

University of Texas Rio Grande Valley

ScholarWorks @ UTRGV

Psychological Science Faculty Publications and
Presentations

College of Liberal Arts

5-2008

Recall Latencies, Confidence, and Output Positions of True and False Memories: Implications for Recall and Metamemory Theories

Jerwen Jou

The University of Texas Rio Grande Valley, jerwen.jou@utrgv.edu

Follow this and additional works at: https://scholarworks.utrgv.edu/psy_fac



Part of the [Psychology Commons](#)

Recommended Citation

Jou, Jerwen, "Recall Latencies, Confidence, and Output Positions of True and False Memories: Implications for Recall and Metamemory Theories" (2008). *Psychological Science Faculty Publications and Presentations*. 22.

https://scholarworks.utrgv.edu/psy_fac/22

This Article is brought to you for free and open access by the College of Liberal Arts at ScholarWorks @ UTRGV. It has been accepted for inclusion in Psychological Science Faculty Publications and Presentations by an authorized administrator of ScholarWorks @ UTRGV. For more information, please contact justin.white@utrgv.edu, william.flores01@utrgv.edu.



Published in final edited form as:

J Mem Lang. 2008 May ; 58(4): 1049–1064. doi:10.1016/j.jml.2007.12.003.

Recall Latencies, Confidence, and Output Positions of True and False Memories: Implications for Recall and Metamemory Theories

Jerwen Jou

University of Texas - Pan American

Abstract

Recall latency, recall accuracy rate, and recall confidence were examined in free recall as a function of recall output serial position using a modified Deese-Roediger-McDermott paradigm to test a strength-based theory against the dual-retrieval process theory of recall output sequence. The strength theory predicts the item output sequence to be in the descending order of memory strength. The dual-retrieval process theory postulates two phases in a free recall, a first direct access phase in which items are output verbatim in the weakest-to-strongest order (cognitive triage) and a second reconstructive phase in which reconstructed items are output in the strongest-to-weakest order. In three experiments, all three indicators of memory strength (latency, accuracy, and confidence) consistently showed a descending-strength order of recall both for true and false memories. Additionally, false memory was found to be output in two phases and subjects' confidence judgment of their own memory to be unaccountable by retrieval fluency (recall latency).

Keywords

recall latencies; recall output position; false memory; metamemory

The main purpose of this study is to examine the relationship between memory strength of remembered items and their sequential output order in free recall. Research results have been inconsistent regarding the relationship between these two variables. Two theoretical perspectives make different predictions. The first view will be termed *memory strength* theories (Anderson, 1976, 2005; Doshier, 1984; Gillund & Shiffrin, 1984; Norman, 2002; Wixted, Ghadisha, and Vera, 1997). These theories hold that the output sequence of items in a free recall follows the decreasing order of memory strength or activation of the items, i.e., recall output order is from the strongest to the weakest item. Also, according to these theories, the retrieval time for stronger memory is shorter than for weaker memory. The other perspective will be termed the *dual-retrieval processes* theory (Brainerd, 1995; Brainerd, Olney, & Reyna, 1993; Brainerd, Reyna, Howe, & Kevershan, 1991; Brainerd, Wright, Reyna, & Payne, 2002). This theory holds that there are two phases in the process of

© 2007 Elsevier Inc. All rights reserved.

Address for correspondence: Jerwen Jou, Department of Psychology & Anthropology, University of Texas - Pan American, 1201 West University Dr., Edinburg, TX 78539-2999, jjou@utpa.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

An earlier version of this study was presented at the 2005 Annual Meeting of the American Psychological Society in Los Angeles.

a free recall, a first verbatim retrieval, or direct access phase, and a second gist-based, constructive or reconstructive phase. Also, according to this theory, the direct retrieval process is subject to output interference, whereas the reconstructive process is not subject to this interference. Moreover, in the first verbatim phase of recall, the item with the weakest memory trace is output first and the item with the strongest memory trace is output last. This strategic recall process of giving the top priority to the weakest item is known as *cognitive triage* (Brainerd, 1995; Brainerd, Reyna, & Howe, 1990; Brainerd, Reyna, Howe, & Kevershan, 1991) and its function is to minimize the output interference for the weakest item. Although the output order in the first phase is from the weakest to the strongest, in the second reconstructive phase, the output order follows the decreasing order of strength, i.e., the strongest constructed item is output first, and the weakest one the last. Brainerd and his associates have demonstrated the cognitive triage phenomenon of the dual-retrieval recall process in many developmental studies (Brainerd, Olney, & Reyna, 1993; Brainerd, Reyna, & Howe, 1990; Brainerd, Reyna, Howe, & Kevershan, 1990, 1991). Similarly, Barnhardt, Choi, Gerkens, & Smith (2006) recently demonstrated the same phenomenon with adult subjects and found no evidence supporting the strength theory. Brainerd (1995; Brainerd et al., 1990) argued that cognitive triage in free call is not just a memory strategy but rather a *basic* memory interference minimizing process because neither children as young as age 6 (before they start to use any memory strategy) nor adults have conscious awareness of using a deliberate strategic control when cognitive triage can be observed in their recall.

On the other hand, in a free recall study using frequency of presentation as a strength manipulation, Wixted et al. (1997) found that items (in a strong and weak items mixed list) with greater memory strength yielded a shorter recall latency than items with weaker memory strength, providing strong evidence supporting the strength theory, and no evidence at all for the dual-retrieval processes theory. Therefore, this issue deserves further investigation. The primary goal of the present study is to further investigate this issue and test the two theories.

In previous studies, memory strength was either measured by the proportion of accurate recall (Brainerd, 1995) or defined by the number of times an item was studied (Doshier, 1984; Wixted et al., 1997) or by study time (Rohrer & Wixted, 1994). Also, confidence rating was found to be negatively correlated with latency but positively correlated with accuracy both in recognition (Jou, Matus, Aldridge, Rogers, & Zimmerman, 2004; Robinson, Johnson, & Herndon, 1997) and recall (Koriat, 1993; Nelson, Gerler, & Narens, 1984; Nelson & Narens, 1990). Hence, accuracy, recall latency, and confidence have been used as indices of memory strength by researchers. In the present study, all three measures were used as indices of memory strength and as converging evidence for testing the theories. In addition, the materials and a modified procedure in the Deese-Roediger-McDermott (DRM) (Deese, 1959; Roediger & McDermott, 1995) paradigm were used because the high rate of false memory this procedure generates insured the occurrence of high frequency of constructed memories in the recall output for testing the dual-retrieval processes theory.

Also to be examined in this study is the question of whether recall latency or output serial position (SP) can account for the confidence judgments subjects make of their memory. One possibility is that subjects heavily rely on the *retrieval fluency* (Kelly & Rhodes, 2002), or the ease and quickness with which information comes to mind as the basis of confidence judgment (Kelly & Lindsay, 1993; Lindsay & Kelly, 1996; Mazzoni & Nelson, 1995). Nelson and Narens (1990) referred to this idea as *confidence-determined-entirely-by-latency* hypothesis. Another possibility is that the retrieval or recall latency cannot account for the confidence or lack of confidence subjects indicate for the recalled words. In that case, subjects may use other cues to evaluate the validity of their memory and they have the

conscious access to the difference in the sources of the correct and incorrect memories (Koriat, 1993, 2007).

Jou, et al. (2004) showed that false memory as defined in the DRM paradigm produced a significantly longer recognition latency than true memory, and suggested that the activation level of false memory is lower than true memory. Therefore, still another purpose of this study is to determine whether such a latency difference between true and false memory also exists in recall. A recall latency was measured in this study as the time elapsed between the end of typing a word and the end of typing the next word in a self-paced sequential free recall test. It is assumed that the pause before typing the word contributes significantly to the word's production time. If recall latency for false memory is indeed longer than for true memory, then the distinction in response time between true and false memories can be generalized across recognition and recall. If false memory can be shown to have longer recognition and recall latencies than true memory, then false memory can be considered a weaker form of memory than true memory from the strength point of view (Wixted et al., 1997).

Experiment 1

Experiment 1 used the DRM materials and a modified DRM procedure to measure the recall latencies and the output serial positions (SPs) of the recalled items. As converging evidence, the rate of correct recall was also examined as a function of output SP. The crucial question asked is whether the recall latency/output SP function first shows a negative slope (weak items output earlier have longer latencies and strong items output later have shorter latencies) prior to the middle point and then a positive slope past the middle point as will be predicted by the dual-retrieval processes theory. The strength theory predicts a monotonically increasing latency/output SP function. A second question is whether false memory has a longer recall latency than true memory and also a different pattern of output SP distribution than true memory.

Method

Subjects—A total of 123 undergraduates at the University of Texas – Pan American participated in the experiments for extra course credit. They all met the criterion of English as their only language or their dominant language if they were bilingual.

Materials and design—The twenty-four lists of semantically associated words that produced the top 24 false memory rates were selected from the Stadler, Roediger, and McDermott's (1999) DRM associated word norms. The theme word or the critical word of each list was not presented (the critical nonpresented words, henceforth CNPW). These 24 lists of words were divided into three blocks, with list 1 to 8 as block 1, 9 to 16 as block 2, and 17 to 24 as block 3. Each subject studied and was tested on two blocks (i.e., 16 lists) with the three blocks rotated across the subjects so that each block was used for an equal number of times across subjects (i.e., one third of the subjects received blocks 1 and 2, one third received blocks 1 and 3, and one third blocks 2 and 3). This design encompassed a larger material sample to enhance the generality of the results while keeping the experiment from running too long. For the last 36 subjects, because of time constraint, each subject studied and was tested only on 12 of the 24 lists. Half of them received the even-numbered, and the other half odd-numbered lists of the 24 lists.

Procedure—A modified DRM procedure was used. Each subject was seated in a cubicle to perform the experimental task individually. The experiment was run by a computer. In the learning phase, the 16 lists of words were presented in a new random order for each individual subject, and so were the 15 words in each list. Each word was displayed for 2.5 s

with a 1 s blank screen in between words. Subjects were asked to pay close attention to the words during presentation. At the end of presentation, subjects performed backward counting for 30 s by steps of 3 starting from a random 3-digit number generated by the computer program. They were asked to count at a reasonably fast pace. At the end of the counting, the recall input screen appeared. The recall prompt was a number (starting from 1) followed by a question mark and a blank space along with a blinking cursor displayed near the upper center location of the screen. Subjects typed the first recalled word into the space and followed it with a press of the enter key. This simultaneously removed the question mark from the first entry (the number “1” and the entered word remained visible in their original positions) and brought on the second word prompt. The recall latency was measured from subjects’ pressing Enter (to start recalling a word) to pressing Enter again (to submit the input word and start the next recall). Subjects were told to submit a word by pressing Enter immediately after they completed typing it. They could skip a prompt by pressing Enter without entering a word. The recall was subject self-paced and they were not told that their recall time was being measured by the computer. However, they were told not to take a break before completing the recall for a list. They were told that they should avoid guessing as much as possible and that they could recall the words in any order they liked. At the end of the recall of each list of words, subjects could take a break and were told how many of the 16 lists they had completed. Because of time constraint, the last 36 subjects studied and were tested on only 12 lists of words (other subjects received 16 lists of words).¹

Results and Discussion

Recalled list words whose plurality was changed from the presented form or which were spelled incorrectly (but could be clearly identified) were scored as wrong by the computer program. Before the data analysis, the author changed the scoring of these words from “wrong” to “correct”. This increased the correct recall rate by about 3.5%. The correct recall rate for the list words was .584. The false recall rate of the CNPW was .363 which was close to the median false recall rate of the CNPW in the Stadler et al.’s (1999) norms.

Recall latency

The mean latency as a function of Experiment, word type (list words vs. CNPW), and recall output SP are presented in Figure 1.

The recall latency functions showed that overall the word recall time increased as more words were being recalled (or as recall progressed with the output SP). Furthermore, over and beyond this output SP main effect, at each output SP except for the first few, the recall was slower for the CNPW than for the list words. This recall latency difference could not have been caused by typing more letters on average for the CNPW than for the list words because on average the CNPW were shorter than the list words. The mean number of letters for the list words was 5.21, as compared with 4.97 for the CNPW. The difference was significant, $F(1, 118) = 26.43$, $MSE = .123$, $p < .001$. So, if other things were equal, the CNPW should on average have produced a shorter recall time if the recall latency merely measured typing time. Therefore, factors other than word length should have caused the difference in recall time. It is suggested that the differential recall onset latency (the pause before typing the word) was the source of the recall latency difference between these two types of words. This finding was consistent with findings in other prior studies showing that the recall latency of commission errors was longer than that of the correct recall (Nelson et al., 1984; also see Nelson et al., 1990). In addition, this difference tended to increase with the output SP.

¹The pattern of results from the last 36 subjects’ data was the same as from subjects who were given 16 lists of words.

An ANOVA conducted on the part of the Experiment 1 data shown in Figure 1 indicated that the main effect of word type was significant, $F(1, 118) = 33.17$, $MSE = 122,540,173$, $p < .001$, with the mean latency for the list words being 7,622 ms, and that of the CNPW 10,900 ms. The output SP main effect was significant, $F(14, 1429) = 73.66$, $MSE = 38,580,382$, $p < .001$, indicating that overall, the recall time increased with the output SP. The word type by output SP interaction was also significant, $F(14, 365) = 11.21$, $MSE = 36,842,289$, $p < .001$. The significant word type by output SP interaction confirmed the visual observation that the difference in time between these two types of words widened as the output SP progressed.

According to the dual-retrieval processes theory, during the first phase, items are output according to the *cognitive triage* principle whereby more vulnerable items (i.e., weaker memory) are output before items of stronger memory and during the second phase, items are output in the reverse of the above order (i.e., decreasing order of strength) (Brainerd, 1995). Assuming that recall latency is indicative of memory strength (Wixted et al., 1997), the latency functions in Figure 1 show no sign of outputting weaker items before stronger items in the first phase (there was no indication that in the first several output SPs, the function was negatively sloped). The functions appeared to be monotonically increasing for the list words and the CNPW. Thus, the latency data contradicted the cognitive triage principle but supported the idea of a strength-based output order.

Recall probability

The mean output SP for the list words was 5.32, and that of the CNPW was 6.34. Thus, the mean of the output SP distribution of the list words was about 1 position earlier than that of the CNPW. The difference was significant, $F(1, 118) = 20.88$, $MSE = 2.45$, $p < .001$. In this experiment, each subject studied and recalled 16 lists of 15 words each. The recall probability at each output SP was the relative frequency of recall for that output SP, i.e., the number of words output at that position divided by 16 (i.e., the total number of chances for which words could be output at that position). The mean recall output probabilities as a function of experiment and the recall output SP for the list words are presented in Figure 2 and those of the CNPW presented in Figure 3.

Figure 2 showed that the probability of the list word recall declined continuously and monotonically over the output SPs. The output SP recall probability function of the CNPW looked very different. The SP output probabilities showed two modes, an early and a later one, although the first peak was a little higher than the second. Again, the probability of output at each output SP was calculated by the summed frequency of recalled CNPWs at that output SP across the 16 lists divided by 16. Because there was only one chance to recall a CNPW for each list (instead of 15 chances as for the list words), the probability of recall for the CNPW was much lower than for the list words as can be seen in Figure 3. An ANOVA conducted on the output probability data of the list words and the CNPW of Experiment 1 showed that the main effect of word type (mean of list words = .569, mean of CNPWs = .024)² was significant, $F(1, 122) = 3019.55$, $MSE = .091$, $p < .001$, as was the main effect of output SP, $F(14, 1708) = 1076.17$, $MSE = .008$, $p < .001$. The crucial word type by output SP interaction was significant, $F(14, 1708) = 750.25$, $MSE = .010$, $p < .001$. The significant interaction confirmed that the two functions were not the same in shape. Although the mean output SP for the CNPW was 1 position later than that of the list words, the first (or the higher) peak of the output SP distribution was actually at position 2 (i.e., a

²These means, .569 and .024, were the average probability of recall per output SP for list words and CNPWs, respectively. Since there was only one chance of recalling a CNPW over the 15 output SPs, the average probability of recalling a CNPW *per output SP* was quite low. However, the mean probability of recalling a CNPW *per list* would be .024 times 15, or .36, which was the number reported at the beginning of the Results section of this experiment.

very early position) which contradicts the idea of the dual-retrieval processes theory that gist-based memory is output in the second constructive phase as well as other prior reports that CNPWs are output relatively late in the recall sequence (Brainerd et al., 2002, 2003; Brainerd et al., 2005; Roediger & McDermott, 1995). Thus, the mean CNPW output SP fails to reveal some important details of the CNPW output distribution such as whether the distribution is unimodal or bimodal and where the mode or modes of the distribution are located. The present data showed that the CNPWs are output in recall in two phases, rather than just in the second phase, and that the mean output position typically reported in prior studies fails to reveal the bimodal nature of the distribution. It should be noted that the gist-based part of the output that occurred in the first phase produced the same recall latency as the list words (see Figure 1).

The trough in the output probability function (roughly position 5) of the CNPWs coincided with the point in the recall latency functions at which the false and the true memory time functions started to diverge from each other. This is consistent with the idea that false memory that is output within the first 4 or 5 positions in the recall sequence is highly accessible just as the list words and may have been generated during the encoding stage, whereas those that are output after position 5 may likely have passed through constructive steps and generated during the retrieving stage (Hicks & Starns, 2005; Brainerd et al., 2003).

Conditional correct recall rate

Besides latency, accuracy has been used as a measure of memory strength (Brainerd, 1995). Therefore, accuracy was examined as a function of output SP to see if it corroborated the latency results. The accuracy rate of recall for each output SP for each subject was calculated by dividing the number of correctly recalled words by the total number of words recalled at that output SP. The proportion was the relative frequency of correct recall conditional on recalled words (note that it is different from the probability of correct recall which is the total number of correctly recalled words divided by 16. The CNPW was excluded from this calculation). The mean rates of correct recall as a function of Experiment and output SP are presented in Figure 4.

As is evident in Figure 4, accuracy of recall monotonically decreased over the output SPs. There was no indication of accuracy starting lower, going higher, then going lower, the nonmonotonic function found in some studies (see Brainerd, 1995). An ANOVA indicated that the decrease over the output SPs was significant, $F(14, 1454) = 62.19$ $MSE = .022$, $p < .001$. Thus, two indicators of memory strength, latency and accuracy, provided converging evidence that the recall output order follows the order of memory strength and not the sequence predicted by the cognitive triage principle.

If accuracy is an indicator of memory strength, then, correctly recalled words should have a lower latency than incorrectly recalled words (not counting the recalled CNPWs). Indeed, the overall mean latency for the correctly recalled words was 5966 ms as compared with 13,606 ms for the incorrect recalled words. The difference was significant, $F(1, 113) = 114.12$, $MSE = 30,266,920$, $p < .001$. Thus, recall accuracy and recall latency were consistent as an indicator of a common underlying construct, namely, memory strength.

Several conclusions can be drawn from Experiment 1. First, consistent with recognition latency (Jou et al., 2004), false memory produces longer recall latency than true memory, strengthening the conclusion that it is a weaker form of memory than true memory, despite its vivid illusion-like characteristics. Second, both the latency and the accuracy data indicated that free recall output sequence follows the order of memory strength, and not the order in which the weakest item is output first and the strongest last during the first direct retrieval phase, and the strongest first and the weakest last during the second gist-based

reconstructive phase as claimed by the cognitive triage principle of the dual-retrieval processes theory. Third, false or constructed memory was output in two phases, rather than only in the second phase as suggested by the dual-retrieval processes theory and other investigators (Brainerd et al., 2002, 2003, 2005; Barnhardt et al., 2006).

Experiment 2

The main purpose of Experiment 2 was to collect converging evidence for further testing the strength-based and dual-retrieval process theory of recall. Confidence has been found to be correlated with memory accuracy (Jou et al., 2004; Nelson et al., 1990; Robinson, Johnson, & Herndon, 1997). Several questions were asked in this experiment. The first question is whether the confidence rating function will show any signs of cognitive triage in operation. Assuming that weaker memory is associated with lower confidence, the confidence rating function should show a positive slope in the first half of the confidence function if weaker memory is output before stronger memory in the first phase of recall. The second question is whether output SPs associated with indicators of greater memory strength are also associated with higher confidence ratings, and vice versa. The third question is whether subjects are conscious (as may be indicated by confidence rating) of the distinction between true and false memory as reflected in recall latency difference. The fourth question is whether the confidence difference, if any, can be accounted for by the latency difference alone (the confidence-determined-entirely-by-latency hypothesis, see Nelson et al., 1990) or by output SP alone. This hypothesis is based on the idea of *output or retrieval fluency* determining confidence judgment, namely, that items that comes to mind quickly or early in recall must be the items that are studied (Koriat, 1993; Kelly & Lindsay, 1993; Lindsay & Kelly, 1996; Mazzoni & Nelson, 1995).

Method

Subjects—Eighty-two undergraduates at the University of Texas-Pan American participated in this experiment for extra course credit. They met the same language criterion as in Experiment 1.

Materials and design—Materials and design were the same as in Experiment 1.

Procedure—The procedure was the same as in Experiment 1 with the following exceptions. First, a confidence rating screen immediately followed the submitting of a recalled word by pressing the Enter key. Subjects were asked how confident they were that the word they entered was on the list of words they studied. They pressed one of the 5 numbers on the numeric keypad to indicate their confidence, with 1 being “not confident at all”, 2 “somewhat confident”, 3 “confident”, 4 “very confident”, and 5 “absolutely positive”. The pressing of a number key within that numerical range would bring back the recall input screen for entering the next word. Second, an instruction was displayed on the screen after the backward counting and before the onset of the recall screen asking the subjects to make sure that they were ready for the recall before they should press the Enter key to start the recall. This instruction was intended to minimize the extra initial word recall latency possibly caused by the task switching (Gopher, Armony, & Greenspan, 2000) from counting to recall.

Results and discussion

The correct recall rate for the list words was .588, and the false recall rate for the CNPW was .348.

Recall latency

The mean recall latency as a function of Experiment, word type, and output SP are presented in Figure 1. The pattern of the latency function paralleled that of Experiment 1 very well. The main effect of word type was significant (mean latency of list words = 7411 ms; mean latency of CNPW = 10,545 ms), $F(1, 79) = 14.62$, $MSE = 173,434,185$, $p < .001$, as was that of output SP, $F(14, 993) = 51.35$, $MSE = 33,775,164$, $p < .001$. The interaction of word type and output SP was also significant, $F(14, 238) = 43.64$, $MSE = 5,032,517$, $p < .001$, again confirming the observation of the increasing difference in latency between the two types of words with the increase of output SP. Thus, the recall latency results of Experiment 1 were replicated in Experiment 2 when a confidence rating was required for a recalled word.

Recall probability

The mean output SP of the distribution for the list words was 5.46, and that for the CNPWs was 6.89. The difference was significant, $F(1, 79) = 36.90$, $MSE = 2.25$, $p < .001$. Again, the distributions of the output of the list words and the CNPWs replicated those of the Experiment 1 quite remarkably. The mean recall probability for the list words as a function of experiment and output SP are presented in Figure 2 and that of the CNPWs in Figure 3. The patterns of results replicated those of Experiment 1 very closely. An ANOVA on the data from Experiment 2 of these two types of words showed that the main effect of word type (mean of list words = .588, mean of CNPWs = .023) was significant, $F(1, 81) = 2147.36$, $MSE = .091$, $p < .001$, as was the main effect of output SP, $F(14, 1134) = 855.58$, $MSE = .006$, $p < .001$. The crucial interaction between these two factors was also significant, $F(14, 1134) = 606.89$, $MSE = .008$, $p < .001$. Again, the output of the true memory seemed to be carried out in one distribution, but that of the false memory in two distributions.

Conditional correct recall rate

The mean rates of conditional correct recall as a function of Experiment and output SP are presented in Figure 4. The decrease in accuracy over the output SPs was significant, $F(14, 1002) = 28.40$, $MSE = .020$, $p < .001$. The overall mean latency for the correctly recalled words was 5989 ms as compared with the incorrectly recalled words' latency of 13,528 ms. The difference was significant, $F(1, 76) = 58.10$, $MSE = 38,848,990$, $p < .001$. Again, the result of the conditional probability of correct recall replicated the counterpart in Experiment 1.

Confidence rating

Confidence rating was found to be negatively correlated with recall latency, $r = -.317$, $t(12814) = -37.79$, $p < .001$, that is, the longer was the recall time, the lower the confidence rating, consistent with the results reported in the literature (Nelson et al., 1984; Robinson et al., 1997). Figure 5 presented the mean confidence ratings as a function of word type and output SP.

An ANOVA showed that the mean confidence rating for the list words (4.46) was significantly higher than that of the CNPW (3.07), $F(1, 79) = 239.64$, $MSE = 2.06$, $p < .001$. The output SP main effect was significant, $F(14, 993) = 20.83$, $MSE = .662$, $p < .001$. The word type by output SP interaction was also significant, $F(14, 238) = 128.20$, $MSE = .050$, $p < .001$. Thus, the confidence rating pattern was consistent with the pattern of recall latency in that higher confidence was associated with list words than the CNPW, that the overall confidence rating decreased with output SP, and that the gap in confidence between these two types of words increased with the output SP.

A visual inspection of Figure 5 suggested that even for the first two output SPs, the confidence gap between the list words and CNPWs seemed substantial. Therefore, the confidence ratings of the list words and the CNPWs for the first 2 output SPs were compared to see if the confidence difference for the first 2 positions was indeed significant. The ANOVA showed that the confidence difference for the first 2 positions was significant (mean confidence rating of 4.77 for the list words versus 4.13 for the CNPW), $F(1, 46) = 24.48$, $MSE = .912$, $p < .001$. The important message provided by the significant difference in confidence rating between the two types of words, and especially by that of the first 2 positions was that subjects were metacognitively aware of the difference between the two types of words even when they recalled CNPWs as early as the first or second word, and even when their recall response was as quick as that of the list words (see Figure 1). This suggested that subjects had some knowledge in their metamemory about the validity of their own memory that was not reflected in latency or in the output SP, and that they had conscious access to memory traces or sources and can base their confidence rating on detecting the difference in strength or source (Hart, 1967; Koriat, 1993, 2007; Nelson, Gerler, & Narens, 1984; Yaniv & Meyer, 1987). Therefore, the confidence-determined-entirely-by-latency (Nelson, 1990) hypothesis was not supported by the present data. To further examine the retrieval fluency hypothesis, the mean confidence ratings were plotted against latency in units of a second as a function of word type in Figure 6.

The number of data points in the very long latency brackets was small, and so the means were more variable in those categories than in lower-latency categories. As was evident, the latency could not account fully for the effect on confidence rating. There was clearly a robust effect on confidence due to the type of words. An ANOVA showed that the main effect of word type (mean confidence for list words = 4.46, and mean confidence for CNPW = 3.23) was significant, $F(1, 79) = 149.85$, $MSE = 2.27$, $p < .001$, as was the main effect of latency, $F(12, 918) = 26.83$, $MSE = .596$, $p < .001$. The word type by latency interaction was also significant, $F(12, 193) = 6.18$, $MSE = .505$, $p < .001$. This result again suggested that subjects had cues to rely on to evaluate their own memory other than how quickly the item came to mind. In other words, they may actually consciously know more about their memory than their recall behavior appeared to indicate. Alternatively, this result can suggest that subjects' recall behavior was not fully governed by their metacognitive knowledge. Finally, consistent with the latency data, the mean confidence rating for the correctly recalled words (4.57) was significantly higher than that of the incorrectly recalled words (2.62) (the incorrectly recalled words did not include the CNPWs), $F(1, 76) = 303.71$, $MSE = .499$, $p < .001$.

Experiment 3

Jou et al. (2004) found that the recognition latency for the critical presented words (CPW) was shorter than that of the list words as well as of the CNPWs and hence suggested that CPWs have higher activation level than both the list words and CNPWs. This finding is consistent with the idea that if the critical word is studied, it should accrue greater activation than other list words because in addition to the activation derived from studying the critical word directly, the activations of all the other list words converge on that word (Brainerd & Wright, 2005; Roediger, McDermott, & Robinson, 1998). Barnhardt et al. (2006) found that the mean output SP of the CPWs and that of the CNPW were both located near the middle point of the recalled array, but with the former being just prior to the middle point, and the latter just past the middle point. Based on this finding, they argued that the dual-retrieval process, but not the strength theory, was supported. The basis for this argument is that, according to the cognitive triage principle, the verbatim item with the greatest memory strength (i.e., CPWs) should be output as the last word in the first direct retrieval phase, and the reconstructed item with the highest strength (i.e., CNPW) should be output as the first

item in the second phase, which was indeed what they found. They indicated that if items were output in the descending order of memory strength as claimed by the strength theory, then the CPW should be output as the first word of the recalled array. There are a few points in the above arguments that may be questionable. First, the finding of the mean CPW output position located earlier than that of the CNPW is also consistent with the prediction of the strength theory. Second, in that study, there was no measure of the memory strength for the items that were output prior to the CPW. Third, the assumption that according to the strength theory, the CPW should be output as the first word of the array may not be valid in that it fails to take into consideration the random variation that accompanies all mental processes. A more tenable version of this assumption (if the strength theory is true) would be that the *distribution* of the output positions of the CPWs should be located earlier than that of the list words, and that of the list words should be located earlier than that of the CNPW (and that the three distributions certainly overlap as in most cases of memory variable distributions).

The purpose of Experiment 3 is to address these issues and further test the strength theory against the dual-retrieval process theory by presenting the critical words in half of the word lists. Regarding recall latency, the strength theory predicts that the CPWs have the shortest mean latency, the list words the next shortest, and the CNPWs the longest mean latency, and regarding the output SPs, the distribution mean of the CPW output positions should be located prior to that of the list words, and that of the list words prior to that of the CNPWs. The dual-retrieval process theory will make the same prediction regarding the relative order of the magnitudes of the mean latencies of CPWs and CNPWs (although it is less clear what it will predict on the relative latency magnitude for the CNPW), but an opposite prediction regarding the relative order of the mean output positions for these two types of words, namely, that the mean output position of the CPWs should be located *after* that of the list words.

Method

Subjects—One hundred and five undergraduates at the University of Texas – Pan American participated in this experiment for extra course credit. They met the same language criterion as in Experiment 1.

Materials and design—Materials and design were the same as in Experiment 1 except that half of the lists (odd-numbered) of presented words included the critical words, and the other half (even-numbered) did not. The two halves of the lists were counterbalanced across the subjects. When the critical word was included in the presentation list, the last word of the original list in the Stadler et al. (1999) norms was dropped to make the total number of words equal to 15 across all lists.

Procedure—The procedure was identical to that of Experiment 1. Subjects were not given any information about the manipulation on the presentation of the critical word.

Results and discussion

The correct recall rate for the list words was .541 (the recall of the CPWs was not counted as list words), that of the CPWs was .830, and the false recall rate for the CNPWs was .348.

Recall latency

The mean recall latencies as a function of word type and output SP are presented in Figure 7.

The recall latency results basically replicated those of Experiments 1 and 2. The only new observation was that the CPW seemed to produce a somewhat lower latency than the other

list words. An ANOVA³ showed that the word type main effect (mean of list words = 6415 ms, mean of CPW = 4186 ms, and mean of CNPW = 7996 ms) was significant, $F(2, 197) = 20.89$, $MSE = 66,166,670$, $p < .001$. A Newman-Keuls post hoc test indicated that each mean was significantly different from the other. However, as indicated in the results section of Experiment 1, the critical words were on average shorter than other list words. Could the shorter latency for the CPWs have derived from the shorter length of the words? A multiple regression that partialled out the effect of the word length showed that CPWs still yielded significantly shorter RTs than other list words, with the partial coefficient being 1299 ms (equivalent to the difference between CPWs and list words after partialing out the word length effect) $t(14218) = 4.67$, $SE = 278.46$, $p < .001$. Thus, the CPW produced the shortest mean recall latency, the list words the second shortest, and the CNPW the longest mean latency. Therefore, the critical words were produced either faster or more slowly than other list words depending on whether they were presented or not in study. The main effect of output SP was significant, $F(11, 1107) = 78.87$, $MSE = 27,881,490$, $p < .001$ as was the word type by output SP interaction, $F(22, 500) = 4.25$, $MSE = 940,841$, $p < .001$. Again, there was no sign in latency of a negative slope in the early portion of the latency function indicative of cognitive triage in the output process.

Recall probability

The mean output SP of the distribution of the list words was 5.34, that of the CPW 3.96, and that of the CNPW 6.03. The difference was significant, $F(2, 198) = 59.80$, $MSE = 1.88$, $p < .001$. A Newman-Keuls post-hoc test indicated that each mean was significantly different from the other. Thus, the CPW produced both the shortest latency and the earliest mean output position, the list word the second shortest latency and the second earliest mean output position, and the CNPW the longest latency and the latest mean output position. This perfect ordinal rank correspondence between the magnitude of latency and mean output positions unequivocally supports the strength theory, and contradicts the cognitive triage principle. If subjects had adopted cognitive triage in the output process, the mean of the output distribution of CPWs (the strongest verbatim item) should be after that of the list words.

The mean probabilities of recall for the list words as a function of experiment and output SP are presented in Figure 2, those of the CNPW in Figure 3, and those of the CPW as a function of output SP presented in Figure 8.

The patterns of results for the list words and the CNPW replicated those of Experiments 1 and 2 quite well. The noticeable discrepancy of the Experiment 3 list-word curve was its relatively lower probabilities of recall at the beginning several output SPs. This was the result of excluding the correctly recalled CPWs from the category of list words in the calculation of mean probability of recall. An ANOVA on the data of the three types of words showed that the word type main effect (mean probability of list word recall = .541, mean of CPW = .055, and mean of CNPW = .024) was significant, $F(2, 204) = 2245$, $MSE = .059$, $p < .001$ (Newman-Keuls test indicated that each mean was significantly different from the other), as was the main effect of output SP, $F(14, 1456) = 906.91$, $MSE = 006$, $p < .001$. The crucial word type by output SP interaction was also significant, $F(28, 2912) = 376.49$, $MSE = 009$, $p < .001$. Thus, the visually apparent different shapes of the recall probability distribution curves was confirmed. More specifically, the shape of the CPW function closely resembled that of the list words and gave no indication of bimodality, which made it very distinguishable from the function of the CNPWs.

³Because of very few data points in each of the last several output SPs for the CPW and the CNPW, one error term in the ANOVA had a zero variance when these output SPs were included in the analysis. For this reason, only the data of the first 12 output SPs were included in this ANOVA to avoid this problem.

Conditional correct recall rate

The mean conditional probabilities of correct recall as a function of output SP are presented in Figure 4. The effect of output SP was significant, $F(14, 1220) = 36.03$, $MSE = .022$, $p < .001$. Again, there was no indication that the accuracy first increased and then decreased as should be the case according to the dual-retrieval process theory. The mean latency for the correctly recalled words was 5345 ms, and that of the incorrectly recalled words was 10,634 ms. The difference was significant, $F(1, 101) = 66.61$, $MSE = 21,734,685$, $p < .001$.

Thus, when distributions of output positions are compared, the order of the mean output positions of the three distributions coincided with that of the latency measure of strength. Although the CPW did not always emerge at the first position of the recalled array, as a distribution, it emerged earlier than the list words as a whole, which contradicted the prediction of the dual-retrieval process theory. Finally, the latency and accuracy results were consistent in that the 3 or 4 items that were output prior to the CPW's mean output position (3.96) showed no signs of going from weak to strong items.

Experiment 4

Experiment 4 was conducted to address a few theoretical and methodological issues in the first three experiments. First, although subjects were instructed to recall the words in any order they liked (with this instruction sentence printed on the monitor screen in all capital letters to make it more noticeable) in the first three experiments, it might still be possible that subjects might interpret the numerical recall prompt "1" as "the strongest in memory", "2" as the "second strongest in memory", etc. To rule out this possibility, only a question mark was used as a recall prompt for each output SP in Experiment 4. Also, in addition to their receiving the same instruction displayed on the screen as in the first three experiments, subjects were given an oral instruction emphasizing that they could recall the words in any order they liked before they entered the testing cubicle to start the experiment.

A second issue addressed in this experiment was the relationship between recall latency and output positions on the one hand and memory strength on the other. The definition of memory strength in the first three experiments was not based on a direct manipulation of strength although the recall latency difference between the CNPWs and list words can be interpreted as resulting from a presentation frequency manipulation (with the CNPWs presented a zero time and the list words presented once). The assumed latency/strength relation will be reaffirmed if a direct manipulation of presentation frequency of list words shows that words that are presented two times produce a shorter recall latency than words presented once. Additionally, the strength theory will be further shored up if it can also be shown that words that are presented two times will be recalled overall earlier than words that are presented once or never presented (CNPW). Finally, the program running the experiment was modified to code the data in such a way that computing the correlation between the input and output orders was made easier. If there is a negative correlation between these two orders, it suggests that items presented later tend to be recalled earlier. This result will be consistent with the strength theory since one can assume that the more recently presented items have higher activation levels than earlier presented items.

Method

Subjects—One hundred and 82 undergraduate psychology students participated in this experiment for course extra credit.

Materials, design and procedure—In general, materials, design, and procedures were the same as in Experiment 1 with the following exceptions. For each list of words, 4 of the

15 words were randomly selected for each individual subject for a repeated presentation in the study phase. The other 11 words were presented once. The two repeated presentations of the 4 words were randomly intermixed with the words that were presented once. Subjects were told that some words would be presented twice, others just once, and that they should still recall the repeatedly presented words once. It was emphasized to them prior to their entering the testing cubicle and during the experiment (in an instruction printed in all capital letters) that they could recall the words in any order they liked. Also, the number before the question mark was removed from the recall prompt at each output SP.

Results and discussion

The correct recall rate for the list words was .585, that of the repeated words was .717, and the false recall rate for the CNPWs was .344. The correlation between the word input order and the output order was $r = -.50$, $t(55192) = -135.70$, $SE = .004$, $p < .001$. This suggested that words seen later in the study phase tended to be recalled earlier in the output and vice versa.⁴

Recall latency

The mean recall latencies as a function of word type and output SP are presented in Figure 9.

An ANOVA showed that the word type main effect (mean of repeated words = 5712 ms, mean of nonrepeated words = 7209 ms, mean of CNPWs = 10,269 ms) was significant, $F(2, 353) = 48.56$, $MSE = 121,083,593$, $p < .001$. A Newman-Keuls post-hoc test showed that each mean was significantly different from the other. Thus, repeated presentation of the words in study did shorten the response latency of the words. The CNPWs can be considered to have been presented zero time. Therefore, the relative lengths of the recall latencies correlated inversely with the number of presentations very well. The output SP main effect was significant, $F(14, 2190) = 120.68$, $MSE = 52,868,752$, $p < .001$, as was the word type by output SP interaction, $F(28, 2374) = 10.96$, $MSE = 20,814,146$, $p < .001$. The interaction reflected the increasing diverging of the RT functions with the increase in output SP (see Figure 9). The finding that a manipulation of memory strength by varying the frequency of presentation for different list words resulted in different recall latencies reinforces the idea that recall latency and memory strength are related causally and that false memory is overall a weaker form of memory than true memory.

Recall probability

A crucial result bearing on testing the strength theory against the cognitive triage account was the relative mean output SPs of the three output distributions for the three types of words. The mean output SP of the repeated words was 4.99, that of the nonrepeated words was 5.79, and that of the CNPWs was 6.47. An ANOVA showed that the differences were significant, $F(2, 353) = 64.84$, $MSE = 1.50$, $p < .001$. The Newman-Keuls test indicated that each mean output SP was significantly different from the other. Thus, consistent with the RT results, the relative order of the mean output SPs of the output distributions was in a perfect inverse relationship with the number of presentations of the words in the study phase, i.e., the mean output SP of the repeated words was earlier than that of the presented-once words which was in turned earlier than that of the CNPWs.

⁴A negative correlation is not evidence that a primacy effect plays no role in the item output order although it reduces such a probability. Buffer items were not used to eliminate the possible primacy effect in this study because including the first couple of items in the data actually provided a more powerful test for the two competing theories if indeed the first couple of items have the strongest memory traces.

The mean probabilities of correct recall of list words (with the repeated words excluded) as a function of output SP are presented in Figure 2, those of the CNPWs presented in Figure 3, and those of the repeated words only in Figure 10.

As can be seen in the figure, the recall probability function for the list words was much lower than those of the first three experiments, especially in the early parts of the function. This was caused by the high frequencies with which the repeated words were recalled in earlier output SPs. An ANOVA on the recall probability of these three types of words showed that the main effect of word type (mean of list words = .393, mean of repeated words = .191, and mean of CNPWs = .023) was significant, $F(2, 362) = 2631.10$, $MSE = .036$, $p < .001$. The Newman-Keuls test showed that each word type mean was significantly different from the other. Both the output SP effect, $F(14, 2534) = 1274.12$, $MSE = .006$, $p < .001$, and the word type by output SP interaction, $F(28, 5068) = 171.44$, $MSE = .012$, $p < .001$, were significant.

In addition, a trend test was conducted on the 4 CNPW recall probability functions from the 4 experiments shown in Figure 3. However, an ANOVA was first conducted to determine if there was an experiment by output SP interaction in the functions. The ANOVA using experiment and output SP as factors indicated that the experiment main effect was not significant, $F < 1$, nor was the experiment by output SP interaction, $F(42, 6832) = 1.14$, $MSE = .002$, $p = .25$, suggesting that the 4 functions had the same overall recall levels as well as the same bimodality pattern. Only the output SP main effect was significant, $F(14, 6832) = 29.61$, $MSE = .002$, $p < .001$. Therefore, a test for bimodality was performed on the average of the 4 functions. However, a standard test for a quartic trend might not be the right test for assessing this bimodal pattern because the 15 output SPs were not equally divided by the two arches and the two arches were far from symmetrical. To solve this problem, four linear regression analyses were performed, with the first one testing the first upward slope (output SP 1 to output SP 2), the second testing the first downward slope (output SP 2 to output SP 5), the third testing the second upward slope (output SP 5 to output SP 8), and the fourth testing the second downward slope (output SP 8 to output SP 15). If all four linear regressions show significant slopes, then the bimodality pattern is confirmed. The first regression test showed that the upward slope of .017 was significant, $t(981) = 4.77$, $SE = .004$, $p < .001$, the second test showed that the downward slope of $-.007$ was significant, $t(1965) = -6.83$, $SE = .001$, $p < .001$, the third test showed that the second upward slope of .005 was significant, $t(1965) = 4.57$, $SE = .001$, $p < .001$, and the fourth test showed that the second downward slope of $-.005$ was significant, $t(3933) = -16.50$, $SE = .0003$, $p < .001$. Thus, the observed bimodality of the 4 recall functions of the CNPWs from the 4 experiments was statistically reliable.

Conditional correct recall rate

The mean rate of conditional correct recall as a function of experiment and output SP are presented in Figure 4. The decrease in accuracy over the output SPs was significant, $F(14, 2224) = 92.18$, $MSE = .021$, $p < .001$. The recall latency for the correctly recalled words (7165 ms) was significantly shorter than that of the incorrectly recalled words (11,320 ms), $F(1, 179) = 18.78$, $MSE = 83,155,616$, $p < .001$.

Experiment 4 removed the sequential numbers from the recall prompts but obtained the same results as in the first three experiments. More important, a manipulation of memory strength produced unequivocal evidence indicating that memory strength and recall latency were causally related, as were memory strengths and the mean output SPs. The consistent data strongly suggest that recall latency and output SPs both reflect the same underlying variable of memory strength.

General Discussion

Some studies reported that the experience of false memory is so realistic that it cannot be distinguished from true memory (Gallo, McDermott, Percer, & Roediger, 2001; Lampinen, Neuschatz, & Payne, 1999; Payne, Elie, Blackwell, & Neuschatz, 1996; Togli, Neuschatz, & Goodwin, 1999), whereas others indicated that the two experiences are not the same (Miller, Baratta, Wynveen, & Rosenfeld, 2001; Johnson, Foley, Suengas, & Raye, 1988; Johnson, Hashtroudi, & Lindsay, 1993; Jou et al., 2004; Mather, Henkel, & Johnson, 1997; Schooler, Gerhard, & Loftus, 1986; Norman & Schooler, 1997). The findings from this study on this issue are that the two types of memory can be distinguished subjectively in confidence rating, and to a large extent, objectively in recall latency, and in the *distribution* of the item output. These results are consistent with the findings in studies using PET (see Norman & Schacter, 1997 for a review), P300 (Miller et al., 2001), lateralized brain potentials (Fabiani, Stadler, & Wessels, 2000), and recognition latency (Jou et al., 2004). It is suggested that a measure (e.g., recognition and recall latencies) taken concurrently with the ongoing memory processes or immediately after the recognition or recall (e.g., the confidence rating in this study) is more sensitive to the differences between these two types of memories than one taken after a substantial delay as in some previous studies cited above.

This study examined several indicators of memory strength for true and false memories as a function of output SP in free recall. All three indicators consistently showed across four experiments that items were output in the order of decreasing memory strength whether recalled items were the presented words or words constructed from gist. There was no indication in latency, or in recall accuracy rate, or in confidence rating, of outputting items from the weakest to the strongest in the first section of the recalled sequence. The nonmonotonic output SP functions reported in Brainerd and his associates' studies (Brainerd, Olney, & Reyna, 1993; Brainerd, Reyna, & Howe, 1990; Brainerd, Reyna, Howe & Kevershan, 1991; see Brainerd, 1995 for a summary) were not found in this study. In addition, when the mean SPs of the CPWs output distribution, of the list word output distribution, and of the CNPW's output distribution were compared in Experiment 3, and those of the repeated words, nonrepeated words, and CNPWs compared in Experiment 4, the order of the three mean output SPs agreed with the prediction of the memory strength theory. How could the fast and early outputs of some portions of CNPWs be reconciled with this conclusion? One idea to accommodate this finding is that some false memories are quite strong or have activation levels equal to those of true memories. Although accuracy is correlated with strength and earliness of output, strength (or activation level) may be dissociated from accuracy under certain circumstances.

Why was Barnhardt et al.'s conclusion not the same as from this study when the methods used in these two studies were generally similar? Actually, the reported part of Barnhardt et al.'s data was consistent with the present data, i.e., the mean output SP of the CPWs was located before that of the CNPWs. The different conclusion reached by Barnhardt et al. was based on the part of their data that was not examined, i.e., the mean output SP of the noncritical list words. Also, there was no position-by-position measure of memory strength between the beginning of the recall output sequence and the near-middle point (i.e., the mean output SP of the CPWs) in their study. Barnhardt et al. assumed that the memory strengths of the recalled items prior to the middle point increased up to the mean CPW output SP. But, that may not be the case. Therefore, it is possible that the results from Barnhardt et al.'s and this study were in fact consistent.

Also important is the question of why Brainerd and his associates found cognitive triage in free recall whereas the present study and others (Wixted et al., 1997) did not. The answer may lie in the study-test-cycles method used in Brainerd et al.'s studies. One mechanism

Brainerd (1995) suggested which can lead to weak-to-strong output order was item position shift from trial to trial. In the study-test-cycles paradigm, the memory strength measure for the current trials (trial *i*) is based on the recall accuracy of the items on the preceding trial (trial *i*-1). It was assumed that when the items are randomly presented for study over the repeated trials, the near-end position items regress toward the mean and the near-middle position items shift toward the two ends from trial *i*-1 to trial *i*. Items presented near the end positions on trial *i* (which tend to be the middle-position items on trial *i*-1 and hence classified as weak items on trial *i*) have the benefit of being recalled early and more accurately. This interpretation is based on the assumption of recalling the recency part of trial *i* items first and hence is a strength-based account for triage. Another alternative interpretation is the selective special processing account. According to this idea, after several cycles of studying and testing, subjects have accumulated some error-success frequency information about the items and will give special processing to the weaker items. That is, items that fail to be recalled on trial *i*-1 (hence classified as weak) will receive special processing on trial *i* and be recalled earlier on trial *i*. In examining these accounts, Brainerd (1995) did not seem to rule out the possibility that the cognitive triage phenomenon may be the product of the study-test cycles and hinted that “strength-ordering relationships of some sort should be apparent on the first free-recall test (before error-success counts have begun to accumulate)” (p. 132).

Although constructed (false) memory was found to be indeed output in two phases in all three experiments, the finding contradicted the claim of the dual-retrieval process theory that constructed memory is output in the second phase of recall. The present data showed robustly that a large portion of the constructed memory is output in the first phase. In fact, all three experiments showed that the major peak of the output distribution of the CNPWs was in the first phase. Can the false recall produced in the first phase be what is referred to as *phantom recollection* (Brainerd et al., 2003; Brainerd, Wright, Reyna, & Mojarin, 2001)? The phantom recollection is a kind of false memory that is experientially so realistic that it cannot be distinguished from true memory phenomenologically. The output positions and the recall latencies of the CNPWs that were produced in the first phase of recall are indeed indistinguishable from those of the memory of the studied words. However, the confidence rating results clearly showed that subjects were conscious of the difference between these two types of words. In that light, these recalls do not meet the criteria of phantom recollection of its subjective isomorphism with true memory. The question for future research is why subjects recall these words so early and so quickly if they were not as confident about these recalls? Thus, recall fluency (or latency), and output earliness (SP) could not account for all the variations in confidence rating. As demonstrated by the first few recalled CNPWs, subjects did not depend on the ease with which items came to mind as the sole cue for making the confidence judgment. They likely had a conscious access to the different sources of the memory.

Contrary to the typical report that false recall is output in later SPs, for example, position 6 (Brainerd et al., 2003; Roediger & McDermott, 1995), or 65% down the recall sequence (Barnhardt et al., 2006), this study has found that false recall is output in two bouts, an early and a later one. In addition, the relative frequency of the early one is higher than that of the later one. More important, the early bout of CNPW output is associated with a shorter recall latency that is indistinguishable from that of the list words, whereas the later bout is associated with a noticeably longer latency than that of the list words. Although the discontinuity in confidence rating function between the first and the second bout of CNPW recall was not as distinct as in the recall latency functions, there was some moderate indication of that transition. The observation of the consistent pattern of discontinuity between the two output clusters of the CNPWs across different dependent measures and across the four experiments leads one to question the idea that CNPWs are recalled late in

the recall sequence (Brainerd et al., 2003; Brainerd et al., 2005; Roediger & McDermott, 1995) as well as the notion of a unitary form of false memory. Although the recalled CNPWs are considered false memory, they are actually a mixture of almost two qualitatively different batches, i.e., the early batch with equal output fluency (in output SP and latency) with the studied words and the late batch with distinctively lower output fluency than the studied words. When the two batches with very different qualities are averaged, important information about the false memory can be obscured. Take the mean output SP as an example. Although the overall mean CNPW output position is quite late in the sequence (e.g., 6.03), the major mode of the bimodal output distribution is actually at a very early position (2).

It is noteworthy that when the output SP is held constant, the additional time needed to produce the CNPW in excess of the time needed to produce a list word can be plausibly attributed to online construction that the subjects may be performing right before outputting the word (Brainerd et al., 2003). The CNPWs that did not take any longer time than the studied words to output within the first three output SPs might not have been constructed during recall. They seemed to be ready in the memory buffer to be output at that time point just as the studied words. The differential output time patterns for the CNPWs that are output early and late in sequence are consistent with the idea that false memory can be generated either in the encoding stage or in the retrieving stage (Smith, Gerkens, Pierce, & Choi, 2002). Based on the latencies of the first phase of false recall, it is difficult to argue that the studied words are retrieved directly from verbatim traces whereas the CNPWs are constructed abstractly from semantic gist (Barnhardt et al., 2006; Brainerd et al., 2002).

Although the CNPWs produced in the first phase of the false memory output seem quite similar to veridical memory in the earliness and quickness of output, the confidence associated with them is lower than with the veridical memory. This suggests that these CNPWs exceed the criterion for outputting, but fall short of a higher standard, (e.g., “very confident”). The different recall latencies and confidence ratings associated with the words within the recalled category suggest that not all recalled words are equal in memory strength. Their strengths seem to vary continuously. These findings are compatible with the concept of a single underlying dimension of memory strength (Donaldson, 1996; Hirshman & Henzler, 1998; Hirshman & Master, 1997; Xu & Bellezza, 2001) along which one or more than one criterion can be set. The higher confidence ratings subjects placed on the list words than on the recalled CNPWs can be considered a second, more conservative criterion they established along the continuum of strength within the “recalled” category.

In conclusion, the results from all four experiments unambiguously supported a strength-based theory (Anderson, 1976; Doshier, 1984; Gillund & Shiffrin, 1984; Norman, 2002; Wixted et al., 1996) rather than the dual-retrieval processes theory of free recall (Brainerd, 1995; Barnhardt et al. 2006). Results from Experiment 2 suggested that subjects likely have a conscious metacognitive access to the internal and external sources of their memory at least in the contexts of the present experiments even when their recall behavior appeared otherwise.

Acknowledgments

The first three experiments of this study were supported by a NIH MBRS-SCORE grant (No. 516863) to J.J. I thank Joseph Foreman, Alma Delarbre, and Jennifer Tsai for collecting data, Jim Neely and two anonymous reviewers for very helpful comments, and Bob Greene for editorial comments, and Professors Armstrong, Buckman, Santiago, Leach, and Guerra for allowing their nonpsychology students to participate in the experiments for extra course credit.

References

- Anderson, JR. Language, memory and thought. Hillsdale, NJ: Erlbaum; 1976.
- Anderson, JR. Cognitive Psychology. New York: Worth; 2005.
- Barnhardt TM, Choi H, Gerkens DR, Smith SM. Output position and word relatedness effects in a DRM paradigm: Support for a dual-retrieval process theory of free recall and false memories. *Journal of Memory and Language*. 2006; 55:445–467.
- Brainerd, CJ. Interference processes in memory development: The case of cognitive triage. In: Dempster, FN.; Brainerd, CJ., editors. *Interference and inhibition in cognition*. New York: Academic Press; 1995. p. 105-139.
- Brainerd CJ, Olney CA, Reyna VF. Optimization versus effortful processing in children's cognitive triage: Criticisms, reanalyses, and new data. *Journal of Experimental Child Psychology*. 1993; 55:353–373.
- Brainerd CJ, Payne DG, Wright R, Reyna VF. Phantom recall. *Journal of Memory and Language*. 2003; 48:445–467.
- Brainerd CJ, Reyna VF, Howe ML. Cognitive triage in children's memory: Optimal retrieval or effortful processing? *Journal of Experimental Child Psychology*. 1990; 49:428–447. [PubMed: 2348160]
- Brainerd CJ, Reyna VF, Howe ML, Kevershan J. The last shall be first: How memory strength affects children's retrieval. *Psychological Science*. 1990; 1:247–252.
- Brainerd CJ, Reyna VF, Howe ML, Kevershan J. Fuzzy-trace theory and cognitive triage in memory development. *Developmental Psychology*. 1991; 27:351–369.
- Brainerd CJ, Wright R. Forward association, backward association, and false-memory illusion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2005; 31:554–567.
- Brainerd CJ, Wright R, Reyna VF, Mojardin AH. Conjoint recognition and phantom recollection. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2001; 27:307–327.
- Brainerd CJ, Wright R, Reyna VF, Payne DG. Dual-retrieval processes in free and associative recall. *Journal of Memory and Language*. 2002; 46:120–152.
- Deese J. On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*. 1959; 58:17–22. [PubMed: 13664879]
- Donaldson W. The role of decision processes in remembering and knowing. *Memory & Cognition*. 1996; 24:523–533.
- Doshier BA. Degree of learning and retrieval speed: Study time and multiple exposures. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1984; 10:541–574.
- Fabiani M, Stadler MA, Wessels PM. True but not false memories produce a sensory signature in human lateralized brain potentials. *Journal of Cognitive Neuroscience*. 2000; 12:941–949. [PubMed: 11177415]
- Gallo DA, McDermott KB, Percer JM, Roediger HL. Modality effects in false recall and false recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2001; 27:339–353.
- Gallo DA, Roberts MJ, Seamon JG. Remembering words not presented in lists: Can we avoid creating false memories? *Psychonomic Bulletin & Review*. 1997; 4:271–276. [PubMed: 21331837]
- Gopher D, Armony L, Greenspan Y. Switching tasks and attention policies. *Journal of Experimental Psychology: General*. 2000; 129:308–339. [PubMed: 11006903]
- Gillund G, Shiffrin RM. A retrieval model for both recognition and recall. *Psychological Review*. 1984; 91:1–67. [PubMed: 6571421]
- Hart JT. Memory and the memory-monitoring process. *Journal of Verbal Learning and Verbal Behavior*. 1967; 6:685–691.
- Hicks JL, Starns JJ. False memories lack perceptual detail: Evidence from implicit word-stem completion and perceptual identification tests. *Journal of Memory and Language*. 2005; 52:309–321.
- Hirshman E, Henzler A. The role of decision processes in conscious recollection. *Psychological Science*. 1998; 9:61–65.

- Hirshman E, Master S. Modeling the conscious correlates of recognition memory: Reflections on the remember/know paradigm. *Memory & Cognition*. 1997; 25:345–351.
- Johnson MK, Foley MA, Suengas AG, Raye CL. Phenomenal characteristics of memories for perceived and imagined autobiographical events. *Journal of Experimental Psychology: General*. 1988; 117:371–376. [PubMed: 2974863]
- Johnson MK, Hashtroudi S, Lindsay DS. Source monitoring. *Psychological Bulletin*. 1993; 114:3–28. [PubMed: 8346328]
- Jou J, Matus YE, Aldridge JW, Rogers DM. How similar is false recognition to veridical recognition objectively and subjectively? *Memory & Cognition*. 2004; 32:824–840.
- Koriat A. How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review*. 1993; 100:609–639. [PubMed: 8255951]
- Koriat, A. Metacognition and consciousness. In: Zelazo, PD.; Moscovitch, M.; Thompson, E., editors. *Cambridge handbook of consciousness*. New York: Cambridge University Press; 2007. p. 289-325.
- Kelly CM, Lindsay DS. Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory and Language*. 1993; 32:1–24.
- Kelly, CM.; Rhodes, MG. Making sense and nonsense of experience: Attributions in memory and judgment. In: Ross, BH., editor. *The psychology of learning and motivation: Advances in research and theories*. New York: Elsevier Science; 2002. p. 293-320.
- Lampinen JM, Neuschatz JS, Payne DG. Source attributions and false memories: A test of the demand characteristics account. *Psychonomic Bulletin & Review*. 1999; 6:130–135. [PubMed: 12199307]
- Lindsay SD, Kelley CM. Creating illusions of familiarity in a cued recall remember/know paradigm. *Journal of Memory and Language*. 1996; 35:197–211.
- Mather M, Henkel LA, Johnson MK. Evaluating characteristics of false memories: Remember/know judgments and memory characteristics questionnaire compared. *Memory & Cognition*. 1997; 25:826–837.
- Mazzoni G, Nelson TO. Judgments of learning are affected by the kind of encoding in ways that cannot be attributed to the level of recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1995; 21:1263–1274.
- Miller AR, Baratta C, Wynveen C, Rosenfeld JP. P300 latency, but not amplitude or topography, distinguishes between true and false recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2001; 27:354–361.
- Nelson TO, Gerler D, Narens I. Accuracy of feeling-of-knowing judgments for predicting perceptual identification and relearning. *Journal of Experimental Psychology: General*. 1984; 113:282–300. [PubMed: 6242754]
- Nelson, TO.; Narens, L. Metamemory: A theoretical framework and new findings. In: Bower, G., editor. *The psychology of learning and motivation: Advances in research and theory*. New York: Academic Press; 1990. p. 125-173.
- Norman KA. Differential effects of list strength on recollection and familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2002; 28:1083–1094.
- Norman KA, Schacter DL. False recognition in younger and older adults: Exploring the characteristics of illusory memories. *Memory & Cognition*. 1997; 25:838–848.
- Payne DG, Elie CJ, Blackwell JM, Neuschatz JS. Memory illusions: Recalling, recognizing, and recollecting events that never occurred. *Journal of Memory and Language*. 1996; 35:261–285.
- Robinson MD, Johnson JT, Herndon F. Reaction time and assessments of cognitive efforts as predictors of eyewitness memory accuracy and confidence. *Journal of Applied Psychology*. 1997; 82:416–425. [PubMed: 9190148]
- Roediger HL, Jacoby JD, McDermott KB. Misinformation effects in recall: Creating false memory through repeated retrieval. *Journal of Memory and Language*. 1996; 35:300–318.
- Roediger HL, McDermott KB. Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1995; 21:803–814.

- Roediger, HL.; McDermott, KB.; Robinson, KJ. The role of associative processes in creating false memories. In: Conway, MA.; Gathercole, SE.; Cornoldi, C., editors. *Theories of memory*. Vol. II. Hove, East Sussex UK: Psychology Press; 1998. p. 187-245.
- Rohrer D, Wixted JT. An analysis of latency and interresponse time in free recall. *Memory & Cognition*. 1994; 22:511–524.
- Schooler JW, Gerhard D, Loftus EF. Qualities of the unreal. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1986; 12:171–181.
- Smith SM, Gerkens DR, Pierce BH, Choi H. The roles of associative responses at study and semantically guided recollection at test in false memory: the Kirkpatrick and Deese hypotheses. *Journal of Memory and Language*. 2002; 47:436–447.
- Stadler MA, Roediger HL, McDermott KB. Norms for word lists that create false memories. *Memory & Cognition*. 1999; 27:494–500.
- Toglia MP, Neuschatz JS, Goodwin K. Recall accuracy and illusory memories: When more is less. *Memory*. 1999; 7:233–256. [PubMed: 10645381]
- Xu M, Bellezza FS. A comparison of the multimemory and detection theories of know and remember recognition judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2001; 27:1197–1210.
- Yaniv I, Meyer DE. Activation and metacognition of inaccessible stored information: Potential bases for incubation effects in problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1987; 13:187–205.

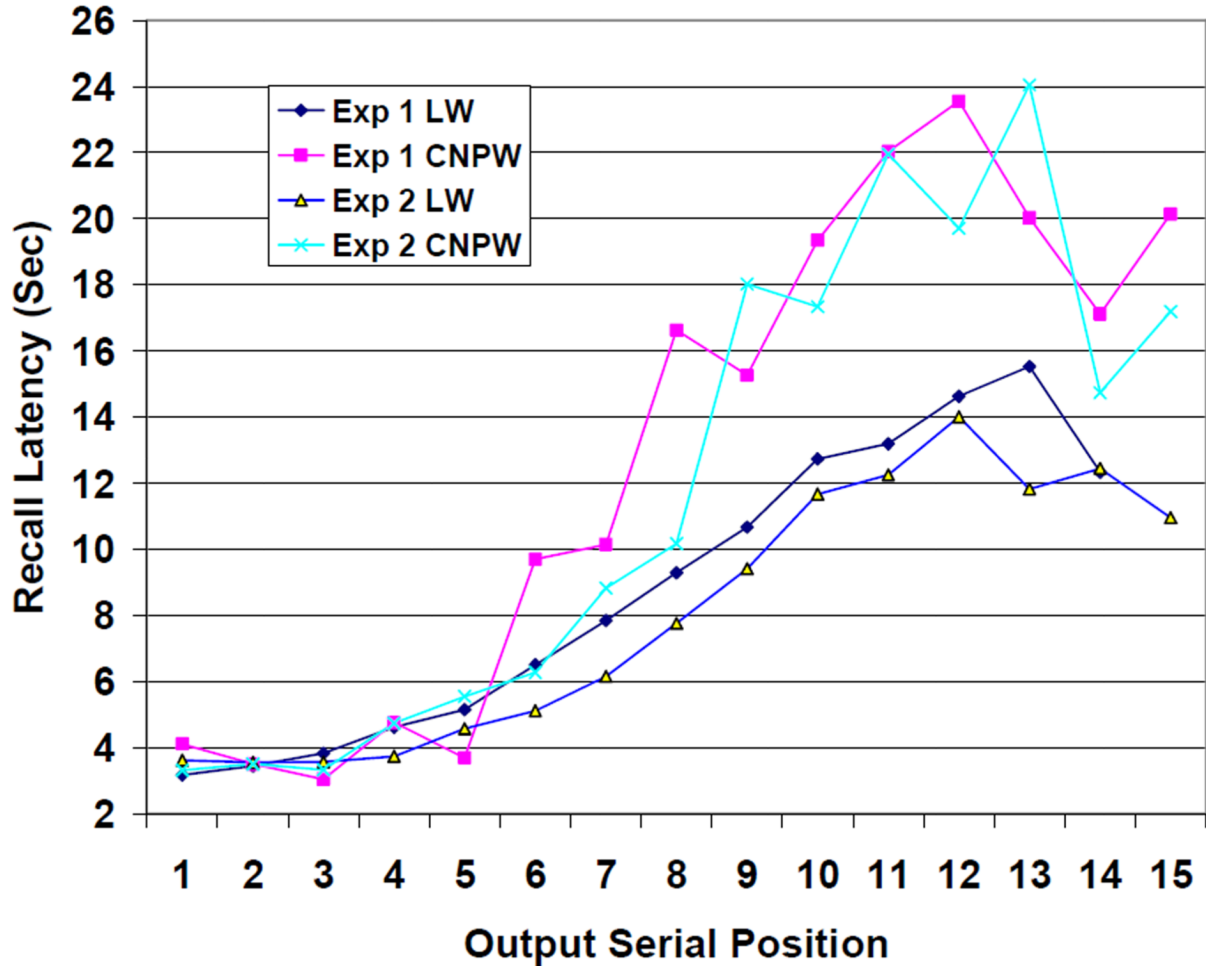


Figure 1.
 Mean recall latencies as a function of experiment, word type, and output serial position of Experiments 1 and 2.

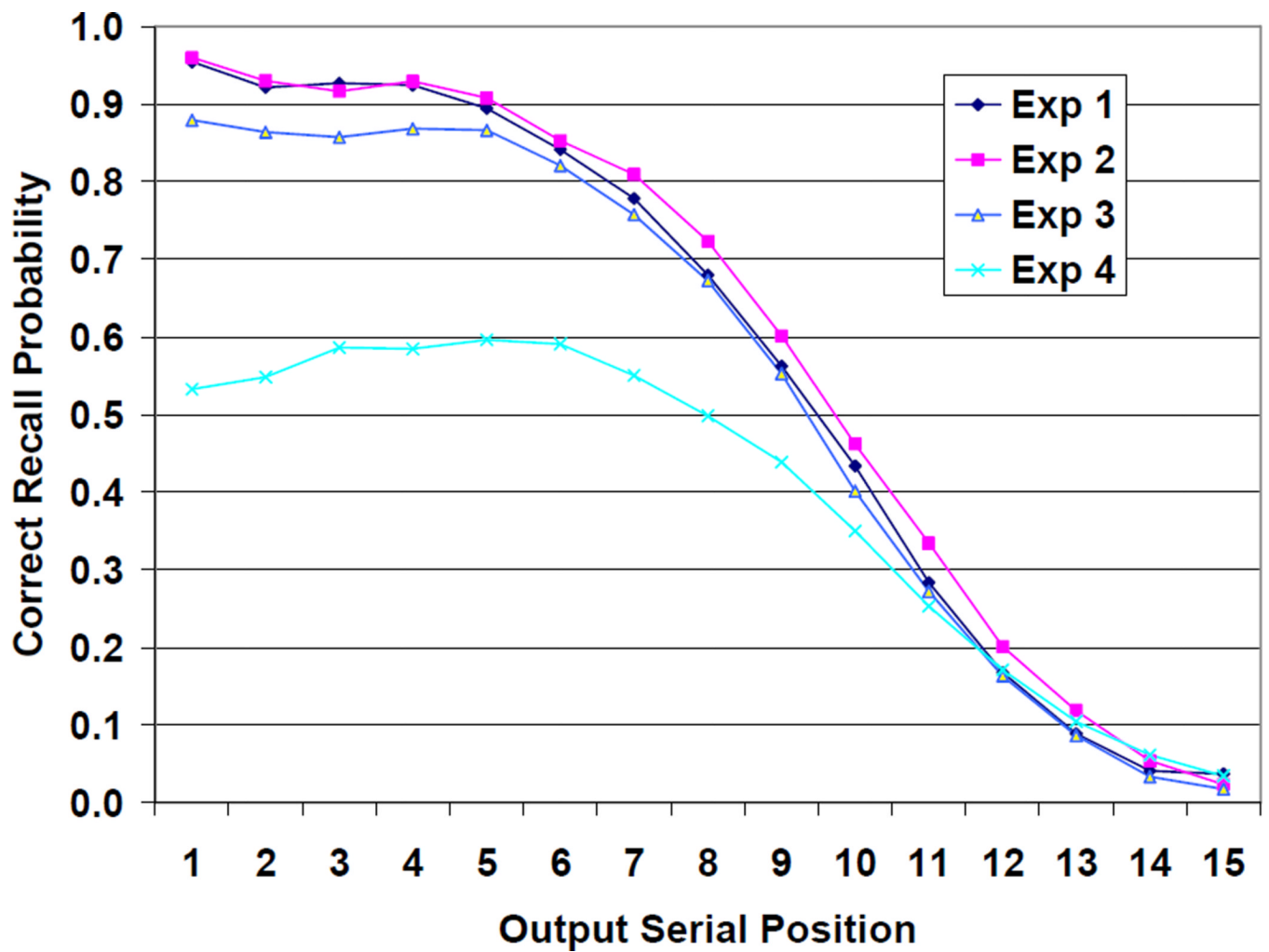


Figure 2. Mean recall probabilities of list words as a function of experiment and output serial position of Experiments 1, 2, 3, and 4.

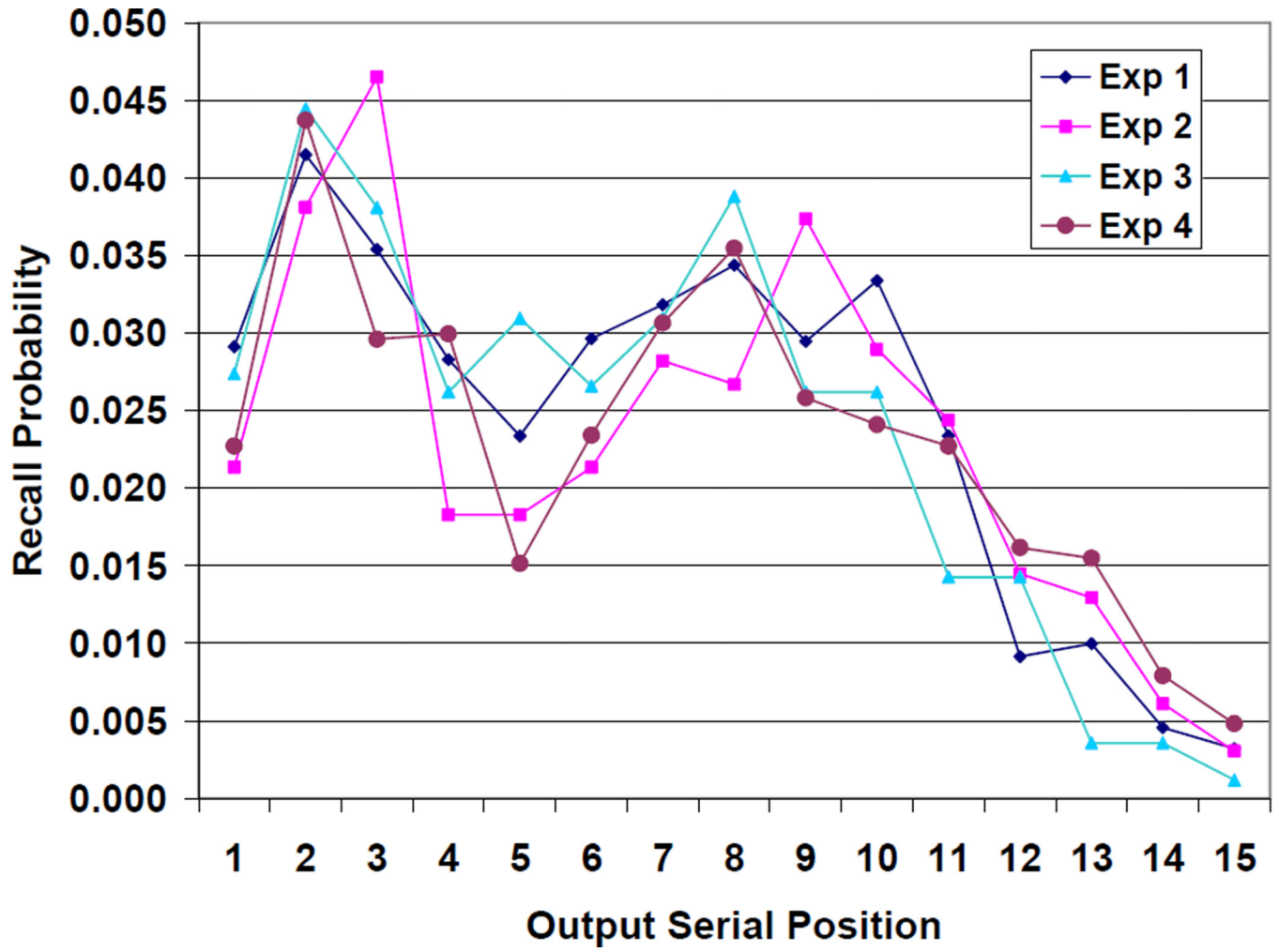


Figure 3. Mean recall probabilities of critical nonpresented words as a function of experiment and output serial position of Experiments 1, 2, 3, and 4.

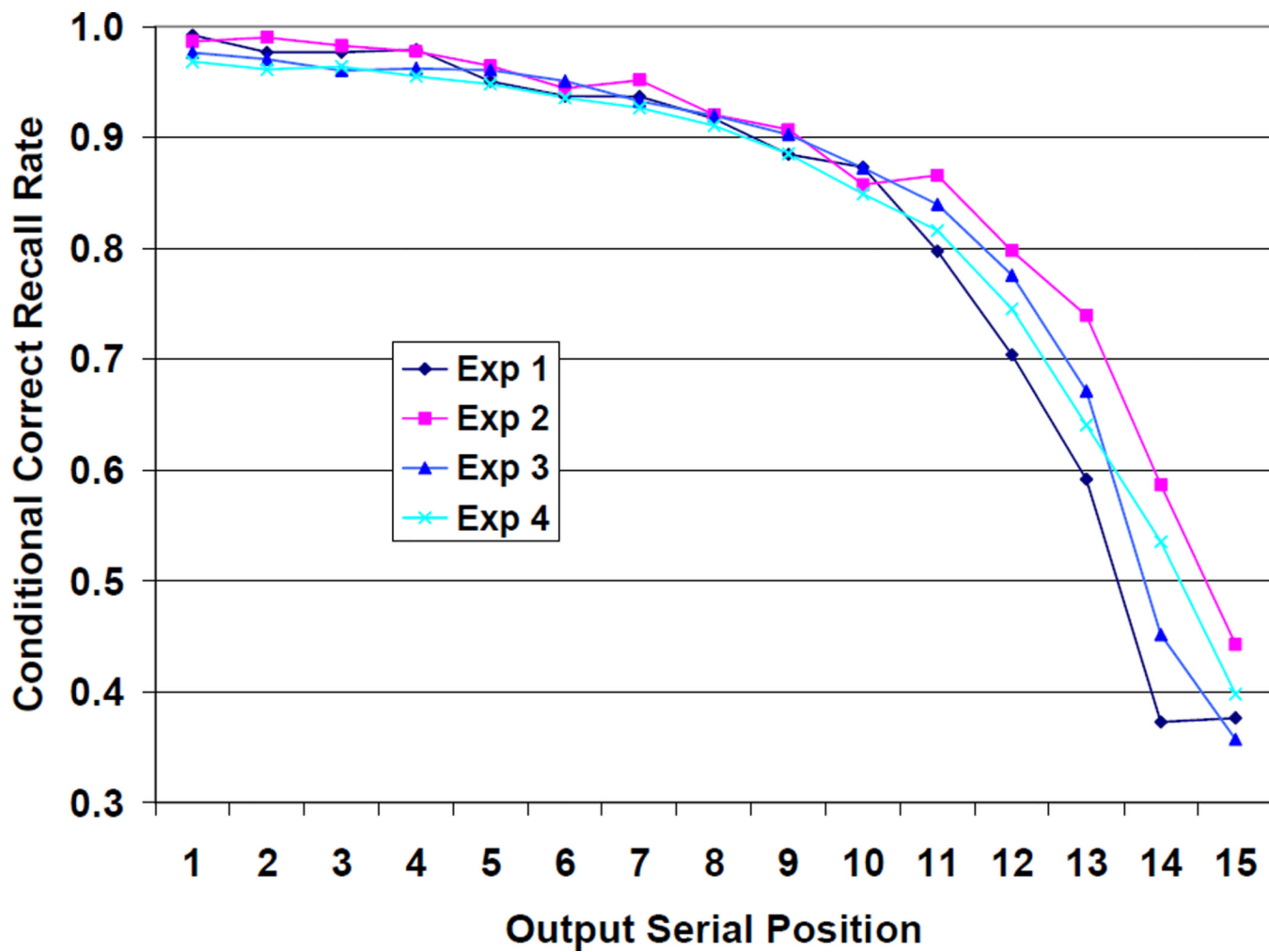


Figure 4. Mean conditional correct recall rates as a function of experiment and output serial position of Experiments 1, 2, 3, and 4.

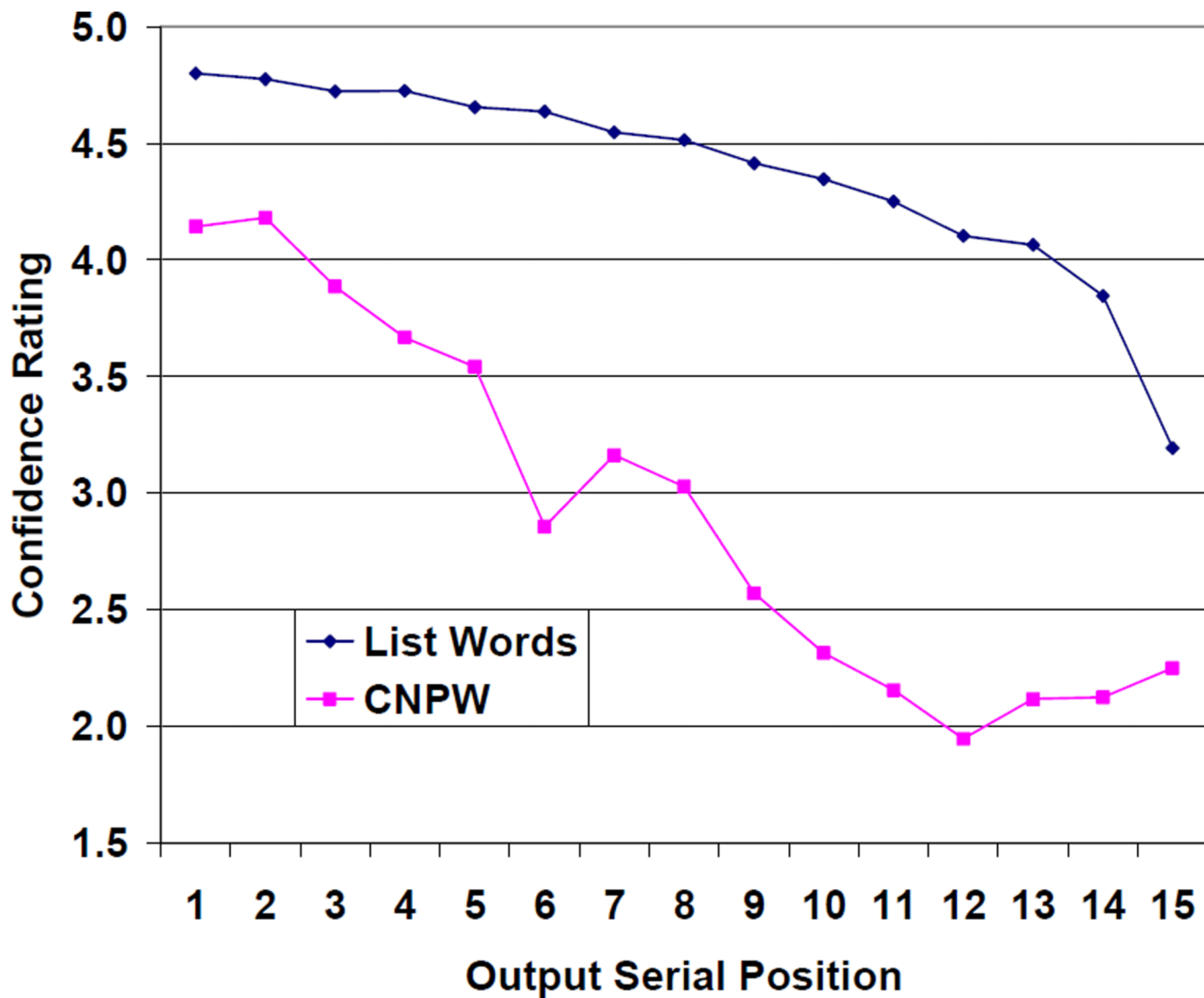


Figure 5.
 Mean confidence ratings as a function of word type and output serial position of Experiment 2.

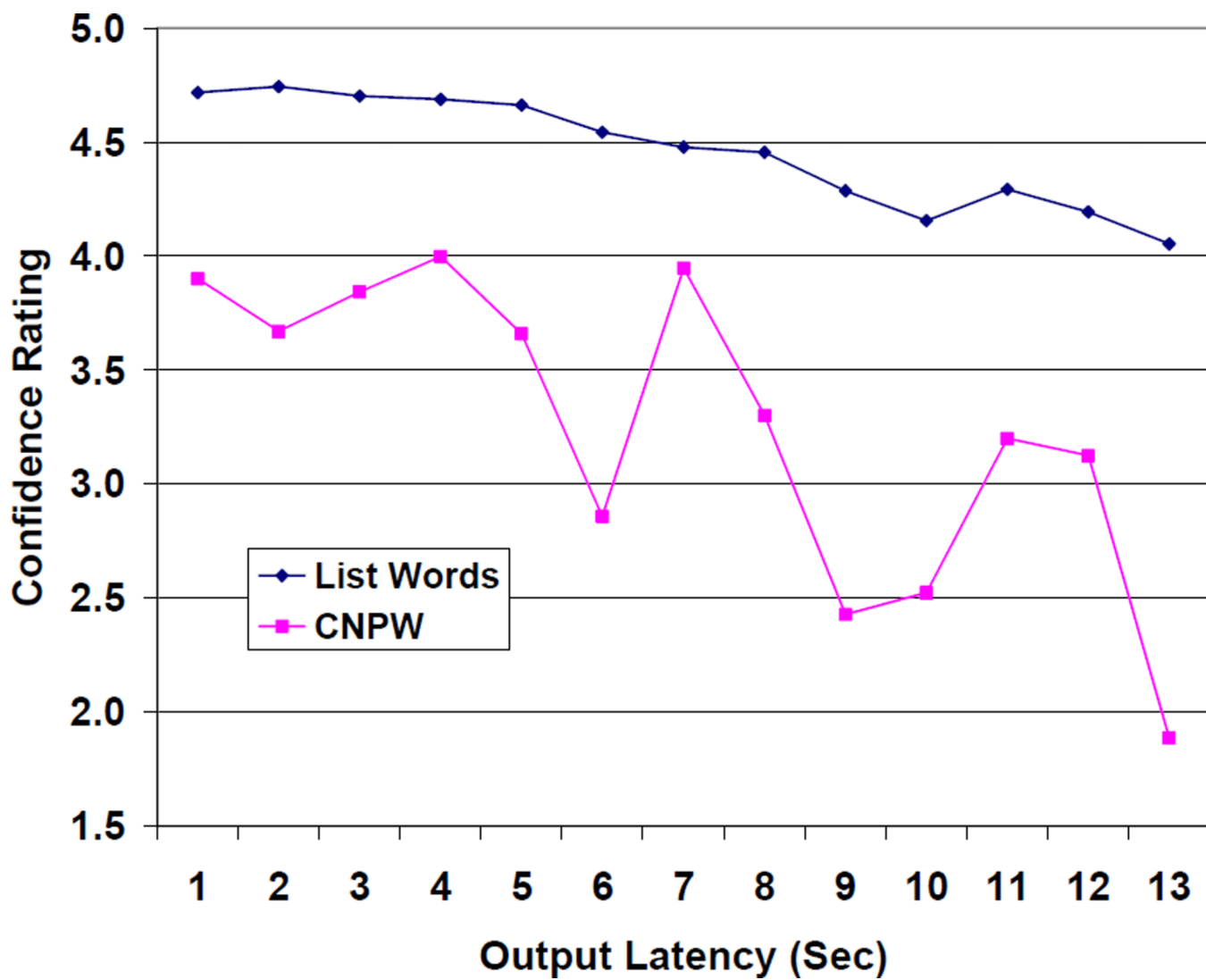


Figure 6.
Mean confidence ratings as a function of word type and recall latency of Experiment 2.

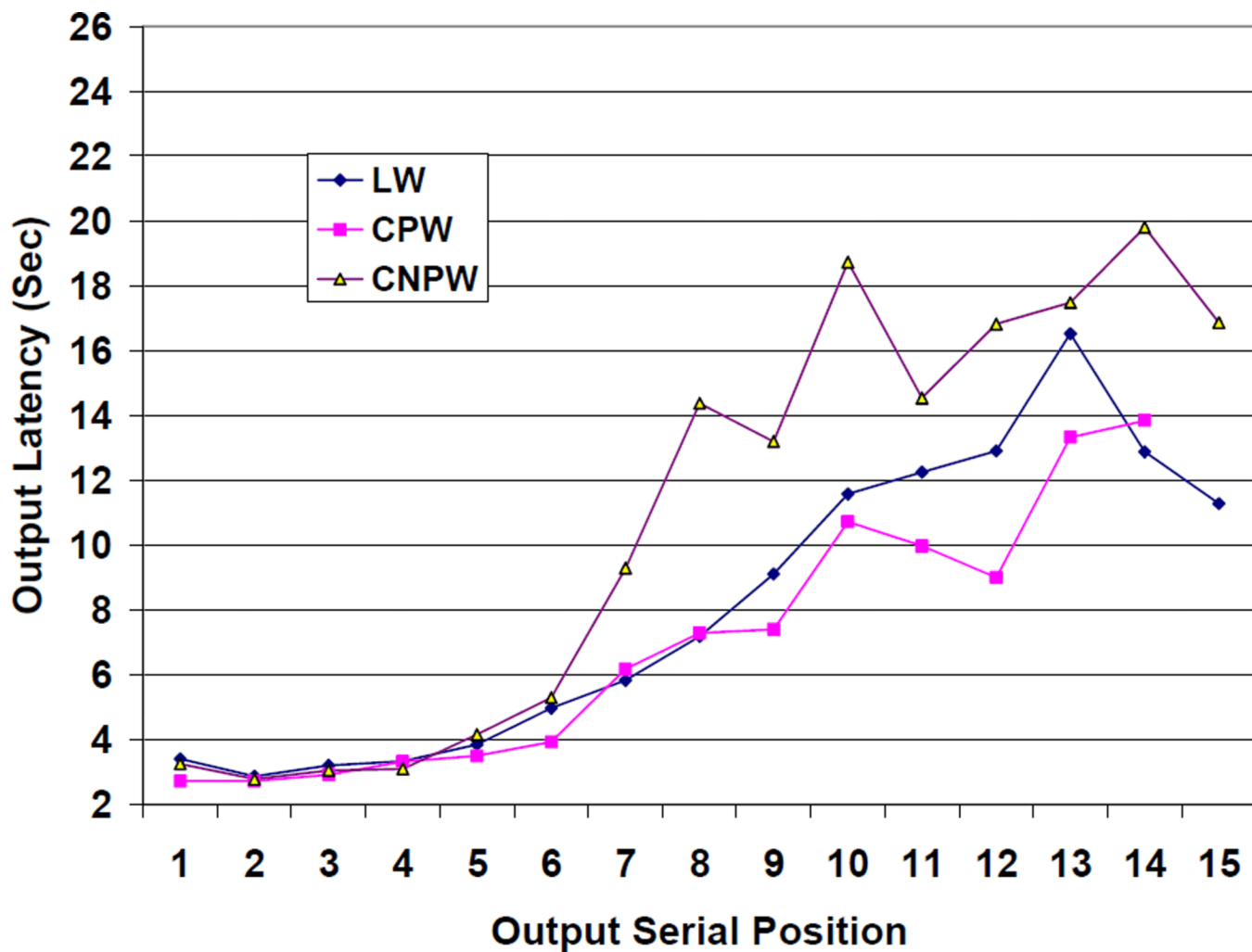


Figure 7.
 Mean recall latencies as a function of word type and output serial position of Experiment 3.

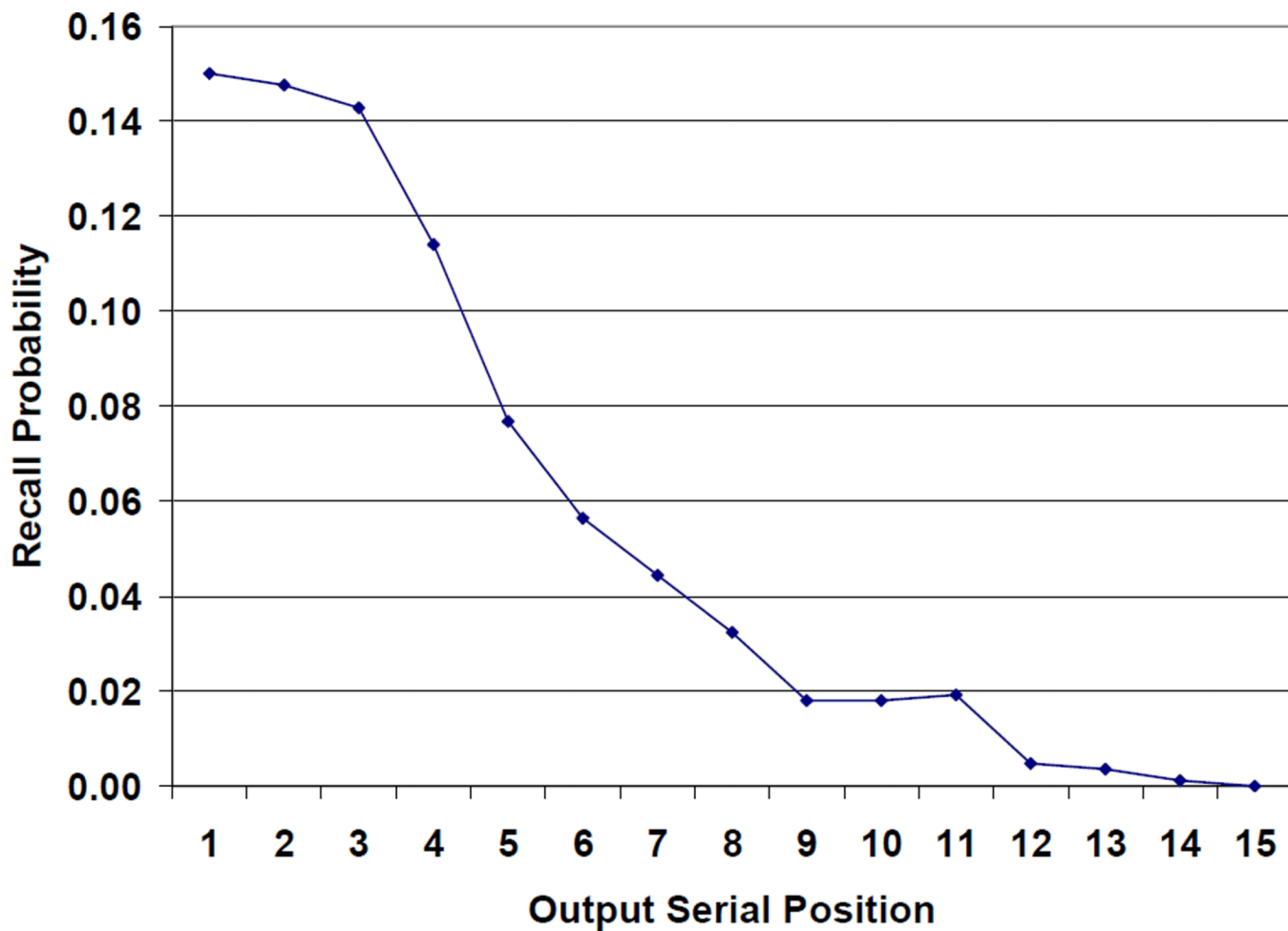


Figure 8. Mean recall probabilities of critical presented words as a function of output serial position of Experiment 3.

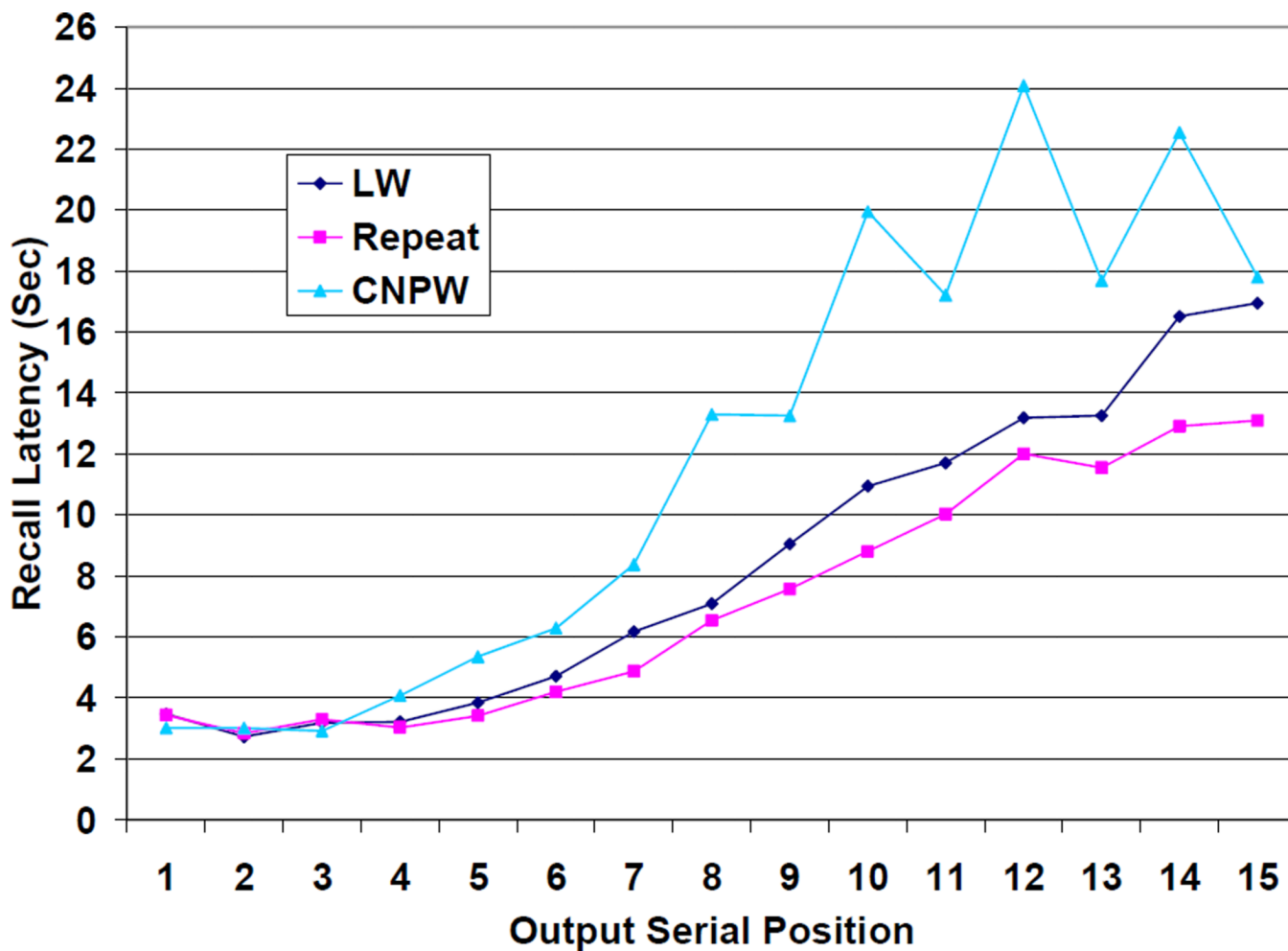


Figure 9.
 Mean recall latencies as a function of word type and output serial position of Experiment 4.

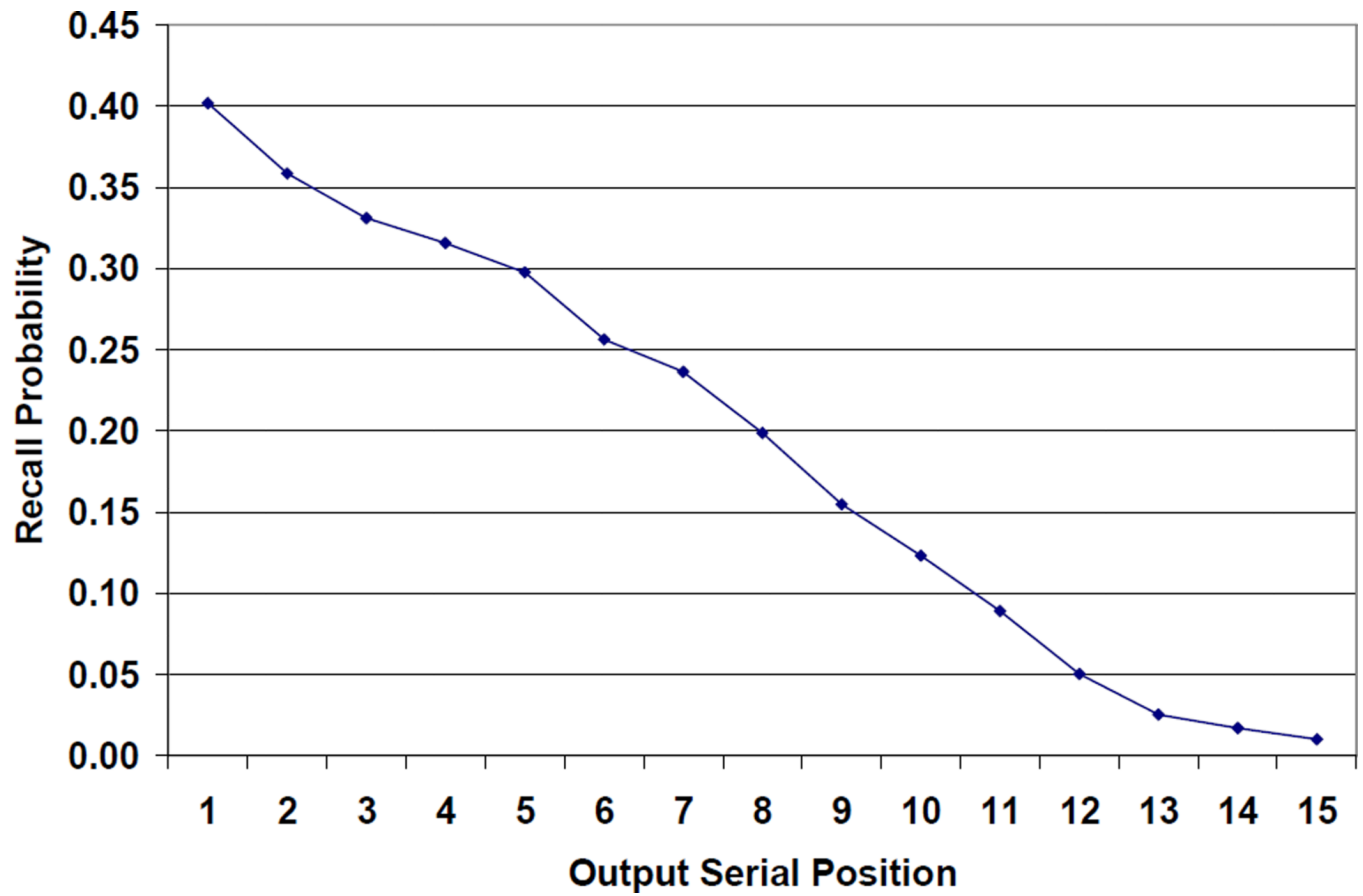


Figure 10. Mean recall probabilities of repeated words as a function of output serial position of Experiment 4.