

2015

Studying on Borrowed Time: How Does Testing Impair New Learning?

Sara D. Davis

Iowa State University, sddavis@iastate.edu

Jason C.K. Chan

Iowa State University, ckchan@iastate.edu

Follow this and additional works at: http://lib.dr.iastate.edu/psychology_pubs

 Part of the [Cognitive Psychology Commons](#), [Educational Assessment, Evaluation, and Research Commons](#), and the [School Psychology Commons](#)

The complete bibliographic information for this item can be found at http://lib.dr.iastate.edu/psychology_pubs/28. For information on how to cite this item, please visit <http://lib.dr.iastate.edu/howtocite.html>.

This Article is brought to you for free and open access by the Psychology at Iowa State University Digital Repository. It has been accepted for inclusion in Psychology Publications by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

Studying on Borrowed Time: How Does Testing Impair New Learning?

Abstract

Retrieving studied materials often enhances subsequent learning of new materials (Pastötter & Bäuml, 2014). However, retrieval has also been shown to impair new learning (Finn & Roediger, 2013). Here we attempted to determine when retrieval enhances and when it impairs new learning. We argue that testing impairs new learning when one intermixes testing with new learning, which biases participants to relearn the tested information at the expense of the new information. We refer to this as the borrowed time hypothesis. Consistent with this idea, we reduced or eliminated test-impaired new learning by discouraging time borrowing. Moreover, testing enhanced new learning only when the test trials and new learning trials were presented in separate blocks. These results suggest that test-impaired new learning and test-enhanced new learning are based on different underlying mechanisms, and that they are not simply the flipped side of the same coin.

Keywords

interpolated testing, test-potentiated learning, self-regulated learning, testing effect, test-enhanced new learning

Disciplines

Cognitive Psychology | Educational Assessment, Evaluation, and Research | School Psychology

Comments

This is a manuscript of an article from *Journal of Experimental Psychology: Learning, Memory, and Cognition* (2015) in press. This article may not exactly replicate the final version published in the APA journal. It is not the copy of record.

Studying on Borrowed Time: How Does Testing Impair New Learning?

Sara D. Davis and Jason C.K. Chan

Iowa State University

In press: Journal of Experimental Psychology: Learning, Memory, and Cognition

Address correspondence to:

Sara D. Davis

sddavis@iastate.edu

Department of Psychology
W112 Lagomarcino Hall
Iowa State University
Ames, IA 50014
USA

Abstract

Retrieving studied materials often enhances subsequent learning of new materials (Pastötter & Bäuml, 2014). However, retrieval has also been shown to impair new learning (Finn & Roediger, 2013). Here we attempted to determine when retrieval enhances and when it impairs new learning. We argue that testing impairs new learning when one intermixes testing with new learning, which biases participants to relearn the tested information at the expense of the new information. We refer to this as the borrowed time hypothesis. Consistent with this idea, we reduced or eliminated test-impaired new learning by discouraging time borrowing. Moreover, testing enhanced new learning only when the test trials and new learning trials were presented in separate blocks. These results suggest that test-impaired new learning and test-enhanced new learning are based on different underlying mechanisms, and that they are not simply the flipped side of the same coin.

Keywords: interpolated testing, test-potentiated learning, self-regulated learning, testing effect, test-enhanced new learning

Studying on Borrowed Time: How Does Testing Impair New Learning?

Extensive empirical evidence has demonstrated that retrieval enhances the retention of studied materials. Relatively less attention has been paid to the facilitating effects of retrieval on future learning. This test-potentiated learning effect (Izawa, 1970) has been examined in both the laboratory and applied contexts, and provides several benefits to learning. For example, interpolated testing can potentiate subsequent learning by helping students sustain attention during a lecture, increasing note-taking, reducing proactive interference (Szpunar, Khan, & Schacter, 2013), and facilitating source discrimination (Chan & McDermott, 2007; Chan, Wilford, & Hughes, 2012).

“Test-potentiated learning” is often used to describe the phenomenon whereby retrieval practice of previously-studied information can enhance future relearning of that information (Izawa, 1970, Arnold & McDermott, 2013). In this report, we focus on a different potentiating effect of retrieval – that taking a test on a set of studied material can enhance subsequent learning of *new* material. To distinguish this phenomenon from test-potentiated (re)learning, we will refer to it as *test-enhanced new learning*. Test-enhanced new learning is typically investigated in the laboratory using multi-list learning paradigms. In an early demonstration of this effect, Tulving and Watkins (1974) had participants learn two lists of paired associates. After learning List 1 (A-B pairs), participants either performed an immediate test or a control, unrelated task. Afterwards, participants learned List 2 (A-C pairs) before their memory was examined in a final test. The most important result for our purposes is that immediate testing of List 1 facilitated learning of the List 2 pairs relative to the control condition. However, in stark contrast to the increasingly prevalent finding of test-enhanced new learning (for a review, see

Pastötter & Bäuml, 2014), Finn and Roediger (2013) reported evidence that interpolated testing, relative to restudying, can *impair* new learning. The purpose of the present study is to identify the mechanism that underlies this interesting reversal. Gaining such an understanding is important from both theoretical and applied perspectives. Theoretically, being able to delineate the conditions under which testing enhances or impairs new learning can shed light on the mechanisms by which testing influences future learning. Practically, such an understanding would aid in the development of optimal teaching and study techniques.

Evidence for Test-Enhanced New Learning

A substantial body of research has shown that testing can enhance future learning of new materials. Szpunar, McDermott, and Roediger (2008) found evidence for test-enhanced new learning using lists of single words. Participants studied five lists of words, and then completed math problems, restudied the material, or free-recalled the items between each list. To assess the effect of interpolated testing on later learning, all participants were tested on the fifth list immediately after its presentation. Szpunar et al. found that participants in the interpolated testing condition remembered more words from the fifth list than participants in the interpolated restudy or interpolated math conditions. Experiments with similar methodologies have demonstrated this same basic finding across a variety of materials (Bäuml & Kliegl, 2013, for unrelated words; Pastötter, Weber, & Bäuml, 2013, for pictures; Weinstein, McDermott, & Szpunar, 2011, for face-name pairs; Szpunar et al., 2013, for video lectures).

The impact of testing on new learning has also been examined using more complex materials like those typically employed in the eyewitness memory literature. A

common component of eyewitness testimony is repeated retrieval (e.g., to police, lawyers, etc., Chan & LaPaglia, 2011), and recent research has examined how earlier retrieval of a witnessed event influences one's susceptibility to later misinformation. Intuitively, one might hypothesize that testing could insulate eyewitnesses from the misinformation effect by strengthening memory for the witnessed event. However, Chan and colleagues (Chan & LaPaglia, 2011, 2013; Chan, Thomas, & Bulevich, 2009; Chan, Wilford, & Hughes, 2012; Wilford, Chan & Tuhn, 2013) have shown that the opposite can often occur. Participants who are tested prior to misinformation exposure can recall misinformation at a higher rate than those who are not tested, (cf., Huff, Davis & Meade, 2013; LaPaglia & Chan, 2012; and Pansky & Tenenboim, 2011). While counterintuitive, these data are easily reconciled with test-enhanced new learning effects, where misinformation may be construed as a special case of new learning similar to learning a new list of items (e.g., Gordon & Thomas, 2014; Chan, Thomas & Bulevich, 2009). Therefore, initial testing can have a similar effect on the learning of misinformation as it would in a typical multi-list learning paradigm.

Testing can enhance learning even when retrieval fails (Kornell, Hays, & Bjork, 2009; Grimaldi & Karpicke, 2012). Kornell et al. (2009) pretested participants on weakly related cue-target pairs by asking participants to generate a target to a given cue (e.g., pond - ?). Because the cue-target pairs were weakly associated (e.g., pond-frog) and had never been studied, participants rarely anticipated the correct target (< 5% of the time). Despite this, the pretests enhanced subsequent learning of the actual targets. It is important to note, though, that this paradigm differs somewhat from traditional paradigms examining test-enhanced new learning. In the pretesting paradigm, retrieval

(from semantic memory) occurs prior to any study episode. In the multi-list learning paradigm, retrieval (from episodic memory) occurs between two study episodes. Therefore, it is possible that the test-enhanced new learning in the two paradigms may have different underlying etiologies. Indeed, important differences exist between the effects found in these paradigms. In the pretesting paradigm, test-enhanced new learning occurs only for related word pairs and tends to disappear after a short delay between “retrieval” and the subsequent study trial (Grimaldi & Karpicke, 2012), while test-enhanced new learning in the intervening testing paradigm is not similarly constrained (Wissman, Rawson, & Pyc, 2011; Chan & Langley, 2011). Nevertheless, both paradigms have provided evidence that testing in some form, such as a pretest or an intervening test, can enhance subsequent learning, even if these paradigms might facilitate learning in different ways.

Evidence for Test-Impaired New Learning

Taken together, there is ample evidence that new learning can be enhanced by prior retrieval. However, Finn and Roediger (2013) demonstrated that, relative to restudy, retrieval can impair new learning. Given the wealth of data suggesting that testing should enhance new learning, this finding is surprising. In the current study, we sought to examine the reasons underlying this discrepancy. Our design closely follows that of Finn and Roediger, so we present their study in detail here.

Figure 1 displays the general paradigm used in our and Finn and Roediger’s (2013) study. Specifically, participants must remember three pieces of information: a face, a name, and a profession. In the initial study phase, participants learned a series of face-name pairs. Following a short distractor activity, participants completed an

intervening task. Participants assigned to the restudy condition were presented with each face-name pair again. Immediately after presentation of each face-name pair, the profession corresponding to that face was introduced. Participants in the test condition were presented with the face and prompted to recall the name. After that, the correct name appeared as feedback, which was followed by the profession for that face. After another short filler task, participants completed a final test for both the name and profession associated with each face.

In the context of this design, the name for each face represents original learning (similar to A-B pairs in Tulving and Watkins, 1974) and the profession represents new learning (similar to A-C pairs). Surprisingly, Finn and Roediger (2013) found that intervening retrieval of the names hindered, rather than facilitated, profession learning. This test-impaired new learning effect appeared to be robust, as it occurred either in the presence or absence of feedback on the intervening test, persisted over a delay of 24 hours, and occurred for arbitrary (e.g., face-name-profession) and pre-existing associations (e.g., word triads such as *ponder-wonder-think*).

Finn and Roediger (2013) proposed, and ultimately rejected, several possible explanations for this counterintuitive finding. One hypothesis that is of particular relevance for present purposes is that after attempting to recall a name, participants focus on relearning the name when the feedback is presented, which interferes with profession learning. They dismissed this account because test-impaired new learning persisted even

when feedback for the name was not administered (Experiment 1a and 4)¹. Ultimately, Finn and Roediger concluded that “a clear theoretical account of [their] findings is not yet at hand” (p. 1678), and that “these results are difficult to accommodate within the current theories that mostly emphasize benefits of retrieval for learning” (p. 1665).

We believe it may be premature to rule out the possibility of selective attention. Critically, participants in Finn and Roediger’s (2013) studies completed the intervening task (restudy or test) *and* the new profession learning for each face before proceeding to the next face. Thus, new (profession) learning and testing/relearning were intermixed during the intervening task. We believe that this intermixing procedure may lead participants in the intervening test condition, but not participants in the restudy condition, to prioritize relearning the names over learning the professions. Here, we refer to this idea as the *borrowed time* hypothesis. Specifically, we argue that testing participants on the name (with or without feedback) leads them to perseverate on learning the name, either because receiving a test on an item increases its perceived importance (Wilford, Chan, & Tuhn, 2014; Dirks, Thoma, Kester, & Kirschner, 2014), its perceived probability of being tested again later (Weinstein, Gilmore, Szpunar, & McDermott, 2014), or because being tested on a item highlights its difficulty (Kang, 2010). As a result, participants continue to study the name (when feedback is given) or continue to attempt retrieval of the name (when feedback is not given) when they were supposed to

¹ Specifically, Finn and Roediger noted that “This impairment did not appear to be due to selective attention to the feedback because no significant differences emerged between the tests with feedback and the test without feedback conditions in the recall of professions” (p. 1669).

study the profession. This can lead to poorer encoding of the profession, which is then manifested as test-impaired new learning. Indeed, the literature is replete with examples in which learners preferentially attend to one class of items at the expense of another. For example, in the spacing effect literature, when presented with mixed lists, participants often allocate their rehearsal time to spaced items over massed items (see Delaney, Verkoeijen, & Spirgel, 2010 for a review).

In fact, Finn and Roediger (2013) provided some evidence that is consistent with this borrowed time hypothesis. Specifically, test-impaired new learning was observed for faces whose names participants failed to recall on the initial test. In contrast, when initial retrieval of the name was successful, testing did not impair new learning. While there are the usual concerns with item selection effects for this type of analysis, these data suggest that performance on the intervening test may be a key factor in determining whether testing impairs new learning. Finn and Roediger proposed that the effect of initial retrieval success on new learning was the result of a context effect on knowledge updating. For successful retrievals, the test episode more closely matches the study context (i.e., the face appears with the correct name), which is more easily updated with the new profession. Conversely, unsuccessful retrievals do not match the study context (i.e., the face now appears with either an incorrect name or no name at all), and is less amenable to updating.

We suggest that there is a more straightforward explanation. Under the framework of the borrowed time hypothesis, successful retrievals reduce the need for participants to persevere on learning the name, because the name is already learned. As such, participants can focus on learning the profession when it appears. Conversely,

when participants failed to recall a name, they may be motivated to relearn it in preparation for a later test. This may cause participants to borrow time from profession learning to restudy the name instead. Importantly, this can occur regardless of whether feedback is presented. If a test contains no feedback, participants may still continue to try to retrieve the name when the profession is presented, leading to the impairment on profession performance.

A final and critical piece of the borrowed time hypothesis is that restudying should not lead participants to borrow time, because the difficulty of retrieval would not be salient. There is considerable evidence to suggest that this is the case. When individuals restudy, metacognitive judgments of learning are often inflated and they overestimate true performance due to the fluent processing afforded by a restudy trial (Koriat et al., 2004; Koriat & Bjork, 2005). Therefore, when participants restudy the face-name pairs during the intervening phase, they are unlikely to realize the difficulty they will have in recalling the name during the final test, thus making them unlikely to borrow time from profession learning. In contrast, when individuals are given an initial test on studied material, metacognitive judgments of learning are often depressed relative to restudying (Agarwal et al., 2008; Kang, 2010; McCabe, 2011). Consequently, testing, relative to restudying, should encourage relearning of the name, leading subjects to borrow time from profession learning. An interesting irony of this hypothesis is that testing impairs new learning partly because it produces better metacognitive monitoring than restudying. As participants become aware of their own memorial shortcomings during name retrieval, they attempt to rectify this poor performance by prioritizing study

of the name and borrowing time from learning the profession, leading to test-impaired new learning.

In fact, this prioritization of name learning does seem to have a positive impact on name recall on the final test. Finn and Roediger (2013) observed a testing effect in all seven of their experiments, despite the low initial test performance (~30%) and short retention interval (60 s) in most of their experiments – two factors that reduce the magnitude of the testing effect (Rowland, 2014). It is therefore quite possible that the observation of a testing effect in this context can be attributed, at least partly, to a heightened focus on name relearning following the intervening test. Consistent with this idea, participants in Finn and Roediger's study demonstrated a reversed testing effect for the names when feedback was not provided (Experiment 1a and Experiment 4).

An important attribute of the borrowed time hypothesis is that it suggests that test-impaired new learning occurs because people willingly prioritize encoding of one type of information over another. Essentially, individuals are regulating their own study strategies. Nelson and Narens (1994) proposed a three-pronged framework for self-regulated learning. When studying a set of items, individuals must decide 1) what material to study, 2) how long to study the selected material, and 3) when to terminate study. In the present context, we predict that participants are more motivated to learn the name in the test condition than in the restudy condition due partly to the salient difficulty of name retrieval. Once individuals determine which items to study (e.g., the names), they may devote more study time to more difficult items (Thiede & Dunlosky, 1999, but see Metcalfe, 2002; Metcalfe & Kornell, 2003; Metcalfe & Kornell, 2005; Kornell & Metcalfe, 2006 for an alternative account). The pattern of the conditional analysis

performed by Finn and Roediger (2013) fits neatly within this framework. Specifically, participants may devote more time to relearn the unrecalled names because these have proven particularly difficult to remember. In contrast, little additional study time is warranted for the recallable names. In sum, participants are motivated and able to focus on the name at the expense of the profession, especially on unsuccessful retrievals, and this causes profession performance to suffer in the test condition relative to the restudy condition.

It is also noteworthy that the specific methodology used by Finn and Roediger (2013) might have encouraged participants to borrow time from profession learning to focus on relearning the name. The vast majority of studies reporting test-enhanced new learning present the test/restudy phase separately from the new learning phase. For example, a participant might learn a list of words, take a free recall test for those words, and then learn a list of entirely new words. Thus, the initial study, retrieval, and new learning occurred in separate phases/blocks of trials (e.g., Chan, Thomas, & Bulevich, 2009, Pastötter et al, 2011, Szpunar, McDermott, & Roediger, 2008, Tulving & Watkins, 1974). In contrast, Finn and Roediger's interpolated test trials were fully intermixed with new learning trials, and this procedure may be a key to enabling time borrowing. When profession learning immediately follows name retrieval, name performance will be particularly salient and more likely to affect encoding of the profession. It is worth noting here that the pretesting paradigm discussed previously (Kornell et al., 2009) produced test-enhanced new learning despite it requiring participants to switch from "test" to study repeatedly. This finding does not pose a problem for the borrowed time hypothesis because the pretest did not require retrieval of studied associates. Rather,

participants were encouraged to guess the identity of a target, and were given the correct target immediately after. Because the “tested” target had never been studied, participants would not be motivated to borrow time to relearn their own “mistake” at the expense of the correct target.

The current series of experiments were designed to test our borrowed time hypothesis. The overall strategy was to reduce test-impaired new learning by discouraging time borrowing. The key differences between each experiment as well as the key findings from each experiment are presented in Table 1.

Experiment 1: Replicating Finn and Roediger (2013)

In our first experiment, we sought to replicate Finn and Roediger’s (2013) finding. Using a different set of faces, we followed their procedure exactly with one exception. On the final test, rather than intermixing profession and name recall, we tested the names and professions in separate trial blocks, with the profession test always occurring first. We made this change to ensure that any impairment observed for professions was not due to output interference from the names. This test procedure was used in all of the following experiments.

Method

Participants, design, materials, and procedure. Seventy-nine undergraduate students at Iowa State University participated for course credit. Four participants were excluded for insufficient language proficiency, yielding 38 participants in the restudy condition and 37 in the test with feedback condition.

The experiment consisted of a 2 (Intervening Task: Restudy vs. Test with Feedback) \times 2 (Item Type: Name vs. Profession) mixed design. Intervening task was manipulated between subjects and item type was manipulated within subjects.

We assembled 20 face-name-profession triads for stimuli. The names and professions were identical to those used by Finn and Roediger (2013), but the faces were selected from Minear and Park (2004). The 10 male and 10 female faces ($M_{\text{age}} = 35.7$ and 35.5 years, respectively) all had a neutral, closed-mouth expression. All faces were paired with names typical of their gender while professions were randomly assigned to each face. The order of presentation for each face was randomized for each participant. The final test consisted of two sets of trials. The first 20 trials tested participants' memory for the professions, and the last 20 trials tested their memory for the names.

During the intentional encoding phase, the 20 face-name pairs were presented for 5 s each, with the name presented directly below the face (see left side of Figure 1). Afterwards, participants completed a 60 s math task. Ten problems (e.g., $15 - 5 \times 2 = ?$) were presented on the screen sequentially for 6 s each, and participants were instructed to type in their responses.

During the intervening phase, participants either restudied the name before learning the profession, or they completed a test for the name before learning the profession. All participants were instructed to pay attention to the professions, as they would be tested later. In the restudy condition, faces were presented with their corresponding names for 5 s, and then the profession appeared below the name for an additional 5 s. In the test with feedback condition, participants were given a face and were asked to recall the name. Participants had unlimited time to respond, but could

press the Enter key to indicate that they could not recall the name. Once a response was entered, the correct name appeared for 2 s as feedback, and then the profession appeared below the name for 5 s as in the restudy condition.

After another 60 s math task with new questions, participants took the self-paced final test. The 20 faces were presented sequentially in a random order, and participants were instructed to recall the profession for each face. After all professions had been tested, participants saw the faces in a new random order and were prompted to enter the names. A 500 ms inter-trial interval was used throughout the experiment.

Results and Discussion

Recall performance on the final and initial tests is presented in Figure 2. During the initial test, participants recalled 31% of the names. This relatively poor performance (similar to results from Finn and Roediger, 2013) suggests that, as expected, face-name associations were difficult to learn. A 2 (Intervening Task: Restudy vs. Test with Feedback) \times 2 (Item Type: Name vs. Profession) analysis of variance (ANOVA) was conducted on final test recall performance. There was a main effect of item type, $F(1, 73) = 22.64, p < .01, \eta_p^2 = .24$, which showed that participants recalled more professions ($M = .53$) than names ($M = .42$). More importantly, there was a crossover interaction between item type and intervening task, $F(1, 73) = 26.38, p < .01, \eta_p^2 = .27$. Initial testing enhanced final name recall ($M = .48$) relative to restudying ($M = .35$), $t(73) = 2.55, p = .01, d = .59$, but it impaired final profession recall ($M = .47$) relative to restudying ($M = .57$), $t(73) = 2.45, p = .02, d = .57$.

These results thus replicated those reported by Finn and Roediger (2013). Once again, we suggest that test-impaired new learning occurs because participants borrow

time from profession learning to relearn the name during the intervening test, which led to poorer encoding of the profession. To further evaluate this hypothesis, we conducted a conditional analysis to examine whether final recall of professions depended on initial retrieval success of the name. We believe that participants would be more inclined to borrow time from profession learning if they failed to recall the name than if they successfully recalled the name. Consistent with this idea, testing impaired profession learning ($M = .43$) relative to restudying ($M = .58$), $t(73) = 3.62$, $p < .01$, $d = .83$, when participants failed to recall the name during the initial test, but not when subjects successfully recalled the name during the initial test ($M = .55$), $t(73) = .50$, $p = .62$, $d = .11$. The data are displayed in Table 2. It is important to note here that conditional analyses must be interpreted cautiously due to item selection effects. That is, items that are more difficult to learn are less likely to be successfully retrieved, and would likely require more processing following retrieval. However, we believe that these analyses can be valuable for identifying when and why time borrowing may occur.

In sum, we extended Finn and Roediger's (2013) findings to a new set of materials and final test in Experiment 1. The remaining experiments were designed specifically to test the borrowed time hypothesis. In Experiment 2, we attempted to discourage participants from borrowing time from profession learning by presenting the profession *before* testing the name during the intervening phase. Specifically, during each trial of the intervening phase, participants learned the profession before they took a test on the name for each face. We predicted that because the new learning occurred before the intervening test, participants could not borrow time from the profession presentation and so test-impaired learning of professions should not occur.

Experiment 2: Reversed Restudy/Test and New Learning

Method

Participants, design, materials, and procedure. Eighty-one undergraduate students at Iowa State University participated in this experiment. Data from six participants were excluded for insufficient language proficiency, yielding 27 participants in the restudy condition and 24 in the test with feedback condition.

The design and materials were similar to Experiment 1, except that the order of testing/restudying and new learning was reversed. A graphical representation of the procedure is displayed in Figure 3. In the restudy condition, the face and profession were presented for 5 s, followed by the face and name for an additional 5 s. In the test condition, the face and the profession were presented for 5 s, after which the profession was removed and participants were tested on the name, which was followed by a 2 s presentation of the name as feedback. Similar to Experiment 1, the duration of profession presentation for each face was held constant across the restudy and test conditions (i.e., 5 s). Intervening task (restudy vs. test with feedback) was manipulated between subjects, and item type (name vs. profession) was manipulated within subjects.

Results and Discussion

Recall performance is displayed in Figure 4. During the initial test, participants recalled 32% of the names. There was a main effect of item type, $F(1, 49) = 6.32, p = .02, \eta_p^2 = .11$, showing that names ($M = .49$) were better recalled overall than professions ($M = .42$). Participants in the restudy condition ($M = .54$) also outperformed those in the test condition ($M = .38$) overall, $F(1, 49) = 14.33, p < .01, \eta_p^2 = .23$. These main effects were qualified by an interaction, $F(1, 49) = 6.92, p = .01, \eta_p^2 = .12$, such that testing ($M =$

.45) and restudying ($M = .53$) led to similar performance for the names during the final test, $t(49) = 1.23$, $p = .23$, $d = .34$. Most importantly, the test-impaired new learning effect for profession persisted, with participants in the test condition ($M = .29$) recalling fewer professions than participants in the restudy condition ($M = .54$), $t(49) = 6.34$, $p < .01$, $d = 1.80$.

It is surprising that test-impaired new learning remained in the present experiment. Although we expected that presenting the profession before name retrieval should eliminate time borrowing, in hindsight, this prediction was flawed, as participants could simply borrow time from profession learning during the next, instead of the current, trial. Moreover, by reversing the order of the initial test and profession learning, we might have further encouraged time borrowing for participants in the test condition. Due to procedural constraints, the name feedback was only presented for 2 s in Experiment 2, relative to 7 seconds in Experiment 1. This drop from 7 s to 2 s of name feedback could have exacerbated time borrowing. Although participants in the restudy condition also received less time to relearn the name in Experiment 2 (a total of 5 s) than in Experiment 1 (a total of 10 s), this drop was proportionally smaller than that of the test condition.²

Again, we performed conditional analyses on final profession performance for the test condition based on initial retrieval success. Similar to Experiment 1, testing impaired new learning when initial retrieval was unsuccessful ($M_s = .54$ and $.26$ for the restudy

² Although the name feedback was presented for a shorter duration in Experiment 2 than in Experiment 1, this change did not alter the total duration of the intervening phase (compare Figure 1 to Figure 3). Note that the name appeared for a shorter period in Experiment 2 because we did not present it *simultaneously* with the profession.

and test conditions, respectively), $t(49) = 7.11, p < .01, d = 2.02$. Interestingly, test-impaired new learning occurred when initial retrieval was successful, too ($M = .34$ for the test condition), $t(49) = 2.99, p < .01, d = .83$, although the effect was smaller when retrieval was successful than when it was not.

One takeaway of the results in Experiment 2 is that the test-impaired learning effect might be attributable to the differences in relearning duration, as participants in the test condition consistently received less time to relearn the name than their restudy counterparts.³ We suspect that this disparity in relearning time is partially responsible for time borrowing. In Experiment 3, we sought to reduce the likelihood that participants would borrow time from profession learning by equating the duration of name relearning between the test and restudy conditions. We hypothesized that giving participants more time to relearn the name would reduce their need to borrow time when the profession was presented, eliminating the test-impaired learning effect.

Experiment 3: Equating Relearning Duration

Method

Participants, design, materials, and procedure. Sixty-six undergraduate students at Iowa State University participated in this experiment. Data from four participants were removed because they were non-English speakers, yielding 31 participants in each condition. The design and materials were identical to Experiment 1, except that the name feedback was presented for 5 s instead of 2 s. This procedure thus equated the amount of time that participants had to relearn the names in the restudy and test conditions

³ This is also true for all the experiments reported in Finn and Roediger (2013).

Results and Discussion.

The key data are presented in Figure 5. During the initial test, participants recalled 31% of the names. There was a main effect of item type, such that participants recalled more professions ($M = .50$) than names ($M = .40$) during the final test, $F(1, 60) = 15.79, p < .01, \eta_p^2 = .21$. Moreover, participants in the test condition ($M = .49$) outperformed their restudy counterparts ($M = .41$), $F(1, 60) = 3.36, p = .07, \eta_p^2 = .05$. These main effects were qualified by an interaction, $F(1, 60) = 30.42, p < .01, \eta_p^2 = .34$. Follow-up comparisons revealed a powerful testing effect for the names, $t(60) = 4.12, p < .01, d = 1.06$. Here, participants in the test condition recalled far more names ($M = .50$) on the final test than those in the restudy condition ($M = .29$). Most importantly, initial testing no longer produced a significant impairment on new (profession) learning, $t(60) = 1.26, p = .21, d = .32$, although the difference was still in the direction of test-impaired new learning.

The reduction of test-impaired new learning is consistent with our hypothesis that participants borrow time from the profession learning trial to relearn the name. By lengthening the feedback duration, we gave participants in the test condition more time to focus on learning the name so the need to borrow time from the profession was diminished. Furthermore, a large testing effect on final name performance was found ($d = 1.09$), which suggests that participants did make use of the additional feedback duration to improve their name performance.

Despite equating the name relearning duration between the test and restudy conditions, a small and nonsignificant test-impaired learning effect remained. We suspect that this occurred because initial testing, relative to restudying, highlighted the

difficulty of name retrieval, thus causing participants to borrow time despite the longer feedback duration. To examine this possibility, we again examined final profession performance contingent on initial retrieval success. Similar to results from Experiments 1 and 2, testing impaired new learning when initial retrieval of the name failed ($M_s = .52$ and $.39$ for the restudy and unsuccessful retrieval conditions, respectively), $t(60) = 2.65$, $p = .01$, $d = .67$, but not when initial retrieval was successful ($M = .57$), $t < 1$. Thus, even though we discouraged time borrowing overall by increasing the duration of the feedback, it appears that participants were still motivated to borrow time from profession learning when the initial retrieval of the name was unsuccessful.

In Experiment 3, we were able to reduce, though not eliminate, test-impaired learning by equating the relearning duration between the restudy and test conditions. As we discussed in the Introduction, we believe that intermixing testing and relearning trials might have encouraged participants to borrow time, and this appears to be the case even when we gave participants ample time to relearn the name. In Experiment 4, we sought to address the idea that intermixing testing and new learning contributed to test-impaired learning by dividing these activities into separate blocks of trials. Specifically, after subjects studied all the face-name pairs, they either restudied or were tested on all the faces with feedback. Note that no new (profession) learning trials occurred during the intervening test/restudy phase. Instead, new learning occurred in an immediately following block of trials. We hypothesized that individuals would be less likely to borrow time under this procedure, as profession learning was not intermixed with test trials. Additionally, by blocking all profession encoding trials together, participants

would have little incentive to try to relearn the name by retrieving it from long-term memory, because the names were never presented during the profession learning phase.

This paradigm also more closely matched the procedure used by studies that demonstrated test-enhanced new learning, where initial testing and new learning occurred in separate blocks of trials (e.g., Szpunar et al., 2008; Tulving & Watkins, 1974).

Therefore, it is possible that participants in the test condition would recall more professions on the final test than those in the restudy condition here.

Experiment 4: Separating New Learning from Restudying/Testing

Method

Participants, design, and materials, and procedure. Seventy-four undergraduate students at Iowa State University participated in this experiment. We excluded data from eight participants for insufficient language proficiency and two for not following instructions. This yielded 32 participants in each intervening task condition. The design and materials were identical to Experiment 1, with one critical difference. For each intervening task condition, restudy/retrieval and profession learning trials were presented in separate trial blocks. Participants were tested on and relearned names for all faces before learning the professions.

The initial encoding phase of this experiment was identical to Experiment 1, and was followed by a 60 s filler task. Next, participants in the restudy condition saw the 20 face-name pairs again for 5 s each. Participants in the test condition were presented with each face and cued to enter its name. After a response was entered, the correct name was displayed below the face for 2 seconds. Participants proceeded to the new learning phase after all 20 faces were restudied/tested. Here, the 20 faces were presented in a new

random order for 5 s each, with the profession displayed beneath each face. Because the purpose of this experiment was to discourage time borrowing, the name for each face was not presented (similar to Experiment 2)⁴. Finally, participants completed the final test.

Results and Discussion

Results are displayed in Figure 6. Name recall on the initial test was 38%. Similar to previous experiments, participants recalled more professions ($M = .61$) than names ($M = .51$) during the final test, $F(1, 62) = 12.09, p = .01, \eta_p^2 = .16$. There was also an interaction between item type and intervening task, $F(1, 62) = 21.98, p < .01, \eta_p^2 = .26$, such that name performance in the test condition ($M = .44$) was lower than that in the restudy condition ($M = .57$), $t(62) = 2.49, p = .02, d = .62$. We address this reversed testing effect in more detail in the General Discussion. More important for the present purpose is that unlike the previous experiments, testing enhanced new learning of the professions, such that initial testing of the name facilitated learning of the profession ($M = .69$) relative to restudying ($M = .54$), $t(62) = 3.25, p = .01, d = .81$. Even more impressively, the conditional analysis showed that the level of test-enhanced new learning on profession recall was not dependent on whether initial retrieval was successful, ($M_{\text{test}} = .63$ vs. $M_{\text{restudy}} = .57$), $t(62) = 2.42, p < .02, d = .61$, or not ($M_{\text{test}} = .64$), $t(62) = 2.51, p < .02, d = .63$, which suggests that participants did not borrow time from profession learning to recall the names, even when initial retrieval failed.

⁴ One might wonder whether having the name presented together with the profession is necessary for test-impaired new learning to occur. We believe that this is not the case, as we observed a powerful test-impaired new learning effect in Experiment 2, where the profession did not appear with the name simultaneously.

The findings from this experiment are notable as they represent a reversal of the effects in Experiments 1-3 and those reported in seven experiments by Finn and Roediger (2013). This result shows that testing can enhance new learning for face-name-profession associations and when participants need to update existing memory structures (i.e., face-name associations) with additional new information (i.e., profession), so that neither the materials themselves nor the requirement of memory updating are responsible for test-impaired new learning.

Interim Summary

In Experiment 1, we showed that intermixing test trials with new learning trials impaired new learning. Experiment 2 showed that reversing the order of new learning and test trials did not reduce this effect. In Experiment 3, we demonstrated that increasing presentation duration of the relearning trials weakened, though did not eliminate, test-impaired new learning. Experiment 4 established that testing can enhance new learning with these materials, as long as initial testing and new learning were split into separate trial blocks.

Heretofore, we have argued that these results support the borrowed time account. However, a closer examination suggests that these findings are also consistent with an alternative explanation. Specifically, intermixing test trials with new learning trials may produce test-impaired new learning because it requires participants to repeatedly switch between encoding and retrieval. Considerable research has shown that rapidly switching between two or more tasks with different processing requirements is associated with a performance cost (Masson et al., 2003). Consistent with this idea, we observed a reduction in test-impaired new learning by increasing the feedback duration, similar to

how task switching costs can be minimized by giving people more time to prepare for an upcoming switch trial (Meiran, 1996). The test-enhanced new learning effect in Experiment 4 is also consistent with this task switching account, because here participants no longer had to repeatedly switch between encoding and retrieval during the intervening phase.

Essentially, the task switching account suggests that test-impaired new learning occurs because intermixing test trials with encoding trials increases the processing requirement of the task. In contrast, the borrowed time account suggests that the intermixing procedure causes a change in encoding preference, such that participants now place a greater emphasis on relearning the name than the profession.

In Experiment 5, we aimed to tease apart the borrowed time and task switching explanations by altering encoding preference during the intervening phase. Here, participants were told explicitly during the intervening phase whether they should prioritize their learning effort on the names or the professions. We believe this manipulation, if successful, would cause a shift in participants' encoding preference. For example, when participants were told to focus on learning the professions, they should be less likely to borrow time from profession learning to relearn the name. Critically, this instructional manipulation did not alter the task switching requirement of the intervening task, as participants still needed to repeatedly switch between encoding and retrieval. Therefore, according to the task switching account, altering participants' learning preference should have no impact on test-impaired new learning. In contrast, according to the borrowed time account, having participants prioritize their learning preference to

the professions should discourage time borrowing and thus reduce or eliminate test-impaired new learning.

In Experiment 5, we utilized an instructional manipulation that prioritized either name or profession learning. To ensure that the learning priority instructions were effective, we offered participants a reward for achieving a high level of performance on the attribute (i.e., name or profession) designated by the instructions. Specifically, participants were told that they could finish the experiment early if they achieved high performance on either names or professions. This incentive manipulation was modeled after Kang and Pashler's (2014). We predicted that emphasizing profession performance, but not name performance, would eliminate test-impaired new learning, because we believe that names were prioritized "by default," at least for participants in the test condition.

Experiment 5: Incentivizing Name or Profession Performance

Method

Participants, materials, design, and procedure. One hundred twenty-seven undergraduate students participated in this experiment. Data from five participants were eliminated for insufficient language proficiency and two for not following instructions, yielding 30 participants in each between-subjects condition. The materials and procedure were identical to Experiment 1, with the exception of the instructions given prior to the intervening task.

This experiment utilized a 2 (Study Priority: Profession vs. Name) x 2 (Intervening Task: Restudy vs. Test) x 2 (Item Type: Name vs. Profession) mixed design.

Study priority and intervening task were manipulated between subjects, and item type was manipulated within subjects.

Prior to the intervening task/profession learning phase, participants were incentivized⁵ to either maximize their performance on the names or professions. These instructions were given immediately before the initial recall test, so that incentive could not impact initial encoding of the faces and names. Half of the participants were told to maximize their name performance, and half were told to maximize their profession performance with the following instructions:

In the next portion of the experiment, you will be relearning the names as well as learning the profession for each face. Later, you will be given a final test for the names and professions. Your performance on these final tests is very important. If you remember more than 14 of the 20 PROFESSIONS/NAMES, you will not be required to complete any further tasks. If you remember fewer than 14 of the 20 PROFESSIONS/NAMES, you will have to complete one more task, which takes approximately 25 minutes. If you are allowed to leave early, you will still receive full credit, so it is in your best interest to study the PROFESSIONS/NAMES carefully in this next part in order to maximize your score.

⁵ We also conducted a pilot experiment in which we told participants to either prioritize learning the names or professions without incentivizing performance. But these instructions were apparently utterly ineffective, as they did not impact performance on either names or professions relative to Experiment 1. For the purpose of full disclosure, the results of this experiment are presented in the Appendix. As can be seen, the pattern here was highly similar to that reported in Experiment 1.

These instructions were first given in person by the experimenter and then reiterated in writing on the instruction screen. Finally, although participants were told that their final performance would determine whether the duration of the experiment was shortened, all participants were actually dismissed after the final test.

Results and Discussion

We first present data from the Name Priority condition; we then present data from the Profession Priority condition. Initial test performance was 42% in the Name Priority condition and 37% in the Profession Priority condition. The initial and final test data are presented in Figure 7.

Name Priority. The Name Priority instructions appeared to be effective, as names ($M = .57$) were recalled more often overall than professions ($M = .51$), $F(1, 58) = 4.59, p < .04, \eta_p^2 = .07$. This main effect was qualified by a crossover interaction between item type and intervening task, $F(1, 58) = 11.41, p < .01, \eta_p^2 = .16$, such that initial testing benefitted final recall of the names ($M = .65$) relative to restudying ($M = .50$), $t(58) = 3.09, p < .01, d = .80$, but it impaired, though not significantly, the recall of professions ($M = .47$) relative to restudying ($M = .54$), $t(58) = 1.60, p = .12, d = .41$

Why did the Name Priority instructions reduce the test-impaired new learning effect, even though we had hypothesized that participants in the test condition were prioritizing name relearning “by default?” One explanation is that the Name Priority instructions shifted the encoding preference for participants in the restudy condition, thus making these participants more likely to borrow time from professions to relearn the names as well, as evidenced by their name recall superiority over profession. By motivating participants in the restudy condition to have the same goal as those in the test

condition (i.e., to maximize name performance), the difference between these two conditions was diminished.

We again conditionalized final profession performance based on initial retrieval success. For the Name Priority condition, testing impaired profession learning marginally when name retrieval was unsuccessful ($M_s = .54$ and $.45$ for the restudy and test conditions, respectively), $t(57) = 1.90, p = .06, d = .49$, but not when initial retrieval was successful ($M = .51$), $t(57) = .68, p = .50, d = .18$.

Profession priority. Similar to the Name Priority instructions, the Profession Priority instructions appeared to be effective, such that participants in this condition recalled more professions ($M = .58$) overall than names ($M = .48$), $F(1, 58) = 11.22, p = .001, \eta_p^2 = .16$. This main effect was qualified by an interaction between item type and intervening task, $F(1, 58) = 16.20, p < .01, \eta_p^2 = .22$. Testing ($M = .60$) increased recall probability for the names relative to restudying ($M = .37$), $t(58) = 3.93, p < .01, d = 1.01$, but it did not influence recall probability for professions at all ($M_s = .58$ and $.59$ for the test and restudy conditions, respectively), $t < 1$. Remarkably, the profession priority instructions abolished test-impaired new learning, regardless of whether initial retrieval was successful ($M_s = .58$ and $.54$ for the restudy and test conditions, respectively), $t(57) = .43, p = .67, d = .11$, or not ($M = .61$), $t(57) = .75, p = .46, d = .19$. This is important because testing consistently impaired new learning when initial retrieval failed — even when the overall effect was not significant (e.g., the Name Priority condition in Experiment 3). This suggests that the Profession Priority instructions might have eliminated time borrowing.

One might wonder why testing did not enhance new learning in this experiment if we were able to eliminate time borrowing. Although we showed in Experiment 4 that the present *materials* are amenable to test-enhanced new learning, the intermixed testing and new learning *procedure* might not be. Szpunar and colleagues (2008) suggested that testing might enhance new learning because taking a test between two learning episodes inoculates the later learning episode against the deleterious effects of proactive interference from prior learning. Proactive interference tends to build up over a series of learning trials. In a restudy condition, this proactive interference is never released, as participants continue with a restudy phase and then a new learning phase. In a test condition, however, the intervening test provides an opportunity for proactive interference to be released, which could be accomplished by an internal context change from encoding to retrieval (Jang & Huber, 2008; Pastötter et al., 2011). But this release from proactive interference may not be possible when one intermixes retrieval trials with new learning trials, as the intervening task now contains further encoding trials, thus allowing proactive interference to build up.

To examine whether intermixing retrieval and encoding trials during the intervening phase causes proactive interference to build up, we compared performance between the first half and second half of the initial test. If proactive interference accumulated across the intervening phase, then we should observe a drop in initial test performance across trials. Combining the data from Experiments 1, 2, 3, and 5, we observed a large drop in performance between the first half ($M = .41$) and the second half ($M = .28$) of the initial test, $t(151) = 7.59, p < .01, d = .66$. However, no detectable drop in performance associated with proactive interference was found for participants in

Experiment 4 (M 's = .38 and .36 for the first and second half, respectively), where retrieval was not intermixed with new learning, $t(31) = .52$, $p = .60$, $d = .09$.

These results are consistent with the idea that intermixing testing and new learning trials contributed partly to Finn and Roediger's (2013) results because of the buildup of proactive interference during the intervening learning phase. However, the intermixing procedure likely also led to test-impaired new learning because it encouraged time borrowing, rather than because of task switching.

General Discussion

Despite a growing body of research showing that testing can enhance new learning, our results, along with those reported by Finn and Roediger (2013), suggest that the opposite can occur. Testing can impart a cost on future learning when the encoding environment encourages strategic processes that favor relearning the old material over the new. We have proposed that test-impaired new learning occurred in Finn and Roediger's study because their paradigm capitalized on participants' metacognitive judgments about their inability to learn the face-name associations, for which participants compensated by allocating more time to relearn. Our borrowed time hypothesis posits that when the intervening test trials are intermixed with new learning trials, it biases participants to focus on relearning the name at the expense of the profession, thus leading to test-impaired new learning.

Following Finn and Roediger's procedure while using a new set of faces and a different final test, we found test-impaired new learning in Experiment 1. We then employed several experimental manipulations that aimed to alleviate time borrowing in Experiments 2-5. In Experiment 2, we presented the new learning trials before the

intervening task, using the logic that the intervening test could not impair new learning as it occurred after new learning. However, a test-impaired new learning effect remained.

In Experiment 3, we increased the initial test feedback duration to discourage participants from borrowing time. This procedure successfully reduced the size of the effect. However, our conditional analyses also revealed a test-impaired new learning effect when initial retrieval had failed, suggesting that even when the experimental conditions discouraged time borrowing, it could still occur on an individual trial level.

In Experiment 4, we attempted to reduce time borrowing by dividing the intervening restudy/test and new learning into separate blocks of trials. Here, we observed test enhanced new learning. Moreover, this effect occurred regardless of whether the name for a given face was successfully retrieved initially. Notably, this test-enhanced new learning effect for professions was accompanied by a reversed testing effect for the names, contrary to our findings from the remaining experiments. We believe that this reversed testing effect lends even more credence to the borrowed time hypothesis. Specifically, although our discussion thus far has focused on the effects of time borrowing on new learning (i.e., testing impairs learning for the professions), time borrowing is also partly responsible for the testing effect for the names (as we mentioned in the Introduction). In essence, the borrowed time account suggests that due to the limited time with which participants had to relearn the name and to learn the profession, some trade-off was necessary. The reversed testing effect in Experiment 4 shows that this is indeed the case. When participants were unable to allocate part of their profession study time to relearn the name, the testing effect disappeared.

In Experiment 5, we examined whether task switching was responsible for test-impaired learning and its reduction/elimination in Experiments 3 and 4. We manipulated participants' learning priority without altering the task-switching requirement of the intervening task. When participants were told to prioritize name learning over profession learning, testing impaired new learning, at least for faces whose name was not recalled initially. In contrast, when participants were told to focus on learning the professions, testing no longer impaired new learning – even when initial retrieval failed. The results from Experiment 5 therefore favor a time borrowing, rather than a task switching, explanation. Furthermore, it is important to note that in contrast to Experiment 4, we found a testing effect for the names. We believe that this occurred in Experiment 5 because our instructional manipulation increased motivation to perform well, which is interesting from a pedagogical perspective.

Test-Impaired New Learning as Self-Regulated Learning

The time borrowing hypothesis suggests that test-impaired new learning occurs because participants willingly direct encoding resources to relearn old information at the expense of new information. In the present study, we believe that participants were prioritizing name relearning over profession learning. But why would the tested participants place more emphasis on the names than the professions in the absence of any explicit instructions to do so?

Metcalf and colleagues (Metcalf, 2002; Metcalf & Kornell, 2003; Metcalf & Kornell, 2005; Kornell & Metcalf, 2006) have proposed that individuals might prioritize study based on regions of proximal learning. That is, individuals will study material that is the closest to being fully learned. When faced with time constraints, participants will

focus on easy- and medium-difficulty items that may most benefit from additional study rather than laboring in vain to learn difficult items. In the present experiments, the most proximally learned materials are likely the names from the participants' perspective. The names have been studied initially and re-encoded (either as a restudy trial or as test feedback), and may be deemed as more worthy of additional study than new information. When faced with a decision to either abandon study of the names or to continue focusing on them, participants may have wagered that their largest gains in performance would occur by studying the names, rather than professions. In addition, participants might have opted to prioritize relearning the names over the profession because they had already committed considerable effort to these names during the initial learning phase. This is essentially a sunk-cost phenomenon (Arkes & Aykon, 1999), whereby decision makers tend to over-invest in tasks that they have committed to previously.

Despite the evidence supporting the region of proximal learning hypothesis, an alternative view is that participants simply prioritized learning of more difficult items, as these items demonstrated the greatest discrepancy between unlearned and learned states (Dunlosky & Hertzog, 1998). In the absence of motivational instructions, names were remembered more poorly overall than professions in our experiments (with the exception of Experiment 2). Thus, under this account, participants should be more inclined to focus on name performance because they were more difficult to remember than professions. Our experiments were not designed to test between these two hypotheses, and the data could support both. Regardless of which stance one takes on how individuals regulate study, however, the literature suggests that individuals are motivated and able to direct

encoding to some types of information at the expense of others, especially when they perceive that doing so will be advantageous for later performance.

We add the caveat that the current experiments were not designed to specifically examine how participants regulated their study time. As we did not directly assess metacognitive judgments in any of our experiments, it is difficult to tell which items participants may have deemed worthy of further study. However, the literature in self-regulated study informs the borrowed time hypothesis by confirming that participants can and do use their metacognitive judgments to guide their study behaviors. The overarching goal of the current experiments was to explore the feasibility of the borrowed time account by utilizing experimental manipulations that would discourage time borrowing. In so doing, we showed that these manipulations largely reduced the test-impaired new learning effect observed by Finn and Roediger (2013).

One might also wonder why would testing impair new learning in the absence of feedback (Finn & Roediger, 2013)? While there is evidence supporting the idea that individuals will shift encoding attention to one class of materials over another, these shifts typically occur at a fixed presentation rate, as in the spacing effect literature (Delaney et al., 2010). In the case of the present paradigm, however, the test trials are self-paced, so participants can, in theory, spend as much time as they need to retrieve each name on the initial test. Based on this logic, participants would have no reason to borrow time during profession learning without initial test feedback. However, we believe that simply because participants are given unlimited time for retrieval does not ensure that they will use it. Most models of memory (e.g., Search of Associative Memory) include criteria for terminating retrieval (Mensink & Raaijmakers, 1988,

Raaijmakers, 2008, Raaijmakers & Shiffrin, 1981). On an initial test trial, participants may quickly terminate their search for the name due to its apparent difficulty, or they may simply indicate that they have terminated a retrieval attempt to move to the profession learning trial even though the search was ongoing. Once the face appears again with the profession, participants may re-initiate the search for the name, perhaps because they erroneously believe that the profession will help them retrieve the name. This can then trigger time borrowing and produce a test-impaired learning effect.

Mechanisms of Test-Enhanced New Learning

In the present experiments, we showed that discouraging time borrowing can weaken (Experiment 3) or eliminate (Experiment 5) the test-impaired learning, but it was not enough to produce test-enhanced new learning. Instead, interpolated testing must be temporally separated from new learning to produce test-enhanced new learning (Experiment 4). We argue that this is because the procedure of intermixing new learning trials with the intervening test does not allow the test trials to reduce proactive interference that builds up during encoding. These results also show that testing enhances and impairs new learning based on separable mechanisms, with time borrowing and proactive interference playing different roles.

This finding has important educational implications, where tests and quizzes are often intermixed with presentation of new material. Szpunar et al. (2013) found that administering periodic quizzes during online lectures can help improve learning outcomes, and that these quizzes promoted engagement with material and helped students sustain attention. However, the lectures used by Szpunar et al. took place on a larger time-scale than in the current study (and in fact was more similar to the procedure used in

Experiment 4), and the number of quizzes administered was far fewer than in the present paradigm. Although it is unclear exactly how many tests (or how frequently they are administered) is necessary to induce test-impaired learning, one possibility is that people might be particularly prone to perseverate on the original material (i.e., the name in the present experiments) in lieu of learning the new material if the former is difficult.

Moreover, our results also show that being reminded of previously studied materials while learning new ones do not necessarily remove the deleterious effects of proactive interference (cf., Wahlheim, 2014). Regardless, the present results suggest that intermixing study and test trials can sometimes have a negative impact on new learning. Ultimately, the frequency of test administration in the classroom should be approached somewhat cautiously, especially when to-be-tested material is difficult and performance is expected to be low.

Conclusion

The time borrowing hypothesis provides a framework for understanding test-impaired new learning, and is couched within participants' metacognitive judgments of performance. Although the borrowed time hypothesis can predict when testing may impair new learning, it cannot determine when testing will enhance new learning. We have provided evidence here that the two effects are driven by separate mechanisms, and that test-enhanced new learning can only occur in paradigms in which the buildup of proactive interference is minimized. An important implication of the current study is that testing provides metacognitive information that can direct encoding effort to some type of information at the expense of another. Despite the empirical evidence that testing can be

a powerful memory enhancer, in some cases it can also have a negative impact on future learning outcomes.

References

- Agarwal, P.K., Karpicke, J.D., Kang, S.H.K., Roediger, H.L., & McDermott, K.B. (2008). Examining the testing effect with open- and closed-book tests. *Applied Cognitive Psychology, 22*, 861-876. doi:10.1002/acp.1391
- Arkes, H.R., & Aykon, P. (1999). The sunk cost and Concorde effects: Are humans less rational than lower animals? *Psychological Bulletin, 125*(5), 591-600. doi:10.1037/0033-2909.125.5.591
- Arnold, K. M., & McDermott, K. B. (2013). Free recall enhances subsequent learning. *Psychonomic Bulletin & Review, 20*(3), 507-513. doi:10.3758/s13423-012-0370-3
- Bäuml, K.H.T., & Kliegl, O. (2013). The critical role of retrieval processes in release from proactive interference. *Journal of Memory and Language, 68*(1), 39-53. doi:10.1016/j.jml.2012.07.006
- Chan, J.C.K., & Langley, M. (2011). Paradoxical effects of testing: Retrieval enhances both accurate recall and suggestibility in eyewitnesses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 248-255. doi:10.1037/a0021204
- Chan, J.C.K., & LaPaglia, J.A. (2011). The dark side of testing memory: Repeated retrieval can enhance eyewitness suggestibility. *Journal of Experimental Psychology: Applied, 17*, 418-432. doi:10.1037/a0025147
- Chan, J. C. K., & LaPaglia, J. A. (2013). Impairing existing declarative memory in humans by disrupting reconsolidation. *Proceedings of the National Academy of*

- Sciences of the United States of America*, 110(23), 9309–9313.
<http://doi.org/10.1073/pnas.1218472110>
- Chan, J. C. K., & McDermott, K. B. (2007). The testing effect in recognition memory: A dual process account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(2), 431–437. doi:10.1037/0278-7393.33.2.431
- Chan, J.C.K., Thomas, A.K., & Bulevich, J.B. (2009). Recalling a witnessed event increases eyewitness suggestibility: The reversed testing effect. *Psychological Science*, 20, 66–73. doi:10.1111/j.1467-9280.2008.02245.x
- Chan, J.C.K., Wilford, M.M., & Hughes, K.L. (2012). Retrieval can increase or decrease suggestibility depending on how memory is tested: The importance of source complexity. *Journal of Memory and Language*, 67, 78–85.
doi:10.1016/j.jml.2012.02.006
- Delaney, P. F., Verhoeijen, P. P. J. L., & Spirgel, A. (2010). Spacing and testing effects: A deeply critical, lengthy, and at times discursive review of the literature. In Ross, B.H. (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 63–147) Elsevier Academic Press, San Diego, CA.
- Dirkx, K.J.H., Thoma, G.B., Kester, L., & Kirschner, P.A. (2014). Answering questions after initial study guides attention during restudy. *Instructional Science*, 1–13.
doi:10.1007/s11251-014-9330-9
- Dunlosky, J., & Hertzog, C. (1998). Training programs to improve learning in later adulthood: Helping older adults educate themselves. In D.J. Hacker, J. Dunlosky, & A.C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 249–276). Mahway, NJ: Erlbaum.

- Finn, B., & Roediger, H.L. (2013). Interfering effects of retrieval in learning new information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(6), 1665-1681. doi:10.1037/a0032377
- Gordon, L.T., & Thomas, A.K. (2014). Testing potentiates new learning in the misinformation paradigm. *Memory and Cognition*, 42(2), 186-197. doi:10.3758/s13421-013-0361-2
- Grimaldi, P.J., & Karpicke, J.D. (2012). When and why do retrieval attempts enhance subsequent encoding? *Memory and Cognition*, 40(4), 505-513. doi:10.3758/s13421-011-0174-0
- Huff, M.J., Davis, S.D., & Meade, M.L. (2013). The effects of initial testing on false recall and false recognition in the social contagion of memory paradigm. *Memory and Cognition*, 41(6), 820-831. doi:10.3758/s13421-013-0299-4
- Izawa, C. (1970). Optimal potentiating effects and forgetting-prevention effects of tests in paired-associate learning. *Journal of Experimental Psychology*, 83, 340-344. doi: 10.1037/h0028541
- Jang, Y., & Huber, D.E. (2008). Context retrieval and context change in free recall: Recalling from long-term memory drives list isolation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(1), 112-127. doi:10.1037/0278-7393.34.1.112
- Kang, S.H.K. (2010). Enhancing visuo-spatial learning. The benefit of retrieval practice. *Memory and Cognition*, 38, 1009-1017. doi:10.3758/MC.38.8.1009

- Kang, S.H.K., & Pashler, H. (2014). Is the benefit of retrieval practice modulated by motivation? *Journal of Applied Research in Memory and Cognition*, *3*, 183-188. doi:10.1016/j.jarmac.2014.05.006
- Koriat, A., & Bjork, R.A. (2005). Illusions of competence in monitoring one's knowledge during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 187-194. doi:10.1037/0278-7393.31.2.187
- Koriat, A., Bjork, R.A., Sheffer, L., & Bar, S.K. (2004). Predicting one's own forgetting: The role of experience-based and theory-based processes. *Journal of Experimental Psychology: General*, *133*, 643-656. doi:10.1037/0096-3445.133.4.643
- Kornell, N., Hays, M.J., Bjork, R.A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(4), 989-998. doi:10.1037/a0015729
- Kornell, N., & Metcalfe, J. (2006). Study efficacy and the region of proximal learning framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 609-622. doi:10.1037/0278-7393.32.3.609
- LaPaglia, J.A., & Chan, J.C.K. (2012). Retrieval does not always enhance suggestibility: Testing can improve eyewitness identification performance. *Law and Human Behavior*, *36*, 478-487. doi:10.1037/h0093931
- Masson, M.E.J., Bub, D.N., Woodward, T.S., Chan, J.C.K. (2003). Modulation of word-reading processes in task switching. *Journal of Experimental Psychology: General*, *132*(2), 400-418. doi:10.1037/0096-3445.132.3.400

- Meiran, N. (1996). Reconfiguration of processing mode prior to task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(6), 1423-1442. doi:10.1037/0278-7393.22.6.1423
- McCabe, J. (2011) Metacognitive awareness of learning strategies in undergraduates. *Memory and Cognition*, 39(3), 462-476. doi:10.3758/s13421-010-0035-2
- Mensink, G., & Raaijmakers, J. G. (1988). A model for interference and forgetting. *Psychological Review*, 95(4), 434-455. doi:10.1037/0033-295X.95.4.434
- Metcalfe, J. (2002). Is study time allocated selectively to a region of proximal learning? *Journal of Experimental Psychology: General*, 131, 349-363. doi:10.1037/0096-3445.131.3.349
- Metcalfe, J., & Kornell, N. (2003). The dynamics of learning and allocation of study time to a region of proximal learning. *Journal of Experimental Psychology: General*, 132, 530-542. doi:10.1037/0096-3445.132.4.530
- Metcalfe, J., & Kornell, N. (2005) A region of proximal learning model of study time allocation. *Journal of Memory and Language*, 52, 463-477. doi:10.1016/j.jml.2004.12.001
- Miner, M. & Park, D.C. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers*, 36, 630-633. doi:10.3758/BF03206543
- Nelson, T.O., & Narens, L. (1994). *Why investigate metacognition?* In J. Metcalfe, & A.P. Shimamura (Eds), *Metacognition: Knowing about knowing* (pp. 1-25). Cambridge, MA: The MIT Press.

- Pansky, A., & Tenenboim, E. (2011). Inoculating against eyewitness suggestibility via interpolated verbatim vs. gist testing. *Memory and Cognition*, 39(1), 155-170.
doi:10.3758/s13421-010-0005-8
- Pastötter, B., Schicker, S., Niedernhuber, J., & Bäuml, K.H.T. (2011). Retrieval during learning facilitates subsequent memory encoding. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 37, 287–297.
doi:10.1037/a0021801
- Pastötter, B., Weber, J., Bäuml, K.H.T. (2013). Using testing to improve learning after severe traumatic brain injury. *Neuropsychology*, 27(2), 280-285.
doi:10.1037/a0031797
- Pastötter B., and Bäuml, K.H.T. (2014) Retrieval practice enhances new learning: the forward effect of testing. *Frontiers in Psychology*. 5(286). doi: 10.3389/fpsyg.2014.00286
- Raaijmakers, J.G. (2008) Mathematical Models of Human Memory. In H. L. Roediger, III (Ed.), *Cognitive Psychology of Memory. Vol. 2 of Learning and Memory: A Comprehensive Reference*, pp. [445-466] Oxford: Elsevier.
- Raaijmakers, J. G., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, 88(2), 93-134. doi:10.1037/0033-295X.88.2.93
- Rowland, C.A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, 140(6), 1432-1463.
doi:10.1037/a0037559

Szpunar, K.K., Khan, N.Y., & Schacter, D.L. (2013). Interpolated memory tests reduce mind wandering and improve learning of online lectures. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 6313-6317.

Szpunar, K.K., McDermott, K.B., & Roediger, H.L. (2008). Testing during study insulates against the buildup of proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(6), 1392-1399.
doi:10.1037/a0013082

Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(4), 1024-1037.
doi:10.1037/0278-7393.25.4.1024

Tulving, E., & Watkins, M.J. (1974). On negative transfer: Effects of testing one list on the recall of another. *Journal of Verbal Learning and Verbal Behavior*, *13*(2), 181-193. doi:10.1016/S0022-5371(74)80043-5

Weinstein, Y., Gilmore, A.W., Szpunar, K.K., & McDermott, K.B. (2014). The role of test expectancy in the build-up of proactive interference in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(4), 1039–1048. doi:10.1037/a0036164

Wahlheim, C. N. (2015). Testing can counteract proactive interference by integrating competing information. *Memory & Cognition*, *42*, 27-38. doi:10.3758/s13421-014-0455-5

Weinstein, Y., Gilmore, A. W., Szpunar, K. K., & McDermott, K. B. (2014). The role of test expectancy in the build-up of proactive interference in long-term memory.

- Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(4), 1039-1048. doi:10.1037/a0036164
- Weinstein, Y., McDermott, K. B., & Szpunar, K. K. (2011). Testing protects against proactive interference in face–name learning. *Psychonomic Bulletin & Review*, 18(3), 518-523. doi:10.3758/s13423-011-0085-x
- Wilford, M.M., Chan, J.C.K., & Tuhn, S.J. (2014). Retrieval enhances eyewitness suggestibility to misinformation in free and cued recall. *Journal of Experimental Psychology: Applied*, 20, 81–93. doi:10.1037/xap0000001
- Wissman, K. T., Rawson, K. A., & Pyc, M. A. (2011). The interim test effect: Testing prior material can facilitate the learning of new material. *Psychonomic Bulletin & Review*, 18(6), 1140-1147. doi:10.3758/s13423-011-0140-7

Table 1

Key Methodological Differences Between Experiments 1-5

Experiment	Feedback Duration	Priority Instructions	Intervening Task	Intervening Task Order	Main Finding
1	2 s	None	Intermixed	Name First	Impairment
2	2 s	None	Intermixed	Profession First	Impairment
3	5 s	None	Intermixed	Name First	No Difference
4	2 s	None	Blocked	Name First	Enhancement
5	2 s	Name or Profession	Intermixed	Name First	No Difference

Table 2

Final Profession Performance for the Restudy Condition and Test Conditions Based on Initial Retrieval Success in Experiments 1-5.

	Restudy	Failed Retrieval	Successful Retrieval
Experiment 1	.58 (.19)	.43 (.17)	.55 (.29)
Experiment 2	.54 (.17)	.26 (.10)	.34 (.28)
Experiment 3	.52 (.19)	.39 (.21)	.57 (.29)
Experiment 4	.54 (.18)	.66 (.21)	.67 (.24)
Experiment 5			
Name Priority	.54 (.20)	.45 (.18)	.51 (.21)
Profession Priority	.58 (.20)	.54 (.22)	.61 (.31)

Note: Standard deviations are in parentheses.

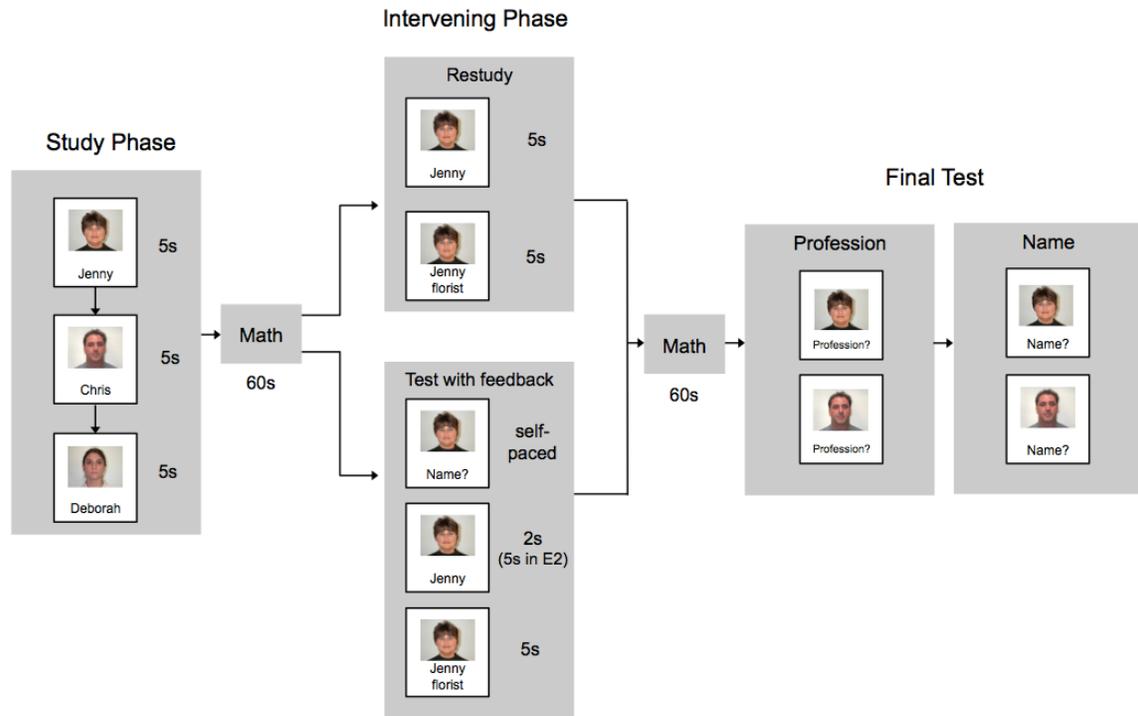


Figure 1. A graphical depiction of the procedure used in Experiments 1-5.

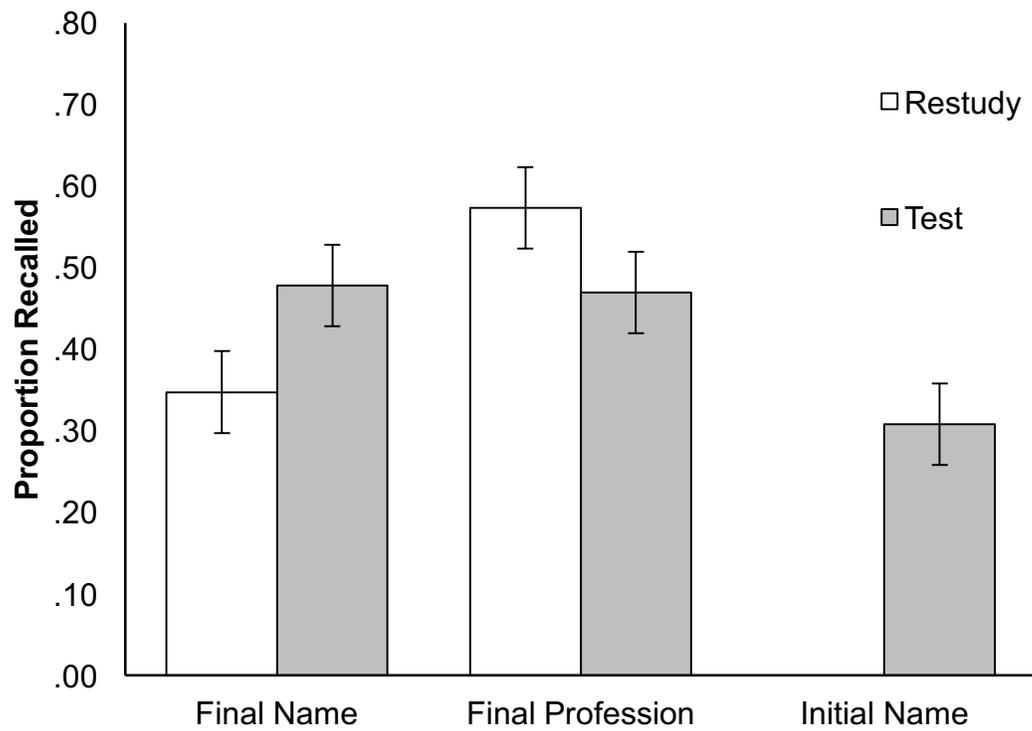


Figure 2. Final test performance for professions and names and initial test performance for names in Experiment 1. Error bars are between-subjects 95% confidence intervals.

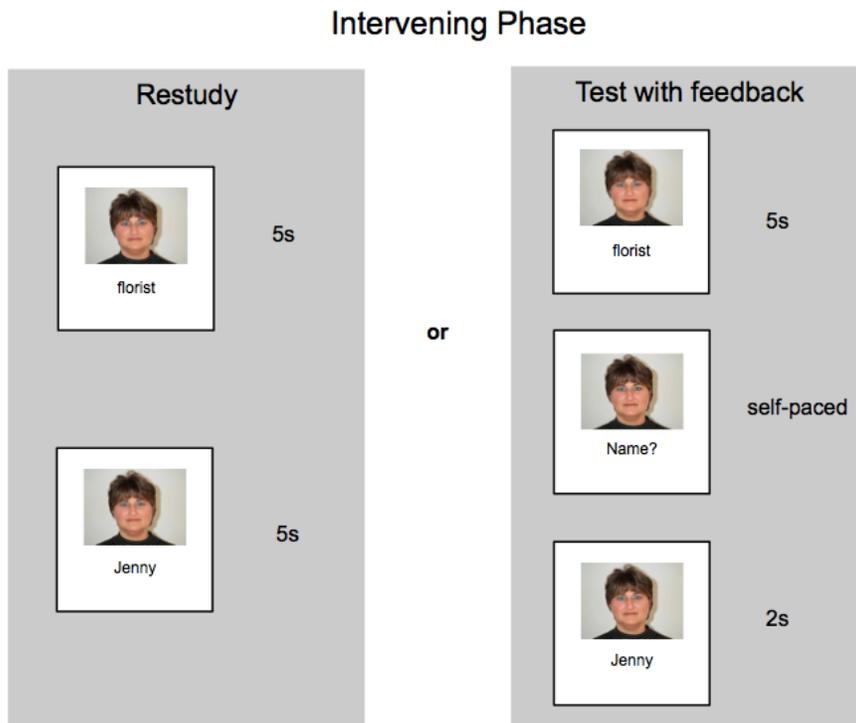


Figure 3. A graphical depiction of the intervening phase used in Experiment 2.

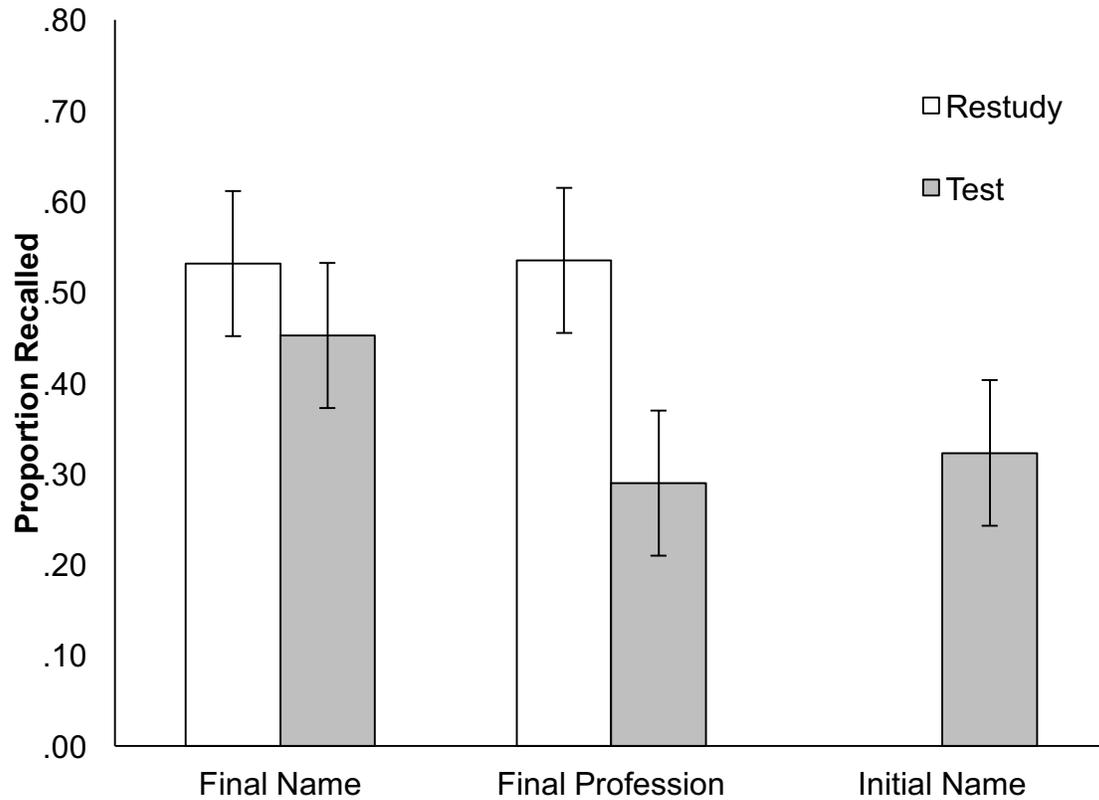


Figure 4. Final test performance for professions and names and initial test performance for names in Experiment 2. Error bars are between-subjects 95% confidence intervals.

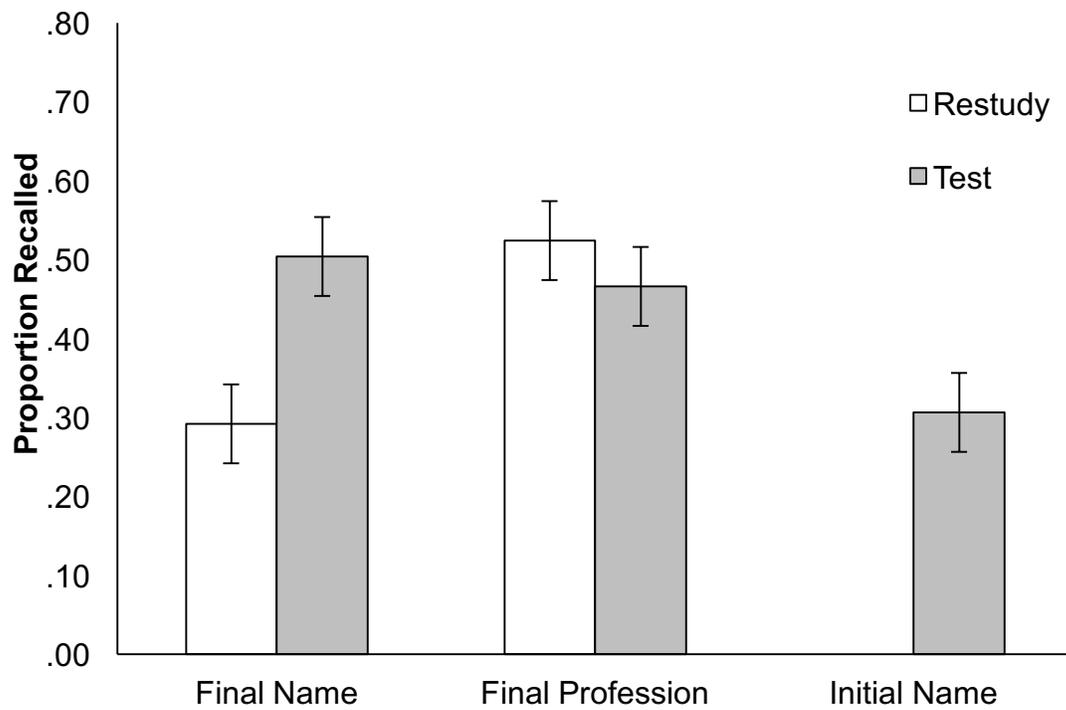


Figure 5. Final test performance for professions and names and initial test performance for names in Experiment 3. Error bars are between-subjects 95% confidence intervals.

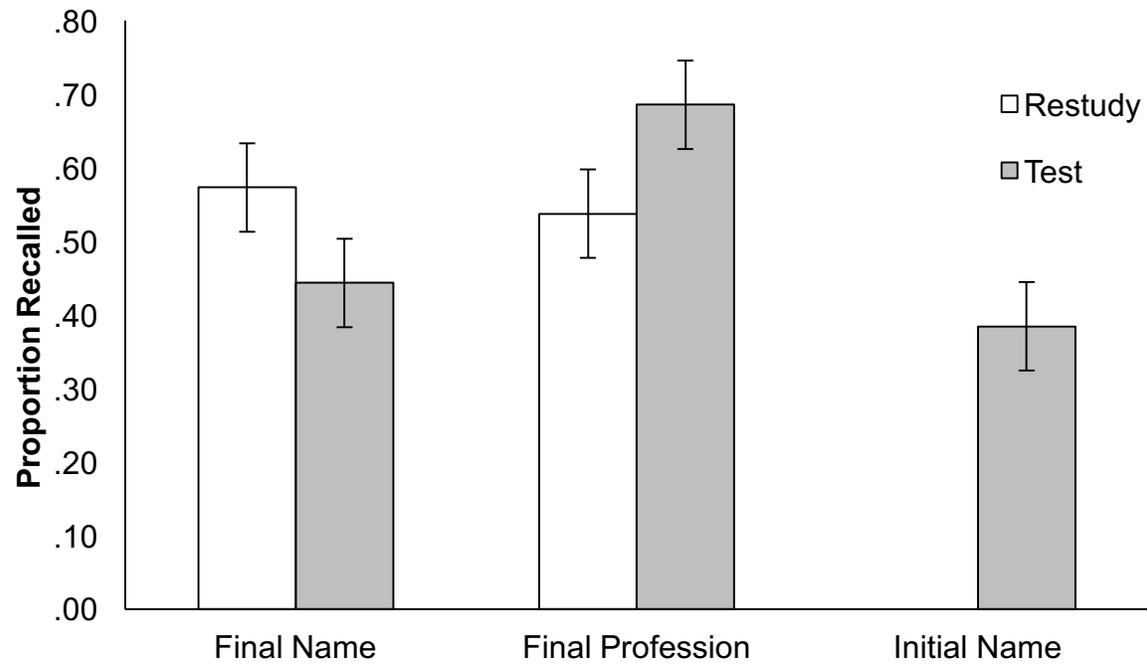


Figure 6. Final test performance for professions and names and initial test performance for names in Experiment 4. Error bars are between-subjects 95% confidence intervals.

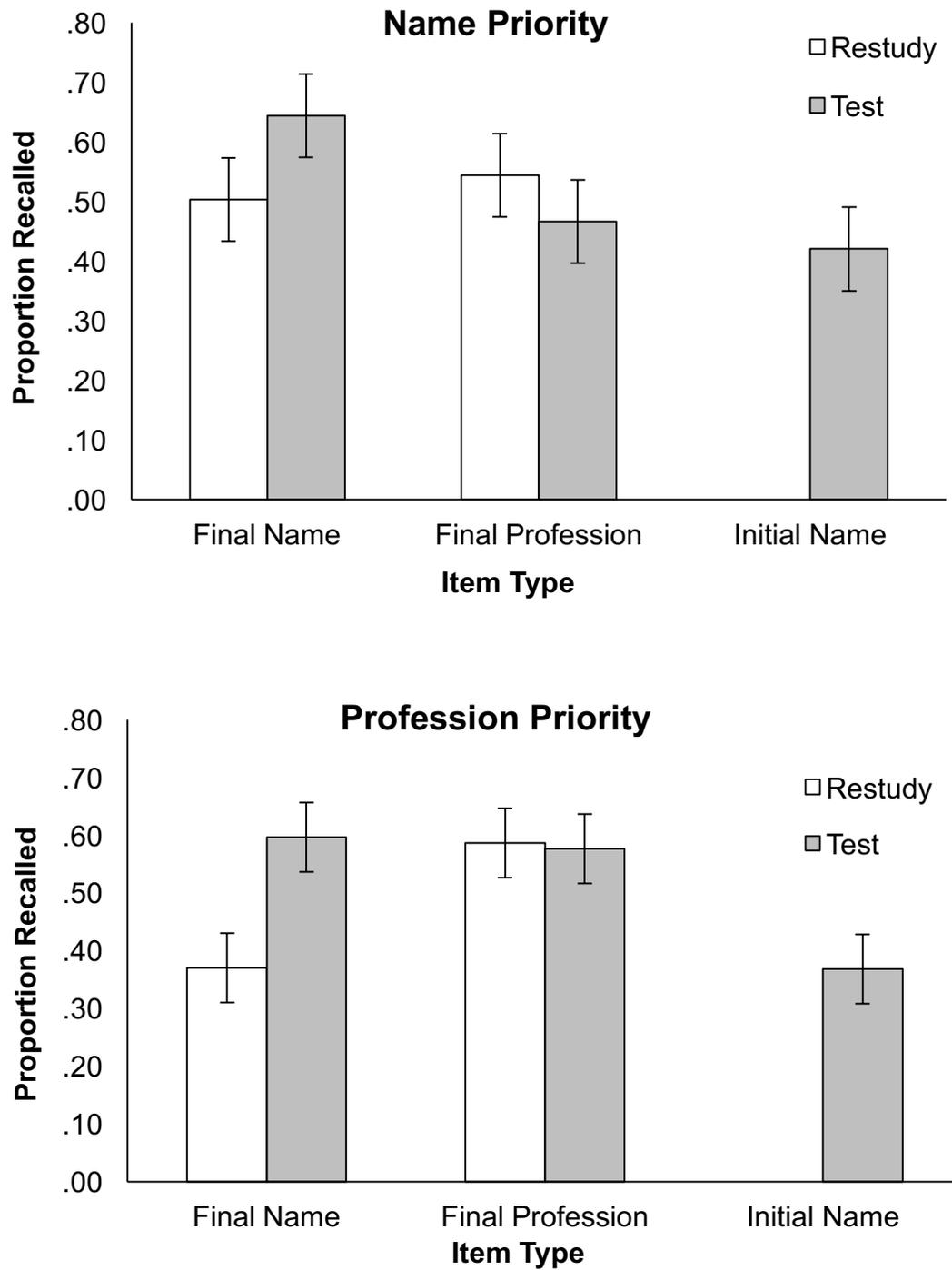


Figure 7. Final test performance for professions and names and initial test performance for names in Experiment 5. The top panel shows results for participants in the name priority condition and the bottom panel shows results for participants in the profession priority condition. Error bars are between-subjects 95% confidence intervals.

Appendix

Data From a Pilot Experiment Where Participants Were Told to Prioritize Names or Professions With No Incentive.

	Final Name	Final Profession	Initial Test
<hr/>			
Name Priority			
Restudy (n=17)	.33 (.17)	.57 (.17)	
Test (n=7)	.27 (.22)	.32 (.10)	.32 (.30)
Profession Priority			
Restudy (n=13)	.38 (.19)	.39 (.22)	
Test (n=15)	.48 (.19)	.29 (.18)	.20 (.17)

Note: Standard deviations are in parentheses.