On the effects of external support on memory and metamemory

by

Megan Olivia Kelly

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Doctor of Philosophy

in

Psychology

Waterloo, Ontario, Canada, 2024

© Megan Olivia Kelly 2024

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner	Sam Gilbert, PhD Professor Institute of Cognitive Neuroscience University College London
Supervisor	Evan F. Risko, PhD Professor Department of Psychology
Internal Members	Colin M. MacLeod, PhD Distinguished Professor Emeritus Department of Psychology
	Clara Colombatto, PhD Assistant Professor Department of Psychology
Internal-external Member	Oliver Schneider, PhD Associate Professor Department of Management Science and Engineering

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

Experiments 1a and 1b (Chapter 2) have been published in *Cognition* (Kelly & Risko, 2022b). Experiments 2 and 3 (Chapter 3) are currently under review (Kelly & Risko, *submitted*). Experiments 4, 5, 6a, 6b, and 7 (Chapter 4) are currently under review (Kelly et al., *submitted*).

Contributor Roles Taxonomy Author Statements (see Brant et al., 2015)

Chapter 2: Experiments 1a and 1b
M.O.K.: Conceptualization, Methodology, Investigation, Data Curation, Formal Analysis, Writing – Original Draft, Writing – Review & Editing, Visualization
E.F.R.: Conceptualization (supporting), Resources, Writing – Review & Editing, Supervision, Funding acquisition

Chapter 3: Experiments 2 and 3
M.O.K.: Conceptualization, Methodology, Investigation, Data Curation, Formal Analysis, Writing – Original Draft, Writing – Review & Editing, Visualization
E.F.R.: Conceptualization (supporting), Resources, Writing – Review & Editing, Supervision, Funding acquisition

Chapter 4: Experiments 4, 5, 6a, 6b, and 7
M.O.K.: Conceptualization, Methodology, Investigation, Data Curation, Formal Analysis, Writing – Original Draft, Writing – Review & Editing, Visualization
B.K.: Investigation, Formal Analysis (supporting)
A.P.: Conceptualization (supporting)
X.L.: Conceptualization (supporting)
E.F.R.: Conceptualization (supporting), Resources, Writing – Review & Editing, Supervision, Funding acquisition

Funding

The work presented in this dissertation was supported by an Alexander Graham Bell Canada Graduate Scholarship from the Natural Sciences and Engineering Research Council of Canada (NSERC) to M.O.K., as well as by an NSERC Discovery Grant (#04091), an Early Researcher Award from the Province of Ontario (#ER14-10- 258), funding from the Canada Foundation for Innovation and Ontario Research Fund (#37872) from the Canada Research Chairs (#950- 232147) program to E.F.R.

Abstract

Our ability to remember has long been aided by various supports in our environment. Although these supports are indeed helpful when available, compared to having expected no support, there is a memory cost of expecting external memory support. This dissertation presents a series of experiments that systematically examined this cost of expecting external memory support across a variety of conditions and contexts. Central to the dissertation is the proposed explanation for the memory cost: that it occurs because expecting external support leads to reduced efforts at committing the to-be-remembered information into internal memory—or study effort. Results do suggest that the costs to memory of expecting external support can be explained, at least in part, by this study effort hypothesis. But results also evidence memory costs of expecting external support that are not easily explained by reduced study effort. In exploring the influences of expecting external support on metamemory, results suggest that people have a general awareness of their performance ability with and without expected support. However, results also suggest areas of difficulty with respect to metacognitive awareness of performance, implying that metacognitive accuracy could be challenging when using external memory supports. In examining the memory cost of expecting external support across a variety of conditions and contexts, this dissertation provides further insights regarding our use of external memory supports and places the study effort hypothesis on solid footing as a partial explanation for the cost.

Acknowledgments

Although much of this dissertation was written far more recently than I intended, its completion represents several years marked by countless doubts, challenges, and even a few insights—a journey that would be impossible without the support of many generous individuals.

First, I am extremely grateful to my exceptional specimen of a supervisor, Dr. Evan F. Risko, for his incredible expertise and literally award-winning mentoring. I will miss dearly the impromptu coffee chats and the special "high praise" of seeming *reasonable* (or even the possibly lesser *not un*reasonable). He can take comfort in knowing that he has assuredly ruined my standards for a "prompt" email response. I cannot overstate how appreciative I am for him taking a chance on me those many years ago (despite my initial accidental addressing him as Dr. *Frisko*).

Thank you to my internal committee members, Dr. Colin MacLeod and Dr. Clara Colombatto. The knowledge and support they have both offered, including beyond this dissertation, have been extremely generous and impactful. I also thank my external committee members, Dr. Sam Gilbert, and Dr. Oliver Schneider, for their great care and valuable insights.

To the members of the *CaNB* Lab and my other friends in the program (some of you have since left), the support and friendships we have cultivated make leaving the program even more difficult. We made sure to remind each other that we were $p < .001$.

To my parents, my brother, and the rest of my loved ones, your endless love and encouragement have been so fundamental to this feat. You continually root for me even if it sometimes means that I must be farther away. Thank you for keeping me grounded yet allowing me to soar. I can finally tell you all that I am done school (I think).

To my collaborator in life, Jon: None of this would have been possible without you. Your "peer review" has made me a better person, and you never get tired of my "revisions." Thank you for everything that you do. I look forward to a lifetime of trial, error, and spontaneous findings together.

Table of Contents

List of Figures ..... *ix*

List of Tables ..... *x*

Chapter 1 Introduction *1*

Chapter 2 ..... *12*

Experiments 1a & 1b 17

Method 17

Results 21

General Discussion 33

Chapter 3 ..... *38*

Experiment 2 45

Method 45

Results 49

Discussion 59

Experiment 3 62

Method 63

Results 63

Discussion 75

General Discussion 78

Chapter 4 ..... *82*

Experiment 4 90

Method 90

Results 93

Experiment 5 99

Method 101
Results 102
Discussion 106
Experiments 6a & 6b 108
Method 108
Results 110
Discussion 117
Experiment 7 120
Method 122
Results 123
Discussion 132
General Discussion 134
Chapter 5 Concluding Remarks *141*
General Conclusions 150
References *152*

List of Figures

Figure 1. General procedure implemented in the experiments of this dissertation. 5

Figure 2. Experiments 1a and 1b: Mean proportion of study items recalled by external store condition, trial type, and experiment. 25

Figure 3. Experiments 1a and 1b: Mean study time per study item by external store condition, trial type, and experiment. 27

Figure 4. Experiments 1a and 1b: Change in study time as a mediator on the relation between external store condition and Trial 4 recall performance. 30

Figure 5. Experiments 1a and 1b: Trial 4 study time as a mediator on the relation between external store condition and Trial 4 recall performance. 31

Figure 6. Experiments 1a and 1b: Strategy use as a mediator on the relation between external store condition and Trial 4 recall performance. 32

Figure 7. Experiments 2 and 3: The general procedure including where our dependent variables of interest were collected. 49

Figures 8A and 8B. Experiment 2: Means and confidence intervals for mean recall proportion (A) and mean predicted recall proportion (B) as a function of support condition and trial type. 52

Figures 9A and 9B. Experiment 3: Means and confidence intervals for mean recall proportion (A) and mean predicted recall proportion (B) as a function of support conditions and trial type. 65

Figure 10. Experiment 7: Study time as a mediator on the relation between external store condition and unaided hit rate. 128

Figure 11. Experiment 7: Study time as a mediator on the relation between external store condition and sensitivity. 129

Figure 12. Experiment 7: Study time as a mediator on the relation between external store condition and percentage correct. 130

List of Tables

Table 1. Experiments 1a and 1b: Percentage of self-reported strategy (or no strategy) by external store condition and experiment. 29

Table 2. Experiment 2: Means and confidence intervals for each study effort index across support conditions on each trial. 53

Table 3. Experiment 2: Correlations between predicted and actual memory performance by Trial Type (Aided vs. Unaided) and Support Condition (Full vs. Partial-3 vs. Partial-2 vs. No-support). 58

Table 4. Experiment 3: Means and confidence intervals for each dependent variable across support conditions on each trial. 66

Table 5. Experiments 2 and 3: Correlations between predicted and actual memory performance by Trial Type (Aided vs. Unaided) and Support Condition (Full vs. Partial-2 vs. No-support) in Experiment 3 and also collapsing across Experiments 2 and 3. 71

Table 6. Experiment 4: Means and Confidence intervals for Hit rate, False alarm rate, and Sensitivity across trials and conditions. 94

Table 7. Experiment 4: 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). 96

Table 8. Experiment 5: Means and Confidence intervals for Hit rate, False alarm rate, and Sensitivity across trials and conditions. 103

Table 9. Experiment 5: 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). 106

Table 10. Experiments 6a and 6b: Means and Confidence intervals for Hit rate, False alarm rate, and Sensitivity across trials and conditions. 112

Table 11. Experiments 6a and 6b: 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). 117

Table 12. Experiment 7: Means and Confidence intervals for Hit rate, False alarm rate, and Sensitivity across trials and conditions. 125

Table 13. Experiment 7: Relation between predictions of memory performance and actual performance indices. 131

Chapter 1

Introduction

We are exposed to far more information than we can process, let alone remember. Although our intentional efforts to remember often yield success, there of course are times when our internal memory systems can and do betray us, often through forgetting. Indeed, we can readily think of times wherein we suffered consequences of various sorts because of forgetting an important date or deadline. To combat this, we can look to memory supports in our environment (i.e., external to our internal biological memory systems). People have long done so, with oft-cited examples including the *yupana* and *quipu* of the Inca people thought to aid in the counting and recording of various data and possibly narratives (Eeckhout, 2012; Urton, 2008). Thus, in addition to engaging in efforts at *internal* remembering (e.g., making mental notes, imagining visual images, or rehearsing information), another way that we may invoke to remember is by enlisting the support of *external* memory aids. That is, we can save information *externally* to available supports in our environment (e.g., planners, calendars, to-do-lists, journals, and even other trusted people; Finley et al., 2018)—a form of *offloading memory* (Gilbert, 2024; Risko & Gilbert, 2016; Risko & Kelly, 2024; Risko et al., 2024).

Integrating external supports is far from a novel approach to remembering, but it is only recently that we have begun concerted empirical efforts toward better understanding remembering in this extended context. Memory research has traditionally focused solely on the triumphs and failures of our *internal* memory systems. This dissertation extends the unit of analysis beyond the internal context to include the use of external memory supports (Clark, 2010; Clark & Chalmers, 1998; Hutchins, 2000). The use of such external supports—*offloading memory*—can take a variety

of forms. For example, we may aim to use the external store more as "replacement" (*substitutive* use) or we may still aim to remember internally and hence aim to use the external support more as "backup" (*duplicative* use; Risko et al., 2024).

In general, enlisting the support of external memory supports is highly effective, often yielding performance superior to that of our internal memory alone (Ball et al., 2022; Boldt & Gilbert, 2019; Fellers et al., 2023; Gilbert, 2015; Gilbert et al., 2023; Grinschgl et al., 2021; Kelly & Risko, 2019a; 2019b; 2022a; 2022b; *submitted*; Kelly et al., *submitted*; Lu et al., 2020; 2022; Morrison & Richmond, 2020; Murphy, 2023; Park et al., 2022; Peper et al., 2023; Richmond et al., 2023; Risko & Dunn, 2015; Scarampi & Gilbert, 2021; see also Runge et al., 2019; 2021; Storm & Stone, 2015; Tsai et al., 2023 for similar benefits of offloading memory). However, how might our ability to consult external memory supports to aid memory matter for or interact with internal memory processes? Through the research in this dissertation (and some research *not* included), I have begun to explore this and related questions.

As mentioned, the enlisting of external memory support often yields superior performance compared to remembering internally alone (i.e., unaided), however, much of the current work traces back to an influential finding by Sparrow et al. (2011) that there can be a *cost* to expecting external memory support. Namely, the unexpected *loss* of such expected external support leads to significantly *poorer* memory performance than had one expected no support and instead relied solely on internal memory. It appears that offloading memory leads to clear costs to unaided memory performance compared to *not* offloading memory. Note that in this dissertation, I refer to this cost using a few different labels (e.g., *the cost of external memory store reliance, the cost of*

external memory store availability, *the cost of expecting external memory support*), however, I intend for them to all refer to the same phenomenon—that when presumably able to expect external memory support, people tend to remember less internally (i.e., when unaided) compared to people told not to expect such support (i.e., also unaided).

In my master's thesis, I developed a novel paradigm to examine this cost of expecting external memory support (Kelly & Risko, 2019a; 2019a; 2022a). In doing so, I found evidence that memory phenomena thought to be *more* dependent on intentional or top-down mnemonic study efforts (e.g., the primacy effect; Fischler et al., 1970; Tan & Ward, 2008) were more likely to show this memory cost than memory phenomena thought to be *less* dependent on such intentional efforts (the recency effect, the von Restorff or isolation effect; Dunlosky et al., 2000; Fabiani & Donchin, 1995; Glanzer & Cunitz, 1966; Kelly & Risko, 2019a; 2019b; 2022a; Köhler & von Restorff, 1995; MacLeod, 2020; Shriffin & Atkinson, 1969; von Restorff, 1933; see also Lu et al., 2022 and Park et al., 2022). Here, top-down mnemonic study efforts are meant to refer to *intentional* efforts that are aimed at supporting internal remembering but are not meant to refer to any *unintentional* or *incidental* processes that might also occur and support internal remembering.

Taken together, results led to a study effort explanation for the cost of expecting external memory support. Specifically, when one expects to be able to use an external memory support at retrieval, this reduces the likelihood of investing intentional efforts toward remembering internally. This study effort hypothesis draws from two key tenets from research on effort mobilization: (i) that effort is costly to expend (Kool et al., 2010; see also Inzlicht et al., 2018), and (ii) that effort

is only expended when the costs of doing so are outweighed by the expected payoffs (Shenhav et al., 2021). In the present context of offloading memory, the expected availability of external memory supports containing to-be-remembered information allows for payoffs (i.e., retrieval success) *without* much required cost in terms of study efforts. A central goal of this dissertation is to examine the extent to which this study effort hypothesis explains the cost of external memory store reliance.

The Current Dissertation

Throughout this dissertation, the memory tasks are designed to test unaided memory performance across two main conditions: a condition in which participants are told that they will have their external memory stores available (*told-store* condition) and a condition in which participants are told that they will *not* have external stores available (*told-no-store* condition). The general procedure contains two main types of memory trials, *trust* trials and *critical* trials. The memory tasks started with three trust trials, in each of which participants were asked to study and store a list of items (study phase) for a recall memory test (testing phase) that would be *aided* with the externally stored items—that is, on these trust trials, *aided* recall is tested. These trials were to enhance trust in the external store in the way that we trust and consult our own calendars and memory aids as needed.

These trust trials were followed by a critical trial (two critical trials for a within-participant design; Chapter 4; Kelly et al., *submitted*) that differed from the trust trials in that the recall test was *not* aided with the externally stored items—hence, *unaided* recall was tested. On the critical trial, one group of participants was given the same instructions as in the trust trials prior to studying

and storing such that they could presumably expect an aided recall test as on the trust trials. The other group was told prior to studying and storing that they would *not* have their external store available at test, unlike in the previous trust trials. Critically, neither group actually had access to their external store. Thus, as mentioned, this paradigm is designed to compare unaided memory performance between a group expecting external store support (told-store) and a group knowing *not* to expect the external store support (told-no-store). Figure 1 demonstrates the general procedure.

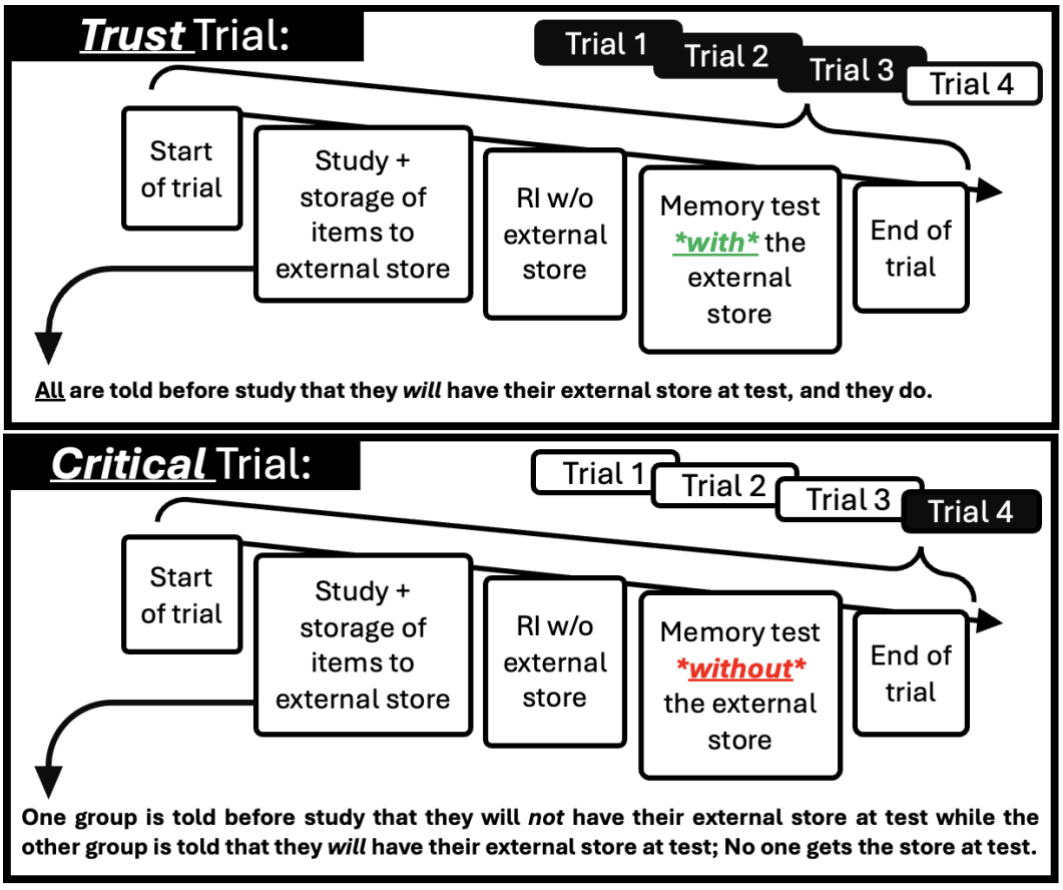


Figure 1. General procedure implemented in the experiments of this dissertation.

Note: RI stands for retention interval.

The following chapters focused on examining the feasibility of the study effort hypothesis as a theoretical account for why people demonstrate poorer unaided memory performance when expecting external memory support.

A direct test of study effort. The dissertation begins with developing a direct investigation of the study effort hypothesis as an explanation of the cost of expecting external memory support (Chapter 2; Kelly & Risko, 2022b). Across two preregistered experiments, I presented a direct test of whether differences (i.e., reductions) in study effort could explain variation in the cost of expecting support on resulting free recall performance. In these experiments, participants controlled their own study time (Ariel et al., 2015; Dunlosky & Thiede, 1998) and reported whether they used any intentional strategies to support remembering (e.g., repeated reading of items, imagery, sentence generation, grouping items in some meaningful way; based on Dunlosky & Kane, 2007; Unsworth, Miller, & Robison, 2019; Unsworth, 2016). Hence, indices of study effort were study time allocation and strategy use (Beaudoin & Desrichard, 2017; Froger et al., 2012; Mazzoni & Cornoldi, 1993; Metcalfe & Kornell, 2005; Son & Metcalfe, 2000). Here, I found evidence that study effort *partially* mediated the cost of expecting external memory support. This will be a recurring finding throughout the dissertation.

Varying the degree of external support. In Chapter 2, the focus was on comparing the effects of expecting *full support* versus *no support* on memory without the support (unaided). For full support, if "apple" is encoded at study, "apple" is expected as support at test; for *no* support, "apple" is not available externally at test. In Chapter 3 (Kelly & Risko, *submitted*), I extended the examination to investigating the effects of varied levels of support by including a *partial* level of

external support. Here, in a *partial-support* condition, if "apple" is encoded at study, "AP" or "APP" was provided at test depending on the partial support condition. To better understand people's expectations in response to varying levels of external support, I elicited predictions of memory performance. This work examined how individual's expectations about external supports influenced memory, perceptions of study effort, and predictions of performance.

Across the two preregistered experiments, results indicated that partial levels of external support (wherein only a portion of each to-be-remembered item is saved) can lead to better recall compared to internal memory alone and, critically, that the unexpected loss of these partial external supports led to memory costs like those demonstrated when expecting full support (as in the case of earlier work and Chapter 2), though smaller than when expecting full support. Unexpectedly, participants provided with partial support reported levels of study effort no different from those expecting no support, thus, study effort could not explain the memory cost of expecting *partial* support. Predictions of memory performance suggested that people were aware of *a* cost of losing an expected external support, although their predictions were not sensitive to the difference between losing a partially versus a fully supportive store. That said, in general, predictions and actual performance demonstrated positive correlations regardless of external support condition. The results of Chapter 3 extended the work of Chapter 2 by adding insights regarding the effects of expecting *varying* levels of external support on memory, expected performance, and study effort. In addition, the results of Chapter 3 support the overall conclusions of Chapter 2—that, although there is evidence for study effort reduction as an explanation for the cost of expecting external support, it cannot fully explain this cost. That is, I found a cost of expecting external

support (i.e., partial support) that was not associated with any reported reductions in effort at encoding.

Investigating recognition memory. The earlier chapters focused exclusively on testing memory in a recall context. In Chapter 4 (Kelly et al., *submitted*), I examined whether offloading memory would also lead to a cost in recognition memory. One key difference between recognition and recall is that the former provides the to-be-remembered items for endorsement (among foils) upon testing whereas the latter requires participants to generate the to-be-remembered items on their own. Borrowing from Craik's (1983; 1990; 1994; 2022) conclusions that environmental support can reduce age-related effects on memory, having the additional environmental support at test in a recognition test could be generally beneficial to memory performance, leading to a smaller observed cost to memory when expecting the external support available in recognition compared to recall.

The work of Chapter 4 also extends the work in earlier chapters by examining the influence of expecting support on recollection versus familiarity (Jacoby, 1991; Yonelinas, 2002) given that the study effort hypothesis makes a clear prediction that recollection processes ought to be affected. Less clear from the study effort hypothesis is whether familiarity processes ought to be affected. One idea is that familiarity is largely a result of unintentional, automatic memory processes (Jacoby, 1991). From this perspective, the study effort hypothesis would predict little to no cost of expecting external support on familiarity. The final experiment of Chapter 4 tests whether study effort can explain any cost to recognition memory, doing so by testing whether study effort indices mediate the relation between expecting support and recognition performance.

Across five preregistered experiments, there was evidence of a cost to recognition memory of expecting support. The cost to observed recognition, however, was smaller and more difficult to reliably detect than the memory cost reported in recall. Along with the previous chapters, this work continues to support the idea that the (small) cost to recognition memory is partially explainable by differential study efforts when expecting external memory support. Thus, study effort appears to be a factor in the resulting cost to unaided recognition memory, although not fully, thus echoing the findings of Chapters 1 and 2. Chapter 4 also reports evidence that recollection was affected negatively by external memory support, as predicted by the study effort hypothesis.

Finally, there were positive correlations between predicted performance and actual performance regardless of whether external support was expected, and predictions of recognition performance did not demonstrate a cost of expecting external memory support. Thus, individuals demonstrated some metacognitive awareness of their performance in a recognition test (regardless of whether they expect support). Furthermore, individuals also seem aware that the cost of expecting support on recognition is, indeed, smaller than on recall, when considering the results of Chapters 3 and 4 together.

As described, the study effort hypothesis is evidenced throughout this dissertation as a partial explanation for the cost of expecting external memory support. This begs the question of what other underlying mechanisms could be contributing to the cost of expecting external memory support. Although the focus of this dissertation is on the feasibility of the study effort hypothesis specifically, I touch on alternative, non-mutually exclusive mechanisms throughout and in the

General Discussion (Chapter 5). Briefly, these include potential feelings of surprise or alarm from learning of the unexpected loss of external support which may disrupt mnemonic processes and, therefore, impair retrieval performance. In addition, the unexpected loss of external memory support could encourage a reduction to retrieval effort at test, therefore leading to costs in memory performance. I also present a kind of transfer-appropriate-processing explanation particularly in the case the cost of losing a *partial* support in Experiment 3 (Chapter 3).

In summary, the nine experiments of this dissertation provide a systematic examination of the cost to memory of expecting external memory support through providing varied tests of a proposed explanation of that cost—the study effort hypothesis. Taken together, the results uphold the idea that a study effort explanation can explain, *in part,* the cost to memory when expecting external support. Additionally, the current experiments clearly indicate that interactions between internal and external memory sources are complex and, as such, deserving of further scrutiny. The importance of doing so is highlighted by the omnipresence in the modern world of external memory supports and the rapid proliferation of their sophistication and availability (e.g., the internet, large language models).

Note that Chapters 2-4 comprise scholarly works published or submitted to peer-reviewed journals so these chapters will contain subtle stylistic differences depending on the reporting format preferred by the relevant peer-reviewed academic outlet. Furthermore, much of the content across chapters, particularly in the case of the discussion sections (e.g., *introduction, general discussion)*, will contain similar information. The projects contained in this dissertation were

preregistered and have accompanying data, analyses code, and other materials openly available on the Open Science Framework as indicated throughout.

Chapter 2

The following work has been published in Cognition (Kelly & Risko, 2022b).

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

Much work has begun to explore the relation between internal memory and external supports, the various ways individuals use external memory supports to support internal memory, and the effects of doing so (e.g., Bulley et al., 2020; Eskritt & Ma, 2014; Finley et al., 2018; Gilbert et al., 2022; Grinschgl et al., 2021; Henkel. 2014; Kelly & Risko, 2019a, 2019b; Lu et al., 2020, 2021; Marsh & Rajaram, 2019; Meyerhoff et al., 2021; Park et al., 2022; Pereira et al., 2021; Risko et al., 2019; Risko et al., 2024; Sachdeva & Gilbert, 2020; Scarampi & Gilbert, 2020; Sparrow, Liu, & Wegner, 2011; Storm & Stone, 2015). Of recent interest is how storing information externally influences the internal memory for that information (Eskritt & Ma, 2014; Grinschgl et al., 2021; Kelly & Risko, 2019a, 2019b; Lu et al., 2020, 2021; Park et al., 2022; Sparrow, Liu, & Wegner, 2011). One consistent finding to emerge from this research is that when individuals can rely on an external memory store, they demonstrate significantly poorer subsequent unaided recall

of the stored information compared to if no external memory store was presumed to be available (Eskritt & Ma, 2014; Kelly & Risko, 2019a, 2019b; Lu et al., 2020; Sparrow, Liu, & Wegner, 2011). One explanation for this relative cost is that when individuals expect access to an external memory store, they forgo top-down, controlled efforts at study. For example, in cost/benefit models of effort allocation (e.g., the Expected Value of Control theory; Shenhav et al., 2021), these kinds of control demanding strategies (e.g., rehearsal, narrative generation) would be viewed as inherently costly and deployed only when the payoff (e.g., in increased accuracy) exceeded that cost (Shenhav et al., 2021). With an external store capable of ensuring (presumably) perfect accuracy available, the expected value of control (i.e., the payoff due to engaging strategies aimed at improving memory accuracy) would be correspondingly low. Thus, the anticipation of having access to an external store effectively acts as a cue to *not* invest such mnemonic effort. We refer to this explanation as the *study effort hypothesis*. In the present investigation, we test this account directly by examining both study time allocation and self-reported strategy use as a function of the availability of an external memory store.

The cost of external store reliance

A variety of different memory paradigms have been used to investigate the influence of relying on an external memory device. Sparrow et al. (2011) examined the influence of saving information on its subsequent recall during a test wherein the saved information was not available. Individuals who were told that their saved information had been erased before test demonstrated significantly better recall than individuals who were given the impression their information was saved. Similarly, Eskritt and Ma (2014) found that when individuals were given the opportunity

to use study notes to assist memory for an array of images, they demonstrated poorer unaided memory for image location than those who were unable to make study notes.

Recent work has focused on explaining the cost of external store availability using the study effort hypothesis. Kelly and Risko (2019a; 2019b) and Lu et al. (2020) found that the instructed availability of an external store led to *reduced* recall for information thought to be *more* reliant on top-down, controlled memory efforts at study and *no* reduced recall for information thought to be *less* reliant on such effort. Specifically, while memory for salient information (Kelly & Risko, 2019b), and the general theme of a list (e.g., when categorized; Lu et al., 2020) appears preserved when individuals have an external memory store available to rely on, recall is often more substantially reduced for beginning-of-list than end-of-list information (Kelly & Risko 2019a; 2019b) and for more item-specific verbatim information (Lu et al., 2020). That said, these studies did not directly attempt to index top-down, controlled efforts made at study. The present work does so through the examination of study time allocation and self-reported study strategy use.

Study time allocation and strategy use

One means of indexing study effort in the context of a memory task is to measure study time allocation, that is, the amount of time individuals spend "studying" a to-be-remembered stimulus when that time is controlled by the individual (i.e., when the task is self-paced; e.g., Mazzoni & Cornoldi, 1993; Metcalfe & Kornell, 2005; Son & Metcalfe, 2000). A number of studies have used study time allocation in this manner. For example, Ariel and colleagues (2015) examined whether the influence of information value on memory (i.e., memory for high-value versus low-value information) reflected differences in study time allocation as a function of value.

They found that high value items received significantly more study and were remembered significantly better than lower value items. Similarly, Dunlosky and Thiede (1998) manipulated the probability that a specific study item would be tested later, finding that items very likely to be tested (i.e., 90% chance) were studied longer and remembered better than items less likely to be tested (i.e., 50% or 10% chance). Thus, under conditions wherein participants might modulate their effort and individuals are given control over study time, study time varies in relation to those putative modulations in effort.

While study time can provide a window into the allocation of study effort, it provides little in the way of insight into what individuals may be doing (or not doing) during that time. For example, increased study time might be associated with the use of time intensive (Beaudoin & Desrichard, 2017; Froger et al., 2012), top-down, controlled memory strategies or mnemonics aimed at increasing later memory performance (e.g., repeated reading of items, imagery, sentence generation, grouping items in some meaningful way; based on Dunlosky & Kane, 2007; Unsworth et al., 2019; Unsworth, 2016). Moreover, study time alone will not distinguish between study strategies if those study strategies do not differ in the time spent engaging in them (Dunlosky & Kane, 2007). Previous findings suggest that engaging in effective controlled memory strategies (e.g., imagery or sentence generation vs. repetition or reading) can effectively increase the probability of recall success (e.g., Bailey et al., 2008; Bower, 1972; Dunlosky & Kane, 2007; Richardson, 1998; Rundus, 1971; Unsworth et al., 2019). For example, Unsworth et al. (2019) found evidence that individual differences in self-reported study strategies predicted recall performance (see also Dunlosky & Kane, 2007). Just as individuals can differ in their controlled

study strategies, aspects of the study situation, such as the current manipulation of external memory store availability, could influence the use of study strategies. Thus, it might be useful to complement investigations of variation in study time with the concurrent examination of strategy use. Asking participants to report on whether they used a study strategy and/or to provide a description of what strategy they employed (if they did so) could provide additional insight into what individuals were doing at study across conditions. Thus, we examine both study time allocation and self-reported strategy use here as a function of whether individuals are told during study that they will be able to use an external memory store at test.

The current investigation

In the present work, we report two preregistered experiments (Experiment 1a and Experiment 1b; the second as a replication of the first) wherein the basic procedure of each follows that of previous work (Kelly & Risko, 2019a), such that participants complete a series of trials wherein they store to-be-remembered information on which they are subsequently tested in a free recall test. On the first three trials, participants are instructed that their memory test will be supported by their externally stored list of study items, and this is, indeed the case. Critically, however, on the fourth trial, they are not supported by their external store at test. Half of the participants are given notice of this (instructed-no-store), and half of the participants are not (instructed-store). Unlike previous studies, participants were given control of their study time, advancing study items as they saw fit. In addition, participants provided post-task reports of whether they used a memory strategy at study. The study effort account predicts that participants who anticipate relying on the external memory store at test (instructed-store condition) should

allocate less time *and* be less likely to report the use of a mnemonic strategy than those who are instructed that they cannot rely on the external memory store at test (instructed-no-store condition). Furthermore, study time allocation and strategy use should each mediate the effect of instruction on recall performance (i.e., explain the cost of relying on an external store). We index the former both in terms of a difference score across Trial 3 (i.e., the last trial wherein participants expect access to the external store) and Trial 4 (i.e., wherein half the participants expected access to the external store and the other half did not) and absolute study time allocation on Trial 4. The former allows us to control for differences in study time allocation across participants (i.e., Trial 3 is identical for all participants).

Experiments 1a & 1b

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

Method

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

Participants

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

Stimuli

Stimuli were four 20-item word lists (available at osf.io/8s75w) derived randomly from a stimulus set of 80 items. Word lengths ranged from four to 10 letters and word frequencies ranged from 26 to 82060 using *FreqCount* from *SUBTLEX-UK* (Van Heuven et al., 2014) and from two

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

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

Procedure

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

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

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

Test phase. In the test phase, participants were instructed to type the items that were presented in the study phase into a large text field onscreen with their saved list as a resource (counterbalanced across participants to appear on the right or left side of the screen). An important feature of our procedure is that on the first three trials, participants had access to their saved list during the test phases. This was to help participants establish a sense of trust in the external store akin to how they might trust their own external stores (e.g., their phones). Critically, participants do not get their external memory support during the test phase of Trial 4. Half of the participants were told of this after they completed Trial 3 (instructed-no-store condition); the other half of participants were not given notice that their external store would not be available during recall (instructed-store condition).

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

In Experiment 1a, Question 1 asked, "In performing tasks like the ones we asked you to do in the current study, individuals sometimes report engaging in strategies they believe might help them remember the presented words. On the final trial, did you engage in any such strategies during the study phase (i.e., when you were typing the to-be-remembered words)? Please elaborate on your strategy or why you did not choose to use one (it is OK if you did not)." Question 1 in Experiment 1b had the same preamble but contained slightly different wording at the end: "In

performing tasks like the ones we asked you to do in the current study, individuals sometimes report engaging in strategies they believe might help them remember the presented words. On the final trial, did you engage in any such strategies during the study phase (i.e., when typing the to-be-remembered word list)? It is completely OK if you did not. Please respond only based on the final trial." If participants responded "Yes" to this question, they were led to a follow-up free response question before advancing: "You responded that you did engage in such memory strategies on the final trial. Please briefly state what you did. Please only comment with respect to the final trial only." If participants responded to the initial question with "No", they advanced straight to Question 2.

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

Results

Data from 19 participants in Experiment 1a and 22 participants in Experiment 1b were excluded and replaced because they did not follow instructions and/or did not meet the preregistered inclusion criteria. In Experiment 1a, two participants requested that their data not be used, three participants reported doing something else during their completion of the study, one participant did not encode at least 80% of items during study across Trials 2–4, and 13 participants

did not recall at least 80% of items during test across trust trials (Trials 1–3). In Experiment 1b, two participants reported doing something else during their completion of the study, four participants did not encode at least 80% of items during study across Trials 2–4, and 16 participants did not recall at least 80% of items during test across trust trials. In the final samples (E1a: $N = 90$; E1b: $N = 110$), the mean reported age was 35.9 years-old in Experiment 1a (information unavailable for two participants) and 38.7 in Experiment 1b (information unavailable for one participant), with 53 participants identifying as female and 36 as male in Experiment 1a (information unavailable for one participant), and with 65 identifying as female and 44 as male in Experiment 1b.

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

For the free-response strategy question (applicable to all participants in Experiment 1a but only applicable to participants responding that they used a strategy in Experiment 1b), two naïve

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

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

variables and were conducted using the *lme4* package in R (Bates et al., 2015). For mixed linear models (e.g., as used for study time data), the package *lmerTest* (Kuznetsova et al., 2017) was used to extract *p* values. While their utility is debated, we have included them here for ease of reading and consistency.

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

The effect of external store availability

Recall. In both experiments, mixed effects logistic regression with external store condition as a predictor of final trial (Trial 4) recall performance (with by-participant and by-item intercepts as random effects) found that those in the told-store condition were significantly less likely to accurately recall items (**E1a:** told-store: .24; told-no-store: .47, $b = –1.31$, $SE = 0.25$, $z = –5.34$, $p < 001$; **E1b:** told-store: .26; told-no-store: .42, $b = –1.04$, $SE = 0.26$, $z = –4.00$, $p < 001$). An analogous analysis using Welch's independent samples t-test found qualitatively similar results in (**E1a:** $t(88.00) = 5.23$, $p < .001$, $d = 1.10$; **E1b:** $t(107.77) = 3.63$ $p < .001$, $d = 0.69$). Figure 2

presents the mean proportion of study items recalled as a function of external store condition for both trial types (trust and critical) in both Experiments 1a and 1b.

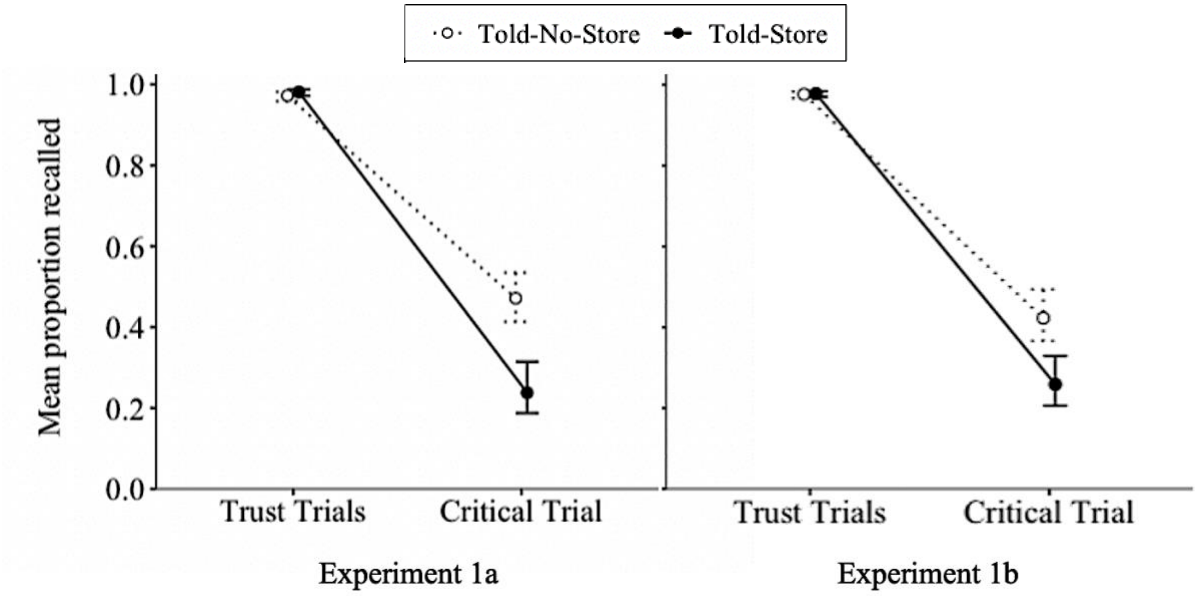


Figure 2. Experiments 1a and 1b: Mean proportion of study items recalled by external store condition, trial type, and experiment.

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

Study time. Mixed effects linear regression with external store condition (told-store vs. told-no-store) and trial (Trial 3 vs. Trial 4) as fixed effects on study time was conducted.[3] This analysis revealed a significant interaction between external store condition and trial such that the difference in study time between Trial 3 and Trial 4 was larger for the told-no-store condition

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

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

Using mixed effects linear regression with by-participant and by-item random intercepts, we examined the fixed effect of external store condition on study time on Trial 4 (these analyses were not preregistered). Those in the told-store condition did not allocate significantly less study time (**E1a:** told-store: 3738 ms, told-no-store: 4317 ms, b = -0.21, SE = 0.15, t = -1.36, p = .178

[told-store: 3968 ms, told-no-store: 4675 ms, b = -0.19, SE = 0.12, t = -1.59, p = .115]; **E1b:** told-store: 3616 ms, told-no-store: 4115 ms, b = -0.21, SE = 0.12, t = -1.79, p = .076 [told-store: 4000 ms, told-no-store: 4632 ms, b = -0.14, SE = 0.08, t = 1.71, p = .091]), although the results are in the predicted direction. Figure 3 presents the mean study time per item as a function of external store condition for both trial types (trust and critical) in both Experiments 1a and 1b.

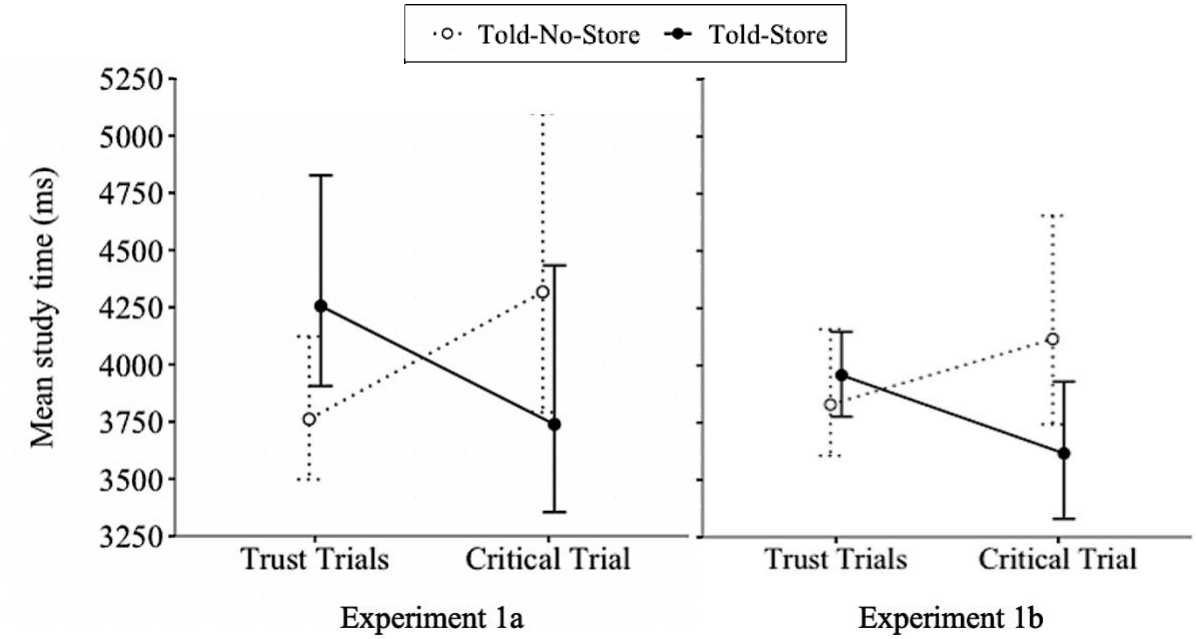


Figure 3. Experiments 1a and 1b: Mean study time per study item by external store condition, trial type, and experiment.

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

Self-reported strategy—yes/no. A chi-square test found that those in the told-store condition were significantly less likely to report using a memory strategy on Trial 4 (**E1a:** told-store: .40, told-no-store: .82, $\chi^2(1) = 15.15$, $p < .001$; **E1b**: told-store: .27, told-no-store: .53, $\chi^2(1)$

= 6.40, p = .011). See Table 1 for a percentage breakdown of reported strategy type by external store condition and experiment.

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

[4] In Experiment 1a, all participants were presented with a free response field to respond with whether they used a strategy or not during final trial (Trial 4) recall, and to describe their strategy—thus all responses in Experiment 1a were qualitative and coded. We streamlined this procedure in Experiment 1b by having participants first respond with either *yes* or *no* options to using a strategy during final trial recall and only if participant responded yes were they given opportunity to describe their strategy. Thus, only data from participants responding *yes* to using a strategy were qualitative and coded in Experiment 1b.

Table 1. Experiments 1a and 1b: Percentage of self-reported strategy (or no strategy) by external store condition and experiment.

	Experiment 1a		**Experiment 1b**	
	Told-no-store	*Told-store*	*Told-no-store*	*Told-store*
1: Read/Type each word	6.7%	4.4%	3.6%	0%
2: Repeated word often or as much as possible	4.4%	6.7%	0%	1.8%
3: Sentence/story generation, including to link multiple words	28.9%	6.7%	18.2%	16.4%
4: Mental imagery/visuals of words	4.4%	0%	5.5%	0%
5: Group/cluster words in some way	20.0%	8.9%	16.4%	3.6%
6: Other (e.g., unlisted strategy and/or multiple strategies)	17.8%	13.3%	9.0%	5.5%
Total proportion reporting (clear) strategy	82.2%	40%	52.7%	27.3%
0: No (clear) strategy reported	17.8%	60.0%	47.3%	72.7%

Mediation analyses

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

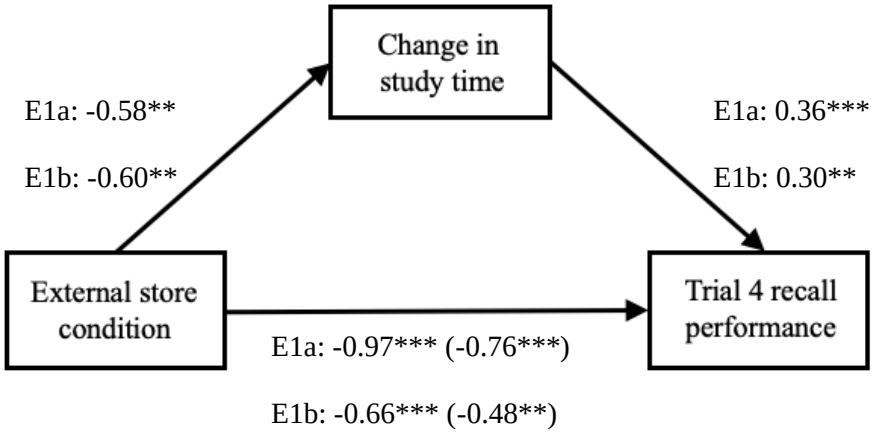


Figure 4. Experiments 1a and 1b: Change in study time as a mediator on the relation between external store condition and Trial 4 recall performance.

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

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

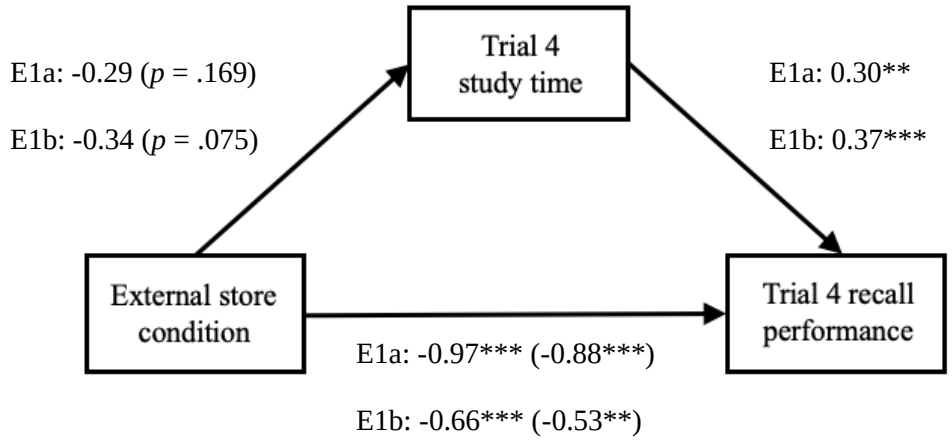


Figure 5. Experiments 1a and 1b: Trial 4 study time as a mediator on the relation between external store condition and Trial 4 recall performance.

Note. Standardized regression coefficients for the relation between external store condition and Trial 4 recall performance as mediated by Trial 4 study time for each experiment. The standardized

regression coefficient between external store condition and Trial 4 recall performance controlling for Trial 4 study time is in parentheses (outlier data of study time not included).

* $p < .05$, ** $p < .01$, *** $p < .001$.

As depicted in Figure 5, in testing study time as a mediator through the effect of external store condition on Trial 4 recall, the indirect effect was not significant despite being in the predicted direction (**E1a:** -0.09, $p = .210$ [-0.13, $p = .110$]; **E1b:** -0.12, $p = .056$, [-0.12, $p = .080$]). Provided the total effect (**E1a:** -0.97; **E1b:** -0.66) reported earlier, the direct effect was significant (**E1a:** -0.88, $p < .001$ [-0.84, $p < .001$]; **E1b**: -0.53, $p = .006$ [-0.55, $p = .002$]).

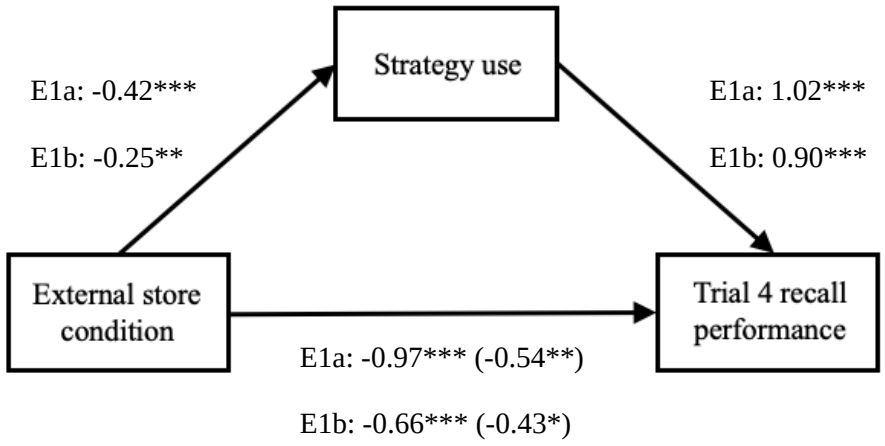


Figure 6. Experiments 1a and 1b: Strategy use as a mediator on the relation between external store condition and Trial 4 recall performance.

Note. Standardized regression coefficients for the relation between external store condition and Trial 4 recall performance as mediated by self-reported strategy use for each experiment. The standardized regression coefficient between external store condition and Trial 4 recall performance controlling for strategy use is in parentheses (outlier data of study time not included).

* $p < .05$, ** $p < .01$, *** $p < .001$.

Finally, as depicted in Figure 6, the effect of strategy use mediated the relation between external store condition and Trial 4 recall performance. The indirect effect was significant (**E1a:** -

0.43, $p < .001$; **E1b:** -0.23, $p = .006$). The direct effect was also significant (**E1a:** -0.54, $p = .006$; **E1b:** -0.43, $p = .010$). The proportion of the effect of external store condition on recall mediated by strategy was .45 in Experiment 1a and .35 in Experiment 1b.

Exploratory

The following analyses were not preregistered.

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

General Discussion

Memory for to-be-remembered information is often worse if, at study, we anticipate the support of an external memory store. The present work provided a direct test of the study effort hypothesis of this cost (Kelly & Risko 2019a; 2019b; Lu et al., 2020; Park et al. 2021) by

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

While the mediation analyses across both experiments revealed that study time allocation and strategy mediated the effect of external store condition on free recall performance, this effect was clearly only partial. Indeed, a sizable portion of the effect appears to *not* be due to these mediators alone. This implies two non-mutually exclusive possibilities: the current indices did not fully capture effort devoted to encoding and/or the cost of external memory store reliance is caused

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

Beyond limitations in measuring effort at study, the partial mediations might also indicate that the cost of external store reliance cannot be explained fully by the withdrawal of such effort.

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

Finally, we would be remiss if we did not address the possibility that the scenario that we constructed in the current empirical work using list memory does not translate perfectly to how individuals use external memory stores in their day-to-day lives. This limitation of the current work could pose some difficulty in interpreting the current findings in the context of our day to day lives. That said, we believe the paradigm used captures key features about how stores are used in the "real world". For example, many of us regularly trust and rely on various external stores to

complete our memory goals. Our trust trials are intended to capture this aspect of regularly using external devices. Many of us also know the feeling of being unable to retrieve some key information kept on those stores in the face of misplacing them. Thus, although we are constructing a lost-phone scenario in the laboratory, we believe the findings from implementing this paradigm could provide deeper understanding of the mechanisms involved (or not) when relying on external memory supports.

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

Conclusion

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

Chapter 3

The following work is under review (Kelly & Risko, submitted).

Remembering occurs under various levels of *environmental support*—or the degree to which external stimuli and context at retrieval share features of those at encoding. To the extent that environmental support is not available, people must rely on "self-initiated processing" (Craik, 1994, p. 156). For example, retrieval in a free recall task is thought to involve less environmental support and, hence, more self-initiated processing than in a recognition test (Craik, 1983; 1990; 1994; 2022; Craik & Bialystok, 2006; Lindenberger & Mayer, 2014; Luo & Craik, 2008). Craik and colleagues use the notion of environmental support to help explain and mitigate deficits in memory associated with healthy aging, but more generally environmental support also plays a central role in navigating the memory demands of our day-to-day lives. In the present investigation, we examine how people's expectations about environmental support influence their memorial behavior.

A common form of environmental support can be found in the use of artifacts as external memory stores (e.g., notepad, computer file) to support remembering—or *offloading* (Risko & Gilbert, 2016; Risko et al., 2024; Risko & Kelly, 2024). For example, if one wants to remember a list of grocery items to be purchased, they may record the needed items externally in their phone and review the list once at the grocery store. In this context, researchers have recently begun to examine how the expectation of such environmental support can influence the monitoring and control of memory (e.g., Eskritt & Ma, 2014; Gilbert et al., 2022; Kelly & Risko 2019a; 2019b;

2022a; Lu et al., 2020; 2022; Park et al., 2022; Pereira et al., 2022; Risko et al., 2019; Risko et al., 2024; Risko & Kelly, 2024; Sparrow et al., 2011; Storm et al., 2022; Storm & Stone, 2015).

Exploring this empirically has typically involved tasking participants with studying to-be-remembered items for a future memory test and telling one group that they will be given access to a memory aid (e.g., a list of the to-be-remembered items) at test and leaving another group to (expect to) rely on their unaided internal memory. The store has typically contained a verbatim list of the to-be-remembered items. Thus, in the aided condition, participants can expect to be, in a sense, *completely* supported at retrieval (i.e., with a *full* list of the to-be-remembered items). In contrast, those in the unaided condition expect *no* such support (i.e., they can rely only on their own internal, self-initiated cues). The differences observed across these conditions provide some insight into how individuals modulate their memorial behavior based on the environmental support that they can expect to aid future retrieval, and the consequences of such modulations.

The influence of external memory supports on internal memory performance

There are two consistent findings in the literature on the influence of external memory stores on memory performance. First, as expected, when individuals have access to an external store with a list of the to-be-remembered verbatim items, recall performance is typically nearly perfect (Kelly & Risko 2019a; 2019b; 2022a; Lu et al., 2020; 2021; Park et al., 2022; Risko & Dunn, 2015). Second, and more interestingly, there is a reliable cost to *unaided* memory performance when individuals expect to have this kind of support at test compared to when there is no such expectation (Eskritt & Ma, 2014; Kelly & Risko, 2019a; 2019b; 2022a; Lu et al., 2020; 2022; Park et al., 2022; Sparrow et al., 2011). According to the study effort hypothesis (e.g., Kelly

& Risko, 2022b), this cost arises because individuals are less likely to invest study effort when they expect to be able to "remember" the information by relying on information externally available at retrieval (i.e., their list of words). That is, individuals consider the contribution of the expected environmental support in their anticipated retrieval success and this consideration guides their effort at study.

This kind of behavior can be understood in terms of models of study time allocation (e.g., Ackerman, 2014; Dunlosky & Thiede, 1998; Metcalfe & Kornell, 2005; Undorf & Ackerman, 2017) wherein individuals invest study time until monitoring processes indicate that a given level of performance will be achieved. From this perspective, if expecting a high level of environmental support, a high level of performance can also be expected without much in the way of study effort. In a similar vein, according to (a modest extension of) the Expected Value of Control model of effort investment (e.g., Shenhav et al., 2013; Shenhav et al., 2021), deploying control processes during study (e.g., intentional mnemonic strategies) should be considered costly and only done when the expected payoff (e.g., high memory performance) exceeds those costs.

When *complete* environmental support is expected to be available, this should *lower* the value associated with expending effort to store that information internally because the same or similar payoff (high memory performance) can be achieved without effort. Consistent with these ideas, individuals tend to allocate less study time and are less likely to use intentional mnemonic strategies (e.g., repetition, imagery) when instructed that they will be fully supported at test (i.e., with the to-be-remembered items) compared to individuals instructed that there will be no such memory support (Kelly & Risko, 2022b).

A comparison of two extremes

Although research on the use of external memory stores suggests that the anticipation of environmental support can lead to the modulation of study effort and consequences for memory performance, much of this research has focused on what can be considered the extreme ends of the environmental support continuum (although see Richmond et al., 2023 for a relevant exception). That is, in one condition the *full*, exact, version of the to-be-remembered information is provided whereas in the other condition *no* support is provided. While some work has used choice-offloading paradigms or limited capacity stores (e.g., Burnett & Richmond, 2023; Burnett & Richmond, 2024; Richmond & Morrison, 2020; Risko & Dunn, 2015; Murphy, 2023), individuals *can* opt to store study items in full or at least fully *intact* at the item-level.

In contrast to this prior work, we extend this research to contexts wherein the environmental support is incomplete—that is, it is only *partially* representative of the original to-be-remembered information at the item-level. There are at least two important motivations for an extension to scenarios of partial storage. First, incomplete or partial environmental support is akin to many of the support we may find ourselves dealing with day-to-day. For example, we may implement creative shorthand or highlight a sentence in a book to use later, but upon later review, the meaning of the shorthand or the purpose of the highlighted sentence can escape us. Second, moving from having no environmental support to having complete environmental support can be viewed as crossing a kind of categorical boundary (Hsee & Zhang, 2010) and, as such, might translate into an easier metacognitive task. That is, the monitoring and control of study behavior when one can expect either *full* or *no* environmental support might be more straightforward than

modulating behavior in anticipation of more intermediate, *partial* levels of support (Hsee & Zhang, 2010). Whether individuals can modulate study effort beyond these "all" or "none" modes remains largely unexplored from the perspective of metacognitive control.

Present investigation

Our goal is to test the influence of anticipated environmental support on study effort, metacognitive judgments, and memory performance. We do so in both the presence and absence of the expected environmental support. Here, environmental support takes the form of a list of items available at test. We manipulated the degree of support between-participants by varying the degree to which the list items at test match the original study items. We included conditions involving full support (i.e., the list of study items themselves) and no support (i.e., no list), as used in previous work, as well as novel *partial* support conditions wherein participants received the first portion of each study item ("word stems" that were either two or three letters long, *Partial-2* and *Partial-3* support conditions, respectively).

In the current procedure, participants completed multiple trials each containing a study phase and a test phase. During study phases, participants were presented with to-be-remembered items that they typed out; in the test phases, they were instructed to recall as many of the study phase words as possible. We randomly assigned participants to support condition. In the full-support condition, the items that were made available at test were the ones they were to recall (e.g., if "apple" was encoded at study, "apple" was provided at test). As described above, in the partial-support conditions, the items that were available at test were word stems of the ones they were to recall (e.g., if "apple" was encoded at study, "ap" or "app" was provided at test depending on

whether participants were in the *Partial-2* or *Partial-3* support condition). In the no-support condition, no stored information was available at test to aid recall. Participants first completed a Practice Trial wherein they were told that they would get their support at test (except for those in the no-support condition) and this was, indeed, the case (i.e., the test on the practice trial was aided). After the practice trial, there were the two critical trials: the *Aided Trial* and the *Unaided Trial*.

Aided Trial

As on the Practice Trial, participants were told that they would get support at test (except if they were in the no-support condition) on the Aided Trial and they did; thus, the memory test on this trial was *aided*. Prior to the memory test, participants made a global prediction about their likely memory performance after completing the study phase. We also measured study effort with three indices: allocated study time, a global effort rating, and a self-report of strategy use.

Based on the amount of environmental support at test, we expected that those in the full support condition would have higher recall than those in the partial and no-support conditions, and that those in the partial support conditions would have higher recall than those in the no-support condition. We were less clear on how the support conditions would differ across the metacognitive predictions and study effort. To the extent that participants (i) are metacognitively aware of how degree of support influences future remembering and (ii) use this information to control study effort, we would expect similar patterns in their global memory predictions (i.e., Full > Partial > No-support) and in study effort indices: allocated study time (i.e., No-support > Partial > Full), global effort rating (i.e., No-support > Partial > Full), and reported use of a strategy (i.e., No-

support > Partial > Full). That is, if participants believe that retrieval will be more difficult as environmental support decreases, then efforts at study should increase. In principle, if study effort were allocated sufficiently to counteract differences in retrieval difficulty, then we might even expect equivalent levels of memory performance as a function of environmental support.

Unaided Trial

The second trial type of interest is the *Unaided Trial*. Here, participants were still told that they would get their support at test (again, except if they were in the no-support condition) but on the unaided trial this was *not* the case. Thus, participants engaged in study under the expectation of receiving the same type of support that they had received up to that point in the study and that, instead, *no support* was provided at test regardless of support condition. Provided that we intended to violate the expectations of participants on the Unaided Trial, the Unaided Trial always occurred after the first two trials (which were aided). Critically, we collected global predictions about *unaided* memory success by asking for predictions *after* individuals expecting support found out that there would be no such support (after completing the study phase, prior to testing). As noted earlier, previous work has found that when individuals expect *full* support to be available at retrieval, but are *not* provided *any* support, their recall performance suffers relative to when there is no such expectation (Kelly & Risko 2019a; 2019b; 2022a; 2022b; Lu et al. 2020; 2022; Park et al., 2022; Sparrow et al., 2011).

To the extent that participants modulate their study effort as a function of expected support (i.e., external support condition) and, critically, to the extent that their study effort is effective (e.g., not in vain), we would expect poorer unaided memory performance in the full support than the no

support condition when tested without external support, as found previously. More novel, we would expect memory performance in the partial support conditions to fall between the full and no support conditions and that predictions of memory performance would follow this qualitative pattern. For the above outlined predictions to be correct, participants' metacognitions must be accurate and translated into control (i.e., study effort), and this control must be effective. That said, inferring the influences of *partial* support on memory performance might be difficult for individuals (i.e., 3/5 vs. 2/5 letters available). Critically, understanding where any such breakdowns occur will provide novel insights into potential limitations in the monitoring and control of memory.

Experiment 2

Method

The preregistration as well as the data and analyses code for Experiment 2 are available respectively at osf.io/fpdsn and osf.io/9qxmc.[5]

Participants

Participants were 180 (102 females, 72 males, 1 prefer not to say, 5 unavailable; M_{age} = 37.70 years, SD_{age} = 12.48 years) individuals from PROLIFIC divided equally over the four groups, and were paid £3.75 for their participation (~20 minutes). This sample size was based on powering for a balanced one-way ANOVA with four between-participants groups, considering a

[5] Note that prior to Experiment 2 of the current work, two preregistered studies were conducted but ultimately excluded due to flaws in experimental design (e.g., ineffective data quality checks) as we were learning on the fly about out how to transition to online data collection due to the onset of COVID-19.

power of .80 to detect a medium effect size of $f = .25$ (two-tailed, alpha = .05) using ANOVA_exact (Lakens & Caldwell, 2021). Participants were replaced if one or more of the following preregistered criteria applied: (1) They indicated that they did not want their data to be included; (2) they indicated that they were not paying attention; (3) they indicated that they used an external store other than that provided by the experimental program; (4) they did not type the full study item out 75% of the time across Trials 2 and 3 (this is a liberal exclusion criterion considering that they are asked to type out all items); and (5) they demonstrated lower than 75% recall on a post-task aided trial (wherein participants were given the full store at test regardless of support condition, thus recall should have been near perfect).

Stimuli

To construct three lists of 20 words, 60 words were drawn randomly from the stem completion stimuli set of Jacoby (1998; outdated or potentially inappropriate terms were replaced). Words were 5 letters in length and 11 to 32670 in frequency (via FREQcount; Brysbaert & New, 2009).[6] Each list appeared in each trial position roughly equivalently across participants. Given the constraints of the item set, there were a total of four instances wherein an item within the word set shared a 2-letter word stem with another item (ar---, tr---, mo---, sw---), but these were split across lists so that item stems within lists were always unique. All 3-letter word stems were unique within and between lists.

[6] Note that four of the study items (i.e., *eagle*, *forge*, *frost*, *tally*) did not appear in the data base, thus, frequencies were not available.

Procedure

Once consent and instruction procedures were completed, the first trial began (this trial was treated as practice). Participants completed the practice trial, then the Aided trial, and then the Unaided Trial. Within each trial, there were two main phases: The study phase and the test/recall phase.

Study. At the beginning of each study/encoding phase, participants were presented with to-be-remembered items one at a time. Participants typed these items fully into a textbox in all conditions (i.e., if the study word was “answer”, all participants were to type “answer” in the text box). After all twenty study items were presented, participants were asked the global effort and strategy questions in a randomized fashion. The global effort question asked, “On a scale from 1-9, how much mental effort did you invest in remembering the items presented during the study phase?” and participants rated the response on a scale from “1 (very, very low mental effort)” to “9 (very, very high mental effort)” (Van Kessel et al., 2024). The strategy question said “In performing tasks like the ones we asked you to do in the current study, individuals sometimes report engaging in strategies they believe might help them remember the presented words. Did you engage in any such strategies during the study phase (i.e., when typing the study words) to help you remember the words? It is completely OK if you did not.” and participants responded with either “Yes”, “No”, or “I prefer not to answer.” If a participant responded with “Yes,” then that participant was presented with a prompt that said “You responded 'yes' to engaging in a memory strategy. Please briefly state what you did.” Following responses to these questions, a brief retention interval of 15 s occurred before the start of the test/recall phase. On the third trial,

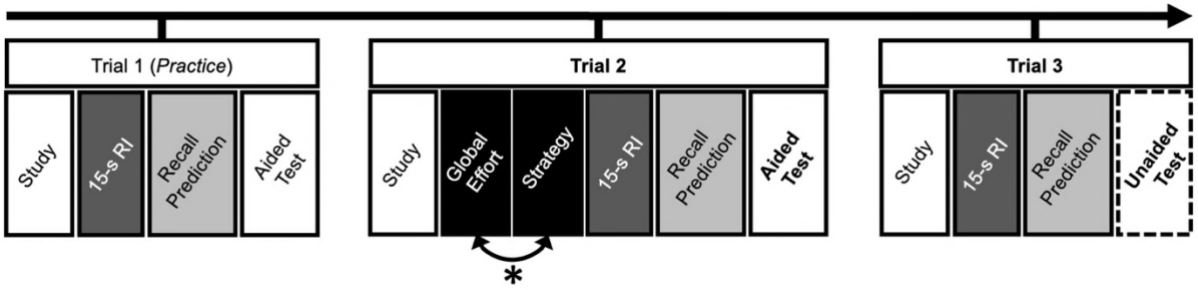
participants in the support conditions (i.e., full or partial) were given notice right before the test/recall phase that they would not have access to their support list, unlike in previous trials.

Test. At the beginning of this phase, participants in a support condition (i.e., full or partial) were first asked, "*In the previous list, 20 words were presented in total. How many words do you anticipate accurately typing into the response box during the memory test? Remember, you will have your saved items available,*" to which they selected a number on a scale from 0 to 20. Those in the No-support condition were presented with the following: "*In the previous list, 20 words were presented in total. How many words do you anticipate accurately recalling without your list available during the memory test?*" to which they responded by selecting a number on a scale from 0 to 20. Critically, on the third trial, participants in a support condition were asked, "*In the previous list, 20 words were presented in total. How many words do you anticipate accurately typing into the response box during test? Remember, you will NOT have your saved items available*." That is, these participants were reminded that they would *not* have their support list available and should predict memory performance without their support available (i.e., unaided memory).

After predicting, the recall test was administered. For those in the support conditions (i.e., full or partial), there was no support available to them unlike in the previous two trials. For those in the no-support condition, the trials were identical such that they were not instructed to expect any support list (though they still typed the items at study) and were not given any support list at test. Finally, participants took part in a post-task trial wherein their external supports were available, thus, on this final post-task trial, near perfect performance would be expected. This trial

was used only for assessing whether participant data quality was too low for inclusion (see exclusion criteria below). Figure 7 presents an overview of the procedure.

Figure 7. Experiments 2 and 3: The general procedure including where our dependent variables of interest were collected.



Note. *****The order of global effort rating and strategy was randomly determined for each participant and was collected in both the Aided Trial (Trial 2) and Unaided Trial (Trial 3) during Experiment 3. Also note, study time allocation was collected per item during each study phase, and there was a fourth trial (not pictured) as part of the data quality checking process.

Results

General analytical approach

Exclusions and outliers. All results reported in this work have had the preregistered exclusions applied. Specifically, we did not have to replace any participants for not wanting their data to be included, for reporting using an external store other than that provided by the experimental program, or for not typing out the full study item 75% of the time across Trials 2 and 3. One participant was replaced due to reporting that they were not attentive/were doing something else during the study, and 34 were replaced due to not recalling at least 75% on the post-task trust trial wherein individuals were given full store at test regardless of support condition (i.e., should have had a performance level near ceiling). Analogous analyses with these individuals included

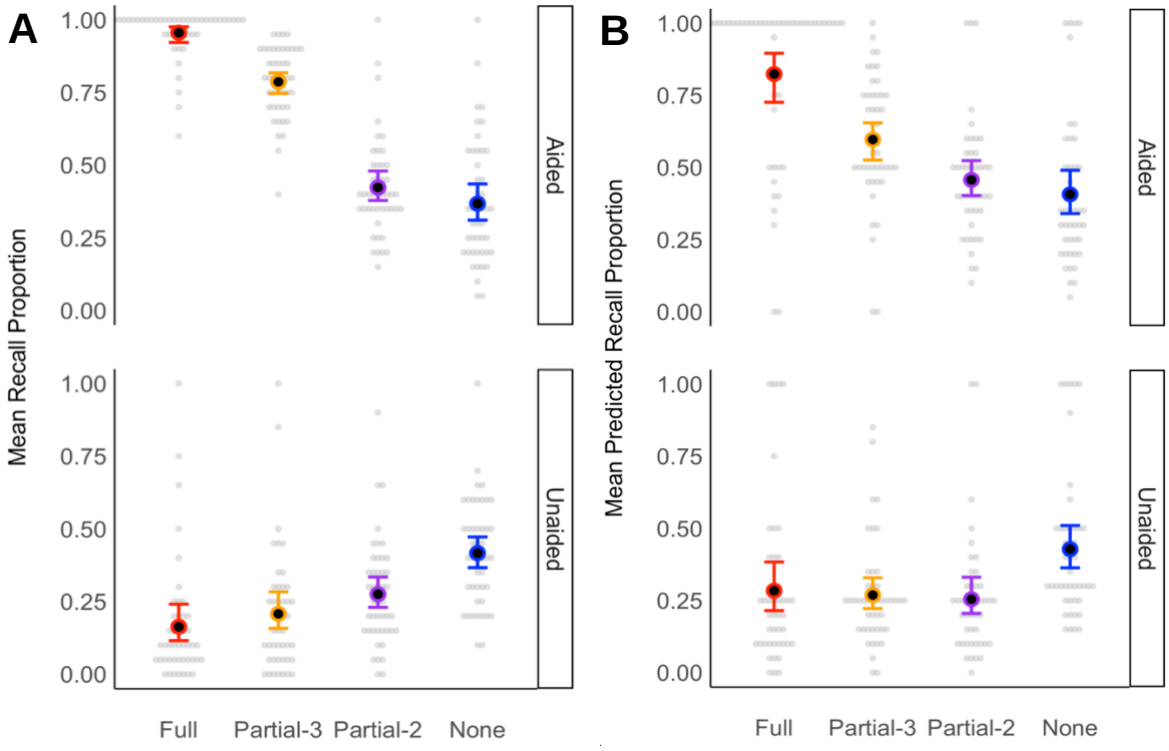
were conducted and revealed qualitatively the same results unless otherwise noted (data and code for these analyses are available at osf.io/9qxmc/). For analyses involving study time, within-participant outliers were detected using a non-recursive z criterion based on the number of trials within each block within each participant (i.e., 20; Van Selst & Jolicoeur, 1994). Analogous analyses with outliers still included revealed qualitatively the same results throughout (data and code for these analyses are available at osf.io/9qxmc/). One participant gave no predictions, and another participant responded only for the Aided Trial, thus, their data were not included in any analysis involving recall predictions.

Mixed models. Mixed models were conducted in R using the 'lmerTest' package (Kuznetsova et al., 2017).[7] When applicable, the random effects structure of models was determined using a model-fitting approach (via the *anova()* function; R Core Team, 2023). Because the environmental support condition is between-participants, participants only have data in *one level* of this manipulation, thus we did not include by-participant slopes for support condition (e.g., Brown, 2021). For logistic models, multiple models varying in which support condition was the reference group were conducted to obtain all paired comparisons. We preregistered a model-fitting approach using mixed models for study time; however, we also preregistered collapsing across trial type (i.e., aided vs. unaided), given that study phases were theoretically identical across trials. We deviated from the preregistration by only examining study time on the aided trial to remain consistent with the other study effort indices which were only

[7] The utility of *p*-values in the context of mixed effects models has been debated; we include them here, however, for the sake of reading ease and consistency with other analyses.

collected on the aided trial. Moreover, this allowed us to use the preregistered model-fitting approach for study time.

The between-participants comparisons reported (e.g., one-way ANOVAs, Welch's t-tests) either did not violate the assumption of homogeneity of variance (as tested using Levene's Test) or did violate it but were well within the recommended ratios of variance as recommended by Howell (2015). All analyses reported were preregistered unless otherwise specified. See Figures 8A and 8B for the means and confidence intervals for memory performance and metacognitive predictions of performance; See Table 2 for the means and confidence intervals for the study effort indices.



Figures 8A and 8B. Experiment 2: Means and confidence intervals for mean recall proportion (A) and mean predicted recall proportion (B) as a function of support condition and trial type.

Note. Participant-level data is represented by the gray points. Error bars are 95% bias-corrected accelerated confidence intervals with 10 000 replications.

Table 2. Experiment 2: Means and confidence intervals for each study effort index across support conditions on each trial.

	Full-support	**Partial-3**	**Partial-2**	**No-support**
Study time (ms)				
Aided Trial	2846	2501	2599	3204
	[2311, 4395]	[2244, 2816]	[2284, 3135]	[2736, 4709]
Unaided Trial	2879	2626	2882	3486
	[2312, 4050]	[2354 2985]	[2531, 3400]	[2846, 5643]
Global effort (1-9)	4.58	6.24	6.73	6.69
	[3.84, 5.29]	[5.64, 6.67]	[6.20, 7.20]	[5.98, 7.20]
Strategy use (% yes)	29	29	62	51
	[18, 47]	[16, 40]	[44, 73]	[33, 64]

Note: Error bars are 95% bias-corrected accelerated confidence intervals with 10,000 replications. Study time is reported as a function of aided versus unaided trial whereas Global effort and Strategy use are not as they were only collected once (i.e., on the Aided Trial).

Anticipated Environmental Support on Memory Performance

Aided Recall. A mixed effects logistic regression (with by-participant and by-item intercepts; structure determined via model-fitting approach) examined the influence of support condition (Full- vs. Partial-3 vs. Partial-2 vs. No-support) on recall for the Aided Trial. The likelihood that an item was recalled was significantly higher in the Full-support condition than in

the other conditions [all *b*s > 2.10, all *z*s > 7.60, all *p*s < .001]. The Partial-3 condition had significantly higher recall than those in the Partial-2 and No-support conditions [$b = 1.89$, $SE = 0.21$, $z = 8.80$, $p < .001$; $b = 2.17$, $SE = 0.22$, $z = 9.99$, $p < .001$] and the Partial-2 condition did not differ significantly from the No-support condition [$b = 0.29$, $SE = 0.21$, $z = 1.37$, $p = .172$].

Unaided Recall. A mixed logistic regression (with by-participant and by-item intercepts; structure determined via model-fitting approach) examined the influence of support condition (Full- vs. Partial-3 vs. Partial-2 vs. No-support) on recall for the Unaided Trial. Here, the likelihood of recall was significantly lower in the Full-support condition than in other conditions [*b*s > 0.88, *z*s > 3.36, *p*s < .001] except for the Partial-3 condition [$b = 0.38$, $SE = 0.27$, $z = 1.42$, $p = .174$]. Those in the Partial-3 condition were less likely to recall items than those in the Partial-2-support condition [$b = 0.51$, $SE = 0.26$, $z = 1.97$, $p = .049$], and significantly less likely than those in the No-support condition [$b = 1.27$, $SE = 0.26$, $z = 4.98$, $p < .001$]. Those in the Partial-2 condition were less likely to recall items than those in the No-support condition [$b = 0.80$, $SE = 0.26$, $z = 3.14$, $p = .002$]. Including the participants who failed the bonus all aided trial, Full and Partial-3 did not significantly differ [$b = 0.32$, $SE = 0.26$, $z = 1.25$, $p = .211$], Partial-3 and Partial-2 marginally differed [$b = 0.47$, $SE = 0.26$, $z = 1.82$, $p = .069$], and Partial-2 and No support significantly differed [$b = 0.63$, $SE = 0.24$, $z = 2.65$, $p < .001$].

Anticipated Environmental Support on Metacognitive Predictions

Predicted Aided Recall. A one-way ANOVA tested the influence of support condition (Full- vs. Partial-3 vs. Partial-2 vs. No-support) on predicted performance for the Aided Trial. There was a significant main effect of support condition [$F(3, 173) = 26.11$, $p < .001$, $\eta_G^2 = .31$].

Because of the six pairwise comparisons, we use an adjusted alpha level of .008 (i.e., .05/6; Bonferroni correction). Predicted recall was significantly higher in the Full support condition than in the other conditions [all *t*s > 4.12, all *p*s < .008, all *d*s > 0.87], significantly higher in the Partial-3 support condition than in the Partial-2 and No-support conditions [Partial-3 vs. Partial-2: $t(85.48) = 2.95$, $p = .004$, $d = 0.63$; Partial-3 vs. No-support: $t(86.33) = 3.91$, $p < .001$, $d = 0.83$], and not different between the Partial-2 and No-support conditions [$t(84.70) = 1.25$, $p = .213$, $d = 0.26$].

Predicted Unaided Recall. A one-way ANOVA examining the influence of support condition on the predicted performance of unaided memory demonstrated a main effect of support [$F(3, 173) = 4.55$, $p = .004$, $\eta_G^2 = .07$]. Again, given the six pairwise comparisons, we use an adjusted alpha level of .008 (i.e., .05/6; Bonferroni correction). Specifically, predictions in the Full support condition were not significantly lower (given the corrected alpha level) than in the no-support condition [$t(82.83) = 2.25$, $p = .012$, $d = 0.48$] and were not different from the Partial-3 and Partial-2 conditions [$t(72.21) = 0.43$, $p = .666$, $d = 0.09$; $t(78.08) = 0.67$, $p = .508$, $d = 0.14$]. The Partial-3 and Partial-2 conditions had significantly lower recall predictions than those in the No-support condition [$t(82.43) = 3.26$, $p = .001$, $d = 0.69$; $t(86.00) = 3.35$, $p = .001$, $d = 0.71$], with no difference between the Partial-3 and Partial-2 conditions [$t(84.92) = 0.32$, $p = .751$, $d = 0.07$].

Anticipated Environmental Support on Study Behavior

Study behavior was indexed by study time allocation per item (ms), global effort rating, and self-reported strategy (the latter two indices were collected between study and test during the Aided trial only).

Study time. A mixed linear regression (with by-participant intercepts included based on model-fitting approaches) examining the influence of support condition on Study time revealed no significant effect of support on study time allocation on the Aided Trial [*b*s < 703.0, *t*s < 1.58, *p*s > .110] or on the Unaided Trial [*b*s < 861.0, *t*s < 1.69, *p*s > .090]. The preregistration for Experiment 2 outlined this analysis be collapsed across the Aided and Unaided Trials (i.e., on the mean study times across trials); however, we opted to keep them separate for clarity and consistency with the structure of our other analyses (e.g., on memory performance and predicted performance). Analogous exploratory (i.e., not preregistered) mixed linear regression on the log-transformed study times (to mitigate the effects of skew) revealed qualitatively the same results [*b*s < 0.18, *t*s < 1.80, *p*s > .070].

Global effort rating. A one-way ANOVA examining the influence of support condition on global effort rating showed a significant effect of support condition [$F(3, 177) = 11.08$, $p < .001$, $\eta_G^2 = .16$]. The effort ratings of the Full support condition were significantly lower than those of the other conditions [*t*s > 3.57, *p*s < .001, *d*s > 0.75]. None of the other paired comparisons were significant [*t*s < 1.41, *p*s > .162, *d*s < 0.30].

Strategy. Logistic regression examined the influence of support condition on Strategy use (yes vs. no). Those in the Full support condition were significantly less likely to report strategy use than those in the Partial-2 and No-support conditions [$b = 1.43$, $SE = 0.45$, $z = 3.19$, $p = .001$; $b = .98$, $SE = 0.44$, $z = 2.20$, $p = .028$] but were no different from those in the Partial-3 condition [$b < .01$, $SE = .46$, $z < 0.01$, $p > .947$]. Those in the Partial-3 condition were less likely to report using a strategy than those in the Partial-2 and No-support conditions [$b = 1.40$, $SE = .45$, $z = 3.11$,

p = .002; b = 0.95, SE = .44, z = 2.13, p = .033]. There was no significant difference in strategy reports between the Partial-2 and No-support conditions [b = 0.45, SE = 0.43, z = 1.06, p = .289]. Including the participants who failed the bonus All Aided Trial, strategy reports in the Full and No-support conditions did not differ significantly [b = 0.61, SE = 0.38, z = 1.60, p = .109].

Secondary Analyses

Relation between study effort indices

We examined the relation among study effort indices that we collected during the Aided Trial—study time, global effort, and strategy—and found that these indices mostly correlated with one another. Specifically, the relation between study time and global effort rating [r(178) = .22, p = .002] and the relation between global effort and strategy [r(178) = .28, p < .001] showed clear positive correlations; only the relation between study time and strategy showed no clear correlation [r(178) = .08, p = .271].

Relation between predicted and actual performance

All conditions showed significant positive correlations between predicted and actual recall performance on the Aided Trial [Full: r(42) = .32, p = .031, Partial-3: r(43) = .30, p = .045; No-support: .85, p < .001] except for the Partial-2 condition [r(42) = .24, p = .113]. This pattern was also found on the Unaided Trial [Full: r(42) = .71, p < .001; Partial-3: r(43) = .61, p < .001; Partial-2: r(42) = .24, p = .120; None: r(43) = .51, p < .001]. Including the participants who failed the bonus All Aided Trial, for the Aided Trial, those in the Partial-3 condition did not show a significant positive correlation [r(45) = .20, p = .169] whereas those in the Partial-2 condition did

[r(46) = .29, p = .048]. Table 3 presents the correlations between predicted and actual memory performance by trial type and support condition for Experiment 2.

Table 3. Experiment 2: Correlations between predicted and actual memory performance by Trial Type (Aided vs. Unaided) and Support Condition (Full vs. Partial-3 vs. Partial-2 vs. No-support).

	Full	Partial-3	Partial-2	No-support
Aided	.32*	.30*	.24	.85***
Unaided	.71***	.61***	.24	.51***

*Note: *** < .001, ** < .01, * < .05.*

We used Fischer's *z* test to compare the predicted-actual recall correlations (from the 'cocor' package in r; Diedenhofen & Musch, 2015) between support conditions separately for the Aided and Unaided trials. Given the number of statistical tests, we used a corrected alpha level of .008 in each case (i.e., .05/6 tests; Bonferroni correction). On the Aided trial, those in the No-support condition had a significantly higher correlation between their predicted and actual recall performance than did those in the other conditions (zs > 4.73, ps < .001). There were no other significant differences (zs < 0.43, ps > .65). Including the participants who failed the bonus All Aided Trial, for the Aided Trial, those with Full support did not have significantly different metacognitive accuracy from those with No support [z = .90, p = .370]. On the Unaided trial, there were no significant differences in correlations using the corrected alpha level of .008 (zs < 2.13, ps > .03).

Apart from the main investigation of Experiment 2, we had two secondary motivations. The first of these was to examine the extent to which study effort mediates the relation between support condition and memory performance (for each study effort index; Kelly & Risko, 2022b).

The second motivation was to examine whether the *set size* of each word stem (the number of accurate completions) or the *base rate* of each study item (the commonness of the study item as the completed word when given its word stem; Jacoby et al., 1993) influenced the likelihood of recall in the partial conditions. A priori, one can imagine that word stems with larger set sizes and study items with lower base rates would be more challenging to recall. Despite being preregistered as Experiment 2 analyses, we opted to combine data across both Experiments 2 and 3 to better address these secondary questions (thereby increasing power and decreasing the number of statistical tests). For these results, see *Secondary Analyses* for Experiment 3.

Discussion

Aided Memory and Predictions

Results demonstrated that aided memory performance does depend on the degree of environmental support such that full support was significantly better than having 3 letters available, which was significantly better than having 2 letters available and no support, which did not differ from one another. Individuals' metacognitive predictions of aided recall followed the same pattern as actual aided recall such that individuals predicted they would do better with full support than having 3 letters, which would be better than having 2 letters or none (i.e., no support), which did not differ from one another. Thus, individuals predicted the positive effect of having external support relative to having no support at all, but also showed some evidence of sensitivity to different levels of support on aided memory performance.

Unaided Memory and Predictions

For unaided memory performance (i.e., recall without support after having expected to receive it in the supported conditions), all of the support conditions recalled fewer items than those in the No-support condition. In addition, those with full and 3-letter support (which did not differ reliably) had significantly lower recall than those with 2-letter support. Thus, expecting partial support modulated the cost of losing an expected store (i.e., the cost was less than when expecting full support). Interestingly, individuals' metacognitive predictions showed a different pattern. While predictions in all the conditions where individuals were expecting support were lower than in the No-support condition (marginally so for those with full support given the conservative Bonferroni correction), there were no differences across those with less, more, and full support. Thus, while the actual unaided memory performance of those expecting support was influenced by the degree of expected support, predictions of *unaided* performance were not. Rather, in general, individuals predicted that they would do worse having had expected any kind of support (i.e., regardless of the level of support) and not receiving it. This dissociation suggests a form of metacognitive difficulty in predicting the consequences of losing an expected store. This dissociation was absent when individuals made predictions about aided memory performance (i.e., when they expected to have support and were, indeed, supported).

Study Effort

Measures of study effort positively correlated with one another (although not significantly in the case of study time and strategy report). For study time, we found marginal evidence that the support conditions (i.e., Full, Partial-3, Partial-2) each showed significantly lower study time than the no support condition. That said, study time increased numerically for each added level of

support. Global effort ratings were lowest in the Full support condition, with the remaining differences not significant. The ratings reported by those with partial support were more numerically close to those with no support than to those with full support. Thus, in terms of self-reported effort, individuals reported investing less effort when fully supported but similar levels of effort when receiving less than full support. Finally, individuals with Full and 3-letter support were less likely to report strategy use (while not differing from one another) than were those with 2-letter and no support (which did not differ from one another). While the effort measures were somewhat inconsistent across condition, it is worth noting that in the partial-2 condition there were no differences in effort relative to the no support condition, but there was a cost of losing partial support. This pattern suggests that expecting an environmental support can lead to costs in the absence of clear evidence of differential effort being allocated across conditions.

Prediction-performance correlation

All support conditions generally demonstrated significant positive correlations between predicted and actual memory performance on both aided and unaided trials, although less clearly for partial support conditions. Those in the no-support condition demonstrated a significantly stronger relation between prediction and performance than those in each of the support conditions, suggesting that there is something about expecting a support that alters our ability to judge memory performance. Upon inspection, most individuals underestimated their aided memory performance (this is also hinted at when comparing the means between performance and prediction). On the Unaided Trial, those expecting Full and No-support (i.e., the extreme comparison groups) demonstrated moderate correlations between predictions and performance. Generally speaking, the prediction-performance correlations tended to be numerically weaker in the support conditions

(compared to no-support). This is consistent with the idea that predicting performance alongside the expectation of support could be more challenging than predicting unaided performance, especially upon the loss of a such an expected support.

In summary, individuals exhibited behavior generally consistent with predictions derived in the *Introduction.* Full and No-support conditions clearly demonstrated the expected patterns such that Full support led to higher aided and lower unaided performance compared to no support, consistent with their lower effort ratings and strategy reports. That said, results still departed from these predictions in somewhat unintuitive ways. For example, compared to those expecting no support, those expecting 2-letter support demonstrated a clear cost to unaided memory, yet there were no clear differences between these groups in study effort as indicated by global effort and strategy reports. Individuals were also rather insensitive to the amount of support when it came to predictions about unaided memory performance, despite there being a modulation of unaided recall by amount of support. These findings are consistent with the idea that the monitoring and control of study behavior is difficult with more intermediate forms of environmental support as opposed to when can anticipate no or extensive environmental support.

Experiment 3

There were multiple motivations for Experiment 3. Mostly, we were interested in replicating the unanticipated findings from Experiment 2. Indeed, results from Experiment 2 hinted that the monitoring and control of study behavior is more challenging in the case of intermediate levels of support compared to when expecting the extremes of full or no support. Replicating the findings of Experiment 2 would help solidify conclusions drawn about metacognitive abilities (i.e., performance prediction) as a function of anticipated environmental support. Aside from replicating

findings, we amended our procedure from Experiment 2 by collecting global effort ratings and strategy on the Unaided Trial (in addition to on the Aided Trial).

Method

In Experiment 3, we included only the 2-letter partial support condition, along with both full and no support, rather than having two partial support conditions like in Experiment 2. As just mentioned, we collected global study effort ratings and self-reported strategy on both the Aided and Unaided trials unlike in Experiment 2.

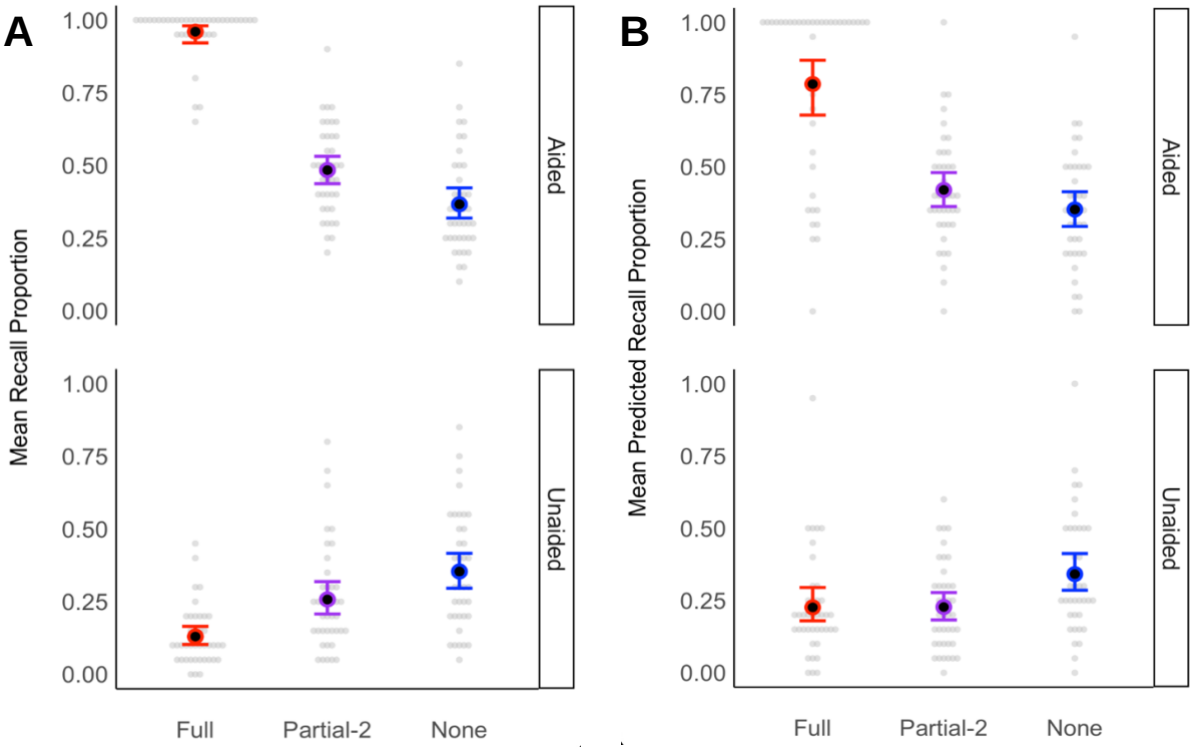
Participants were 120 (n = 40; 52 females, 67 males; M_{age} = 37.94 years, SD_{age} = 12.79 years) individuals from Prolific divided equally across the four groups, and were paid £3.75 for their participation (~20 minutes). No participants from Experiment 3 had also taken part in Experiment 2. This sample size was based on powering for a balanced one-way ANOVA with three between-participants groups, considering a power of over .80 to detect a medium effect size of f = .30 (two-tailed, alpha = .05) and considering a power of ~.80 to detect a between-participants pairwise comparison of d = 0.50 using ANOVA_exact (Lakens & Caldwell, 2021).

All other aspects of Experiment 3 were the same as Experiment 2 except for earlier mentioned difference of collecting global effort ratings and strategy reports on *both* the Aided and Unaided Trials, rather than just on the Aided trial (as in Experiment 2). The preregistration for Experiment 3 is available at osf.io/tp8xm. Data and analyses code are available at osf.io/9qxmc.

Results

As in the previous experiment, all results reported have the preregistered exclusions applied. No participants were replaced for not wanting their data to be included, not being attentive, reporting using an external store other than that provided by the experimental program, or not

typing the full study item out 75% of the time across Trials 2 and 3. Eighteen were replaced due to lower than 75% recall accuracy on the bonus trial (wherein all were fully aided, thus, wherein performance should have been near ceiling). Again, analogous analyses with these individuals included revealed qualitatively the same results unless otherwise noted. For analyses involving study time, within-participant outliers were detected using a non-recursive *z* criterion based on the number of trials within each block within each participant (i.e., 20; Van Selst & Jolicoeur, 1994) and study time analyses including these outliers revealed qualitatively the same results. Figures 9A and 9B present the means and confidence intervals for memory performance and metacognitive predictions of performance. See Table 4 for the means and confidence intervals for the study effort indices.



Figures 9A and 9B. Experiment 3: Means and confidence intervals for mean recall proportion (A) and mean predicted recall proportion (B) as a function of support conditions and trial type.

Note. Participant level data is represented by the gray points. Error bars are 95% bias-corrected accelerated confidence intervals with 10,000 replications.

Table 4. Experiment 3: Means and confidence intervals for each dependent variable across support conditions on each trial.

	Full-support	**Partial-2**	**No-support**
Study time (ms)			
Aided Trial	2514	3253	3107
	[2159, 3123]	[2664, 4641]	[2604, 3828]
Unaided Trial	2487	3193	3057
	[2180, 3142]	[2635, 4297]	[2627, 3635]
Global effort (1-9)			
Aided Trial	5.05	6.63	6.83
	[4.18, 5.83]	[5.95, 7.15]	[6.15, 7.30]
Unaided Trial	4.78	6.78	6.78
	[3.93, 5.60]	[6.03, 7.30]	[5.90, 7.42]
Strategy use (% yes)			
Aided	28	53	33
	[13, 40]	[35, 65]	[18, 45]
Unaided Trial	28	53	35
	[13, 40]	[35, 65]	[20, 48]

Note: Error bars are 95% bias-corrected accelerated confidence intervals with 10,000 replications. Study time is reported as a function of aided versus unaided trial whereas Global effort and Strategy use are not as they were only collected once (i.e., on the aided trial).

Anticipated Environmental Support on Memory Performance

Aided Recall. Mixed effects logistic regression (with by-participant and by-item intercepts; structure determined via model-fitting approach) examined the influence of support condition (Full- vs. Partial-2 vs. No-support) on recall for the Aided Trial. The likelihood that an item was recalled was significantly higher in the Full-support condition than in the other two conditions [all *b*s > 3.75, all *p*s < .001], and recall in the Partial-2 condition was significantly higher than in the No-support condition [$b = 0.62$, $SE = 0.20$, $z = 3.07$, $p = .002$].

Unaided Recall. Mixed effects logistic regression (with by-participant and by-item intercepts; structure determined via model-fitting approach) examined the influence of support condition (Full vs. Partial-2 vs. No-support) on recall for the Unaided Trial. The likelihood of recall was significantly lower in the Full-support condition than in the other two conditions [*b*s > .90, *p*s < .001] and recall in the Partial-2 condition was significantly lower than in the No-support condition [$b = 0.52$, $SE = 0.21$, $z = 2.51$, $p = .012$].

Anticipated Environmental Support on Metacognitive Predictions

Predicted Aided Recall. A one-way ANOVA examining the influence of support condition (Full- vs. Partial-2 vs. No-support) on the predicted performance for aided memory demonstrated a significant main effect of support condition [$F(2, 117) = 38.23$, $p < .001$, $\eta_G^2 = 0.40$] such that those in the Full support condition predicted significantly higher aided memory performance than did those in both the Partial-2 and No-support conditions [Full vs. Partial-2: $t(65.68) = 6.39$, $p < .001$, $d = 1.43$; Full vs. No-support: $t(66.80) = 7.50$, $p < .001$, $d = 1.68$], which did not significantly differ from each other [$t(77.94) = 1.54$, $p = .127$, $d = 0.34$].

Predicted Unaided Recall.[8] A one-way ANOVA tested the influence of support condition on predicted performance for the Unaided Trial and revealed a significant main effect of support condition [$F(2, 117) = 5.39$, $p = .006$, $\eta_G^2 = .08$]. Specifically, Full- and Partial-2 support conditions did not differ [$t(75.96) = .03$, $p = .973$, $d < 0.01$] but each predicted significantly lower recall than in the No-support condition [Full vs. No-support: $t(76.80) = 2.68$, $p = .009$, $d = 0.60$; Partial-2 vs. No-support: $t(72.21) = 2.83$, $p = .006$, $d = 0.63$].

Anticipated Environmental Support Influence on Study Behavior

The following analyses on the influence of anticipated environmental support on study behavior were preregistered to compare only partial support with no support for Experiment 3, however, we followed suit from the Experiment 2 preregistration wherein we compared all support conditions (i.e., full, partial, and no-support) to allow for easier comparison to Experiment 2. Unlike in Experiment 2, we collected each study effort index on both the Aided and Unaided trials and therefore, report the findings in each case.

Study time. Using linear regression, we found no clear difference between the conditions on either the Aided or Unaided Trials [*b*s < 706.0, *t*s < 1.68, *p*s > .080]. When including the participants who failed the bonus All Aided Trial, those in the Full-support condition had marginally lower study time than those in the No-support condition on both the Aided Trial and the Unaided Trial [Aided Trial: $b = 753.10$, $SE = 423.00$, $t = 1.78$, $p = .077$; Unaided Trial: $b =$

[8] Note that this analysis was mistakenly not specified in the preregistration of Experiment 3 analyses. It should have been preregistered given our intention to replicate the unexpected result in Experiment 2 that the predicted cost to unaided performance of expecting external support was not sensitive to degree of external support.

738.80, *SE* = 373.60, *t* = 1.98, *p* = .050], and marginally lower study time than those in the Partial-2 condition on the Unaided Trial [*b* = 680.90, *SE* = 391.8, *t* = 1.74, *p* = .084]. Exploratory mixed effects linear regression (not preregistered) found qualitatively the same results when comparing the log-transformed study times [*b*s < 0.18, *t*s < 1.72, *p*s > .080]. Including the participants who failed the bonus All Aided Trial, those in the Full-support condition had significantly lower log transformed study time than those in the No-support condition on both the Aided Trial and Unaided Trial [Aided Trial: *b* = 0.19, *SE* = 0.10, *t* = 2.01, *p* = .046; Unaided Trial: *b* = 0.23, *SE* = 0.09, *t* = 2.45, *p* = .016].

Global effort rating. A one-way ANOVA comparing Full, Partial-2, and No-support conditions in global effort rating found a significant main effect of condition on both the Aided and Unaided trials [Aided: $F(2,117) = 7.88$, $p < .001$, $\eta_G^2 = .12$; Unaided: $F(2,117) = 9.03$, $p < .001$, $\eta_G^2 = .13$]. On both the Aided and Unaided trials, participants in the Full-support condition rated effort significantly lower than those in the Partial-2 [Aided: $t(70.12) = 3.02$, $p = .004$, $d = 0.67$; Unaided: $t(76.85) = 3.46$, $p = .001$, $d = 0.77$] and No-support conditions [Aided: $t(69.36) = 3.42$, $p = .001$, $d = 0.77$); Unaided: $t(72.76) = 3.68$, $p < .001$, $d = 0.82$)], with no difference between the Partial-2 and No-support conditions [Aided: $t(77.97) = 0.47$, $p = .637$, $d = 0.11$); Unaided: $t(76.27) < 0.00$, $p > .999$, $d < 0.00$)].

Strategy. Logistic regression compared Full, Partial-2, and No-support conditions in strategy use (yes vs. no). Those in the Full condition were less likely to report strategy use than those in the Partial-2 condition on both the Aided and Unaided Trials [Aided: *b* = 1.07, *SE* = 0.48, *z* = 2.25, *p* = .024; Unaided: *b* = 1.07, *SE* = 0.48, *z* = 2.25, *p* = .024]. Those in the Partial-2 condition were marginally more likely to report using a strategy than those in the No-support condition on

the Aided trial [b = 0.83, SE = 0.46, z = 1.80, p = .073]. Including the participants who failed the bonus All Aided Trial, the difference between the Full and Partial-2 conditions was marginal on the Unaided Trial [b = 0.79, SE = 0.45, z = 1.78, p = .075]. No other comparisons within either the Aided or Unaided Trials were significant.

Relation between study effort indices

We examined the relation among study effort indices—study time, global effort, and strategy. Unlike in Experiment 2, all study effort indices were collected on both the Aided and Unaided trials so we obtained the correlations between these indices separately for each the Aided and Unaided Trial. We found that all indices significantly positively correlated with each other on both the Aided and Unaided Trials. Indeed, study time positively correlated with both global effort rating [Aided: $r(118) = .24$, $p = .008$; Unaided: $r(118) = .25$, $p = .006$] and strategy use [Aided: $r(118) = .20$, $p = .026$; Unaided: $r(118) = .30$, $p < .001$] with these latter indices also positively correlated [Aided: $r(118) = .41$, $p < .001$; Unaided: $r(118) = .28$, $p = .002$].

Relation between predicted and actual performance

Those in the Full support condition demonstrated a marginal positive correlation between their predicted and actual recall performance on the Aided Trial [$r(38) = .29$, $p = .071$] and no correlation on the Unaided Trial [$r(38) = .13$, $p = .408$]. Those in the Partial and No-support conditions demonstrated significant positive correlations between predicted and actual recall performance on both the Aided Trial [Partial: $r(38) = .59$, $p < .001$; No-support: $r(38) = .56$, $p < .001$] and the Unaided Trial [Partial: $r(38) = .59$, $p < .001$; No-support: $r = .68$, $p < .001$]. Including the participants who failed the bonus All Aided Trial, those in the Full condition demonstrated a significant positive correlation on the Aided Trial [r(43) = .46, p = .002]. Table 5 presents the

correlations between predicted and actual memory performance by trial type and support condition for Experiment 3 and collapsing across Experiments 2 and 3.

Table 5. Experiments 2 and 3: Correlations between predicted and actual memory performance by Trial Type (Aided vs. Unaided) and Support Condition (Full vs. Partial-2 vs. No-support) in Experiment 3 and also collapsing across Experiments 2 and 3.

	Full	Partial-2	No-support
Aided	.29	.59***	.56***
Unaided	.13	.59***	.68***
E2 + E3 combined			
Aided	.30**	.36***	.70***
Unaided	.42***	.37***	.64***

*Note: *** < .001, ** < .01, * < .05*

As in Experiment 3, we used Fischer's z test to compare the predicted-actual recall correlations between support conditions for each of the Aided and Unaided trials. On the Aided Trial, there was no significant difference in the magnitude of predicted-actual correlations among conditions (zs < 1.62, ps > .107). On the Unaided Trial, those in the Full-support condition had a significantly lower predicted-actual correlation than those in both the Partial-2 and the No-support conditions (Full vs. Partial-2: $z = 2.33$, $p = .020$; Full vs. No-support: $z = 2.99$, $p = .002$), which did not significantly differ from each other ($z = 0.65$, $p = .511$).

Effort as a mediator (Partial vs. No-support)

Previous work provides evidence that study effort partially mediates the relation between support condition and unaided recall performance (Kelly & Risko, 2022b). In Experiment 3, we

were interested in examining whether study effort explained any of the differences in memory performance, but particularly in the case of comparing Partial and No-support conditions. We intended to examine this separately for each study effort index (i.e., study time, global effort rating, strategy) given a significant total effect—that is, those in the Partial-2 condition had significantly lower unaided memory performance than those not supported at all [b = 0.10, SE = 0.04, t = 2.28. p = .025]. That said, there was no significant effect of support condition on any of the study effort indices, thus, we did not continue with these mediation analyses [*Study time:* b = 54.10, SE = 498.7, t = 0.11, p = .914; *Global effort rating*: b < .01, SE = 0.51, t < .01, p > .999; *Strategy*: b = 0.72, SE = 0.46, z = 1.57, p = .117].

Secondary Analyses

Study effort as a mediator continued

As noted earlier, one secondary interest of ours was to replicate the previous finding that study effort partially mediates the relation between support condition and unaided memory performance (Kelly & Risko, 2022b). To do so, we collapsed across both Experiments 2 and 3 to compare between Full and No-support conditions, the conditions most analogous to those of Kelly and Risko (2022b). The total effect of support condition on recall performance was significant [b = 1.11, SE = 0.13, p < .001] and is the same for all three mediation analyses of interest, thus it is only reported here once. Starting with study time, there was no significant relation between support condition and study time, so we did not continue with the mediation analysis [b = 0.23, SE = 0.15, p = .132]. That said, the results were in the predicted direction, such that those with no support demonstrated longer study time.

For the global effort rating, those in the no-support condition had higher ratings of effort compared to those in the full support condition [b = 0.78, SE = 0.14, p < .001], therefore we tested global effort rating on recall performance controlling for support condition and found that the effort rating remained significantly higher for those in the no-support condition [b = 0.28, SE = 0.07, p < .001] when controlling for the effect of support condition [b = 0.89, SE = 0.13, p < .001]. The average indirect effect of effort rating was significant [0.22, p < .001], hence study effort significantly mediated the relation between support condition and recall performance. Critically, the direct effect remained significant even when including the mediator of study effort rating in the model [0.89, p < .001].

For strategy use, those in the no-support condition were more likely to report the use of a strategy [b = 0.15, SE = 0.07, p = .038], therefore we tested strategy report on recall performance controlling for support condition and found that the effect of strategy on recall performance was significant [b = 0.60, SE = 0.13, p < .012], when controlling for the effect of support condition on recall performance [b = 1.02, SE = 0.12, p < .001]. The average indirect effect of strategy use was significant [.09, p = .028], hence strategy report significantly mediated the relation between support condition and recall performance. Like for global effort rating, the direct effect of external support condition remained significant when including the mediator of strategy use [1.02, p < .001]. Including the participants who failed the bonus All Aided Trial, the effect of condition on strategy was not significant [b = 0.11, SE = 0.07, p = .112].

The influence of set size and base rate on recall

The following procedures were completed for the partial support conditions across Experiments 2 and 3 and based on work by Migo et al. (2010) and Jacoby (1998) as these were the

conditions that were presented word stems as partial external memory support during the Aided Trial. When given a word stem (APP—or AP---), the set size is defined as the number of legal completions for that word stem (e.g., APPLE, APPLY; or APPLE, APART). In the current work, all stimuli were 5 letters in length, hence, the word stems in the current work were defined as the number of 5-letter legal completions and we based this on the SUBTLEX-US corpus (corpus used available at osf.io/9qxmc/).

Base rates were computed from data collected from N = 200 participants on PROLIFIC paid £1.25 for their participation (~ 10 min) and who did *not* take part in either Experiment 2 or 3. These participants were presented with the unique word stems for the words of Experiments 2 and 3 such that half the participants completed the 2-letter stem sets while the other half completed the 3-letter stem sets. Participants were told that they had to complete the word stems with the 5-letter words that came to mind easiest, and to avoid names (e.g., proper nouns) or plurals (4-letter words with "s" added). If they could not think of a valid word, participants were told not to use external resources (e.g., the internet) and to, instead, type a question mark, '?'. Indeed, participants could only advance to the next word stem if they provided a response that was exactly 5 letters in length (or a question mark in the case that they could not think of a word). An independent research assistant coded whether the responses were valid completions based on the criteria. Base rates were then derived by dividing the number of times participants responded with the study set items in response to the relevant word stems out of the total number of times they responded with a legal word stem completion.

To relate set size and base rate to memory performance, we computed item-level means for recall performance and conducted correlation analyses relating recall performance, set size, and

base rate. We found no relation between recall performance and set size [$r(115) = .14$, $p = .132$], nor between recall performance and base rate [$r(115) = -.06$, $p = .496$], though unsurprisingly there was a significant negative correlation between set size and base rate [$r(115) = -.67$, $p < .001$].

Discussion

Aided Memory and Predictions

Those with full support had significantly better aided memory than those with *partial* support (i.e., 2/5 letters) and those with partial support had significantly better aided memory than those with no support. This pattern was also found in Experiment 2 but, unlike in Experiment 2, those in the Partial-2 condition had significantly higher memory performance than those in the No-support condition. For the metacognitive predictions of performance, those with Full support predicted significantly higher memory than did those in both the Partial-2 and No-support conditions, which did not significantly differ. Like Experiment 2, Experiment 3 demonstrated clear evidence that individuals are sensitive to the positive effect of having external support relative to having no support at all, but the evidence is less clear regarding sensitivity to the influence of varied levels of support on aided memory performance.

Unaided Memory and Predictions

Unaided recall performance was significantly lower in the Full support condition than in the Partial-2 support condition and significantly lower in the Partial-2 condition than in the No-support condition, like in Experiment 2. We also replicated the unexpected finding from Experiment 2 that metacognitive predictions in the conditions with support (i.e., Full and Partial) were lower than in the No-support condition, with no difference between these supportive

conditions. Indeed, as we started to conclude in Experiment 2, it seems that people were insensitive to the effects of varied levels of support on unaided memory. Instead, they only predicted that memory would be generally worse when they no longer had access to a support compared to when they expected no support.

Study Effort

With respect to the indices of study effort—study time, global effort rating, and strategy (which consistently positively correlated with each other)—results were a bit mixed regarding the original predicted pattern of Full < Partial < No-support, but consistent with Experiment 2. First, study time was numerically lower in the Full-support condition than in the Partial-2 and No-support conditions. It is not surprising that this was the case on both the Aided and Unaided trials, given that the study phases of each are theoretically treated the same by participants (in both cases, the support conditions expect support, and the no-support condition do not). Like in Experiment 2, the global effort ratings reported by those with partial support were essentially equal to those with no support, with these effort ratings significantly higher than provided by those with full support. Thus, as we concluded in Experiment 2, people report using less effort when fully supported but similar levels of effort when receiving less than full support. Also like in Experiment 2, those in the Full support condition were significantly less likely to report strategy use than those in the Partial-2 and No support conditions (which only marginally differed on the Aided Trial).

We expected that if participants were modulating the monitoring and control of memory processes proportionally (or perhaps more precisely, inversely proportionally) to expected support, then effort ratings would align accordingly with respect to level of support. Instead, while

expecting Full support clearly led to less effort than expecting no support, expecting partial support led to equivalent effort (or possibly more) compared to no support. Nonetheless there was still a clear cost to losing such partial support. Thus, these findings provide evidence of a cost to memory that is not as easily explained by effort reduction. We return to this in the General Discussion.

A final goal of the current investigation was to test a study effort explanation for the cost of external memory store reliance, as we have done previously (Kelly & Risko, 2022b). When we collapsed across both experiments to compare the Full and No-support conditions, we found some evidence that study effort partially mediates the cost of memory store reliance, though the degree to which study effort indices supported this varied by index. Specifically, we found clear evidence for global effort rating as a significant mediator (a novel measure not collected by Kelly & Risko, 2022b), evidence for strategy use as a mediator (though not when including those who failed the all aided bonus trial), and no conclusive evidence for study time as a mediator.

Prediction-performance correlation

We found that those in the Partial and No-support conditions demonstrated clear positive correlations between predicted and actual memory performance; those in the Full condition showed a weaker (if present) correlation. This differed from Experiment 2 wherein the Partial support conditions seemed to show weaker correlations between predicted and actual memory performance. Indeed, in Experiment 2, it seemed that those in the Full support and No support conditions were making predictions reliably correlated with their actual performance. While these differences could suggest some sample-to-sample variability, both experiments suggest that predictions made by those expecting *no* support will strongly correlate with actual performance

rather consistently and that this is less likely the case when expecting a form of environmental support.

General Discussion

Across two preregistered experiments, we examined how individuals' expectations about environmental support influenced their memory performance, study effort, and predictions of memory performance. We turn first to the contrast between expecting full and no support, then to the partial support results.

Individuals who expected full support demonstrated *higher* memory performance when they were, indeed, supported compared to those expecting and receiving *no* support. When individuals expecting full support did *not* receive such support, they demonstrated *lower* performance compared to those expecting and receiving no support. These findings are in line with the results of previous research comparing groups expecting full forms of support versus no support (Kelly & Risko 2019a; 2019b; 2022a; 2022b; Lu et al., 2020; 2021; Park et al., 2022; Risko et al., 2024; Risko & Kelly, 2024; Sparrow et al., 2011). The resulting differences between these support conditions (e.g., in memory performance) have been proposed to be, at least in part, due to differential study efforts (Kelly & Risko, 2022b).

To further the examination of a study effort explanation, we incorporated study effort indices of study time, strategy use, and—novel here—self-reported effort ratings. We found that study effort indices were clearly consistent with the findings of memory performance, such that when expecting full support, study time, effort rating, and strategy use all were lower compared to when expecting no support. These findings comparing Full and No support conditions are

consistent with a study effort explanation of the cost to memory when there is "full" external support available. A study effort explanation is further supported by our secondary analyses collapsing across experiments, wherein we found that effort rating and strategy use partially mediated the relation between environmental support (full vs. no-support) and unaided recall performance. In sum, when comparing the more extreme forms of environmental support—full vs. no support—individuals seem in tune with the influence of expected environmental support on memory performance.

We also obtained predictions of memory performance to examine how external support affected expectations of memory performance. We found that individuals were able to predict that their performance would be higher when fully supported compared to when not supported. There was also evidence that they were able to predict that their performance would suffer upon the unexpected loss of full support compared to when expecting *no* support. In addition, predictions of performance tended to correlate positively with actual performance in each condition collapsed across experiments, suggesting some metacognitive ability regardless of condition (see *E2+E3 combined* in Table 5). Taken together, the full and no support conditions support past empirical and theoretical conclusions. Expecting full support leads to a withdrawal of study efforts and this withdrawal leads to a cost when that expected support is unavailable. This withdrawal of effort does not, however, appear to explain the entire cost.

Partial Environmental Support

Critical to the current work is our inclusion of intermediate forms of support, *partial* support. This allowed an extension of earlier work to a context wherein the level of support was

intermediate rather than "all or nothing," and revealed multiple noteworthy results. First, like when unexpectedly losing a fully supportive external store, we found evidence in both experiments that the unexpected loss of a partially supportive store leads to a cost to unaided memory relative to when expecting no support. This cost, however, was smaller than that observed in the case of losing full support. Perhaps counterintuitively, however, this cost appeared *not* to be driven by a *withdrawal* of study effort as proposed by a study effort explanation. This is most clear when comparing the Partial-2 and No-support conditions, such that there was no significant influence of support condition on any effort indices. Thus, the cost to memory in this context does not seem easily explained with study effort differences.

One speculative explanation for this result is that the *partially* saved information led to investing study effort via a strategy that was ineffective in the *absence* of the partially supportive store compared to that of the No-support condition. Thus, when individuals lose that support, their performance suffers, but not because they had not invested effort in the first place. This demonstration is theoretically significant as it establishes a clear instance of a cost of external memory support that is not a product of withdrawing effort. These conclusions are less clear when comparing the results between the Partial-3 and No-support conditions, however, it is possible that expecting 3/5 letters (most of the item) was enough for participants to behave more like those expecting Full support.

Predictions of Unaided Memory

As noted above, the cost in the partial conditions was smaller than that observed in the full support condition. However, participants' metacognitive predictions did not follow this pattern in

either experiment. Instead, after learning about the loss of their environmental support, individuals in the Full and Partial support conditions all predicted similar unaided memory performance at the aggregate level, with these predictions significantly lower than those of the no support condition. One potential explanation for the inaccurate predictions is that, all else being equal, we hold a general belief that the unexpected loss of a memory support has negative consequences for memory performance, and that individuals rely on this belief as opposed to memorial cues that would correctly lead to predicting a differential cost. The latter might reflect a tendency to ignore such cues in light of a cue as salient as a large shift in expectations (i.e., going from expecting a store to not receiving it) or might reflect the absence of such cues. That said, it is worth noting that there were positive correlations among memory predictions and memory performance across all conditions. Thus, participants clearly have available enough information to make predictions that are to some extent accurate, just not the kind of information that would capture between condition differences in performance.

Conclusion

In two preregistered experiments, we found evidence that individuals can modulate their study efforts in response to the degree of environmental support that they can expect. Results support a study effort based explanation for the cost observed when expecting (but not receiving) full external memory support and a cost to memory that is not entirely explainable by study effort modulation. Insights such as these are important for bettering our understanding of the monitoring and control of memory and are broadly important as we continue to incorporate various environmental supports into our day-to-day lives.

Chapter 4

The following work is under review (Kelly et al., submitted).

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, 2023; Richmond et al., 2023; Risko et al., 2019; 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 0.80 or higher; e.g., Cohen, 1988; Kelly & Risko, 2019a; 2019b; 2022a; 2022b).[9] 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, 2023; 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

[9] 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) weare 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…*"

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, and 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 versus intentional learning, Anderson & Bower, 1972; Dorbush & Winnich, 1964; Eagle & Leiter, 1964; Postman, Adams, & Phillips, 1955, although see Popov & Dames, 2023a; 2023b; list-method directed forgetting, Sahakyan et al., 2009; 2013), 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 both directed forgetting and the current offloading memory procedures, internal memory is tested for items that presumably need 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; Dorbush & Winnich, 1964;

Eagle & Leiter, 1964; Postman, Adams, & Phillips, 1955; though see Craik, 2023, Dames & Popov, 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; Dorbush & Winnich, 1964; Eagle & Leiter, 1964; Postman, Adams, & Phillips, 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*[10] but not in items reported as *Know*. According to a

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

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).

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 5, 6a, and 6b; 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 4

Experiment 4 was preregistered at osf.io/e2bxm.

Method

Participants

Data was 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 .80 ($\alpha = .05$, 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 4 to 10 letters and word frequencies ranged from 3 (*inartistic*) to 82060 (*second*) using FREQCount from SUBTLEX-UK (Van Heuven et al., 2014) and from 1 (*inartistic*) to 21384 (*friend*) using FREQCount from SUBTLEX-US (Brysbaert & New, 2009).[11] Lists were counterbalanced across trial position (i.e., first through fifth) and across target and foil designations.

[11] 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.*

Procedure

The experimental task was administered on desktop computers via E-Prime version 3.0 stimulus presentation software. Participants were invited in groups of 3-4 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. Participants engaged in a total of 5 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 a list of 30 words one at a time; these were displayed in white on a black background at the centre 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). 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 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 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, 2023). For relevant analyses on sensitivity (*d'*), extreme values of 0 or 1 for hit rate and false alarm rate were corrected by adding .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 6 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 Experiment 4 are available at osf.io/y36p2.

Table 6. Experiment 4: Means and Confidence intervals for Hit rate, False alarm rate, and Sensitivity across trials and conditions.

	Hit rate	**False alarm rate**	**Sensitivity**
Trust Trial 1	.85 [.75, .88]	.05 [.02, .17]	2.85 [2.12, 3.11]
Trust Trial 2	.86 [.83, .89]	.05 [.03, .08]	2.86 [2.61, 3.10]
Trust Trial 3	.85 [.81, .88]	.05 [.03, .08]	2.91 [2.63, 3.19]
Told-no-store	.85 [.84, .91]	.08 [.05, .11]	2.66 [2.38, 2.94]
Told-store	.82 [.80, .87]	.09 [.07, .14]	2.43 [2.14, 2.74]

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

Hit Rate

A paired samples t-test revealed no difference between the told-no-store condition and the told-store condition [$t(38) = 1.38$, $p = .174$ $d = 0.22$]. Analogous mixed effects logistic regression revealed that participants in the told-no-store condition were significantly more likely to respond correctly to a target [$b = 0.39$, $SE = 0.17$, $z = 2.37$, $p = .018$]. 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), paired-samples t-tests revealed no statistically significant differences between the final trust trial hit rate and either external store condition (trust: .85; told-no-store: .84; told-store: .82) [trust vs. told-no-store: $t(38) = 0.24$, $p = .809$, $d = 0.04$; trust vs. told-store: $t(38) = 1.86$, $p = .071$, $d = 0.30$] although hit rate was numerically higher on the trust trial in both cases.

False Alarm Rate

Paired samples t-test found no significant difference between external support conditions for false alarm rate [$t(38) = 1.11$, $p = .274$, $d = 0.18$]. Analogous mixed effects logistic regression revealed the same result [$b = .01$, $SE = 0.30$, $z = 0.03$, $p = .974$]. The final trust trial false alarm rate was significantly lower than the false alarm rate for either external store condition (trust: .05; told-no-store: .08; told-store: .09) [trust vs. told-no-store: $t(38) = 2.70$, $p = .010$, $d = 0.43$; trust vs. told-store: $t(38) = 2.46$, $p = .019$, $d = 0.39$].

Sensitivity (d')

There was no significant difference between external support conditions for discriminating between targets and foils [$t(38) = 1.64$, $p = .110$, $d = 0.26$]. 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: $t(38) = 2.48$, $p = .018$, $d = 0.40$; trust vs. told -store: $t(38) = 3.65$, $p = .001$, $d = 0.59$].

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) [$t(38) = 2.23$, $p = .032$, $d = 0.36$].

Post-task prompts

As mentioned at the end of the procedure, 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 [$t(37) = 0.47$, $p = .644$, $d = 0.08$]. 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 7 below presents the breakdown for internal-external ratings for Experiment 4.

Table 7. Experiment 4: 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%

Discussion

Experiment 4 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 4 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 4 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 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 et al., 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 5

Experiment 4 provided some (but certainly not overwhelming) support for an effect of expecting external support on unaided recognition memory performance. In Experiment 5, 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, Kazanas, & Altarriba, 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[12] whereas items reported as Know showed no directed forgetting effect. Experiment 5 was preregistered at osf.io/ap2tr.

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

Method

The method of Experiment 5 was the same as that of Experiment 4 apart from the implementation of the Remember/Know/New procedure instead of the standard Yes/No recognition test used in Experiment 4. Here, participants were asked to make a "remember," "know," or "new" response by pressing "R", "K," or "N" on the keyboard. 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 .80 ($\alpha = .05$, 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 4 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 = [p("remember")]; and familiarity estimates reflect the *proportion* of "know" responses divided by the *proportion* of *non*-remember responses: F = p("know") / [1 – 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 4, 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) x 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 x 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 6a and 6b as well despite not preregistering these analyses initially). Table 8 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 Experiment 5 are available at osf.io/y36p2.

Table 8. Experiment 5: Means and Confidence intervals for Hit rate, False alarm rate, and Sensitivity across trials and conditions.

	Hit rate	**False alarm rate**	**Sensitivity**
Trust Trial 1	.84 [.80, .87]	.10 [.07, .16]	2.52 [2.32, 2.70]
Trust Trial 2	.84 [.81, .86]	.08 [.05, .14]	2.65 [2.46, 2.82]
Trust Trial 3	.85 [.82, .88]	.07 [.04, .14]	2.73 [2.51, 2.94]
Told-no-store	.86 [.85, .90]	.14 [.11, .24]	2.53 [2.26, 2.79]
Told-store	.82 [.81, .87]	.14 [.11, .24]	2.29 [2.05, 2.51]

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

Hit Rate

A paired samples t-test revealed that the told-no-store condition had a higher hit rate than the told-store condition [$t(54) = 2.15$, $p = .036$, $d = 0.30$] and analogous mixed effects logistic regression confirmed the same result [$b = 0.40$, $SE = 0.17$, $z = 2.42$, $p = .016$]. Exploratory analyses revealed no difference between the final trust trial and told-no-store critical trial hit rate (trust: .85; told-no-store: .86; told-store: .82) [$t(54) = 0.50$, $p = .617$, $d = 0.07$], although the trust trial hit rate was significantly higher than that of the told-store condition [$t(54) = 2.16$, $p = .035$, $d = 0.29$].

False alarms

A paired samples t-test revealed no significant difference between conditions for false alarm rate [$t(54) = 0.47$, $p = .642$, $d = 0.06$]. Analogous mixed effects logistic regression confirmed the same result [$b = 0.12$, $SE = 0.18$, $z = 0.68$, $p = .498$]. The trust trial false alarm rate was significantly lower than the false alarm rate for either external store condition (trust: .07; told-no-store: .14; told-store: .14) [trust vs. told-no-store: $t(54) = 4.44$, $p < .001$, $d = 0.60$; trust vs. told-store: $t(54) = 4.04$, $p < .001$, $d = 0.54$].

Sensitivity (d')

Those in the told-no-store condition demonstrated greater sensitivity than those in the told-store condition [$t(54) = 2.06$, $p = .045$, $d = 0.28$]. 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: $t(54) = 2.36$, $p = .022$, $d = 0.32$; trust vs. told-store: $t(54) = 4.48$, $p < .001$, $d = 0.60$].

Correct response time (ms)

There was no significant difference in correct response times between external support conditions (told-no-store: 1119; told-store: 1162) [$t(55) = 1.19$, $p = .238$, $d = 0.16$].

Recollection and Familiarity

Targets. A 2 (told-no-store vs. told-store) x 2 (recollection vs. familiarity) ANOVA demonstrated no main effect of condition (told-no-store: .53; told-no-store: .53) [$F(1, 48) < .01$, $p = .984$, $\eta_G^2 < .01$], no significant main effect of process (recollection: .58; familiarity: .48) [$F(1, 48) = 2.12$, $p = .152$, $\eta_G^2 = .02$], and no significant interaction such that the effect of external store expectations was not significantly different between recollection and familiarity estimates (recollection: .04; familiarity: = -.04) [$F(1, 48) = 3.43$, $p = .070$, $\eta_G^2 < .01$]. 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 [$t(48) = 1.68$, $p = .100$, $d = 0.24$] or familiarity [$t(48) = 1.05$, $p = .299$, $d = 0.15$] estimates.

Foils. A 2 (told-no-store vs. told-store) x 2 (recollection vs. familiarity) ANOVA demonstrated no main effect of condition (told-no-store: .09; told-store: .09), $F(1, 48) < 0.01$, $p = .970$, $\eta_G^2 < .01$, a main effect of process such that the estimates for recollection were lower than for familiarity (recollection: .06; familiarity: .11) [$F(1, 48) = 10.43$, $p = .002$, $\eta_G^2 = .03$] and no interaction (recollection: < .01; familiarity: < .01) [$F(1, 48) = 0.23$, $p = .633$, $\eta_G^2 < .01$]. Paired samples t tests (not preregistered) on recollection and familiarity estimates revealed no effect of expecting support on recollection [$t(48) = .48$, $p = .634$, $d = 0.07$] or familiarity [$t(48) = 0.28$, $p = .781$, $d = 0.04$] 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 5 [$t(51) = 0.47$, $p = .640$, $d = 0.07$]. Table 9 below presents the breakdown for internal-external ratings for Experiment 5.

Table 9. Experiment 5: 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%

Discussion

In Experiment 5, we found results consistent with Experiment 4 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 4, there was a significant effect of condition on sensitivity, and no effect on correct response times. That is, in Experiment 5, we found additional evidence of a small cost ($d = \sim .30$) 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 4 by breaking recognition memory down further into recollection and familiarity processes in Experiment 5. This was based on the idea that a study effort hypothesis would predict that recollection estimates would be lower for those 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 (.04 vs. -.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 4, 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 4, 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 4, 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 6a & 6b

Experiment 5 demonstrated clearer evidence than Experiment 4 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 6a and 6b, we aimed to put these results on stronger footing via two replications with larger sample sizes. Experiment 6a was replicated in Experiment 6b, thus, the method and results are described together. Experiments 6a and 6b were preregistered at osf.io/5p8m4 and osf.io/kx835 respectively.

Method

Participants

Participants were undergraduate psychology students taking part for course credit. For Experiment 6a, data from N = 90 participants were collected and analyzed based on an a priori power analysis with a desired power of .80 (α = .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 6b, we preregistered N = 120 participants based on an a priori power analysis with a desired power of .80 (α = .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 6a were those of earlier experiments, just with 5 fewer words per list (randomly determined) such that lists were 25 items long. In Experiment 6b, however, we returned to using 30-item word lists using the same 300 words from Experiments 4 and 5 but redrawing items randomly to generate 10 novel lists. Unlike Experiments 4 and 5, the remaining experiments were collected online due to work-from-home protocols given the status of COVID-19 pandemic. The procedure for Experiments 6a and 6b 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 to 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 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 to follow.

Results

All analyses followed the preregistration unless otherwise specified. As mentioned, due to moving to online data collection for Experiments 6a and 6b, we preregistered the following replacement of participants who met these exclusion criteria: (1) did not type at least 85% of words for trials 2 to 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 give 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 6a, 13 participants were replaced due to not encoding at least 85% of study words across trials 2 through 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 6b, 7 participants were removed due to not encoding at least 85% of study words across trials 2 through 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 6b, 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 5. Eight participants in Experiment 6a and five participants in Experiment 6b 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 10 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 6a and 6b are available at osf.io/y36p2.

Table 10. Experiments 6a and 6b: Means and Confidence intervals for Hit rate, False alarm rate, and Sensitivity across trials and conditions.

	Hit rate	False alarm rate	Sensitivity
Trust Trial 1	E6a: .92 [.90, .93] E6b: .92 [.91, .93]	E6a: .03 [.02, .05] E6b: .04 [.02, .06]	E6a: 3.28 [3.16, 3.41] E6b: 3.33 [3.19, 3.45]
Trust Trial 2	E6a: .92 [.90, .93] E6b: .90 [.89, .92]	E6a: .02 [.01, .04] E6b: .03 [.02, .05]	E6a: 3.29 [3.16, 3.41] E6b: 3.19 [3.07, 3.30]
Trust Trial 3	E6a: .92 [91, .94] E6b: .90 [.88, .91]	E6a: .03 [.02, .05] E6b: .04 [.03, .07]	E6a: 3.26 [3.12, 3.39] E6b: 3.13 [2.99, 3.26]
Told-no-store	E6a: .85 [.83, .88] E6b: .86 [.84, .88]	E6a: .11 [.08, .14] E6b: .13 [.11, .17]	E6a: 2.49 [2.37, 2.79] E6b: 2.46 [2.25, 2.59]
Told-store	E6a: .83 [.80, .86] E6b: .83 [.82, .86]	E6a: .12 [.08, .14] E6b: .15 [.12, .19]	E6a: 2.47 [2.28, 2.65] E6b: 2.22 [2.08, 2.42]

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

Hits

There was no significant difference in hit rate between the told-store and told-no-store instruction conditions in Experiment 6a (told-no-store: .85; told-store: .83) [$t(89) = 1.69$, $p = .094$, $d = 0.18$] but there was in Experiment 6b (told-no-store: .86; told-store: .83) [$t(120) = 2.86$, $p = .005$, $d = 0.26$]. Analogous mixed effects models revealed the same results [E6a: $b = 0.15$, $SE = 0.12$, $z = 1.23$, $p = .219$; E6b: $b = 0.29$, $SE = 0.10$, $z = 2.80$, $p = .005$]. In Experiment 6a, exploratory analyses revealed hit rate was higher in the trust trial than for either external store condition (trust:

.92; told-no-store: .85; told-store: .83) [trust vs. told-no-store: $t(89) = 6.18$, $p < .001$, $d = 0.65$; trust vs. told-store: $t(89) = 7.24$, $p < .001$, $d = 0.76$. The same was found for Experiment 6b [trust vs. told-no-store: $t(120) = 3.42$, $p = .001$, $d = 0.31$; trust vs. told-store: $t(120) = 5.40$, $p < .001$, $d = 0.49$] (trust: .90; told-no-store: .86; told-store: .83).

False alarms

There was no significant difference in false alarm rate between the told-no-store and told-store instruction conditions in Experiment 6a [$t(89) = 1.31$, $p = .195$, $d = 0.14$] or Experiment 6b [$t(120) = 1.96$, $p = .052$, $d = 0.18$]. Analogous mixed effects logistic regression revealed qualitatively the same result in Experiment 6a [$b = 0.19$, $SE = 0.11$, $z = 1.73$, $p = .083$] but for Experiment 6b, revealed that those in the told-no-store were more likely to incorrectly identify a foil as "old" [$b = 0.31$, $SE = 0.11$, $z = 2.83$, $p = .005$]. For Experiment 6a, false alarm rate was lower in the trust trial than for either external store condition (trust: .03; told-no-store: .11; told-store: .12) [trust vs. told-no-store: $t(89) = 6.74$, $p < .001$, $d = 0.71$; trust vs. told -store: $t(89) = 6.65$, $p < .001$, $d = 0.70$]. The same was found for Experiment 6b (trust: .04; told-no-store: .13; told-store: .15) [trust vs. told-no-store: $t(120) = 8.21$, $p < .001$, $d = 0.75$; trust vs. told-store: $t(120) = 8.96$, $p < .001$, $d = 0.81$].

Sensitivity

There was no significant difference in sensitivity between the told-no-store and told-store instruction conditions in Experiment 6a [$t(89) = 1.41$, $p = .161$, $d = 0.15$], but there was in Experiment 6b [$t(120) = 3.04$, $p = .003$, $d = 0.28$]. For Experiment 6a, 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: $t(89) = 8.97$, $p < .001$, $d = 0.95$; trust vs. told -store: $t(89) = 9.76$, $p < .001$, $d = 1.03$]. Again, the same was found for Experiment 6b (trust: 3.13; told-no-store: 2.46; told-store: 2.22) [trust vs. told-no-store: $t(120) = 8.28$, $p < .001$, $d = 0.75$; trust vs. told-store: $t(120) = 11.14$, $p < .001$, $d = 1.01$].

Correct response time (ms)

There was no significant difference in response times for correct responses between conditions in Experiment 6a (told-no-store: 1512; told-store: 1534) [$t(89) = 1.42$, $p = .159$, $d = 0.15$] but there was a significant difference in Experiment 6b such that those in the told-no-store condition had slower correct response times (told-no-store: 1603; told-store: 1478) [$t(120) = 2.17$, $p = .032$. $d = 0.20$].

Process estimates

Targets. In Experiment 6a, a 2 (told-no-store vs. told-store) x 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 the told-store condition (told-no-store: .59; told-store: .55) [$F(1, 89) = 4.29$, $p = .041$, $\eta_G^2 < .01$]. There was also a main effect of process such that the estimates were lower for recollection than for familiarity (recollection: .51; familiarity: .62) [$F(1, 89) = 5.76$, $p = .019$, $\eta_G^2 = .04$]. The effect of external store did not differ between processes (recollection: .05; familiarity: .02) [$F(1, 89) = 1.05$, $p = .309$, $\eta_G^2 < .01$]. Paired samples t tests (not preregistered) on recollection and familiarity estimates revealed a significant effect of expecting support on recollection estimates [$t(84) = 2.86$, $p = .005$, $d = 0.31$] but not on familiarity estimates [$t(84) = 0.72$, $p = .382$, $d = 0.08$].

In Experiment 6b, a 2 (told-no-store vs. told-store) x 2 (recollection vs. familiarity) ANOVA demonstrated no main effect of condition on estimates (told-no-store: .57; told-store: .55) [$F(1, 115) = 2.12$, $p = .148$, $\eta_G^2 < .01$], and no main effect of process on estimates (recollection: .57; familiarity: .54) [$F(1, 115) = 0.39$, $p = .532$, $\eta_G^2 < .01$]. The effect of external store condition did not differ between processes (recollection: .03; familiarity: 0.01) [$F(1, 115) = 0.35$, $p = .555$, $\eta_G^2 < .01$]. Paired samples t tests (not preregistered) on estimates revealed a significant effect of expecting support on recollection [$t(115) = 2.17$, $p = .032$, $d = 0.20$] but not on familiarity [$t(115) = 0.24$, $p = .662$, $d = 0.04$].

Foils. In Experiment 6a, a 2 (told-no-store vs. told-store) x 2 (recollection vs. familiarity) ANOVA demonstrated no main effect of condition (told-no-store: .05; told-store: .06) [$F(1, 84) = 1.96$, $p = .166$, $\eta_G^2 < .01$]. There was a main effect of process such that the estimates were lower for recollection than for familiarity (recollection: .03; familiarity: .09) [$F(1, 84) = 27.45$, $p < .001$, $\eta_G^2 = .09$]. The effect of external store did not differ between processes (recollection: -.01; familiarity: -.01;) [$F(1, 84) < 0.01$, $p = .957$, $\eta_G^2 < .01$]. Paired samples t tests (not preregistered) on estimates revealed no effect of expecting support on recollection or familiarity estimates in Experiment 6a [recollection: $t(84) = 1.52$, $p = .132$, $d = .17$; familiarity: $t(84) = 0.88$, $p = .382$, $d = 0.10$].

In Experiment 6b, a 2 (told-no-store vs. told-store) x 2 (recollection vs. familiarity) ANOVA demonstrated no main effect of condition (told-no-store: .07; told-store: .08) [$F(1, 115) = 2.32$, $p = .130$, $\eta_G^2 < .01$]. There was a main effect of process such that the estimates were lower for recollection than for familiarity (recollection: .06; familiarity: .10) [$F(1, 115) = 8.33$, $p = .004$,

η_G^2 = .02]. External store condition did not differ between processes (recollection: -.01; familiarity: - .01) [$F(1, 115) = 0.06$, $p = .815$, $\eta_G^2 < .01$]. Paired samples t tests (not preregistered) on estimates revealed no effect of expecting support on recollection or familiarity estimates in Experiment 6b [recollection: $t(115) = 0.95$, $p = .343$, $d = 0.09$; familiarity: $t(115) = 1.20$, $p = .233$, $d = 0.11$].

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 6a [$t(88) = 6.32$, $p < .001$, $d = 0.67$] and 6b [$t(119) = 9.61$, $p < .001$, $d = 0.88$]. Many participants still reported a reliance mostly or only on internal memory in both Experiments 6a (53.3%) and 6b (52.9%) for the told-store condition. Table 11 below presents the breakdown for internal-external ratings for Experiments 6a and 6b.

Table 11. Experiments 6a and 6b: 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
E6a						
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%
E6b						
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%

Discussion

In Experiments 6a and 6b, we aimed to elucidate the effects of external support on recognition memory performance replicating the method of Experiment 5 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 6a but there was a significant cost to hit rate and sensitivity in Experiment 6b, the latter paralleling the results of Experiments 4 and 5. There was no effect of external store condition on false alarm rate across Experiments 6a or 6b, except for the mixed model analysis of Experiment

6b 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 6a suggested no significant effect of expecting support on correct response time (like Experiment 5), which was not the case in Experiment 4 (wherein expecting no support demonstrated significantly quicker correct response times) or in Experiment 6b (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 6a and 6b, we continued our examination from Experiment 5 of the potential influence of expecting external support on recollection and familiarity more specifically and with increased power. Neither Experiments 6a or 6b 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 5, 6a, and 6b 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.

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 6a and 6b 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 6a and 6b. When rating the chosen strategy from external (relying on list only) to internal (relying on own memory only), ratings in Experiments 6a and 6b 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 6a and 6b 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 7

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 7 to examine the extent to which study effort can explain this observed cost. Thus, in Experiment 7 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; Kelly & Risko, 2022b; Mazzoni & Cornoldi, 1993; Metcalfe & Kornell, 2005; Son & Metcalfe, 2000; Dunlosky & Thiede, 1998). Hence, in Experiment 7, 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 7 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 7 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 7 is available at osf.io/qjgct.

Method

Data from 118 participants 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 .80 ($\alpha = .05$, 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 5, 6a, and 6b and returned to the Old/New response options from Experiment 4. 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 to 100% in 5% increments. Word items were identical to those of Experiments 4, 5, and 6b, with item *lists* identical to that of Experiment 6b. As is the case throughout these experiments, lists appeared in all trial positions (one through five) 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 4 and 5).

Results

We had the same exclusion criteria for participants as in Experiments 6a and 6b (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 and Hayes (2016) 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, 1991; Stanislaw & Todorov, 1999).

The preregistration for Experiment 7 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 12 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 7 are available at osf.io/y36p2.

Table 12. Experiment 7: Means and Confidence intervals for Hit rate, False alarm rate, and Sensitivity across trials and conditions.

	Hit rate	False alarm rate	Sensitivity
Trust Trial 1	.93 [.92, .94]	.02 [.02, .03]	3.33 [3.23, 3.43]
Trust Trial 2	.89 [.88, .90]	.03 [.02, .04]	3.14 [3.03, 3.25]
Trust Trial 3	.91 [.90, .93]	.02 [.01, .03]	3.32 [3.21, 3.42]
Told-no-store	.81 [.79, .84]	.15 [.13, .18]	2.13 [1.98, 2.29]
Told-store	.75 [.73, .78]	.16 [.14, .19]	1.83 [1.69, 1.97]

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

Hit rate

There was a significant effect of condition on hit rate such that those in the told-no-store condition were more likely to correctly identify a target as "old" than those in the told-store condition [b = 0.47, SE = 0.08, z = 5.61, p < .001; $t(119)$ = 5.31, p < .001, d = 0.49]. Exploratory analyses revealed that trust trial hit rate was significantly higher than hit rate for either external store condition (trust: .91; told-no-store: .81; told-store: .75) [trust vs. told-no-store: trust vs. told-no-store: $t(117)$ = 7.95, p < .001, d = 0.73; trust vs. told-store: $t(117)$ = 11.14, p < .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 = .066; $t(117)$ = 1.57, p = .120, d = 0.14]. Trust trial false alarm rate was significantly lower than false alarm rate for either external store condition (trust:

.02; told-no-store: .15; told-store: .16) [trust vs. told-no-store: $t(117) = 9.95$, $p < .001$, $d = 0.92$; trust vs. told-store: $t(117) = 12.71$, $p < .001$, $d = 1.17$].

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 [$t(117) = 4.23$, $p < .001$, $d = 0.40$]. 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: $t(117) = 14.05$, $p < .001$, $d = 1.29$; trust vs. told -store: $t(117) = 18.83$, $p < .001$, $d = 1.73$].

Correct response time (ms)

There was no significant effect of external store condition on correct response time (told-no-store: 1319; told-store: 1269) [$b = 54.03$, $SE = 47.80$, $t = 1.15$, $p = .251$].

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: .09; told-store: .18) [$t(117) = 2.85$, $p = .005$, $d = 0.26$]. 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: $t(117) = 2.71$, $p = .008$, $d = 0.25$; told-store: $t(117) = 4.94$, $p < .001$, $d = 0.46$]

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, $p < .001$; $t(117) = 4.53$, $p < .001$, $d = 0.42$].

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) [$t(117) = 3.74$, $p < .001$, $d = 0.34$]. 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: 3342 ms vs. told-store: 3020 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) [$t(117) = 3.74$, $p < .001$, $d = 0.34$].

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 [$r(116) = .23$, $p = .011$] and sensitivity [$r(116) = .25$, $p = .007$], but was not significantly correlated with false alarm rate [$r(116) = -.09$, $p = .347$]. For the told-store condition, study time

was significantly correlated with false alarm rate [$r(116) = -.19$, $p = .035$], and sensitivity [$r(116) = .25$, $p = .006$] but not with hit rate [$r(116) = .10$, $p = .304$].

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*: .006, CI_{95} (-.001, .017)] and the effect of external store condition on hit rate [*total effect*: $b = .06$, $SE = .01$, $t = 5.04$, $p < .001$] remained robust when controlling for the mediator [*direct effect*: $b = .05$, $SE = .01$, $t = 4.24$, $p < .001$] (see Figure 10).

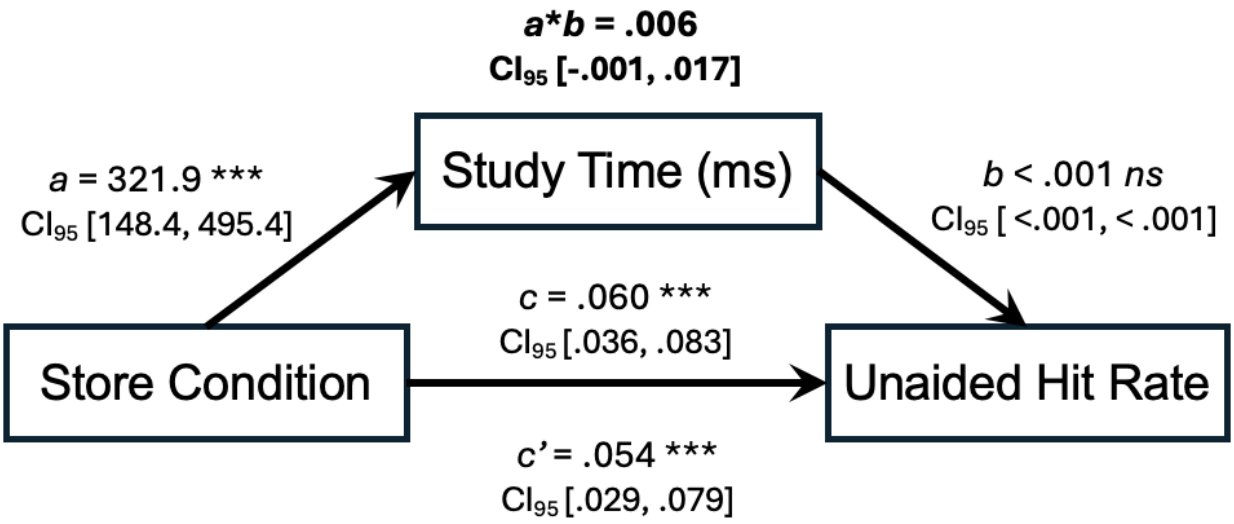


Figure 10. Experiment 7: Study time as a mediator on the relation between external store condition and unaided hit rate.

Note. * $p < .05$, ** $p < .01$, *** $p < .001$, *ns: nonsignificant.*

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*: .070, CI_{95} (.015, .142)] and the effect of external store condition on hit rate [*total effect*: b = .31, CI_{95} (.159, .451), SE = .07, t = 4.13, p < .001] remained robust when controlling for the mediator [*direct effect*: b = .24, CI_{95} (.081, .390), SE = .08, t = 3.02, p = .003] (see Figure 11).

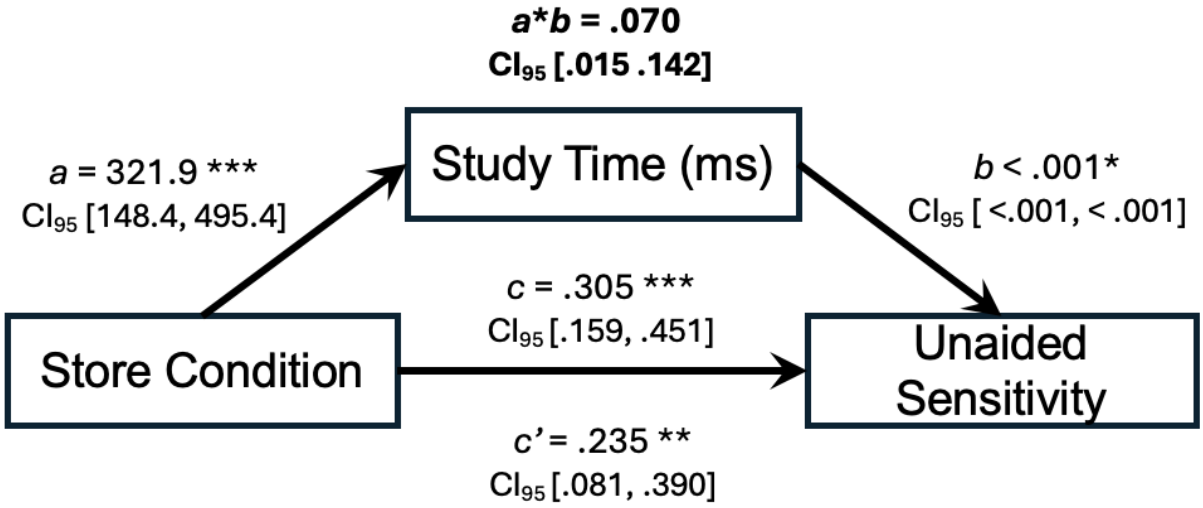


Figure 11. Experiment 7: Study time as a mediator on the relation between external store condition and sensitivity.

Note. * p < .05, ** p < .01, *** p < .001, *ns: nonsignificant.*

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*:

.008, CI_{95} (.001, .016)] and, again, the effect of external store condition on hit rate [*total effect*: *b* = .04, CI_{95} (.021, .057), *SE* = .01, *t* = 4.31, *p* < .001] remained robust when controlling for the mediator [*direct effect*: *b* = .03, CI_{95} (.013, .051), *SE* = .01, *t* = 3.27, *p* = .001] (see Figure 12).

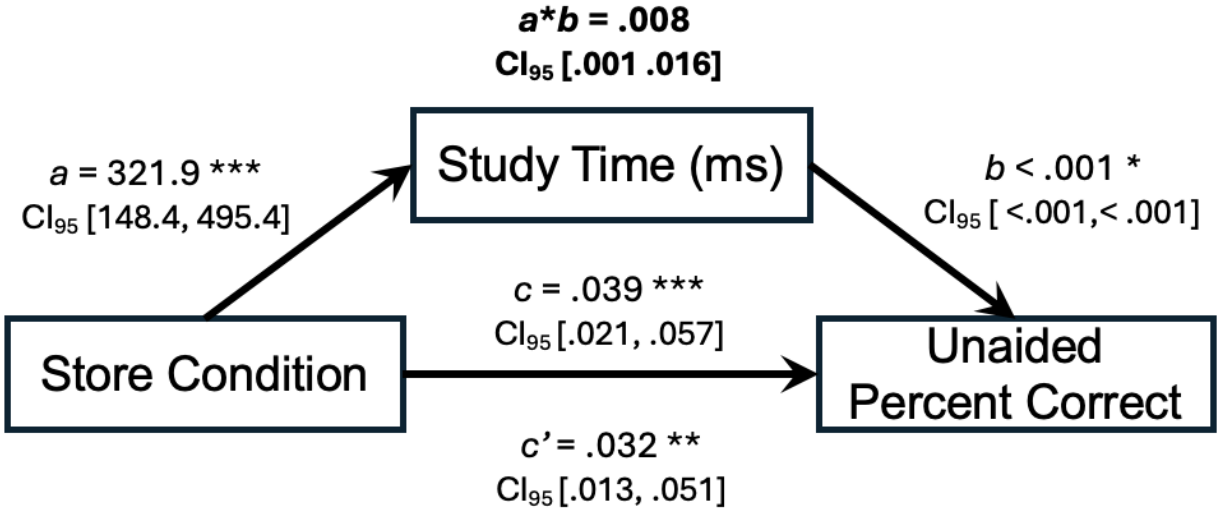


Figure 12. Experiment 7: Study time as a mediator on the relation between external store condition and percentage correct.

Note. * *p* < .05, ** *p* < .01, *** *p* < .001, *ns: nonsignificant.*

Predicted percentage correct

There was no significant effect of condition on predictions, *t*(115) = 1.08, *p* = .281, *d* = 0.10 (told-no-store: .52; told store: .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* < .001, *d* = 0.55 (told-no-store: .56; told store: .48). In contrast, those in the told-store-first order condition predicted better memory upon

losing expected support compared to expecting no support $t(57) = 3.58$, $p = .001$, $d = 0.47$ (told-no-store: .48; told store: .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 13 below. 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.

Table 13. Experiment 7: 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	.32 ***	.18 *ns*	-.34 ***	.35 ***
Told store	.39 ***	.25 **	-.28 **	.37 ***
Told no store	.33 ***	.15 *ns*	-.37 ***	.32 ***
Statistical comparison between external store conditions	*ns*	*ns*	*ns*	*ns*

Note. * $p < .05$, ** $p < .01$, *** $p < .001$, *ns: nonsignificant.*

Discussion

In Experiment 7, 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 7 provided clear evidence of a cost to unaided recognition memory from expecting external support. Given this cost of external memory support expectations to recognition memory, one of the main aims in Experiment 7 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 7, 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, 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 7, we did find evidence that individuals are using the external support during the trust trials. That is, like in Experiments 6a and 6b, it was clear in Experiment 7 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, 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 5, 6a, and 6b).

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).

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 7 of the current work and earlier work by Kelly and Risko (2022b). In Experiment 7 of the current work, correlations between study time and performance indices were r = .23 for hit rate, r = -.08 false alarm rate (not significant) and r = .25 for sensitivity in the told-no-store condition. For the told-store condition, correlations between study time and performance indices were r = .10 for hit rate (not significant), r = -.19 for false alarm rate and r = .25 for sensitivity. Although not reported, using the recall data of Kelly and Risko (2022b) posted to OSF, the correlations between study time and performance were r = .28 (E1a) and r = .38 (E1b) for the told-no-store group and r = .41 (E1a) and r = .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 et al., 2019a; 2019b; Kelly & Risko 2022a; 2022b, Lu et al., Park et al., 2022). Furthermore, throughout Experiments 4 to 6b, 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, individual's 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. We focus on gathering predictions participants make on a trial where they had just found out that they had lost their external store. Predictions of performance when individuals expect and do receive support would provide insight into how beneficial individuals view the use of the external store in the present task.

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.

Chapter 5

Concluding Remarks

Many of our memory goals would not be as easily attainable without the incorporation of external memory supports. Within this dissertation, I have extended the unit of analysis beyond the internal memory context of traditional memory research and have provided insights into how the use of external memory supports affect memory and metamemory. It is clear from the current experiments that memory performance can be *enhanced* by the *availability* of external supports across different levels of support and different memory tasks, consistent with previous work (Ball et al., 2022; Boldt & Gilbert, 2019; Fellers et al., 2023; Gilbert, 2015; Gilbert et al., 2023; Grinschgl et al., 2021; Kelly & Risko, 2019a; 2019b; 2022a; 2022b; Lu et al., 2020; 2022; Morrison & Richmond, 2020; Murphy, 2023; Park et al., 2022; Peper et al., 2023; Richmond et al., 2023; Risko & Dunn, 2015; Scarampi & Gilbert, 2021).

Of critical interest, memory performance is *reduced* by the *unexpected unavailability* of such supports across different levels of support and different memory tasks. This finding is consistent with the idea that individuals are using the external support more as a replacement for their internal memory—a *substitutive* use of external memory support—as opposed to as an external backup for internal memory—a *duplicative* use (Risko et al., 2024). A major focus of the dissertation was testing the degree to which this reduced memory performance from expecting external support could be explained with reduced study efforts: the *study effort hypothesis*. The study effort explanation for the cost to memory was based on prior work of mine suggesting that phenomena thought to be *more* dependent on intentional mnemonic study efforts exhibited a larger

cost compared to phenomena thought to be *less dependent* on such intentional study efforts (Kelly & Risko, 2019a; 2019b; 2022b; see also Lu et al., 2022 and Park et al., 2022).

In Chapter 2, I presented a direct test of the study effort hypothesis as an explanation of the cost to unaided memory of expecting external support. Across two preregistered experiments, participants controlled their own study time and reported whether they used any intentional strategies (Ariel et al., 2015; Dunlosky & Thiede, 1998; Dunlosky & Kane, 2007; Unsworth, Miller, & Robison, 2019; Unsworth, 2016). Here, participants who were told not to expect external support invested more study time on the critical trial compared to the prior trust trial, which was not the case for those told that they *could* expect external support. Moreover, participants who were told not to expect external support were significantly more likely to report use of a strategy to facilitate remembering than were those told that they could expect support. These findings support the idea that those told not to expect the external support invested greater study efforts on the critical trial compared to those told that they could expect support. Hence, these results are consistent with the study effort hypothesis. Furthermore, change in study time and strategy use significantly mediated the relation between external support condition and unaided recall performance on the critical trial, however, the direct effect of external support condition on recall performance remained significant. Thus, the mediation analyses provided further support for the study effort hypothesis, but also provided support for the idea that there could be other factors contributing to the cost of expecting external support.

In Chapter 3, I extended the work of Chapter 2 by adding an intermediate level of external memory support and eliciting expectations of memory performance through predictions.

Specifically, the external support conditions included full support (study APPLE, expect APPLE as support) and no support (study APPLE, expect *nothing* as support) as in prior work and Chapter 2, as well as a novel partial support condition (study APPLE, expect AP or APP as support).

As in earlier work, there was a clear cost of unexpectedly losing full support compared to expecting no support, a cost that was partially mediated by study effort (as indexed by global study effort rating and strategy report but not by study time), thus, the results were generally consistent with a study effort hypothesis. With respect to the addition of the partial support condition, the presence of partial support yielded generally better memory performance compared to expecting and receiving no support, although the benefit was not as great as receiving full support. Critically, when participants *lost* partial support, there was a cost to unaided memory performance, but this cost was smaller than when participants had expected full support. Furthermore, there were no differences across the study effort indices in the condition where participants received only two letters of support and the no-support condition despite a clear cost to losing this support compared to the no-support condition. This suggests the cost to memory in the partial-2 condition was not due to reduced study efforts—hence, a novel cost of external memory support not consistent with the study effort hypothesis.

In Chapter 3, predicted memory performance demonstrated a clear cost of expecting external memory support indicating that participants expected to perform worse overall if an external support was unexpectedly lost. In addition, across participants, there were positive correlations between predicted and actual performance, suggesting some metacognitive insights into how they would perform with and without the external support. Interestingly, in Chapter 3,

predictions of unaided performance did not differ with degree of expected external support (full vs. partial) despite observed unaided performance showing a graded cost as a function of support level (i.e., losing full support demonstrated a greater cost compared to losing partial support). That is, participants seemed to estimate a generic cost of losing expected support. Thus, while individuals seem generally aware of the negative impact of losing an expected memory store, this metacognitive insight appears to have blind spots.

In the final experiments of this dissertation (Chapter 4), I built upon the previous chapters by testing whether the cost of external memory support occurred in recognition memory. This work was motivated by the potential for the different levels of environmental support offered by recognition compared to recall to differentially influence the cost to unaided memory, and predictions made by the study effort hypothesis with respect to the effect of study effort on *recollection* in particular (Basden & Basden, 1996; Jacoby et al., 1997; Mangels et al., 2001; Ochsner, 2000; Ozubko et al., 2012; Yonelinas, 2002; Yonelinas & Jacoby, 1995). Across the five experiments, I found evidence of a general cost to recognition memory performance of expecting external support. Experiments 6a and 6b suggested that the cost to recognition memory was driven by reduced recollection (versus familiarity; this finding would benefit from additional methods of examining recollection vs. familiarity, e.g., ROC curve analysis; Ozubko et al., 2012; Yonelinas, 1994; 1997). In examining the effects of expecting support on study effort, I found a reduction in study time during study that partially mediated the relation between expecting support and memory performance indices. Hence, these results are, like earlier chapters, supportive of the idea that (i)

study effort reductions can partially explain the cost observed to unaided memory of expecting external support and (ii) more than study effort is contributing to these costs to unaided memory.

In examining the effects of external memory support on aspects of metacognition, performance predictions were positively correlated with actual performance in both external store conditions, suggesting a general awareness of unaided recognition memory ability regardless of expecting external support. Performance predictions were mixed in that some participants (those who experienced the told-no-store condition before the told-store condition) predicted a cost while others predicted a benefit (those who experienced the told-store condition prior to the told-no-store condition). More work is needed to understand the clearly complex consequences of losing support on expectations of memory. These results contrast to some degree with the previous findings in recall wherein individuals predict a clear robust cost of expecting support (Chapter 3; Kelly & Risko, *submitted*; see also Lu et al., 2022; Park et al., 2022).

In Chapter 4, exploratory analysis of performance on trust trials compared to the critical trials found that performance was similar (especially in Experiments 4 and 5 of Chapter 4), suggesting that participants may not have been relying much on the external support during trust trials—that is, when it was available. This was further supported by analyzing responses to two post-task prompts meant to elicit ratings of the degree to which participants intended to rely on their saved list (external memory) versus own memory (internal memory). Regardless of external memory store condition, most participants across the experiments of Chapter 4 (~50% to ~ 70% depending on experiment) reported that they intended to *rely mostly* or *completely* on their own internal memory. This contrasts with earlier work by Kelly and Risko (2022a) where participants

were asked the same question but in a recall test context. There, only 13% of the participants in the told-store condition reported intending to rely mostly or completely on their own internal memory. The results also appear to contrast with prior work suggesting a potential preference to offload in a short-term recall memory task: Risko and Dunn (2015) found that individuals frequently elected to save information externally for a set size of 2 items despite no observable benefit to performance of doing so (similarly, see set size of 4 items from Morrison & Richmond, 2020). Overall, the results suggest that, for recognition memory, people seemed less likely to rely on external support, contrasting with findings in a recall context and highlighting the idea that the type of remembering required could greatly affect the use of an external memory support. This could be due to differing retrieval demands between recall and recognition (e.g., environmental support), and/or differing expectations of the cost of external support between recall and recognition contexts, as hinted at by the clear costs predicted for recall versus the lack of clear cost predicted for recognition.

Explaining the cost of expecting external memory support

Throughout the current work, study effort has consistently been able to account for the cost of expecting external support—at least partially. This was found across multiple study effort indices (Chapters 2-4; Kelly & Risko, 2022b; *submitted*; Kelly et al., *submitted*), varied levels of external support (Chapter 3; though not Partial-2 in Experiment 3; Kelly & Risko, *submitted*), and extended to recognition memory (Chapter 4; Kelly et al., *submitted*). Taken together, the findings of this dissertation put the study effort hypothesis on strong footing as an explanation for the memory cost. Another recurring finding was that the study effort hypothesis did not *fully* account

for the cost observed to unaided memory when expecting external memory support. That is, as it stands, study effort *partially* explains the cost to memory of expected external support. There are at least two ways to interpret this result. One interpretation is that there could be additional contributions to the cost that are not based on study effort. Another interpretation is that the effort indices implemented throughout the dissertation did not completely capture "study effort" and, hence, study effort could account for the entire cost if it were captured "completely." I discuss each interpretation more below.

With respect to the idea that there are additional contributions to the cost of expecting external support, there are multiple feasible explanations that could be considered either in isolation or in some combination with one another. One potential explanation brought up in Chapter 2 was that the unexpectedly losing external support, as a violation of expectations, could elicit strong affect such as surprise or alarm that could compromise one's ability to recover the studied information (e.g., Wessel et al., 2016). Another potential contribution to the cost is that expecting support and learning of its unavailability reduce efforts at retrieval. For example, individuals may invest less effort at retrieval because they are anticipating that their unaided memory will be poor. Indeed, as noted above, individuals expect a clear cost of losing an expected support, at least in a recall context. This belief could lead them to reduce retrieval efforts, thereby reducing memory performance. Testing this retrieval effort explanation would be a productive area for future work.

The partial support data in Chapter 3 suggest another potential additional contribution to the cost of expecting an external support on unaided memory. In Chapter 3, I offered a kind of

transfer-appropriate processing explanation (e.g., Blaxton, 1989; Morris et al., 1977; Tulving & Thomson, 1973) of the cost observed in the partial-2 condition despite no apparent difference in study effort. I speculated that this memory cost could be due to study efforts being invested in a strategy reliant upon the presence of partial support at test that did not transfer to the test without support. This would explain the superior memory performance of partial-2 in the aided trials (when they received the support) compared to the no-support condition and the cost in the unexpectedly unaided condition despite equivalent effort. Extending this to the full support condition, this transfer-appropriate processing type of account would suggest the effectiveness of the study strategy in the full-condition was also highly dependent upon receiving the external support. Specifically, this account would predict a cost to unaided memory given that the strategy at study required the support to be available at test, even if people invested equal effort. This would also yield the observed results of higher performance when the full support was available on the aided trials compared to having no support, but worse performance when the full support was unexpectedly unavailable.

As stated above, the unexplained part of the cost to memory of expecting support could also reflect the present indices of study effort not being able to *fully* capture "study effort." The effort indices used throughout were varied in that they included behavioral (e.g., study time) and self-report (global ratings of effort and strategy) measures, however, it of course remains possible that effort at study is not completely captured by such indices. Additional measures of study effort could include behavioral and self-report indices like in the current work, as well as more comprehensive effort scales (e.g., NASA Task load Index; Hart, 2006; Hart & Staveland, 1988),

psychophysiological indices such as pupillometry (da Silva Castanheira et al., 2021; Laeng & Alnaes, 2019), or neural correlates (e.g., Skvortsova et al., 2014). Increasing the range of study effort indices in future work could help determine whether the current indices were not capturing study effort *fully,* whether such additional indices of effort are needed to fully capture "study effort" and, if so, whether these indices then fully account for the cost to memory of expecting external support.

Despite the cost of expecting support not being entirely explained by the study effort hypothesis, the current collection of work has reliably provided evidence that reduced study efforts can account for at least some of the cost of expecting external support and, importantly, has done so across a range of support contexts and remembering conditions. Thus, the research in this dissertation supports the study effort hypothesis as one valid explanation of the resulting memory cost.

Effects of expecting external memory support on metacognition

In addition to effects on memory, the current work also provided insights into the influence of expecting external support on aspects of metacognition (i.e., predictions of performance and the accuracy of those predictions). At the aggregate level, people generally demonstrated positive correlations between predictions and performance when the support was available (i.e., trust trials), and when support was expected or not expected but unavailable (critical trials). Moreover, they generally seemed aware that there is a cost to expecting support in free recall. However, results also revealed potential gaps in metacognitive awareness. Individuals did not predict the graded cost to memory performance as a function of degree of expected support (Chapter 3; Kelly &

Risko, *submitted*) and did not, in *general*, predict the significant cost to recognition memory (Chapter 4; Kelly et al., *submitted*). That is, across Chapters 3 and 4, there was evidence for a common inability for predictions of performance to reflect subtler patterns in observed performance.

Despite some metacognitive issues raised above, those who unexpectedly lost their support were still able to make predictions of unaided performance that positively correlated with unaided performance, suggesting a general awareness of unaided performance despite the loss of the expected support. That said, in this dissertation, the experiments of Chapter 3 found that individuals expecting full support demonstrated significantly lower correlations between predicted and observed performance compared to those expecting no support. In contrast, the findings of Chapter 4 did not indicate any effects of expecting support on the correlation between predicted and observed performance. Of course, the work of Chapter 4 was in the context of recognition memory and deviated in a number of other ways compared to prior research focused on free recall. The suggestion that expecting support (i.e., full support, in the case of Chapter 3) reduces metacognitive accuracy compared to no support aligns with other related work supporting the idea that expecting external support reduces metacognitive accuracy for unaided performance (e.g., Lu et al., 2022; Park et al., 2022). Further work is needed to establish how general this negative effect of expecting external support is on metacognitive accuracy.

General Conclusions

Through a systematic examination of the cost of expecting external memory support, this dissertation paints a nuanced picture of a common but largely unexplored aspect of human

remembering. When available, external memory supports lead to benefits in memory performance relative to internal memory alone, but when unexpectedly *unavailable*, they lead to costs in memory performance compared to expecting no support. These costs to memory can be explained, at least in part, by a study effort hypothesis. Results also suggest, however, that some of the memory cost is not easily explainable with reduced study efforts. In examining the influence of expecting support on metacognition, individuals have an overall awareness of their performance and that the cost to memory is moderated by the type of test (recall vs. recognition). At the same time, metacognitive awareness of performance in these more distributed memory contexts shows limitations. Thus, across a variety of conditions and contexts, this dissertation has provided a number of theoretical insights regarding our use of these external supports and has established the study effort hypothesis as a partial explanation for the memory cost of expecting external support.

References

Ackerman, R. (2014). The diminishing criterion model for metacognitive regulation of time investment. *Journal of Experimental Psychology: General*, *143*(3), 1349.

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.

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

Ball, H., Peper, P., Alakbarova, D., Brewer, G., & Gilbert, S. J. (2022). Individual differences in working memory capacity predict benefits to memory from intention offloading. *Memory*, *30*(2), 77-91.

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.

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

Blaxton, T. A. (1989). Investigating dissociations among memory measures: Support for a transfer-appropriate processing framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(4), 657.

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

Boldt, A., & Gilbert, S. J. (2019). Confidence guides spontaneous cognitive offloading. *Cognitive Research: Principles and Implications*, *4*(1), 1-16.

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

Brand, A., Allen, L., Altman, M., Hlava, M., & Scott, J. (2015). Beyond authorship: Attribution, contribution, collaboration, and credit. *Learned Publishing*, *28*(2).

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

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

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

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.

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

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

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

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

Craik, F. I. (2022). 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.

Craik, F. I. (1990). Changes in memory with normal aging: a functional view. *Advances in Neurology*, *51*(210-205).

Craik, F. I. (1983). On the transfer of information from temporary to permanent memory. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *302*, 341–359. doi:10.1098/ rstb.1983.0059

Craik, F. I., & Bialystok, E. (2006). Cognition through the lifespan: mechanisms of change. *Trends in Cognitive Sciences*, *10*(3), 131-138.

Danckert, S. L., & Craik, F. I. (2013). Does aging affect recall more than recognition memory? *Psychology and Aging*, *28*(4), 902.

da Silva Castanheira, K., LoParco, S., & Otto, A. R. (2021). Task-evoked pupillary responses track effort exertion: Evidence from task-switching. *Cognitive, Affective, & Behavioral Neuroscience, 21*, 592-606.

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.

Diedenhofen, B. & Musch, J. (2015). cocor: A Comprehensive Solution for the Statistical Comparison of Correlations. *PLoS ONE, 10*(4): e0121945.

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., Hunt, R., & Clark, E. (2000). Is perceptual salience needed in explanations of the isolation effect? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 649–657.

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

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

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.

Eeckhout, P. (2012). Inca Storage and Accounting Facilities at Pachacamac. *Andean Past*, *10*(1), 12.

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

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

Fabiani, M., & Donchin, E. (1995). Encoding processes and memory organization: A model of the Von Restorff effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 224–240.

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.

Fellers, C., Miyatsu, T., & Storm, B. C. (2023). Remembering what to do when the time comes: The effects of offloading in a complex prospective memory task. *Journal of Experimental Psychology: Applied*, *29*(3), 631.

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

Fischler, I., Rundus, D., & Atkinson, R. C. (1970). Effects of overt rehearsal processes on free recall. *Psychonomic Science, 19,* 249-250.

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

Gilbert, S. J. (2015). Strategic use of reminders: Influence of both domain-general and task-specific metacognitive confidence, independent of objective memory ability. *Consciousness and Cognition*, *33*, 245-260.

Gilbert, S. J., Boldt, A., Sachdeva, C., Scarampi, C., & Tsai, P. C. (2023). Outsourcing memory to external tools: A review of 'intention offloading'. *Psychonomic Bulletin & Review*, *30*(1), 60-76.

Glanzer, M., & Cunitz, A. R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior, 5,* 351–360.

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

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

Händel, M., de Bruin, A. B., & Dresel, M. (2020). Individual differences in local and global metacognitive judgments. *Metacognition and Learning*, *15*(1), 51-75.

Hart, S. G. (2006, October). NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 50, No. 9, pp. 904-908). Sage CA: Los Angeles, CA: Sage publications.

Hart, S. G. & Staveland, L. E. (1988) Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati (Eds.) Human Mental Workload. Amsterdam: North Holland Press.

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

Howell, D. C. (2015). *Statistical methods for psychology*. Cengage Learning.

Hsee, C. K., & Zhang, J. (2010). General evaluability theory. *Perspectives on Psychological Science*, *5*(4), 343-355.

Hutchins, E. (2000). Distributed cognition. *International Encyclopedia of the Social and Behavioral Sciences. Elsevier Science, 138*, 1-10.

Inzlicht, M., Shenhav, A., & Olivola, C. Y. (2018). The effort paradox: Effort is both costly and valued. *Trends in Cognitive Sciences*, *22*(4), 337-349.

Laeng, B., & Alnaes, D. (2019). Pupillometry. *Eye Movement Research: An Introduction to its Scientific Foundations and Applications*, 449-502.

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. (1998). Invariance in automatic influences of memory: Toward a user's guide for the process-dissociation procedure. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24(1).*

Jacoby, L. L., Toth, J, P., & Yonelinas, A. P. (1993). Separating conscious and unconscious influences on memory: measuring recollection. *Journal of Experimental Psychology: General.* 122(2), 139 – 154.

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). Hillsdale: 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. *Submitted/Under Review.*

Kelly, M. O., Karimjee, B., Pereira, A. E., Lu, X., & Risko, E. F. (2024). Does expecting external memory support cost recognition memory? *Submitted/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.

Köhler, W., & von Restorff, H. (1995). An analysis of the processes in the trace field. (Dorsch, A., Trans.) Retrieved from http://www.utsa.edu/mind/von Restorff translation.htm (Original work published 1933)

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.

Lakens D., Caldwell A. R. (2021). Simulation-Based Power Analysis for Factorial Analysis of Variance Designs. *Advances in Methods and Practices in Psychological Science*, *4*(1).

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

Lindenberger, U., & Mayr, U. (2014). Cognitive aging: is there a dark side to environmental *support? Trends in Cognitive Sciences, 18*(1), 7-15.

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.

Luo, L., & Craik, F. I. (2008). Aging and memory: A cognitive approach. *The Canadian Journal of Psychiatry*, *53*(6), 346-353.

MacLeod, C. M. (2020). Zeigarnik and von Restorff: The memory effects and the stories behind them. *Memory & Cognition*, *48*, 1073-1088.

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.

Marsh, E. J., & Rajaram, S. (2019). The Digital Expansion of the Mind: Implications of Internet Usage for Memory and Cognition. *Journal of Applied Research in Memory and Cognition*.

Mazancieux, A., Pereira, M., Faivre, N., Mamassian, P., Moulin, C. J., & Souchay, C. (2023). Towards a common conceptual space for metacognition in perception and memory. *Nature Reviews Psychology*, *2*(12), 751-766.

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

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

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

Migo, E. M., Roper, A., Montaldi, D., & Mayes, A. R. (2010). British English norms for the spontaneous completion of three-letter word stems. *Behavior Research Methods*, *42*(2), 470-473.

Montoya, A. K., & Hayes, A. F. (2017). Two-condition within-participant statistical mediation analysis: A path-analytic framework. *Psychological Methods*, *22*(1), 6.

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

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, *16*(5), 519-533.

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. (2023). Strategic offloading: How the value of to-be-remembered information influences offloading decision-making. *Applied Cognitive Psychology*, *37*(4), 749–767.

Nestojko, J. F., Finley, J. R., & Roediger III, H. L. (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.

Peper, P., Alakbarova, D., & Ball, B. H. (2023). Benefits from prospective memory offloading depend on memory load and reminder type. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *49*(4), 590.

Pereira, A., Kelly, M.O., Lu, X., & Risko, E.F. (2022). On our susceptibility to external memory store manipulation: Examining the influence of perceived reliability and expected access to an external store. *Memory, 30*(4), 412-428.

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

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

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*.

Rouault, M., Dayan, P., & Fleming, S. M. (2019). Forming global estimates of self-performance from local confidence. *Nature Communications*, *10*(1), 1141.

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

Runge, Y., Frings, C., & Tempel, T. (2019). Saving-enhanced performance: Saving items after study boosts performance in subsequent cognitively demanding tasks. *Memory*, *27*(10), 1462-1467.

Runge, Y., Frings, C., & Tempel, T. (2021). Specifying the mechanisms behind benefits of saving-enhanced memory. *Psychological Research*, *85*, 1633-1644.

Sachdeva, C., & Gilbert, S. J. (2020). Excessive use of reminders: Metacognition and effort-minimisation in cognitive offloading. *Consciousness and Cognition, 85*, 103024.

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.

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

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., Botvinick, M. M., & Cohen, J. D. (2013). The expected value of control: an integrative theory of anterior cingulate cortex function. *Neuron, 79*(2), 217-240.

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.

Shiffrin, R. M., & Atkinson, R. C. (1969). Storage and retrieval processes in long-term memory. *Psychological Review, 76*(2), 179.

Skvortsova, V., Palminteri, S., & Pessiglione, M. (2014). Learning to minimize efforts versus maximizing rewards: computational principles and neural correlates. *Journal of Neuroscience, 34*(47), 15621-15630.

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

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.

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

Storm, B. C., Yamashiro, J. K., & Hamilton, K. A. (2022). Rethinking cognition in a digital environment. *Applied Cognitive Psychology*, *36*(1).

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.

Tan, L., & Ward, G. (2008). Rehearsal in immediate serial recall. *Psychonomic Bulletin & Review, 15*(3), 535-542.

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

Tsai, P. C., Sachdeva, C., Gilbert, S. J., & Scarampi, C. (2023). An investigation of the saving-enhanced memory effect: The role of test order and list saving. *Applied Cognitive Psychology*, *37*(4), 736-748.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*(5), 352.

Undorf, M., & Ackerman, R. (2017). The puzzle of study time allocation for the most challenging items. *Psychonomic Bulletin & Review, 24*, 2003-2011.

Unsworth, N. (2016). Working memory capacity and recall from long-term memory: Examining the influences of encoding strategies, study time allocation, search efficiency, and

monitoring abilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42,* 50–61.

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

Urton, G. (2008). The Inca khipu: knotted-cord record keeping in the Andes. In *The Handbook of South American Archaeology* (pp. 831-843). New York, NY: Springer New York.

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

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

Van Kessel, K., Ashburner, M., & Risko, E. F. (2024). Dissociations between data-driven and goal-driven effort reports: Performance, metacognition, and affect. *Quarterly Journal of Experimental Psychology*, *77*(5), 983-993.

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

von Restorff, H. (1933). Über die Wirkung von Bereichsbildung im Spurenfeld [On the effects of the formation of a structure in the trace field]. *Psychologische Forschung, 18,* 299–342.

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

Woodward Jr, A. E., Bjork, R. A., & Jongeward Jr, R. H. (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. (1994). Receiver-operating characteristics in recognition memory: evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(6), 1341.

Yonelinas, A. P. (1997). Recognition memory ROCs for item and associative information: The contribution of recollection and familiarity. *Memory & Cognition*, *25*, 747-763.

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. doi: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.