Available online at www.sciencedirect.com
SCIENCE DIRECT®

Journal of Memory and Language 52 (2005) 595–617

Journal of Memory and Language

www.elsevier.com/locate/jml

Generation failure: Estimating metacognition in cued recall ☆

Philip A. Higham *, Helen Tam

School of Psychology, University of Southampton, Highfield, Southampton SO17 1BJ, UK

Received 9 August 2004; revision received 20 January 2005
Available online 23 February 2005

Abstract

Three experiments examined generation, recognition, and response bias in the original encoding-specificity paradigm using the type 2 signal-detection analysis advocated by Higham (2002). Experiments 1 (pure-list design) and 2 (mixed-list design) indicated that some guidance regarding the strength of the associative relationship between the test cue and target greatly improved strong-cue target production relative to no guidance, and that this effect was attributable to improved generation, as well as recognition. Problems with generating candidates for response during standard cued recall was further shown in Experiment 3, where despite having the opportunity to provide multiple responses for each cue, participants' ability to produce the targets remained poor. The results are discussed in terms of traditional and modern generate-recognize theory, metacognition, and dual-route models of recall.

Keywords: Cued recall; Generate-recognize; Metacognition; Encoding specificity

Since Tulving and colleagues introduced the encoding specificity principle in the early 1970s (e.g., Thomson & Tulving, 1970; Tulving & Thomson, 1973), most students of memory have viewed generate-recognize theory as a straw man. It is considered by many to be an old-fashioned theory, with a particular failing when it comes to explaining context reinstatement effects in cued recall. In this paper, we revisit both generate-recognize theory and the classic cued-recall paradigm that provided the initial support for the encoding-specificity principle. However, let us be clear at the outset that we are not attempting to resurrect traditional generate-recognize theory. Indeed, as will become apparent, the data from the experiments that we report are quite inconsistent with those early models, and, if anything, they support many aspects of Tulving's message. On the other hand, we will argue that a more modern generate-recognize model of cued recall that maintains the crucial distinction between memory access (generation) and metacognitive monitoring (recognition) processes is still a useful framework for cued-recall performance, and performance on many other tasks as well. In this way, our message is similar to that of other metacognitive researchers who have promoted two-stage models involving separate stages of access and memory

☆ Philip A. Higham and Helen Tam, School of Psychology, University of Southampton. Preparation of this article was supported by a research grant from the British Academy. Portions of this research were presented at the 43rd annual meeting of the Psychonomic Society, November, 2002 in Kansas City, Missouri, USA, and at the 45th annual meeting of the Psychonomic Society, November, 2004 in Minneapolis, Minnesota, USA. We thank Nina Eskriett, Wendy Kneller and David Brook for research assistance. We thank Harry Bahrick, Chuck Brainerd, Morris Goldsmith, Asher Koriat, Steve Lindsay, Doug Nelson, and Mike Watkins for helpful comments on earlier drafts of this article.
* Corresponding author. Fax: +44 23 8059 4597.
E-mail address: higham@soton.ac.uk (P.A. Higham).

0749-596X/$ - see front matter
doi:10.1016/j.jml.2005.01.015

monitoring (e.g., Barnes, Nelson, Dunlosky, Mazzoni, & Narens, 1999; Goldsmith, Koriat, & Weinberg-Eliezer, 2002; Higham, 2002; Kelley & Sahakyan, 2003; Klatzky & Erdelyi, 1985; Koriat & Goldsmith, 1996). Before describing our unique method of analyzing the underlying generation and recognition processes, we will first review the traditional generate-recognize models, and some of their variants.

Early generate-recognize theory and encoding specificity

Early generate-recognize theory (e.g., Anderson & Bower, 1972; Bahrick, 1969, 1970) proposed that cued recall could be achieved by covertly generating associates of test cues, and then attempting to recognize the sought-after target from amongst the generated candidates. Support for the theory came, in part, from experiments demonstrating that the associative strength between the test cue and the target affected the probability of recall (e.g., Bahrick, 1970).

In the early 1970s, Tulving and colleagues (e.g., Thomson & Tulving, 1970; Tulving & Thomson, 1973; Wiseman & Tulving, 1976) argued that generate-recognize models were an insufficient account of recall for two reasons. First, Thomson and Tulving (1970) demonstrated that extralist retrieval cues that are strong associates of the target words (based on free association norms) are not very effective retrieval cues, particularly if the target words were encoded in relation to some other (weak) associate during study. If cued recall is accomplished by first generating candidates from the test cue, and then recognizing the target from amongst the candidates, one would expect that strong associates would be excellent retrieval cues because the probability of accomplishing the first step in the process (generating the target) is high. Second, Tulving and Thomson (1973) demonstrated recognition failure of recallable words. In this demonstration, participants were first given weak associate-target pairs to study. Next they were provided with strong associates of the targets and asked to free associate. Unsurprisingly, copies of the targets were often generated during this phase of the experiment. Following free association, participants were asked to circle those generated items that were targets from the study list. Finally, they attempted to recall the targets in the presence of the weak cues that were encoded specifically with the targets during study. Recognition failure was revealed in that targets not recognized during the generate-recognize phase of the experiment were often recalled later in the presence of reinstated weak cues. This basic finding has been replicated many times (e.g., Bartling & Thompson, 1977; Gardiner, 1988; Postman, 1975; Reder, Anderson, & Bjork, 1974; Sikstrom & Gardiner, 1997; Tulving, 1974; Watkins & Tulving, 1975; Wiseman & Tulving, 1975, 1976; see Nilsson & Gardiner, 1993 for a review) and forms the basis of the Tulving–Wiseman law. Recognition failure is problematic for generate-recognize theory because recall is limited by two bottlenecks, whereas recognition is only limited by one (i.e., the target item has already been "generated" in recognition). Therefore, Tulving and Thomson reasoned that it is impossible for recall to be superior to recognition.

Tulving and colleagues argued, instead, that their results were best explained in terms of the encoding specificity principle. According to this principle, the effectiveness of retrieval cues is determined by the extent to which the cues are encoded specifically with the to-be-remembered (TBR) information. Thus, strong extralist cues are generally not effective for retrieval, despite the fact that they elicit the TBR information with a high probability, because they were not encoded specifically with it. Similarly, recognition failure occurs because the cues available during recognition differ from those that were present at encoding. In contrast, the (weak) cues in recall are reinstated from study. The difference in the reinstatement status of the strong versus weak cues renders recall performance that is superior to recognition.

Variants of early generate-recognize theory

It is important to point out that Tulving and colleagues' data and criticisms only pertain to one class of generate-recognize models: those that assume "trans-situational identity of words" (Tulving & Thomson, 1973, p. 358). If a given target word is considered to have only a single representation in memory—for example, a node in a stable, abstract associative network—then the target representation generated in the context of a strong extralist associate at test must match the target representation activated in the context of a reinstated weak associate. Consequently, under this assumption, there is no way for generate-recognize models to explain either poor performance with strong extralist cues, or recognition failure, as Tulving and colleagues correctly pointed out. However, since the publication of the first encoding specificity papers, several authors have argued that the one-representation-per-word assumption need not necessarily hold, and that encoding-specificity effects can be incorporated into generate-recognize theory under different assumptions. For example, Reder et al. (1974; see also Martin, 1975) proposed that a given word can have more than one "sense," and that the sense of the target evoked at study, when it is presented with a weak associate, is different from the sense of the target when it is generated to a new, strong associate. For example, LIGHT, when generated in the context of the strong associate *dark* at test, has a different sense than the word LIGHT when presented in the context of the weak associate *head* at study; the former means LIGHT