Journal of Memory and Language 103 (2018) 19–27

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Testing enhances memory for context

Melisa Akan[a,*], Sarah E. Stanley[a], Aaron S. Benjamin[a,b]

[a] University of Illinois at Urbana-Champaign, United States
[b] Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, United States

ARTICLE INFO

Keywords:
Testing effect
Context memory
Episodic retrieval

ABSTRACT

The beneficial effect of retrieval practice on memory is a well-established phenomenon. Despite the wealth of research on this *testing* effect, it is unclear whether the benefits of testing extend beyond the tested information to include memory for the context in which the memoranda were encountered. Three experiments examined the effect of testing on memory for context using a standard variant of a traditional item-context memory task, in which cue-target word pairs (the items) were presented on the computer screen in varying locations (the contexts). All experiments revealed an enhancement to memory for context following retrieval practice of the items, regardless of whether that retrieval took place in a neutral (Experiments 1 and 2) or in an interfering (Experiment 3) location. These results support the view that retrieval practice elicits retrieval of relatively comprehensive prior episodes, rather than of only semantic aspects of the prior episodes relevant to the practice cues.

Introduction

Retrieval of information from memory is a powerful means of enhancing long-term retention (Bjork, 1975). It is often more effective than additional study of the same information, a phenomenon called the *testing effect* (Roediger & Karpicke, 2006a; see Nunes & Karpicke, 2015; Rowland, 2014, for recent reviews). The benefits of testing have been demonstrated both in the lab and in classroom settings, using a variety of learning materials, including prose passages (e.g., Roediger & Karpicke, 2006b), single words (e.g., Hogan & Kintsch, 1971), paired associates (Carrier & Pashler, 1992), as well as nonverbal material (e.g., Wheeler & Roediger, 1992). Testing has even been promoted for wider use within educational settings as a means of enhancing, and not simply assessing, knowledge (Benjamin & Pashler, 2015; Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Pashler et al., 2007).

These important applications notwithstanding, there is lack of a consensus within the field as to what actually causes the benefits of retrieval practice on memory. Some theoretical positions include a prominent role for the episodic context of the original encoding, as well as of the retrieval practice event (e.g., Karpicke, and Lehman, & Aue, 2014; Lehman, Smith, & Karpicke, 2014), whereas others include no role for context (e.g., Carpenter, 2009) or are mute to its effects (e.g., Bjork, 1975; Kornell, Bjork, & Garcia, 2011). Within those theories that do allow a role for some kind of context, there are ones that attribute similarity between study and test circumstances as key (e.g., via transfer-appropriate processing, Landauer & Bjork, 1978; Roediger & Karpicke, 2006a, 2006b) and others in which variability across contexts is important (Karpicke et al., 2014; McDaniel & Masson, 1985). Clearly, experiments directly assessing the degree to which memory for contextual elements is enhanced, retarded, or unaffected by testing will be central to developing a thorough understanding of the causes of the testing effect. Here we report three experiments using traditional tests of context memory within a testing-effect paradigm, and demonstrate consistent enhancement to memory for context following retrieval practice. These benefits persist even when that retrieval practice introduces a context that would be expected to interfere with memory for the original encoding context.

A point that is highly relevant to the applied potential of testing effects is that the benefits of testing sometimes extend beyond the tested information itself to include conceptually related but nontested information presented in the same episode with the tested information (e.g., Butler, 2010; Carpenter, Pashler, & Vul, 2007; Chan, 2009, 2010; Chan, McDermott, & Roediger, 2006; but see Pan, Gopal, & Rickard, 2015). With semantically related materials it can be hard to tell whether such benefits reflect the incidental retrieval of untested aspects of the material or complex knock-on effects of enhancing memory for the tested material. For this reason, it is critical to evaluate this question using materials for which the untested elements are purely episodically related and devoid of larger meaning, thereby minimizing influences of semantic encoding and retrieval strategies. However, there are only a

* Corresponding author at: Department of Psychology, University of Illinois at Urbana-Champaign, 603 E. Daniel St., Champaign, IL 61820, United States.
E-mail address: makan2@illinois.edu (M. Akan).

https://doi.org/10.1016/j.jml.2018.07.003
Received 10 January 2018; Received in revised form 3 July 2018

handful of studies that examined whether the benefits of testing extend to contextual information under such circumstances, despite much research indicating a crucial role for context in episodic retrieval (Divis & Benjamin, 2014; Howard & Kahana, 2002; Jang & Huber, 2008; Lehman & Malmberg, 2013). Currently, there is no conclusive evidence as to whether the testing effect generalizes to memory for incidental source or context.

In one relevant study, Rowland and DeLosh (2014) found that the benefits of testing were not limited to untested items that were semantically related to the tested items, but also generalized to untested items that had no designated association with the tested items other than being presented as part of the same list. There is also evidence that participants are better able to identify the list membership of previously studied items following testing (Brewer, Marsh, Meeks, Clark-Foos, & Hicks, 2010; Chan & McDermott, 2007; Verkoeijen, Tabbers, & Verhage, 2011), and that tested materials are more likely to elicit a *Remember* response in the remember/know paradigm (Jones & Roediger, 1995). However, neither of these data are dispositive. List membership is only a very coarse measure of context and is confounded with recency of exposure. And *remember* judgments may not accurately reveal retrieval of contextual information (Benjamin, 2005; Dunn, 2008).

To our knowledge, there is only one published study (Brewer et al., 2010) that examined whether testing leads to an enhancement in memory for context beyond episodic temporal information. In that task, participants studied two lists of words, and each word was presented in either a male or female voice. Both lists were followed by either retrieval practice (free recall) or a math distractor task. In the final test, participants indicated, for each studied item, either whether it had been presented in list 1 or in list 2, or whether the word had been spoken by a female or by a male voice. The results revealed an enhancement in memory for list membership but not for speaker gender. However, in another experiment, when participants were asked to also additionally indicate gender source information as they recalled each word during retrieval practice, testing also enhanced gender discrimination performance on the final test. These results would seem to indicate that temporal information is naturally accessed in the processes underlying cued recall, but that other contextual aspects of the original presentation are not unless the retrieval cue specifically promotes their involvement. Such a claim is buttressed by findings that temporal information is often automatically encoded, even under incidental learning conditions (Hintzman & Block, 1971; Proctor & Ambler, 1975). Yet, in contradiction with this claim, an unpublished thesis by Rowland (2011) reported that retrieval practice resulted in a small advantage in recalling which of the two possible colors a word was presented in (Experiment 1), and the order in which the individual words in semantically unrelated word pairs were presented (Experiment 2), even though the retrieval practice did not involve reporting either of these details.

In short, the few studies that directly examined memory for contextual information have not provided conclusive evidence. The present experiments used a time-uncorrelated contextual dimension and used more varied contextual information than previous work—the items could appear in one of eight possible locations on the screen. The spatial configuration of these screen locations was circular, as shown in Fig. 1a. Unlike a linear display, a circular arrangement mitigates against easy translation into a temporal code (Fischer-Baum & Benjamin, 2014; Hitch, 1974). Spatial information would seem to be more difficult to associate with words semantically than a gendered voice or temporal information (which allows for story-building and retrieval strategies that involve seriation), and may not be encoded automatically (Naveh-Benjamin, 1987, 1990). We directly compared testing with both a restudy condition and a control condition in which items did not receive any additional exposure.

Experiments 1A and 1B

Experiments 1A and 1B sought to test if the benefits of retrieval practice would extend to contextual details by having participants study word pairs presented in different locations on the screen. During the study session, words were presented in one of eight possible locations, and the review (either restudy or retrieval practice) occurred in the center of the screen. The stimuli in these experiments were low-association word pairs. Participants were asked to study the word pairs and were informed that there would be a later test in which they would be given the first word of the pair (the cue) and asked to provide the second word in the pair (the target). The only difference between Experiment 1A and Experiment 1B is the number of studied word pairs—96 and 48 respectively. The reason for reducing the number of word pairs for Experiment 1B was a concern over potential floor effects during the process of collecting data for Experiment 1A. Rather than starting over, we continued collecting data for Experiment 1, after reducing the number of items, until we achieved our planned sample size. This was determined to be n = 52 to achieve 80% power to detect an effect size of d = 0.40 for a paired-sample *t*-test.[1] As we did not know the effect size for retrieval practice on memory for context, we made a conservative estimate based on prior work on the effect of retrieval practice on item memory (see Rowland, 2014 for a meta-analysis).

Method

Participants

Twenty-nine undergraduate students from the University of Illinois at Urbana-Champaign (UIUC) participated in Experiments 1A and 1B, each, in partial fulfillment of a course requirement. Three participants in Experiment 1A had incomplete data and were excluded from analyses (two due to a failure to attend Day 2 of the experiment, and one due to computer difficulties). Four participants in Experiment 1B had incomplete data and were excluded from analysis (all due to a failure to attend Day 2 of the experiment). For all experiments reported in this paper except Experiment 2B, demographic information collected from participants was not connected to the particular experiments that they participated in. Here we provide the overall demographic profile of the subject pool from which the participants were drawn. Participants from this pool ranged from 18 to 35 years of age, and 91% of the participants were between the ages 18 and 21. Females constituted 63% of the subject pool and the percentage of native speakers was 78%.

Materials

Ninety-six weakly associated word pairs (cue to target association of 0.028–0.030) were selected from the University of South Florida Free Association Norms database (Nelson, McEvoy & Schreiber, 2004). For Experiment 1B, only 48 of the original 96 word pairs were used. We reduced the number of study pairs in Experiment 1B; the experiments were otherwise identical. The materials are included in the raw data files, which can be accessed online on our main project page at Open Science Framework (OSF; https://osf.io/bqr5f/).

Design

The experiment used a 3 (Review Type) × 2 (Type of Final Test) within-subject design. The three review types consisted of retrieval practice, restudy, and a control condition of no review. The final test was either a cued recall task that required retrieval of the target item given the cue item, or an 8-alternative-forced-choice (8-AFC) test on memory for the word location context. All conditions had an equal number of word pairs. Both review condition and test condition were manipulated between-item (i.e., no item was reviewed or tested more

[1] All power analyses were performed using G*Power (Faul, Erdfelder, Buchner, & Lang, 2009).