Journal of Experimental Psychology: Learning, Memory, and Cognition

Not All Order Memory Is Equal: Test Demands Reveal Dissociations in Memory for Sequence Information

Tanya R. Jonker and Colin M. MacLeod

Online First Publication, September 22, 2016. http://dx.doi.org/10.1037/xlm0000301

Journal of Experimental Psychology:
Learning, Memory, and Cognition

0278-7393/16/$12.00 http://dx.doi.org/10.1037/xlm0000301

Not All Order Memory Is Equal: Test Demands Reveal Dissociations in Memory for Sequence Information

Tanya R. Jonker and Colin M. MacLeod
University of Waterloo

Remembering the order of a sequence of events is a fundamental feature of episodic memory. Indeed, a number of formal models represent temporal context as part of the memory system, and memory for order has been researched extensively. Yet, the nature of the code(s) underlying sequence memory is still relatively unknown. Across 4 experiments that manipulated encoding task, we found evidence for 3 dissociable facets of order memory. Experiment 1 introduced a test requiring a judgment of which of 2 alternatives had immediately followed a word during encoding. This measure revealed better retention of interitem associations following relational encoding (silent reading) than relatively item-specific encoding (judging referent size), a pattern consistent with that observed in previous research using order reconstruction tests. In sharp contrast, Experiment 2 demonstrated the reverse pattern: Memory for the studied order of 2 sequentially presented items was actually better following item-specific encoding than following relational encoding. Experiment 3 reproduced this dissociation in a single experiment using both tests. Experiment 4 extended these findings by further dissociating the roles of relational encoding and item strength in the 2 tests. Taken together, these results indicate that memory for event sequence is influenced by (a) interitem associations, (b) the emphasized directionality of an association, and (c) an item's strength independent of other items. Memory for order is more complicated than has been portrayed in theories of memory and its nuances should be carefully considered when designing tests and models of temporal and relational memory.

Keywords: memory, order, temporal, interitem associations, item-order account

A key feature of the episodic memory system is the ability to place events in time. Indeed, many formal models include a representation of temporal features or contextual features that could involve temporal characteristics (e.g., Kahana, 1996; Polyn, Norman, & Kahana, 2009; Raaijmakers & Shiffrin, 1981; Sederberg, Howard, & Kahana, 2008; Shiffrin & Steyvers, 1997), and these are theorized to aid not just the retrieval of the target item (e.g., Jonker, Seli, & MacLeod, 2013; Sahakyan & Kelley, 2002), but also the retrieval of other items that were temporally proximal to the target item (e.g., Howard & Kahana, 2002).

Although generally there is agreement that temporal information plays an important role in memory, the exact mechanisms underlying the encoding and use of this information are not well understood. One of the oldest accounts of order memory is that of item-to-item associations (e.g., chaining; Ebbinghaus, 1885; Lewandowsky & Murdock, 1989), which can be used to reconstruct the relative order of events in a sequence (other accounts include item-to-context associations, e.g., Ebenholtz, 1972; Mensink & Raaijmakers, 1988; hierarchical organization, e.g., Lashley, 1951; and retrieval-related fluctuations in temporal/contextual representations, e.g., Howard & Kahana, 2002; Polyn et al., 2009). Despite being one of the oldest accounts of order memory, this interitem association hypothesis certainly has not fallen out of favor. A more recent theory involving interitem associations—the *item-order account*—emphasizes the important role that interitem associations play in guiding free recall (Nairne, Riegler, & Serra, 1991; Serra & Nairne, 1993). According to this account, memory for order is the direct product of the encoding of relational information between items, which is greater during common or passive encoding tasks, such as silent reading. Specifically, when participants study lists of words by silently reading or passively viewing, they encode strong *interitem associations* between list items, which can later be used to reconstruct the order of a series of events and/or to guide free recall. Indeed, numerous studies have demonstrated superior performance following relational encoding on tests of order reconstruction (studied items are presented during test in a scrambled order), as well as stronger evidence of preserved order in free-recall outputs (for a review, see McDaniel & Bugg, 2008). In contrast, when participants perform atypical encoding tasks, such as generating study words from cues, they encode more *item-specific information* about each item independently, which benefits

Tanya R. Jonker and Colin M. MacLeod, Department of Psychology, University of Waterloo.

This research was supported by Discovery Grant A7459 from the Natural Sciences and Engineering Research Council of Canada (NSERC) to Colin M. MacLeod, as well as by an NSERC Vanier Canada Graduate Scholarship to Tanya R. Jonker. We thank Drs. Ori Friedman, Evan Risko, and William Hockley for feedback on Experiment 1 as part of the PhD dissertation of Tanya R. Jonker. We also thank Sai Kavapalle, Madalina Oancea, and Fiqir Worku for their assistance with data collection, and Iulia Banica, Arielle Levy, and Graham Schwoob for their assistance with data collection and coding.

Correspondence concerning this article should be addressed to Tanya R. Jonker, who is now at Center for Neuroscience, University of California, Davis, 1544 Newton Court, Davis, CA 95618. E-mail: tanyarjonker@gmail.com

memory for the individual item itself but not memory for order. In fact, atypical encoding tasks—even simple generic response tasks—appear to have a disruptive effect on relational encoding (Jonker & MacLeod, 2015).

Although it is thought that interitem associations play a key role in guiding memory for sequence (McDaniel & Bugg, 2008), the nature of these associations is not well understood. In the present study, to examine interitem associations in memory for order, we drew on the robust finding that silent reading promotes more relational encoding than does semantic judgment. Specifically, we examined whether memory that two events occurred nearby in time (i.e., X was temporally proximal to Y) is distinct from memory for the order in which those two events occurred (i.e., X occurred before or after Y). If order memory results from a single underlying representation, then encoding techniques that improve memory for one type of order (e.g., X was temporally proximal to Y) should also benefit other types of order memory (e.g., X occurred before Y). In other words, if order is intrinsic in interitem associations, then tests that separately probe associations and the order of associations should both reveal superior performance following silent reading compared to semantic judgment. Alternatively, if these types of memory result from different processes, then we might see diverging results following different encoding tasks.

There is some evidence supporting the possibility that the association between two items is distinct from the order of these associations. For example, during free recall of word lists, given that a participant has just recalled item i from a list, the next item output will most likely be $i + 1$ (Healey & Kahana, 2016; Kahana, 1996; Polyn et al., 2009; Sederberg et al., 2008); indeed, the subsequent recall of $i + 1$ is more probable than $i - 1$. From this result, one could conclude that the association between sequential items is asymmetric and biased in a forward direction. However, there are several problems with drawing this conclusion from free recall data, the most relevant of these being that free recall is guided not only by interitem associations, but also by item-context associations. According to the Temporal Context Model (Howard & Kahana, 2002) and the more generalized Context Maintenance and Retrieval Model (Polyn et al., 2009), the context associated with $i + 1$ includes the preexperimental knowledge of item i, such as semantic information, whereas the context associated with $i - 1$ does not include semantic information for i because $i - 1$ occurred prior to the presentation and activation of i and its semantic information. Therefore, when i is recalled and its context is reactivated, that context is more similar to that of $i + 1$ than that of $i - 1$, and so subsequent recall is biased to $i + 1$. Thus, when examining the order of associated items, it is preferable to use methods that rely on interitem associations alone, rather than both interitem association and item-context associations.

To examine memory for order, we designed two novel tests to probe different features of order memory. In our first experiment, we validated a recognition test of memory for interitem associations and replicated the previously observed pattern of superior memory for order following silent reading compared to semantic judgments (Jonker & MacLeod, 2015). Our novel test provided a conceptual replication of the robust results found using the order reconstruction test and laid the groundwork for the subsequent experiments. In Experiments 2 and 3, we isolated memory for the order of consecutively presented pairs of items. If memory for the order of items is intrinsic to interitem associations, then silent reading should result in superior performance on a test that probes memory for the order of two consecutive items. However, these experiments resulted in the opposite finding: Memory for the order of associated events was superior following semantic judgments, revealing a dissociation between interitem associations and the order of associated events. Therefore, in Experiment 4, we furthered our investigation by using a relational encoding task that involved explicit order judgments to examine the roles of item strength and directional encoding in memory for order.

Experiment 1

The aim of Experiment 1 was to validate a new test of interitem associations. Previously, the majority of the work examining memory for item order has relied on tests of order reconstruction. An order reconstruction test typically involves the presentation of the studied items in a random order and requires participants to reconstruct the studied order from the scrambled list. Although order reconstruction tests produce consistent results, relying predominantly on only one type of test severely limits the generalizability of the conclusions drawn from the experiments[1] and might lead to false conclusions about a single source of information underlying memory for order. There might be particular features of the order reconstruction test that produce the observed pattern of results such that changing features of the test might well lead to an entirely different pattern. Therefore, we designed our test to share important conceptual features with the order reconstruction test but to involve unique perceptual and contextual features.

Like the order reconstruction test, our novel test emphasized interitem associations. In our test, participants were shown a target word from the study list along with two other words from the study list and were to indicate which of these two words immediately followed the target word. However, our interitem association test differed from the order reconstruction test in a number of important ways. First, it involved only a subset of the study list (three items) rather than the entire list, eliminating the reinstatement of the entire study list; second, it placed no emphasis on where in the list the target word occurred (i.e., serial position); third, it involved a single trial with a single key press, which provided more control and allowed for more straightforward response-time analyses. Therefore, the primary goal of this experiment was to determine whether the superiority of order memory for silent reading relative to semantic judgment would replicate using a novel test of order memory.

To provide discriminant validity, Experiment 1 also employed a second test—a speeded semantic test, which was predicted to be

[1] Performance on an order reconstruction task has sometimes been linked to measures of order from free recall output protocols (for examples of order scoring methods, see Asch & Ebenholtz, 1962; Kahana, 1996). These two measures of memory for order typically yield converging results (e.g., Jonker, Levene, & Macleod, 2014; McDaniel & Bugg, 2008; Nairne, Riegler, & Serra, 1991), which suggests that the benefit of common encoding over uncommon and elaborative encoding is a reliable result. Free recall is, however, an uncontrolled method of testing memory for order and consequently it is difficult to determine whether measures of order reflect memory for relational information, context drift (Polyn, Norman, & Kahana, 2009), or some other mechanism. Therefore, as a controlled examination of memory for order, the order reconstruction test has been the primary method.

facilitated by semantic processing over silent reading. This prediction relies on the principle of transfer-appropriate processing, which is the idea that memory is superior when the processes engaged during encoding are similar to the processes engaged during test (Morris, Bransford, & Franks, 1977). If item-specific processing occurred during encoding, then responses on a test emphasizing item-specific information should be facilitated (i.e., faster); if relational processing occurred during encoding then responses on the novel interitem association test should be facilitated. As such, a crossover interaction in response times between encoding task and test performance was predicted.

Method

Participants. Participants in this and all subsequent experiments were recruited from the University of Waterloo; they participated in exchange for partial course credit and were eligible for the study only if they reported fluency in written and spoken English, normal or corrected-to-normal vision, and normal color vision. Experiment 1 consisted of 30 participants (9 male, 21 female) who reported an average age of 19.9 years.

Materials and procedure. A large set of 276 common nouns with word frequency scores lower than 500 was selected from the MRC psycholinguistic database (Coltheart, 1981). For each participant, 24 lists of eight words each were constructed by randomly selecting from the set of nouns, without replacement. The experiment consisted of 24 blocks, each with a study phase, a distractor task, and a test.

During study, participants were either to read words silently (in blue), or to indicate whether the word's referent was larger or smaller than the size of a microwave (in white). In the latter case, participants were to respond orally by saying "larger" or "smaller" and a research assistant was present to ensure compliance. Lists consisted of eight items presented serially for 2 s each with a 500-ms interstimulus interval. Following study of each list, participants performed a 30-s distractor task in which they were to make odd/even judgments on single digits. Memory for items from the study list was tested in one of two ways in each block—via an order test or via a size test (see Figure 1).

Interitem association test. For this test, participants were shown a single word in the center of the screen (i.e., the target word), along with two other words from the study list, one in each of the bottom left and bottom right corners of the screen. Of these two words, one had occurred immediately following the target in the study list (*helmet* in Figure 1), and the other had occurred four positions later (*jacket*). Participants were to indicate with a key press which of the two items had immediately followed the target during study; they were encouraged to respond as quickly and as accurately as possible. There were two test trials per block, and target words were always selected from Positions 2 and 4 of the study list, meaning that response options were from Positions 3 and 6 and 5 and 8, respectively. This was done to avoid recycling items within a block and to avoid the item from serial Position 1.

Unfortunately, an error in the experiment program resulted in participants always receiving one of these two serial positions first; in other words, 14 of the 30 participants always received the trial from Position 2 first, and the others always received the trial from Position 4 first. This programming error was corrected for Experiment 3 and the results replicated, suggesting that this unintentional consistency did not influence the results of Experiment 1.

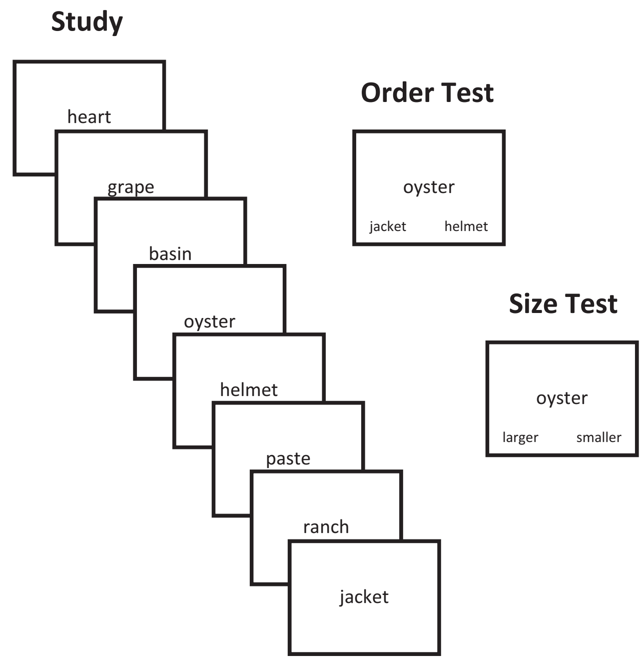

Figure 1. Example of the two different types of tests in Experiment 1. For the interitem association test, participants were to indicate which of the two items displayed at the bottom of the screen followed the target item displayed in the center of the screen. For the size test, participants were to indicate whether the target item is larger or smaller than an average chair.

Size test. For this test, participants were shown a single word from the study list in the center of the screen and they were to indicate with a key press whether that object is larger or smaller than an average chair. The surface features of the test were made to differ from the study features: The reference object differed (microwave vs. chair) and test responses were key presses (study responses were oral). There were eight test trials for the size test and all items from the study list were tested in a random order.

Each test phase began with a 2-s reminder to respond as quickly and as accurately as possible, and each test trial was preceded by a screen (2 s) with a question to cue the participant to the test type ("which came next?" for the order test, and "larger or smaller than a chair?" for the size test). There were 24 lists in this experiment, half allocated to each encoding type (silent reading, semantic judgment). For each set of 12 studied lists, four were followed by the size test and eight were followed by the order test. This was done to compensate for the fact that the size test involved more trials (8) than the order test (2), thus resulting in 32 data points for the size test and 16 data points for the order test for each encoding condition for each participant.

Prior to the 24 study-test blocks, to ensure understanding, participants completed a practice block involving each of the tasks under the guidance of a research assistant. The items used during practice did not recur in the experimental blocks.

Results and Discussion

The responses to the size judgment tests were subjective and therefore accuracy was not measureable for this test; instead, for this test, the dependent variable of interest was response time (RT) only. Thus, here we report accuracy results for the order test and a direct comparison of RT results for both the size and order tests.

Accuracy. Mean accuracy on the order tests was superior following silent reading compared to semantic judgments, $t(29) = 2.20$, $SE = .04$, $p = .04$, $d = 0.40$[2], shown in Figure 2, Panel A. This pattern replicates that produced on the order reconstruction test (Jonker & MacLeod, 2015), and demonstrates that participants were better able to recognize consecutive items after having encoded a list through silent reading than they were after having encoded a list while making semantic judgments. This in turn provides further support for the claim that reading silently leads to better order memory by establishing stronger interitem associations. Importantly, these results demonstrate that the item-order results generalize beyond the order reconstruction test.

Response time. The pattern observed in accuracy was also observed for RT on the interitem association test. RTs were included only for correct trials from the order test; all responses from the size test were included. Prior to analyzing response time data, anticipatory responses to test trials were removed (<400 ms), resulting in the removal of four trials from all order test data and one trial from all size test data. Furthermore, any responses that were 2.5 *SD*s slower than the mean for the relevant condition for each participant were removed. This resulted in the removal of 0.8% of the accurate trials from the order test (no more than one trial was removed as an outlier for any one participant); no data were removed from the size test. After these exclusions, one participant was excluded from the RT analysis because this participant had only two eligible trials in the semantic condition; all other participants had at least six trials per encoding condition for the order test ($M = 9.9$).

A 2 × 2 repeated-measures ANOVA assessed the effects of encoding condition (silent, semantic) and test type (order, size) on test RT. There was a main effect of test type, with faster responses overall to the size test than to the order test, $F(1, 28) = 96.40$, $MSE = 429644.84$, $p < .001$, $\eta_p^2 = .78$, but there was no main effect of encoding condition, $F(1, 28) = 0.21$, $MSE = 75127.36$, $p = .65$, $\eta_p^2 = 01$. However, as can be seen in Figure 2, Panel B, this absence of a main effect of encoding condition was due to a crossover interaction between encoding condition and test type, $F(1, 28) = 18.35$, $MSE = 59983.54$, $p < .001$, $\eta_p^2 = .40$. For correct responses on the order test, participants responded more quickly after having encoding through silent reading compared to size judgments, $t(28) = 2.52$, $SE = 86.61$, $p = .02$, $d = 0.47$, whereas on the size test, participants responded more quickly after having encoded by making size judgments compared to reading silently, $t(28) = 4.02$, $SE = 42.63$, $p < .001$, $d = 0.75$. Importantly, the fact that a difference was observed on the size test demonstrates that participants were indeed encoding semantically when prompted to do so. Therefore, any results on the interitem association test cannot be attributed to an encoding strategy that benefits silent reading only. Instead, the crossover interaction makes a compelling case for specialized information as a result of different types of encoding.

This experiment demonstrates that the order memory effect that is typically found using the order reconstruction test (silent reading > elaborative encoding) can also be found on other types of order tests, at least on one that emphasized interitem associations between successively studied items. This result extends the generalizability of the pattern of results because the interitem association test used in the present experiment differs from the order reconstruction test in a number of ways. For example, our interitem association test places no emphasis on the serial position of the target word, whereas an order reconstruction test requires that the participant place each item in its correct serial position. This experiment demonstrates that the order reconstruction results are not the product of participants' knowledge of generally where in the study list the item was presented (i.e., temporal memory) and instead supports the argument that order reconstruction is the result of superior memory for associations among consecutively presented items.

Experiment 2

The primary aim of this article was to examine whether relational encoding improves various facets of memory for order. Having replicated in Experiment 1 the finding that silent reading results in stronger interitem associations between successive items, we turned our attention to memory for the order of these associated items. That is, having established that participants remember consecutively presented items best following relational encoding, we sought to determine whether they also better recognize the *order* of two consecutively presented items following relational encoding.

If memory for associated items and memory for the order of associated items are one and the same, then encoding tasks that strengthen interitem associations (e.g., silent reading) should also produce superior performance on a test of the order of associated items. To examine this possibility, we presented participants at test with two items from the study list and asked them to indicate whether the two items were presented in the correct order (i.e., top item was studied prior to the bottom item). The test items could have been encoded back-to-back (i.e., consecutively) or they could have been separated by several items (i.e., spaced). If interitem associations are inherently ordered, then the superiority of silent reading over elaborative encoding should be moderated by the spacing of the two test items: Items presented consecutively during study are likely to be strongly associated because their representations were activated nearby in time, whereas items spaced apart during study are less likely to have strong interitem associations because of the time and interference between their presentations. In other words, silent reading should enhance performance on our order judgment test but this should be specific to—or more apparent for—test trials involving consecutive items. Put simply, we predicted an interaction: Silent reading should lead to superior order memory performance for consecutively studied items, but this effect would be attenuated, eliminated, or possibly even reversed for items separated during study. Observing this result would demonstrate that encoded interitem associations are inher-

[2] Cohen's d was calculated using an online calculator (http://www.cognitiveflexibility.org/effectsize/) based on Equation 8 from Morris and DeShon (2002).

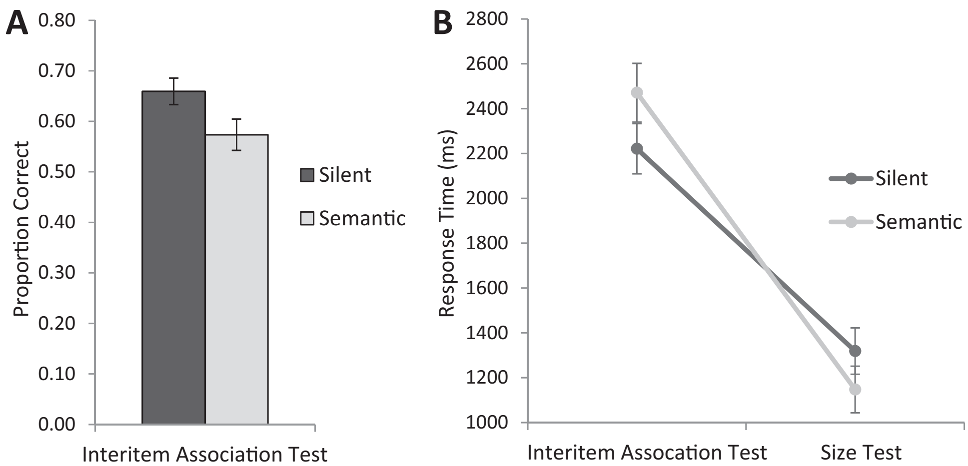

Figure 2. Experiment 1. Panel A displays the proportions of trials on which participants correctly identified the subsequently presented item on the order recognition test. Panel B displays the response times for correct trials from the order test and for all trials from the size test. Error bars represent one standard error of the mean.

ently ordered, such that participants can remember which of two test items appeared earlier.

Method

Participants. Thirty-six students participated in Experiment 2A and 38 students participated in Experiment 2B. Participants who responded at chance across all test trials (average performance < = .5) were excluded from analyses. Five were excluded from Experiment 2A, resulting in a final sample size of 31 participants (12 male, *M* age = 21.6) and four from Experiment 2B, resulting in a final sample size of 34 participants (11 male, *M* age = 19.9).

Materials and procedure. For each participant, 32 lists of eight words each were constructed by randomly selecting from the set of nouns described in Experiment 1. The experiment consisted of 32 blocks, each with a study phase, a distractor task, and a test. Words were not repeated across lists.

During study, participants were shown eight words, each for 3 s with a 250-ms interstimulus interval. In Experiment 2A, participants were instructed to make a pleasantness judgment (pleasant or unpleasant) for words that appeared in blue, and to silently read words that appeared in yellow. In Experiment 2B, participants were instructed to read blue words aloud, and to read yellow words silently. Half of the lists involved silent reading; the other half involved the item-specific encoding task (semantic judgment, reading aloud).

Following study of each list, participants performed a 15-s distractor task in which they were to make odd/even judgments on single digits. They were then given a single test trial. Two words from the most recent study list were presented vertically on the screen, and participants were to indicate whether those two studied words had occurred during study in the order shown on the test or in the reverse order (called "scrambled order" on the test; see Figure 3). Test items were selected from Positions 3 to 8 (Positions 1 and 2 were not used to avoid influence from the strong primacy effect). For the spaced tests, the earlier item was selected randomly from Positions 3, 4, or 5 and the later item was the item that occurred three positions later (i.e., Positions 6, 7, or 8, respectively). For consecutive tests, the earlier item was selected randomly from Positions 3 to 7, and the later item was selected from the subsequent position (i.e., 4 to 8). Half of the lists were tested with two consecutively studied words and half were tested with two spaced words. For each of these types, half involved presentation of the two test words in the correct order and half in reversed order, such that there were four lists for each possible combination

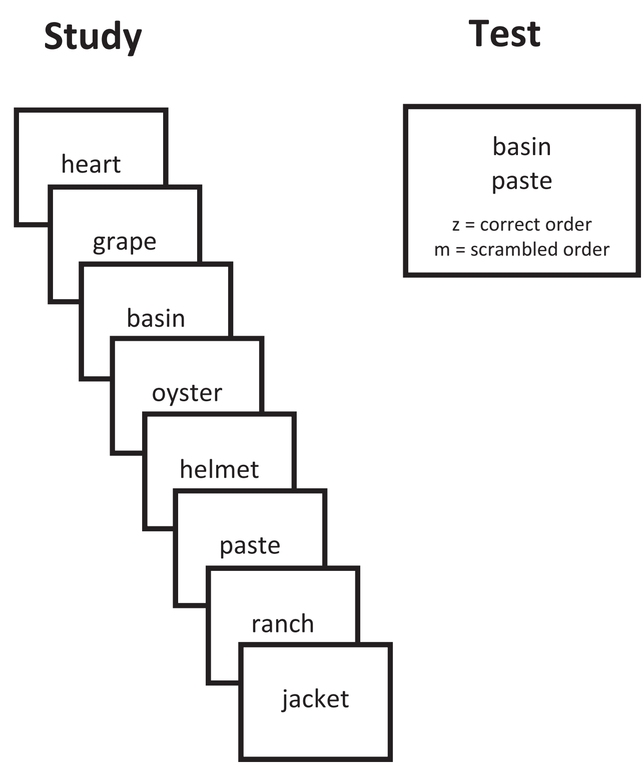

Figure 3. Example of the study and test from Experiment 2.

of encoding type (item-specific encoding vs. silent reading), spacing (consecutive vs. spaced), and test order (correct vs. reversed). The two test items remained on the screen until the participant pressed a key to indicate that the items were "shown in correct order" (z) or "shown in scrambled order" (m).

Prior to the 32 study-test blocks, to ensure understanding, participants completed a practice block involving each of the tasks under the guidance of a research assistant.

Results and Discussion

Experiment 2A. A 2 × 2 repeated-measures ANOVA assessed the effects of encoding task (semantic, silent reading) and test spacing (consecutive, spaced) on proportion correct on the order judgment test. The main effects of encoding task and test spacing were both significant, $F(1, 30) = 7.83$, $MSE = .03$, $p = .009$, $\eta_p^2 = .21$, and, $F(1, 30) = 17.30$, $MSE = .01$, $p < .001$, $\eta_p^2 = .37$, respectively. These differences are both readily apparent in Figure 4A. However, no significant interaction was observed, $F(1, 30) = 0.07$, $MSE = .01$, $p = .80$, $\eta_p^2 = .00$, demonstrating that the advantage of having made semantic judgments rather than reading silently did not differ when the test involved consecutive items, $t(30) = 1.93$, $SE = .04$, $p = .06$, $d = 0.35$, versus spaced items, $t(30) = 2.77$, $SE = .03$, $p = .01$, $d = 0.50$. The effect for consecutive items was only marginally significant, but the trend was in the direction opposite to that found in Experiment 1 and also opposite to that predicted.

The absence of any interaction suggests that silent reading does not enhance encoding of the relative order of consecutive items or spaced items. Thus, despite previous research having demonstrated that silent reading increases the encoding of interitem associations (e.g., Experiment 1 of the present work; Jonker & MacLeod, 2015), there is no evidence to suggest that these associations involve order. In other words, the fact that, following silent reading participants better remember *oyster* and *helmet* as having occurred together in time (i.e., an interitem association), does not necessitate that they also have better memory that *oyster* was presented prior to *helmet*.

Experiment 2B. A 2 × 2 repeated-measures ANOVA assessed the effects of encoding task (aloud, silent) and test spacing (consecutive, spaced) on proportion correct. As can be seen in Figure 4B, the main effect of encoding task was marginally significant, $F(1, 33) = 3.77$, $MSE = .03$, $p = .06$, $\eta_p^2 = .10$, and the effect of spacing was significant, $F(1, 33) = 9.50$, $MSE = .02$, $p = .004$, $\eta_p^2 = .22$. As in Experiment 2A, no significant interaction was observed, $F(1, 33) = 0.39$, $MSE = .02$, $p = .54$, $\eta_p^2 = .01$. Although the effects were smaller in Experiment 2B, possibly due to the pure list production effect being smaller than the levels of processing effect, the patterns observed in Experiment 2A were replicated in Experiment 2B using a different elaborative encoding task. This provides a conceptual replication of Experiment 2A.

Assessing the likelihood of the null hypothesis. The key finding in this experiment was the null interaction between encoding condition and spacing. To provide an assessment of the strength of the conclusion that the effect of encoding task on memory for order does not differ when items are consecutive versus spaced, a Bayesian approximation procedure (Wagenmakers, 2007) was applied to the results of the critical test of the encoding-condition-by-spacing interaction. For this analysis, the

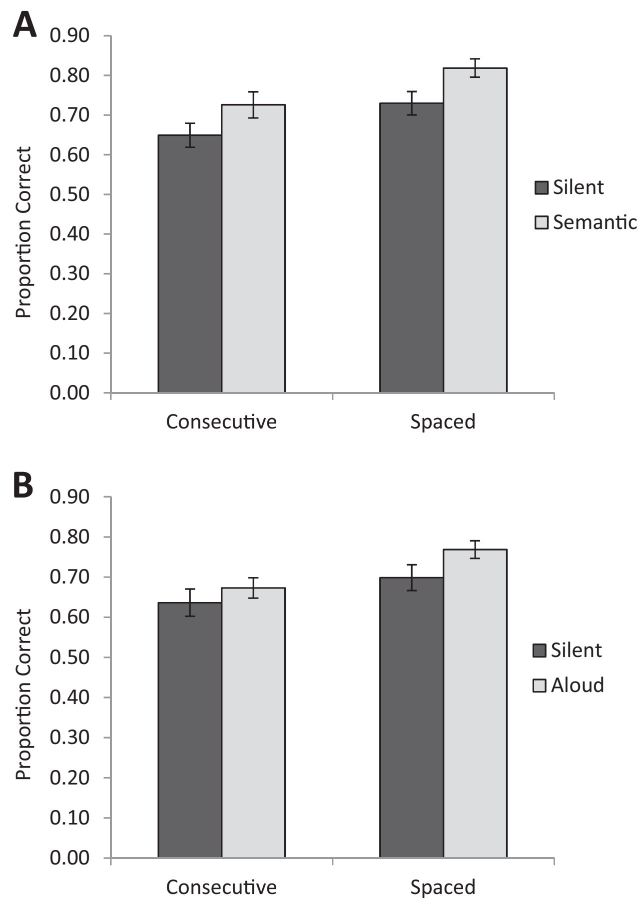

Figure 4. Experiment 2. Proportion correct on the order judgment test for (A) Experiment 2A and (B) Experiment 2B. Error bars reflect one standard error of the associated mean.

posterior odds of the null hypothesis over the alternative were estimated from the sums of squares (from a 2 × 2 ANOVA collapsing across experiment) using a calculator provided by Masson (2011). The posterior odds can be converted into p_{BIC}, which provides a measure of the support favoring the null on a scale of 0 to 1. According to Raftery's (1995) system for labeling strength of the evidence, the analysis yielded "positive" evidence in favor of the null hypothesis, $p_{BIC} = .88$, suggesting that a model assuming no interaction between encoding condition and spacing is preferred. In a second analysis, each experiment was entered sequentially, and the p_{BIC} result from Experiment 2A was used as the prior odd for Experiment 2B. This method produced "strong" evidence in favor of the null hypothesis, $p_{BIC} = .98$. Under either approach, the Bayes model estimation procedure demonstrates that the null hypothesis is more likely than the alternative, suggesting that there is no meaningful interaction between encoding condition and spacing.

Taken together, Experiments 2A and 2B demonstrate that memory for the order of associated items was superior following semantic encoding rather than following relational encoding. This surprising result occurred despite the fact that silent reading led to better memory for interitem associations (Experiment 1). Importantly, these results suggest that order memory is not the product of a single representation of sequence—if it were, then all tests of order memory should yield consistent patterns. Instead, memory for interitem associations and memory for the order of episodically associated events were found to be dissociable.

Experiment 3

A possible drawback of Experiments 2A and 2B is that each participant completed multiple study-test cycles involving only one test type. Thus, a potential explanation for the results of Experiment 2 is that participants somehow shifted their encoding strategy in anticipation of the order test, and that this new encoding strategy benefited performance following elaborative encoding. It is not clear what this new encoding strategy might entail, but it is a plausible explanation because each participant performed only one type of test in Experiment 2. To rule out this possibility, Experiment 3 combined the interitem association test of Experiment 1 with the order test of Experiment 2: Each study session was followed by either the interitem association test or the order test such that test type was not predictable. If encoding strategy underlay the surprising results of Experiment 2, then the results should not replicate in a case where the two test types occur within subjects and are not predictable.

Method

Participants. Of the 31 participants, four were excluded from analyses for responding at chance on the tests (average performance $< = .5$), resulting in a final sample size of 27 participants (8 male, *M* age = 20.3).

Materials and procedure. For each participant, 32 lists of eight words each were constructed by randomly selecting from the set of nouns described in Experiment 1. Participants completed 32 study-test blocks; half of the blocks involved silent reading (in blue), and half involved the semantic judgment described in Experiment 1 (in white). The study and distractor task parameters were identical to those described in Experiment 2. Of the 32 blocks, a random half involved the interitem association test described in Experiment 1, and the remaining half involved the order test described in Experiment 2. Because using both spaced and consecutive pairs would reduce the number of each type of test, decreasing power, only spaced pairs were used in Experiment 3. Spaced pairs were preferred over consecutive pairs because they produced a more consistent effect in Experiment 2.

After completing the 32 study-test blocks, participants completed a surprise recognition test for items from some of the blocks. The test involved 16 items of each encoding task-test type pairing (e.g., silent reading with order test), totaling 64 old items, along with 20 new items that had not occurred during the experiment. For each item, participants were to indicate whether it had been presented during the course of the experiment (i.e., old) or was new. The recognition test was included as a manipulation check to confirm that a levels-of-processing effect was observed.

Results and Discussion

On the final recognition test, semantic encoding resulted in better performance than did silent reading, $t(26) = 6.13$, $SE = .03$, $p < .001$, $d = 1.19$, confirming the effectiveness of the levels-of-processing manipulation (independent $M = .87$, silent $M = .68$).

To examine whether encoding task interacted with type of memory for order, a 2 × 2 repeated-measures ANOVA assessed the effects of encoding task (semantic, silent reading) and test type (interitem association test, order test) on proportion correct on the test. Trials on which participants pressed a key faster than 400 ms from onset were considered anticipatory responses and these trials were removed from the analysis. This resulted in the removal of four trials for one participant on the interitem association test, and one trial for a different participant on the order test. The analysis revealed no main effect of either encoding task or test type, $F(1, 26) = 0.61$, $MSE = .02$, $p = .61$, $\eta_p^2 = .01$, and, $F(1, 26) = 0.82$, $MSE = .03$, $p = .37$, $\eta_p^2 = .03$, respectively. The variables did, however, interact, $F(1, 26) = 19.52$, $MSE = .01$, $p < .001$, $\eta_p^2 = .42$, demonstrating that performance on different tests depended on whether study had involved making semantic judgments or reading silently. As shown in Figure 5, performance on the order test was superior following semantic encoding, $t(26) = 2.23$, $SE = .03$, $p = .04$, $d = 0.50$, whereas performance on the interitem association test was superior following silent reading, $t(26) = 2.55$, $SE = .04$, $p = .02$, $d = 0.49$. Thus, in an experiment where the predictability of the test was removed, the patterns observed in Experiments 1 and 2 were both reproduced.

The lack of superiority for silent reading on the order test was a surprising result. Common encoding tasks, such as silent reading and passive viewing, have consistently produced benefits on order reconstruction tests, yet the results of Experiments 2 and 3 reveal a type of order memory that does not benefit from common encoding: the order of associated items. Item-specific encoding enhanced order memory over and above silent reading. This dissociation suggests that performance on the order test might rely on a different source of information in memory. One possibility is item strength. In previous research, estimates of temporal information have been linked to the strength of the item, such that stronger items are perceived to have occurred more recently (Hintzman, 2005). That is, perhaps when participants are shown

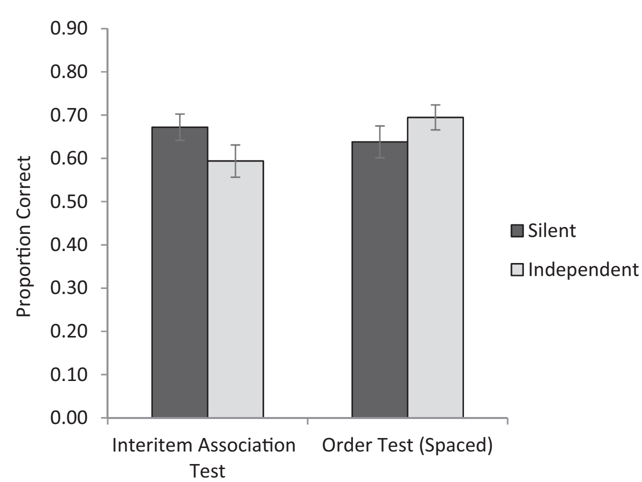

Figure 5. Experiment 3. Proportion correct on the interitem association test and the order test. Error bars reflect one standard error of the associated mean.

two items, they assess the strength of the memory trace of each and then deem the weaker trace to be the older item. In Experiments 2 and 3, item memory was stronger following semantic encoding compared to silent reading, which might have resulted in more accurate assessments of the older versus more recent item. Experiment 4 was designed to explore the possible roles of memory strength and relational encoding in determining memory for the order of associated items.

Experiment 4

In Experiment 4, we sought to identify the type of memory relied on for the order test. Informed by the findings of Experiments 1 to 3, we examined whether item strength alone was linked to performance on the order test, irrespective of whether items were strengthened through relational encoding versus item-specific encoding. To do so, we used an encoding task that we previously linked to superior order reconstruction performance (Jonker & MacLeod, 2015). This task involved a semantic judgment, which should strengthen item-specific memory, but critically this semantic judgment was relational in nature: Participants were to indicate the size of the object in relation to the previously presented item (“Is this object typically larger or smaller than the previously presented object?”). This judgment can be contrasted with a similar judgment made relative to a constant (i.e., “Is this object typically larger or smaller than a microwave?”). Importantly, both the former (a *relational* semantic judgment) and the latter (an *independent* semantic judgment) involve deep processing, which should strengthen memory for that item, but only the former involves relational encoding.

We had not before tested the degree to which our relational semantic task would improve item memory as measured by a recognition test. However, given that both the independent and relational semantic judgments involve activating the same item-specific information (i.e., size information), we hypothesized that item memory would be similarly strengthened by these two semantic encoding tasks—a hypothesis that was supported by the results of a surprise recognition test following all encoding blocks in the present experiment. Thus, these two tasks allowed us to determine whether relational encoding would have an effect on the order test when memory strength was equated.

If performance on the order test for both spaced and consecutively presented items is guided by item strength alone, then performance should be equivalent following both independent and relational semantic judgments because memory strength is equated. This pattern of results would demonstrate that the relational nature of judgments does not matter. Alternatively, it is possible that a task that explicitly emphasizes directional relations between items could benefit order memory in addition to item strength. If so, then performance on the order test should be superior following relational semantic judgments compared to independent semantic judgments because the only difference between these two encoding tasks was that the relational semantic judgments emphasized directional relations between list items, rather than the relation between a list item and a constant (i.e., a microwave). Importantly, this effect might be particularly evident for items that occurred in consecutive serial positions on the study list because the relational semantic judgment creates a unique directional relation between consecutive items (the subsequently presented item is always compared to the preceding item). The distinct direction of this encoding task could enhance memory for the order of the two items, which would benefit performance on consecutive trials on the order test more than would independent semantic judgments.

In summary, this experiment allowed us to test two main hypotheses. First, we could determine whether performance on the order test is dependent on item strength alone; if it is, then performance on the order test should not differ following relational versus independent semantic judgments. Second, we could determine whether directional encoding similarly affects consecutive and spaced order tests (i.e., an interaction). This would be informative with respect to understanding the type of relational information that affects memory for order. One possibility is that performance on the order test is affected by the encoded directional comparison of consecutive items; if this were true, then performance following relational semantic encoding should be superior only on the consecutive tests. Alternatively, it is possible that performance on the order test is affected by a more general sense of the temporal position of items, a process that might be enhanced with strong relational encoding; if this were true, then performance following relational semantic encoding should be superior on both consecutive and spaced tests.

Method

Participants. Two groups of participants were recruited for Experiments 4A and 4B. One participant was excluded from each experiment for chance responding on the tests (average performance $< = .5$); additionally, two participants were excluded from Experiment 4A for failing to comply with encoding instructions (complied on less than 75% of the encoding trials). This resulted in final sample sizes of 28 for Experiment 4A (12 male, M age = 22.2), and 29 for Experiment 4B (7 male, M age = 19.9; 3 participants did not report age).

Materials and procedure. For each participant, 32 lists of eight words each were constructed by randomly selecting from the set of nouns described in Experiment 1. The procedure was very similar to that of Experiment 3. Participants completed 32 study-test blocks, half involving the nonrelational semantic judgment described in Experiment 1 (in red), and the other half involving a relational semantic judgment in which participants were to determine whether the presented object is typically larger or smaller than the previously presented object (in blue). The study and distractor task parameters were identical to those described in Experiment 2. Of the 32 blocks, half involved the interitem association test and half involved the order test, as was the case in Experiment 3. Experiment 4A involved consecutive pairs, and Experiment 4B involved spaced pairs. Spacing was manipulated between-subjects to maintain a sufficient number of trials per condition. After completing the 32 study-test blocks, participants completed a surprise recognition test for items from some of the blocks. For each item, participants were to indicate whether it had been presented during the course of the experiment (i.e., old) or was new.

Results and Discussion

A 2×2 repeated-measures ANOVA assessing the effects of encoding task (relational, independent) and experiment (4A, 4B) on proportion correct on the final recognition test revealed no main effect of encoding task, $F(1, 55) = 2.79$, $MSE = .01$, $p = .10$, $\eta_p^2 = .05$, nor an interaction between encoding task and experiment, $F(1, 55) = 1.65$, $MSE = .01$, $p = .20$, $\eta_p^2 = .03$, demonstrating that the two semantic tasks strengthened item memory to a similar degree (independent $M = .87$, relational $M = .85$).

Trials on which participants produced a response in fewer than 400 ms were considered anticipatory responses and these trials were removed from the analysis. This resulted in the removal of six trials from each of the two tests.

To assess whether the superiority of relational encoding on the interitem association test replicated with our new relational encoding task, a 2 × 2 mixed-measures ANOVA assessed the effects of encoding task (relational, independent) and experiment (4A, 4B) on performance on the interitem association test. This analysis revealed that the effects observed on the interitem association test using silent reading in Experiments 1 and 3 were replicated using a semantic relational encoding task (see Figure 6): Participants were far more likely to correctly identify the item from the subsequent serial position following relational semantic encoding than following independent semantic encoding, $F(1, 55) = 29.47$, $MSE = .57$, $p < .001$, $\eta_p^2 = .35$. There was no interaction between encoding task and experiment, $F(1, 55) = 0.38$, $MSE = .01$, $p = .54$, $\eta_p^2 = .01$; an interaction was not expected because all features of the encoding and interitem association test tasks were identical across Experiments 4A and 4B.

To address the key question of whether relational encoding would improve memory on the order test given equal item strengthening, a 2 × 2 mixed-measures ANOVA assessed the effects of encoding task (independent, relational) and order type (consecutive, spaced) on proportion correct on the order test. There were marginal effects of encoding task, $F(1, 55) = 3.28$, $MSE = .02$, $p = .08$, $\eta_p^2 = .11$, and order type, $F(1, 55) = 2.96$, $MSE = .03$, $p = .09$, $\eta_p^2 = .05$, but these were superseded by a significant interaction between encoding task and spacing, $F(1, 55) = 6.43$, $MSE = .10$, $p = .01$. As can be seen in Figure 6, relational encoding resulted in superior performance when tested items had occurred consecutively in the encoding list, $t(27) = 2.89$, $SE = .03$, $p = .01$, $d = 0.55$. When they had been spaced, however, relational and independent semantic encoding did not differ, $t(28) = 0.55$, $SE = .03$, $p = .59$, $d = 0.10$.

When considering performance following independent semantic judgments, memory was superior for spaced pairs, $t(55) = 3.00$, $SE = .04$, $p = .004$, $d = 0.81$, replicating the effects found in Experiment 2 (consecutive < spaced). This is not a surprising finding given that it is likely easier to identify which of the two items in a spaced pair occurred earlier because they have more disparate strength information. Interestingly, however, there was no difference in performance on consecutive versus spaced pairs following relational judgments, $t(55) = 0.06$, $SE = .04$, $p = .95$, $d = 0.02$, suggesting that explicitly encoding a directional association uniquely improved memory for order for consecutively presented items.

These findings demonstrate two things. First, performance on the order test is influenced by item strength. A relational task that resulted in weaker item memory (i.e., silent reading) resulted in poorer performance on the order test when compared to item-specific encoding, but when a stronger relational task was used (i.e., relational semantic judgment), this difference was no longer observed. This demonstrates the importance of item memory when it comes to judging the relative order of two memories: The stronger the two memories, the more accurately one can assess which of the two occurred more recently. In other words, increasing strength increases the fidelity of relative order judgments (cf. Hintzman, 2005). However, this claim must be conditionalized by our second finding, and that is that directional encoding between list items—as in the case of the relational semantic judgment—results in strong memory for the order of those associated items, but this effect was limited to cases where the order was made explicit by the judgment, as spaced items did not benefit from relational semantic encoding.

General Discussion

In four experiments, we found informative dissociations among types of memory for order. Experiments 2 and 3 revealed that—contrary to previous assumptions—relational encoding does not always benefit memory for order. In fact, although in many previous studies silent reading has been shown to benefit performance on an order reconstruction test when compared to item-specific encoding tasks (e.g., Jonker, Levene, & MacLeod, 2014; Jonker & MacLeod, 2015; Mulligan, 2002), silent reading resulted in poorer performance on our novel order test in three separate samples (Experiments 2A, 2B, and 3). This was observed despite the fact that relational encoding (silent reading or relative semantic judgments) was found to reliably benefit memory for interitem asso-

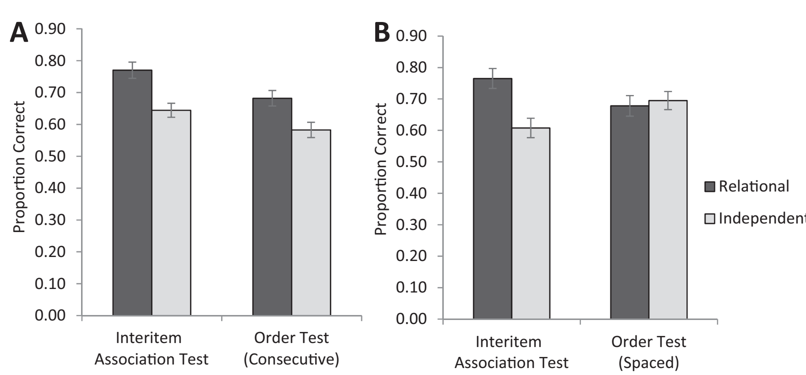

Figure 6. Experiments 4A and 4B. Proportion correct on the interitem association test and the order test. Error bars reflect one standard error of the associated mean.

ciations, a robust pattern observed in four separate samples (Experiments 1, 3, 4A, and 4B). This dissociation demonstrates that measures that probe seemingly similar representations (i.e., interitem associations vs. the order of associated items) can produce divergent results.

The results from the present experiments suggest that memory for order is the product of multiple sources of information. One of these sources is interitem associations, which were assessed independently of serial position using our interitem association test. Importantly, our interitem association test generalizes the pattern of results found using order reconstruction. Because the order reconstruction test requires resequencing of the entire list, it tests memory for both interitem associations and the order of those associations (Neath, 1997). The interitem association test used here allows one to circumvent these issues because successful performance requires no memory for the order of the associated list items.

A second type of information related to memory for order is item-specific information regarding the age of the memory. Research suggests that familiarity decays linearly over time (e.g., Yonelinas & Levy, 2002); people might use the familiarity of a memory to determine how old that memory is, with less familiar items being deemed older than more familiar ones. In support of this notion, Hintzman (2004, 2005) has demonstrated that stimuli that are more strongly encoded are also perceived as having occurred more recently. Thus, perhaps when people assess the relative order of two items—as with our order test—they independently assess the familiarity signal of each of the items and then contrast the signals, deeming the weaker one the older memory. In cases where this information is of higher quality—such as following semantic encoding—assessments of relative familiarity and therefore of order will be more accurate. This interpretation is supported by the novel result of Experiment 4: Here it was found that equating the strength of item memory resulted in equivalent performance on the order test, but only for items that had occurred in distant serial positions. The finding that performance on the order test was improved by increasing item strength demonstrates the important role of item strength for judgments of relative order.

A third source of information is directional relations encoded among items. In Experiment 4, we found that an encoding task that emphasized a specific directional relation among items (i.e., "Is this object typically larger or smaller than the previously presented object?") benefited memory on the order test, but this was the case for consecutive test trials only; spaced trials showed no such effect. It is interesting that these directional relations were not beneficial for order memory overall. The finding that they did not benefit memory on the consecutive test suggests that they did not enhance participants' temporal or serial-order memory in general, but instead that these types of encoded relations are only beneficial for cases where they are directly relevant to both items.

One thing that is unclear, however, from the present results is whether interitem associations encoded during silent reading and other passive tasks could also benefit performance on the order test of consecutive items. In other words, must the task include specific directional relations to benefit performance on the order test? We were not able to determine this from the present results because the case where we had interitem associations without explicit directionality (i.e., silent reading in Experiments 2A and 2B) was confounded with lower memory strength. Perhaps if memory strength for silently read items had been equated to that for semantic items, then we would have been able to determine whether interitem associations on their own—without an explicit directional relation—improve memory for relative order. However, it could be argued that if interitem associations in the absence of explicit directional encoding benefited memory for relative order of consecutive items, then an interaction between encoding task and consecutive versus spaced ordering should have been observed in Experiment 2. This argument could be made because the interitem associations would have uniquely benefited the consecutive items following silent reading, even despite lower memory strength overall. This is a tentative conclusion, however, and remains an open question requiring further research.

It is interesting to compare the results of our order test with those of a very similar test: the recency test. During a recency test, participants are presented with two items from the study list and are to indicate which item was studied most recently. A recency test has been previously used for studying design effects typically explained by the item-order account. For example, Greene, Thapar, and Westerman (1998) had participants study short lists of items, either by reading aloud or by generation from a cue. During test, participants were shown two words that had been studied four serial positions apart, and participants were to indicate which was shown later in the list. Using their method, they found no differences between words that had been read versus generated. Similarly, Mulligan (2000, 2001) found no differences on this test for items that had been masked with perceptual interference versus those that had not been masked, nor for high- versus low-frequency words. This was the case even though generation, perceptual interference, and low frequency all result in poorer order reconstruction relative to reading or high-frequency words (for a review, see McDaniel & Bugg, 2008).

Our outcome is surprising when contrasted with those of Mulligan and Greene and colleagues because, at face value, the tests are very similar: Both involve two study words, participants must detect relative order, and items are spaced apart a similar distance. Furthermore, these findings were reported in three different articles coming from two different labs, and ours replicated several times in the present article, suggesting robustness in both cases. There are a few methodological differences that might underlie the divergent results. One candidate is the presentation of items: In our experiments, test items were presented in a stacked orientation, which might facilitate fluency, especially when the top item had occurred earlier in the list compared to the bottom item. In the Greene et al. (1998) and Mulligan (2000, 2001) experiments, items were presented beside each other. Arguably, this orientation could also facilitate fluency as reading occurs left to right, so a direct comparison between these differing displays would be informative. A second candidate, and perhaps the more likely one, is the differences in orienting instructions: In our task, we had participants report whether the items were presented in the preserved order versus a scrambled order, whereas Greene et al. and Mulligan had participants identify the item that occurred more recently. It is possible that our instruction resulted in participants treating items together as a whole, whereas Greene et al.'s and Mulligan's instruction encouraged participants to contrast items. A final explanation could be that the recency test is less sensitive to differences in relative order than is our order recognition test. This is an interesting domain for future research.

Taken together, our experiments demonstrate that memory for the sequence of events can be influenced by (a) interitem associations, (b) the memory strength of an item independently of other items, and (c) the emphasized directionality of a relation. Thus, we conclude that memory for order could be the product of at least three codes, and tests that emphasize one of them over others will result in diverging patterns of results. If a test probes item-specific temporal information, then encoding techniques that enhance item-specific strength will yield superior memory; if a test probes interitem associations—as with the order reconstruction test—then relational encoding techniques will yield superior memory; and if a test probes relative order, then encoding techniques that emphasize directional relational encoding will yield superior memory. As a consequence, we encourage researchers to carefully consider their chosen tests of order memory. Controlled tests that target different types of order memory, such as the tests we have introduced here, will produce valuable insights into the sources of memory for order and the subtleties of how different encoding tasks influence retention of order information.

References

Asch, S. E., & Ebenholtz, S. M. (1962). The process of free recall: Evidence for non-associative factors in acquisition and retention. *The Journal of Psychology: Interdisciplinary and Applied, 54,* 3–31. http://dx.doi.org/10.1080/00223980.1962.9713093

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology, Section A: Human Experimental Psychology, 33,* 497–505. http://dx.doi.org/10.1080/14640748108400805

Ebbinghaus, H. (1885). *Memory: A contribution to experimental psychology*. New York, NY: Dover.

Ebenholtz, S. M. (1972). Serial learning and dimensional organization. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 267–314). New York, NY: Academic Press.

Greene, R. L., Thapar, A., & Westerman, D. L. (1998). Effects of generation on memory for order. *Journal of Memory and Language, 38,* 255–264. http://dx.doi.org/10.1006/jmla.1997.2552

Healey, M. K., & Kahana, M. J. (2016). A four-component model of age-related memory change. *Psychological Review, 123,* 23–69. http://dx.doi.org/10.1037/rev0000015

Hintzman, D. L. (2004). Time versus items in judgment of recency. *Memory & Cognition, 32,* 1298–1304. http://dx.doi.org/10.3758/BF03206320

Hintzman, D. L. (2005). Memory strength and recency judgments. *Psychonomic Bulletin & Review, 12,* 858–864. http://dx.doi.org/10.3758/BF03196777

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology, 46,* 269–299. http://dx.doi.org/10.1006/jmps.2001.1388

Jonker, T. R., Levene, M., & Macleod, C. M. (2014). Testing the item-order account of design effects using the production effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40,* 441–448. http://dx.doi.org/10.1037/a0034977

Jonker, T. R., & MacLeod, C. M. (2015). Disruption of relational processing underlies poor memory for order. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41,* 831–840. http://dx.doi.org/10.1037/xlm0000069

Jonker, T. R., Seli, P., & MacLeod, C. M. (2013). Putting retrieval-induced forgetting in context: An inhibition-free, context-based account. *Psychological Review, 120,* 852–872. http://dx.doi.org/10.1037/a0034246

Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition, 24,* 103–109. http://dx.doi.org/10.3758/BF03197276

Lashley, K. S. (1951). The problem of serial order in behavior. In A. Jeffress (Ed.), *Cerebral mechanisms in behavior* (pp. 112–136). New York, NY: Wiley.

Lewandowsky, S., & Murdock, B. B. (1989). Memory for serial order. *Psychological Review, 96,* 25–57. http://dx.doi.org/10.1037/0033-295X.96.1.25

Masson, M. E. J. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behavior Research Methods, 43,* 679–690. http://dx.doi.org/10.3758/s13428-010-0049-5

McDaniel, M. A., & Bugg, J. M. (2008). Instability in memory phenomena: A common puzzle and a unifying explanation. *Psychonomic Bulletin & Review, 15,* 237–255. http://dx.doi.org/10.3758/PBR.15.2.237

Mensink, G.-J., & Raaijmakers, J. G. (1988). A model for interference and forgetting. *Psychological Review, 95,* 434–455. http://dx.doi.org/10.1037/0033-295X.95.4.434

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16,* 519–533. http://dx.doi.org/10.1016/S0022-5371(77)80016-9

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7,* 105–125. http://dx.doi.org/10.1037/1082-989X.7.1.105

Mulligan, N. W. (2000). Perceptual interference and memory for order. *Journal of Memory and Language, 43,* 680–697. http://dx.doi.org/10.1006/jmla.2000.2721

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. http://dx.doi.org/10.3758/BF03195760

Mulligan, N. W. (2002). The generation effect: Dissociating enhanced item memory and disrupted order memory. *Memory & Cognition, 30,* 850–861. http://dx.doi.org/10.3758/BF03195771

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. http://dx.doi.org/10.1037/0278-7393.17.4.702

Neath, I. (1997). Modality, concreteness, and set-size effects in a free reconstruction of order task. *Memory & Cognition, 25,* 256–263. http://dx.doi.org/10.3758/BF03201116

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review, 116,* 129–156. http://dx.doi.org/10.1037/a0014420

Raaijmakers, J. G., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88,* 93–134. http://dx.doi.org/10.1037/0033-295X.88.2.93

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology, 25,* 111–163. http://dx.doi.org/10.2307/271063

Sahakyan, L., & Kelley, C. M. (2002). A contextual change account of the directed forgetting effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 1064–1072. http://dx.doi.org/10.1037/0278-7393.28.6.1064

Sederberg, P. B., Howard, M. W., & Kahana, M. J. (2008). A context-based theory of recency and contiguity in free recall. *Psychological Review, 115,* 893–912. http://dx.doi.org/10.1037/a0013396

Serra, M., & Nairne, J. S. (1993). Design controversies and the generation effect: Support for an item-order hypothesis. *Memory & Cognition, 21,* 34–40. http://dx.doi.org/10.3758/BF03211162

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM-retrieving effectively from memory. *Psychonomic Bulletin & Review, 4,* 145–166. http://dx.doi.org/10.3758/BF03209391

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review, 14,* 779–804. http://dx.doi.org/10.3758/BF03194105

Yonelinas, A. P., & Levy, B. J. (2002). Dissociating familiarity from recollection in human recognition memory: Different rates of forgetting over short retention intervals. *Psychonomic Bulletin & Review, 9,* 575–582. http://dx.doi.org/10.3758/BF03196315

Received January 1, 2016
Revision received April 26, 2016
Accepted May 15, 2016 ■