Reinstating effortful encoding operations at test enhances episodic remembering

Stephen A. Dewhurst

Department of Psychology, Lancaster University, England

Karen R. Brandt

School of Psychology, University of Keele, England

Short title: Reinstating encoding operations

Word count = 3981

Please address correspondence to:

Stephen A. Dewhurst

Department of Psychology

Lancaster University

Lancaster LA1 4YF

England

Tel: 01524 593835

Fax: 01524 593744

Email: s.a.dewhurst@lancaster.ac.uk

1

Two experiments investigated the effects of reinstating encoding operations on remember and know responses in recognition memory. Experiment 1 showed that reinstating an effortful encoding task (generating words from fragments) increased remember responses at test but reinstating an automatic encoding task (reading intact words) did not. This pattern was confirmed in Experiment 2 in which words were either read intact or generated from anagrams. These findings show that repeating effortful (but not automatic) encoding operations at test cues not only the recognition of the information that was acquired via those operations but also the conscious recollection of the encoding episode.

According to many influential theories of memory, successful retrieval depends on the degree to which the operations performed at retrieval overlap with those performed at encoding. For example, the encoding specificity principle (Tulving & Thompson, 1973), the transfer-appropriate-processing framework (Morris, Bransford, & Franks, 1977), and the proceduralist approach to memory (Kolers, 1973) all emphasise the importance of the compatibility of encoding and retrieval operations (see Roediger & Guynn, 1996, for a review). According to these theories, details of the operations performed at encoding form part of the representation of the information acquired via those operations. Repeating the same operations at test cues the retrieval of that information. The question addressed in the present study was whether reinstating encoding operations at test facilitates the conscious recollection of the encoding episode or simply increases the familiarity of the material learned during that episode.

Previous research has shown that recognition memory is enhanced when participants perform the same orienting task at study and test. For example, Glisky and Rabinowitz (1985) investigated the generation effect in recognition memory by asking participants to read some of the study items and generate the others from fragments. They replicated the generation effect, whereby participants were more likely to recognise words that were generated at study rather than read (Slamecka & Graf, 1978). Test items were also presented as words or fragments, with half appearing in their study format and half in the alternative format. The recognition of words that were generated at encoding was enhanced when they were also generated at test.

A similar pattern was reported by Engelkamp et al. (1994) using the enactment paradigm. Participants were presented with a series of action phrases, such as "close the book", and instructed either to read the phrase only or to read the phrase and perform the action. Engelkamp et al. replicated the enactment effect, whereby participants remembered the enacted phrases better than the read phrases. Engelkamp et al. also found that the enactment effect was enhanced when the same phrases were enacted again at test.

These findings indicate that repeating encoding operations at test aids the recognition of items encoded via those operations. The present study used the remember-know procedure (Gardiner, 1988; Tulving, 1985) to investigate whether this effect occurs by increasing the conscious recollection of the studied items or by increasing their familiarity. In the remember-know procedure, participants are instructed to categorise each positive recognition decision as either a remember (R) or a know (K) response on the basis of their subjective experience. They make an R response if their recognition features the conscious recollection of episodic details such as thoughts, images, and associations that occurred when the item was encoded, or a K response if their recognition is based on a feeling of familiarity in the absence of conscious recollection. Findings from studies that have used the remember-know procedure indicate that R and K responses are functionally distinct (see Gardiner & Richardson-Klavehn, 2000).

Previous research suggests that both R and K responses can be influenced by the reinstatement of encoding processes. For example, Gregg and Gardiner (1994) investigated the effects of study-test modality shifts on R and K responses. Participants in Experiment 2 studied a series of words presented visually at rapid

presentation rates and their task was to check whether any of the letters were blurred. This task was chosen to direct participants' attention to perceptual attributes of the stimuli. The test items were then presented either visually or auditorially. Gregg and Gardiner found that the shift from visual processing at study to auditory processing at test significantly reduced K responses but not R responses. They suggested that the reduction in K responses occurred because the modality shift impaired the fluency with which the test items were processed.

In contrast to the findings of Gregg and Gardiner (1994), Macken (2002) found that R but not K responses were reduced by changes in context. Context in this study was determined by a combination of the colour and location of a word on a display monitor and the colour of the background. Test items were presented either in their study context or in a novel context. Words presented in the novel context were associated with significantly fewer R responses relative to words presented in their original context. Macken suggested that reinstating the encoding context at test increased R responses by cueing the conscious recollection of item-context associations formed at encoding.

If reinstating context at test increases the conscious recollection of item-context associations, the same pattern should be observed when the orienting task performed at study is repeated at test. The present study investigated this possibility. Experiment 1 adopted the read-generate manipulation used by Glisky and Rabinowitz (1985) in which words were either read or generated from fragments. Test items were presented either in the same condition as at study or in the alternative condition. Based on Macken's findings, we predicted that reinstating encoding operations at test would selectively enhance R responses, as repeating the orienting task should cue the

conscious recollection of the encoding episode. This prediction is also supported by the findings of Java, Gregg, and Gardiner (1997) that the recollection of operations carried out at encoding is often the basis for making an R response.

This prediction is at odds with recent findings by Mulligan and Lozito (2006) that reinstating encoding operations at test *reduced* recognition accuracy. They investigated the revelation effect, in which items presented in a recognition test are more likely to be selected as old if they are obscured and have to be revealed in order to be processed (Watkins & Peynircioglu, 1990). The revelation effect is similar to the generation effect except that presentation format is manipulated at test rather than at study. Mulligan and Lozito presented participants with eight-letter words or eight-letter anagrams at both study and test, with stimulus type (intact or anagram) manipulated between participants. Test items were presented in either their study format or the alternative format. Using *d'* to measure recognition accuracy, Mulligan and Lozito found that items presented as anagrams at study were associated with greater accuracy than items presented intact. In contrast, the manipulation of stimulus type at test produced the opposite effect, whereby items presented intact at test were associated with greater recognition accuracy than items presented as anagrams.

Mulligan and Lozito's findings appear to contradict theories of memory that emphasise the importance of the overlap between encoding and retrieval processes. We therefore conducted a second experiment in which we attempted to reconcile their findings with those of previous studies showing that recognition memory is enhanced by the repetition of encoding operations.

Experiment 1

Experiment 1 was essentially a replication of Glisky and Rabinovitz (1985) with the addition of a remember/know decision. Participants studied five-letter words or generated them from four-letter fragments and test items were presented in either their study format or in the alternative format. Test items presented as fragments had to be generated before a recognition decision was made.

Method

Participants. Participants were 40 undergraduates from Lancaster University, all of whom were native English speakers. They were tested individually and received a payment of £3.

Stimuli and Design. Stimuli consisted of 80 common five-letter words, divided into two sets of 40. Four-letter fragments of each word were created by deleting one letter and indicating its position with an underscore (e.g. ADU_T). Each fragment had only one possible solution. Participants studied 40 items, of which 20 were presented intact and 20 were presented as fragments. Intact words and fragments were presented in a different random order for each participant.

Procedure. Study items were presented individually on Apple Macintosh computers. Participants were instructed to read the intact words aloud. For the fragments, they were instructed to identify the word and then speak it aloud. Each item remained on the screen for three seconds. Participants were informed they would later be given a memory test for the items. The recognition test consisted of the 40 studied items plus the unstudied set. Study and lures were counterbalanced across participants. Of the 40 studied items, 20 were presented in their study format (intact or as fragments) and 20 were presented in the alternative format. Participants were again instructed to read the intact words aloud and to generate and say aloud the fragments

before making their recognition decisions. Lure items were also presented either intact or as fragments.

Participants made old/new and remember/know/guess decisions for each item and indicated their decisions by pressing labelled response keys. We included a separate "guess" (G) response in order to remove guesses from the K category. Participants were instructed to make an R response if their recognition featured the recollection of contextual details such as thoughts and images experienced at study, a K response if their decision was based on a feeling of familiarity, and a G response if they were uncertain whether or not the item had appeared at study.

Results and Discussion

Following Glisky and Rabinowitz (1985), statistical analyses consisted of 2x2 (Study Format x Test Format) repeated-measures Analyses of Variance (ANOVA). Separate analyses were conducted on R and K responses, with alpha set at .05. Guesses were not analyzed as they are typically made at chance levels. However, their mean proportions are included in the tables. Table 1 shows mean proportions of correct and false R, K, and G responses as a function of study and test formats.

(Table 1 about here)

Correct R responses showed a significant main effect of study format, whereby generated items were associated with more R responses than read items, *F* (1,39) = 60.29, *MSE* = 2.93. The main effect of test format was not significant, *F* (1,39) = 1.19, *MSE* = 2.55, *p* = .29. This was qualified by a significant study x test interaction, *F* (1,39) = 5.53, *MSE* = 2.39. Pairwise comparisons showed that the advantage for items generated at study was reliable regardless of whether they were generated or read at test, *t* = 6.99 and 3.99 respectively. The source of the interaction

was in the effects of test format. Items that were generated at test were associated with more R responses than items that were read at test, but only when the generated items had also been generated at study, $t = 2.38$. In other words, the generation effect in R responses was enhanced when the same items were generated at test. Test format did not reliably affect items that were read at study, $t = 0.84$.

The effect of study format on K responses was in the opposite direction to that observed in R responses but did not reach statistical significance, $F(1,39) = 2.92$, $MSE = 2.20$, $p = .10$. Neither the main effect of test format nor the interaction between study and test were reliable, $F < 1$. False alarms were analyzed in a 2x2 (test format x response type) ANOVA which showed that they were more likely to be categorised as K than as R responses, $F(1, 39) = 9.47$, $MSE = 3.16$. Neither the effect of test format nor the interaction with response type were reliable, $F < 1$.

Experiment 1 replicated the generation effect, whereby words that were generated at encoding were more likely to be recognised than words that were read at encoding. The presence of this effect in R but not K responses is consistent with previous findings by Gardiner (1988). The main finding from Experiment 1, however, was that R responses were also enhanced by generation at test, but only when the items had been generated at study. In other words, reinstating the same generation operations at test selectively increased R responses. These findings are consistent with those of Glisky and Rabinowitz (1985) and indicate that reinstating generation at test influences the recollection component of recognition memory. In contrast, reinstating the Read condition at test did not enhance R responses, suggesting that the reinstatement effect occurs only for effortful encoding operations. This is considered further in the General Discussion.

In contrast to the findings of Experiment 1, Mulligan and Lozito (2006) found that repeating the generation task at test *reduced* recognition accuracy (as measured by *d'*). There are a number of possible reasons for the discrepancy between their findings and ours. One is that Mulligan and Lozito manipulated study and test format between-groups, whereas we used a fully within-groups design. A second is that we did not replicate the standard revelation effect in Experiment 1, in that test format did not reliably influence recognition. It is possible that the pattern of results obtained by Mulligan and Lozito depends on the presence of a revelation effect. In order to address these two issues, we conducted Experiment 2 in which we replicated the anagram procedure used by Mulligan and Lozito but manipulated study and test format in a within-groups design.

<p style="text-align:center;">*Experiment 2*</p>

Method

The Method was the same as Experiment 1 with the following modifications: A new group of 40 undergraduates were presented with 20 eight-letter words and 20 eight-letter anagrams. Following Mulligan and Lozito (2006), a number appeared below each letter indicating its position in the solution. Each anagram had the same solution of 54687321. Each item was preceded by the instruction "read" or "anagram" and remained on the screen for 10 s or until the participants produced the correct response. Once the correct response was produced the presentation program was advanced manually by the experimenter. If the correct response was not produced within the 10 s period the experimenter provided the solution. The study list was preceded by six practise trials (three words and three anagrams). Test items (the 40 old items plus 40 new items, fully counterbalanced) were presented in either their

study format or the alternative format. Participants made oral old/new decisions followed by remember/know/guess decisions for each item judged as old.

Results and Discussion

Table 2 shows mean proportions of correct and false R, K, and G responses as a function of study and test formats. Correct and false R and K responses were analyzed in separate 2x2 (study format x test format) ANOVAs. R responses showed a significant main effect of study format, $F(1,39) = 103.71$, $MSE = 2.20$, whereby more R responses were made to generated than to read items. The main effect of test format was also significant, $F(1,39) = 8.28$, $MSE = 2.81$, and showed that more R responses were made to items that were generated at test. The interaction between study and test format was also significant, $F(1,39) = 28.92$, $MSE = 2.21$. Pairwise comparisons revealed the same pattern as was found in Experiment 1. Words generated at study were associated with more R responses, regardless of whether they were generated or read at test, $t = 11.01$ and 3.39 respectively. The source of the interaction was again located in the effects of test format. Words generated at test were associated with more R responses than words read at test, but only if they had been generated at study, $t = 5.40$. Test format did not reliably affect words that were read at study, $t = 1.33$.

(Table 2 about here)

Correct K responses were not reliably affected by study format, $F < 1$. There was a significant main effect of test format, $F(1,39) = 12.87$, $MSE = 1.52$, whereby items that were read at test were associated with more K responses than items that were generated at test. The interaction between study and test format was not significant, $F < 1$. A 2x2 (test format x response type) ANOVA on false alarms

showed that they were more likely to be categorised as K rather than R responses, F (1,39) = 23.01, MSE = 1.45. Neither the main effect of test format nor the interaction were significant, $F < 1$.

Experiment 2 replicated Experiment 1 in that repeating the generation task at test increased R responses to words that were generated at study. As in Experiment 1, reinstating the Read condition at test did not enhance memory. In contrast to Experiment 1, a revelation effect was observed whereby participants made more old responses for items generated at test than for items read at test. This finding indicates that the discrepancy between the results of Experiment 1 and those of Mulligan and Lozito (2006) cannot be due to the absence of a revelation effect. The presence of a revelation effect in Experiment 2 but not Experiment 1 is likely due to the fact that generating from anagrams is more effortful than generating from fragments.

General Discussion

The main finding from the present study was that generating words at retrieval increased R responses to words that had been generated at encoding. Previous studies have shown that recognition memory is enhanced by the repetition of encoding operations such as generation (Glisky & Rabinowitz, 1985) and enactment (Engelkamp et al., 1994). The present findings suggest that this enhancement occurs because repeating encoding operations cues the conscious recollection of items encoded via those operations, rather than simply increasing their familiarity. The findings from the Generate condition are thus consistent with theories of memory that emphasise the importance of the overlap between encoding and retrieval operations, such as the encoding specificity principle (Tulving & Thompson, 1973), the transfer appropriate processing framework (Morris, Bransford, & Franks, 1977), and the

proceduralist view of memory (Kolers, 1978). The proceduralist view of memory is particularly relevant to the present findings as it emphasises the overlap between the content of knowledge and the means by which that knowledge was acquired.

The findings from the Read condition, however, are inconsistent with the above theories in that reading words at test did not enhance the recognition of words that were read at encoding. This asymmetry was also found by Glisky and Rabinowitz (1985) and Engelkamp et al. (1994) and indicates that retrieval is not enhanced by the reinstatement of all encoding operations, but only by the reinstatement of effortful operations that enhance recognition when instantiated at encoding. Reinstating an encoding task that does not increase recognition, such as reading relative to generation, does not produce the same memory enhancement. Our findings suggest that the effortful nature of the generation task produces an episodic representation that features both the target and the operations that led to its identification. Repeating the same operations at test cues the recollection of the target. This is less likely to occur in a task such as reading, in which the operations that lead to the identification of the target are relatively automatic.

The findings from the Generate condition contrast with those of Mulligan and Lozito (2006) who found that generating words at both study and test reduced recognition accuracy. The difference between their findings and ours may be due in part to their use of a between-groups design, in which one group of participants studied words and another studied anagrams. Within each group, half the participants performed the same task at test whereas the remaining participants performed the alternative task. In contrast, the present study used a fully within-groups design. It is possible that reinstating encoding operations at test increases the relative

memorability of test items, in that test items that appear in their study format are more easily recognised in comparison to items that appear in the alternative format. Such an effect is more likely to occur in a within-groups design. It is also notable that the reduction in recognition accuracy reported by Mulligan and Lozito was located entirely in the false alarms, which were greater for anagrams than for intact words. Hicks and Marsh (1998) found that the increase in false alarms following revelation was more likely to occur in a between-groups design than a within-groups design.

Mulligan and Lozito claimed that their findings are at odds with theories of memory in which "successful memory retrieval recapitulates encoding processes" (p.10). We believe the operative phrase here is "successful memory retrieval". It is possible that the Read and Generate orienting tasks led to list-wide context effects, influencing both hits and false alarms. However, the findings from the present study suggest that the effect of reinstating encoding operations occurs primarily on an item-specific basis, in that it cues the conscious recollection of the items presented at study. If this is the case, manipulating the overlap between encoding and retrieval operations is meaningless as far as false alarms are concerned, as lures in a recognition test are by definition not presented at study. It is therefore questionable whether d' is an appropriate measure for such data, given that it includes the false alarm rates.

Other investigations of the revelation effect have also found it to be larger in false alarms than in hits (e.g., LeCompte, 1995; Watkins & Peynircioglu, 1990). LeCompte's study is particularly relevant to the present experiments as it included R and K measures (Experiment 2). Although LeCompte found no evidence of a revelation effect in overall hits, the analysis of R and K responses showed that correct R responses were reduced by revelation while correct K responses were increased. At

first glance this appears to contradict the present findings. However, in LeCompte's study the read-versus-generate manipulation was conducted only on test items (as is usually the case in investigations of the revelation effect) whereas the present study manipulated the orienting task at both study and test. Considered together, these findings suggest that when the orienting task is manipulated only at test, generated items are associated with elevated levels of familiarity, as indicated by increases in correct and false K responses. When the orienting task is manipulated at both study and test, R responses are enhanced for items that were generated at both stages, at least in a within-groups design.

Our use of the remember/know procedure implies a tacit assumption that R and K responses reflect separate processes in recognition memory. This contrasts with the view expressed in unidimensional models that R and K responses reflect different thresholds on a single underlying dimension of confidence or trace strength (e.g., Dunn, 2004). According to such models, participants make R responses to items that exceed the higher threshold and K responses to items that fall between the two thresholds. Although the present findings do not arbitrate between these accounts, our view is that unidimensional models can mimic the decision process in recognition memory but say nothing about the psychological processes on which the decision is based. They also fail to explain differences between R and K responses in terms of subjective experience and patterns of brain activity (see Gardiner & Richardson-Klavehen, 2000, for a review).

To summarise, our findings show that reinstating the generation task at test cues both the recognition of the generated items and the conscious recollection of the encoding episode. These findings are consistent with the view that R responses reflect

episodic memories that include details of the encoding context, such as the operations carried out on the stimuli, as well as the information acquired via those operations (Gardiner, 1988; Java et al., 1997; Tulving, 1985). In contrast, repeating the Read condition at test did not enhance memory. This asymmetrical pattern (see also Glisky & Rabinowitz, 1985, and Engelkamp et al., 1994) suggests that the enhancement of R responses by the reinstatement of encoding operations occurs only with effortful orienting tasks that enhance recollection when instantiated at encoding.

References

Craik, F.I.M., & Lockhart, R.S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior, 11*, 671-684.

Dunn, J.C. (2004). Remember-know: A matter of confidence. *Psychological Review, 111*, 524-542.

Engelkamp, J., Zimmer, H. D., Mohr, G. & Sellen, O. (1994). Memory of self-performed tasks: Self-performing during recognition. *Memory & Cognition, 22*, 34-39.

Gardiner, J.M. (1988). Functional aspects of recollective experience. *Memory and Cognition, 16*, 309-313.

Gardiner, J.M., & Richardson-Klavehn, A. (2000). Remembering and knowing. In E. Tulving & F.I.M. Craik (Eds), *Handbook of memory* (pp. 229-244). New York: Oxford University Press.

Glisky, E.L., & Rabinowitz, J. (1985). Enhancing the generation effect through repetition of operations. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 11*, 193-205.

Gregg, V. H., & Gardiner, J.M. (1994). Recognition memory and awareness: A large effect of study test modalities on "know" responses following a highly perceptual orienting task. *European Journal of Cognitive Psychology, 6*, 131-147.

Hicks, J.L., & Marsh, R.L. (1998). A decrement-to-familiarity interpretation of the revelation effect from forced-choice tests of recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 24*, 1105-1120.

Jacoby, L.L. (1983). Remembering the data: Analyzing interactive processes in reading. *Journal of Verbal Learning & Verbal Behavior, 22*, 485-508.

Java, R.I., Gregg, V.H., & Gardiner, J.M. (1997). What do people actually remember (and know) in "remember/know" experiments? *European Journal of Cognitive Psychology, 9*, 87-197.

Kolers, P.A. (1973). Remembering operations. *Memory & Cognition, 1*, 347–355.

LeCompte, D.C. Recollective experience in the revelation effect: Separating the contributions of recollection and familiarity. *Memory & Cognition, 23*, 324-334.

Macken, W.J. (2002). Environmental context and recognition: The role of recollection and familiarity. *Journal of Experimental Psychology: Learning, Memory and Cognition, 28*, 153-161.

Morris, C.D., Bransford, J.D., & Franks, J.J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning & Verbal Behavior, 16*, 519-533.

Mulligan, N.W., & Lozito, J.P. (2006). An asymmetry between memory encoding and retrieval. *Psychological Science, 17*, 7-11.

Roediger, H.L., & Guynn, M.J. (1996). Retrieval processes. In E.L. Bjork & R.A. Bjork (Eds.), *Human Memory*. (pp. 197-236). San Diego: Academic Press.

Slamecka, N.J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 13*, 501-518.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychologist, 26*, 1-12.

Tulving, E., & Thompson, D. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*, 352-272.

Watkins, M.J., & Peynircioglu, Z.F. (1990). The revelation effects: When disguising test items induces recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 16*, 1012-1020.

Table 1. Mean Proportions (with Standard Errors) of Correct and False Remember, Know, and Guess Responses as a Function of Study and Test Format in Experiment 1.

Remember responses

| Test Format | Study Format | | | |
	Generate	Read	Mean	False Alarms
Generate	.67 (.04)	.40 (.03)	.53 (.04)	.04 (.01)
Read	.58 (.04)	.43 (.03)	.51 (.04)	.04 (.01)
Mean	.63 (.04)	.42 (.03)		

Know Responses

| Test Format | Study Format | | | |
	Generate	Read	Mean	False Alarms
Generate	.21 (.03)	.24 (.02)	.22 (.03)	.08 (.01)
Read	.21 (.03)	.26 (.03)	.23 (.03)	.08 (.01)
Mean	.21 (.03)	.25 (.03)		

Guess Responses

| Test Format | Study Format | | | |
	Generate	Read	Mean	False Alarms
Generate	.04 (.01)	.08 (.02)	.06 (.02)	.06 (.05)
Read	.05 (.01)	.07 (.02)	.06 (.02)	.07 (.03)
Mean	.05 (.01)	.07 (.02)		

Table 2. Mean Proportions (with Standard Errors) of Correct and False Remember, Know, and Guess Responses as a Function of Study and Test Format in Experiment 2.

Remember responses

Test Format	Study Format			
	Generate	Read	Mean	False Alarms
Generate	.65 (.03)	.28 (.03)	.46 (.03)	.03 (.01)
Read	.44 (.03)	.33 (.03)	.39 (.03)	.02 (.01)
Mean	.54 (.03)	.31 (.03)		

Know Responses

Test Format	Study Format			
	Generate	Read	Mean	False Alarms
Generate	.25 (.03)	.22 (.02)	.23 (.02)	.07 (.01)
Read	.31 (.02)	.30 (.02)	.30 (.02)	.07 (.01)
Mean	.28 (.03)	.26 (.02)		

Guess Responses

Test Format	Study Format			
	Generate	Read	Mean	False Alarms
Generate	.05 (.01)	.11 (.02)	.08 (.02)	.06 (.01)
Read	.09 (.01)	.12 (.02)	.10 (.02)	.09 (.02)
Mean	.07 (.01)	.11 (.02)		