Psychological Research (2003) 67: 280–290
DOI 10.1007/s00426-002-0118-1

ORIGINAL ARTICLE

Johannes Engelkamp · Petra Jahn · Kerstin H. Seiler

The item-order hypothesis reconsidered: The role of order information in free recall

Received: 4 April 2002 / Accepted: 23 July 2002 / Published online: 25 February 2003

Abstract According to the item-order approach of free recall, in pure short lists the free recall of unrelated items is organized according to their order of presentation in the study list. The approach was applied in the present study to experimenter-performed tasks (EPTs) and subject-performed tasks (SPTs). It claims that EPTs provide better serial order information than SPTs. Consequently, free recall of EPTs should be more organized along the presentation order of the items than the free recall of SPTs. In three experiments, some specific aspects of this approach were studied. Firstly, it was demonstrated that serial retrieval is not strongly used spontaneously and that its use is overestimated in the literature because it is usually evoked by an order reconstruction test which follows free recall testing. Secondly, a serial retrieval strategy in free recall can be encouraged by explicit instructions. Finally, the present experiments showed that a serial output strategy alone does not allow one to predict performance in free recall. The implications of these findings for the item-order approach will be discussed.

Episodic memory as a subsystem of memory is designed specifically for storing temporally and spatially bound events and relations between such events (e.g. Tulving, 1985). In daily life, it occurs frequently that we try to remember episodic events. If we remember episodic events, we consciously reinstate them and recognize the retrieved memories as belonging to the past (Wheeler, Stuss & Tulving, 1997): Such reinstatements are based on retrieval processes which follow different strategies. Three important strategies use temporal, spatial or conceptual relations among the events in order to retrieve them. In the present paper, we will deal with the item-order hypothesis which claims that short lists of unrelated events tend to be retrieved along their temporal serial order and that the events are recognized on the basis of their item information in free recall (e.g. Nairne, Riegler & Serra, 1991). An important and interesting aspect of this hypothesis is that it claims that encoding conditions might differ in the amount of item and serial order information which they provide. Some conditions provide better order information than others, and the situation for item information is reversed. Another important aspect is that it is claimed that order encoding depends on the type of design. The condition providing better order information does so only if encoding condition is treated as a between-subjects variable. If the encoding conditions are varied within subjects, there are no longer differences in order encoding. Item information, on the other hand, is claimed to be by and large independent of the type of design (e.g. Burns, Curti & Lavin, 1993; Nairne, Riegler & Serra, 1991; Serra & Nairne, 1993. See, however, DeLosh & McDaniel, 1996, for slightly different assumptions). Finally, it is consistently claimed that free recall, at least in pure lists, is often based on a serial output strategy and that order information like other types of relational information contributes to performance in free recall (cf. Hunt & Einstein, 1981; Hunt & McDaniel, 1993).

With the assumptions of the item-order hypothesis, it can be explained why a number of contrasted encoding conditions such as read vs. generate (Burns et al., 1993; Serra & Nairne, 1993), intact perception of items vs. perceptual interference (e.g. Mulligan, 1999), common vs. bizarre items (McDaniel et al., 1995), high vs. low word frequency (DeLosh & McDaniel, 1996), or encoding of actions by watching the experimenter performing them vs. by performing the actions oneself (Engelkamp & Dehn, 2000) all show a common pattern of findings in free recall. There is a recall advantage of that situation which provides better item information in the within-subjects design. However, this recall advantage disappears (or reverses) in a between-subjects design. The basic explanation of this design-dependent recall effect is that a serial output strategy is used. In a

J. Engelkamp (✉) · P. Jahn · K.H. Seiler
FR Psychologie, Saarland University,
D-66123 Saarbruecken, Germany
E-mail: engelkamp@rz.uni-sb.de

within-subjects design in which order encoding does not differ, there is an advantage of the condition which provides better item-specific information. In a between-subjects design, the condition with better serial order encoding benefits relatively more from the serial output strategy than the other condition, so that the result in recall is often a balance between both encoding conditions.

The item-order hypothesis has been submitted to some constraints in the course of time. It has been shown that it is confined to short lists of 6 – 8 items, at least if action phrases serve as items (Engelkamp & Dehn, 2000). Serial order information does not play a decisive role in long lists of about 15 or more action phrases (Engelkamp & Dehn, 2000). Furthermore, arguments and evidence have been brought forward that it is unlikely that free recall is based on an order-based output strategy in a within-mixed list design (Engelkamp & Dehn, 2000; McDaniel, DeLosh & Merritt, 2000). McDaniel et al. assume, for instance, that participants focus on differences between the item subsets and use a distinctiveness-based retrieval strategy instead of an order-based one in a mixed-list design.

We will show here that even with regard to a between-subjects pure list design, the evidence of an order-based retrieval strategy in free recall is less impressive than generally assumed. In detail, we will demonstrate that (a) the spontaneous use of order information is very small, (b) that the order information as measured in an order reconstruction task is induced by using this task, and (c) that order information, although it may contribute to recall, will do less so than might be expected according to the item-order hypothesis. We will demonstrate these constraints with the example of action phrases which are either encoded in subject-performed tasks (SPTs) or in experimenter-performed tasks (EPTs). In SPTs, the participants mimic performing the denoted actions (i.e. without using real objects), and in EPTs they watch how the experimenter mimics performing the actions (without real objects).

Engelkamp and Dehn (2000) have applied the item-order hypothesis to memory for action phrases assuming that watching the experimenter performing the task in EPTs provides better serial order information than self-performing the task in SPTs and that SPTs provide better item information than EPTs. This assumption is based on the argument that self-performing actions upon their verbal descriptions force the participants to concentrate on the individual items. That is, SPTs focus the encoding processes on item information and at the same time distract from encoding the inter-item relations, such as their serial order. The situation is different if a series of actions is observed as in EPTs. The task of perceiving a series of actions is less specific and leaves it more up to the participant how the items are encoded. As a consequence, in EPTs participants pay relatively more attention to the serial order of the items and relatively less attention to the individual items than in SPTs (cf. Engelkamp, 1995). Hence, order encoding is assumed to be better in EPTs than in SPTs, and the reverse holds true for item information if only one task is required for a list. If some items of the same list are to be observed and others to be performed as in mixed lists, the situation is different. In the latter case, attention may be primarily drawn to the differences between perceived and performed items, and particularly the order encoding of perceived items will suffer (e.g. McDaniel et al., 2000).

As postulated by the item-order hypotheses, interactions of type of task and type of design were observed in free recall (FR) and in order reconstruction (OR). In order reconstruction, the participants had to reorder the items which were presented to them randomly to the positions which they had during the study of the list. For order reconstruction, there was an EPT advantage over SPT in the between-subjects design and no difference in the within-subjects design. For free recall, there was an SPT advantage over EPT in the within-subjects design and no difference in the between-subjects design.

Hence a central point of the item-order hypothesis for the present study that focuses on the between-subjects design is that EPTs provide better order information than SPTs and that free recall is based on this order information. However, these assumptions were not adequately tested in most studies for the following reason: Order reconstruction was used as a test of order information (e.g. Burns et al., 1993; DeLosh & McDaniel, 1996; Engelkamp & Dehn, 2000; Serra & Nairne, 1993). This test followed free recall, and both tests were applied repeatedly because participants learned a series of about 6 to 8 short lists with free recall and order reconstruction being tested successively after each list. According to Greene, Thapar and Westerman (1998), this procedure might be critical for the use of order information. They argue that the exact positions of test items required in the order reconstruction task are probably reconstructed by means of serial order recall. The presence of all items in order reconstruction may elicit increased dependence on inter-item associations. If Greene et al. (1998) are right, then it is significant whether order reconstruction follows free recall, a situation typically realized in item-order studies (e.g. Engelkamp & Dehn, 2000; DeLosh & McDaniel, 1996). Since participants know, after having been tested for the practice list, that they have to expect an order reconstruction test, it is likely that they develop a serial order encoding and a serial order retrieval strategy in order to cope with the situation. Using order reconstruction at test would then encourage participants to rely on inter-item associations (cf. also Li & Lewandowski, 1993, 1995).

If this assumption holds true, there are important implications for the item-order studies. One is that the use of order reconstruction tests does not allow for assessing how much serial order encoding takes place spontaneously and is used in free recall retrieval. To explore the spontaneous use of serial information during encoding and during retrieval in free recall, only free recall must be tested. The common testing of free recall and order reconstruction means that serial information encoding and retrieval are probably overestimated in most experiments testing the item-order hypothesis. We assume that encoding and use of order information in free recall is negligible if no order reconstruction tests are used and that order information is induced if order reconstruction follows free recall.

Another implication is that if order reconstruction testing can induce better encoding of serial information and its use at test, the same effect should be achieved if participants are requested to focus on encoding of serial information and its use at test (cf. McDaniel, DeLosh and Merritt, 2000).

A third implication is that in the relevant experiments, one condition – EPTs in our case – should nevertheless provide better serial information than the other condition – SPTs in our case.

A fourth implication is that the finding concerning equal free recall for EPTs and SPTs in standard free recall tests of short lists (e.g. Cohen, 1981, 1983; Engelkamp & Zimmer, 1997) without being followed by an order reconstruction and without repeated testing cannot easily be explained by order-based retrieval because there is little order information available. According to the item-order hypothesis, free recall depends to a large extent on the availability of serial order information, and it should increase if order information increases. Although other retrieval strategies are not ignored – for instance, if categorically structured lists are used (e.g. Nairne et al., 1991), alternative retrieval strategies seem to be more flexibly used than the item-order hypothesis suggests. Such a view is held by Greene (1992) who suggests that for the explanation of the so-called "long-term recency effect" in free recall, we have to keep in mind that the discrimination hypothesis (e.g. Bjork & Whitten, 1974; Glenberg et al., 1980) assumes "that at least two retrieval processes are employed in recall" (p. 63). Similar considerations about retrieval processes in free recall are formulated by Zimmer, Helstrup and Engelkamp (2000). They also distinguish two retrieval strategies which are flexibly applied depending on the differential availability of item and relational information, whereby relational information is defined more broadly as any kind of inter-item association. Although these claims were formulated in the context of recall of long lists, there is no reason not to expect more flexible retrieval processes for the recall of short lists, also. A consequence of such flexible retrieval strategies would be that no clear predictions concerning free recall are possible on the basis of order information.

In order to test these implications, we have to abstain from using order reconstruction to test how much order information is available and used in free recall. What alternatives are there?

In any case, indicators of serial output strategies in free recall should be directly derived from free recall performance, and the scores should assess how good the item order from the study list is preserved during recall. The score which was most often used in studies on the item-order hypothesis (besides order reconstruction) is the input-output correspondence according to Asch and Ebenholz (1962). It measures the number of times adjacently recalled items preserve the relative order of input with chance performance at .50. This index also takes into account order relations of items which were not immediate neighbors during study. A problem of this index is that it does not systematically co-vary with the findings of order reconstruction and it does not show the effects which are expected due to the item-order hypothesis. Sometimes, the input-output correspondence is close to chance level, particularly for that condition which should provide less order information than the other condition (e.g. Burns et al., 1993; McDaniel et al., 1995; McDaniel, DeLosh & Merritt, 2000). In other studies, it is clearly greater than chance if it was expected according to the item-order hypothesis (e.g. DeLosh & McDaniel, 1996). A supplementary inspection of our own study (Engelkamp & Dehn, 2000) showed that input-output correspondence was at chance and did not differ between EPTs and SPTs. Hence, the input-output correspondence does not seem to be capable of revealing differences in order information.

Moreover, it might be possible that order-based retrieval relies on associations between direct item neighbors rather than on indirect associations over greater distances between items. If this was the case, it would explain why input-output correspondence is not particularly appropriate to measure the use of order information in free recall. For this reason, an index which is based on direct item neighborhood might be more appropriate. We decided to use the uni-directional ARC' score (Pellegrino, 1971) as a score which considers how often items which are neighbored at recall were also neighbors in the same order at study. This score is sensitive to the degree to which serial neighborhood relations at study are preserved in recall. According to the item-order hypothesis, we expected, for instance, that the ARC' score would be better in EPTs than in SPTs.[1]

In addition to this score, we also considered directly how much of the presentation order during study was preserved in free recall. For this purpose, we analyzed free recall performance as serial recall performance (SR). In this case, we only considered recalled items correct when they were recalled in strictly the same order as they were presented during study. This score is stricter than

[1]A number of different measures have been proposed which can indicate the extent to which the inspection of the output order by pairs matches with the input order, e.g. Pair Frequency Score (PF, Anderson & Watts, 1969) and Intertrial Repetition Measure (ITR, Bousfield & Bousfield, 1966). Since PF and ITR have been considered by many to be two fundamental output adjacency measures (Sternberg & Tulving, 1977), these scores as well as the uni- and bidirectional version of the ARC' scores were calculated for the present data. Even if it is possible to analyze higher order units with the ARC' scores, we only considered item pairs. All scores showed similar effects in all experiments. ARC', which can be viewed as an advancement of ITR, has the advantage in that it states a fixed chance level which is 0 and a maximum of perfect organization which is 1.0. Therefore, in the present paper we decided to present only unidirectional ARC' scores.

the ARC' score because it takes the value of zero as soon as a person does not start the recall with the first list item. However, it will be interesting to explore whether even this score shows higher values in EPTs than in SPTs and to test how strong the tendency is to begin the recall with the first word of a list. A serial order output strategy would predict that a high proportion of participants will do so. Moreover, this tendency should be higher in EPTs than in SPTs. For control purposes and in order to evaluate the input-output correspondence score and to compare it with the ARC' score, we also computed the serial recall score. It mainly serves to get additional information on the suitability of this score in the context of studying whether free recall of short lists is order-based.

Taken together, the above considerations led us, first of all, to compare performance in the direct serial information measures as described above for EPTs and SPTs in a basic condition in which only free recall is tested. Under these conditions, we expect low scores for order information. However, as far as order information is encoded and used, it should still be better in EPTs than in SPTs.

Secondly, the above considerations led us to push serial information encoding and its use in free recall by instructing participants to focus on serial information and to use it during recall. This instruction should enhance serial order information, and the corresponding scores should be enhanced compared to the basic condition. Again, the indices of order information should be better for EPTs than for SPTs.

Thirdly, we introduced a condition in which we tested order reconstruction after free recall. This condition should – as should the serial instruction – enhance the serial order information scores as compared to the basic condition. Again, the EPT advantage should be preserved. Such an advantage should also be observed for order reconstruction and positional recall.

Finally, we had no firm expectations as to free recall performances. However, assuming that different retrieval strategies might be used and the retrieval strategies flexibly adapt to the information available, we do not expect that the free recall level is as strongly dependent on the degree of order encoding as assumed by the item-order hypothesis.

Overview of the experiments

To test the above assumptions, we conducted three experiments in which we manipulated the encoding condition (EPT vs. SPT) in a between-subjects design. In Experiment 1 we used a standard instruction informing the participants that they were participating in a memory experiment without specifying the type of memory test. In Experiment 2 we used a specific instruction which requested the participants to pay particular attention to the serial order of the items and to use the serial order actively during the recall tests. In Experiment 3 we again used a standard instruction, only informing the participants that they were taking part in a memory experiment. This time, however, each free recall was followed by a positional recall and an order reconstruction test.

Experiment 1

Experiment 1 served to test whether serial recall and ARC' scores were better in EPTs than in SPTs. For this purpose, EPTs and SPTs were presented in a between-subjects design. Performance was tested in a free recall test. From the free recall performance, serial recall performance as well as the ARC' scores and input-output correspondence were calculated.

Method

Participants Seventy-two students at Saarland University participated in the experiment. They were paid for their participation. Half of them were assigned to EPT-, half to SPT-encoding conditions.

Materials Sixty-four action phrases were selected from a larger pool of action phrases. From these items, eight lists of eight phrases were constructed. For every list, the phrases were selected in such a way that they were phonologically and semantically unrelated. In addition, there was one list of another eight items which served as a practice list. All items were presented to the participants visually on a computer screen.

To balance possible confounding of study positions within a single list and item material, items were assigned to study positions in such a way that for all participants each phrase appeared equally often at one of the eight study positions. The order of lists (i.e. the position of a list within the set of all lists) was the same for all participants.

Procedure Participants were tested individually in single experimental sessions. The experiment consisted of nine study-recall sequences, of which the first sequence only served the purpose of training.

More specifically, the structure of the experiment was as follows: At the beginning of the experiment, participants were given a booklet containing the general instructions, nine sheets of paper for the free recall tests, and nine sheets of paper for a task that had to be performed during the retention interval, the "d2" (Brickenkamp, 1962). The booklet informed the participants that they were participating in an experiment on memory for actions. They were told that they would be presented lists of simple action phrases (e.g. "break the stick") on the computer screen, and that they should try to memorize the items for a later retention test. In addition, participants in the EPT condition were instructed to watch the experimenter performing all phrases symbolically. Participants in the SPT condition were instructed to enact all phrases symbolically.

Immediately after the instruction was given, participants were presented the first (practice) list of phrases. The phrases were presented for 5,000 ms each on the computer screen together with the instruction to "watch" or "enact" the phrases. The interstimulus interval was set at 1,000 ms. After the presentation of a study list of eight items, participants had to work through the d2 test (Brickenkamp, 1962), a test for measuring attentional capacities. This test requires the participant to detect as many *d* s with two dashes as possible among a row of *d* s and *p* s with one to three dashes. The time for performing the d2 was determined to be 30 s. After the 30 s had elapsed, participants were instructed to write down all the phrases they could remember in any order. For this free recall test they were given 60 s. The above sequence

– presentation of study list, d2, and free recall – was then carried out eight times.

Results and discussion

Data analyses in all experiments were carried out separately for the different memory scores. The first score was free recall (FR). It represents the number of correctly recalled items independent of whether the items were recalled in the correct position or not. A phrase produced was classified as ‘‘correct’’ whenever both the verb and the object corresponded to the ones used in the study list. A further score was serial recall (SR) which refers to the number of items that were recalled in the correct order starting from the first item of the study list. Furthermore, serial information in free recall was assessed by uni-directional ARC’ scores, and for control purposes input-output correspondence (IO) was calculated. An overview of the results is given in Table 1.

The probabilities of free recall of EPTs (.61) and SPTs (.58) did not differ significantly ($t < 1$). The serial recall performance was very low. However, the probabilities of serial recall differed significantly between EPTs and SPTs ($t(70) = -2.28$, $p < .05$). It was .08 for EPTs and .04 for SPTs.

The ARC’ scores measuring to what degree the neighborhood associations from the study list were preserved in recall were slightly higher in EPTs than in SPTs (.09 vs. .03, $t(70) = -1.92$, $p = .06$). The input-output correspondence scores did not differ between EPTs (.48) and SPTs (.43), and both scores were at chance level.

As predicted and in line with the item-order hypothesis, serial recall which directly reflects the use of serial information in recall was better in EPTs than in SPTs. At the same time, the low probabilities for serial recall make it obvious that only a very small proportion of free recall performance is strictly order-based. Out of approximately 5 items, which were on average free recalled, less than one was reproduced in the series corresponding to the presentation order. If one considers the proportion of the participants who started their recall on average, i.e. across the eight lists, with the first item of the list, it turns out that 31% of the participants in EPTs and 22% of those in SPTs did so. Both findings, the relatively low number of participants starting their recall at all with the first list item and the very low level of serial recall, at least attenuates the weight of the serial retrieval assumption of the item-order hypothesis (cf. McDaniel et al., 2000) and indicates that very little order information is encoded and used spontaneously in free recall (i.e. if no order reconstruction test is applied after free recall).

However, although the items were not substantially recalled strictly in their presentation order, neighboring items at study also appeared as neighbors at test, as the ARC’ scores indicated. Although these scores were also low, they were higher in EPTs than in SPTs. On the other hand, input-output correspondence scores were at chance level and did not differ between EPTs and SPTs. The different findings of ARC’ scores and input-output correspondence scores indicate that EPTs allow for better pair associations and their use during recall than SPTs, but that greater distance associations (e.g. between items which are separated by other items) might play a minor role as indicated by the chance level performance in input-output correspondence scores.

In order to obtain some idea about how much serial information contributed to the performance in free recall, we calculated the Pearson correlations between free recall and the various measures of serial information in free recall. The results are presented in Table 2.

As the results show, there was a weak but positive correlation between free recall and serial recall and between free recall and ARC’ for EPTs, but not between free recall and input-output correspondence, and there were no significant correlations in the same analyses for SPTs. However, these findings must be interpreted cautiously because the data range of serial recall scores and ARC’ scores are very restricted. Nevertheless, it is interesting to observe that serial information and free recall seem to be stronger correlated in EPTs than in

Table 1 Mean performance and SD in free recall (FR), serial recall (SR), input-output correspondence (IO) and adjusted ratio of clustering (ARC’) as function of type of encoding (EPT, SPT) in Experiments 1, 2 and 3

		EPT		SPT	
		M	SD	M	SD
FR	Exp. 1	.61	.10	.58	.13
	Exp. 2	.60	.15	.57	.08
	Exp. 3	.65	.13	.62	.13
SR	Exp. 1	.08	.08	.04	.04
	Exp. 2	.24	.15	.16	.06
	Exp. 3	.22	.17	.14	.13
ARC’	Exp. 1	.09	.16	.03	.09
	Exp. 2	.39	.23	.26	.12
	Exp. 3	.35	.22	.19	.21
IO	Exp. 1	.48	.14	.43	.15
	Exp. 2	.91	.11	.84	.09
	Exp. 3	.73	.18	.65	.19

EPT = experimenter-performed task; SPT = subject-performed task

Table 2 Pearson correlation coefficients between free recall (FR) and serial recall (SR), between free recall and input-output correspondence (IO) and between free recall and ARC’ Scores. The correlations were calculated separately for EPTs and SPTs in Experiment 1. For each condition, N was 36

	EPT		SPT	
FR-SR	r = .32	p = .06	r = .22	p = .21
FR-ARC’	r = .32	p = .06	r = .17	p = .33
FR-IO	r = .14	p = .41	r = .29	p = .08

EPT = experimenter-performed task; SPT = subject-performed task

SPTs. This finding corresponds nicely to the findings of the direct comparisons of SR and ARC' scores between EPTs and SPTs.

Moreover, these low correlations also indicate that something other than associative information is used in free recall and that this other type of information can compensate for the smaller amount of serial information in SPTs compared to EPTs. The implication of the free recall findings for the item-order hypothesis will be dealt with in more detail in the general discussion.

Altogether, Experiment 1 showed: little order information is spontaneously encoded and used, that is, if no order reconstruction test is involved. However, EPTs still allow better encoding and use of order information than SPTs on this generally low level. Input-output correspondence scores prove to be – as expected – insensitive to measure the variation of order information. Free recall seems to be largely independent of order information. It did not differ between EPTs and SPTs.

Experiment 2

Experiment 2 served one main purpose, namely to push encoding and use of serial information in recall. For this purpose, participants were explicitly instructed to pay attention to the serial order of items during study and to use serial order during test when remembering the items. We expected that this instruction would increase the serial recall scores as well as the organizational scores in comparison to Experiment 1.

Method

Except for the afore mentioned instruction, Experiment 2 was identical to Experiment 1. The same eight experimental lists were presented and recalled as in Experiment 1.

Participants Two groups of eighteen participants took part in this experiment. One group learned EPTs, the other group SPTs. The participants were students at Saarland University, and they were paid for their participation.

Procedure Procedural details were identical to Experiment 1 except that the participants of this experiment were told that while studying the lists they should pay attention to the presentation order of the items and that they should use this information to memorize the items during study and to retrieve the items during test.

Results and discussion

The same scores as in Experiment 1 were computed. For an overview, see Table 1.

As in Experiment 1, there was no difference in free recall between EPTs (.60) and SPTs (.57, $t < 1$). However, EPTs led to better serial recall than SPTs (.24 vs. .16, $t(34) = 2.30, p < .05$). The ARC' scores were better in EPTs (.39) than in SPTs (.26, $t(34) = 2.10, p < .05$). The input-output correspondence scores were high, .91 for EPTs and .84 for SPTs, and clearly above chance level. Moreover, the difference between EPTs and SPTs approached significance ($t(34) = 2.01, p = .06$).

Hence, the pattern of results was by and large as expected. It corresponded to that of Experiment 1. The only obvious difference was that the serial order scores increased as compared to Experiment 1, whereas the free recall performance was unchanged. The increase of serial order scores showed that the instruction was efficient. In this experiment, even input-output correspondence scores approached the 5% level of significance, and the input-output correspondence scores were high and clearly above chance level. This finding might be due to the fact that if neighborhood associations are better retained, ordinal ordering also improves.

The unchanged free recall scores again show that free recall performance is largely independent of order information. Increasing order information does not necessarily increase free recall.

However, Pearson correlations showed that in Experiment 2 all correlations between free recall and the organizational scores (except the correlations between free recall and input-output correspondence) were positive and significant. These correlations are presented in Table 3.

The fact that in spite of the high input-output correspondence scores in this experiment, the correlations between free recall and input-output correspondence scores were not significant, whereas the correlations were significant between free recall and serial recall as well as between free recall and ARC' scores, confirms the conclusions from Experiment 1 that longer distance associations are not the critical basis of serial order encoding and that input-output correspondence scores are not appropriate to measure serial order information. This information rather seems to be based on pair associations.

As Table 3 also shows, all correlational coefficients (except the one for FR-IO) were increased by the explicit instructions to use order information, and they were substantially higher for EPTs than for SPTs. These findings show that in Experiment 2, serial recall contributes to recall performance and more so in EPTs than in SPTs. On the other hand, recall level was not increased from Experiment 1 to Experiment 2. The latter finding indicates that free recall depends on more than a serial retrieval strategy. A good candidate for this other

Table 3 Pearson correlation coefficients between free recall (FR) and serial recall (SR), between free recall and input-output correspondence (IO) and between free recall and ARC' scores. The correlations were calculated separately for EPTs and SPTs in Experiment 2. For each condition, N was 18

	EPT		SPT	
FR-SR	r = .92	p < .001	r = .70	p < .01
FR-ARC'	r = .88	p < .001	r = .57	p < .05
FR-IO	r = .34	p = .16	r = .37	p = .13

EPT = experimenter-performed task; SPT = subject-performed task determinant of free recall might be item information. We will discuss this point further in the general discussion.

The differences between Experiments 1 and 2 were analyzed directly in analyses of variance including experiment and type of task as factors. The analysis of free recall scores showed no effects (type of encoding $F = 1.3$, experiment and interaction both $F < 1$).

However, the analysis of serial recall showed effects of type of task (EPT = .16, SPT = .10, $F(1,104) = 12.68$, $MSE = .0072$, $p < .001$) and of experiment (Exp. 1 = .06, Exp. 2 = .20, $F(1,104) = 66.96$, $MSE = .0072$, $p < .001$). The interaction of both factors was not significant ($F(1,104) = 2.36$, $MSE = .0072$, $p < .13$).

The analysis of ARC' scores yielded effects of type of task (EPT = .24 vs. SPT = .15, $F(1,104) = 9.24$, $MSE = .0226$, $p < .01$) and of experiment (Exp. 1 = .06 vs. Exp. 2 = .33, $F(1,104) = 73.88$, $MSE = .0226$, $p < .001$).

In addition, the analyses of input-output correspondence scores yielded two main effects. Input-output correspondence was higher in EPTs (.69) than in SPTs (.63), $F(1,102) = 4.51$, $MSE = .0178$, $p < .05$, and it was higher in Experiment 2 (.86) than in Experiment 1 (.45), $F(1,102) = 235.08$, $MSE = .0178$, $p < .001$.

As is obvious from Table 1 and from the analyses, the serial order instruction was efficient. It enhanced those scores which are directly based on serial order information. However, as the free recall scores show, increasing serial information and using it at retrieval does not necessarily mean that free recall performance is increased. Focusing explicitly on serial order information may have the side effect of decreasing encoding and/or using other types of information during recall. In any case, it seems that manipulating the encoding of serial information changes retrieval strategies during recall as the comparison of Experiments 1 and 2 shows. Hence, it is important to know whether using an order reconstruction test after free recall increases order information similarly as the explicit instruction to use order information at study and test.

Experiment 3

Because a number of studies used order reconstruction following free recall (e.g. DeLosh & McDaniel, 1996; Engelkamp & Dehn, 2000; Serra & Nairne, 1993), a critical question was whether the experience of order reconstruction testing stimulates order-based retrieval processes in free recall. In order to explore this question, we introduced order reconstruction testing in Experiment 3. If order reconstruction testing increases order encoding, we should expect similar serial recall and organizational scores as in Experiment 2. These scores should also be higher in Experiment 3 than in Experiment 1. In addition, order reconstruction scores should be higher in EPTs than in SPTs, indicating that order reconstruction taps serial order information. The same assumption should hold true for a positional recall test in which the participants have to recall the items and assign them to the positions they appeared in at study.

Experiment 3 was identical to Experiment 1 with the exception that each free recall was succeeded by two tests: a positional recall and a reconstruction test. The positional recall was introduced to explore how much of the order reconstruction performance was due to presenting the items at test and what role the fact that the items must also be retrieved plays for order memory.

Method

Participants Forty participants, all students at Saarland University, took part in the experiment. They were paid for their participation. Twenty of them received EPTs and twenty received SPTs.
Procedure Experiment 3 followed the procedure of Experiment 1 with the exception that free recall was followed by the task to mark each recalled item with a number corresponding to the position of that item during study list presentation. This positional recall was then followed by the use of an order reconstruction test for every list. For the reconstruction test, the booklet with the general instructions, d2 s, and free recall tests contained an additional page after each free recall sheet. On top of this page, all items of the previous study list were presented in random order. Participants were instructed to arrange the items according to the order in which they were presented during study. There was no time restriction for the reconstruction task. It should further be noted that the memory instructions were unspecific, as in Experiment 1.

Results and discussion

The positional recall (PR) indicates the proportion of free recalled items marked with correct positions divided by the maximum number of eight items. An overview about the findings except for order reconstruction and positional recall is given in Table 1.

Again, there was no difference in free recall between EPTs and SPTs (.65 vs. .62, $t < 1$). The serial recall was numerically better in EPTs than in SPTs (.22 vs. .14). However, this difference failed to reach the usual level of significance ($t(38) = 1.69$, $p = .10$).

Positional recall (EPT = .43 vs. SPT = .32).[2] as well as order reconstruction (EPT = .63 vs. SPT = .50) were higher in EPTs than in SPTs (PR, $t(38) = -1.98$, $p < .06$; OR, $t(38) = 2.54$, $p < .05$). Also, the ARC' scores showed significant differences. The ARC' score was higher for EPT (.35) than for SPT (.19, $t(38) = 2.31$, $p < .05$). However, input-output correspondence scores did not differ between EPT (.73) and SPT (.65, $t(38) = 1.32$, $p = .20$).

To compare performances in Experiments 1 and 3 directly for those scores which were comparable, analyses of variance were carried out, including type of task (EPT, SPT) and experiment (Exp. 1, Exp. 3) as factors (see Table 1).

[2]The positional recall findings are unchanged if the performance is computed as the proportion of recalled items. In this case, the score is .63 for EPTs and .50 for SPTs ($t(38) = -2.22$, $p < .05$). This finding shows that it does not depend on recall level.

The analyses yielded no effects in free recall, neither for type of task, nor for experiment. However, all other scores showed significant effects of type of task and of type of experiment. The effects of type of task were significant for serial recall (EPT = .15, SPT = .09, $F(1,108) = 7.97$, $MSE = .01$, $p < .01$); for ARC' scores (EPT = .22, SPT = .11, $F(1,108) = 10.91$, $MSE = .03$, $p < .001$); and for input-output correspondence scores (EPT = .60, SPT = .54, $F(1,108) = 4.03$, $MSE = .03$, $p < .05$). The effects of experiment were significant for serial recall (Exp. 1 = .06, Exp. 3 = .18, $F(1,108) = 36.23$, $MSE = .01$, $p < .01$); for ARC' scores (Exp. 1 = .06, Exp. 3 = .27, $F(1,108) = 40.06$, $MSE = .03$, $p < .001$), and for input-output correspondence scores (Exp. 1 = .45, Exp. 3 = .69, $F(1,108) = 58.23$, $MSE = .023$, $p < .001$). None of the interactions was significant.

The data show that the inclusion of positional recall and order reconstruction testing had an effect on all scores which are directly based on serial order information. Positional recall as well as order reconstruction show the typical EPT over SPT advantage. This finding replicates the finding of Engelkamp and Dehn (2000) as well as more generally the advantage of those conditions that are presumed to provide better order information than the corresponding control conditions (e.g. DeLosh & McDaniel 1996; Serra & Nairne, 1993). These EPT effects of positional recall and order reconstruction correspond to the EPT advantages observed for serial recall and ARC' scores. Although the serial recall score in Experiment 3 only approached significance, it was significant if Experiments 1 and 3 were analysed together. This finding and the finding that the EPT advantage did not interact with the factor experiment suggests that Experiment 3 failed to show the effect due to lack of power.

The same argument can be applied to the input-output correspondence scores which were not different between EPT and SPT in Experiment 3, but they were in a common analysis of Experiments 1 and 3 with more test power. Altogether, the conclusion that input-output correspondence scores are less sensitive to measuring differences in order information than ARC' scores, for instance, is supported again.

If compared with the data of Experiment 1, it becomes clear that the introduction of the positional recall and the order reconstruction test had a very similar effect as introducing an instruction which makes participants pay explicit attention to serial order. In both cases, the encoding and use of serial order in recall is increased.

At the same time, free recall shows little variation across tasks (EPT, SPT) as well as across experiments (Exp. 1, 2, 3). However, again there were significant correlations between free recall performances and organizational scores (except for input-output correspondence scores) in Experiment 3. They are summarized in Table 4.

The correlations indicate clearly that serial information contributes to free recall, and the correlations between free and serial recall which are higher for EPTs than for SPTs show that serial information contributes more to EPT recall than to SPT recall. This contribution is again restricted to pair associations and does not hold true for associations between items which are not neighbors in the study list – as the non-significant correlation between free recall and input-output correspondence shows. On the other hand, the relative constant level of free recall performance shows that in spite of the variable amount of serial order information, information other than just serial information is involved. It also seems as if the other information compensates for possible decreases in serial information.

Table 4 Pearson correlation coefficients between free recall (FR) and serial recall (SR), between free recall and input-output correspondence (IO) and between free recall and ARC' scores. The correlations were calculated separately for EPTs and SPTs in Experiment 3. for each condition, N was 20

	EPT		SPT	
FR-SR	r = .72	p < .001	r = .46	p < .05
FR-ARC'	r = .57	p < .01	r = .62	p < .01
FR-IO	r = .29	p = .22	r = .14	p = .57

EPT = experimenter-performed task; SPT = subject-performed task

General discussion

The reported experiments yielded four clear-cut results.

Firstly, they showed that in free recall of EPTs more serial information is used than in recall of SPTs if both tasks are manipulated between subjects.

Secondly, our findings showed that the degree of serial information used in free recall of EPTs and SPTs can vary. Serial information can be enhanced by the explicit instruction to use it. However, it can also be enhanced if a positional recall and an order reconstruction follow free recall.

Thirdly, serial information contributes to free recall as the correlations between free recall and organizational scores show. The relatively invariant level of free recall across the experiments, however, also makes clear that something other than serial information contributes to free recall, too.

Fourthly, input-output correspondence is not a particularly appropriate score to measure serial order information in the free recall of short lists.

These findings have clear implications for the item-order approach. They confine this approach further. On the one hand, in line with the approach, they show that encoding conditions – here EPT and SPT – differ with respect to the amount of order information which they provide and to the degree with which order information is used in free recall.

Furthermore, our data propose that a substantial part of the serial information in the experiments from the literature (e.g. DeLosh & McDaniel, 1996; Engelkamp & Dehn, 2000; McDaniel et al., 1995; Mulligan, 2000a; Serra & Nairne, 1993) supporting the item-order hypothesis is probably due to the fact that order reconstruction was measured following free recall testing. Order reconstruction might have induced serial order encoding. This finding means that the spontaneous use of order information is greatly overestimated, even with short lists.

Our data also show that the degree to which serial order information is encoded and use can be influenced by explicit serial order encoding instructions (cf. also McDaniel et al., 2000).

So far, two assumptions of the item-order hypothesis are basically supported, but also differentiated: items and tasks may vary with regard to order information, and order information is used in free recall.

However, it must be mentioned that the present findings are based on lists of action phrases which consist of more than one word. The item-order account was initially based on experiments which used lists of single nouns. Whether our findings can be generalized to lists of nouns remains to be explored. Nevertheless, the item-order account is not explicitly confined to lists of nouns, and, as the comparison of EPTs and SPTs shows, these conditions differ, and the difference depends on the type of design as predicted by this account (Engelkamp & Dehn, 2000).

However, one assumption of the item-order approach seems to be unjustified and not in line with the present findings: the assumption that retrieval in free recall is predominantly order-based. That this assumption does not hold true for a mixed list condition has been shown before (e.g. Engelkamp & Dehn, 2000; McDaniel et al., 2000). Mc Daniel et al. (2000) extended this objection to bizarre material in general. They concluded that "the variations in order encoding across bizarre and common materials cannot completely account for the bizarre imagery patterns in free recall because the encoded order information does not appear to be involved in free recall of bizarre sentences" (p. 1053). In order to explain their findings, they proposed a differential retrieval view. According to this view, retrieval can be based on item distinctiveness. This retrieval strategy is particularly applied when mixed lists are used. However, retrieval can, as assumed by the item-order hypothesis, also be based on serial order. This strategy is particularly used in pure lists with common items.

Our findings suggest that retrieval in free recall of pure lists is also not necessarily order-based. As Experiment 1 of the present study showed, free recall is almost independent of order information if order information is not specifically induced by either an explicit serial encoding instruction or implicitly by an order reconstruction test following the free recall across repeated trials. However, free recall level in this case was as high as in Experiments 2 and 3 in which encoding and use of order information in free recall was specifically induced. Furthermore, it did not differ between EPTs and SPTs, although the amount of order information was too small to explain the null effect along the lines of the item-order hypothesis.

However, also the findings of Experiments 2 and 3 which – in line with previous findings (Engelkamp & Dehn, 2000) – show no differences in free recall of EPTs and SPTs, but clearly more order information with EPTs than with SPTs do not prove that free recall of pure lists is mainly based on order information. If this was the case, the recall level of Experiment 1 should have been lower than that of Experiment 2, which was clearly not the case (see above).

In order to account for the free recall findings, we suggest that more than one retrieval strategy can be used in free recall and that these strategies are used flexibly (e.g. Greene, 1992; Zimmer et al., 2000). Moreover, we suggest that the retrieval processes vary depending on the relative availability of item and order information. That means that SPTs do not only provide more item information and less order information than EPTs, but that as a consequence of this difference, retrieval of SPTs is more dominated by a distinctiveness-based process such as the pop-out proposed by Zimmer et al. (2000), whereas retrieval of EPTs may induce relatively more order-based retrieval processes. The result might be that the recall level of EPTs and SPTs does not differ, as was repeatedly observed.

However, so far, these assumptions do not explain why the recall level for EPTs and SPTs, respectively, do not differ between Experiment 1 and Experiments 2 and 3. In Experiment 1, there was clearly less order information available and used in free recall than in Experiments 2 and 3. In order to explain why these differences in order information in recall did not result in different recall levels, two explanations are conceivable. Firstly, variations in the availability of serial order information might influence the availability of item information (e.g. DeLosh & McDaniel, 1996). If serial order information decreases, item information could increase. However, Mulligan (2000b, 2001) showed that a decrease of serial information does not necessarily entail an increase of item information. Therefore, the second explanation seems to be more likely. Serial order information might vary independently of item information and retrieval. Retrieval strategies could change depending on the availability of serial order and item information. The higher the amount of order information is relative to item information, the more likely is an order-based retrieval strategy. If the amount of order information is small relative to item information, a distinctiveness-based retrieval strategy could be more likely. As a consequence, it should be possible to keep the recall on a relatively constant level across changes in serial order information by giving more weight to item information. Other types of information could also be involved but are not discussed here.

Taken together, retrieval processes must be more flexible than suggested by the item-order hypothesis, and item information may play a more important role in retrieval even of short and pure lists than the item-order hypothesis suggests (cf. Hunt & McDaniel, 1993).

Of course, the question arises as to why the assumption that free recall can be predicted from the item-order framework was not questioned earlier. The reason is obviously that most studies focused on the design effect in free recall. The typical pattern is that there is a free recall advantage for one encoding condition in a within-subjects mixed list design which disappears in a between-subjects pure list design. Moreover, order information was mainly measured by an order reconstruction test (e.g. DeLosh & McDaniel, 1996; Engelkamp & Dehn, 2000; Mulligan, 1999; Serra & Nairne, 1993). The typical finding for order reconstruction is that in pure lists order encoding is better in one than in the other condition and that this effect disappears in mixed lists. The explanation that given equal order information with mixed lists, a free recall advantage for the condition with better item information occurs, is in line with the item-order approach. This advantage should be reduced, disappear or even reverse with pure lists because the condition providing poorer item information now provides better order information than the other condition. That the findings so far are consistent with the item-order hypothesis does not mean that other explanations can not explain this pattern of findings. For mixed lists, the item order explanation was questioned, and other explanations have been offered due to some inconsistent findings (e.g. Engelkamp & Dehn, 2000; McDaniel et al., 2000). However, little effort has been made to test the central assumption of the item-order framework that at least in short pure lists, free recall is determined by order information and order-based retrieval. In order to test this assumption more directly, two manipulations were necessary: first to calculate order-based retrieval directly from free recall performance, and second to vary the amount of order information within one encoding condition. These manipulations were realized in the present study.

Although in some studies (e.g. Burns et al., 1993; McDaniel et al., 1995; DeLosh & McDaniel, 1996) order information in free recall was tested directly by computing input-output correspondences, this score proved to be rather unreliable. This unreliability was shown again in the present study. Therefore we used the ARC' score and computed serial recall in addition to free recall.

We know of only one study in which the amount of order information was also varied with pure lists. McDaniel et al. (2000, Exp. 2) compared memory for common and bizarre items under standard and explicit serial instructions. However, in their experiment there was always an order-reconstruction test following free recall. They observed no change in free recall for common items, but an increase of recall for bizarre items from standard to serial instructions. Hence only half the findings support the assumptions of the item-order hypothesis. However, it is yet unclear why the recall of bizarre items benefitted from an explicit serial encoding instruction and common items did not. Therefore, further research is desirable to improve our understanding of the mechanisms underlying free recall.

Acknowledgements This research was supported by a grant from the Deutsche Forschungsgesellschaft (En 124/13). We would like to thank Nicola Ferdinand, Jochen Glössner, Hedda Janssen and Annabell Saffran for their assistance with data collection.

References

Anderson, R.C. & Watts, G.H. (1969). Bidirectional associations in multi-trial free recall. *Psychonomic Science, 15*, 288–289.

Asch, S.E. & Ebenholz, S.M. (1962). The process of free recall: Evidence for non-associative factors in acquisition and retention. *Journal of Psychology, 54*, 3–31.

Bjork, R.A. & Whitten, W.B. (1974). Recency-sensitive retrieval processes in long-term free recall. *Cognitive Psychology, 6*, 173–189.

Bousfield, A.K. & Bousfield, W.A. (1966). Measurement of clustering and of sequential constancies in repeated free recall. *Psychological Reports, 19*, 935–942.

Brickenkamp, R. (1962). *Test d2. Aufmerksamkeits-Belastungs-Test* (Test d2. Test of attentional capacities). Göttingen: Hogrefe.

Burns, D.J., Curti, E.T. & Lavin, J.C. (1993). The effects of generation on item and order retention in immediate and delayed recall. *Memory & Cognition, 21*, 846–852.

Cohen, R.L. (1981). On the generality of some memory laws. *Scandinavian Journal of Psychology, 22*, 267–281.

Cohen, R.L. (1983). The effect of encoding variables on the free recall of words and action events. *Memory & Cognition, 11*, 575–582.

DeLosh, E.L. & McDaniel, M.A. (1996). The role of order information in free recall: Application to the word-frequency effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1136–1146.

Engelkamp, J. (1995). Visual imagery and enactment in memory of actions. *British Journal of Psychology, 86*, 227–240.

Engelkamp, J. & Dehn, D.M. (2000). Item and order information in subject-performed tasks and experimenter-performed tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 671–682.

Engelkamp, J. & Zimmer, H.D. (1997). Sensory factors in memory for subject-performed tasks. *Acta Psychologica, 96*, 43–60.

Glenberg, A.M., Bradley, M.M., Stevenson, J.A., Kraus, T.A., Tkaschuk, M.J., Gretz, A.L., Fish, J.H. & Turpin, B.A.M. (1980). A two-process account of long-term serial position effects. J*ournal of Experimental Psychology: Human Learning and Memory, 6*, 355–369.

Greene, R.L. (1992). Human memory. Paradigms and paradoxes. Hillsdale, N.J.:Erlbaum.

Greene, R.L., Thapar, A. & Westerman, D.L. (1998). Effects of generation on memory for order. *Journal of Memory and Language, 38,* 255–264.

Hunt, R.R. & Einstein, G.O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning and Verbal Behavior, 20*, 497–514.

Hunt, R.R. & McDaniel, M.A. (1993). The enigma of organization and distinctiveness. *Journal of Memory and Language, 32*, 421–445.

Li, S.C. & Lewandowski, S. (1993). Intralist distractors and recall direction: Constraints of models of memory for serial order. *Journal of Experimental Psychology: Learning, Memory & Cognition, 19,* 895–908.

Li, S.C. & Lewandowski, S. (1995). Forward and backward recall: Different retrieval processes. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 21,* 837–847.

McDaniel, M.A., Einstein, G.O., DeLosh, E.L., May, C. & Brady, P. (1995). The bizarreness effect: It's not surprising, it's complex. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 422–435.

McDaniel, M.A., DeLosh, E.L. & Merritt, P. (2000). Order information and retrieval distinctiveness: Recall of common versus bizarre material. *Journal of Experimental Psychology: Learning, Memory and Cognition, 26,* 1045–1056.

Mulligan, N.W. (1999). The effects of perceptual interference at encoding on organization and order: Investigating the roles of item-specific and relational information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 54–69.

Mulligan, N.W. (2000a). Perceptual interference and memory for order. *Journal of Memory and Language, 43,* 680–697.

Mulligan, N.W. (2000b). Perceptual interference at encoding enhances item-specific encoding and disrupts relational encoding: Evidence from multiple recall tests. *Memory & Cognition, 28,* 539–546.

Mulligan, N.W. (2001). Word frequency and memory: Effects on absolute versus relative order memory and on item memory versus order memory. *Memory & Cognition, 29*, 977–985.

Nairne, J.S., Riegler, G.L. & Serra, M. (1991). Dissociative effects of generation on item and order retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 702–709.

Pellegrino, J.W. (1971). A general measure of organization in free recall for variable unit size and internal sequential consistency. *Behavior Research Methods & Instrumentation, 3,* 241–246.

Serra, M. & Nairne, J.S. (1993). Design controversies and the generation effect: Support for an item-order distinction. *Memory & Cognition, 21*, 34–40.

Sternberg, R.J. & Tulving, E. (1977). The measurement of subjective organization in free recall. *Psychological Bulletin, 84*, 539–556.

Tulving, E. (1985). How many memory systems are there? *American Psychologist*, *40*, 385–398.

Wheeler, M.A., Stuss, D.T. & Tulving, E. (1997). Toward a theory of episodic memory: the frontal lobes and the autonoetic consciousness. *Psychological Bulletin, 121*, 331–354.

Zimmer, H.D., Helstrup, T. & Engelkamp, J. (2000). Pop-out into memory: A retrieval mechanism that is enhanced with the recall of subject-performed tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 658–670.