Journal of Memory and Language 92 (2017) 114–127

Contents lists available at ScienceDirect

Journal of Memory and Language

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

Late consequences of early selection: When memory monitoring backfires

Katarzyna Zawadzka *, Maciej Hanczakowski, Edward L. Wilding [1]

School of Psychology, Cardiff University, UK

ARTICLE INFO

Article history:
Received 18 September 2015
revision received 31 May 2016

Keywords:
Retrieval orientation
Monitoring
Memory for foils

ABSTRACT

At retrieval, people can adopt a retrieval orientation by which they recreate the mental operations used at encoding. Monitoring by retrieval orientation leads to assessing all test items for qualities related to the encoding task, which enriches foils with some of the qualities already possessed by targets. We investigated the consequences of adopting a retrieval orientation under conditions of repeated monitoring of the same foils. Participants first processed foils in the context of one of two tests encouraging different retrieval orientations. The foils were then re-used on a subsequent test in which retrieval orientation either matched or mismatched that adopted on the first test. In the aggregate data, false alarms for repeated foils were higher when there was a match between the retrieval orientations on both tests. This demonstrates that when retrieval orientation enriches foils with target-like characteristics, it can backfire when repeated monitoring of the same foils is required.

Introduction

When asked to retrieve some information from memory, people can employ a variety of monitoring strategies to improve the quality of their memory report. One strategy is to mentally recreate the operations performed at the time of encoding. This mentally recreated information becomes embedded in the retrieval cue and every item in a memory test is then assessed with respect to the degree of match with this cue. Since only studied items are associated with diagnostic details now embedded in the retrieval cue, this form of monitoring allows for effective rejection of non-studied items (foils). This monitoring strategy can be viewed as a consequence of having adopted a *retrieval orientation* (e.g., Gray & Gallo, 2015; Herron & Rugg, 2003a; Pierce & Gallo, 2011; Rugg & Wilding, 2000) or as an example of *early selection* (e.g., Guzel & Higham, 2013; Jacoby, Kelley, & McElree, 1999).[2]

Research conducted to date has shown the benefits of applying early selection mechanisms for memory reports (e.g., Bridger, Herron, Elward, & Wilding, 2009; Pierce & Gallo, 2011). The present study breaks with this tradition by delineating the conditions in which the use of such a monitoring strategy comes at a cost when repeated monitoring of the same foils is required.

Evidence for monitoring by retrieval orientation comes from two strands of research which, although distinct, share a common approach: they infer the operation of a monitoring strategy from the ways in which foils are

* Corresponding author at: Division of Psychology, Nottingham Trent University, Burton Street, Nottingham NG1 4BU, UK.

E-mail address: katarzyna.zawadzka@ntu.ac.uk (K. Zawadzka).

[1] Present address: School of Psychology, University of Nottingham, UK.

[2] Other terms that have been used to describe this kind of monitoring process include *front-end control* (e.g., Halamish, Goldsmith, & Jacoby, 2012), and *source-constrained retrieval* (e.g., Alban & Kelley, 2012; Jacoby, Shimizu, Daniels, & Rhodes, 2005; Jacoby, Shimizu, Velanova, & Rhodes, 2005). Generally, early selection monitoring strategies are contrasted with strategies referred to as *late correction* (Jacoby et al., 1999), such as the distinctiveness heuristic (Dodson & Schacter, 2002; Hanczakowski & Mazzoni, 2011) or response withholding (Koriat & Goldsmith, 1996), which operate on the information already retrieved from memory.

http://dx.doi.org/10.1016/j.jml.2016.06.003

processed at test. Both strands capitalize on the premise that as non-studied foils are by definition unaffected by the study phase, any difference between the processing of the foils must be caused by the monitoring strategy adopted at test. The first strand of research uses measures of neural activity such as event-related potentials (ERPs). It is possible to compare ERPs elicited by foils in two tasks differing with respect to the kind of information that needs to be retrieved in order to answer a memory question. The differences between ERPs are assumed to demonstrate the operation of distinct retrieval orientations depending on the type of queried information. Past research has shown differences between ERPs elicited by new items when study items were presented as pictures versus words (e.g., Herron & Rugg, 2003a; Robb & Rugg, 2002), were studied with a pleasantness versus an animacy judgment (Herron & Wilding, 2004, 2006), or with a shallow versus deep processing task (Rugg, Allan, & Birch, 2000), among others. These studies clearly demonstrate that non-studied foils are processed differently under various retrieval orientations.

The second strand of research uses behavioral methods to gain insights into when and how monitoring by retrieval orientation is employed. In the *memory-for-foils paradigm* (Jacoby, Shimizu, Daniels, & Rhodes, 2005; Jacoby, Shimizu, Velanova, & Rhodes, 2005; Shimizu & Jacoby, 2005), participants first learn a list of words with two different orienting tasks: one deep (for example, a pleasantness judgment task) and one shallow (e.g., counting the number of letters or vowels in each studied word). Following the study phase, they are given two old/new recognition tests. On the deep test, only deeply processed words are presented among foils. On the shallow test, only words studied with the shallow task are among foils. Finally, an additional memory test for unstudied foils is administered. The final test list consists of three types of items: deep foils (foils presented on the deep test), shallow foils (foils presented on the shallow test), and new words (not presented on any of the tests). Participants are instructed to distinguish new words from those that were presented earlier during any phase of the experiment. The main finding in this paradigm is a task-dependent difference in correct endorsements for previously encountered foils: deep foils are more often indicated as having been seen during the course of the experiment than shallow foils. This is taken as evidence that the retrieval orientation adopted on the deep test benefits the subsequent memorability of the foils on that task to a greater degree than the orientation adopted on the shallow test. Crucially, this line of research goes beyond demonstrating that non-studied foils are subjected to different assessments. The novel focus here is on the *consequences* of adopting a retrieval orientation: Jacoby, Shimizu, Daniels, and Rhodes (2005) argued that the assessment of foils with the use of retrieval orientation on the deep test enriches these foils with diagnostic information embedded in the retrieval cue.

The findings in subsequent studies support the explanation that the better memory for deep foils observed by Jacoby, Shimizu, Daniels, and Rhodes (2005) stems from their enrichment with details diagnostic of study. Marsh et al. (2009; see also Danckert, MacLeod, & Fernandes, 2011; Gray & Gallo, 2015) added a remember/know task (e.g., Gardiner, 1988; Tulving, 1985) to the final test for foils. Foils presented on a deep test were later assigned 'remember' responses more often than foils first encountered on a shallow test. This is consistent with the assumption that on deep tests new words are deeply processed in the context of the orienting task. Danckert et al. (2011) substituted the final test for foils with a second judgment phase, in which the task was to perform on foils the same judgments that were earlier made for targets at study. They predicted that if participants processed foils in the context of a retrieval orientation on the deep test, then a judgment had already been made for deep foils when they were first presented. If this was the case, then a deep orienting task during the second judgment phase would be completed faster for deep than for shallow foils. The results were consistent with that prediction. Recently, Gray and Gallo (2015) demonstrated that the deep > shallow difference in memory for foils occurs at all levels of foil strength, ruling out an alternative explanation that this effect is due to a post-retrieval monitoring process employed specifically for items yielding ambiguous evidence that does not allow a determination of whether an item was studied or not.

The research conducted to date allows a clear conclusion: monitoring by early selection can change the way foils are processed in a memory test. Specifically, foils are considered in light of the adopted retrieval orientation which, if the test requires deep processing, leads to their enrichment with the details that are embedded in the particular retrieval cue. In other words, deeply processed foils become associated with the details that are diagnostic of previous study. If a memory test for foils is later given, such enriched foils are remembered better than foils that were monitored on a shallow test.

However, if monitoring by retrieval orientation can enrich foils with features diagnostic of earlier study, it means that foils may start to resemble studied items. A straightforward question thus arises: what would be the consequences of early selection if the same retrieval orientation was applied twice to the same foils? If applying a retrieval orientation makes foils more similar to targets, would people mistake these foils for targets on a subsequent test requiring the adoption of the same retrieval orientation? In other words, could adopting a potentially beneficial monitoring strategy at test ultimately lead to impairment in performance on a future test if the same to-be-rejected materials are assessed again?

We tested this hypothesis in four experiments by having participants complete two study-test blocks. The first block consisted of a single study phase followed by two test phases. In the study phase, single words were studied in two different deep encoding tasks. The test phases were both exclusion tasks. Each required endorsement of words studied in one of the two encoding tasks, and rejection of new words as well as those studied in the alternate encoding task. The encoding task associated with words requiring endorsement was changed across the two exclusion test phases.

The second study-test block had the same structure. The study phase was as for the first block, and all of the words presented were new to the experiment. The two