Available online at www.sciencedirect.com
SCIENCE DIRECT®
Acta Psychologica 119 (2005) 107–121
acta psychologica
www.elsevier.com/locate/actpsy

Encoding specificity manipulations do affect *retrieval* from memory

René Zeelenberg *

Department of Psychology, Erasmus Universiteit Rotterdam, Woudestein, J5-65, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands
Department of Psychology, Indiana University, Bloomington, IN 47405, USA

Received 20 April 2004; received in revised form 27 November 2004; accepted 1 December 2004
Available online 8 January 2005

Abstract

In a recent article, P.A. Higham (2002) [Strong cues are not necessarily weak: Thomson and Tulving (1970) and the encoding specificity principle revisited. *Memory & Cognition, 30*, 67–80] proposed a new way to analyze cued recall performance in terms of three separable aspects of memory (retrieval, monitoring, and report bias) by comparing performance under both free-report and forced-report instructions. He used this method to derive estimates of these aspects of memory in an encoding specificity experiment similar to that reported by D.M. Thomson and E. Tulving (1970) [Associative encoding and retrieval: weak and strong cues. *Journal of Experimental Psychology, 86*, 255–262]. Under forced-report instructions, the encoding specificity manipulation did not affect performance. Higham concluded that the manipulation affected monitoring and report bias, but not *retrieval*. I argue that this interpretation of the results is problematic because the Thomson and Tulving paradigm is confounded, and show in three experiments using a more appropriate design that encoding specificity manipulations do affect performance in forced-report cued recall. Because in Higham's framework forced-report performance provides a measure of retrieval that is uncontaminated by monitoring and report bias it is concluded that encoding specificity manipulations do affect retrieval from memory.

* Corresponding author. Tel.: +31 10 408 9560; fax: +31 10 408 9009.
E-mail address: zeelenberg@fsw.eur.nl

0001-6918/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.actpsy.2004.12.001

PsycINFO classification: 2343
Keywords: Cued recall; Retrieval processes; Encoding specificity

1. Introduction

Human memory is not perfect. Everyone has experienced failures to remember certain information, such as the name of a familiar person or what one had for dinner last night. Such failures to retrieve information from memory are called errors of omission. Another type of errors are the errors of commission: the tendency to remember things that did not happen. Not everything we 'remember' is necessarily accurate. For example, participants in a memory experiment may recall words that were not presented on the study list. It is often assumed that, in an attempt to limit the number of such errors, participants monitor the products of retrieval before emitting a response. Thus, the observed performance of a participant in a memory experiment is not only a function of what is retrieved from memory but also of metacognitive processes: Participants may choose not to report a word retrieved from memory if they are not confident that the word was indeed presented on the study list.

To separate the contribution of retrieval processes and metacognitive processes Higham (2002) recently proposed a new way to analyze performance in a cued recall task. According to Higham, cued recall performance is determined by three processes: retrieval, monitoring effectiveness (i.e., the ability to discriminate correct from incorrect products of retrieval) and report bias (i.e., willingness to report an answer). Higham's analysis of memory performance assumes that a 'best-candidate' answer is retrieved from memory. Next, the monitoring mechanism estimates the probability that the best candidate is the correct answer. If the estimated probability surpasses the report criterion, the candidate is reported, otherwise it is withheld. Thus, in standard *free-report* conditions in which participants are not forced to report an answer to every cue, performance is supposedly due to the combination of these three aspects of memory.

Higham's (2002) analysis is based on type 2 signal detection theory (e.g., Healy & Jones, 1973; Lockhart & Murdock, 1970). To derive estimates for the three parameters (i.e., retrieval, monitoring and report bias), participants perform a cued recall task under both free-report and forced-report instructions. Under standard free-report instructions, participants will not generate an answer to every cue because, for some of the generated candidate answers, the probability correct assigned by the monitoring mechanism will not surpass the report criterion. Cues to which no answer was generated under free-report instructions are then presented under *forced-report* instructions (i.e., an answer must be generated to every single cue). Some of the answers initially withheld under free-report instructions will be correct in forced-report whereas other answers will be incorrect. By observing the frequencies in the four cells of a 2 (correct answer/incorrect answer) $\times$ 2 (response initially reported/response initially withheld) contingency table, estimates of report bias and monitoring effectiveness can be calculated. Details about the exact procedure used to calculate these estimates can be found in Higham (2002). Most important for the purpose of the present article is that, within this framework, forced-report conditions yield a measure of retrieval that is uncontaminated by monitoring and report bias.

This new way to analyze memory performance was used by Higham (2002) to revisit the encoding specificity principle, a basic principle of memory described in many textbooks. According to the encoding specificity principle, retrieval from memory in a cued recall task is affected by the extent to which information relating the cue and target was stored during study of the target, or as Tulving and Thomson (1973) stated "the target item must be encoded in some sort of reference to the cue for the cue to be effective" (p. 359). Several results (e.g., Barclay, Bransford, Franks, McCarrel, & Nitsch, 1974; Light & Carter-Sobell, 1970; Thomson & Tulving, 1970) reported in the memory literature are consistent with the encoding specificity principle and mechanisms compatible with the encoding specificity principle have been incorporated in many memory models (e.g., Humphreys, Bain, & Pike, 1989; Nelson, McKinney, Gee, & Janczura, 1998; Raaijmakers & Shiffrin, 1981). Within the framework of Higham, one can ask the question which of the three memory parameters are affected by encoding specificity manipulations. According to the encoding specificity principle, such manipulations are expected to affect the accessibility of memory traces (i.e., retrieval from memory). In contrast to this basic principle of memory, Higham reported an experiment suggesting that the effect of encoding specificity is on monitoring and report bias, but not on retrieval.

Higham (2002) investigated performance in an experiment similar to that of Thomson and Tulving (1970). In the study phase of the experiment, participants studied cue–target pairs consisting of pre-experimentally weakly associated words (e.g., *glue*-*CHAIR*). At test, participants were given either the weakly associated cue (e.g., *glue*) or a strongly associated cue (e.g., *table*) to assist recall of the target (e.g., *CHAIR*). The weak cue was always the word with which the target had been paired during study; the strong cue had not been presented during study. According to the encoding specificity principle, performance is expected to be better for the weak reinstated cue than for the strong extralist (i.e., nonstudied) cue because the target has been encoded in reference to the weak cue.

Under standard free-report conditions, Higham's (2002) results replicated those of Thomson and Tulving (1970): Cued recall performance was better for weak cues (mean recall = 25.8%) than for strong cues (mean recall = 7.3%). Under forced-report instructions, however, there was no difference between the weak cue (mean recall = 27.9%) and strong cue (mean recall = 25.8%) conditions. Or stated differently, the increase in target recall under forced-report instructions was rather small ($M = 2.1\%$) for weak cues, but rather large ($M = 18.6\%$) for strong cues. Because performance for the weak-and strong-cue conditions was equivalent under forced-report instructions, Higham concluded that the encoding specificity manipulation did not affect retrieval.[1]

[1] To be fair, it should be noted that Higham was careful to mention that he did not argue that his experiment undermines the encoding specificity principle in general. Nevertheless, his paper raises questions about the validity of the encoding specificity principle.

One problem with this conclusion is that the large increase in performance under forced-report instructions (compared to that under free-report instructions) might be largely due to pure guessing on the basis of *pre-experimental* associations between the cue and the target. In the weak-cue condition, the mean probability of generating the target word to the cue (according to free association norms) was 1% and, consequently, the probability of correctly guessing the target without any episodic information was very low. In the strong-cue condition, however, the probability of guessing the target based on the pre-experimental association between the cue and the target was rather high (the free association probability for the strong cue–target pairs was on average 35%). Thus, for the strong-cue condition, the increase in the proportion correct responses under forced-report conditions may have been largely due to guessing. Because study conditions (i.e., reinstated cue vs. extralist cue) and pre-experimental associative strength (i.e., weak vs. strong) were confounded, no strong conclusions can be drawn from this experiment.

To recapitulate, there are two different explanations for the results reported by Higham (2002). One possibility is that encoding specificity manipulations do not affect retrieval from memory. If this were true, the encoding specificity principle would be undermined. Another possibility, however, is that encoding specificity manipulations do affect retrieval from memory, but in the Higham study the encoding specificity effect was masked by the larger probability of guessing the target in the strong extralist cue condition than in the weak reinstated cue condition. Encoding specificity effects have been obtained in a number of paradigms but, to the best of my knowledge, no study, except for the Higham study, has investigated encoding specificity effects in forced-report cued recall. Existing studies, therefore, do not address the question whether encoding specificity manipulations affect retrieval, because in the Higham framework any difference between two conditions under free-report conditions may be due not only to differences in retrieval but also to differences in monitoring effectiveness and differences in the setting of the report criterion. Hence, it is important to establish whether in paradigms other than the weak-strong paradigm encoding specificity manipulations do or do not affect retrieval from memory (and hence are absent under forced-report conditions). The present study therefore reexamined the question whether or not encoding a target in reference to the test cue affects retrieval of the target stimulus.

2. Experiment 1

The experimental design of Experiment 1 was very similar to that of Higham (2002). Participants studied cue–target pairs (e.g., *pull-ROPE*) and were tested in a forced-report cued recall task.[2] In the reinstated cue condition, the cue presented

[2] No free-report condition was included. Such a condition would be necessary to obtain estimates of monitoring effectiveness and report bias. However, because the question addressed by the present study is whether there is an encoding specificity effect on retrieval the inclusion of a free-report condition was not necessary.

during recall was the same word (e.g., *pull*) that had been presented with the target during study. In the extralist condition, the cue presented during recall was a different nonstudied word (e.g., *jump*). The main difference with the Higham study was that the cues in the reinstated and extralist conditions were both strong cues, thereby removing the confounding present in the weak-strong paradigm. If encoding the cue–target relation during study does not affect retrieval, a possibility suggested by the Higham study (2002), no difference should be observed between the reinstated cue and extralist cue conditions under forced-report instructions. Such a result would be contradictory to the encoding specificity principle.

2.1. Method

2.1.1. Participants

Sixteen students and faculty members at the Erasmus University Rotterdam participated in the experiment. The students received course credit for their participation.

2.1.2. Stimulus material and design

A set of 40 triplets was selected from published free association norms (de Groot, 1980; van Loon-Vervoorn & Van Bekkum, 1991). Each triplet consisted of a target word (e.g., *ROPE*) and two cue words (e.g., *pull*, *jump*).[3] The triplets were selected from the norms in such a way that the target was a strong associate of both cue words (mean cue-to-target association frequency = 30.6%, range = 17–55%), with the restriction that the two cues were not associated to each other (mean cue-to-cue association frequency = 0.0%).

The experimental design consisted of two conditions. In the reinstated cue condition, the word (e.g., *pull*) that had been presented with the target (e.g., *ROPE*) during study was also presented as a cue in the recall task. In the extralist cue condition, the cue presented during recall was different from the word presented with the target during study; that is, the other word of the triplet (e.g., *jump*), which had not been presented during study, was presented as a test cue. For counterbalancing purposes two different study lists and two different test lists were constructed. Across the four resulting study list–test list combinations each cue and each target were presented equally often in the reinstated cue and extralist cue conditions. Hence, any possible difference between these conditions cannot be attributed to different pre-experimental cue-to-target associative strengths and is instead due to the experimental manipulation.

2.1.3. Procedure

In the study phase 40 words pairs (e.g., *pull-ROPE*) were presented for intentional study. The word pairs were printed under each other on five sheets of paper.

[3] The examples are translations of the cues and targets used in the experiment. The actual stimulus materials of Experiments 1 and 2 consisted of Dutch words.

Participants received one of eight different randomly ordered study lists. The cue words were printed in lowercase and the target words in uppercase letters. Three blank lines separated one pair from the next one to promote encoding of the relation between the two members of a pair instead of encoding relations between target items. Participants were informed that their memory for the uppercase words would be tested. They were instructed to pay attention to the words printed to the left of each uppercase word because these might be provided in the memory test to assist recall of the uppercase words. Participants were given 3 min for studying the word pairs and were informed about the remaining study time half way through the 3-min period and 30 s prior to the end of the 3-min period.

In the test phase a cued recall task was given (in all experiments reported in the present article, the study phase was immediately followed by the test phase). Forty cue words were printed on two sheets of paper. Each participant received one of eight different randomly ordered test lists. Participants were instructed to use the cue word to recall an uppercase word that had been presented in the study phase and write it down next to the cue word. They were informed that each cue word was related to a word from the study list and told that some of the cues had been presented with the target word during study whereas other cue words had not been presented during the study phase. Participants were instructed to write down an answer next to each cue even if they were not sure or had to guess.

2.2. Results and discussion

For each participant, the percent recalled target words in the test phase was calculated for the two conditions. Forced-report performance in the reinstated cue condition (mean recall = 82.5%) was much higher than forced-report performance in the extralist cue condition (mean recall = 47.5%), $t(15) = 6.78$, $p < .0001$. A closer look at the results showed that out of the 16 participants, 15 participants recalled more target words in the reinstated cue condition than in the extralist cue condition and 1 participant showed no effect ($p < .0001$, *sign* test).

Experiment 1 showed an encoding specificity effect using a cue reinstatement procedure. Perhaps it is not very surprising that cued recall performance is better if the retrieval cue has been presented with the target during study than if the cue has not been presented with the target during study. Nevertheless, Higham (2002) suggested the possibility that cue reinstatement does not affect performance in forced-report conditions. An absence of a difference in performance between the reinstated cue and extralist cue conditions in the present experiment would have been inconsistent with the encoding specificity principle.

In the remaining two experiments, I investigated whether encoding specificity manipulations affect retrieval by comparing performance between two extralist cue conditions (i.e., the cue was never presented during study) that differed in the extent to which information relating the cue and target was encoded in memory. The main motivation for these experiments was that it seems more interesting and surprising to obtain an encoding specificity effect using only extralist cuing conditions in which encoding of the cue–target relation is implicit (rather than through explicit presentation of the cue with the target during study). One may wonder how the relation between the cue and target can be encoded in memory when the cue is not explicitly presented with the target during study. Several possible mechanisms have been proposed that may result in encoding of the cue–target relation. A first possibility is that during study of a word related words consciously (but covertly) come to mind (Underwood, 1965) creating a possibility for cue–target learning (e.g., Tulving & Osler, 1968; Zeelenberg, Shiffrin, & Raaijmakers, 1999). Another possibility is that cue–target associations are strengthened not because participants consciously think of the cue word during study of the target, but because of unconscious automatic strengthening of the links between words in memory (Nelson, McKinney, et al., 1998). Finally, encoding of the cue–target relation during the study may depend on the storage of semantic features common to the cue and the target (e.g., Roediger & Adelson, 1980; Zeelenberg, Pecher, Shiffrin, & Raaijmakers, 2003). Any of these mechanisms could explain why an extralist cue could assist in retrieval of the target from memory.

3. Experiment 2

In Experiment 2 encoding of the cue–target relation during study was manipulated by using word pairs with either pre-existing *bidirectional* or pre-existing *unidirectional* associations. The bidirectionally associated word pairs (e.g., *spider-WEB*) were associated in both directions. That is, according to published free association norms, the word *spider* evokes the response *WEB* and the word *WEB* evokes the response *spider*. The unidirectionally associated word pairs (e.g., *frog-GREEN*) were associated in the forward (i.e., cue-to-target) direction only. Thus, according to the norms, the word *frog* often evokes the response *GREEN*, but *GREEN* never evokes the response *frog*. Because, for the unidirectionally associated word pairs, there is no association from the target to the cue it is unlikely that the relation between the cue and the target will be encoded during study of the target. In the bidirectional condition, however, there is an association from the target to the cue and hence the relation between the cue and the target may be encoded in memory. This may happen because during study of the word *WEB* participants think of or activate *spider*. However, when studying *GREEN* the participant will not likely think of or activate *frog*. Thus, the encoding specificity principle predicts better extralist cued recall performance for bidirectionally associated word pairs than for unidirectionally associated word pairs. Consistent with this prediction, Nelson, McKinney, et al. (1998) observed that extralist cued recall was better for bidirectionally associated cue–target pairs than for unidirectionally associated cue–target pairs that had no backward target-to-cue association.[4] The study by Nelson et al., however, used stan-

[4] Humphreys and Galbraith (1975) and Ley (1977) also manipulated backward associative strength in an extralist cued recall study. In their studies, however, cue–target pairs were either associated in the forward direction only or in the backward direction only. Hence, forward and backward associative strength were confounded.

dard free-report instructions and hence we cannot infer with certainty that the manipulation of backward strength affected retrieval.

3.1. Method

3.1.1. Participants

Sixteen persons participated in the main experiment and another 16 participated in a pilot study. Participants were students or staff members at the University of Amsterdam or Erasmus University Rotterdam and received a small monetary award for their participation.

3.1.2. Stimulus materials

The critical stimuli consisted of 30 unidirectionally associated cue–target pairs (e.g., *profit-MONEY*, *frog-GREEN*, *poison-DEATH*) and 30 bidirectionally associated cue–target pairs (e.g., *book-READ*, *spider-WEB*, *milk-COW*) taken from Zeelenberg et al. (1999). The cue–target pairs were selected from published free association norms (de Groot, 1980; Lauteslager, Schaap, & Schievels, 1986; van Loon-Vervoorn & Van Bekkum, 1991). The two types of cue–target pairs were matched on forward association frequency and log frequency of the target words. The mean forward association frequency was 19.2% (SD = 9.6) for the bidirectionally associated pairs and 19.8% (SD = 11.9) for the unidirectionally associated pairs. The log frequencies of the target words were derived from the CELEX database (Baayen, Piepenbrock, & van Rijn, 1993). The mean log frequency per million of the target words was 1.92 (SD = 0.64) for the bidirectionally associated pairs and 1.92 (SD = 0.55) for the unidirectionally associated pairs. The mean backward association frequency was 39.4% (SD = 22.4) for the bidirectionally associated pairs and 0.0% (SD = 0.0) for the unidirectionally associated pairs.

A pilot study was run to make sure that the mean forward associative strengths of the bidirectionally and unidirectionally associated pairs did not differ from each other. Although the two types of pairs were matched on forward associative strength according to published norms, I wanted to make sure that the forward associative strengths did not differ for the subject population used in the present experiment. This would ensure that any possible difference between the two conditions observed in the forced-report cued recall task of the main experiment could not be attributed to different probabilities of guessing the target word on the basis of pre-experimental associations. Sixteen participants participated in a short paper and pencil experiment in which 60 cue words were presented in a free association task. Next to each cue word, participants wrote down the first related word that came to mind. In the bidirectional condition, 22.7% of the responses corresponded to the target word. In the unidirectional condition, 22.1% of the responses corresponded to the target word. The difference between the two conditions was not significant, $t(15) = 0.20$, $p = .84$.

3.1.3. Procedure

The experiment consisted of a study phase and a test phase. In the study phase an incidental learning task[5] was given in which participants gave pleasantness ratings to the 60 target words of each pair. The target words were printed under each other on three sheets of paper. Each participant received one of four different random orders. Next to each target word, the numbers 1–5 were printed and participants gave pleasantness ratings to each word by circling the number that indicated best how much they liked the word (with 1 indicating 'very unpleasant' and 5 indicating 'very pleasant').

In the test phase, an extralist cued recall task was given. Sixty cue words were printed on three sheets of paper. Each participant received one of four different random orders. Next to each cue word, participants wrote down a word from the study phase. Participants were instructed to use the cue to recall a word that had been presented in the pleasantness rating task. They were informed that the cue words were related to the words they had to recall and were given some examples. Participants were instructed that they had to write down an answer next to each cue even if they were not sure or had to guess.

3.2. Results and discussion

For each participant, the percent recalled target words in the test phase was calculated for the two conditions. Forced-report performance in the bidirectional condition (mean recall = 60.2%) was higher than forced-report performance in the unidirectional condition (mean recall = 32.5%), $t(15) = 10.47$, $p < .0001$. A closer look at the results showed that all 16 participants recalled more target words in the bidirectional condition than in the unidirectional condition ($p < .0001$, *sign* test).

To obtain further information on the role of pre-experimental associations in extralist cued recall performance additional correlational analyses were performed on the results for the bidirectionally associated word pairs.[6] The correlation ($r = .47$) between forward associative strength and cued recall performance was significant, $t(29) = 2.85$, $p < .05$. The positive correlation indicates that participants recalled more target words the stronger the forward association. The correlation ($r = .42$) between backward associative strength and cued recall was also significant, $t(29) = 2.42$, $p < .05$, indicating that participants recalled more target words the stronger the backward association.

[5] In Experiments 2 and 3 incidental study instructions were given to keep the study procedure as similar as possible to that of other studies (e.g., Zeelenberg et al., 1999, 2003) using similar manipulations of encoding the cue–target relationship.

[6] Only bidirectionally associated pairs were included in these analyses because the significant difference in performance between the unidirectional and bidirectional conditions already showed that the absence of a backward association negatively affects cued recall performance. One aim of the correlation analyses was to determine whether differences in backward associative strength affect cued recall for those pairs for which a backward association is present.

A multiple regression analysis with forward and backward associative strength as predictor variables for cued recall performance showed that these variables together accounted for approximately 33% ($R = .58$) of the variance in cued recall performance, $F(2, 27) = 6.69$, $p < .01$. Semipartial correlation (also called part correlation) coefficients were calculated to see whether each variable made a unique contribution to the prediction of cued recall performance. The squared semipartial correlation coefficient between a predictor variable and the criterion variable equals the amount of unique variability in the criterion variable (e.g., cued recall performance) accounted for by the predictor variable (e.g., forward associative strength) after taking into account the variability accounted for by the other variable (e.g., backward associative strength). The semipartial correlation ($r = .40$) between forward strength and cued recall performance was significant, $t(27) = 2.53$, $p < .05$, indicating that forward strength accounted for an unique amount of variability (i.e., 16%) in cued recall performance after taking into account the variability accounted for by backward associative strength. Backward associative strength also contributed uniquely to cued recall performance ($r = .33$, hence approximately 11% of the variance in performance was accounted for by backward strength after taking into account the contribution of forward strength), $t(27) = 2.08$, $p < .05$. These results not only confirm that backward associative strength plays a role in extralist cued recall, they also point to the role of forward associative strength. The fact that forward associative strength plays a role in extralist cued recall performance is maybe not very surprising, but it emphasizes the need to take this factor into account when designing experiments (as was done in the present experiment by matching the different conditions on forward associative strength).

4. Experiment 3

The aim of Experiment 3 was to obtain converging support for the claim that encoding of the cue–target relation affects retrieval from memory by using a manipulation different from the one used in Experiment 2. Storage of information relating the cue and target was manipulated by presenting target words in sentences that biased their interpretation. For example, during the study phase, the word *BEACH* could be presented in one of the following two sentences: (1) 'He had a nice tan after a warm day on the BEACH', or (2) 'Children like to play with scoops and buckets on the BEACH'. The sense of the word *BEACH* emphasized by sentence 1 was highly congruent with the test cue *sun*. The sense emphasized by sentence 2 was less congruent with the cue *sun*. In previous studies, similar manipulations have been shown to affect memory performance (e.g., Barclay et al., 1974; Roediger & Adelson, 1980; Roediger & Payne, 1983). However, these studies did not use a forced-report cued recall task. According to the encoding specificity principle, cued recall is expected to be better in the more congruent condition, in which the study context emphasizes a sense of the target word that is highly related to the test cue, compared to the less congruent condition, in which the study context emphasizes a sense of the target that is less related to the test cue. A methodological advantage of Experiment 3 over

Experiment 2 is that the exact same cue–target pairs were used in the two conditions of the experiment. In Experiment 2 this was impossible because in that experiment encoding of the cue–target relation was manipulated by relying on pre-experimental associations between words. In the present experiment, however, we had control over the encoding of relational information by manipulating the sentence context in which the target was presented.

4.1. Method

4.1.1. Participants

Sixteen Indiana University students participated for course credit. All participants were native speakers of English.

4.1.2. Stimulus materials and design

A set of 46 critical triplets were taken from Zeelenberg et al. (2003). Each triplet consisted of a target word (e.g., *BEACH*) and two cue words (e.g., *sun*, *sand*) that were related to a different sense or meaning aspect of the target. The mean association frequency (Nelson, McEvoy, & Schreiber, 1998) from the cue to the target was 12.8% (SD = 14.2). The mean association frequency from the target to the cue was 14.2% (SD = 13.8).

For each of the 46 triplets two sentences were created. This resulted in 92 different study sentences. Within a pair, each of the two sentences emphasized a different sense or meaning aspect of the target word. For example, for the triplet *sun–sand–BEACH*, the following two sentences were created: (1) 'He had a nice tan after a warm day on the BEACH', (2) 'Children like to play with scoops and buckets on the BEACH'. Sentence 1 provided the more congruent study context for the cue *sun* and sentence 2 provided the less congruent study context for the cue *sun*. The reverse was true for the cue *sand*. The test cue itself was never presented in the study sentence.

For counterbalancing purposes, four study list–test list combinations were constructed. Across the four resulting study list–test list combinations, each test cue appeared once and each target twice in each of the two experimental conditions. Note that because each cue–target pair appeared equally often in the more congruent condition and the less congruent condition, any difference between these conditions cannot be due to differences in pre-experimental associative strength. Likewise, because each sentence appeared equally often in the two experimental conditions, any difference cannot be due to one of the sentences resulting is a stronger or more elaborate encoding of the target word.

Four additional sentences were selected to serve as filler sentences in the study phase. Two additional words were selected to serve as practice trials in the test phase.

4.1.3. Procedure

The experiment consisted of a study phase and a test phase. In the study phase, an incidental learning task was given in which participants completed word stems embedded in a sentence context. In each sentence the target word was missing. The first letter of the missing target was presented in uppercase followed by four dots (participants were informed that the number of dots did not correspond to the number of missing letters). The participant's task was to complete the word so that the result would be a meaningful sentence. For example, for the sentence 'Children like to play with scoops and buckets on the B...' the word *BEACH* would be the correct completion. Sentences were presented one at a time on the computer screen. Participants were instructed to enter their completion on the keyboard. After the response of the participant, the correct completion (i.e., the target word) was presented for 5 s in uppercase letters on the center of the screen two lines below the sentence (the feedback ensured that participants were exposed to all target words even if they were unable to generate the target from the sentence). Participants were instructed to check whether their response was correct and to see how the correct completion resulted in a meaningful sentence. A total of 4 filler sentences and 46 critical sentences were presented. Two filler sentences were inserted at the beginning and two at the end of the list to control for recency and primacy effects. The 46 critical sentences were presented in the middle of the list. Sentences were presented in a different random order for each participant.

In the test phase, an extralist cued recall task was given. Forty-eight words were presented one at a time on the screen and participants were instructed to use the cue word to recall one of the 'missing' words from the first part. They were informed that the cue words were related to the words they had to recall and were given an example. Participants were instructed that they always had to write down an answer even if they were not sure or had to guess. The next cue word was presented after the participant pressed the spacebar. The first two cue words in the extralist cued recall task were practice words that were given to make sure the participants understood the instructions (these two cue words were related to two of the filler words presented in the study phase). The two practice cues were followed by the 46 critical cue words presented in a random order. For 23 of the 46 critical cues, the target word had been presented in the more congruent sentence context during the study phase. For the remaining 23 critical cues, the target word had been presented in the less congruent sentence context.

4.2. Results and discussion

Overall, 84% of the word stems presented in the study phase were completed with the target word. Recall performance was scored irrespective of whether the word stem was completed with the target word during the study phase. For each participant, the percent recalled target words in the test phase was calculated for the two conditions. More targets were recalled in the more congruent condition (mean recall = 64.4%) than in the less congruent condition (mean recall = 42.4%), $t(15) = 5.03$, $p < .0001$. A *sign* test also showed that the congruency effect was significant ($p < .001$); out of the 16 participants, 13 participants recalled more target words in the more congruent condition than in the less congruent condition and three participants showed no effect.

5. General discussion

The present study investigated encoding specificity effects on retrieval in a forced-report extralist cued recall task. In accordance with the encoding specificity principle, participants recalled more targets in the reinstated cue condition in which the cue and target had been presented as a pair during study than in the extralist condition in which the test cue had not been presented during study. In Experiment 2 participants recalled more words in the bidirectionally associated condition that in the unidirectionally associated condition. In Experiment 3, extralist cued recall performance was much better when the study context emphasized meaning aspects of the target that were highly related to the test cue compared to when the study context emphasized meaning aspects that were less related to the test cue. These results stand in contrast with the results obtained by Higham (2002) who obtained no evidence for encoding specificity under forced-report conditions in the weak-strong paradigm of Thomson and Tulving (1970). However, as mentioned in the introduction, interpretation of forced-report results in the weak-strong paradigm is troublesome, because the pre-experimental associative strengths between the cue and the target words are confounded with study condition. In the introduction to his article, Higham argued that previous studies may have led to some misleading interpretations of encoding specificity effects because these studies failed to control report option. In the present study, however, withholding responses was not an option. Yet, the present experiments still showed rather large encoding specificity effects.

In accordance with the encoding specificity principle, many memory theories (e.g., Humphreys et al., 1989; Nelson, McKinney, et al., 1998; Raaijmakers & Shiffrin, 1981) assume that retrieval is affected by the extent to which information relating the cue and target stimuli is encoded during study. As mentioned in the introduction, several mechanisms have been proposed by which information relating the cue and target may be encoded even if the cue and target are not presented as a pair during study. One possibility is that for associatively related cue–target pairs participants (sometimes) think of the test cue during study of the target (e.g., Tulving & Osler, 1968; Zeelenberg et al., 1999) and that as a result the cue–target relation is stored in memory. Another possibility is that cue–target associations are automatically strengthened, that is, without the need to consciously think of the cue–target association (Nelson, McKinney, et al., 1998). Yet another possibility is that the strengthening of cue–target associations depends on the strengthening of semantic features shared by the cue and the target (Barclay et al., 1974; Roediger & Adelson, 1980; also see Zeelenberg et al., 2003). Several researchers (e.g., Reder, Anderson, & Bjork, 1974) have advocated such a semantic interpretation of the encoding specificity effect. The present study, however, was not designed with the intention to discriminate between different explanations of encoding specificity effects. Whatever the exact mechanism(s) responsible for encoding of the cue–target relation during study, the important point is that most current theories of memory assume that such encoding increases the strength of the match between the test cue and the target word in memory and, therefore, affects the probability that the target stimulus is retrieved from memory. The present results are consistent with this notion.

Acknowledgements

I thank Jeroen Raaijmakers, Romke Rouw and Niels Smits for helpful discussions. Phil Higham, Doug Nelson, Diane Pecher and Eric-Jan Wagenmakers provided valuable comments on earlier versions of this manuscript. Priscilla van Leeuwen assisted in the construction of stimulus materials and testing of participants for Experiment 1.

References

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX lexical database (CD-ROM), linguistic data consortium*. Philadelphia, PA: University of Pennsylvania.

Barclay, J. R., Bransford, J. D., Franks, J. J., McCarrel, N. S., & Nitsch, K. (1974). Comprehension and semantic flexibility. *Journal of Verbal Learning and Verbal Behavior, 13*, 471–481.

de Groot, A. M. B (1980). *Mondelinge Woordassociatienormen: 100 woordassociaties op 460 Nederlandse zelfstandige naamwoorden* [Oral word association norms: 100 word associations to 460 Dutch nouns]. Lisse, The Netherlands: Swets & Zeitlinger.

Healy, A. F., & Jones, C. (1973). Criterion shifts in recall. *Psychological Bulletin, 79*, 335–340.

Higham, P. A. (2002). Strong cues are not necessarily weak: Thomson and Tulving (1970) and the encoding specificity principle revisited. *Memory & Cognition, 30*, 67–80.

Humphreys, M. S., Bain, J. D., & Pike, R. (1989). Different ways to cue a coherent memory system: a theory for episodic, semantic and procedural tasks. *Psychological Review, 96*, 208–233.

Humphreys, M. S., & Galbraith, R. (1975). Forward and backward associations in cued recall: predictions from the encoding specificity principle. *Memory & Cognition, 6*, 702–710.

Lauteslager, M., Schaap, T., & Schievels, D. (1986). *Schriftelijke woordassociatienormen voor 549 Nederlandse zelfstandige naamwoorden* [Written word association norms for 549 Dutch nouns]. Lisse, The Netherlands: Swets & Zeitlinger.

Ley, R. (1977). Encoding specificity and unidirectional associates in cued recall. *Memory & Cognition, 5*, 523–528.

Light, L. L., & Carter-Sobell, L. (1970). Effects of changed semantic context on recognition memory. *Journal of Verbal Learning and Verbal Behavior, 9*, 1–11.

Lockhart, R. S., & Murdock, B. B. (1970). Memory and the theory of signal detection. *Psychological Bulletin, 74*, 177–196.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). *The University of South Florida word association, rhyme, and word fragment norms*. Available from http://www.usf.edu/FreeAssociation/.

Nelson, D. L., McKinney, V. M., Gee, N. R., & Janczura, G. A. (1998). Interpreting the influence of implicitly activated memories on recall and recognition. *Psychological Review, 105*, 299–324.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88*, 93–134.

Reder, L. M., Anderson, J. R., & Bjork, R. A. (1974). A semantic interpretation of encoding specificity. *Journal of Experimental Psychology, 102*, 648–656.

Roediger, H. L., & Adelson, B. (1980). Semantic specificity in cued recall. *Memory & Cognition, 8*, 65–74.

Roediger, H. L., & Payne, D. G. (1983). Superiority of free recall to cued recall with "strong" cues. *Psychological Research, 45*, 275–286.

Thomson, D. M., & Tulving, E. (1970). Associative encoding and retrieval: weak and strong cues. *Journal of Experimental Psychology, 86*, 255–262.

Tulving, E., & Osler, S. (1968). Effectiveness of retrieval cues in memory for words. *Journal of Experimental Psychology, 77*, 593–601.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*, 352–373.

Underwood, B. J. (1965). False recognition produced by implicit associative verbal responses. *Journal of Experimental Psychology, 70*, 122–129.

van Loon-Vervoorn, W. A., & Van Bekkum, I. J. (1991). *Woordassociatie lexicon* [Word association lexicon]. Amsterdam/Lisse: Swets & Zeitlinger.

Zeelenberg, R., Pecher, D., Shiffrin, R. M., & Raaijmakers, J. G. W. (2003). Semantic context effects and priming in word association. *Psychonomic Bulletin & Review, 10*, 653–660.

Zeelenberg, R., Shiffrin, R. M., & Raaijmakers, J. G. W. (1999). Priming in a free association task as a function of association directionality. *Memory & Cognition, 27*, 956–961.