MEMORY, 2005, 13 (3/4), 267–273

Psychology Press
Taylor & Francis Group

Fill-in and infill errors in order memory

Aimée M. Surprenant, Matthew R. Kelley, Lisa A. Farley, and Ian Neath
Purdue University, IN, USA

Many current models of memory are specified with enough detail to make predictions about patterns of errors in memory tasks. However, there are often not enough empirical data available to test these predictions. We report two experiments that examine the relative frequency of fill-in and infill errors. In immediate serial recall tasks, subjects sometimes incorrectly recall item N too soon, placing it in position $N-1$. The error of interest is which item is recalled after this initial mistake. A fill-in error is the tendency to recall item $N-1$ next, whereas an infill error is the tendency to recall item $N+1$ next. Both experiments reveal more fill-in than infill errors, not only overall but at each possible error location throughout the list. The overall ratio is approximately 2:1. We conclude that none of the currently existing models adequately accounts for fill-in and infill errors.

How we remember a sequence of items has been a focus of research and theory since Ebbinghaus (1885) and current models are highly sophisticated, accounting not only for accuracy data but also for much of the error data. In immediate serial recall and serial reconstruction of order tasks, the most common error data reported are position error gradients, which show the proportion of times that each item was recalled at each possible output serial position. The key characteristics are that when an item is recalled in the incorrect position, it is most likely to be recalled in an immediately adjacent position (e.g., Healy, 1974; Nairne, 1991).

One type of order error has recently received a lot of theoretical attention. Henson (1998; see also Henson, Norris, Page, & Baddeley, 1996; Page & Norris, 1998) has identified fill-in errors as one way of testing a model's ability to account for the complete pattern of error data. A fill-in error is most easily described by example. Suppose correct recall of a list is ABCDE, where A is the first item recalled in the first position, B is the second item recalled in the second position, and so on. Suppose further that the subject recalls B first (an error). A fill-in error occurs when the second response is A; that is, the subject is said to ''fill in'' the item that has been missed. Using x to represent the other responses, a fill-in error is BAxxx when correct recall is ABxxx.

Infill errors also require that B is given (incorrectly) as the first response, but this time the subject responds next with C (the third item) rather than A (the first item). The subject ''infills'' the position that would have been vacant had B been recalled in its correct position. Both types of errors require the same initial mistake (i.e., recalling item B first) and differ only in which item is recalled next, the item that originally preceded B or the item that originally followed B.

Although previous discussions have focused only on positions 1 and 2, fill-in and infill errors can also occur at later list locations. A fill-in error would occur if the subject recalled ACBxx and an infill error would occur if the subject recalled ACDxx. For errors at later list locations, one can distinguish between a strict and lenient criterion: strict scoring requires that all responses preceding the error are correct whereas lenient scoring ignores whether the previous responses are

Correspondence should be addressed to Aimée M. Surprenant, Purdue University, Department of Psychological Sciences, 703 Third Street, West Lafayette, IN 47907-2004, USA. Email: aimee@psych.purdue.edu
We thank R. N. A. Henson for providing detailed error analyses from his dissertation research.
Matthew R. Kelley is now at Lake Forest College, IL, USA.

http://www.tandf.co.uk/journals/pp/09658211.html DOI:10.1080/09658210344000396

correct. ACBxx is a fill-in error by the standards of the strict criterion (because all responses prior to CB, the fill-in error, are correct). DCBxx is not a fill-in error by the strict criterion (because an error occurred prior to the fill-in error) but is a fill-in error by the lenient criterion. In many ways the lenient way of scoring is the more complicated one. For the response beginning DCB there is first the issue of what has happened to A; moreover, for an infill error to occur after the C, the list would have to begin DCD (or perhaps DCE) but this is unlikely because the D is unlikely to be repeated. We report only the data scored according to the strict criterion. Analyses using the lenient criterion did not affect the interpretation of the data.[1]

There are actually very few empirical data available on fill-in and infill errors, especially those involving positions other than the first two. The one detailed analysis available (Henson, 1996, Table 4.1) suggests that fill-in errors are approximately twice as common as infill errors. However, this finding is based on data from several different types of experiments that included variations in stimuli, methodology, etc., and were collapsed over serial position.

Fill-in and infill errors are potentially important in distinguishing among different models of memory. Henson et al. (1996) have argued that the predominance of fill-in over infill errors poses a major problem for chaining models of serial order. If a model represents order through item to item associations, then a general prediction must surely be that infill errors should be much more frequent. When the subject incorrectly recalls B first, the strong item–item association between B and C should make C far more likely as the next response than A.

The examination of fill-in and infill errors has the potential for eliminating other models of serial order. Before a theoretical analysis can be performed, however, more data are needed. In particular, it is necessary to answer several basic questions: Do fill-in errors predominate? Is the relative frequency of fill-in and infill errors constant over list location? Does the relative frequency change as a function of presentation modality? Does the relative frequency change as a function of set size? The two experiments reported below were designed to answer these questions.

[1] The lenient data may be obtained from the first author.

EXPERIMENT 1

The purpose of Experiment 1 was to measure the frequency of fill-in and infill errors at each location in the list where these errors could occur. Subjects saw a list of seven words and were then immediately tested on their memory for order. There were 100 lists per condition to ensure sufficient frequency of errors. Even if fill-in errors occur on only 10% of trials, there would still be 10 observations per subject. Unique items were used on every trial to avoid potential interpretation problems noted by Page and Norris (1998, p. 777): if a small pool of items were used, it could be impossible to tell what sort of error occurred. For example, an error on Trial *N* might be a fill-in error, or it might be a protrusion error where an item from a previous list ‘‘protrudes’’ into the current list. With unique items on every trial, such protrusions are not possible. A serial reconstruction of order task was used to avoid another problem, that of intrusions. This type of test allows responses only of items from the current trial. Serial reconstruction of order yields the same types of results as strict serial recall, including modality and set-size effects (Neath, 1997) and position error gradients (Neath, 1999).

Method

Subjects. A total of 66 Purdue University undergraduates participated in exchange for credit in introductory psychology courses and were arbitrarily assigned to one of two groups. All identified themselves as native speakers of American English.

Design. Presentation modality (visual or spoken aloud) was a between-subjects variable. The only within-subject variable was serial position.

Stimuli. The stimuli were sequences of 7 nouns drawn randomly without replacement from a pool of 716 nouns. These were themselves drawn randomly from Francis and Kucera (1982) with the selection criteria as follows: minimum frequency 10; maximum frequency 40; minimum length 6 letters; and maximum length 8 letters. The mean frequency was 19.71 (standard deviation 8.32), and the mean length was 6.95 (standard deviation 0.80). No word appeared more than once during the entire experiment. Each word was shown for 1 s in the middle of the computer screen in lower-case 18-point Geneva bold, with the next word immediately following.

Procedure. Subjects were tested individually. Each sat in front of an Apple Macintosh LC computer. Subjects were informed that they would see a list of words appear on the screen and would be tested on how accurately they could remember the presentation order. Subjects in the auditory group were asked to read each word out loud as it appeared. This condition is functionally equivalent to an auditory condition (Crowder, 1970). Subjects in the visual group were asked to read each word silently. An experimenter remained in the room to ensure compliance with the instructions. After all seven words had been shown, seven buttons appeared on the computer screen; the buttons were labelled (alphabetically) with the seven possible words. Subjects were asked to click on the buttons to indicate presentation order. This test ensured strict serial recall: Once a response was made, it could not be changed; the first response indicated which word the subject thought had occurred first, the second response indicated which word occurred second, and so on. No feedback of any kind was given. All subjects reported having used a mouse before and no subject reported any problems with the method of recall. There was one practice trial (using colour names as the to-be-remembered items) followed by 100 experimental trials. Several times during the experiment the subject was allowed to take a short, self-paced break.

Results

Accuracy. The proportion of words correctly recalled in position was 0.457 for the auditory condition compared to 0.513 for the visual, $F(1, 64) = 2.72$, $MSE = 0.123$, $p > .10$. There was a main effect of serial position, $F(6, 384) = 157.52$, $MSE = 0.006$, $p < .01$, and an interaction between presentation modality and serial position, $F(6, 384) = 42.48$, $MSE = 0.006$, $p < .01$, with better recall of the final item in the auditory condition than in the visual. Thus, standard serial position and modality effects were observed.

Position error gradients. Table 1 shows the position error gradients for both conditions. The serial position functions can be derived by looking at the outlined cells in the diagonal starting in the upper left. These functions show a normal modality effect and replicate Crowder's (1970) finding that active vocalisation (reading out loud) leads to slightly poorer performance for the first few serial positions. The position error gradients yield no information on fill-in and infill errors. For the visual condition, Table 1 shows that item 3 was more likely to be recalled in Position 2 than was

TABLE 1
Proportion of times each item was recalled in each position in the visual and auditory conditions of Experiment 1

	Postion 1	*Position 2*	*Position 3*	*Position 4*	*Position 5*	*Position 6*	*Position 7*
				Visual			
Item 1	0.768	0.058	0.065	0.037	0.034	0.020	0.013
Item 2	0.063	0.635	0.095	0.085	0.053	0.039	0.029
Item 3	0.045	0.095	0.499	0.124	0.102	0.068	0.063
Item 4	0.030	0.068	0.111	0.456	0.135	0.116	0.078
Item 5	0.034	0.051	0.090	0.118	0.392	0.173	0.127
Item 6	0.037	0.050	0.076	0.095	0.161	0.375	0.182
Item 7	0.030	0.044	0.062	0.077	0.099	0.179	0.469
				Auditory			
Item 1	0.659	0.110	0.106	0.058	0.039	0.021	0.006
Item 2	0.122	0.463	0.171	0.125	0.067	0.037	0.015
Item 3	0.073	0.160	0.342	0.184	0.128	0.086	0.026
Item 4	0.059	0.102	0.166	0.312	0.177	0.131	0.048
Item 5	0.044	0.082	0.104	0.154	0.349	0.184	0.069
Item 6	0.035	0.057	0.072	0.112	0.165	0.412	0.113
Item 7	0.014	0.019	0.029	0.040	0.055	0.097	0.666

Although the ratio of fill-in to in-fill errors was uniformly greater than 1, this cannot be determined from this summary data.

Item 1 (0.095 vs 0.058), which might lead one to think that infill errors predominate. What these data do not show, however, are the conditional frequencies, the proportion of times Items 1 and 3 were recalled in Position 2, given that Item 2 had been recalled in Position 1. Position error gradients obtained from models will have the same problem.

Errors. The experiment yielded an appreciable number of the errors of interest. The total number of fill-in and infill errors in the visual condition was 452 and 228, respectively, a ratio of 1.98:1. In the auditory condition, the corresponding numbers were 616 and 319, a ratio of 1.93:1.

Each error involves two serial positions and there are five possible fill-in and five possible infill errors per list. The term ''location'' is used to refer to pairs of positions; for example, the fill-in error BA and in-fill error BC occur at location 1 which involves positions 1 and 2. Table 2 shows the mean number of fill-in and infill errors observed in each location as a function of presentation modality. These data were analysed with a two presentation modality (read aloud and visual) × two error types (fill-in and infill) × 5 locations analysis of variance.

There were more fill-in (3.24) than infill (1.66) errors, $F(1, 64) = 125.97$, $MSE = 3.26$, $p < .01$. There was a main effect of modality, $F(1, 64) = 29.59$, $MSE = 3.33$, $p < .01$, with more errors in the auditory (2.83) than in the visual (2.06) condition. There was an effect of location, $F(4, 256) = 11.75$, $MSE = 4.68$, $p < .01$, which was due to a change in the number of errors over locations. The only interaction that was reliable was modality by location, $F(4, 256) = 15.63$, $MSE = 4.68$, $p < .01$, due primarily to a larger decrease in the number of number of errors over location in the auditory than in the visual condition. This is to be expected given that the modality effect refers to better accuracy in recalling the final item when the presentation modality is auditory. The only non-reliable interaction with an *F*-value over 1 was the interaction between modality and error type: $F(1, 64) = 2.47$, $MSE = 3.26$, $p > .10$.

Discussion

The results of Experiment 1 demonstrate that (1) there are more fill-in errors than infill errors not only overall, but at each location in the list; (2) this predominance of fill-in errors holds for both visual and auditory presentation; and (3) there are more errors with auditory than with visual presentation. The results of Experiment 1 also demonstrate that one can infer nothing whatsoever about fill-in and infill errors simply by examining position error gradients. Whereas Page and Norris (1998) mentioned errors only at serial positions 1 and 2, and Henson (1996) examined fill-in and infill errors without respect to position, Experiment 1 examined these errors at all possible locations.

EXPERIMENT 2

Experiment 2 was designed as a partial replication. The sole change was to use a closed set of items rather than an unlimited pool to make the

TABLE 2

The absolute number of fill-in and infill errors and the mean number of errors for each subject in Experiment 1 as a function of presentation modality when *strict* scoring is used

	Error type									
	Positions 1–2		*Positions 2–3*		*Positions 3–4*		*Positions 4–5*		*Positions 5–6*	
	BA	*BC*	*CB*	*CD*	*DC*	*DE*	*ED*	*EF*	*FE*	*FG*
	Visual									
Frequency	76	39	100	49	92	46	74	34	110	60
Mean Freq.	2.303	1.182	3.030	1.485	2.788	1.394	2.242	1.030	3.333	1.818
Ratio:	1.949		2.041		2.000		2.176		1.833	
	Auditory									
Frequency	170	100	161	111	135	64	83	34	67	10
Mean Freq.	5.152	3.030	4.879	3.364	4.091	1.939	2.515	1.030	2.030	0.303
Ratio:	1.700		1.450		2.109		2.441		6.700	

Correct recall is ABCDEFG. For Positions 1 and 2, the fill-in error ''BA'' is when Item 2 is recalled first and Item 1 is recalled second; the infill error ''BC'' is when Item 2 is recalled first and Item 3 is recalled second.

experiment more comparable to the many short-term memory tasks in which a fixed set of items are used on every trial. For each subject, seven words were drawn randomly from the main pool used in Experiment 1. The same seven items were used on every trial for a given subject, but each subject had a different random set of items.

Subjects. A total of 66 different Purdue University undergraduates participated in exchange for credit in introductory psychology courses and were arbitrarily assigned to one of two groups. All identified themselves as native speakers of American English.

Design. The design was identical to that in Experiment 1.

Stimuli. For each subject, 7 nouns were drawn randomly from the main pool of 716 nouns. The same seven items were used on every trial for a given subject, with their order on each trial determined randomly.

Procedure. The procedure was identical to that followed in Experiment 1.

Results and discussion

Accuracy. Performance was slightly higher in Experiment 2 than in Experiment 1, a common finding when the set size is reduced (e.g., Neath, 1997). The proportion of items correctly recalled in position was 0.542 in the auditory condition compared to 0.632 in the visual, the same finding as in Experiment 1 but this time a reliable difference, $F(1, 64) = 5.75$, $MSE = 0.162$, $p < .01$. There was the usual effect of serial position, $F(6, 384) = 128.35$, $MSE = 0.008$, $p < .01$ and an interaction between modality and position, $F(6, 384) = 23.12$, $MSE = 0.008$, $p < .01$. Recall of the last item was better following auditory than visual presentation, the standard modality effect. The position error gradients are not reported, but do not differ substantively from those observed in Experiment 1.

Errors. As in Experiment 1, there were an appreciable number of errors. The total number of fill-in and infill errors in the visual condition was 473 and 170, respectively, a ratio of 2.78:1. For the auditory condition, the numbers were 607 and 258, respectively, a ratio of 2.35:1.

The pattern of fill-in and infill errors was broadly similar to the pattern seen in Experiment 1 (compare Table 3 with Table 2). There were again more fill-in than infill errors—3.27 vs 1.30; $F(1, 64) = 130.58$, $MSE = 4.93$, $p < .01$. There was again a main effect of modality, $F(1, 64) = 11.60$, $MSE = 6.44$, $p < .01$, with more errors in the auditory than in the visual condition (2.62 vs 1.95). As in Experiment 1, there was a reliable effect of location, $F(4, 256) = 5.69$, $MSE = 7.05$, $p < .01$, and an interaction between modality and location, $F(4, 256) = 15.63$, $MSE = 4.68$, $p < .01$. The location by error type interaction was also reliable, $F(4, 256) = 7.83$, $MSE = 4.29$, $p < .01$, again due primarily to the large increase in fill-in errors at locations 2 and 4. The modality by error type

TABLE 3
The absolute number of fill-in and infill errors and the mean number of errors for each subject in Experiment 2 as a function of presentation modality when *strict* scoring is used

	Error type									
	Positions 1–2		*Positions 2–3*		*Positions 3–4*		*Positions 4–5*		*Positions 5–6*	
	BA	*BC*	*CB*	*CD*	*DC*	*DE*	*ED*	*EF*	*FE*	*FG*
	Visual									
Frequency	44	29	92	30	71	23	64	32	202	56
Mean Freq.	1.333	0.879	2.788	0.909	2.152	0.697	1.939	0.970	6.121	1.697
Ratio:	1.517		3.067		3.087		2.000		3.607	
	Auditory									
Frequency	117	55	163	82	130	53	86	47	111	21
Mean Freq.	3.545	1.667	4.939	2.485	3.939	1.606	2.606	1.424	3.364	0.636
Ratio:	2.127		1.988		2.453		1.830		5.286	

Correct recall is ABCDEFG. For Positions 1 and 2, the fill-in error "BA" is when Item 2 is recalled first and Item 1 is recalled second; the infill error "BC" is when Item 2 is recalled first and Item 3 is recalled second.

interaction was not reliable, $F(1,64) < 1$, but the three-way interaction was reliable, $F(4,256) = 2.73$, $MSE = 4.29$, $p < .05$.

Comparison of Experiments 1 and 2. Because of the design of Experiments 1 and 2, an analysis of variance on the error data could be performed with stimulus set size as a factor. There was no main effect of set size, $F(1,1288) = 1.78$, $MSE = 4.88$, $p > .10$, with approximately the same number of errors in the large (2.45) and small (2.29) set sizes. Set size did interact with location, $F(4,512) = 7.16$, $MSE = 5.86$, $p < .01$, and there was also a three-way interaction with set size, location, and error type $F(4,512) = 4.32$, $MSE = 3.56$, $p < .01$. Both of these had to do with the increase of errors at locations 2 and 3, but also reflected more errors at the early locations with a large set size but more errors at later positions with a small set size. No other interactions involving set size approached significance.

GENERAL DISCUSSION

The results of both Experiments 1 and 2 are similar and straightforward: There are more fill-in than infill errors at every location in the list where such errors are possible regardless of the location within the list, presentation modality, or set size. The major effect of presentation modality is that there are more errors of both kinds with auditory presentation, but fill-in errors still predominate over infill errors even at the end of the list. Set size manipulations do not affect the overall relative frequency of the errors much but do interact with location. One can think of this as due to proactive interference: In the small set size, proactive interference can play a larger role, especially in those parts of the list where performance is not especially good.

The absolute number of errors is of the same general order of magnitude as that reported by Henson (1996). Henson found that of 207 possible instances where Item N+1 was recalled in position N, the proportion of fill-in errors was 0.53 and the proportion of infill errors was 0.21, a ratio of approximately 2.52. In comparison, the ratio of fill-in to infill errors varied between a low of 1.45 and a high of 6.70, with an overall ratio of 1.93 to 2.78. Perhaps the most salient differences between Henson's analysis and the one reported here are that the current studies had a fixed list length, did not manipulate any variables other than serial position within subjects, looked at each error location separately, and had a much larger number of trials per subject.

The current data go beyond Henson's by demonstrating that (1) there are more fill-in than infill errors at each location throughout the list; (2) this pattern of results is the same when presentation modality is visual as when it is auditory, although more errors of both types occur with auditory presentation; and (3) the pattern is the same when unique items are used on every trial as when the same set of seven items are repeated on every trial.

Most current models of serial order correctly predict accuracy data and most correctly predict position error gradients. These results have been well established, and so models could be built to accommodate them. Two types of errors not previously well known are fill-in and infill errors. The two experiments reported here consistently found more fill-in than infill errors both overall and at each location throughout the list, and also more errors for auditory than visual presentation. There was no evidence either of a reversal or of a decrease in the ratio over lists. Neath and Surprenant (2003, p. 307) summarise 10 current models of serial order and the mechanisms/processes that govern order information. They conclude that none of these models could account for the basic pattern of results reported in this paper, which suggests that the mechanisms used in the models are either not correct or not complete. It is certainly possible to alter the models so that they do make more accurate predictions; however, we leave that as an exercise for the models' authors.

REFERENCES

Crowder, R. G. (1970). The role of one's own voice in immediate memory. *Cognitive Psychology*, *1*, 157–178.

Ebbinghaus, H. (1885). *Uber das Gedachtnis: Untersuchungen zur experimentellen Psychologie.* Leipzig: Duncker and Humboldt. [Reprinted as H. E. Ebbinghaus (1964). *Memory: A contribution to experimental psychology.* H. A. Ruger, (Trans.). New York: Dover.]

Francis, W. N., & Kucera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston: Houghton Mifflin.

Healy, A. F. (1974). Separating item from order information in short-term memory. *Journal of Verbal Learning & Verbal Behavior*, *13*, 644–655.

Henson, R. N. A. (1996). *Short-term memory for serial order. Unpublished doctoral thesis*. University of Cambridge, UK.

Henson, R. N. A. (1998). Short-term memory for serial order: The start-end model. *Cognitive Psychology*, *36*, 73–137.

Henson, R. N. A., Norris, D. G., Page, M. P. A., & Baddeley, A. D. (1996). Unchained memory: Error patterns rule out chaining models of immediate serial recall. *Quarterly Journal of Experimental Psychology*, *49*A, 80–115.

Nairne, J. S. (1991). Positional uncertainty in long-term memory. *Memory & Cognition*, *19*, 332–340.

Neath, I. (1997). Modality, concreteness, and set-size effects in a free reconstruction of order task. *Memory & Cognition*, *25*, 256–263.

Neath, I. (1999). Modelling the disruptive effects of irrelevant speech on order information. *International Journal of Psychology*, *34*, 410–418.

Neath, I., & Surprenant, A. M. (2003). *Human memory: An introduction to research, data, and theory* (2nd ed.). Belmont, CA: Wadsworth.

Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review*, *105*, 761–781.