concepts and the structure of memory. New York: Wiley.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.

Craik, F. I. (2023). Reducing age-related memory deficits: The roles of environmental support and self-initiated processing activities. *Experimental Aging Research, 48*(5), 401–427.
Craik, F. I. (1994). Memory changes in normal aging. *Current Directions in Psychological Science, 3*(5), 155–158.
Danckert, S. L., & Craik, F. I. (2013). Does aging affect recall more than recognition memory? *Psychology and Aging, 28*(4), 902.
de Leeuw, J. R., Gilbert, R. A., & Luchterhandt, B. (2023). jsPsych: Enabling an open-source collaborative ecosystem of behavioral experiments. *Journal of Open Source Software, 8*(85), 5351.
Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition, 24*, 523–533.
Dornbush, R. L., & Winnick, W. A. (1967). Short-term intentional and incidental learning. *Journal of Experimental Psychology: General*, 73.
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.
Dunn, J. C. (2004). Remember-know: A matter of confidence. *Psychological Review, 111*(2), 524.
Eagle, M., & Leiter, E. (1964). Recall and recognition in intentional and incidental learning. *Journal of Experimental Psychology, 68*(1), 58.
Eskritt, M., & Ma, S. (2014). Intentional forgetting: Note-taking as a naturalistic example. *Memory & Cognition., 42*, 237–246.
Fellner, M. C., Waldhauser, G. T., & Axmacher, N. (2020). Tracking selective rehearsal and active inhibition of memory traces in directed forgetting. *Current Biology, 30*(13), 2638–2644.
Gilbert, S. J. (2024). Cognitive offloading is value-based decision making: Modelling cognitive effort and the expected value of memory. *Cognition, 247*, 105783
Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30*(5), 513–541.
Jacoby, L. L., Yonelinas, A. P., & Jennings, J. M. (1997). The relation between conscious and unconscious (automatic) influences: A declaration of independence. In J. D. Cohen & J. W. Schooler (Eds.), *Scientific approaches to consciousness* (pp. 13–47). Erlbaum.
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. (2022a). Revisiting the influence of offloading memory on free recall. *Memory & Cognition*, 1–12.
Kelly, M. O., & Risko, E. F. (2022b). Study effort and the cost of external store availability. *Cognition, 288*, 105228.
Kelly, M. O., & Risko, E. F. (2024). Examining levels of external memory support on memory performance, performance predictions, and study effort. *Subtmitted/Under Review.*
Kintsch, W. (1966). Recognition learning as a function of the retention interval and changes in the retention interval. *Journal of Mathematical Psychology, 3*, 413–433.
Kool, W., McGuire, J. T., Rosen, Z. B., & Botvinick, M. M. (2010). Decision making and the avoidance of cognitive demand. *Journal of Experimental Psychology: General, 139*(4), 665.
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.
Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4*, 863.
Lawrence, M. A. (2016). _ez: Easy Analysis and Visualization of Factorial Experiments_. R package version 4.4–0, https://CRAN.R-project.org/package=ez
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. (2022). The gist of it: Ofoading memory does not reduce the benefit of list categorisation. *Memory, 30*(4), 1–16.
Macleod, C. M. (1999). The item and list methods of directed forgetting: Test differences and the role of demand characteristics. *Psychonomic Bulletin & Review, 6*(1), 123–129.
MacLeod, C. M. (1975). Long-term recognition and recall following directed forgetting. *Journal of Experimental Psychology: Human Learning and Memory, 1*, 271–279.
Macmillan, N. A., & Creelman, C. D. (1990). Response bias: Characteristics of detection theory, threshold theory, and" nonparametric" indexes. *Psychological Bulletin, 107*(3), 401.
Mangels, J. A., Picton, T. W., & Craik, F. I. M. (2001). Attention and successful episodic encoding: An event-related potential study. *Cognitive Brain Research, 11*, 77–95.
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.
Montoya, A. K., & Hayes, A. F. (2017). Two-condition within-participant statistical mediation analysis: A path-analytic framework. *Psychological Methods, 22*(1), 6.
Morrison, A. B., & Richmond, L. L. (2020). Offloading items from memory: Individual differences in cognitive offloading in a short-term memory task. *Cognitive Research: Principles and Implications, 5*, 1–13.
Murphy, D. H. (2023a). Strategic offloading: How the value of to-be-remembered information influences offloading decision-making. *Applied Cognitive Psychology, 37*(4), 749–767.
Murphy, D. H. (2023b). Does value structure influence measures of memory selectivity? *Memory, 31*, 1074–1088.
Nestojko, J. F., Finley, J. R., & Roediger, H. L., III. (2013). Extending cognition to external agents. *Psychological Inquiry, 24*(4), 321–325.
Ochsner, K. N. (2000). Are affective events richly recollected or simply familiar? The experience and process of recognizing feelings past. *Journal of Experimental Psychology. General, 129*, 242–261.
Olson, G. M. (1969). Learning and retention in a continuous recognition task. *Journal of Experimental Psychology, 81*, 381–384.
Ozubko, J. D., Gopie, N., & MacLeod, C. M. (2012). Production benefits both recollection and familiarity. *Memory & Cognition, 40*, 326–338.
Park, J., Kelly, M. O., Hargis, M., & Risko, E. F. (2022). Influence of external store reliance on predicted and actual value-directed remembering. *Psychonomic Bulletin & Review, 29*(4), 1367–1376.
Popov, V., & Dames, H. (2023). Intent matters: Resolving the intentional versus incidental learning paradox in episodic long-term memory. *Journal of Experimental Psychology: General, 152*(1), 268.
Postman, L., Adams, P. A., & Phillips, L. W. (1955). Studies in incidental learning: II. The effects of association value and method of testing. *Journal of Experimental Psychology, 49*, 1–10.
Rajaram, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory & Cognition, 21*, 89–102.
R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.Rproject.org/
Richmond, L. L., Kearley, J., Schwartz, S. T., & Hargis, M. B. (2023). Take a load off: Examining partial and complete cognitive offloading of medication information. *Cognitive Research: Principles and Implications, 8*(1), 12.

Risko, E. F., & Dunn, T. L. (2015). Storing information in-the-world: Metacognition and cognitive offloading in a short-term memory task. *Consciousness and Cognition, 36*, 61–74.

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

Risko, E. F., & Kelly, M. O. (2024). Cognitive Offloading and Memory: A Review. *Learning and Memory: A Comprehensive Reference*, *3rd Edition* (Eds. J. Wixted & L. Mickes). Elsevier. *In press.*

Risko, E. F., Kelly, M. O., Lu, X., & Pereira, A. (2024). 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. *In press.*

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

Rodd, J. M. (2024). Moving experimental psychology online: How to obtain high quality data when we can't see our participants. *Journal of Memory and Language, 134*, 104472.

Rundus, D. (1971). Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology, 89*(1), 63–77.

Sahakyan, L., Waldum, E. R., Benjamin, A. S., & Bickett, S. P. (2009). Where is the forgetting with list-method directed forgetting in recognition? *Memory & Cognition, 37*, 464–476.

Sahakyan, L., Delaney, P. F., Foster, N. L., & Abushanab, B. (2013). List-method directed forgetting in cognitive and clinical research: A theoretical and methodological review. *Psychology of Learning and Motivation, 59*, 131–189.

Schwartz, F., & Rouse, R. D. (1961). The activation and recovery of associations. *Psychological Issues, 3.*

Sheard, E. D. & MacLeod, C. M. (2005). Listmethoddirected forgetting: Return of the selective rehearsal account. In N. Ohta, C. Macleod, & B. Uttl (Eds.), Dynamic cognitive processes (pp. 219–248). Tokyo: Springer-Verlag.

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

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior, 6*, 156–163.

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

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31*(1), 137–149.

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.

Tan, P., Ensor, T. M., Hockley, W. E., Harrison, G. W., & Wilson, D. E. (2020). In support of selective rehearsal: Double-item presentation in item-method directed forgetting. *Psychonomic Bulletin & Review, 27*, 529–535.

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.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1–48.

Woodward, A. E., Jr., Bjork, R. A., & Jongeward, R. H., Jr. (1973). Recall and recognition as a function of primary rehearsal. *Journal of Verbal Learning and Verbal Behavior, 12*(6), 608–617.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46*(3), 441–517.

Yonelinas, A. P., & Jacoby, L. L. (1995). The relation between remembering and knowing as bases for recognition: Effects of size congruency. *Journal of Memory and Language, 34*, 622–643. https://doi.org/10.1006/jmla.1995.1028

Yonelinas, A. P., Aly, M., Wang, W. C., & Koen, J. D. (2010). Recollection and familiarity: Examining controversial assumptions and new directions. *Hippocampus, 20*(11), 1178–1194.