Testing the Myth of the Encoding-Retrieval Match

Sharon H. X. Lu (sharonhxlu@gmail.com)
Department of Psychology, National University of Singapore
9 Arts Link, Singapore 117570, Singapore

Winston D. Goh (psygohw@nus.edu.sg)
Department of Psychology, National University of Singapore
9 Arts Link, Singapore 117570, Singapore

Abstract

The encoding specificity principle, first proposed by Thomson and Tulving (1970), holds that successful memory performance depends importantly on the extent to which there is a match between encoding and retrieval conditions. However, Nairne (2002) proposed that the principle is a myth because one cannot make unequivocal predictions about memory performance by appealing to the encoding-retrieval match; instead, what matters is the relative diagnostic value of the match, and not the absolute match. Two experiments varied the diagnostic value of the cue by manipulating the degree of cue overload in terms of the extent to which the retrieval cues subsumed other items in the study list and the level of the encoding-retrieval match. Results support Nairne's (2002) assertion that the diagnostic value of retrieval cues is a better predictor of memory performance than the absolute encoding-retrieval match.

Keywords: Encoding specificity; cue overload; cued recall; retrieval processes; memory.

Introduction

One of the fundamental ideas in modern memory research is that the match between encoding and retrieval conditions affects memory performance, an idea that was first proposed in a series of studies conducted by Tulving and colleagues (Tulving & Osler, 1968; Thomson & Tulving, 1970; Tulving & Thomson, 1973; Tulving, 1983). These studies suggested that successful remembering is a joint function of encoding and retrieval processes, and subsequently led to the encoding specificity principle, which states that a retrieval cue will be effective to the extent that it was specifically encoded at the time of learning. Thus, if the target word *NAIL* is encoded and stored in the context of the word *FINGER*, a subsequent retrieval cue such as *a tool* will be ineffective, but a cue such as *a part of the human body* will probably be quite effective.

Recently, Nairne (2002) argued against the case that similarity between encoding and retrieval conditions (or the encoding-retrieval match) is all that matters in cue effectiveness. He in fact argued that knowing the status of the encoding-retrieval match by itself predicts next to nothing about subsequent retention. It is not matching features from the encoding and retrieval conditions *per se* that are needed; it is the presence of features that help discriminate the correct target from incorrect competitors, i.e., the distinctiveness of the retrieval cue. Using the *NAIL* example, if its competitors included *TOE* and *HAND*, the effectiveness of the cue *a part of the human body* would probably be diminished. Adopting Watkins and Watkins' (1975) terminology, the cue can be said to be "overloaded" as it does not provide any diagnostic information about the target occurrence because the competitors are also subsumed by the cue.

As an analogy, Nairne (2002) likened the relationship between the encoding-retrieval match and cue distinctiveness to intensity and brightness. What determines the perception of brightness of a light is the amount of light falling in the centre relative to the surroundings, not the absolute amount of light. In a similar vein, he argued that it is not the absolute encoding-retrieval match that is critical, but rather the relative diagnostic value of the match, which is the extent to which the cue uniquely specifies the target. It is in this sense that he argued that the encoding specificity principle is a "myth" because although the principle specifies that the successful retrieval of a target depends importantly on the extent that to which there was a match between encoding and retrieval conditions, it is not the *absolute* match per se that predicts memory performance, but the *relative distinctiveness* of the retrieval cue.

The question that then arises is whether these arguments render the encoding specificity principle, which claims that "a retrieval cue is effective if, and only if, the information about its relation to the to-be-remembered (TBR) item is stored at the same time as the TBR item itself" (Thomson & Tulving, 1970, p. 255), irrelevant or redundant. The aim of the present study is to examine these arguments in more detail and to empirically test the predictions articulated in Nairne's (2002) thought experiments.

Experiment 1

One thought experiment can be summarised as follows. Participants memorise a series of events E_1, E_2, E_3 ... E_n, and are asked to recall E_1. E_1 has features X_1 and X_2 which could be provided as retrieval cues. From the perspective of the encoding specificity principle, providing both X_1 and X_2 as cues should logically increase the degree of the encoding-retrieval match compared to a single retrieval cue. Hence, one would expect better recall for the condition with two retrieval cues than the condition with just one cue.

However, suppose X_2 is also found in E_2, E_3 … E_n. X_2 now provides no diagnostic value for differentiating E_1 from its competitors. In this case, the two cue condition would not have any advantage over a single cue. Performance may even decline since X_2 subsumes all of the memorised events.

These predictions were tested in Experiment 1 using a cued recall task. In each trial, participants studied 10 semantically unrelated cue-target pairs such as *ABORT-DONKEY*. The degree of cue overload and the degree of the encoding-retrieval match were varied. In the high-overload conditions, all 10 cue-target pairs had targets that were from the same semantic category as *DONKEY*, e.g. *TIGER*, *ELEPHANT*. In the low-overload conditions, the targets were unrelated to *DONKEY*, e.g. *FINGER*, *PAPAYA*. In the high-match conditions, two retrieval cues were provided at test, the original studied cue *abort*, and a second cue that was the name of the semantic category that *DONKEY* belonged to, *a four-footed animal*. In the low-match conditions, only the original studied cue was provided.

If the predictions of the thought experiment are correct, only the low-overload condition would benefit from an increase in the encoding-retrieval match by the provision of a second retrieval cue, since that cue uniquely specifies the target *DONKEY* and maximises the diagnostic value of the retrieval environment.

Method

Participants Forty introductory psychology students participated for course credit.

Design and materials A 2 (Overload: high, low) x 2 (Match: high, low) within-subjects design was employed**.**

Forty semantic categories were selected from the Van Overshelde, Rawson, and Dunlosky (2004) and Yoon *et al*. (2004) norms. The category name was used as the second retrieval cue in the high-match conditions. From each category (e.g, *a four-footed animal*), an exemplar with a low response frequency (e.g. *DONKEY*) was selected to be a critical target, i.e. a target that would be tested in the recall phase. Response frequency refers to the proportion of responses that produced that exemplar out of the total responses for that category. High response exemplars were not selected as critical-targets in order to minimise guessing when given the category name as a retrieval cue. From each category, nine high response exemplars (e.g. *TIGER*, *ELEPHANT*) were selected to be foil-targets, i.e. targets that would be studied but not tested during recall. When studied within the same list, these foil-targets would generate high cue overload as they are subsumed by the category retrieval cue. Each target was then paired with a semantically unrelated cue word. Thus, there were altogether 40 category-lists of 10 cue-target pairs (1 critical-pair and 9 foil-pairs) each.

All words were rated for familiarity by participants who did not take part in the study but were from the same population as the experimental sample. Unfamiliar words were replaced and checked again. The lists were then divided into 4 sets of 10 lists each, which were equated for average response frequency, word frequency, and number of syllables.

Procedure A balanced latin-square procedure was used to rotate the sets across the 4 conditions in the study. For any one participant, a set was assigned to a single condition and was never repeated across conditions. In the high-overload conditions, all cue-target pairs were sampled from within the same category-list. In the low-overload conditions, each cue-target pair was sampled from a different category-list within the set.

A single trial comprised a study phase where ten cue-target pairs were displayed one at a time, with the cue appearing to the left of the target, at a rate of two seconds per pair centred on the computer monitor. To avoid primacy and recency effects, the critical cue-target pair always appeared randomly in either the fourth, fifth, sixth or seventh serial position within the sequence. After the study phase, the test cue(s) appeared and participants typed in their response before moving to the next trial. Participants were told that on some trials, an additional cue would be provided to assist recall.

The 40 trials, 10 from each condition, were randomly interspersed throughout the experiment. Eighteen filler trials that were similar to the experimental trials were also randomly interspersed. These fillers were created such that their critical-pairs were presented in either the first three or last three positions of the sequence, so as to prevent participants from noticing that the position of the critical-pair to be tested always occurred in the middle of the sequence. No subsequent analyses were performed on these fillers.

Results and discussion

The correct recall proportion is summarised in Table 1. Analysis of variance (ANOVA) revealed a significant interaction, $F(1,39) = 13.89$, $MSe = 0.01$, $p < .01$. Planned comparisons of the simple effects showed that when cue overload was low, recall was better in the high-match condition than the low-match condition, $F(1,39) = 17.14$, $MSe = 0.02$, $p < .001$. In the high-match conditions, recall was also better with low-overload than high-overload, $F(1,39) = 8.92$, $MSe = 0.02$, $p < .01$. No other simple effects were reliable.

Table 1: Recall probabilities across overload and match conditions.

	High-overload		Low-overload	
Match	*M*	*SD*	*M*	*SD*
High	.28	.20	.37	.19
Low	.28	.20	.24	.20

The pattern of results is consistent with the predictions of Nairne's (2002) thought experiment. Providing a second retrieval cue in the form of the category name of the target should increase the encoding-retrieval match relative to a single cue. However, this advantage only occurs if the second cue confers additional diagnostic information. If the second cue subsumes competing targets, as is the case in the high-overload conditions, then performance is no better than having just a single cue. This supports the contention that it is the relative diagnostic value of the cue that is important in predicting cue efficacy, rather than the absolute value of the encoding-retrieval match.

Experiment 2

Experiment 1 demonstrated that high match can be negated by high overload. Is it possible to demonstrate a situation in which the encoding-retrieval match still matters despite a large amount of overload?

Nairne (2002) argued that memory performance can be conceptualised as a joint function of the degree of the encoding-retrieval match and the degree of cue overload, where performance is proportional to the former and inversely related to the latter. If this is true, then one can hypothesise that increasing the encoding-retrieval match should cause performance to increase if the degree of cue overload does not change across conditions.

If we consider the original paradigms used to investigate encoding specificity (e.g. Newman *et al*., 1982; Thomson & Tulving, 1970), it may be easy to see why they were able to demonstrate specificity effects. The general procedure required participants to memorise a set of cue-target pairs that were pre-experimentally weakly associated, such as *TRAIN-BLACK*. Weakly associated pairs were studied so that recall by guessing is minimised. At test, some participants were re-presented with the originally studied weak cue *train*; others were given an extra-list cue that was pre-experimentally strongly associated with the target *BLACK*, such as *white*; and a final group was either given no cue at all or a weakly associated extra-list cue. Results showed that the original studied cue *train*, though weakly associated with the target, was much better in eliciting recall of *BLACK* than the other conditions, including strongly associated cues such as *white*. This was the primary basis for the encoding specificity argument – when the target *BLACK* was studied in the context of *TRAIN*, the target word was encoded in a specific manner that is distinct from the pre-experimental encoding of *BLACK* in the context of *WHITE*. This demonstrated that for a cue to be effective, it must be stored during the original encoding.

It is important to note that most experimental work on encoding specificity did not make any explicit control for the degree of cue overload. Although each cue-target pair was associated, there were presumably no associations across pairs similar to the manipulations that were done for Experiment 1. It is then entirely possible that the degree of cue overload was essentially held constant across the experimental conditions. Hence, re-presenting the original studied cue would maximally increase the encoding-retrieval match and result in the best performance.

Experiment 2 introduced a novel manipulation of cue overload to a design adapted from Newman *et al*. (1982). Participants studied weakly associated cue-target pairs such as *TEA-POT*, *AIRPLANE-BIRD*, and *ROOF-TIN*. At test, participants would be re-presented with the originally studied weak intra-list cue, e.g. *tea* for recalling *POT*; a strong extra-list cue, e.g. *feather* for recalling *BIRD*; or a weak extra-list cue, e.g., *armour* for recalling *TIN*. In the high-overload conditions, the foil cue-target pairs that were studied with the critical-pair had targets that were strongly associated with the recall cue in the respective test conditions. For example, *COFFEE*, *ICE* for the intra-list cue *tea*; *LIGHT*, *SOFT* for the strong extralist cue *feather*; and *KNIGHT*, *PROTECTION* for the weak extralist cue *armour*. In the low-overload conditions, foil-targets were unrelated to the recall cues.

At first glance, one might expect to obtain results similar to Experiment 1, in that the high-overload condition should eliminate any advantage of an encoding-retrieval match that would be obtained by providing the intra-list cue at test. However, we make the somewhat counter-intuitive prediction that while overall recall rates would drop in the high-overload conditions due to the subsuming of foil-targets under the recall cues, there should still be some evidence of facilitation due to the use of an originally studied cue relative to the extra-list cues, i.e. the encoding specificity effect, even in the high-overload conditions. This is because when the encoding-retrieval match was increased (i.e. from no match in the extra-list conditions to a match in the intra-list condition), the degree of overload did not change across recall conditions. Overload remained either high or low because all retrieval cues were manipulated to either subsume all the targets of the foil-pairs or were unrelated to those foil-targets. In this situation, the use of an intra-list cue should provide enough diagnostic information relative to extra-list cues to facilitate recall.

Method

Participants Forty introductory psychology students who did not take part in the previous experiment participated for course credit.

Design and materials A 2 (Overload: high, low) x 3 (Cue-Type: intra-list, strong extra-list, weak extra-list) within-subjects design was employed**.**

A total of 54 critical cue-target pairs, 9 in each of the 6 experimental conditions, were created based on the Nelson, McEvoy, and Schreiber (1998) norms. For the intra-list and weak extra-list conditions, a target word (e.g. *POT*) that was weakly associated with each recall cue (e.g. *tea*) was first selected. For the strong extra-list conditions, a strongly associated target was selected. In the three high-overload conditions, 7 words (e.g. *COFFEE*, *ICE*) strongly associated with each recall cue were then selected to be the subsumed targets of the foil-pairs. In the three low-overload conditions, the 7 foil-targets were not associated with the recall cues. All targets were then paired with a weakly associated word to form the cue-target pairs (e.g. *TEA-POT*, *AROMA-COFFEE*, *GLIDE-ICE*). Following Newman *et al*.'s (1982) study, a further constraint was that for both extra-list conditions, each studied cue (e.g. *ROOF*) of a critical cue-target pair was semantically unrelated to the corresponding extra-list recall cue (e.g. *armour*). This was to minimise indirect retrieval of the target via associations between the studied and recall cues. Each of the 6 experimental conditions thus had 9 lists of cue-target pairs, with each list comprising 1 critical-pair and 7 foil-pairs.

The words were then checked for familiarity using the same method as Experiment 1. Due to various constraints faced in selecting the words, it was not possible to divide them into lists and rotate them through the conditions using latin-square procedures. Lists were thus fixed across conditions and were equated on average associative strength between the recall cue and subsumed targets for the high-overload conditions. This was to ensure that recall differences cannot be attributed to differences in the amount of overload generated by the subsumed targets across the 3 cue-type conditions. The low-overload conditions were already equated as it was ensured that recall cues and foil-targets were not associated.

Procedure Instructions to participants followed those of Newman *et al*. (1982). They were asked to memorise all target words and to pay attention to the cue word that accompanied each target as it might help them to remember the target. They were also informed that recall cues at test were related to the targets but they may or may not be from the study lists. They were to recall the target that they think was related to the cue.

Eighteen trials were required to present the 54 experimental-lists without replacement. Within each list of 1 critical-pair and 7 foil-pairs, the presentation order was fixed such that the critical-pair is presented first followed by the foil-pairs in descending order of their targets' associative strength to the corresponding condition's eventual recall cue. This was done to maximise the degree of activation of the recall cue, and thus maximise cue overload at test.

A single trial comprised a study phase where 6 filler cue-target pairs were first presented, followed by a random selection of 3 lists from 3 of the 6 experimental conditions. The purpose of the fillers was to minimise primacy effects, since the critical-pair of each list was always the first pair in the within-list sequence. This resulted in 30 pairs to be studied on each trial, which closely approximated the original procedure of Tulving and Thomson (1970), which had 24 pairs for each study trial. Each pair was presented one at a time on the monitor at a rate of two seconds per pair.

At the end of each trial, the 3 recall cues for the presented conditions were shown one at a time in a random order. There was also a recall cue from one of the filler trials, otherwise participants may notice that the first few pairs were never tested. Participants typed their responses to each cue before the next one was shown. Responses to the filler cues were not analysed. After all responses to the 4 cues were made, the study phase of the next trial of 30 pairs was initiated.

Results and discussion

The correct recall proportion is summarised in Table 2. An ANOVA revealed a significant main effect of overload, $F(1,39) = 90.24$, $MSe = 0.02$, $p < .001$. Recall was better with low-overload ($M = .31$, $SD = .11$) than high-overload ($M = .14$, $SD = .07$). The main effect of cue-type was also reliable, $F(2,78) = 38.33$, $MSe = 0.03$, $p < .001$. Planned contrasts showed that the intra-list cues ($M = .32$, $SD = .16$) elicited better recall than both the strong extra-list cues ($M = .25$, $SD = .11$), $F(1,39) = 4.83$, $MSe = 0.02$, $p < .05$, and the weak extra-list cues ($M = .09$, $SD = .08$), $F(1,39) = 62.40$, $MSe = 0.02$, $p < .001$. Strong extra-list cues were better than weak extra-list cues, $F(1,39) = 66.58$, $MSe = 0.01$, $p < .001$. The interaction was not significant, $F < 1$.

Table 2: Recall probabilities across overload and cue-type conditions.

	High-overload		Low-overload	
Cue-type	*M*	*SD*	*M*	*SD*
Intra-list	.24	.21	.41	.17
Strong extra-list	.16	.10	.35	.16
Weak extra-list	.01	.04	.17	.15

The pattern of results replicated the basic findings of Newman *et al*. (1982) and Thomson and Tulving (1970). The low-overload conditions were essentially similar to these previous studies, and showed that re-presenting the original studied cue, even though it was weakly associated with the target, elicited better recall than extra-list cues, even those with strong associations with the target. In the novel high-overload conditions, the same pattern emerged, even though overall recall dropped significantly. This demonstrates that if cue overload can be experimentally held constant across retrieval conditions, then the difference in the encoding-retrieval match between extra-list and intra-list cues would lead to a facilitative recall advantage for the latter, even when cue overload is high.

General Discussion

The present experiments were designed to evaluate whether Nairne's (2002) claims that the encoding specificity principle is a myth is valid. The principle specifies that successful memory performance depends importantly on the similarity, or match, between encoding and retrieval conditions. However, Nairne (2002) argued that to improve performance, it is not really the absolute encoding-retrieval match that is critical, but rather the presence of diagnostic features that help discriminate the target from competitors. Stated differently, what matters is not the absolute encoding-retrieval match, but the relative diagnostic value of the match, which is the extent to which the retrieval cue uniquely specifies the target. Nairne (2002) supported his claims through a few thought experiments, which suggested that under the right circumstances, increasing the encoding-retrieval match can improve performance, produce no effect, or even lower performance. In this respect, he argued that the encoding-specificity principle is a myth. Specifically, the myth is that recall performance will improve to the extent that features of the retrieval cue match those present during original encoding.

Experiment 1 demonstrated that an increase in the encoding-retrieval match does not necessarily result in a recall advantage. Providing a second retrieval cue that subsumes the target should theoretically increase the functional match because there are now two features, the original studied cue and an additional feature that is present in the target. However, if this second cue also subsumes the target's competitors, the increased match is now countered by an increase in cue overload. This essentially reduced the diagnostic value of the two-cue retrieval condition, leading to recall levels that were no better than the one-cue conditions. Only when the second cue uniquely specified the target was there an increase in the relative distinctiveness of the two-cue retrieval condition over the single cue condition, leading to better recall.

Experiment 2 demonstrated that the encoding and retrieval conditions could be manipulated such that increases in the encoding-retrieval match would result in improved recall in spite of increases in cue overload. This occurred only because the degree of overload, regardless of whether it was high or low, was held constant across the retrieval conditions. Therefore, the increased match found in the intra-list cues provided information that overlapped with the original encoding, which conferred a recall advantage over the extra-list cues, and the advantage was still evident in spite of the general reduction of recall performance due to high cue overload.

At this point, it is possible to argue that having demonstrated that other factors such as cue distinctiveness also play important roles in recall need not necessarily render the encoding-specificity principle a myth. Both encoding-specificity and cue distinctiveness could play equally important roles in retrieval. Nairne (2002) addressed this point and claimed that it is misleading to give equal weight to both encoding specificity and cue distinctiveness because the main controller of performance is the distinctiveness of the retrieval cue (or diagnostic value of the encoding-retrieval match), rather than the absolute encoding-retrieval match. He suggested that when an encoding-retrieval match leads to an improvement in recall, it could be because the matching features happened to possess diagnostic features which may help one to discriminate the target item from its competitors. This was precisely what happened in Experiment 2.

In conclusion, it is apparent from the present findings that increasing the encoding-retrieval match need not necessarily facilitate recall. It could facilitate recall, when cue overload was kept constant (Experiment 2), but it could also have no effect on recall when the match increase was countered by a cue overload increase (Experiment 1). It can be argued, therefore, that an encoding-retrieval match is not intrinsically or universally beneficial, but should be effective in so far as the matching features do not overlap with the encoded features of other possible retrieval candidates. The critical factor for successful retrieval thus appears to depend on the extent to which a retrieval cue can provide diagnostic information about the target. For example, if one was asked to search for a particular individual and was told that this person is a boy in school uniform, this information would most likely be very useful for identifying this person in a room full of adults. However, the same cue would practically provide no distinctive information about that particular person in the setting of a school cafeteria full of school boys (Goh & Tan, 2006).

Is the encoding specificity principle a myth? The present findings suggest a 'yes' because in line with Nairne's (2002) proposal, an encoding-retrieval match by itself cannot be used to make unequivocal predictions about memory performance. As argued throughout, it is the relative extent to which a retrieval cue uniquely specifies a given target (i.e. cue distinctiveness) that determines successful memory performance.

References

Goh, W. D., & Tan, H. (2006). Proactive interference and cuing effects in short-term cued recall: Does foil context matter? *Memory & Cognition*, *34*, 1063-1079.

Nairne, J. S. (2002). The myth of the encoding-retrieval match. *Memory*, *10*, 389-395.

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

Newman, S. E., Cooper, M. H., Parker, K. O., Sidden, J. A., Gonder-Frederick, L. A., Moorefield, K. M., & Nelson, P. A. (1982). Some tests of the encoding specificity and semantic integration hypotheses. *American Journal of Psychology*, *95*, 103-123.

Tulving, E. (1983). Elements of episodic memory. New York: Oxford University Press.

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

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

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

Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the Battig and Montague (1969) norms. *Journal of Memory and Language*, *50*, 289-335.

Watkins, O. C., & Watkins, M. J. (1975). Buildup of proactive inhibition as a cue-overload effect. *Journal of Experimental Psychology: Human Learning and Memory*, *104*, 442-452.

Yoon, C., Feinberg, F., Hu, P., Gutchess, A. H, Hedden, T., Chen, H., Jing, Q., Cui, Y. & Park, D. C. (2004). Category norms as a function of culture and age: Comparisons of item responses to 105 categories by American and Chinese adults. *Psychology and Aging, 19,* 379-393.