Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Memory and Language 50 (2004) 113–133
Journal of Memory and Language
www.elsevier.com/locate/jml

Separating item and order information through process dissociation

James S. Nairne[a,*] and Matthew R. Kelley[b]

[a] Department of Psychological Sciences, Purdue University, West Lafayette, IN 47907-1364, USA
[b] Department of Psychology, Lake Forest College, Lake Forest, IL 60045, USA

Received 22 January 2003; revision received 22 August 2003

Abstract

In the present paper, we develop and apply a technique, based on the logic of process dissociation, for obtaining numerical estimates of item and order information. Certain variables, such as phonological similarity, are widely believed to produce dissociative effects on item and order retention. However, such beliefs rest on the questionable assumption that item and order memory can be measured through performance on a particular kind of retention test (e.g., order can be measured through a reconstruction test). Retention measures are probably not process pure, but instead recruit multiple kinds of information and memory processing. Across four experiments, we show that our derived estimates conform generally to expected trends, although surprising results emerged in some instances. We discuss the implications of our analysis for traditional beliefs about the item–order distinction, and we show that a popular immediate memory model—the perturbation model—is capable of handling most of the empirical trends.

Keywords: Memory; Item and order memory; Process dissociation; Perturbation model; Serial recall

To build a successful general account of memory, it is important to distinguish between the recovery of item and order information (e.g., Healy, 1974; Murdock, 1976; Nairne, 1991). Item information refers to our ability to remember that a particular item occurred in an experimental context (i.e., by recalling or recognizing the item). Order information is measured by the successful reconstruction of an item's absolute or relative position in a sequence or array. For example, you might remember that the word CARROT occurred in a list without remembering either its absolute serial position or its position relative to another item (item information). Alternatively, you could remember that something was presented in a particular serial position, or after CARROT, but fail to remember its identity (order information).

Traditional retention tests typically tap both kinds of information. Serial recall, for instance, requires memory for items as well as their respective positions of occurrence. Recognition is purportedly a test of item information, although people may use order information to help decide whether an item occurred (e.g., HORSE must be on the list because I remember it occurring after CARROT). Free recall, at least nominally, does not demand order memory, but people often rely on serial order as an output strategy (e.g., DeLosh & McDaniel, 1996; Nairne, Riegler, & Serra, 1991; Postman, 1972). Finally, even reconstruction tests, which are often seen as pure tests of order (i.e., the items are returned and the task is simply to order them correctly), show considerable sensitivity to individual item characteristics (e.g., concreteness, set size, and modality; Neath, 1997).

The fact that memory tests are not "process pure," but instead recruit multiple mnemonic processes, is

* Corresponding author. Fax: 1-765-496-1264.
E-mail address: nairne@psych.purdue.edu (J.S. Nairne).

0749-596X/$ - see front matter
doi:10.1016/j.jml.2003.09.005

probably not surprising to most researchers (Dunn & Kirsner, 1989; Jacoby, 1991). However, significant problems remain for those interested in building a general account of mnemonic phenomena. To model the processes (or process) that are responsible for a given type of memory, such as item or order information, it is important to understand how empirical variables selectively influence each kind of memory. Yet, because of the process purity problem, these influences cannot be measured directly; instead, we can only hope that a particular retention test provides a reasonably accurate estimate of a given information type. Thus, recognition is seen as a "relatively" pure test of item information and reconstruction as a "relatively" pure test of order (Nairne & Neumann, 1993; Neath, 1997). What would be preferable, of course, is a more direct, or uncontaminated, technique for measuring information—a technique that does not hinge on an association between tasks and processes.

In the present paper, we provide one solution to the test purity problem by applying the process dissociation framework. We introduce a procedure for obtaining estimates of item and order information, and we then begin an exploration of how relevant independent variables (e.g., similarity, word frequency, and generation) affect those estimates. In addition, partly because any application of the process dissociation framework relies on assumptions that are model-dependent (Buchner, Erdfelder, & Vaterrodt-Plünnecke, 1995; Curran & Hintzman, 1995; Jacoby, 1998; Ratcliff, Van Zandt, & McKoon, 1995), we frame our results in terms of a particular memory model—the perturbation model of Estes (e.g., Lee & Estes, 1981; Nairne, 1991). As traditionally employed, the perturbation model assumes that item and order memories are controlled by separate, and independent, mnemonic processes. Our intent is not to provide quantitative fits to data, but merely to show how a particular process model, one assuming independence, might handle the basic data patterns.

Process dissociation and the item/order distinction

The process dissociation procedure was originally designed to separate automatic and conscious contributions to retention task performance (e.g., Jacoby, 1991, 1998). It assumes that both conscious and automatic processes contribute to performance and operate independently. Estimates of each can be obtained by comparing performance across two conditions—one in which both processes are expected to facilitate performance (called an Inclusion condition) and one in which the processes act in opposition (the Exclusion condition). Suppose you are asked to study a list of words and at test you are given word stems to complete (mot__ for motel). In one condition—the Inclusion condition—you are told to use the stem as a cue to recall an old word or to complete the stem with the first word that comes to mind. In such a case, "motel" could be given as a stem response because it is consciously recollected as an earlier study item (with probability C), or because it automatically comes to mind (with probability A). Assuming independence, the probability that a studied word will be produced is therefore given by: $C + A - CA$.

Now consider a second condition in which you are told never to complete a stem with one of the studied words. In this "Exclusion" condition, "motel" can still be given as an answer but it would be incorrect, and presumably would occur only if you fail to consciously recollect it as one of the earlier study items. Thus, the probability of producing "motel" in this condition can be expressed as: $A(1 - C)$. Calculating the difference between performance in the Inclusion and Exclusion conditions then provides an estimate of C ($C = \text{Inclusion} - \text{Exclusion}$); similarly, one can obtain an estimate of the automatic component in the following way: $A = \text{Exclusion}/(1 - C)$. Jacoby and others have shown that these derived estimates track experimental manipulations in consistent ways. For example, dividing attention at study seems to have a selective effect on C whereas lexicality differences influence A (see Jacoby, Yonelinas, & Jennings, 1997, for a review).

To apply the process dissociation logic to the item/order distinction, it is necessary to craft comparable Inclusion and Exclusion conditions. Consider immediate serial recall: To recall an item in its correct position requires memory for the item as well as for its position in the sequence. Assuming independence, this should be equal to the probability of remembering the item (I) multiplied by the probability that its ordered position is remembered (O_R). In this case, which corresponds to an Inclusion condition, increases in either item or order memory should facilitate performance. Now consider an "Exclusion" condition with the following instructions: Recall all of the items from the just-presented list *except* the item that occurred in a particular position X. In this case, you will only recall the item from position X if you remember the item (with probability I) but fail to remember its ordered position correctly (with probability $1 - O_R$). We now have two conditions in which the probability of recalling the item from position X can be measured:

$$\text{Inclusion} = IO_R,$$

$$\text{Exclusion} = I(1 - O_R).$$

Through simple algebra, we can then solve for I, which turns out to be the sum of performances in the Inclusion and Exclusion conditions; O_R can be calculated by simply dividing Inclusion performance by I.

$I = \text{Inclusion} + \text{Exclusion},$

$O_R = \text{Inclusion}/I.$

In this way, by comparing observable performance across these Inclusion and Exclusion conditions, estimates of item and order information can be obtained that do not suffer from the test purity problem.[1]

The independence assumption and perturbation theory

Of course, as noted earlier, any application of process dissociation rests on assumptions that are model dependent (Curran & Hintzman, 1995; Jacoby, 1998; Ratcliff et al., 1995). For example, we have assumed that memory for an item's occurrence on a particular trial (item retention) is independent of memory for its serial position within the trial (order retention). We have assumed as well that the quantities I and O_R remain the same across the Inclusion and Exclusion conditions—that is, the contribution of each quantity does not depend on the type of retention test employed. Violations of these assumptions could produce distorted estimates of I and O_R (see Curran & Hintzman, 1995).

The independence assumption is actually well-grounded in the short-term memory literature. Researchers have commonly assumed independence, in part, because item and order retention are easy to dissociate empirically. For example, similarity (Crowder, 1979), category membership (Murdock, 1976), word frequency (DeLosh & McDaniel, 1996; Whiteman, Nairne, & Serra, 1994), and generation (Nairne et al., 1991) all produce dissociative effects (e.g., generating a word from a fragment improves overall recognition memory for the item but hurts one's memory for its particular sequential position in the list; Nairne et al., 1991). In addition, tests which purportedly measure item and order memory can produce different serial position curves: Item functions tend to be flat, at least for very short lists, whereas order functions tend to be more bow-shaped (Drewnowski, 1980; Healy, 1974). There is also evidence suggesting that item and order information may be forgotten at different rates (e.g., Bjork & Healy, 1974).

Unfortunately, by themselves, such dissociations cannot establish independence in any kind of definitive way. In fact, questions about the viability of the independence assumption may not be solvable through any kind of simple empirical analysis. Part of the problem is that task dissociations (even cross-over interactions) can be mimicked by single process models (e.g., Hirshman & Master, 1997; Ratcliff et al., 1995); moreover, independent processes can produce data that look dependent, and dependent processes can lead to data that look independent (see Hintzman, 1990). There appear to be conceptual problems with the independence assumption as well. For example, it seems reasonable to assume that we can remember an item without remembering its corresponding position, but how can the converse be true? In what sense is it possible to remember the position of an item without remembering the item itself?

Some of these concerns can be allayed by considering how independence works in an existing model of immediate retention, the perturbation model of Estes (1997; Lee & Estes, 1981). In the perturbation model, items are encoded in terms of their positions along list and within list dimensions; e.g., an item might be coded as having occurred in the third serial position in the fourth list of the session. These representations are then assumed to drift or "perturb," along the encoding dimensions over time (see Estes, 1997; Nairne, 1991 for details). An order error is created when an item drifts, with probability θ_1, into an adjacent position along the within list dimension—e.g., an item that was presented originally in position 3 is now remembered as having occurred in positions 2 or 4. Item errors are created when drifting occurs along the list dimension. There is a second probability, θ_2, that an item will perturb into a different list, leading to an omission—or item—error at the point of recall. In this way, then, it is possible to retain accurate memory for an item's within list position (order information), but fail to recall the item itself (item information)—the loss of item, but not order, information occurs because the item is remembered as having occurred on a different trial.

To perform correctly in a ordered recall context, it is necessary to remember both the item and its within list position of occurrence. The probability that an item will be remembered as having occurred in a particular list at time $n+1$ can be expressed as follows:

$$X_{i,n+1} = [1 - (\theta_2/2)]X_{i,n} + (\theta_2/2)X_{i-1,n}.$$

In words, the item will be remembered as having occurred on the current trial if the item was already represented correctly at time n and no perturbation occurred in the interval $n+1$, or if the item was represented incorrectly as having occurred on the immediately preceding trial and a perturbation did, in fact, occur. Note that the perturbation rate is set as $\theta_2/2$ in this case because items are unable to perturb into the

[1] For clarification, these equations apply to the retention of a particular item in the list. So, *Inclusion* is measured by noting the observable retention of the item from position X in standard serial recall and *Exclusion* is measured by noting how often the item is incorrectly recalled when it is the to-be-excluded item in the Exclusion condition. Similarly, I and O_R are quantities specific to that particular item. Over trials, as each list item is sampled in an Exclusion condition, one can calculate I and O_R for each serial position in a list.

future; an item can only drift backward into preceding trials.[2] Movement along the within list dimension is represented in a similar way. The probability that an item will be remembered as occupying a particular interior position i at time $n+1$ is the following:

$$X_{i,n+1} = (1-\theta_1)X_{i,n} + (\theta_1/2)X_{i-1,n} + (\theta_1/2)X_{i+1,n},$$

where θ_1 is the probability that a within list perturbation will occur. Slightly different equations govern memory for items in the first and last list positions because items in these position can only perturb in an inward direction—e.g.,

$$X_{1,n+1} = [1-(\theta_1/2)]X_{1,n} + (\theta_1/2)X_{2,n}.$$

These boundary conditions for the primacy and recency items lead to bow-shaped serial position curves of the type often found in studies of ordered recall.

Assuming independence, performance in an Inclusion condition—that is, serial recall—will simply be equal to the product of the respective list and within list probabilities (see Lee & Estes, 1981). We can also calculate the probability that an item will be placed in the current list, but remembered as having occurred in an incorrect serial position, for any experimental situation. Thus, it should be possible to fit data generated empirically with the model for both an Inclusion and Exclusion condition by varying the two main item and order parameters, θ_1 and θ_2. The perturbation model has been successfully fit to a variety of empirical settings, but not to the manipulations employed in the present experiments.

The present experiments

In this paper, we report the results of four experiments exploring how various independent variables affect the estimates of I and O_R derived through process dissociation. Each manipulation is historically relevant to the item–order distinction, and our initial goal was to check whether the derived estimates of I and O_R would show expected trends. For example, in Experiment 1 we manipulated phonological similarity across Inclusion and Exclusion conditions. Phonological similarity is widely believed to impair order retention, as evidenced by serial recall and reconstruction performance, but to increase retention for item information (e.g., Watkins, Watkins, & Crowder, 1974). Experiment 2 investigates semantic similarity which is also assumed to produce dissociative effects (e.g., Crowder, 1979; Murdock, 1976; although see Saint-Aubin & Poirier, 1999). Experiments 3 and 4 investigate word frequency and generation, respectively. In each case, once again, the goal was to see how a given independent variable affected the estimates of I and O_R. Finally, to extend its generality, we show that the perturbation model provides reasonable qualitative fits to the data as well.

Experiment 1

In immediate serial recall, it is difficult to perform well when memory lists are composed of similar-sounding items (e.g., Baddeley, 1966; Conrad, 1964). This phonological similarity effect is generally viewed as a benchmark finding in the immediate memory literature, and virtually all models of serial order include mechanisms to account for the effect (e.g., Baddeley, 1986; Burgess & Hitch, 1999; Henson, 1998; Nairne, 1990; Page & Norris, 1998). Empirically, though, the detrimental effect tends to be restricted to ordered recall; when the task is free recall, rather than ordered recall, phonological similarity can lead to a beneficial effect (Watkins et al., 1974; Wickelgren, 1965). As a result, many researchers have concluded that phonological similarity has opposite effects on item and order retention, improving item memory and impairing order memory.

Of course, this conclusion rests on the dubious link between task and process: Neither serial recall nor free recall is process pure with respect to item and order information. One solution is to restrict attention only to serial recall and conditionalize the data in an attempt to obtain separate estimates of item and order memory. For example, one can compute the number of order errors per item recalled, or calculate the proportion of items recalled irrespective of recall position (see Murdock, 1976; Saint-Aubin & Poirier, 1999; Wickelgren, 1965). Here, too, however, the measures are subject to the idiosyncrasies of one task, serial recall, so task-dependent strategies probably influence the derived measures. There is, in fact, a direct connection between conditionalizing and the process dissociation procedure employed here (both are based on the same underlying model); we consider this relationship in more detail in the general discussion.

In Experiment 1, we obtain estimates of item and order memory by comparing across two tasks that are acknowledged to rely on item and order retention. The Inclusion condition corresponds to traditional serial recall—subjects were asked to recall list items in their original order of presentation; in the Exclusion condition, subjects were asked to recall all the items from a list *except* the item that occurred in a particular serial position. Of main interest here are the instances in which the subjects continue to recall the to-be-excluded item,

[2] If one tests retention after multiple lists, as in Nairne (1991), items can potentially drift into either the preceding or following list (with probability θ_2)—in the present context, however, it is only possible for an item to drift into the immediately preceding list (with probability $\theta_2/2$).

reflecting item memory but impaired order memory. By comparing across these two conditions, as outlined earlier in the introduction, it should be possible to obtain estimates of I and O_R that are uncontaminated by the test purity problem.

Method

Subjects and apparatus

Sixty Purdue undergraduates participated for course credit. Subjects were tested individually and all stimulus materials were presented and controlled by personal computers.

Materials and design

Sixty lists of five rhyming nouns, medium to high frequency, were constructed from sets reported by Libkuman (1994). The lists were divided into two sets of 30, serving as the similar and dissimilar lists across participants. The dissimilar lists were formed by simply recombining words from similar lists in the following manner: For a given set, the 30 similar lists were divided into six groups of five lists. Within each group, five words, one from each list, were selected and recombined, creating a list of dissimilar sounding items. By forming dissimilar lists in this fashion, we ensured that everyone would receive the same set of 300 words in the experiment. The words were simply grouped differently, serving in similar lists for one group of participants and in dissimilar lists for another.

Inclusion versus Exclusion instructions were manipulated within-subjects, but in a blocked format. Subjects received a block of 30 trials with Inclusion instructions and a block of 30 trials with Exclusion instructions; presentation order was counterbalanced across subjects. Within each block, 15 of the lists contained similar-sounding items and 15 lists contained dissimilar items. The order of the similar and dissimilar lists was randomly determined within a block, as was the ordering of the words within each list.

Procedure

Each trial began with the word READY accompanied by a tone, followed by the presentation of a five item list. Each list item was presented for 750 ms with a 250 ms interval separating the offset of one item from the onset of the next. Participants were instructed to say each item aloud as it appeared on the screen. Immediately after the last item, there was a short retention interval during which subjects were asked to complete ten simple addition problems (e.g., $1 + 4 = ?$). After the tenth problem was completed correctly, recall began. The delay was introduced in an effort to reduce overall performance, given that the lists were relatively short and performance approached the ceiling in pilot work using immediate serial recall.

For the *Inclusion* condition, standard serial recall instructions were given. Subjects were told to recall the items in order, from first to last, by typing each word into ordered boxes highlighted on the CRT screen. Specifically, five boxes were displayed on the screen. At the beginning of the recall period, the first box was highlighted and subjects responded by typing in the first word on the list; next, the second box was highlighted, and so on. If a particular item could not be remembered, subjects were told to type an "x" in the highlighted box. For the *Exclusion* condition, subjects were told to recall all of the items except for the item that occurred in a designated serial position (e.g., the second list position). Free recall was allowed—that is, subjects could recall the items in any order—and subjects were asked to type an "x" for any unrecalled items (including the to-be-excluded item). To prevent selective encoding, information about the to-be-excluded item was presented after the last addition problem, just prior to the initiation of recall. Across trials, the to-be-excluded item was sampled equally often from each of the five serial positions. In all other respects the procedure mimicked the one used in the Inclusion condition.

Results and discussion

Of main interest is performance in the Inclusion and Exclusion conditions as a function of similarity. Turning first to the Inclusion data, which are shown on the upper-left side of Fig. 1, we expected performance to be better for the dissimilar lists, reflecting the standard phonological similarity effect. An overall analysis of variance (ANOVA) on the data confirmed our expectations. Significant effects were found for similarity [$F(1, 59) = 28.95$; $MSE = .04$, $p < .001$] and for serial position [$F(4, 236) = 65.79$; $MSE = .03$, $p < .001$]; the interaction was also significant [$F(4, 236) = 11.83$, $MSE = .01$; $p < .001$]. As shown, the serial position curves were generally bow-shaped and the size of the similarity effect increased over serial position. Comparable patterns are typically found in immediate serial recall, suggesting that our addition of a filled delay (to lower performance) did not change the effects of similarity in any fundamental way.

Performance in the Exclusion condition is shown on the upper-right side of Fig. 1. These data depict Exclusion errors—that is, each data point represents the proportion of times that a to-be-excluded item was actually recalled, contrary to the instructions, for a given serial position and condition. (On a given trial, of course, only one item was to-be-excluded, so the curve shown in Fig. 1 comes from collapsing across trials.) Performance in this case was assumed to reflect primarily the loss of order information, so we expected more Exclusion errors for similar lists. The ANOVA confirmed these expectations: There was a significant main effect of

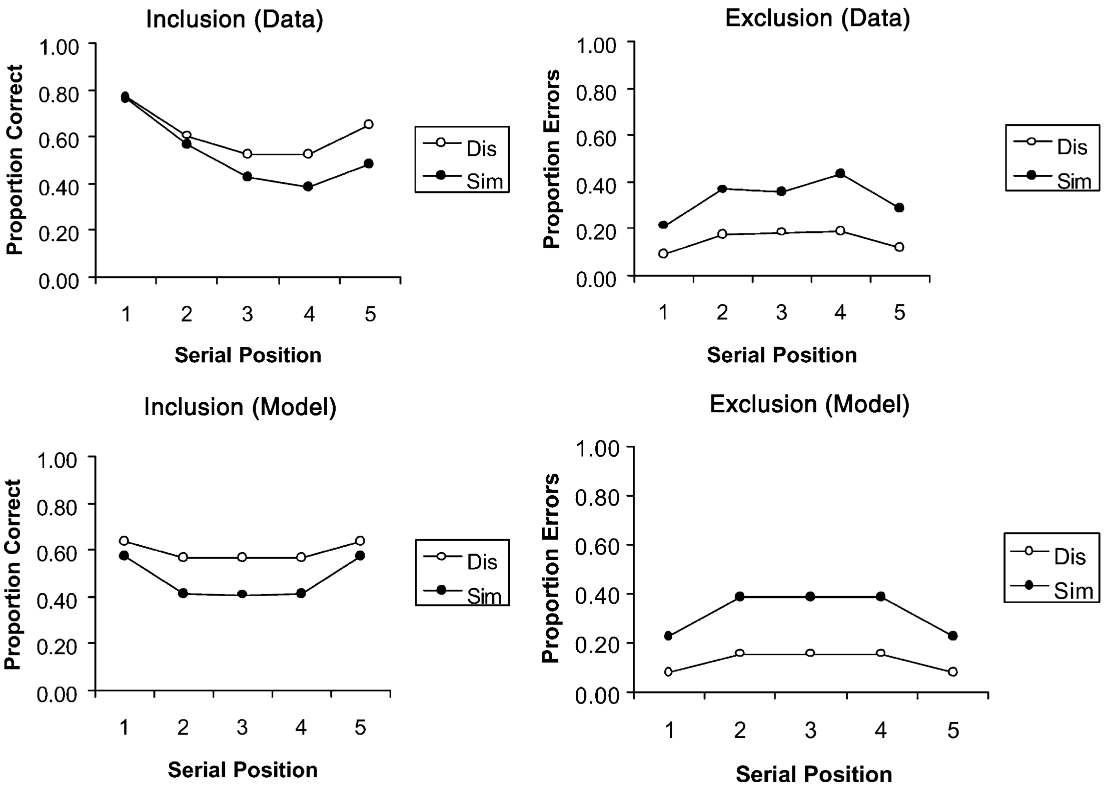

Fig. 1. Data from the Inclusion and Exclusion conditions, plotted as a function of phonological similarity condition and serial position. The lower two panels show data generated from an application of the perturbation model.

similarity [$F(1,59) = 81.00$; $MSE = .06$; $p < .001$] as well as a significant effect of serial position [$F(4,236) = 7.58$, $p < .001$]; the interaction was not significant [$F(4,236) < 1$]. The serial position functions showed an inverted form—fewer exclusion errors occurred for the first and last items in the list.

We also examined the individual error patterns for both instruction conditions. These data are shown in

Table 1
Mean proportion of errors, collapsed across serial position, for similar and dissimilar lists in the Inclusion and Exclusion conditions of Experiment 1

Error type	Condition	Similar	Dissimilar
Transposition	Inclusion	.26	.10
	Exclusion	.30	.16
	All	.28	.13
Omissions	Inclusion	.23	.31
	Exclusion	.18	.32
	All	.20	.31
Intrusions—prior lists	Inclusion	.01	.05
	Exclusion	.01	.04
	All	.01	.04
Intrusions—out of experiment	Inclusion	.05	.06
	Exclusion	.07	.08
	All	.06	.07
Repetitions	Inclusion	.01	.00
	Exclusion	.00	.00
	All	.00	.00

Table 1 and show consistency across conditions. A transposition error was recorded when a list item was recalled, but was placed in new serial position during output (e.g., the third item on the list was placed in the second serial position at test). Because people were free to recall items in any order in the Exclusion condition, there were no transposition "errors" in a technical sense. However, we felt it was still informative to record how often people recalled items in their original serial positions, and to compare this tendency across the similar and dissimilar conditions. As expected, more transposition errors occurred for similar lists, for both the Inclusion and Exclusion conditions. Presumably, at least for the Inclusion data, this reflects a greater loss of order information for the similar condition.

A quite different pattern emerged for omission errors. An omission error occurred when a list item was omitted—that is, not recalled in any serial position. In this case, fewer omission errors were found when lists were composed of similar items (again, for both instruction conditions). This cross-over pattern—more transposition errors and fewer omission errors for similar lists—suggests that similarity has dissociative effects on item and order memory, improving the former and impairing the latter (for a similar pattern, see Coltheart, 1993). Table 1 also shows the percentage of trials in which nonlist items were intruded, either from a prior list in the session or from outside of the experimental pool. Intrusion rates were low, but more intrusions tended to occur during the recall of dissimilar lists. This is a potentially important trend, for reasons that will be discussed in our application of the perturbation model.

Estimates of I and O_R

As noted in the introduction, our main goal was to derive uncontaminated estimates of item and order information by comparing performance across the Inclusion and Exclusion conditions. The equations introduced earlier were applied to the data for each individual subject, giving estimates of I and O_R for both list types. Average values are shown in Fig. 2, collapsed across serial position. Based on conclusions widely held in the literature, and on the error data described above, we expected to find that phonological similarity enhanced I and impaired O_R. This pattern is reflected in the mean data and was supported by an ANOVA. The analysis revealed significant main effects for I versus O_R [$F(1,59) = 8.01$; $MSE = .03$; $p < .01$] and similarity [$F(1,59) = 19.91$; $MSE = .01$; $p < .001$]; the interaction was significant as well [$F(1,59) = 47.20$; $MSE = .02$; $p < .001$]. Subsequent planned analyses showed that the differences between similar and dissimilar conditions were significant for both the estimates of I [$F(1,59) = 11.81$; $MSE = .02$; $p < .01$] and O_R [$F(1,59) = 97.91$; $MSE = .01$; $p < .001$]. As anticipated, these estimates confirm that phonological similarity has dissociative effects on item and order information. Similarity increases the available item information, making it easier to remember that a particular item occurred on a trial, but impairs the retention of order information.

Application of the perturbation model

The derived estimates, along with the error data, present a consistent picture, but it is still useful to see how well the basic data can be fit by an established memory model assuming independence between item and order information. To fit a model to data, of course, does not mean that the particulars of the model are correct, or establish independence, but it does show that the model's assumptions are capable of handling the data. In the present case, we decided against seeking quantitative fits because we wanted to minimize the

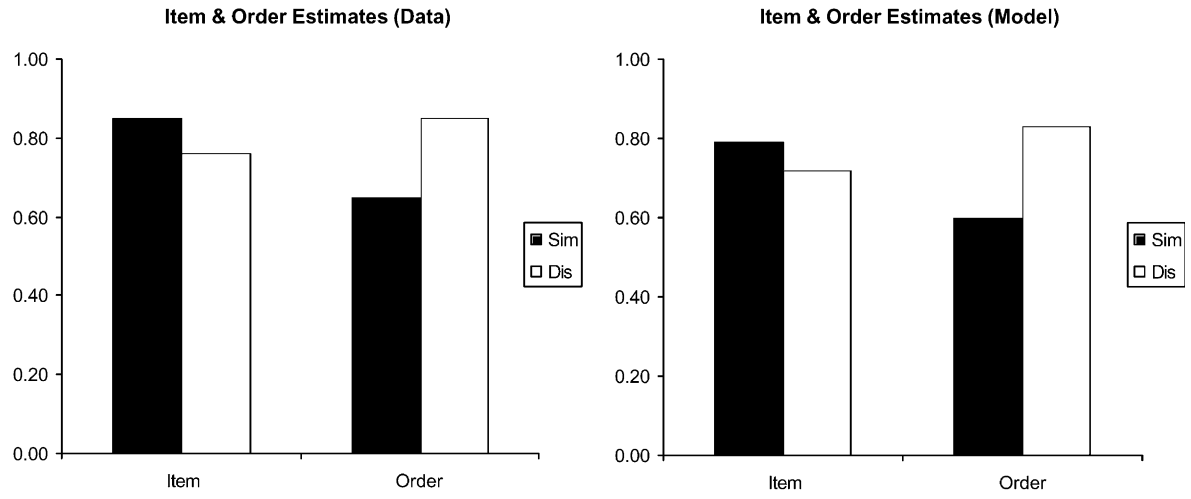

Fig. 2. Average Item and Order estimates generated from the process dissociation application in Experiment 1. Estimates in the right panel are derived from data generated by the perturbation model.

complexity of the model (particularly the number of model parameters). As will be clear momentarily, a bare-bones version of the perturbation model is capable of mimicking the empirical dissociations shown in Experiment 1 reasonably well.

To apply the version of the perturbation model outlined in the introduction requires the setting of two parameters in each of the relevant conditions: the within-list perturbation probability, θ_1, and the across-list perturbation probability, θ_2. Based on the transposition error patterns shown in Table 1, the within-list perturbation probability was assumed to be *higher* for similar lists—that is, similarity was simply assumed to increase the likelihood of drifting along the within list dimension (θ_1 was set at 0.15 for similar lists and 0.05 for dissimilar lists). This assumption cannot be used to explain the effects of phonological similarity in general—e.g., when similar items occur only in alternating list positions, order errors do not increase substantially for the non-confusable items (see Page & Norris, 1998). However, it does an adequate job in the present context and is used here for convenience. The across-list perturbation probability, which drives omission errors in the model, was set *lower* for similar lists (0.10 for similar lists and 0.15 for dissimilar lists). Again, this assumption is justified by the error data of Experiment 1, which show fewer omission errors for the similar lists. Interestingly, Nairne and Kelley (1999; see also Nairne & Neumann, 1993) used similar reasoning to explain how phonological similarity can, at times, lead to improvements in the reconstruction of order (as opposed to the more typical decrement). When different items are used on every list, as in Experiment 1, similarity can make it more difficult to retain an item's position within the list (because every item on a given list sounds similar), but those same unique rhyming qualities can make it easier to identify the item's proper list representation in memory (e.g., the list with words ending in the -ock sound). Whether overall performance will be enhanced or impaired by similarity, then, depends on a trade-off between these two types of memory: Similarity helps in discriminating one list from another, but hurts in discriminating within-list position.

It is also necessary to establish a rule for determining the number of perturbation opportunities, which we set equal to the number of items in the list. Thus, each item was assumed to be represented accurately—in its proper position—at the end of list presentation, but five within-list and across-list perturbation opportunities were allowed to occur during the retention interval. At the point of test, proportion correct in the Inclusion condition was computed by taking the product of the probabilities that an item was remembered in its correct within-list and across-list position. Performance in the Exclusion condition was determined in a similar fashion, although in this case we calculated the probabilities that an item occurred in the list but was not remembered in its correct serial position. [Because we were interested only in determining whether the model was capable of handling the general patterns found in Experiment 1, we made no attempt to achieve a "best fit" of the model.]

The relevant data from the model are shown in the bottom half of Fig. 1. The model produces a strong effect of similarity, in both the Inclusion and Exclusion conditions, and the serial position curves show a symmetrical bow-shaped form. The subject data in the Inclusion condition show significantly more primacy, and are not symmetrical in form, but this is likely due to output interference (or decay) which we did not model for the sake of simplicity.[3] The estimates of I and O_R derived from the model's data are shown in Fig. 2. Most importantly, the critical cross-over interaction is apparent in the estimates: Similarity led to improved item memory and impaired order memory. The effect of similarity was even greater for O_R which was found in the estimates for the results of Experiment 1 as well. Thus, the perturbation model, which assumes independent mechanisms for item and order retention, is capable of reproducing the important qualitative interaction found in Experiment 1.

Experiment 2

Experiment 1 explored the effects of phonological similarity on the retention of item and order information. Historically, phonological similarity has produced consistent results—the phonological similarity effect—but other manipulations of similarity, particularly semantic similarity, have produced equivocal results. Researchers have traditionally assumed that semantic similarity enhances item information and impairs order, just like phonological similarity, although it is widely recognized that its effect may be harder to detect in a sound-based short-term memory environment (Crowder, 1979; Murdock, 1976). Murdock and vom Saal (1967) showed that lists composed of items from the same semantic category are easier to recall (when scored without respect for order) but produce more item-to-item transpositions; Saint-Aubin and Poirier (1999) replicated the similarity advantage in item recall, but found no order decrement (see also Murdock, 1976). Crowder (1979) reported that semantic similarity impairs performance in a immediate reconstruction of order test, but both Nairne and Neumann (1993) and Saint-Aubin and Poirier (1999) found equivalent performance between similar and dissimilar lists.

[3] One could potentially mimic the observed data more closely by assuming that perturbations occur as well during output (see Estes, 1997).

Questions about whether semantic similarity enhances or impairs memory are important because some models of order predict impairment (contingent on the assumption that semantic features are encoded and used in trace reconstruction; e.g., Nairne, 1990). Semantic similarity should make it easier for a subject to discriminate the list in which an item occurred (e.g., the list containing types of weapons rather than furniture), but within-list discriminability should be impaired when items contain overlapping features. In their recent review of the literature, Saint-Aubin and Poirier (1999) concluded that semantic similarity does indeed enhance item memory, but has little, or no, effect on the retention of order information.

Experiment 2 is essentially a replication of Experiment 1, except that semantic rather than phonological similarity was manipulated. Subjects received lists containing words drawn from the same or different semantic categories. Both Inclusion and Exclusion conditions were included, and estimates for both I and O_R were calculated based on the assumptions of process dissociation.

Method

Subjects and apparatus

Sixty Purdue undergraduates participated in exchange for course credit. Subjects were tested individually and all materials were presented and controlled by personal computers.

Materials and design

Experiment 2 mimicked Experiment 1 in all respects except for the stimulus materials. Sixty lists were constructed based on semantic rather than phonological similarity. Each list contained five items drawn from the same semantic category (e.g., fish, colors, fruits, etc.); the materials were constructed from norms published by Hunt and Hodge (1971). As in Experiment 1, each word participated in both similar and dissimilar lists across subjects. Inclusion and Exclusion instructions were again manipulated within-subjects in a blocked, but counterbalanced, fashion.

Procedure

The procedure for Experiment 2 was the same as the one employed for Experiment 1.

Results and discussion

The data from the Inclusion and Exclusion conditions are shown in Fig. 3, once again as a function of serial position. Turning first to the Inclusion data, shown on the upper-left side of the figure, an overall ANOVA revealed highly significant effects of similarity [$F(1, 59) = 76.10$; $MSE = .03$; $p < .001$] and serial position [$F(4, 236) = 90.70$; $MSE = .02$; $p < .001$]. The serial position curves were generally bowed-shaped, as expected, and there was a consistent recall advantage for the similar lists. The fact that there was no similarity decrement contrasts sharply with the results of Experiment 1, but other researchers have reported the same similarity advantage in immediate serial recall (see Saint-Aubin & Poirier, 1999); as we will discuss shortly, the advantage is usually attributed to enhanced item memory. Note the interaction of similarity by serial position was significant as well [$F(4, 236) = 4.80$; $MSE = .01$; $p < .001$], although no obvious pattern is apparent in the mean data.

The results from the Exclusion condition are shown on the upper-right side of Fig. 3. A separate ANOVA on these data revealed a significant effect of serial position [$F(4, 236) = 4.64$; $MSE = .05$; $p < .001$], matching the general inverted form of Experiment 1, but the effect of similarity did not quite meet conventional levels of significance [$F(1, 59) = 3.00$; $MSE = .04$; $p < .08$]. Similarity did interact significantly with serial position [$F(4, 236) = 4.29$; $MSE = .05$; $p < .01$], and the interaction appeared to be due to a similarity advantage existing primarily at the second serial position. Compared to Experiment 1, where phonological similarity produced consistently more Exclusion errors, semantic similarity produced only a slight, and inconsistent, effect on performance.

The individual error data are shown in Table 2 and, once again, similar patterns were found for the Inclusion and Exclusion conditions. Slightly more transposition errors occurred in the similar lists, although the differences were small compared to the patterns found for phonological similarity in Experiment 1. However, large differences were again found for omission errors: Subjects were much more likely to omit an item from the dissimilar list. This suggests that similarity may indeed enhance memory for the individual items that occurred on a trial. There was also a slightly greater tendency to intrude items from prior lists, or from outside of the experiment, when lists were composed of dissimilar items.

Estimates of I and O_R

The average estimates for I and O_R, derived from process dissociation, are shown in Fig. 4. An overall ANOVA, collapsed across serial position, revealed significant effects of I versus O_R [$F(1, 59) = 6.29$; $MSE = .075$; $p < .02$] and similarity [$F(1, 59) = 80.80$; $MSE = .004$; $p < .001$]; the interaction was significant as well [$F(1, 59) = 23.93$; $MSE = .015$; $p < .001$]. Of main interest are the individual similarity comparisons for I and O_R. Planned comparisons revealed a significant difference between similar and dissimilar lists for the estimates of I [$F(1, 59) = 52.49$; $MSE = .07$; $p < .001$], but not for O_R [$F(1, 59) < 1$]. This empirical pattern is consistent

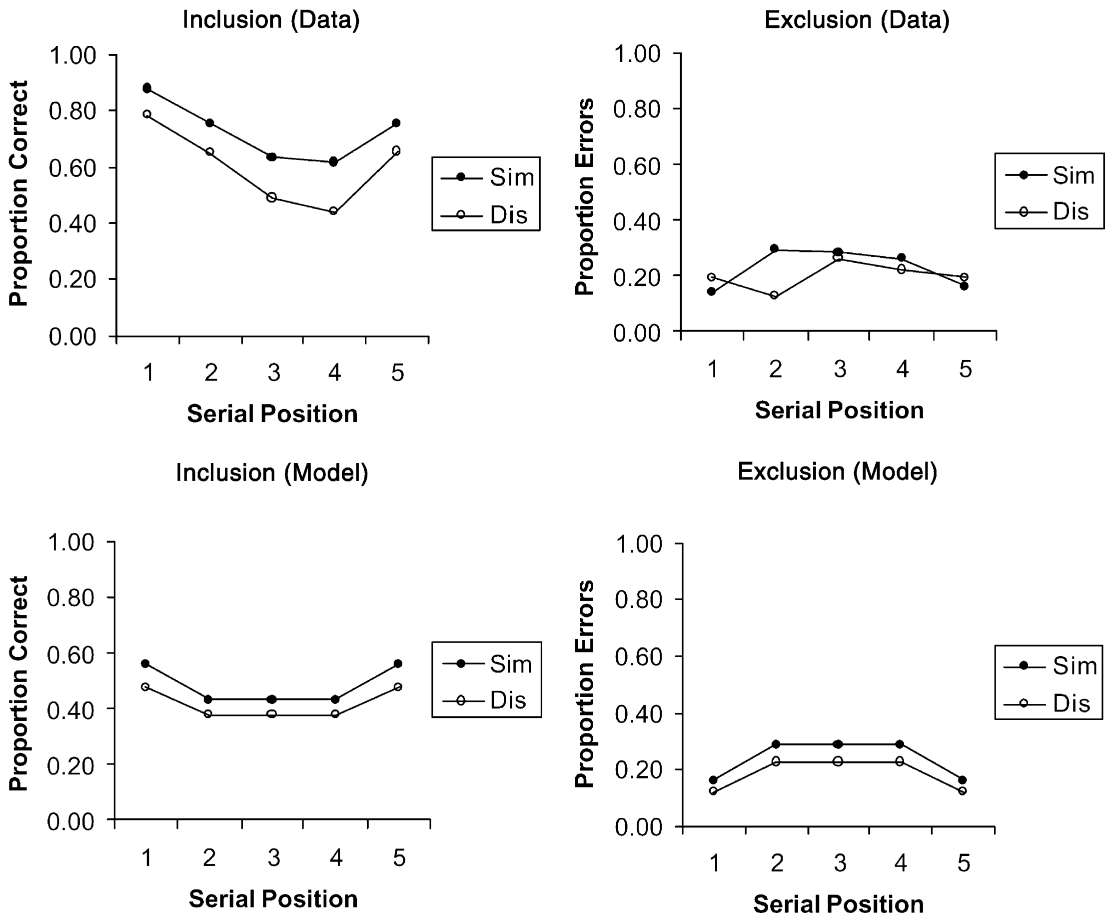

Fig. 3. Data from the Inclusion and Exclusion conditions, plotted as a function of semantic similarity condition and serial position. The lower two panels show data generated from an application of the perturbation model.

Table 2
Mean proportion of errors, collapsed across serial position, for similar and dissimilar lists in the Inclusion and Exclusion conditions of Experiment 2

Error type	Condition	Similar	Dissimilar
Transposition	Inclusion	.15	.11
	Exclusion	.24	.20
	All	.20	.15
Omissions	Inclusion	.13	.30
	Exclusion	.11	.28
	All	.12	.29
Intrusions—prior lists	Inclusion	.00	.02
	Exclusion	.00	.02
	All	.00	.02
Intrusions—out of experiment	Inclusion	.03	.03
	Exclusion	.03	.05
	All	.03	.04
Repetitions	Inclusion	.00	.00
	Exclusion	.00	.00
	All	.00	.00

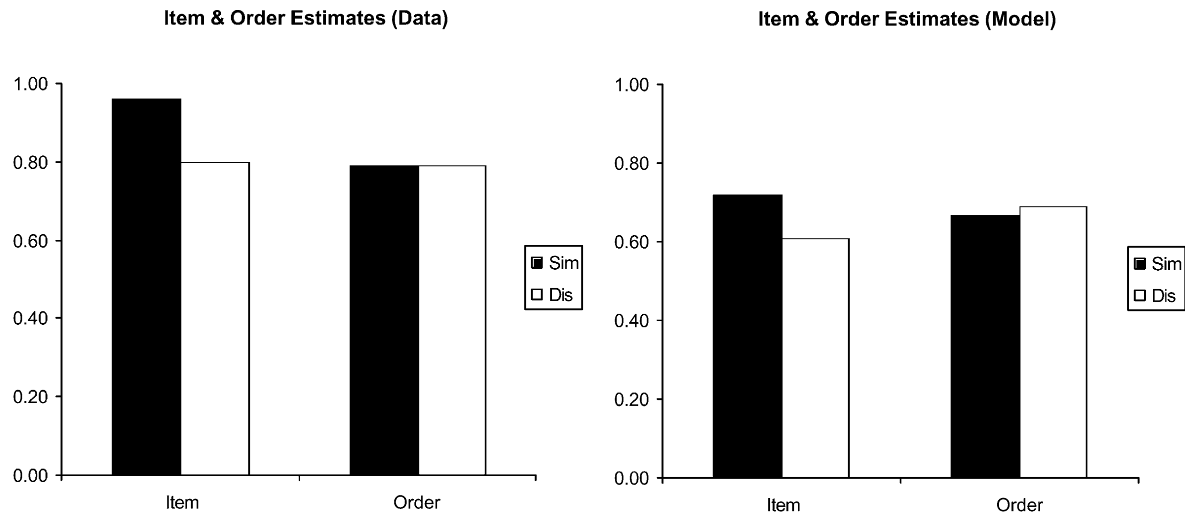

Fig. 4. Average Item and Order estimates generated from the process dissociation application in Experiment 2. Estimates in the right panel are derived from data generated by the perturbation model.

with Saint-Aubin and Poirier's (1999) proposal that semantic similarity produces a robust effect on item retention, leading presumably to the overall advantage in serial recall, but has no significant effect on order. The item advantage found for the similar lists replicates the pattern found in Experiment 1 but, unlike phonological similarity, semantic similarity does not produce impairment in order information.

Application of the perturbation model

Once again, it is useful to see how well the perturbation model, which assumes independence between I and O_R, can handle the basic data patterns. The choice of parameter values for θ_1 and θ_2 was driven mainly by the microscopic error patterns, as in Experiment 1. Because the within-list transposition error rates were comparable for similar and dissimilar lists (see Table 2), the within-list perturbation probability, θ_1, was set to roughly the same value for both list types (θ_1 was set to 0.11 for similar lists and 0.10 for dissimilar lists). On the other hand, similar lists led to significantly fewer across-list intrusions, presumably because it was easier for the subjects to identify the items from a list, so θ_2 was set significantly higher for dissimilar lists (0.15 for similar lists and 0.25 for dissimilar lists). All other modeling decisions followed the earlier application in Experiment 1.

The relevant data from the simulations are shown in Figs. 3 and 4. Again, no attempt was made to provide a precise quantitative fit, to keep the model simple, but it is easy to see that the model is capable of producing the general data patterns from Experiment 2. Even with comparable within-list perturbation probabilities for similar and dissimilar lists, an advantage appeared for the similar lists in the Inclusion condition. A comparable pattern was found in the Exclusion condition, although the inverted form found in the empirical data is evident and the similarity advantage is numerically smaller. The estimates of I and O_R, shown in Fig. 5, showed the expected patterns: Semantic similarity led to an increase in I, but no differences were found in the estimates of O_R.

What accounts for the differential effect of similarity in Experiments 1 and 2? Subjects are clearly sensitive to both phonological and semantic similarity because both kinds of similarity significantly affected performance in these experiments. One possibility is that within-list discriminations (e.g., did the word occur in position two or three?) are mediated primarily by sound-based features. The fact that errors in short-term recall tend to be sound-based, not meaning-based, supports the claim that phonological features receive more weight (e.g., Conrad, 1964). On the other hand, semantic information is potentially quite useful as a cue for list membership (Nairne & Neumann, 1993). Knowing that a list was composed entirely of flowers, for example, provides distinctive information that should be useful in guiding recall; however, that same information provides no information about whether an item occurred in the second or third serial position. To determine within-list position, subjects may need to rely on residual trace information in short-term memory that exists primarily in phonological form (see Nairne, 1990, 1999; Saint-Aubin & Poirier, 1999). We return to this issue in more detail in the general discussion.

We can also assume that similarity, in some instances, helps the subject guess the correct answer. Knowing that the list consists only of words that rhyme, or flowers, presumably restricts the guessing space, although how much is difficult to determine. Neither the process dissociation analysis nor the perturbation application explicitly models guessing; instead, guessing is assumed to

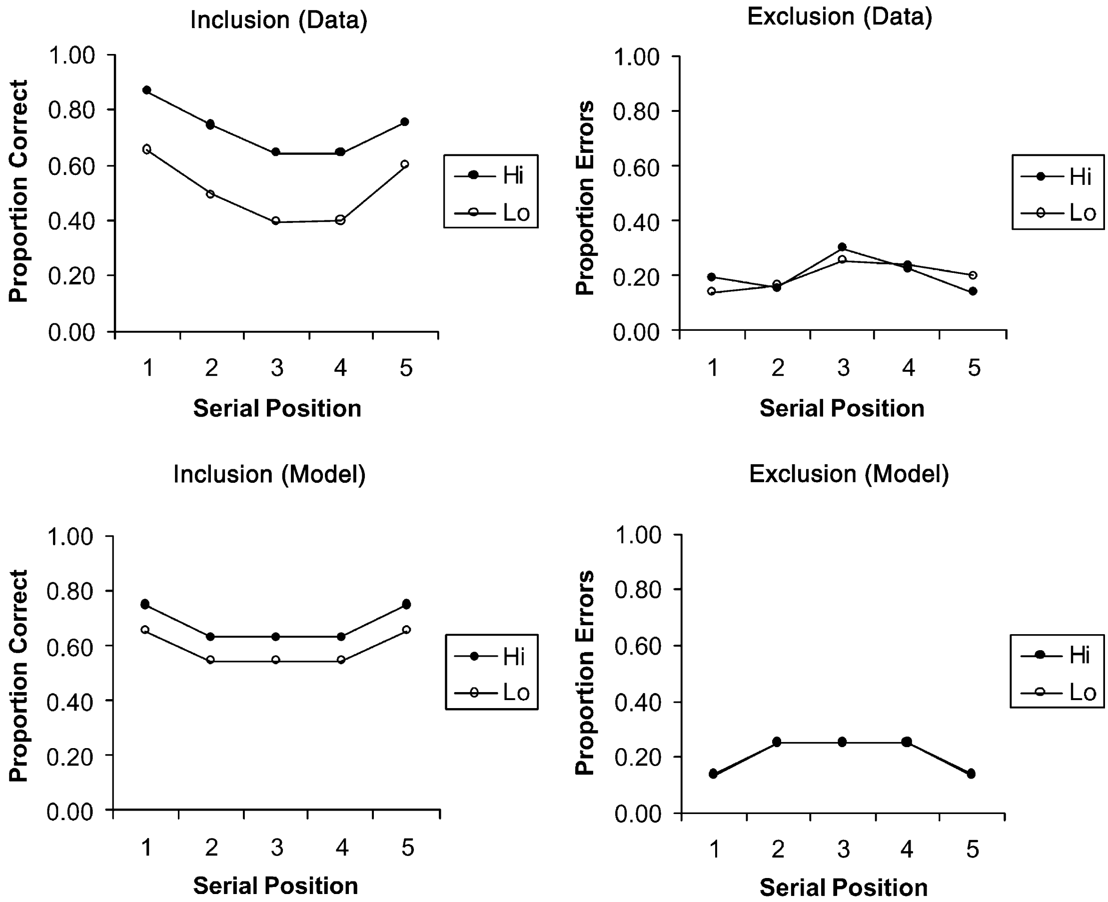

Fig. 5. Data from the Inclusion and Exclusion conditions, plotted as a function of word frequency and serial position. The lower two panels show data generated from an application of the perturbation model.

contribute—or add noise—to the retention estimates in an undetermined way. One could attempt to add guessing parameters, and it may be profitable to do so at some point. However, unfortunately, guessing is not independent of item and order information—i.e., remembering that list items rhyme, a type of item information, presumably modulates guessing in a direct way. Without an explicit model of guessing, and how it differs (if at all) from item-based retrieval, there is no simple way of interpreting guessing parameters when added. We return to this issue as well in the general discussion.

Experiment 3

Word frequency is another independent variable that is widely believed to differentially affect the retention of item and order information. In immediate serial recall, high frequency words generally lead to superior performance, a result that has been used to argue for a trace interpretation, or redintegration, component in short-term recall (e.g., Hulme et al., 1997; Poirier & Saint-Aubin, 1996). It has been suggested that people are better able to interpret the degraded, or blurry, traces of high-frequency words because they have richer associations in long-term memory, or phonological representations that are easier to access in lexical memory.

Empirically, of course, a main effect of frequency in serial recall tells us little about how frequency selectively influences item and order information. Some researchers have proposed that high frequency words lead to improved order retention because high frequency words are easier to associate within a list (Deese, 1960; DeLosh & McDaniel, 1996). DeLosh and McDaniel (1996) provided some support for this view by showing that subjects are sometimes better able to reconstruct the order of lists containing high frequency words. This is a potentially important finding, as noted by DeLosh and McDaniel (1996), because it provides one way to interpret classic word frequency effects in free recall and recognition (see also, Whiteman et al., 1994). Unfortunately, neither Whiteman et al. (1994) nor Poirier and

Saint-Aubin (1996) found comparable results. Whiteman et al. (1994) generally found no effect of word frequency on the reconstruction of order (although there were slight advantages for the high frequency words in some instances); Poirier and Saint-Aubin (1996) found that increasing word frequency led to better item recall, but produced no differences in conditionalized order recall (i.e., transposition errors; see also Hulme et al., 1997). More recently, Mulligan (2001) found an advantage for high frequency words on a test of absolute serial position, but not on a test of relative order.

Experiment 3 uses the process dissociation logic to derive estimates of item and order information for lists composed of high versus low frequency words. Once again, as in Experiments 1 and 2, subjects recalled short lists of words under Inclusion and Exclusion instruction conditions and estimates of I and O_R are calculated from the results.

Method

Subjects and apparatus

Sixty Purdue undergraduates participated for course credit. Everyone was tested individually and all materials were presented and controlled by personal computers.

Materials and design

The stimulus materials were taken from Kucera and Francis (1967) and consisted of words four to seven letters in length. High frequency words had occurrence rates of greater than 100 per million (high frequency mean = 205.6 per million); low frequency words had an occurrence rate of one per million. Design features again matched the previous experiments: Sixty lists were presented in Inclusion and Exclusion blocks of 30 trials each. Within each block, half of the lists contained five unique high frequency words and half contained five unique low frequency words. Individual words occurred only once in the experiment. Within each list, the words were randomly assigned to positions.

Procedure

Procedural details (e.g., Inclusion/Exclusion instructions; timing, etc.) matched those employed in Experiments 1 and 2.

Results and discussion

Data from the Inclusion and Exclusion conditions are presented in the upper portions of Fig. 5, plotted as a function of serial position and frequency level. Turning first to the Inclusion condition, representing traditional serial recall, the results showed the expected trends. The serial position curves were bow-shaped [$F(4,236) = 51.15$; $MSE = .02$; $p < .001$], and there was a highly significant effect of frequency [$F(1,59) = 224.10$; $MSE = .03$; $p < .001$]. High frequency words led to superior serial recall performance, a result that has been reported many times in the immediate serial recall literature (e.g., Hulme et al., 1997; Kausler & Puckett, 1979; Poirier & Saint-Aubin, 1996; Watkins & Watkins, 1977). There was also a significant interaction between serial position and frequency [$F(4,246) = 4.49$; $MSE = .01$; $p < .01$]: The high frequency advantage was slightly reduced at the first and last serial positions (see Hulme et al., 1997, for a discussion of similar interaction forms). Again, the patterns reported here resemble those typically found in immediate serial recall, despite the fact that recall followed a filled delay.

The results from the Exclusion condition showed a different pattern. An overall ANOVA revealed a significant effect of serial position [$F(4,236) = 5.59$; $MSE = .06$; $p < .001$], but neither frequency [$F(1,59) < 1$] nor the interaction of frequency by serial position was statistically significant [$F(4,236) = 1.39$; $MSE = .05$; $p > .05$]. The serial position curves showed the same inverted form found in Experiments 1 and 2, but word frequency failed to affect performance. Turning to the individual error patterns, shown in Table 3, there was essentially no effect of word frequency on transposition errors in either the Inclusion or the Exclusion conditions. Frequency exerted its main effect on omission errors—significantly more omission errors occurred for the low-frequency word lists, which supports the claim that frequency affects primarily item rather than order retention.

Estimates of I and O_R

The item and order estimates, derived from process dissociation, are shown in Fig. 6. An overall ANOVA revealed a main effect of frequency [$F(1,59) = 183.00$; $MSE = .03$; $p < .001$] and a significant frequency by item/order interaction [$F(1,59) = 27.48$; $MSE = 08$; $p < .001$]; the main effect of I versus O_R was not significant [$F(1,59) = 2.88$; $MSE = .19$; $p > .10$]. Planned comparisons showed that the effect of word frequency was significant for both I [$F1,59) = 114.30$; $MSE = .07$; $p < .001$] and O_R [$F(1,59) = 8.93$; $MSE = .05$; $p < .01$], although clearly the effect was considerably larger for I. Based on these estimates, word frequency seems to affect primarily item information, supporting the proposal that the degraded traces of high frequency words may be easier to interpret, or deblur (e.g., Hulme et al., 1997; Poirier & Saint-Aubin, 1996). At the same time, word frequency did affect the estimate of O_R, with order retention enhanced for high frequency words, indicating that it may also be easier for people to form interitem associations in high-frequency word lists. The fact that the effect on order was relatively small may help to explain why the effects of frequency on empirical tests of

Table 3
Mean proportion of errors, collapsed across serial position, for high and low frequency lists in the Inclusion and Exclusion conditions of Experiment 3

Error Type	Condition	High Freq.	Low Freq.
Transposition	Inclusion	.10	.10
	Exclusion	.15	.17
	All	.13	.14
Omissions	Inclusion	.17	.39
	Exclusion	.16	.37
	All	.16	.38
Intrusions—prior lists	Inclusion	.02	.02
	Exclusion	.02	.02
	All	.02	.02
Intrusions—out of experiment	Inclusion	.03	.09
	Exclusion	.04	.10
	All	.03	.09
Repetitions	Inclusion	.00	.00
	Exclusion	.00	.00
	All	.00	.00

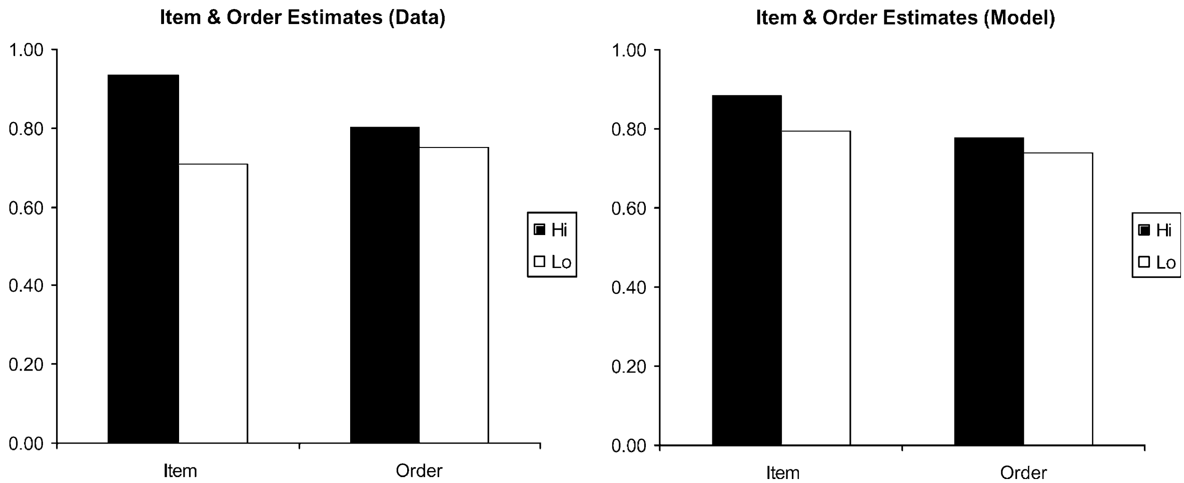

Fig. 6. Average Item and Order estimates generated from the process dissociation application in Experiment 3. Estimates in the right panel are derived from data generated by the perturbation model.

order, such as reconstruction, have been inconsistent across studies (e.g., DeLosh & McDaniel, 1996; Whiteman et al., 1994).

Application of the perturbation model

Once again, we tried to get the perturbation model to mimic the data and we used the error rates shown in Table 3 as a guide for setting the perturbation parameters. As in Experiment 2, the within-list transposition error rates did not differ much so we kept θ_1 roughly the same for high and low frequency words ($\theta_1 = 0.07$ for high and 0.08 for low). Across-list intrusion rates differed substantially so θ_2 was set higher for the low frequency words ($\theta_2 = 0.05$ for high and 0.10 for low). The relevant data for the Inclusion and Exclusion conditions are shown in the lower half of Fig. 5. Again, the major finding is a significant effect of frequency, favoring high frequency words, in the Inclusion condition and essentially no frequency difference in the Exclusion condition. There was a greater effect of frequency in the subject data, and there is no hint of the serial position by frequency interaction in the model data. Again, output interference may well be responsible, in part, for this latter interaction.

Of main interest are the estimates of I and O_R which are shown in Fig. 6. The estimates derived from the subject data showed a relatively large effect of frequency on I and a small effect on O_R. The same general pattern is apparent in the model's estimates, although the effect of frequency on I is smaller. It is clear that the model is capable of capturing the form of the interaction, which is important because it shows that the model can support a variety of interactive patterns. The model produced a cross-over interaction in Experiment 1, an effect on I but not on O_R in Experiment 2, and an effect on both I and O_R in the present experiment. Notably, each of these patterns was obtained with a model assuming that the main mechanisms for item and order information are essentially independent.

Experiment 4

In our final experiment, we examined the effect of generation on item and order retention. The generation effect refers to the finding that self-generation of materials (e.g., generating the word APPLE from a word fragment_PPLE) typically improves retention (e.g., Slamecka & Graf, 1978). However, generation is actually believed to have dissociative effects on item and order memory: Whereas generation improves performance on tests tapping primarily item information (e.g., recognition), it significantly impairs performance in some recall environments (e.g., Schmidt & Cherry, 1989) and specifically on reconstruction of order tests (Burns, 1996; Greene, Thapar, & Westerman, 1998; Mulligan, 2002; Nairne et al., 1991). This dissociation has been important theoretically because it has helped to explain why the generation advantage can depend on whether within- or between-subjects designs are employed (see DeLosh & McDaniel, 1996; Nairne et al., 1991; Serra & Nairne, 1993).

Empirically, the dissociation pattern is clear and has been replicated many times. However, once again, the reasoning rests on the connection between measures of memory and particular retention tests. In Experiment 4, we apply the process dissociation logic to the effects of item generation: Subjects generated or read lists of items under Inclusion and Exclusion instructions. We predicted that generation would enhance estimates of I and impair estimates of O_R, a pattern similar to the one found in Experiment 1 for phonological similarity.

Method

Subjects and apparatus

Sixty Purdue undergraduates participated for course credit in introductory psychology. Subjects were tested individually and all stimuli were presented and controlled by personal computers.

Materials and design

The stimulus materials were drawn from the Paivio, Yuille, and Madigan (1968) norms. The words were four to seven letters in length and rated relatively highly in imageability (mean = 5.9), concreteness (mean = 6.05), and meaningfulness (mean = 5.65). Across subjects, each word appeared in both a generate (word fragment) and read (intact) form and was presented under Inclusion and Exclusion instructions. Single-solution word fragments were created by removing a single letter and replacing it with an underscore. A given word, or word fragment, occurred only once in the experiment.

In other respects, the general design matched the earlier experiments. Subjects received a total of 60 lists broken into two blocks of 30 Inclusion and Exclusion trials. Within each block, 15 lists contained intact words (the Read condition) and 15 lists consisted entirely of word fragments (the Generate condition). The order of the read and generate trials was determined randomly within a block, as was the assignment of words to individual list positions.

Procedure

The instructions, timing, and other procedural variables matched those employed in the earlier experiments. Subjects were told to complete each word fragment, aloud, as it appeared. An experimenter was present at all times to record the occurrence of generation failures.

Results and discussion

Very few generation errors were made during list presentation. Correct item generation occurred approximately 93% of the time. However, we analyzed the data described below in two ways, either conditionalized on correct generation or unconditionalized. Because both analyses yielded essentially the same results, we report only the unconditionalized data below.

The results from the Inclusion and Exclusion conditions are shown in the upper-portion of Fig. 7, presented as a function of serial position and presentation condition (Read versus Generate). For the Inclusion condition, as expected, performance was significantly better in the Read condition [$F(1,59) = 60.30$; $MSE = .03$; $p < .001$]. This result supports earlier work showing that generation can impair retention in some circumstances (e.g., Nairne et al., 1991). There was also a significant effect of serial position [$F(2,236) = 50.00$; $MSE = .03$; $p < .001$] and the interaction was significant as well [$F(4,236) = 5.05$; $MSE = .01$; $p < .01$]. The size of the Generate/Read difference appeared to decrease over serial position.

A quite different pattern of results was found for the Exclusion condition. The ANOVA revealed a significant effect of serial position [$F(4,236) = 4.91$; $MSE = .06$; $p < .001$], but neither Generate/Read nor the interaction approached significance [Fs < 1]. The overall pattern shown in Fig. 7 is reminiscent of the pattern found for

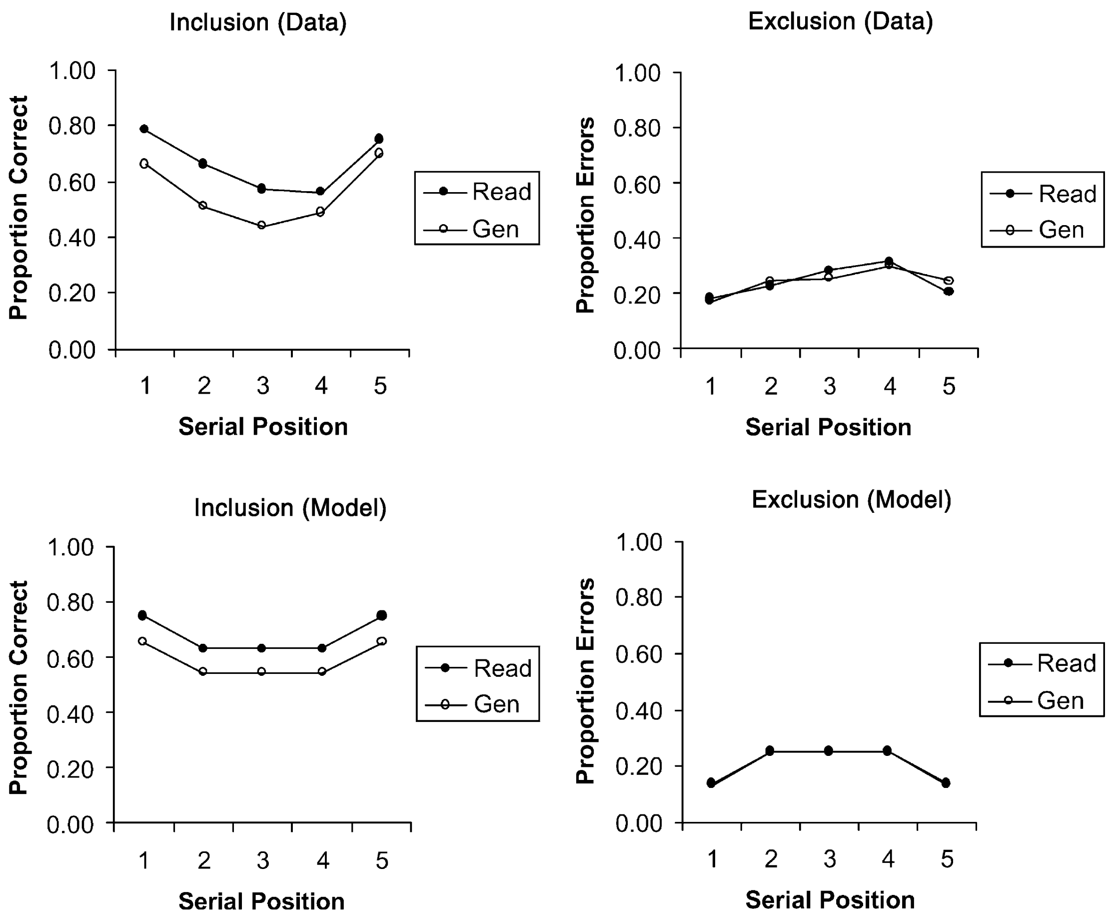

Fig. 7. Data from the Inclusion and Exclusion conditions, plotted as a function of the read/generate variable and serial position. The lower two panels show data generated from an application of the perturbation model.

word frequency in Experiment 3: The main presentation variable (frequency versus generation) affected performance only in traditional serial recall. Moreover, as in Experiment 3, there were no significant differences in transposition errors (shown in Table 4) for the Generate and Read conditions; instead, the main difference was in omission errors. Generation led to greater omission error rates than did reading.

Estimates of I and O_R

The derived estimates for I and O_R are shown in Fig. 8. Not surprisingly, the pattern is very similar to the one reported in Experiment 3. There was a significant effect of item versus order [$F(1,59) = 5.56$; $MSE = .10$; $p < .02$], a significant effect of Generate versus Read [$F(1,59) = 37.90$; $MSE = .01$; $p < .001$], and a significant interaction between the two main variables [$F(1,59) = 4.23$; $MSE = .02$; $p < .05$]. Individual planned comparisons revealed significant effects of Generate/Read for both I [$F(1,59) = 18.50$; $MSE = .02$; $p < .001$] and O_R [$F(1,59) = 4.23$; $MSE = .02$; $p < .05$]. The effect of generation, although negative for both item and order retention, appeared larger for item than order information.

These data patterns are quite clear, but unexpected in some respects. Generation was predicted to enhance I and impair O_R, a dissociative pattern. Instead, generation impaired both kinds of mnemonic information, much like the effect of frequency in Experiment 3. The findings for O_R do replicate the negative effect of generation found in reconstruction of order tests (e.g., Burns, 1996; Greene et al., 1998; Nairne et al., 1991), but the decrement is numerically small. The fact that generation impaired I is more troubling, especially given that generation has traditionally been assumed to enhance item information. In an experimental context similar to the present one, for example, generation was shown to impair reconstruction of order while enhancing the recognition of individual items (Burns, Curti, & Lavin, 1993; Nairne et al., 1991). Simple item recognition, however, is not necessarily a pure index of item information (that is, it suffers from the process

Table 4
Mean proportion of errors, collapsed across serial position, for generated and read lists in the Inclusion and Exclusion conditions of Experiment 4

Error Type	Condition	Generate	Read
Transposition	Inclusion	.12	.11
	Exclusion	.21	.21
	All	.17	.16
Omissions	Inclusion	.32	.23
	Exclusion	.31	.20
	All	.31	.21
Intrusions—prior lists	Inclusion	.02	.02
	Exclusion	.02	.02
	All	.02	.02
Intrusions—out of experiment	Inclusion	.04	.03
	Exclusion	.06	.04
	All	.05	.03
Repetitions	Inclusion	.00	.00
	Exclusion	.00	.00
	All	.00	.00

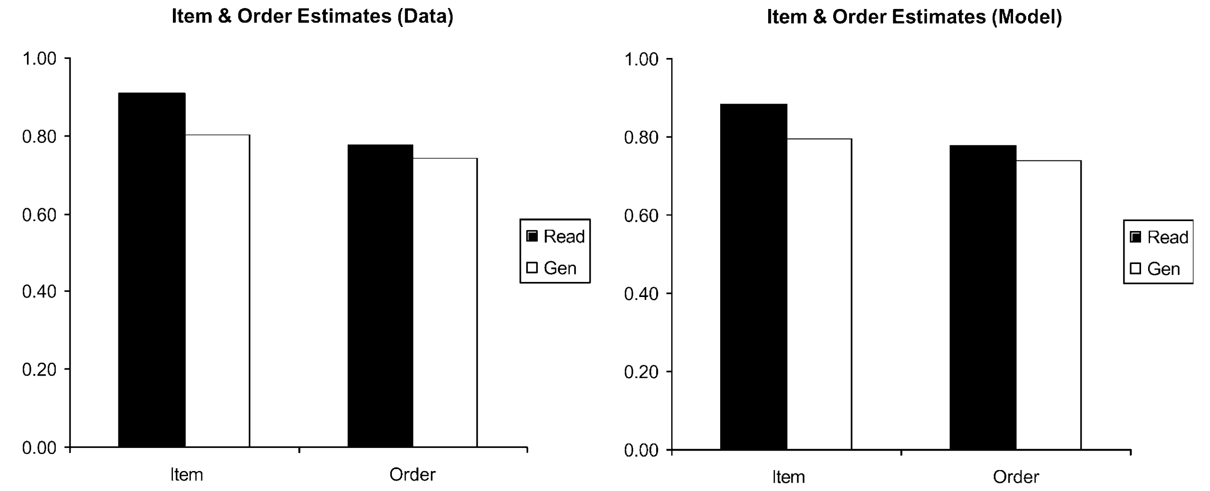

Fig. 8. Average Item and Order estimates generated from the process dissociation application in Experiment 4. Estimates in the right panel are derived from data generated by the perturbation model.

purity concern). Moreover, recognition judgments may be based on familiarity, rather than directed retrieval, and generation may have its main effects on familiarity in contexts such as the present one. We return to this point in the general discussion.

Application of the perturbation model

Because the results of Experiment 4 are quite similar to those found in Experiment 3, it is clear that the perturbation model is capable of reproducing the general data patterns. In fact, we found that the parameter values used in Experiment 3 did a reasonable job of reproducing the results for Experiment 4. In the present case, of course, θ_2 was higher for generated than for read items because the former led to substantially higher omission errors. The model data are shown in Figs. 7 and 8. The fact that the data patterns for Experiments 3 and 4 could be modeled with the same set of parameters, of course, does not imply that frequency and generation rely on the same psychological processes. In this context, they simply affect item and order information in similar ways.

General discussion

The purpose of the present experiments was to introduce and apply a new technique for the measurement of item and order information. As noted throughout, the item–order distinction is important theoretically, and traditional variables can affect item and order information in unique ways. For example, in our current Experiment 1, phonological similarity led to increases in the estimates of item information and to decreases in the estimates of order information. In Experiment 2, semantic similarity increased item information, but had no significant effect on order. In Experiments 3 and 4, word frequency and a generate/read manipulation had parallel effects on item and order (although larger effects were found for item information). Empirically, then, item and order information are clearly dissociable, and any complete model of remembering will need to account for the different patterns of results.

The unique contribution of the present analyses, of course, is the technique for deriving estimates of I and O_R. In work conducted up to this point, researchers have generally employed particular retention tests to gauge how item and order retention operate. Thus, tests such as recognition and free recall are commonly used to tap item information; reconstruction, or some kind of probed position test, is used to measure order. However, as noted in the introduction, it is reasonable to question whether any retention test is really process pure—instead, most tests probably rely on multiple kinds of mnemonic processing (Dunn & Kirsner, 1989; Jacoby, 1991). Serial recall, for instance, requires one to remember the items from the list as well as their positions of occurrence. Our measures of I and O_R were derived from situations in which both item and order information were needed, although sometimes they were placed in opposition. For example, remembering the correct serial position for an item was designed to increase recall of that item in the Inclusion condition and sometimes decrease its recall in the Exclusion condition.

Generally, the derived estimates I and O_R conformed to expected trends. The major unexpected finding occurred in Experiment 4, where generation was discovered to produce a negative effect on I. Most researchers have assumed that generation enhances item information, based on results drawn primarily from item recognition tests. However, in between-list designs, where read and generated items always occur in different lists, generation often produces no advantage (or even a decrement) compared to reading, at least in tests other than recognition (e.g., Slamecka & Katsaiti, 1987). The absence of a generation effect in such circumstances has been attributed to the loss of order information, disrupting the subject's output strategy (e.g., the item–order hypothesis; DeLosh & McDaniel, 1996; Nairne et al., 1991; Serra & Nairne, 1993), but the present results suggest an alternative explanation. It may be that the act of generation in a between-list design disrupts recollection for both item and order information, but leads to an enhancement of an item's general familiarity. Familiarity can act as an effective mnemonic cue in a recognition test, but most other tests (e.g., recall and reconstruction) require the subject to access specific occurrence information or to self-generate the item from cues. The current estimates of I were derived from test environments in which familiarity could not have played much of a role, which may explain why no enhancing effects of generation were detected. (For other variables that selectively dissociate estimates of item and order information after generation, such as comparing words and nonwords, see Mulligan (2002).)

There was also a relatively close correspondence between the error data, both transposition errors and omission errors, and the item/order estimates. In some cases, such as in Experiments 3 and 4, there were no apparent differences in transposition errors across the main conditions, suggesting that neither frequency nor generation affected order, yet both variables produced significant differences in the estimates of O_R. However, these discrepancies disappear when omission errors are taken into account—that is, one can compute the number of transposition errors per item recalled, or the proportion of correct item placements per item recalled (e.g., Murdock, 1976; Saint-Aubin & Poirier, 1999).[4] Deriving measures of item and order memory through conditionalizing relies on the same logic as process dissociation: Order is defined as the number of items recalled in their proper serial positions divided by the number of items recalled irrespective of position (although, to our knowledge, the underlying model is rarely, if ever, specified when conditionalized measures are reported).

In the present application, of course, the derived measures of item and order are based on comparisons across two tasks, Inclusion and Exclusion, rather than simply on traditional serial recall. Relying exclusively on transposition errors in serial recall is problematic because early errors prevent later items from being positioned correctly—that is, the correct positions of later items are already occupied. This problem exists in our Inclusion task as well, but the Exclusion condition was designed, in part, to minimize such contamination by allowing subjects to focus on position information for only a single item—the to-be-excluded item (other list items could be recalled irrespective of their original

[4] Conditionalized measures of order in Experiment 3, like the estimates derived from process dissociation, showed an advantage for the high frequency words (.88) compared to the low frequency words (.84); in Experiment 4, an advantage occurred for the Read (.86) compared to the Generate (.82) conditions after conditionalizing.

positions of occurrence). In addition, Exclusion performance, and therefore the contribution of O_R, could be sampled equally for each serial position across trials, which is something that cannot be assured in serial recall (early output errors can impair measurement of O_R for later items). Conservatively, of course, one should probably rely on multiple measures of item and order information at this point; the present estimates add to the portfolio of possible indices that are currently employed, including conditionalizing from a single serial recall task as well as ostensibly "order-only" tasks such as reconstruction.

The independence assumption

Any application of process dissociation tends to be controversial, in part, because of its underlying assumptions (e.g., Curran & Hintzman, 1995). In the present case, we have assumed that item and order information operate independently, and the equations used to derive I and O_R follow from this assumption. Any violations of independence could, in principle, distort the estimates of I and O_R and render our application suspect. Unfortunately, there is no unequivocal way to assess independence, although the independence of item and order information is commonly assumed in the literature. The fact that our estimates conformed generally to the patterns revealed by other indices of item and order memory provides one kind of validation, but it is certainly not definitive evidence.

We also showed that a model assuming independence could match the empirical patterns, and derived estimates, produced from our experiments. We found that the perturbation model of Estes (Estes, 1972, 1997; Lee & Estes, 1981) did an acceptable job of fitting the diverse data patterns, including the cross-over interaction of Experiment 1, without any major changes in its assumptions. Furthermore, we grounded our choices for parameter values by using the microscopic error patterns from the experiments to help pick the values. We have therefore extended the range of this model to an entirely new set of empirical phenomena. To our knowledge, the perturbation model has never previously been applied to data of the sort collected in these experiments.

It is of interest to note that although the perturbation model assumes that item loss is independent of order retention (and vice versa), the same forgetting mechanism is applied to each. That is, items are initially encoded as values along a dimension—either list (item) or within-list (order) position—and over time the values drift, or perturb, along the dimension. Prior analyses of error distributions lend support for this proposal (see Estes, 1997; Nairne, 1991); in addition, Nairne (1991) was able to show that overall performance could be predicted reasonably well, but not perfectly, by combining performance on each dimension separately (in accordance with an independence assumption). One can therefore conceptualize the difference between item and order retention as essentially one of the grain at which order or position information needs to be accessed—both fundamentally involve the loss of order, but at different levels (list versus within-list position).

Finally, one might criticize the current application because we have generally ignored the role of guessing in the reported data. Again, we might be able to add guessing parameters to the applications, and it could be beneficial to do so in a future application, but without an explicit model of guessing (and retrieval) it would be difficult to determine the meaning of those parameters. Guessing itself can be conceptualized as a type of cue-driven retrieval—we use trial information (list length, item characteristics) to drive the guessing process and it is unclear what relative mix of trial-specific mnemonic and non-mnemonic information is involved (that is, how exactly (or when) is guessing different from normal cue-driven retrieval?). Moreover, even if there is a non-mnemonic component to the derived estimates of I and O_R, it does not necessarily follow that this component is affecting the qualitative patterns in any significant way.

We believe that our technique for obtaining estimates of item and order information is likely to have value across a variety of empirical environments. At the very least, as noted earlier, it adds to the portfolio of estimates that is currently available. Furthermore, the technique itself is quite easy to apply. Virtually any independent variable that has been applied to the immediate retention of serial order can be used in the context of Inclusion and Exclusion conditions. Thus, our technique can help to verify existing data patterns, as in the current experiments, or it can be used to isolate the effects of new variables on item and order retention.

References

Baddeley, A. D. (1966). The influence of acoustic and semantic similarity on long-term memory for word sequences. *Quarterly Journal of Experimental Psychology, 18*, 302–309.

Baddeley, A. D. (1986). *Working memory*. Oxford, England: Oxford University Press.

Bjork, E. L., & Healy, A. F. (1974). Short-term order and item retention. *Journal of Verbal Learning and Verbal Behavior, 13*, 80–97.

Buchner, A., Erdfelder, E., & Vaterrodt-Plünnecke, B. (1995). Toward unbiased measurement of conscious and unconscious memory processes within the process dissociation framework. *Journal of Experimental Psychology: General, 124*, 137–160.

Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review, 106*, 551–581.

Burns, D. J. (1996). The item–order distinction and the generation effect: The importance of order information in long-term memory. *American Journal of Psychology, 109*, 567–580.

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.

Coltheart, V. (1993). Effects of phonological similarity and concurrent irrelevant articulation on short-term memory recall of repeated and novel word lists. *Memory & Cognition, 21*, 539–545.

Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology, 55*, 75–84.

Crowder, R. G. (1979). Similarity and order in memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 13, pp. 319–353). New York: Academic Press.

Curran, T., & Hintzman, D. L. (1995). Violations of the independence assumption in process dissociation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 531–547.

Deese, J. (1960). Frequency of usage and number of words in free recall: The role of association. *Psychological Reports, 7*, 337–344.

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.

Dunn, J. C., & Kirsner, K. (1989). Implicit memory: Task or process? In S. Lewandowsky, J. C. Dunn, & K. Kirsner (Eds.), *Implicit memory: Theoretical issues*. Hillsdale, NJ: Erlbaum.

Drewnowski, A. (1980). Attributes and priorities in short-term recall: A new model of memory span. *Journal of Experimental Psychology: General, 109*, 208–250.

Estes, W. K. (1972). An associative basis for coding and organization in memory. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory* (pp. 161–190). New York, NY: Halsted Press.

Estes, W. K. (1997). Processes of memory loss, recovery, and distortion. *Psychological Review, 104*, 148–169.

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

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

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

Hintzman, D. L. (1990). Human learning and memory: Connections and dissociations. *Annual Review of Psychology, 41*, 109–139.

Hirshman, E., & Master, S. (1997). Modeling the conscious correlates of recognition memory: Reflections on the remember-know paradigm. *Memory & Cognition, 25*, 345–351.

Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D. A., Martin, S., & Stuart, G. (1997). Word frequency effects on short-term memory tasks: Evidence for a multinomial processing tree model of immediate serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 1217–1232.

Hunt, K. P., & Hodge, M. H. (1971). Category-item frequency and category-name meaningfulness (m): Taxonomic norms for 84 categories. *Psychonomic Monograph Supplements, 4*, 97–121.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30*, 513–541.

Jacoby, L. L. (1998). Invariance in automatic influences of memory: Toward a user's guide for the process-dissociation procedure. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 3–26.

Jacoby, L. L., Yonelinas, A. P., & Jennings, J. (1997). The relation between conscious and unconscious (automatic) influences: A declaration of independence. In J. Cohen & J. W. Schooler (Eds.), *Scientific approaches to the question of consciousness*. Hillsdale, NJ: Erlbaum.

Kausler, D. H., & Puckett, J. M. (1979). Effects of word frequency on adult age differences in word memory span. *Experimental Aging Research, 5*, 161–169.

Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

Lee, C. L., & Estes, W. K. (1981). Item and order information in short-term memory: Evidence for multilevel perturbation processes. *Journal of Experimental Psychology: Human Learning and Memory, 7*, 149–169.

Libkuman, T. M. (1994). Norms for words that rhyme. *Behavior Research Methods, Instruments, and Computers, 26*, 278–322.

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.

Mulligan, N. W. (2002). The generation effect: Dissociating enhanced item memory and disrupted order memory. *Memory & Cognition, 30*, 850–861.

Murdock, B. B. (1976). Item and order information in short-term serial memory. *Journal of Experimental Psychology: General, 105*, 191–216.

Murdock, B. B., Jr., & vom Saal, W. (1967). Transpositions in short-term memory. *Journal of Experimental Psychology, 74*, 137–143.

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition, 18*, 251–269.

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

Nairne, J. S. (1999). A functional analysis of primary memory. In H. L. Roediger, J. S. Nairne, I. Neath, & A. M. Surprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder*. Washington, DC: American Psychological Association.

Nairne, J. S., & Kelley, M. R. (1999). Reversing the phonological similarity effect. *Memory & Cognition, 27*, 45–53.

Nairne, J. S., & Neumann, C. (1993). Enhancing effects of similarity on long-term memory for order. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 329–337.

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.

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

Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review, 4*, 761–781.

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph, 76*(1, Pt. 2).

Poirier, M., & Saint-Aubin, J. (1996). Immediate serial recall, item identity and item position. *Canadian Journal of Experimental Psychology, 50*, 408–412.

Postman, L. (1972). A pragmatic view of organization theory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 3–48). San Diego, CA: Academic Press.

Ratcliff, R., Van Zandt, T., & McKoon, G. (1995). Process dissociation, single-process theories, and recognition memory. *Journal of Experimental Psychology: General, 124*, 352–374.

Saint-Aubin, J., & Poirier, M. (1999). Semantic similarity and immediate serial recall: Is there a detrimental effect on order information? *Quarterly Journal of Experimental Psychology, 52A*, 367–394.

Schmidt, S. R., & Cherry, K. (1989). The negative generation effect: Delineation of a phenomenon. *Memory & Cognition, 17*, 359–369.

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

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory, 4*, 592–604.

Slamecka, N. J., & Katsaiti, L. T. (1987). The generation effect as an artifact of selective rehearsal. *Journal of Memory and Language, 26*, 589–607.

Watkins, O. C., & Watkins, M. J. (1977). Serial recall and the modality effect: Effects of word frequency. *Journal of Experimental Psychology: Human Learning and Memory, 3*, 712–718.

Watkins, M. J., Watkins, O. C., & Crowder, R. G. (1974). The modality effect in free and serial recall as a function of phonological similarity. *Journal of Verbal Learning and Verbal Behavior, 13*, 430–447.

Whiteman, H. L., Nairne, J. S., & Serra, M. (1994). Recognition and recall-like processes in the long-term reconstruction of order. *Memory, 2*, 175–194.

Wickelgren, W. A. (1965). Acoustic similarity and intrusion errors in short-term memory. *Journal of Experimental Psychology, 70*, 102–108.