Available online at www.sciencedirect.com
ScienceDirect
Journal of Memory and Language 56 (2007) 592–607
Journal of Memory and Language
www.elsevier.com/locate/jml

Metamnemonic control over the discriminability of memory evidence: A signal detection analysis of warning effects in the associative list paradigm

Jeffrey J. Starns *, Sean M. Lane *, Jill D. Alonzo, Cristine C. Roussel

Department of Psychology, Louisiana State University, Baton Rouge, LA 70810, USA

Received 24 February 2006; revision received 8 August 2006
Available online 11 October 2006

Abstract

According to signal detection theory (SDT), retrieval warnings may decrease false memory in the associative list paradigm either by inducing a conservative criterion shift or by decreasing the amount of evidence that critical theme words were studied. Fitting a SDT model to 12 existing datasets revealed suggestive evidence that warnings impact critical theme evidence, and two new experiments confirmed this conclusion. We argue that this pattern of results is consistent with warned participants relying less on relational and more on item-specific forms of information at retrieval as compared to unwarned participants. We conclude that warnings enhance metamnemonic awareness, thus allowing participants to select a retrieval strategy that capitalizes on discriminative forms of evidence.

Keywords: Signal detection theory; False memory; Warning

Memories sometimes do not correspond to any objective past experience, and a wealth of memory research has explored the factors contributing to false memories (Roediger, 1996). The associative-list paradigm (also known as the "DRM" paradigm; Deese, 1959; Roediger & McDermott, 1995) has been adopted by many researchers as an effective way to induce false memories in a laboratory setting. In this paradigm, participants study lists of words that are organized in terms of their shared association to a non-presented word called the critical theme of the list. On a recognition test, participants are very likely to claim to have studied critical themes, and sometimes show little or no ability to discriminate non-presented critical themes from studied words (e.g., Roediger & McDermott, 1995).

One significant goal of false memory research is to identify strategies that people can use to avoid or edit false memories at retrieval. Some studies in the associative list paradigm have pursued this goal by warning participants about the nature of the paradigm just before a memory test (Anastasi, Rhodes, & Burns, 2000; Gallo, Roberts, & Seamon, 1997; Gallo, Roediger, & McDermott, 2001; McCabe & Smith, 2002; Neuschatz, Payne, Lampinen, & Toglia, 2001). Warnings typically inform participants that the lists that they studied were highly associated to non-presented words, and that they should be careful not to falsely remember these words. It is also typical to provide participants with an example of an associative list and the non-presented critical theme. Warnings enhance participants' metamnemonic

* Corresponding authors. Fax: +1 225 578 4125.
E-mail addresses: starns.4@osu.edu (J.J. Starns), slane@lsu.edu (S.M. Lane).

0749-596X/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2006.08.013

knowledge regarding both the task presented to them (i.e., that the recognition test will contain non-presented words that are highly associated to many studied items) and a property of their memory systems (i.e., that they are likely to experience illusory memories for associated words). Whereas unwarned participants are likely to greatly underestimate the difficulty of a test requiring the discrimination of studied items from critical theme lures, warned participants have the opportunity to form more realistic expectations about the test. Warned participants may be able to translate their increased metamnemonic awareness into a more effective retrieval strategy, resulting in a heightened ability to distinguish true from false memories.

The results of prior research employing a retrieval warning have been somewhat mixed. Some studies suggest that warnings are only effective when they are provided before encoding, but not after encoding and before retrieval (e.g., Gallo et al., 1997). Other studies show a robust effect of retrieval warnings on false recognition (e.g., McCabe & Smith, 2002). Identifying the factors that contribute to warning effectiveness, as well as the mechanisms of false memory reduction following an effective warning, will provide useful information regarding how retrieval strategies promote or discourage false memories.

The purpose of this paper is to clarify the effects of warnings on false memory by applying an appropriate signal detection model. According to Signal Detection Theory (SDT; Wickens, 2002), recognition decisions are determined by both the evidence stored in memory and the decision processes that are used to translate this evidence into specific responses. The memory evidence used in recognition decisions is represented as a single continuous variable that is usually called familiarity.[1] Both targets and lures vary in the amount of familiarity they inspire, but targets are more familiar on average based on the memory evidence encoded for these items in the study phase. It is typically assumed that the familiarity values of both targets and lures are normally distributed with the mean of the target distribution above the mean of the lure distribution on the familiarity continuum. Also, the variance of the target distribution is regularly greater than the variance of the lure distribution (Glanzer, Adams, Iverson, & Kim, 1993). μ is the distance between the means of the target and lure distributions measured in terms of the lure distribution's standard deviation, and this parameter provides a measure of the gain in memory evidence resulting from the encoding episode. σ is the ratio of the target and lure distributions' standard deviations. When the target and lure distributions have equal standard deviations (i.e., $\sigma = 1$), μ is equal to the commonly used memory measure d'. To make recognition decisions, participants set a criterion for the amount of memory evidence they require to claim that an item was studied, and any test item that exceeds this criterion value receives a positive recognition response. The parameter λ expresses the distance of the response criterion from the mean of the lure distribution in terms of the lure distribution's standard deviation.

Applying SDT to the associative list paradigm necessitates a decision space containing three distributions for the three item types on the test (targets, lures, and critical themes; Wixted & Stretch, 2000). Thus, the model requires five parameters to describe recognition performance: μ_T, the distance between the unrelated lure and target distributions; σ_T, the standard deviation of the target distribution relative to the unrelated lure distribution; μ_{CT}, the distance between the unrelated lure and critical theme distributions; σ_{CT}, the standard deviation of the critical theme distribution relative to the unrelated lure distribution; and λ, the position of the response criterion. The μ_{CT} parameter reflects the gain in memory evidence for critical theme words as a result of the presentation of associates in the study phase. This model is graphically displayed in the top panel of Fig. 1. For simplicity, this figure displays a situation in which the standard deviations of all distributions are equal.

A signal detection analysis reveals that there are two ways that retrieval warnings can reduce false memory for critical theme words: warnings may lead to a higher response criterion or warnings may decrease the distance between the unrelated lure and critical theme distributions. A change in response criterion would indicate that warnings prompt participants to use more conservative standards for the evidence required to claim that a retrieval candidate was studied. A change in the position of the critical theme distribution would indicate that warned participants actually retrieve less evidence that critical themes were studied than do unwarned participants. Of course, warnings can simultaneously impact both memory evidence and response criteria, but we will separately consider the implications of a distribution shift and a criterion shift for clarity of exposition.

A criterion shift explanation is displayed in Fig. 1. The top panel in this figure shows recognition performance without a warning, and the bottom panel shows recognition performance following a warning. In the figure, a warning induces a shift in response criterion to a higher, more conservative value. This would result if, following a warning, participants decided to avoid errors by requiring a great deal of evidence to claim that an item was studied. In Fig. 1, a smaller proportion of

[1] We use the term familiarity only to refer to evidence on a single continuum (Wickens, 2002), not as an alternative to recollection as in dual-process theories (e.g., Yonelinas, 1997). Further, we assume that evidence of various types, some of which would be considered "recollective" (e.g., contextual detail), can be integrated into a single continuous variable for use in decision-making (see Wixted & Stretch, 2004).