Item and Order Information in Subject-Performed Tasks and Experimenter-Performed Tasks

Johannes Engelkamp
Saarland University

Doris M. Dehn
Technical University of Berlin

This study investigated the enactment effect from the perspective of the item–order hypothesis (e.g., M. Serra & J. S. Nairne, 1993). The authors assumed that in subject-performed tasks (SPTs), item encoding is improved but order encoding is disrupted compared with experimenter-performed tasks (EPTs), that order encoding of EPTs is only better in pure lists, and that the item–order hypothesis is confined to short lists. Item information was tested in recognition memory tests, order information in order reconstruction tasks, and both item and order information in free-recall tests. The results of 5 experiments using short (8 items) and long lists (24 items) in a design with list type (pure, mixed) and encoding condition (EPT, SPT) as factors supported the hypotheses.

In recent years, some intriguing parallels in the pattern of results for such disparate memory phenomena such as the generation effect (Slamecka & Graf, 1978), the word-frequency effect (Gregg, 1976), the bizarreness effect (e.g., Einstein & McDaniel, 1987), and the perceptual interference effect (Hirshman & Mulligan, 1991) have been established. These parallels refer to the conditions under which a certain type of item or a certain type of encoding condition yields better memory performance than its control counterpart. For example, the generation of a study item usually enhances memory performance compared with a situation in which the study item is provided by the experimenter. This generation effect, however, is only stable if memory performance is tested in a recognition test (Begg & Snider, 1987; Hirshman & Bjork, 1988; Slamecka & Katsaiti, 1987; Watkins & Sechler, 1988) or, for a free-recall test, if generated and nongenerated items are intermixed in the study list (Begg & Snider, 1987; Hirshman & Bjork, 1988; McDaniel, Waddill, & Einstein, 1988; Nairne, Riegler, & Serra, 1991; Slamecka & Katsaiti, 1987). In the same way, the word-frequency effect is limited to certain experimental conditions. Low-frequency words are usually better recognized than high-frequency words (Gorman, 1961; Mandler, Goodman, & Wilkes-Gibbs, 1982; Shepard, 1967; Woodward & Freund, 1970). Nevertheless, if memory is tested in a free recall, memory performance for high-frequency words exceeds memory performance for low-frequency words, provided that unmixed lists of high- and low-frequency words are used (e.g., Duncan, 1974; Gregg, 1976; May & Tryk, 1970). Similarly, bizarreness effects in free recall only emerge if bizarreness is manipulated within subjects and mixed lists are used (e.g., McDaniel & Einstein, 1986, 1989; Merry, 1980; Pra Baldi, de Beni, Cornoldi, & Cavedon, 1985; Wollen & Cox, 1981); the effects disappear if bizarreness is manipulated in a between-subjects design with pure lists (e.g., Cox & Wollen, 1981; McDaniel & Einstein, 1986; Wollen, Weber, & Lowry, 1972). Also, perceptual interference effects in free recall depend on whether type of encoding is manipulated in pure or mixed study lists (Mulligan, 1999).

The Item–Order Hypothesis

Thus, there appear to be two variables that moderate the emergence of some memory effects: the type of memory test and the experimental design used. To account for the pattern of results pertaining to the generation effect, Nairne et al. (1991; see also Serra & Nairne, 1993) proposed a theoretical account that is based on the distinction between item-specific and relational information (e.g., Einstein & Hunt, 1980; Hunt & Einstein, 1981). Item-specific information relates to unique features of the study item (e.g., specific associations that are activated on presentation of a particular item), whereas relational information associates different items in a study list with each other. One specific type of relational information is information on the presentation order of items at study.

In their item–order hypothesis, Nairne et al. (1991) claimed that some encoding conditions, such as generation, have opposite effects on item-specific and order information. A generated item provides better item-specific information, but it does not encourage the encoding of serial order. A read item, on the other hand, does not yield as good item-specific information, but it encourages order encoding to a larger
extent. Given the assumption of a trade-off between item-specific and order information, it becomes evident why the design type and the memory test used can moderate the emergence of a generation effect. With regard to the design type, nongenerated items can only yield better order information if they are presented in pure study lists. If generated and nongenerated items are presented in mixed study lists, the generation of some items in the list also hampers the encoding of serial order for nongenerated items. With regard to memory tests, different types of tests are assumed to draw to a different extent on item-specific and order information. A recognition task is thought to provide a relatively process-pure measure of item-specific information. This is due to the fact that the identification of target items among a set of possible candidates does not require information on serial order. However, a free recall additionally requires the generation of output candidates that can be done, for example, by using a serial output strategy. For this reason, the better item-specific information for generated items can be masked by the better order information for nongenerated items in a free recall, but not in a recognition test.

The item-order hypothesis has been shown to adequately explain the findings not only for the generation effect (Nairne et al., 1991; Serra & Nairne, 1993) but also for the bizarreness effect (McDaniel, Einstein, DeLosh, May, & Brady, 1995), the word-frequency effect (DeLosh & McDaniel, 1996), and the perceptual interference effect (Mulligan, 1999). In the present study, the item-order hypothesis is applied to the enactment effect. On the basis of the enactment effect, we demonstrate that the item-order hypothesis in its present formulation is restricted to the encoding and retrieval of relatively short study lists.

The Enactment Effect

When individuals are requested to learn a list of action phrases (e.g., "lift the pen," "smoke the pipe"), memory performance is usually enhanced by enacting the denoted action in addition to verbally processing the phrase. The memory-enhancing effect of self-performing a phrase is referred to as the enactment effect (e.g., Engelkamp, 1998; Engelkamp & Zimmer, 1994). If, however, a condition in which the individuals perform the actions themselves (i.e., a subject-performed task [SPT]) is contrasted with a condition in which the individuals observe the experimenter performing the actions (an experimenter-performed task [EPT]), then the pattern of results becomes more complex. For example, Cohen (1981, 1983; Cohen & Bean, 1983; Cohen, Peterson, & Mantini-Atkinson, 1987) observed no differences in free recall of EPTs and SPTs. In contrast, Dick, Kean, and Sands (1989) as well as Engelkamp and Zimmer (1983) observed a clear-cut recall advantage of SPTs over EPTs. This difference in results led Engelkamp (1990) to speculate that the use of different list lengths might be the crucial variable. However, as demonstrated by Engelkamp and Zimmer (1997), it is both the type of design and the list length that determine the pattern of results: For short study lists of up to 18 items, the SPT advantage was only observed if the encoding condition was manipulated within subjects.

In a between-subjects design, there was no recall difference between SPTs and EPTs. In contrast, for long study lists of more than 30 items, there was generally an SPT effect regardless of whether the encoding manipulation was carried out within or between subjects.

To explain this pattern of results, Engelkamp and Zimmer (1997) applied the item-order hypothesis to the enactment effect. Specifically, they assumed the following: First, SPTs provide better item-specific information than EPTs, whereas EPTs provide better order information than SPTs. Second, order encoding of EPTs suffers when EPTs occur in the context of SPTs, that is, when EPTs and SPTs are presented in mixed study lists.

According to Engelkamp (1995), item-specific information should be better for SPTs than for EPTs because the execution of an action forces the individual to focus on action-relevant information and to ignore action-relevant context information. Focusing solely on action-relevant information is necessary to guarantee smooth enactment. In contrast, it is assumed that EPTs do not require such an attentional restriction on item information but allow for context encoding. This latter claim is supported by the findings of Koriat, Ben-Zur, and Druch (1991). They showed that although recognition memory of action phrases was better for SPTs than for EPTs, memory for context (the room in which the items were studied) was better for EPTs than for SPTs. Hence, watching the experimenter performing actions should be better suited to encode order information than self-performing actions.

Given a test that captures both item and order information—such as a free recall—the better item information for SPTs will yield an enactment effect in free recall whenever order information does not differ between SPTs and EPTs (such as in the within-subjects design); however, if order information is better for EPTs than for SPTs (such as in the between-subjects design), the enactment effect will be masked. Exact predictions as to whether the enactment effect will increase, will completely vanish, or will be reversed in the between-subjects design are not possible though. Depending on the differences in order information and the differences in item information, there can be a recall advantage for either condition or no difference in performance at all. Hence, the only tenable prediction to be made is that any recall advantage for SPTs in mixed lists has to decrease in pure lists, but it cannot be augmented.

A Further Moderating Variable: Length of the Study List

The assumptions concerning short lists merely constitute an application of the item-order hypothesis to the enactment effect. However, the item-order hypothesis cannot fully explain the pattern of results for the comparison of SPTs and EPTs in long lists. The interaction of design type and encoding conditions in free-recall performance is only observed if relatively short study lists are used. To account for the effect of list length, Engelkamp and Zimmer (1997) added a further assumption to those contained in the item-order hypothesis, namely that for long study lists, a
serial output strategy is no longer feasible because order encoding and, consequently, order-based retrieval become too difficult. For this reason, alternative retrieval strategies are used. These strategies rely on relational information that is part of semantic long-term knowledge as, for example, reflected in organizational scores of categorically structured lists (e.g., Engelkamp & Zimmer, 1996) or in scores of subjective organization of unrelated lists (Engelkamp, Zimmer, & Mohr, 1990; Koriat, Pearlman-Avnon, & Ben-Zur, 1998). The crucial point for the present argument is that semantic long-term knowledge is not influenced by the encoding condition and, hence, does not differ for SPTs and its control conditions (e.g., Koriat et al., 1998). For this reason, differences in free recall of SPTs and EPTs are mainly determined by item information, and the memory-enhancing effect of enactment is no longer moderated by the experimental design.

Overview of the Experiments

To test the above assumptions, we conducted a series of five experiments in which we manipulated the encoding condition (self-performed vs. experimenter-performed task) and the experimental design (within subjects vs. between subjects). Memory performance was assessed on the basis of three tests. A recognition test was used as a measure of item memory (cf. Hunt & Einstein, 1981). An order reconstruction was used as a measure of order information (e.g., Serra & Nairne, 1993). Finally, a free recall was assumed to reflect both item and order information. A further manipulation concerned the length of the study lists: In Experiments 1–3, study lists consisted of 8 action phrases, whereas in Experiments 4 and 5, study lists consisted of 24 action phrases.

Experiment 1

Experiment 1 was conducted to examine the effects of the encoding manipulation (SPT vs. EPT) and the design type (manipulation of encoding within subjects vs. between subjects) on performance in two memory tests: a free-recall and a recognition test. Design type was manipulated between subjects, whereas test type was manipulated within subjects. The manipulation of the encoding condition depended on the level of the design type: For participants in the within-subjects design, the encoding condition served as a within-subjects variable, whereas for participants in the between-subjects design, encoding condition served as a between-subjects variable.

Method

Participants. Forty-eight undergraduate students at Saarland University participated in the experiment. They were paid for participation. Sixteen participants were randomly assigned to the pure SPT group, 16 were assigned to the pure EPT group, and the remaining 16 were assigned to the mixed EPT–SPT group.

Material. Sixty-four action phrases were selected from a larger pool of phrases. From these items, eight lists of 8 phrases were constructed. For every list, the phrases were selected in such a way that they were phonologically and semantically unrelated. In addition to the 64 test items, an additional practice list with 8 items was created. All items were presented to the participants visually on a computer screen.

For the final recognition test, 64 distractor phrases were constructed by replacing each verb with a motorically dissimilar one. Thus, a distractor phrase only differed from the corresponding target phrase in its verb, but it shared the object. For example, if the target phrase reads “eat the egg,” the corresponding distractor phrase would read “peel the egg.” Thus, in the whole set of 128 critical phrases, each verb only occurred once, whereas each object was used twice: once in the target set and once in the distractor set.

To balance possible confoundings of study positions (within a single list) and item material, items were assigned to study positions in such a way that, over all participants, each phrase appeared equally often on 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.

For the mixed EPT–SPT condition, the sequence of encoding conditions was randomly determined in advance, with the restriction that no more than two subsequent phrases were encoded in the same way. To avoid confoundings of type of encoding and material, the whole set of 64 study items was randomly divided into two sets, with each set serving as a watch item for half of the mixed EPT–SPT participants and enact items for the other half of the mixed EPT–SPT participants.

Procedure. Participants were tested individually in a single experimental session lasting, on average, 35 min. Generally, the experiment consisted of nine study-recall sequences, of which the first sequence only served for training purposes. After the presentation of all study-recall sequences, a final recognition test was conducted that included all the items from the eight critical study lists plus the distractor set.

In more detail, 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 with simple action phrases (e.g., “throw the ball”) that they should try to memorize. Participants in the SPT condition were instructed to enact all the phrases symbolically. Participants in the EPT condition were instructed to watch the experimenter performing all phrases symbolically. Finally, participants in the mixed SPT–EPT condition were instructed to either perform the phrase themselves or to observe the experimenter performing the phrase. For all encoding conditions, participants were further informed that in between study trials, a small “concentration test” would be carried out.

Immediately after the instruction was given, participants were presented with the first (training) 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 phrase (depending on the encoding condition). The next phrase then appeared on the screen after 1,000 ms. After the presentation of a study list (i.e., the presentation 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 “d” with one to three dashes. For the present experiment, the d2 served two purposes: first, to avoid ceiling effects and, second, to suppress recency effects. 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 free recall as well as...
the d2 were administered as paper-and-pencil tests, with the timing being supported by the computer. The above sequence—presentation of a study list, d2, and free recall—was then repeated eight times.

After the last free recall had been completed, the final recognition test took place. Participants were informed that they would be presented with all the phrases from the preceding study lists, along with a set of distractor items. They should press the right shift key for “old,” if they identified the phrase as exactly one from the study lists (same verb and same object), or, otherwise, the left shift key for “new.”

In the recognition test, each phrase was presented for up to 5,000 ms on the computer screen. Key strokes were recorded from 400 ms after the phrase onset until the phrase disappeared. As soon as the participant pressed a key, the phrase vanished. The next phrase then appeared after a pause of 1,000 ms.

Results and Discussion

Data analyses in all experiments were carried out separately for the different memory tests.

Free-recall data. A phrase produced in free recall was classified as “correct” whenever both the verb and the object corresponded to the ones used in the study list. The probabilities of free recall for both encoding types and design types are summarized in Table 1.

Numerically, there were more correct reproductions for EPTs than for SPTs in the between-subjects design. In the within-subjects design, in contrast, the probability of a correct reproduction was higher for SPTs than for EPTs. To test this interaction statistically, we used an analysis of variance (ANOVA) developed by Erlebacher (1977) for experiments that contrast within- and between-manipulations of an independent variable. Applying this procedure to the data of Experiment 1, we obtained a significant interaction of type of encoding and design, $F(1, 44) = 5.02, MSE = 0.01, p < .05$. The main effects of type of encoding and design did not achieve significance (both $Fs < 1$). To decide whether the differences in performance for EPTs and SPTs in the two design types were significant, simple contrasts, one for each design type, were calculated. For the between design, the recall advantage of EPT was marginally significant, $F(1, 44) = 3.22, MSE = 0.01, p < .08$, whereas for the within design, the SPT advantage was nonsignificant ($p = .19$).

Recognition data. Table 2 reports the probabilities of hits, false alarms, and correct recognition (measured in terms of hits − false alarms) depending on the mode of encoding and design type.

For both the between- and the within-manipulation groups, SPTs were better recognized than EPTs. Correspondingly, the Erlebacher (1977) procedure showed a significant main effect of the encoding manipulation on recognition performance, $F(1, 45) = 21.38, MSE = 0.01, p < .05$. Neither the main effect of the design nor the interaction of design and type of encoding reached significance ($F < 1$).

The findings supported the hypotheses stated above. We expected better item information for SPTs than for EPTs independently of the design type. The findings supported the expectation: Recognition memory for enacted phrases exceeded memory for observed phrases, regardless of whether the encoding manipulation was carried out within or between subjects.

As opposed to this finding, a trade-off between item and order information should determine the findings in free recall: Observing a phrase provides better order information than enacting a phrase, but only if the participant is not required to switch between encoding conditions (like in the within-subjects groups). Item information, on the other hand, is better for enacted than for observed phrases and is not hampered by the requirement to switch between encoding tasks. Consistent with these considerations, there was a significant interaction between type of encoding and design type for the free-recall data. In the between-subjects design, there was a marginally significant EPT advantage. The

Table 1
Probability of Free Recall as a Function of Type of Encoding (EPT, SPT) and Type of Design (Between, Within)

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Design type</th>
<th>EPT</th>
<th>SPT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Between</td>
<td>.60</td>
<td>.53</td>
</tr>
<tr>
<td></td>
<td>Within</td>
<td>.52</td>
<td>.57</td>
</tr>
<tr>
<td>2</td>
<td>Between</td>
<td>.52</td>
<td>.47</td>
</tr>
<tr>
<td></td>
<td>Within</td>
<td>.46</td>
<td>.54</td>
</tr>
<tr>
<td>3</td>
<td>Between</td>
<td>.33</td>
<td>.39</td>
</tr>
<tr>
<td></td>
<td>Within</td>
<td>.35</td>
<td>.41</td>
</tr>
<tr>
<td>4</td>
<td>Between</td>
<td>.36</td>
<td>.38</td>
</tr>
<tr>
<td></td>
<td>Within</td>
<td>.33</td>
<td>.38</td>
</tr>
</tbody>
</table>

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

Table 2
Probability of Hits, False Alarms, and Pr Scores (Hits Minus False Alarms) as a Function of Type of Encoding (EPT, SPT) and Type of Design (Between, Within)

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Design type</th>
<th>Hits</th>
<th>FAs</th>
<th>Prs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>EPT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Between</td>
<td>.91</td>
<td>.04</td>
<td>.87</td>
</tr>
<tr>
<td></td>
<td>Within</td>
<td>.95</td>
<td>.05</td>
<td>.90</td>
</tr>
<tr>
<td>4</td>
<td>Between</td>
<td>.95</td>
<td>.02</td>
<td>.93</td>
</tr>
<tr>
<td></td>
<td>Within</td>
<td>.97</td>
<td>.02</td>
<td>.95</td>
</tr>
</tbody>
</table>

Note. EPT = experimenter-performed task; SPT = subject-performed task; FAs = false alarms; Prs = probability of recognition scores.

1 For details of calculating the quasi-$F$s and degrees of freedom for the specific effects, see Erlebacher (1977). The simple main effects in each of the design types were tested against the error variance of the interaction term of the Erlebacher analysis.
expected SPT advantage in the within-subjects design, though, failed to reach significance. Because the latter effect did not reach significance, we replicated the pattern of results for the free recall in Experiment 2.

**Experiment 2**

In Experiment 2, we wanted to contrast free-recall performance with performance in a test that is thought to mainly draw on order information: a reconstruction test. As stated in the hypotheses, we expected superior reconstruction performance for EPTs than for SPTs if participants were not required to switch between tasks, that is, in the between-subjects group. If, in contrast, participants were required to switch between tasks, like in the within-subjects group, order encoding should be hampered, thus leveling out the original encoding advantage of EPTs over SPTs.

The design of Experiment 2 was identical to the one used in Experiment 1, with the exception that each free recall was succeeded by a reconstruction test and no final recognition test was given. That is, there were two independent variables, encoding condition (SPT vs. EPT) and design type (within-subjects vs. between-subjects), with the latter variable referring to the manipulation of the encoding condition.

**Method**

**Participants.** Sixty-three individuals with different educational backgrounds and professions participated in the experiment. They were assigned to three groups with 21 participants each. One group received the EPT instruction, the second group received the SPT instruction, and the third received the mixed EPT–SPT instruction.

**Material.** The material consisted of the 64 target phrases plus the 8 training phrases from Experiment 1.

**Procedure.** Experiment 2 essentially followed the procedure of Experiment 1. The main difference compared with Experiment 1 was the use of reconstruction tests for every list instead of one final recognition test. Thus, Experiment 2 contained nine sequences, each involving the presentation of a study list, the distractor task, a free recall, and a reconstruction test. A further modification focused on the presentation of the study items. In Experiment 2, the study items together with their encoding instruction were presented on cards, rather than on a computer screen. As a consequence, presentation times and interstimulus intervals were not taken as exactly as in the previous experiment. Further, items were not shifted systematically over study list positions. The confounding of items and study positions, though, was prevented by generating a new random order of study items for every 4th participant. As in Experiment 1, participants were tested individually.

The instructions given to the participants up to the free-recall test were identical to the ones given to participants in Experiment 1. In contrast to Experiment 1, participants were given 90 s (rather than 60 s) to write down the phrases. After the free-recall test, participants were asked to reconstruct the order of the study items. For this purpose, the booklet with the general instructions, d2s, and free-recall tests contained an additional page after each free-recall sheet. On the top of this page, all items of the previous study list were presented in a random order. Participants were instructed to arrange the items according to the order in which they were presented at study. There was no time restriction for the reconstruction task. The sequence of list presentation, d2, free recall, and reconstruction task was realized nine times, with the first sequence only serving for practice purposes.

**Results and Discussion**

**Free-recall data.** The probabilities of free recall for both encoding types and both design types are summarized in Table 1. As in Experiment 1, free-recall performance was numerically better for EPTs than for SPTs if the encoding condition was manipulated between subjects. If the encoding condition was manipulated within subjects, free recall of SPTs exceeded free recall of EPTs. The interaction between design and type of encoding was tested with the modified ANOVA suggested by Erlebacher (1977). According to this procedure, there was a significant interaction of design and type of encoding, $F(1, 59) = 4.65, MSE = 0.02, p < .05$. Neither the main effect of the design nor the main effect of the encoding condition was significant ($F < 1$). Simple contrasts for each design type showed that, for the within-manipulation group, memory performance for SPTs was, in fact, better than for EPTs, $F(1, 59) = 3.87, MSE = 0.02, p = .05$. For the between-manipulation groups, however, the difference in memory performance for SPTs and EPTs was nonsignificant ($p = .29$). Thus, with regard to the free-recall data, the results of Experiment 2 corresponded to our theoretical expectations.

**Reconstruction data.** Responses were classified as correct reconstructions if the test phrase was written down on the same list position on which it was presented in the study phase. The probability of assigning the correct study position to a phrase depending on the type of encoding and design is given in Table 3.

As expected, performance in the reconstruction task was better for EPTs than for SPTs if the encoding condition was manipulated between subjects. According to simple contrasts, this difference was statistically significant, $F(1, 57) = 9.30, MSE = 0.02, p < .01$. In contrast to this result, there was no difference in reconstruction performance for EPTs and SPTs if the encoding condition was manipulated within subjects ($F < 1$). The Erlebacher (1977) procedure revealed a marginally significant interaction of the encoding condition and design type, $F(1, 57) = 2.84, MSE = 0.02, p < .10$. Because the interaction between encoding condition and design did not reach a conventional alpha level, we postpone

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Design type</th>
<th>EPT</th>
<th>SPT</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Between</td>
<td>.52</td>
<td>.40</td>
</tr>
<tr>
<td>3</td>
<td>Within</td>
<td>.41</td>
<td>.39</td>
</tr>
<tr>
<td>3</td>
<td>Between</td>
<td>.64</td>
<td>.54</td>
</tr>
<tr>
<td>3</td>
<td>Within</td>
<td>.52</td>
<td>.53</td>
</tr>
</tbody>
</table>

*Note. EPT = experimenter-performed task; SPT = subject-performed task.

*List length: 8 items.*
a discussion of the reconstruction data until the results of Experiment 3 are reported.

**Experiment 3**

The pattern of results obtained for the reconstruction test in Experiment 2 corresponded to our expectations in that there was an EPT advantage in the between-subjects design but an equal level of performance for EPTs and SPTs in the within-subjects design. Although this finding is in line with our theoretical assumptions on the item versus order basis of remembering EPTs and SPTs, the interaction failed to reach a conventional level of significance. Further, there might be some methodological concerns with the order of tests: In Experiment 2, the reconstruction test always followed a free-recall test. Although there is evidence that order reconstruction is not influenced by a preceding recall test (cf. DeLosh & McDaniel, 1996), it cannot be completely ruled out that the order of tests influenced the results for the reconstruction test. For this reason, we conducted another experiment in which the presentation of a study list was only followed by a reconstruction test.

**Method**

**Participants.** Sixty-six students at Saarland University participated in the experiment. They were paid for participation. A third of them (i.e., 22) was assigned to the EPT condition, another third to the SPT condition, and the last third to the mixed EPT–SPT condition.

**Material.** The material was identical to that used in Experiments 1 and 2.

**Procedure.** Experiment 3 followed the same procedure as Experiment 2, with the exception that each study list was only succeeded by a reconstruction test. Hence, a single sequence consisted of the presentation of a study list, the d2, and a reconstruction test. This sequence was presented nine times (eight critical trials plus one test trial). As in Experiment 1, study items were presented with computer. Each phrase was presented for 5,000 ms on the computer screen, together with the instruction to “watch” or “enact” the phrase (depending on the encoding condition). The next phrase then appeared on the screen after 1,000 ms. Again, items were shifted over study list positions.

The instructions given to the participants were identical to the ones given to participants in Experiment 2 for the learning of study lists and the reconstruction tests. The time allocated for working through the reconstruction test was set to 120 s.

**Results and Discussion**

As in Experiment 2, responses were classified as correct reconstructions if the test phrase was written down on the same list position it was presented on in the study phase. Table 3 shows the probability of assigning the correct study position to a phrase depending on the type of encoding and design.

An ANOVA that was based on the Erlebacher (1977) procedure showed a significant interaction of encoding condition and type of design, $F(1, 59) = 4.29, MSE = 0.01, p < .05$. Simple contrasts specified this interaction. Performance was better for EPTs than for SPTs if the encoding condition was manipulated between subjects, $F(1, 59) = 7.55, MSE = 0.01, p < .01$. In contrast to this finding, there was no difference in reconstruction performance for EPTs and SPTs if the encoding condition was manipulated within subjects ($F < 1$).

Thus, Experiment 3 succeeded in replicating the pattern of results found for the reconstruction test in Experiment 2. This points to the fact that if there was any influence of the free-recall test on reconstruction performance, it was negligible. Moreover, in Experiment 3 the theoretically expected interaction of encoding condition and design type reached a conventional level of significance.

**Discussion of Experiments 1–3**

The pattern of results obtained in the Experiments 1–3 basically conforms to the theoretical expectations. As expected, recognition memory was better for SPTs than for EPTs independently of the design type. This finding supports the assumptions that SPTs provide better item information than EPTs and that item information is not affected by the context in which the item is encountered.

Also, as theoretically expected, order reconstruction was better for EPTs than for SPTs if a between-subjects design was used. According to our assumptions, the execution of an action hinders order encoding by making the participants focus on item information, that is, on information that is relevant for a particular action and its smooth enactment (e.g., Engelkamp, 1995). The necessity to concentrate on the individual item is smaller when actions are observed, and, hence, encoding of neighboring items (i.e., order encoding) is easier for EPTs than for SPTs. The order reconstruction data for the between-subjects design support these assumptions. In line with this finding, Engelskamp, Zimmer, and Denis (1989) as well as Denis, Engelkamp, and Mohr (1991) observed that cued recall of semantically unrelated actions in a paired-associate learning experiment was worse under the instruction to imagine oneself performing an action than under the instruction to imagine somebody else performing the action, given a between-subjects manipulation of the instruction. In a cued recall of paired-associate learning, participants are given one element of a pair and are required to recall the neighboring element. This requirement is similar to the one involved in order reconstruction, and, thus, cued recall findings can be expected to closely resemble the findings for the reconstruction test.

Also in line with our expectation, the better order reconstruction for EPTs in the present experiments was limited to the between-subjects design. That is, if EPTs were presented in the context of SPTs, there was no difference in order reconstruction for the two encoding conditions.

Finally, it was assumed that the free-recall data reflect both item and order information. In mixed encoding conditions, where order information is equally poor for EPTs and SPTs, the differences in item information determine the pattern of results. This should give rise to a free-recall advantage for SPTs in a within-subjects design. The results of Experiments 1 and 2 supported these assumptions, although the SPT advantage was significant only in Experi-
ment 2. In pure lists, where order information is better for
EPTs than for SPTs, the item-based SPT advantage should
be minimized or even reversed. Corresponding to this,
Experiments 1 and 2 showed a numerical EPT advantage for
free recall of pure lists that approached significance in
Experiment 1.

According to the item-order hypothesis, free-recall perfor-
ance is based on the use of a serial output strategy. If it is
further assumed that order reconstruction draws on the same
kind of information, that is, information on the serial order
of items, then there should be a correlation between order
reconstruction and free-recall performance. To test these
assumptions, we computed the correlations between order
reconstruction and free-recall performance for the data of
Experiment 2. For pure lists, there were highly significant
(correlations \( r = .83 \) for EPTs and \( r = .68 \) for SPTs,
\( p < .001 \)), whereas for mixed lists, the correlations failed to
reach significance (\( r = .33 \) for EPTs and \( r = -.13 \) for
SPTs). Thus, the correlations are only partially in line with
the claim that order reconstruction and free recall both draw
on the same kind of information. Although the high cor-
relation for EPTs in pure lists corresponds to the assumption of a
serial output strategy, the other correlations do not.

Recently, the strong assumption of the item-order hypo-
thesis that free recall (at least with short lists) is based on a
serial output strategy has been questioned by McDaniel,
DeLosh, and Merritt (1999). The authors proposed a modi-
fied item-order hypothesis, according to which different sets
of items within a retrieval set (for our purpose, EPTs and
SPTs) vary with regard to their distinctiveness. The distinc-
tiveness is used later to guide retrieval in free recall. Hence,
not for mixed lists, non-item-based retrieval strategies
might be pursued. This claim is perfectly in line with our
finding of nonsignificant correlations between order recon-
struction and free-recall performance in mixed lists. The
significant correlation for SPTs learned in pure lists indicates
that participants used a serial output strategy, although order
information for SPTs was fairly low.

Experiment 4

In Experiment 4, our goal was to demonstrate that, for
long study lists, the encoding of order information generally
becomes too difficult in order to serve as the basis for
subsequent recall. Consequently, a serial order output stra-
gy should be replaced by retrieval strategies that rely on
associative links of semantic knowledge. Because this kind of
relational information does not differ between EPTs and
SPTs (e.g., Koriat et al., 1998), item information becomes
the determining factor for free-recall differences. Therefore,
free-recall performance should mirror the pattern of results
for recognition memory, that is, there should be a design-

independent SPT advantage. We decided to use lists that
consisted of 24 phrases as long study lists. The choice of this
length was guided by the following considerations: Because
Engelkamp and Zimmer (1997) observed for lists of 15 and
18 items a similar free-recall pattern to the one observed in
Experiments 1 and 2, we wanted to use lists with more than
18 items. However, we also wanted to keep the lists as short
as possible to avoid a complete breakdown in order recon-
struction in this experiment.

Experiment 4 essentially replicated Experiment 1, in that
the encoding condition (SPT vs. EPT) was either manipu-
lated within or between subjects and memory performance
was assessed using free recall and recognition tests. In
contrast to Experiment 1 though, four study lists with 24
items each instead of eight lists with 8 items each were used.
We expected that with this increase in list length, the interac-
tion of encoding task and design type obtained for the
free-recall data in Experiments 1 and 2 would disappear.

Method

Participants. Forty-eight undergraduate students at Saarland
University participated in the experiment. They were paid
for participation. Sixteen participants were randomly assigned to
the pure SPT group, 16 were assigned to the pure EPT group, and
the remaining 16 were assigned to the mixed EPT–SPT group.

Material. In addition to the 64 phrases used as study items in
the previous experiments, 32 other phrases were selected from
a pool of action phrases. From these 96 items, four lists with 24
phrases each were constructed. The initial training list was identical
to the one used in Experiments 1–3 and, hence, only comprised 8
items. As in Experiment 1, all items were presented to the
participants visually on a computer screen.

The distractor set consisted of the 64 distractor phrases from
Experiment 1 plus 32 new distractors. Again, these distractors
were created by replacing each verb of the 32 new target items with a
motologically dissimilar one. For the 8 training list items, 8 distrac-
tors were constructed analogously. Thus, there was a total of 104
target phrases and 104 distractor phrases in Experiment 4 (with 96
items for the critical lists and 8 items for the training list in each set).

The study items were moved over study positions in blocks of
three. For this purpose, the 24 study items as well as the 24 study
positions were first divided into eight blocks of 1–3, 4–6, . . . ,
21–24. Each of the eight study blocks was then rotated over the 8
study positions in such a way that each block occurred equally
often on 1 of the 8 study positions. The order of lists (i.e., the
position of a list within the whole experiment) was the same for all
participants.

The assignment of encoding conditions to study items and
positions in the mixed EPT–SPT condition was accomplished in
the same way as in Experiments 1 and 2. The total set of 96 critical
study items was randomly divided into two sets, with each set
serving as watch items for half of the mixed EPT–SPT participants
and exact items for the other half. The sequence of encoding
conditions was determined in advance, with the restriction that no
more than 2 subsequent phrases were encoded in the same way.

Procedure. Participants were tested individually on a personal
computer with a single experimental session lasting, on average, 45
min. Experiment 4 essentially constituted a replication of Experi-
ment 1 with two modifications: First, list length was increased and
the number of lists was reduced. Second, each free recall was
succeeded by a recognition test, instead of a single recognition test
conducted at the end of the experiment. Hence, the experiment
consisted of four critical sequences of study list presentation,
distractor task, free recall, and recognition test, with each study list
comprising 24 items. An initial sequence with 8 study items served
for training purposes.

The instructions given to the participants were identical to the
ones used in Experiment 1, as were the phrase presentation times
for both the study lists and the recognition test and the time allocated for carrying out the distractor task. The only difference concerned the time given to recall and write down the study items. Because of the larger number of phrases per list, the recall time was extended to 180 s. The instructions for the recognition tests immediately followed each free recall.

**Results and Discussion**

**Free-recall data.** Table 1 outlines the probabilities of free recall for both encoding and design types. In both design types (between and within subjects), free-recall performance for SPTs was better than for EPTs. An ANOVA that was based on the Erlebacher (1977) procedure revealed a significant main effect of the encoding condition on the probability of a correct reproduction, $F(1, 35) = 6.59$, $MSE = 0.02$, $p < .05$. In contrast, neither the main effect of the design nor the interaction of design type and encoding condition was significant.

**Recognition data.** Table 2 summarizes the probabilities of hits, false alarms, and of recognition (hits – false alarms) as a function of encoding and design type. In both design types, the recognition probability for SPTs exceeded the recognition probability for EPTs. The Erlebacher (1977) ANOVA showed a significant main effect of the encoding condition, $F(1, 35) = 21.10$, $MSE = 0.01$, $p < .05$. The design type and the interaction of encoding condition and design type did not influence recognition performance.

Thus, the recognition data from Experiment 4 replicated the findings of Experiment 1, in that enacting a phrase yielded better recognition memory than observing a phrase. This finding is independent of whether the encoding task was manipulated between or within subjects and, as the joint results from Experiment 1 and 4 show, of whether short or long lists have to be learned (cf. Engelkamp, Zimmer, & Biegelmann, 1993).

With regard to the free-recall data, the findings of Experiment 4 differed from those of Experiment 1 in that no interaction of encoding condition and design type was found. Rather, as with recognition memory, enacted phrases were consistently remembered better than observed phrases. The lack of an interaction of encoding condition and design type, though, was in line with our expectations.

**Experiment 5**

In Experiment 5, we investigated how the list length influences the results for the reconstruction test. We expected that the large increase in list length would severely impair order information. This impairment of order information should equally affect enacted and observed phrases presented in both pure and mixed study lists. Therefore, scores of order reconstruction should be low, and the interaction of type of encoding and type of design should disappear.

**Method**

Participants. Participants were 48 undergraduate students at Saarland University who were paid for participation. Sixteen participants were randomly assigned to the pure EPT group, 16 were assigned to the pure SPT group, and the remaining 16 were assigned to the mixed EPT–SPT group.

**Material.** The material consisted of the 96 target phrases from Experiment 4 plus the 8 training phrases. Eight more training phrases were constructed for an additional training list (see below).

**Procedure.** The procedure of Experiment 5 closely followed the procedure of Experiment 4. A modification concerned the way the tests were carried out: In Experiment 5, each of the four study lists was tested either in a free recall or in a reconstruction test. In particular, for the first and the second list, a reconstruction test was carried out, whereas for the third and the fourth list, a free-recall test was given. This modification made it necessary to use two training lists, one for the free-recall tests and the other for the reconstruction tests. Hence, the sequence of lists presented throughout the experiment consisted of a training list for the reconstruction task, two critical lists tested in a reconstruction test, a training list for the free recall, and two critical lists tested in free recall.

The instructions given to the participants for the free recall and the reconstruction tests were identical to the ones used in previous experiments. Presentation times for study phrases and time allocated for carrying out the distractor task matched those of previous experiments. The maximum times for carrying out the free recall and the reconstruction task were set to 240 s and 480 s, respectively.

**Results and Discussion**

**Free-recall data.** The probabilities of free recall as a function of encoding and design types are shown in Table 1. In both design types (between and within subjects), free-recall performance for SPTs was numerically better than for EPTs. An ANOVA that was based on the Erlebacher (1977) procedure yielded a marginally significant main effect of the encoding condition, $F(1, 41) = 3.16$, $MSE = 0.01$, $p < .09$. All other effects were not significant.

**Reconstruction data.** A data analysis that was based on the exact positioning of items on their study positions would have yielded too low scores for lists with 24 items. Following Burns (1996), we computed a more lenient score for reconstruction performance that was based on correct relative positioning. For this purpose, the number of instances in which participants put items in adjacent output positions that were presented in adjacent input positions was divided by the number of items on the list minus one. For pure lists, this procedure yielded a mean score of 0.14 for EPTs and 0.15 for SPTs. For mixed lists, the corresponding score was 0.10. Because the scoring was based on adjacent pairs of items, there was only one reconstruction score for mixed lists. The scores were submitted to a between-subjects ANOVA, with EPT pure list, SPT pure list, and mixed list as levels of the independent variable. This ANOVA yielded no significant effects ($F < 1.20$).

After plotting order-reconstruction scores as a function of study positions, we discovered that even with the weaker criterion, almost all scores from the fifth position onward were smaller than 0.20 and about two thirds approached 0. Therefore, we calculated two separate ANOVAs, one for the first quarter of positions and the other for the remaining three quarters. For the first quarter, the mean reconstruction scores
were 0.33 for EPTs and 0.40 for SPTs in pure lists and 0.18 for mixed lists. A between-subjects ANOVA yielded a significant effect of encoding conditions, $F(2, 45) = 4.42$, $MSE = 0.05$, $p < .05$. Simple contrasts showed that there was no difference between EPTs and SPTs in pure lists ($F = 1$); however, both scores differed from the one for mixed lists, $F(1, 45) = 7.83$, $MSE = 0.05$, $p < .01$. A corresponding ANOVA for the last three quarters of study positions yielded, as expected, no effect ($F < 1$). The mean reconstruction scores in all three conditions were lower than 0.10 (EPT, pure lists = 0.08; SPT, pure lists = 0.09; mixed lists = 0.08).

The low level of reconstruction performance made it highly unlikely that, except for the first quarter of the list, order information served as a basis for generating output candidates in recall. We also computed the correlations between free recall and order reconstruction for the first quarter of the list. The correlations were .38 for EPTs in pure lists, .45 for SPTs in pure lists, and .01 for mixed lists, with none of them being significant.

Taken together, we can conclude that order information does not play a crucial role in the free recall of long study lists. Instead, alternative retrieval strategies are used. We assume that these alternative strategies are based on long-term relational knowledge that does not differ between encoding conditions (EPT–SPT). For this reason, recall differences for long study lists are probably due to better item information for SPTs than for EPTs.

**General Discussion**

The present experiments showed that the pattern of results for the enactment of short lists can be explained within the item–order hypothesis proposed by Nairne et al. (1991). The core assumptions of this framework are that (a) encoding conditions or item types may have opposing effects on item and order information, (b) differential effects on order information only emerge in pure study lists, and (c) memory tests draw to a different extent on item and order information. One specific assumption within this hypothesis concerns the encoding and use of serial order: It is assumed that participants encode information on the presentation order of items and subsequently use this information to guide retrieval during free recall (e.g., DeLosh & McDaniel, 1996; Serra & Naim, 1993). However, this assumption seems to be questionable (see also McDaniel et al., 1999, who proposed a modified order hypothesis).

In this study, we applied the item–order hypothesis to the comparison of EPTs and SPTs. We assumed that for pure study lists, SPTs provide better item information than EPTs, whereas EPTs provide better order information than SPTs. For mixed study lists, order information for EPTs is lowered to the level of the intervening SPTs. Item information, on the other hand, is uninfluenced by the list type.

These assumptions led to the following predictions for short study lists, all of which were supported by empirical evidence:

1. Independent of design type, recognition performance should be better for SPTs than for EPTs. A recognition test draws on item information, which is boosted by enactment and uninfluenced by design type.

2. Reconstruction performance should be better for EPTs than for SPTs in a between-subjects design, but there should be no difference in performance in a within-subjects design. A reconstruction test reflects order information that is superior for EPTs presented in pure study lists, but decreases to the level of SPTs if EPTs are presented in mixed study lists.

3. In free recall of mixed study lists, where order information for EPTs is lowered to the level of SPTs, the better item information for SPTs should produce an enactment effect. This enactment effect, however, was only statistically significant in Experiment 2.

4. In free recall of pure study lists, EPTs should benefit from good order information, which also provides the basis for their retrieval. The better order information for EPTs in pure lists should oppose the enactment effect observed in mixed lists. Corresponding to our expectation, there was a numerical EPT advantage in pure lists that approached significance in Experiment 1.

So far, the predictions are in line with those formulated, for example, for the generation effect (Nairne et al., 1991; Serra & Naim, 1993). However, the above predictions are restricted to the retention of short study lists. For long study lists, we expected that order encoding would be less efficient and, therefore, an order-based retrieval would be replaced by retrieval strategies that rely to a lesser extent on the presentation order. The effect of study list length on the use of order information has not been directly addressed in the literature so far (cf. DeLosh & McDaniel, 1996). To fill this gap, we also tested free recall, reconstruction, and recognition memory for long study lists. For long study lists, the following predictions were made:

1. Again, there should be a recognition advantage for SPTs over EPTs because item information is not changed by list length. In line with this, recognition performance was higher for SPTs than for EPTs in both design types.

2. Reconstruction performance is expected to be generally low because, for long study lists, encoding of serial order becomes highly demanding. Because of the low level of reconstruction performance, we only expected small differences between EPTs and SPTs. Corresponding to our expectation, reconstruction performance was fairly low and did not differ among the experimental conditions except for the first few positions of the lists.

3. For free recall of long lists, we assumed that retrieval strategies that are based on (relational) long-term knowledge structures are used. Because these knowledge structures differ little between SPTs and EPTs (Engelkamp et al., 1990; Kort et al., 1998), the availability of item information determines the pattern of results. In line with this, we found a design-independent recall advantage for SPTs. Relating the findings for order reconstruction to the ones for free recall supports the conclusion that free recall of long lists is scarcely based on order information.

For short study lists, the assumptions and findings formulated for the generation effect, the word-frequency effect, the bizarreness effect, the perceptual interference effect, and
now the enactment effect appear to be largely consistent. Even for short lists, however, there are subtle differences in both the hypotheses and the pattern of results for the various memory phenomena. These differences pertain to the availability of order information in mixed study lists and its effect on free recall.

The Effect of the Design Type on Order Information and on Free Recall

There are two theoretical positions concerning the effect of the design type on order information. According to one position (e.g., Serra & Nairne, 1993), the use of mixed study lists solely hampers order encoding for the condition that, if presented in pure study lists, supports the encoding of serial order. Corresponding to their expectations, Serra and Nairne obtained a semi-ordinal interaction of encoding condition and design type for order reconstruction. Our assumptions and findings also conform to this position. According to the second position (DeLosh & McDaniel, 1996), mixing study lists not only hampers order encoding for the conditions that normally support the encoding of serial order but also raises order encoding for the condition that does not naturally support the encoding of serial order. Corresponding to their expectation, DeLosh and McDaniel obtained an ordinal interaction of encoding condition and design type for order reconstruction.

For the free-recall data, an interaction between type of encoding and type of design was expected. The prediction of the exact pattern of results (i.e., whether an experimental condition that yielded inferior performance in mixed lists would rise to the same level as or even to a higher level than the other experimental condition in pure lists), though, required assumptions on the exact level of item and order information for the two experimental conditions. Again, there were slight differences in the expected pattern of results for the two contrasting views on order information: If the level of order information in the mixed encoding condition is assumed to lie in between the two levels for the separate encoding conditions (DeLosh & McDaniel, 1996), then changes in recall performance from mixed to pure lists should be more pronounced. If, in contrast, the level of order information in the mixed encoding condition is assumed to be equal to the level of order information for the inferior encoding condition (Serra & Nairne, 1993; the present study), then these changes should be less pronounced. This is due to the fact that, given a constant level of item information in both design types, a larger difference in order information for the two encoding conditions in pure lists can compensate to a greater extent for differences in item information between the two encoding conditions. Thus, the assumptions presented by DeLosh and McDaniel render it more likely that the effect under investigation (e.g., the word-frequency effect) would not only disappear from mixed to pure lists, but would even be reversed. This is exactly what the authors observed in their experiments.

A potential explanation for the different patterns of results concerns the ways of manipulating order and item information in the various studies. DeLosh and McDaniel (1996) manipulated item and order encoding by using different kinds of items, namely high- and low-frequency items. Thus, participants in each experimental condition learned different sets of items. Therefore, differences in encoding of item and order information might have been due to differences in the semantic content of items. In contrast, in the study by Serra and Nairne (1993) as well as in the present study, identical items were to be learned, albeit under different encoding instructions. Thus, item content does not differ between conditions.

The general problem of making exact predictions as to which encoding condition will yield higher performance in free recall of pure lists is even enhanced if differences in retrieval strategies are taken into account. Differences in retrieval strategies are indicated by differences in the probabilities with which item and order information, regardless of their availability, contribute to free recall. Consequently, a precise prediction on the free-recall pattern requires not only knowledge on the differences in item and order information but also on the probability with which one or the other type of information influences recall.

List Length as a Boundary Condition for the Item-Order Hypothesis

One goal of this study was to demonstrate that the order-encoding framework in its present form does not account for the pattern of results for long study lists but is confined to recall patterns for short study lists. Our interest in list length as a further moderating factor was fueled by the observation that the free-recall pattern for observed and performed actions changes with list length. For short lists (with up to 18 phrases), an interaction of type of encoding (EPT vs. SPT) and type of design (between-subjects vs. within-subjects) was observed, whereas for long lists a design-independent SPT advantage was found (Engelkamp & Zimmer, 1983, 1997; this study). The typical finding when comparing EPTs and SPTs is that, given a between-subjects design, there is no recall difference in SPTs and EPTs for short lists (up to 18 items), and there is a clear SPT advantage over EPTs with long lists (for an overview, see Engelkamp, 1998).

In our view, the crucial point is that strategies other than order-based retrieval strategies must be taken into account when long study lists are used. For long lists, the availability of order information is so restricted that order-based retrieval generally becomes inefficient. In Experiment 5 of this study as well as in Experiment 1 by Burns (1996), order reconstruction for lists of 24 and 40 items, respectively, was very low. The low level of order information in long lists renders order information as the basis of retrieval highly unlikely.

In the present study, we did not test directly which alternative retrieval strategies were used. However, we
assumed that the retrieval strategies used relied primarily on semantic long-term knowledge that supports the construction of idiosyncratic associations between study items. These associations are relatively independent of type of encoding, as subjective organization scores show (Engelkamp et al., 1990; Kornat et al., 1998). In the same way, there is evidence that the provision of categorical relational information replaces order information as a retrieval strategy for short lists (Nairne et al., 1991, Experiment 3; Mulligan, 1999, Experiment 6). In addition, the choice of a particular retrieval strategy has been shown to be influenced by the length of the retention interval (Burns, Curtis, & Lavin, 1993).

Conclusion

The present study provides another example for the successful application of the item–order hypothesis to a memory phenomenon. We were able to show that the distinction between item and order information adequately describes the pattern of results for the enactment effect, as investigated in short study lists. The fact that such different memory phenomena like the generation effect, the bizarreness effect, the perceptual interference effect, the word-frequency effect, and the enactment effect can all be explained within one theoretical framework points to the integrative power of the item–order hypothesis. In spite of the elegance of such a parsimonious explanation, however, we demonstrated that further refinements of the item–order hypothesis are necessary. The necessity for these refinements arises partly from the observation that there are subtle differences in the findings for the various memory effects. These differences mainly concern the influence of presenting study items in mixed rather than in pure lists on order information. The need for further specifications also arises from the observation that the scope of the item–order hypothesis is restricted to short study lists. For long study lists, other retrieval strategies than seriation can guide free recall. To accommodate all the empirical evidence, more fine-grained assumptions on the contribution of different sorts of relational information to performance in a memory test have to be formulated.

References


Received January 8, 1999
Revision received August 9, 1999
Accepted August 17, 1999