Order of items within associations

Kenichi Kato and Jeremy B. Caplan
Department of Psychology, University of Alberta

Abstract

Association-memory is a major focus of verbal memory research. However, experimental paradigms have only occasionally tested memory for the order of the constituent items (AB versus BA). Published models of association-memory, implicitly, make clear assumptions about whether associations are learned without order (e.g., convolution-based models) or with unambiguous order (e.g., matrix models). Seeking empirical data to test these assumptions, participants studied lists of word-pairs, and were tested with cued recall, associative recognition and constituent-order recognition. Order-recognition was well above chance, challenging strict convolution-based models, but only moderately coupled with association-memory. Convolution models are thus insufficient, needing an additional mechanism to infer constituent order, in a manner that is moderately correlated with association-memory. Current matrix models provide order, but over-predict the coupling of order- and association-memory. In a simulation, when we allowed for order to be wrongly encoded for some proportion of pairs, order-recognition could be decoupled from cued recall. This led to the prediction that participants should persist with their incorrect order judgement between initial and final order-recognition, but this was not supported by the data. These findings demand that current models be amended, to provide order-memory, while explaining how order can be ambiguous even when the association, itself, is remembered.

Keywords: associations; order; cued recall; associative recognition; mathematical models; verbal memory

Introduction

Association-memory has been a major focus of empirical and mathematical modelling studies of verbal memory, but has generally been studied separately from another topic considered important for behaviour, memory for order (e.g., Kahana, 2012; Lashley, 1951; Murdock, 1974; Neath & Surprenant, 2003). An important question, then, is whether

Corresponding author: Jeremy Caplan, jcaplan@ualberta.ca. Department of Psychology, Biological Sciences Building, University of Alberta, Edmonton, Alberta T6G 2E9, Canada, Tel: +1.780.492.5265, Fax: +1.780.492.1768.
The authors thank Michael Dawson, Clayton Dickson, Yang Liu and Tomi Ann Limcangco for helpful feedback on the manuscript. Supported by the Natural Sciences and Engineering Research Council of Canada.

associations are remembered with or without order. That is, after studying a pair such as croissant–coffee, can the participant determine that the studied pair was croissant–coffee, not coffee–croissant? The dominant behavioural paradigms used to quantify association-memory do not test memory for the order of the constituent items: In cued recall, one item is given as a cue and the other is requested as the response. To answer croissant ? ("forward" cue), the participant need only remember that croissant and coffee were paired together; if the wrong constituent-order is retrieved, the participant is at no disadvantage. Likewise, given the cued-recall probe coffee ? ("backward" cue), the participant still need only remember the pairing; constituent-order is irrelevant. Associative recognition, also used to test association-memory, typically includes "intact" probes, such as croissant–coffee, along with "rearranged" probes. If a second studied pair were apple–soup, the probe croissant–soup would be an example of a rearranged pair. Constituent-order is not explicitly tested with these two probe types because items remain in their original positions in both rearranged and intact probes.

This is a problem for the development of models of association-memory. As Rehani and Caplan (2011) noted, published models always need to adopt an assumption about how constituent-order is stored (or not stored), even though the authors of those models did not intend to make any predictions about memory for constituent-order. We dig deeper into existing models in the General Discussion, but here we illustrate the problem, contrasting two major mathematical operations that are at the heart of a large number of vector-models of association-memory: convolution and matrix outer-product. First, we note that we know of no published implementation of order-recognition in a model of association-memory. However, existing models do present obvious ways one might implement order-recognition.

In convolution-based models (Longuet-Higgins, 1968; Metcalfe Eich, 1982; Murdock, 1982; Plate, 1995), two item-vectors, $\mathbf{a}$ and $\mathbf{b}$ (column-vectors are depicted in boldface), are convolved together, denoted $\mathbf{a} \circledast \mathbf{b}$, before being added to a memory vector, $\mathbf{w}$. Because convolution, $\circledast$, is commutative, $\mathbf{a} \circledast \mathbf{b} = \mathbf{b} \circledast \mathbf{a}$. This means that after the association is stored, the model has no way to differentiate whether the pair was AB or BA. A model that stores associations only with convolution would, therefore, predict chance performance at judging constituent-order. As discussed below, prior results have suggested participants are above-chance on tests of constituent-order, as our results will also show. A pure convolution-based model must be rejected. However, given the success of convolution models at fitting a wide range of memory phenomena (e.g., Lewandowsky & Murdock, 1989; Metcalfe Eich, 1982; Murdock, 1982, 1995; Neath & Surprenant, 2003), it could be the case that convolution provides a good account of many association-memory phenomena, but that whenever constituent-order is needed, a different source of information is used. Admittedly less parsimonious, this leads to a specific prediction that we test in the present experiments: order-memory should be somewhat uncoupled from association-memory. That is, associations may be remembered without order, and possibly, order might be judged correctly even when the association cannot be remembered.

In contrast to convolution-based models, matrix models store associations by computing the outer product between two item-vectors, denoted $\mathbf{b}\mathbf{a}^{\mathrm{T}}$, where $^{\mathrm{T}}$ denotes the transpose operation, before being added to a memory matrix, M. Unlike convolution, the outer product is non-commutative: $\mathbf{b}\mathbf{a}^{\mathrm{T}} \neq \mathbf{a}\mathbf{b}^{\mathrm{T}}$. In fact, the forward and backward association are directly related to one another— one is the transpose of the other: $\mathbf{b}\mathbf{a}^{\mathrm{T}} = \left(\mathbf{a}\mathbf{b}^{\mathrm{T}}\right)^{\mathrm{T}}$. Because of this property, there are several ways, in the matrix-model framework, in which the order of constituent items might be distinguished. For example, multiplying a memory of one pair, $M = \mathbf{b}\mathbf{a}^{\mathrm{T}}$ from the right (Pike, 1984), $M\mathbf{a} \simeq \mathbf{b}$, but assuming $\mathbf{a}$ and $\mathbf{b}$ are dissimilar (very small dot product), probing in the opposite direction, $M\mathbf{b} \simeq 0$. Thus, assuming the model has access to this order information, one would predict that, if cued recall is successful, the model also has unambiguous knowledge of constituent-order. The matrix model also suggests that constituent-order and association-memory will be tightly coupled, and covary with one another both across pairs, and across participants, because order information is embedded within the association itself. This is in contrast to the modified convolution model, which implies independence. Pure matrix models may thus be insufficient if participants cannot always accurately judge constituent-order, whenever they successfully remember the association. Given the similar success of matrix-based models at fitting a wide range of phenomena (e.g., Anderson, 1970; Humphreys, Bain, & Pike, 1989; Pike, 1984; Willshaw, Buneman, & Longuet-Higgins, 1969), modifications to the basic operation of the matrix-model must be considered, as we elaborate in the General Discussion.

One could argue that the question of within-pair order has been overlooked by researchers because it does not correspond to an ecologically valid task. Indeed, our croissant–coffee example demonstrates that in many situations, constituent order is not important; one will receive both croissant and coffee, and any order (spatial or temporal) is acceptable. It is not difficult to come up with examples for which order does matter. For example, when first learning a person's name, note that first and last names are often drawn from different stimulus pools (e.g., Gordon Brown), in which case order may not need to be explicitly stored, but can be inferred from item-stimulus properties. Some names, however, are reversible (e.g., Simon Dennis versus Dennis Simon), in which case order must be explicitly stored. Compound words in English, which may be at the end of a continuum with novel associations (Caplan, Boulton, & Gagné, 2014), must be eventually learned with order, because they typically have a modifier–head relationship (e.g., Dressler, 2006). Thus, for example, a Jail–Bird means something different than a Bird–Jail; a Turtle-Neck must be something different than a Neck-Turtle. However, as we just showed, even models developed to explain association-memory for which order is irrelevant, implicitly lead to predictions about whether or not participants could perform well or poorly on order-judgement tests. Thus, our first goal was to measure order-memory ability when, during study, participants had no incentive to consider order, corresponding to the target-data models of association-memory have been tested on (refer to the Order–Ignore groups in all three experiments). Our second goal was to see if order-judgements would be improved if order were made relevant, by instructing participants to attend to order and testing them with order-recognition on each iteration of the task (Order–Attend groups in all three experiments).

We identified a handful of studies that shed some light on the question of memory of constituent-order. First, research based on the so-called "double-function list" procedure (Primoff, 1938) has provided evidence that, in an association-memory task, participants have some moderate ability to discriminate order within associations. In double-function lists, each left-hand item of one pair is a right-hand item of another pair (AB, BC, CD, . . .). Primoff (1938) found that the backward association (B→A) interfered with participants' ability to retrieve the forward association (B→C). Because participants were unable to completely rule out the backward association, memory for the order of constituents of a pair must not be perfect in that paradigm. Rehani and Caplan (2011) gave participants equal numbers of forward and backward cued-recall tests, of both double-function pairs and control pairs for which each item was present in only one pair, termed "single-function" pairs. If we assume the probability of recalling each associate (i.e., A, given B as the cue) were the same for double-function and for single-function pairs, the participant would presumably need to make a guess between the forward and backward associate if no order information were available. The prediction is that accuracy of double-function pairs should be one-half the accuracy of single-function pairs. If, at the other extreme, constituent-order were reliably stored (given that the association itself were stored), accuracy should be equivalent for double- and single-function pairs. In fact, accuracy was mid-way between these upper and lower bounds, suggesting that participants had some capacity to distinguish forward from backward associations, but imperfectly. It should be noted, however, that double-function pairs may have had one advantage over single-function pairs: each double-function item was presented twice, whereas for single-function pairs, each item was presented only once. It is possible that double-function pairs had greater item-memory, increasing the likelihood of retrieving the correct target item (cf. Criss, Aue, & Smith, 2011; Madan, Glaholt, & Caplan, 2010). If this item-memory advantage were large enough, it would inflate the level of double-function relative to single-function accuracy. Challenging this, Caplan, Rehani, and Andrews (2014) found that, in a similar paradigm that allowed participants to respond with both associates, accuracy was nearly identical for double-function as for single-function pairs, arguing against an item-memory advantage for double-function pairs. Still, the results from Rehani and Caplan (2011) are thus not entirely conclusive on the question of order-memory.

Also with a procedure based on paired-associate learning, Mandler, Rabinowitz, and Simon (1981) showed that, when asked to free-recall a list of pairs and report them in order when possible, participants were remarkably accurate at reconstructing constituent-order. This result suggests that, at least in some circumstances, constituent-order might be near-maximal.

A direct way to examine memory for constituent-order is with an order-recognition test— that is, asking participants to discriminate probes that present the paired items in the same order as in study (AB, "intact") or in the opposite order (BA, "reversed"). At least three groups have reported order-recognition data (Kounios, Smith, Yang, Bachman, & D'Esposito, 2001; Kounios, Bachman, Casasanto, Grossman, & Smith, 2003; Greene & Tussing, 2001; Yang et al., 2013), and have consistently shown that participants have above-chance ability to distinguish the constituent-order ($d' > 0$). Some dependence on experimental parameters is suggested; Yang et al. (2013) produced a relatively high d' value (2.27 for unrelated pairs in their experiment 3), similar to Kounios et al. (2003)[1] (d'=2.12 for their so-called "fused" pairs), whereas Greene and Tussing (2001) had lower values (0.59 for the unrelated pairs in their experiment 5). The range of performance levels may be due to numerous differences in the procedures used across these three studies. For example, Kounios et al. instructed their participants to fuse a pair into a single concept and Yang et al. used Chinese words that belonged to specific categories. These may have, for some reason, been easier to remember in order than Greene and Tussing's stimuli, which were English antonyms.

The Rehani and Caplan (2011) results have the advantage that the task was primarily an association-memory task, and order-memory would have been helpful in improving participants' accuracy. However, they never tested order-memory directly. The order-recognition studies (Greene & Tussing, 2001; Kounios et al., 2003; Yang et al., 2013) have the advantage that they tested order-memory directly; however, because association-memory was never demanded of participants, it was possible that participants in those experiments

[1]Because Kounios et al. (2003) did not report d' values, these were calculated from their reported mean accuracy values.

used a non-associative strategy to study for order-recognition tests (as may be the case in Experiment 3 here). To better inform mathematical models, one needs to be able to compare order-memory performance to both an empirical upper- and lower-limit. In addition, none of these prior studies tests the prediction of models such as the matrix model, that order-recognition will be accurate, given that the association is remembered. By testing pairs both with order-recognition and cued recall, we test this prediction. Across the three experiments presented here, we ask, what is the level of order-memory when participants study associations, and how do order- and association-memory relate to one another?

Finally, we consider a known property of word pairs, associative symmetry, which is diagnostic of models. Initially, associative symmetry referred to the finding that forward and backward cued recall accuracy were equivalent on average, which has been replicated many times (e.g., Asch & Ebenholtz, 1962; Horowitz, Brown, & Weissbluth, 1964; Kahana, 2002; Köhler, 1947; Murdock, 1962). This was thought to test the hypothesis that associations lacked direction, or were learned as some sort of Gestalt. However, Kahana (2002) noted that even a model with completely independent associations could produce equal forward and backward accuracy *on average*; thus, symmetry of mean accuracy does not speak to whether or not the underlying associations are direction-specific. Kahana proposed that one should test each pair with cued recall twice, known as "successive testing," using all combinations of cue directions on test 1 and test 2 (Figure 1). For the cases in which test 1 and test 2 differ in cue direction (forward/backward or backward/forward), the correlation between accuracy on test 1 and test 2 (quantified with Yule's Q; see Methods), should be very high, near 1, if associations are Gestalt-like (or bidirectional). In contrast, this correlation, which we call Q_{DIFF}, should be lower, near-zero, if forward and backward associations are learned independently of one another. Numerous studies have found this correlation, Q_{DIFF}, to be quite high and close to 1 (Caplan, Glaholt, & McIntosh, 2006; Caplan, Rehani, & Andrews, 2014; Caplan, Boulton, & Gagné, 2014; Kahana, 2002; Madan et al., 2010; Rizzuto & Kahana, 2000, 2001; Rehani & Caplan, 2011; Sommer, Schoell, & Büchel, 2008), suggesting associations are bidirectional.

Although it may not help with evaluating Gestalt theory of association-memory (Caplan, Boulton, & Gagné, 2014), the finding of a high Q_{DIFF}, suggestive of associative symmetry, has implications for our consideration of models here. First, convolution-based models can only produce a high Q_{DIFF}, since forward and backward cued recall test the same stored association. The same variability in encoding strength influences forward and backward cued recall. A matrix model that stores forward and backward associations separately would, without further assumptions, produce independent forward and backward cued recall performance, because different (independent) encoding strengths would influence forward and backward cued recall of a single association. However, to accommodate the high Q_{DIFF} finding, matrix models must assume that the encoded forward and backward association strengths must be highly correlated, even if, in principle, they could be independent, as demonstrated by Rizzuto and Kahana (2000, 2001). Some matrix models would imply a tradeoff: order-memory may come at a cost of associative symmetry. That is, the better a participant is able to distinguish constituent-order, the more associative symmetry may need to be reduced, a prediction we test here. On the other hand, specific modifications to the matrix model, for example, expanding on Rizzuto and Kahana's model, could accommodate high performance on order-recognition without compromising Q_{DIFF}, an alternate outcome we also test for here. Thus, in the three experiments presented here, we included successive cued recall tests, to enable us to measure Q_{DIFF} and test if increased order-memory results in a breakdown of associative symmetry.

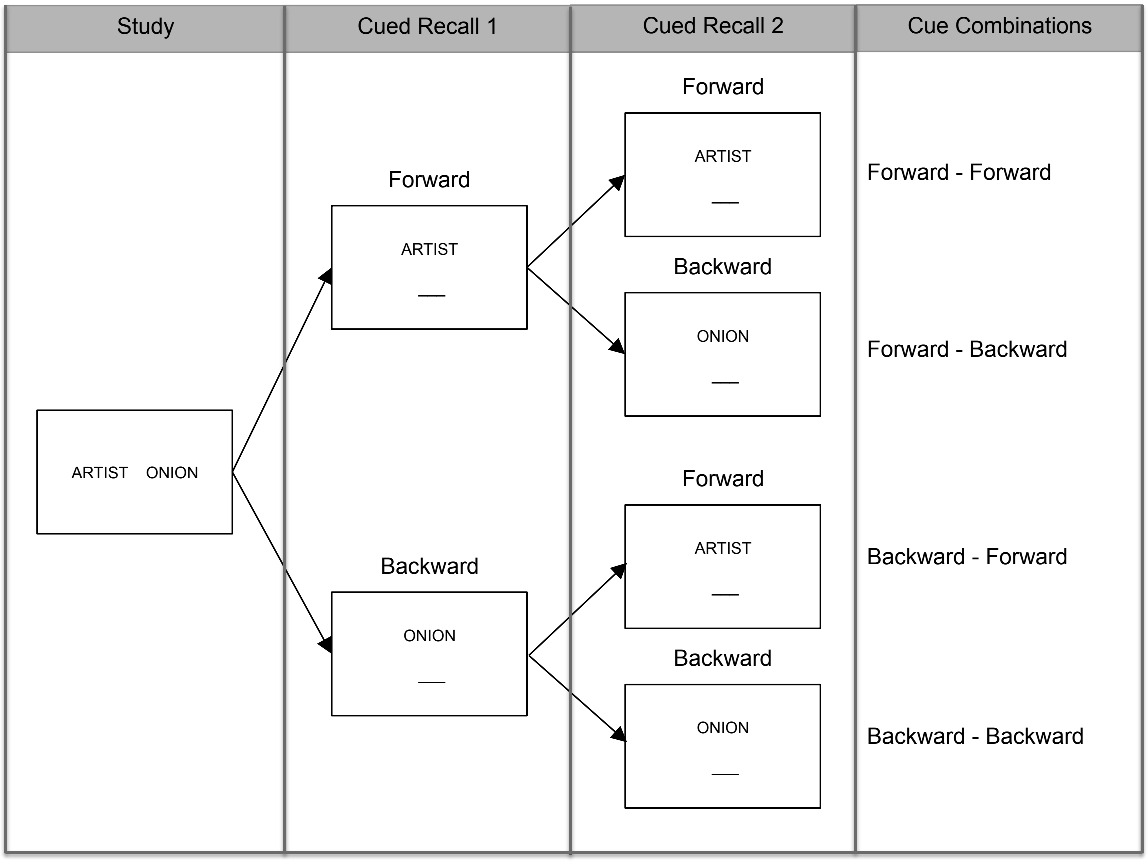


Figure 1. Illustration of cued recall with forward and backward probes and successive testing. Each pair is tested twice: once in test set 1 and a second time in test set 2. The direction of probe can be the same (Forward - Forward, Backward - Backward) or different (Forward - Backward, Backward - Forward). All combinations are illustrated here for one hypothetical pair, ARTIST–ONION.

As an aside, there are many types of information that may need to be coded as constituent order within associations. In our examples, and the procedure in all three experiments, order is presented as spatial, left-to-right order. Given that participants typically read English from left to right, it is plausible that the left-to-right arrangement results in a temporal order as well. Other order information may arise in other ways, such as first/last names and modifier–head relationships as noted in the previous examples. All these sources of "order" between constituent items have one thing in common: not only must pairings be remembered, but the participant must also have some way to break the symmetry and distinguish AB from BA. However, it is also plausible that order memory is learned and judged differently when it is spatial versus temporal versus functional, and future studies should investigate this. Here, we focus only on word-pairs presented left-to-right, and judged based on their left-to-right spatial arrangement, with the possibility that temporal, reading order could be the primary source of order-information in our tasks. Previous studies have found little difference in associative symmetry (differences in mean performance, as well as the correlation between forward and backward probes) between simultaneous and sequential presentation, where words are presented centrally, providing no spatial-order information (e.g., compare Experiments 1 and 2 in Madan et al., 2010). Thus, our findings may generalize to pure temporal-order, but this remains to be tested in future studies using sequential presentation.

We present data from three experiments, with the goal of obtaining empirical data to evaluate the assumptions of existing models, in three ways: by providing empirical measures of within-pair order-memory ability, by determining whether participants can willingly optimize their order-memory while studying pairs, and by characterizing the relationship between association-memory and within-pair order-memory. In the General Discussion, we consider how our findings speak to existing models of association-memory.

In Experiment 1, we designed a procedure such that the primary goal for participants was to learn associations as in a conventional cued-recall experiment (Figure 2). After being tested with cued recall, all pairs were then tested with either order-recognition or associative recognition. By placing cued recall first, and including two full sets of cued recall (although the primary reason for successive testing was to test associative symmetry), we hoped participants would treat association-learning as their primary task. This would address our concern: that participants might find a shortcut that would enable them to perform well on order-recognition, but would have compromised association-memory, as might have been the case in prior order-recognition studies (Greene & Tussing, 2001; Kounios et al., 2001; Yang et al., 2013). The initial recognition test was a between-subjects manipulation. Group "Order-Attend" had initial order-recognition, and group "Order-Ignore" had initial associative recognition. To amplify the group manipulation, the Order-Attend participants were warned that they would be tested on the constituent-order, whereas for the Order-Ignore group, the instructions made no mention of constituent-order. The hope was that Order-Attend participants would tune their study strategy to maximize their order-recognition ability, while Order-Ignore participants would study as in a typical association-memory experiment. After several cycles (each with a new set of pairs) of study, cued recall and order- or associative recognition, participants had an unanticipated set of final recognition tests, testing all studied pairs. Half the participants in each group had final order-recognition, and the remaining participants had final associative recognition. That is, the test types, order- versus associative recognition, over initial and final recognition were a 2×2 between-subjects manipulation (Table 1). This enabled us to test the effect of the intention to study for order (Order-Attend versus Order-Ignore) on final order-recognition.

Effects of initial cued-recall direction on order-recognition in Experiment 1 led us to withhold half the pairs of each list from cued recall and initial recognition tests in Experiment 2, to assess how testing effects might have influenced the results. Finally, in Experiment 3, we included the same Order-Attend versus Order-Ignore manipulation, but omitted cued recall from the initial tests, to ask if participants are able to adjust their study strategy to optimize their order-memory when association-memory is not a concern.

Experiment 1

Refer to Table 1 and Figures 1 and 2 for illustrations of the experimental design.

Methods

Participants. Participants were 74 undergraduate students enrolled in an introductory psychology course at the University of Alberta, who participated in exchange for partial course credit. For all experiments, sample sizes were not determined *a priori*, because the expected magnitudes of the main effects were not straight-forward to estimate.

Main Group	Recognition Test		N		
	Initial	Final	Exp. 1	Exp. 2	Exp. 3
Order-Attend	Order	Order	17	59	30
		Associative	19	55	30
Order-Ignore	Associative	Order	18	58	33
		Associative	16	56	31

Table 1
Sample sizes for the four sub-groups in Experiments 1, 2 and 3. Order-Attend group had the order-recognition whereas Order-Ignore group had associative recognition in the initial recognition test. In the final recognition phase, about half of each group had order-recognition whereas the remaining participants had associative recognition. Order - order-recognition; Associative - associative recognition.

Rather, data-collection proceeded as experimenter time permitted and participants signed up, and more participants were sought for Experiment 2, for which sensitivity was expected to be reduced, due to half the pairs being withheld from initial tests. To address concerns about sample sizes having been too large or too small, we carefully consider effect sizes, avoid over-interpreting small but significant effects, and pair classical analyses with Bayesian follow-up analyses to check null effects. Participants in all experiments were required to have English as their first language and had normal or corrected-to-normal vision. Sex and age were not collected in any of the experiments reported here. The procedures for all experiments were approved by a University of Alberta's ethical review board. Data from four participants were excluded from the analyses due to <10% cued-recall accuracy, which makes the calculation of Q unstable, leaving $N = 70$.

Groups. Assignment to groups in all experiments was arbitrary; groups were assigned by room number in a facility with fifteen testing rooms, and participants arrived and chose a testing room without knowledge of condition. Due to no-shows, sample sizes are not precisely equal across groups. Participants were divided into two main groups and each main group into two subgroups (Table 1): Participants in the Order-Attend main group were instructed to pay attention to the order of words within a pair and that they should expect to be tested on order. By the end of the practice cycle (see below), these participants would have experienced initial order-recognition tests, reinforcing the instruction. Participants in the Order-Ignore main group were not given any instructions about order, and were given associative recognition rather than order-recognition tests of each list. These two main groups were further subdivided into four sub-groups; after six full cycles of the procedure, about half the participants in each group responded to order-recognition tests of all studied pairs, and the remaining participants, associative recognition. The final recognition test was not mentioned before it occurred.

Materials. Stimuli were 478 nouns from the Toronto Word Pool (Friendly, Franklin, Hoffman, & Rubin, 1982), ranging from four to eight letters in length. Words were always assigned to pairs and lists pseudo-randomly (i.e., by the computer's random-number generator) for each participant. Study pairs and recognition test probes were presented in the center of the computer screen in capital letters, in the Courbd (Courier bold) font, with one word to the left of center and the other word to the right of center. Each cued recall probe was a single word presented centrally and a response field mark with an underline, centered just below the cue word. The experiment was run in Python in conjunction with the Python Experiment-Programming Library (Geller, Schleifer, Sederberg, Jacobs, & Kahana, 2007).

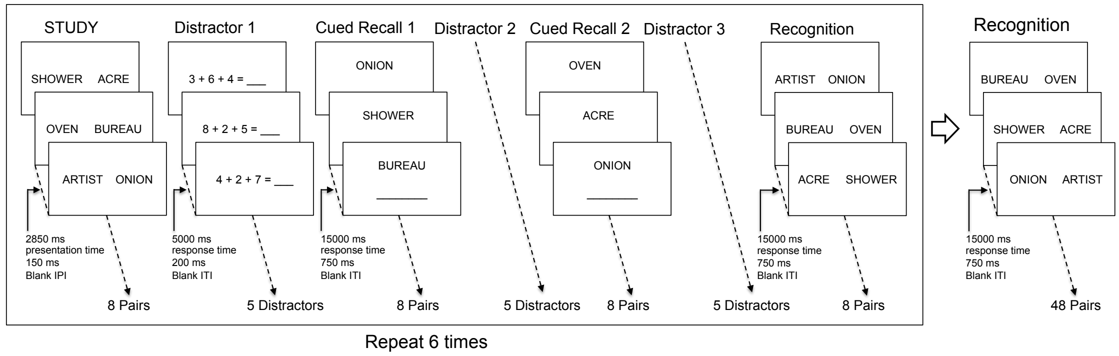


Figure 2. Schematic depiction of the experimental procedure in Experiment 1. For each list, the participant studied eight word pairs, completed two full sets of cued recall, followed by an initial recognition test (order-recognition or associative recognition, for Order-Attend and Order-Ignore, respectively) with trials of the distractor task interleaved between tasks. After six cycles of this procedure, each time with a study set comprised of new words, the participant had a final recognition test of all studied pairs ($48 = 8$ pairs $\times$ 6 lists), again, either order-recognition or associative recognition, depending on group.

Procedure. Illustrated in Figure 2, there were six cycles, where each cycle started with a new study set comprised of new words, consisting of a study phase, two successive sets of cued-recall tests, and an initial recognition test, with a distractor task interleaved between all pairs of tasks. Following those six cycles, participants had a final recognition test of all the studied pairs. About half the participants in each group had a final order-recognition and the remaining participants had final associative recognition test (Table 1).

Distractor. The distractor task consisted of five trials in which the participant had to calculate the sum of three single digits randomly drawn from 2 to 8, with a fixed response interval of 5000 ms and a 200-ms inter-trial interval (ITI). The question was displayed centrally on the screen, and participants typed the answer, pressing the Enter key to submit their response. Upon pressing enter, the colour of the response digits changed to signal to the participants that their response registered, but the procedure still waited until the 5000 ms were up before proceeding.

Study phase. Participants viewed the eight pairs of each list in sequence. The two words in a pair were horizontally presented (side by side) in the center of the screen for 2850 ms, with a 150-ms ITI.

Cued recall. Each list was tested with one set of cued-recall questions in which each pair was tested once, followed by a block of distractor trials and another complete set of cued-recall trials (Figure 2). Cue directions were counter-balanced over the trials within a test set. For the two successive cued-recall tests of a given pair, the directions of the cued-recall tests consisted of four combinations, Forward-Forward, Forward-Backward, Backward-Forward, and Backward-Backward (Figure 1). These successive-testing conditions were pseudo-randomly assigned, and test order was also pseudo-random. The cue word was presented in the center of the screen and an underline representing the area for the response was also presented just below the cue word. Thus, the cue did not indicate cue-direction. The typed letters appeared above the underline as the participant pressed each key. After pressing ENTER to indicate the end of a response, the next cued-recall trial started 750 ms later. The ENTER key was ignored until the participant typed more than two letters, to prevent participants from speeding through the experiment. If 15,000 ms passed without the ENTER key being pressed, the trial was automatically ended and the next trial started; such trials were scored as incorrect in all experiments. Recalls were considered correct only if spelled correctly.[2]

Initial recognition. Initial recognition was either order-recognition (Order-Attend group) or associative recognition (Order-Ignore group). The two probe words were presented horizontally (side by side) in the center of the screen as in the study phase. The number key, 1, was assigned to intact, and 2 to lure (reverse or rearranged). Other key presses were ignored. The number of intact trials and lure trials were counter-balanced. Rearranged pairs were only rearranged with other pairs within the current list, and a pair probed with an intact probe was never used to create a rearranged probe. The presentation order of the test pairs was pseudo-random. The trial automatically ended after 15,000 ms if neither key was pressed; such trials were scored as incorrect in all experiments. The next recognition trial began 750 ms later.

Final recognition. The final recognition procedure was the same as for initial recognition, except that the participants were tested on all the studied pairs over the six blocks at once. The rearranged probes could be rearranged from pairs regardless of which list they came from. The trial automatically ended after 15,000 ms if neither key was pressed; such trials were scored as incorrect.

Practice list. At the very start of the session, participants had one practice list to familiarize themselves with all the tasks, with the same materials and procedures as in the main experiment, except for the final recognition test. This list was excluded from the analyses.

Data analysis. Our chief measure of performance on order- and associative recognition tests was $d' = z(\text{hit rate}) - z(\text{false alarm rate})$. To avoid infinities, whenever the hit rate or false alarm rate were zero or one, one-half an observation was added or subtracted, respectively (Macmillan & Kaplan, 1985). To test whether performance was near-maximal when pairs were correctly stored, d' was also re-calculated only including pairs for which both cued-recall tests were correct. Because of the correction for zeroes and ones, the expected maximum d' depends on the number of trials included, and thus, was not constant across participants. For each participant, d'_{max} was calculated, and the observed d' values were compared with those maximum values with paired t-tests.

Our measure of correlation for dichotomous measures was Yule's Q (Bishop, Fienberg, & Holland, 1975). Q is calculated from the 2×2 contingency table of outcomes between two tests. Cell a tallies the number of times test 1 and test 2 were both correct; cell b tallies incidences of test 1 correct and test 2 incorrect; cell c tallies incidences of test 1 incorrect and test 2 correct; and cell d tallies incidences of both tests incorrect. Then, $Q = (ad - bc)/(ad + bc)$. Similar to Pearson correlation, Q ranges from -1 to $+1$. $Q > 0$ indicates the two tests are positively correlated, $Q = 0$ indicates independence, and $Q < 0$ indicates a negative correlation.

We tested associative symmetry following Kahana (2002). Our main interest is in the value of Q_{DIFF}, computed for the cases in which cued recall switches direction between

[2]Spelling errors were ignored because they were relatively infrequent. To estimate these, for each cued-recall trial, we computed the generalized Levenshtein distance (using the R function adist.r from the utils library) between the response and the correct spelling of what would be the accurate response. As an estimate of the upper limit of the number of misspelled words that might be considered correct, we used a threshold of a distance of 2. With this criterion, the proportions of responses suspected to be correct but misspelled were, M (SD): 0.033 (0.023), 0.027 (0.028) and 0.017 (0.021) for experiments 1, 2, and 3, respectively.

the first and second test. A high Q_{DIFF} would suggest associative symmetry, indicating that forward and backward cues test the same underlying variability in memory. A low value of Q_{DIFF} would suggest a breakdown of associative symmetry, as it would indicate that forward and backward cues test different sources of variability in memory. Computing Yule's Q for the "same" direction cases gives us Q_{SAME}, which estimates the correlation due to test/re-test reliability, and places an upper-limit on Q_{DIFF}. Finally, to obtain an empirical estimate of the lowest correlation value expected between memory tests that are presumably independent, we compute a bootstrap calculation, $Q_{CONTROL}$ (Caplan, 2005). $Q_{CONTROL}$ is measured by re-pairing test 1 and test 2 of different pairs within a given list, using all combinations of test 1 and test 2 within a given list of pairs, excluding cases for which both tests are of the same pair. $Q_{CONTROL}$ is expected to be somewhat positive due to Simpson's Paradox (Hintzman, 1980). That is, because a participant may perform better on some lists than others, all pairs of memory tests will tend to covary somewhat positively with one another.

Finally, to determine how null effects of interest should be interpreted, we used JASP (JASP Team, 2016) to evaluate the corresponding t test or ANOVA, always assuming uniform prior probabilities. For t tests, the Bayes Factor is a ratio of evidence, where by convention, when $BF_{10} > 3$, the difference is considered supported, and when $BF_{10} < 0.3$, the difference should be considered to be not present. For ANOVAs, we report the Bayes factors known as $BF_{inclusion}$, which summarizes across all factorial models and quantifies whether each model fits better with the effect (each particular main effect or interaction) included versus excluded. By convention, when $BF_{inclusion} > 3$, the effect is supported, and when $BF_{inclusion} < 0.3$, the effect is considered to be not supported— i.e., the model fits better without the effect. Our choice to report $BF_{inclusion}$ is because these values parallel the way that classical ANOVAs are reported: one verdict for each effect.

Results

Due to the complexity of the experimental design, we first report a comprehensive set of analyses of the data here, and consider and interpret the most pertinent results in the Discussion section. First we check whether studying for order (Order-Attend group) compromised cued-recall accuracy, symmetry (equal accuracy for forward and backward probes on average) or the property of associative symmetry (high correlation between forward and backward cued recall, Q_{DIFF}) compared to studying pairs when order was not relevant (Order-Ignore group). Next, we quantify performance on initial order-recognition asking whether performance is above chance and below ceiling. We test the relationship between order-recognition and cued recall, to provide evidence as to whether constituent-order might be encoded within the association itself. Finally, we evaluate performance on the final recognition test to examine whether the group manipulation might have affected the level of association and order memory learned during study, and perform additional tests of the relationship between order-recognition and prior association-memory performance.

Cued recall. A three-way, mixed, repeated-measures ANOVA on cued-recall accuracy (Figure 3), with design Group (Order-Attend, Order-Ignore) × Test (1, 2) × Cue Direction (forward, backward) found only a significant main effect of Test, $F(1, 68) = 41.63$, $MSE = 0.002$, $p < 0.0001$, $\eta_p^2 = 0.38$, with a small advantage for test 2. For all other effects, $F < 1$. To check the null main effect and interactions involving Group, we ran a Bayesian ANOVA, with the same design, in JASP (see methods). The main effect of Test had a large Bayes Factor for inclusion, $BF_{inclusion} = 10.7$, providing clear sup-

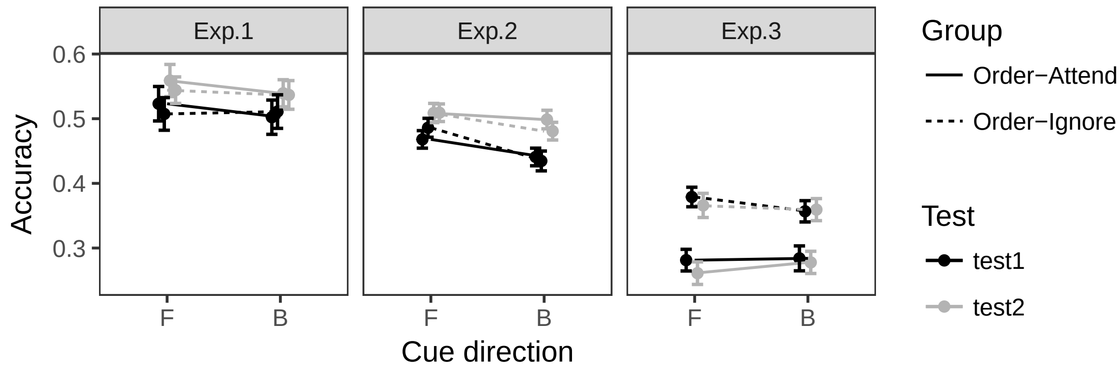


Figure 3. Mean accuracy of cued recall as a function of successive-testing set (test), cue direction (F - forward, B - backward), and group (Order-Attend, Order-Ignore). Note that in Experiments 1 and 2, cued recall was the first test of each studied list, whereas in Experiment 3, cued recall came at the end of the session, after all lists had been studied and tested. Error bars plot 95% confidence intervals based on standard error of the mean, corrected for between-subjects variability (Loftus & Masson, 1994). Note that the x values have been staggered so that all means are visible.

port for this main effect being present. For all other effects, the null was clearly favoured ($BF_{inclusion} < 0.3$). This included the main effect of Group ($BF_{inclusion} = 0.21$) and all interactions involving Group ($BF_{inclusion} < 0.13$), as well as the main effect and interactions involving Direction.

To assess associative symmetry, we computed Q_{DIFF} (see methods), reflecting the correlation between successive cued-recall tests of a pair, for cases in which the direction changes between tests (Figure 4). Also plotted are two control correlations that set the upper and lower limit of the expected range of Q_{DIFF}. First, Q_{SAME} is computed for cases in which cue direction was the same in both tests (both forward or both backward), which estimates the maximum correlation expected if the two tests tested the same learned information. Second, $Q_{CONTROL}$ is a bootstrap formed by re-pairing different pairs from test 1 and test 2 (see methods) which estimates the minimum correlation expected if test 1 and test 2 tested unrelated information in memory. As is typical, Q_{SAME} was close to 1. Q_{DIFF} was less than Q_{SAME}, but closer to Q_{SAME} than to $Q_{CONTROL}$, in line with previously reported values that indicate a very high underlying correlation between forward and backward encoding strengths (Rizzuto & Kahana, 2001). A two-way, mixed, repeated-measures ANOVA on the log-odds-transformed Q values, with design Group (Order-Attend, Order-Ignore) $\times$ Cue Combination (SAME, DIFF), yielded a significant main effect of Cue Combination, $F(1, 68) = 137.9,\ MSE = 1.46,\ p < 0.0001,\ \eta_p^2 = 0.67$, but neither the main effect of Group, $F(1, 68) = 0.018,\ MSE = 1.46,\ p{>}0.5,\ \eta_p^2 < 0.0001$, nor the interaction, $F(1, 68) = 0.19,\ MSE = 1.46,\ p{>}0.5,\ \eta_p^2 = 0.0027$, were significant. Finally, a Bayesian t-test of Q_{DIFF} between Groups also favoured the null hypothesis, $BF_{10} = 0.26$. Thus, studying for order does not appear to undermine associative symmetry.

In sum, speaking to our main questions, studying for order (Order-Attend versus Order-Ignore) did not affect cued-recall accuracy, mean-symmetry, or associative symmetry (Q_{DIFF}).

Initial recognition. Table C1 lists proportion correct for targets and lures, as well as d' (sensitivity) and C (bias) for initial recognition. For order-recognition, mean $d' = 1.36$.

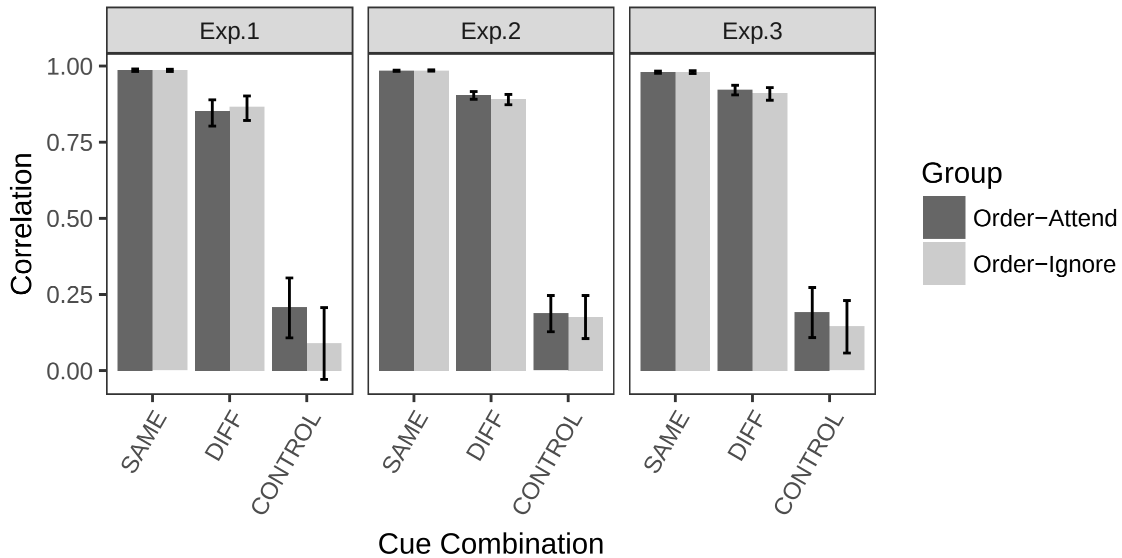


Figure 4. Correlations (Yule's Q) between successive cued-recall tests, as a function of groups, computed via the log-odds transform. The error bars represent 95% confidence intervals, which were computed via log-odds transform as well. SAME - Q_{SAME}, computed for cases in which a pair was tested in the same direction on test 1 and test 2; DIFF - Q_{DIFF}, computed for pairs for which cue direction changed from test 1 to test 2; CONTROL - $Q_{CONTROL}$, computed for test 1 and test 2 of different pairs (see main text for an explanation).

This value lies within the range that has previously been reported for order-recognition procedures that did not include initial cued recall (Greene & Tussing, 2001; Kounios et al., 2001, 2003; Yang et al., 2013). It was significantly greater than zero (chance),[3] $t(35) = 9.18,\ p < 0.0001,$ Cohen's $d = 1.53$, but also significantly less than maximal, d'_{max}=4.07 (see methods for details on how d'_{max} was calculated), $t(35) = 18.36,\ p < 0.0001,\ d = 3.06$.

Although d' was less than its maximum possible value, when the association cannot successfully be retrieved, order judgements may be much less accurate. The specific prediction implied by matrix models is that when the association *can* be retrieved, memory for constituent-order should be correct. We analyzed d' separately for pairs for which both cued-recall tests had already been responded correctly (denoted CC) and pairs for which both cued-recall tests had been incorrect (denoted II).[4] In an ANOVA on order-recognition d', with design Cued-Recall-Accuracy[CC,II]×Task[Order,Associative recognition], the main effects of Cued-Recall-Accuracy and Task were both significant,

[3]Negative values of d', although mathematically possible, would be strange, as $d' = 0$ is what one expects if the participant has no information to differentiate the two strength distributions. Because of this, one might argue that the test here should be one-tailed, which would increase our apparent confidence, which is already quite high. On the other hand, arguably, we should be comparing with an expectation of d' that is greater than zero. It is unclear to us exactly what that expectation should be, and given that a model with identical distributions would produce an expectation of $d' = 0$ with a symmetric distribution about that mean value, we note this concern but do not address it. Because the outcome of these t tests was a very robust rejection of the null in all cases, we judge this to be a minor concern.

[4]For associative recognition, this analysis was more complicated because each item in a rearranged probe is derived from a different pair, which could have had a different outcome in cued recall. As a quick check, d' was broken down in terms of cued-recall correctness of left-item versus right-item pairs. Luckily, this had little influence on the value of d', as is evident in Figure 6.

$F(1,65) = 109.7,\ MSE = 0.44,\ p < 0.0001,\ \eta_p^2 = 0.63$ and $F(1,65) = 44.0,\ MSE = 0.76,\ p < 0.0001,\ \eta_p^2 = 0.40$, respectively, as was the interaction, $F(1,65) = 14.9,\ MSE = 0.44,\ p < 0.0001,\ \eta_p^2 = 0.19$ (Figure 6). The interaction was explained by d' for CC minus d' for II pairs being significantly greater for associative recognition than for order-recognition, $t(65) = -3.84,\ p < 0.001$. For both order- and associative recognition, when the cued recall previously failed, d' was much lower (although still above chance). When cued recall succeeded, associative recognition was close to ceiling, although still significantly below d'_{max}, $t(33) = -4.67,\ p < 0.0001,\ d = 0.31$, but for order-recognition, d' for CC pairs was much lower than d'_{max}, $t(32) = -9.36,\ p < 0.0001,\ d = 1.96$, inconsistent with our matrix-model prediction.

As another test of the relationship between order-memory and association-memory, we asked if cued recall and initial recognition covaried across participants. If order information is incorporated into association-memory, performance in the order-recognition task would be expected to be highly correlated with performance in the cued-recall task. Alternatively, if order information is stored outside association-memory, then some participants might be better at learning order than others, and this skill might be somewhat unrelated (decoupled) from participants' ability to learn unordered associations. Thus, the correlation between order-recognition and cued recall would be expected to be relatively weak, and in particular, weaker than the correlation between associative recognition and cued recall. As shown in Figure C2, participants who performed better in cued recall tended to perform better in order-recognition (Order-Attend group), Pearson correlation, $r(35) = 0.71,\ p < 0.0001$. Although nominally weaker than the correlation between associative recognition and cued recall, $r(33) = 0.84,\ p < 0.0001$, these correlations were not significantly different according to a Williams' test, $z = 1.08,\ p > 0.1$. Because $z > 1$, this result should be considered inconclusive.

Testing effects. Because cued recall preceded initial recognition, we looked for effects of cued recall on initial-recognition. If testing direction in cued recall influences subsequent recognition, that would indicate the presence of potentially confounding testing effects. Figure 5 shows that cue direction did affect order-recognition, confirmed by a two-way repeated-measures ANOVA on cued-recall probe directions between test 1 and test 2, (F, B)×(F, B); there was a significant main effect of test 1 direction, $F(1,35) = 31.05,\ MSE = 0.43,\ p < 0.0001,\ \eta_p^2 = 0.47$, and of test 2 direction, $F(1,35) = 7.40,\ MSE = 0.54,\ p < 0.05,\ \eta_p^2 = 0.17$, but the interaction was not significant, $F(1,35) = 0.43,\ MSE = 0.49,\ p > 0.5,\ \eta_p^2 = 0.012$, suggesting the effects of the two cued-recall tests simply summate.

For associative recognition, pairs were further broken down in terms of the cue directions of the pair derived from the left- versus right-sided item of rearranged pairs (see previous section). A $2 \times 2 \times 2$ within-subjects ANOVA on associative recognition d', with design Item Side (left- vs. right-item) × Cue Direction Test 1 (forward, backward) × Cue Direction Test 2 (forward, backward), produced no significant effects (all $p > 0.1$). The three-way interaction was the only effect with $F > 1$: $F(1,33) = 2.19,\ MSE = 0.11,\ p = 0.155,\ \eta_p^2 = 0.060$, but in a Bayesian ANOVA, $BF_{inclusion} < 0.0001$ for this term, suggesting it may be safely ignored. Thus, no indication of any direction-specific testing effects was found for associative recognition.

When both tests had been in the forward direction (FF), d' for order-recognition was high, but still not as high as for associative recognition. However, the difference was not significant, $t(68) = -1.27,\ p = 0.21,\ d = 0.30$, and the Bayesian version of this test suggests the outcome was inconclusive, $BF_{10} = 0.48$.

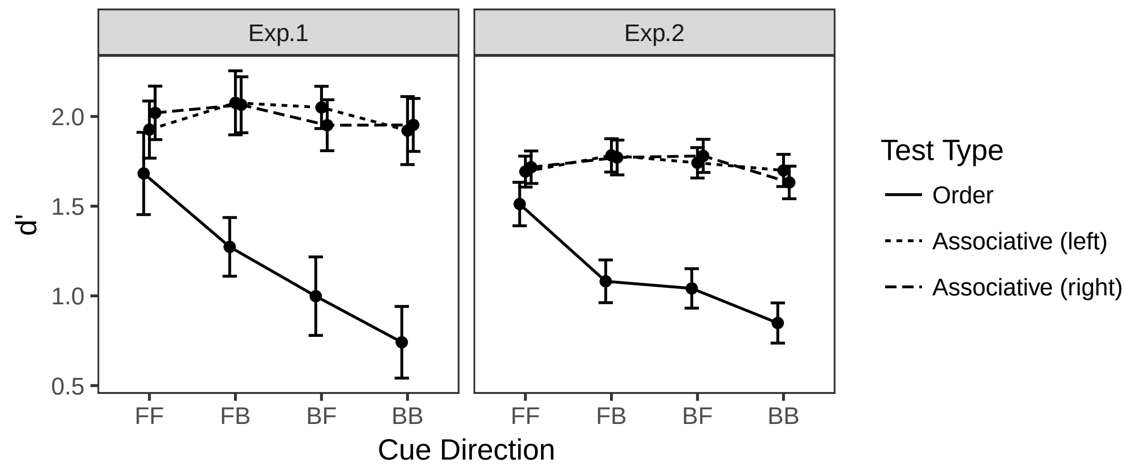


Figure 5. Testing effect: d' in initial recognition as a function of the cue directions in the prior two cued-recall probes. FB - Test 1 was forward, Test 2 backward, etc. For associative recognition, d' is broken down by the cue directions of associations from which the left, and right items were derived. Error bars plot 95% confidence intervals based on standard error of the mean. Also note that the x values have been staggered so that all means are visible.

Final recognition. The final recognition test was placed at the end of the experimental session to test whether, when studying for associations without regard to order, the Order-Ignore group nonetheless might have had some ability to retrieve order, and conversely, whether studying for order might have compromised the Order-Attend group's ability to perform associative recognition. Secondly, this test gave us an additional opportunity to interrogate the relationship between order- and association-memory.

Both groups performed better than chance on order-recognition; average d' was greater than zero (Order-Attend: $M = 0.75$, $t(16) = 4.12$, $p < 0.001$, $d = 1.00$; Order-Ignore: $M = 0.70$, $t(17) = 4.51$, $p < 0.001$, $d = 1.06$). However, d' did not differ significantly between groups (Table C2 and Figure 7): A 2 × 2 between-subjects ANOVA on Group (Order-Attend, Order-Ignore) × Test Type (order- vs. associative recognition), on d' found only a significant main effect of Test Type, $F(1, 66) = 78.08$, $MSE = 0.69$, $p < 0.0001$, $\eta_p^2 = 0.54$. The main effect of Group and the interaction were not significant, $F(1, 66) = 1.41$, $MSE = 0.69$, $p > 0.1$, $\eta_p^2 = 0.0069$ and $F(1, 66) = 0.19$, $MSE = 0.69$, $p > 0.5$, $\eta_p^2 = 0.0028$, respectively. To check the null main effect and interactions involving Group, we ran a Bayesian ANOVA with the same design. The main effect of Test Type had a $BF_{inclusion} > 1000$, providing clear support for this main effect being present. The null was favoured for the main effect of Group ($BF_{inclusion} = 0.28$). The interaction between Group and Test Type was inconclusive, $BF_{inclusion} = 0.32$. However, even if it had been reliable, the form of the interaction would have indicated Order-Ignore participants more impaired on associative recognition than on order-recognition, the opposite interaction than what we had predicted. The lack of main effect of Group suggests first, that participants have some substantial amount of order-memory available even when order is not relevant during study, and second, order-memory may not be able to be strengthened by intentional effort, at least with the current procedures.

Just like we did with initial recognition, we asked whether final order-recognition and cued recall covary across participants. First, comparing groups, Pearson correlations

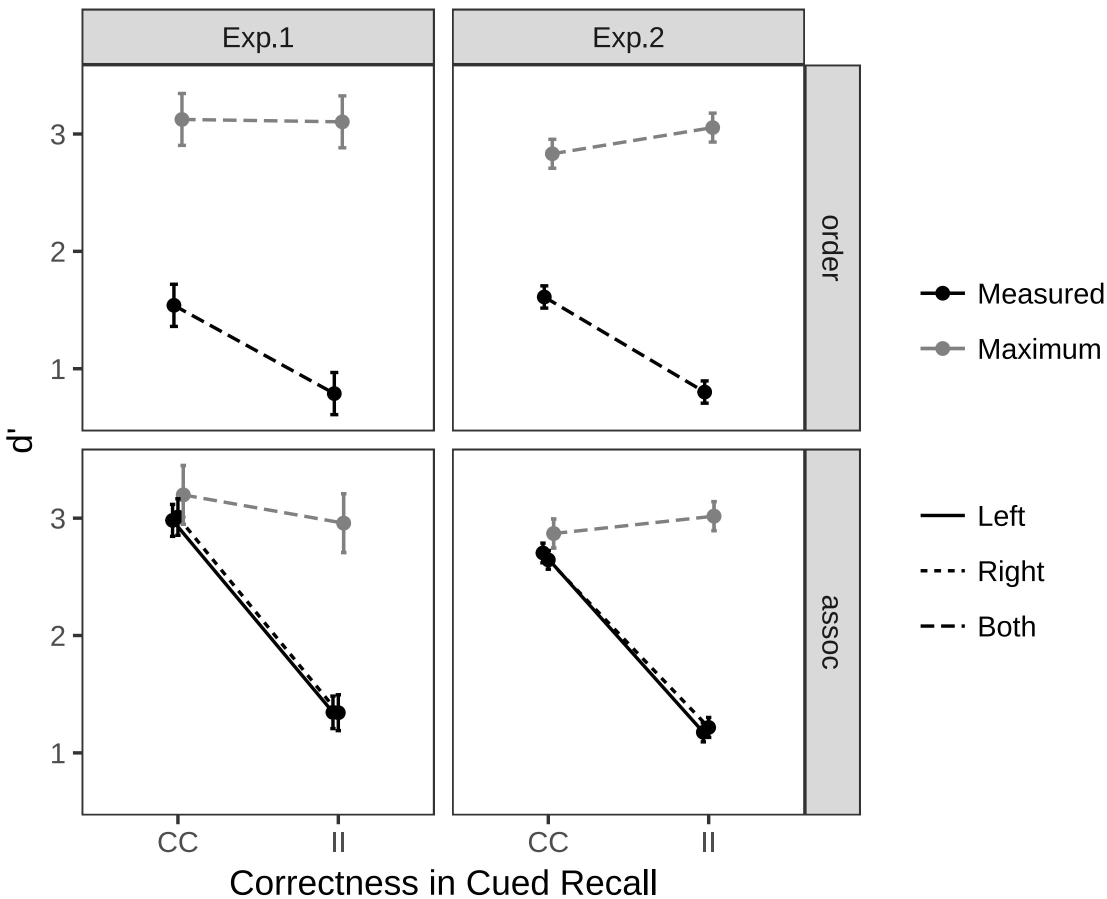


Figure 6. Performance on initial recognition, contingent on cued-recall outcome. d' in initial recognition for pairs for which both cued-recall tests had been correct (CC) or both incorrect (II). For associative recognition, the correctness (CC vs. II) was computed for either the left side item or the right side item, in separate calculations. "Left" denotes d' in terms of the correctness of the left side item, and "Right" denotes d' for the right side item. "Maximum" denotes an average of a maximum d' in terms of the corresponding trials for each condition. Error bars represent 95% confidence intervals based on standard error of the mean, corrected for between-subjects variability (Loftus & Masson, 1994). Note that the x values have been staggered so that all means are visible.

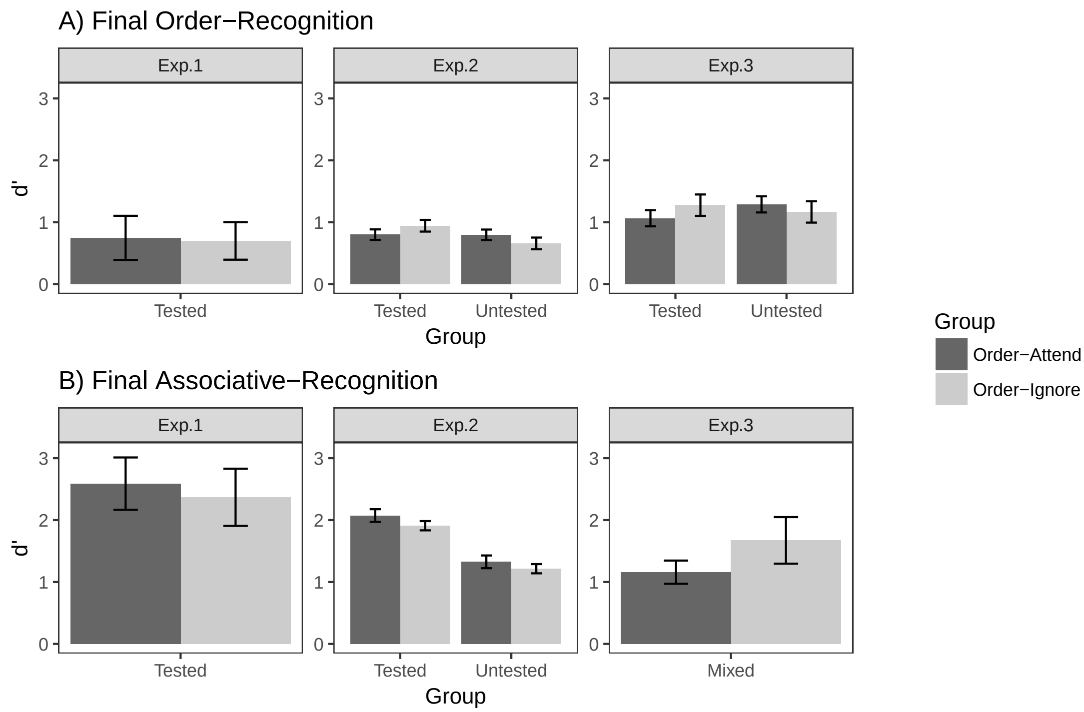


Figure 7. d' in final recognition, as a function of group, for all experiments. "Tested" denotes that the pairs were tested in both the cued-recall and the initial recognition test whereas "Untested" denotes that the pairs were not tested in those tests. "Mixed" denotes that the performance was calculated with both "Tested" and "Untested" pairs. Error bars represent 95% confidence intervals based on standard error of the mean.

between cued-recall accuracy and d' in final order-recognition (Figure C3) were nearly significant for the Order-Attend group, $r(16) = 0.48,\ p = 0.052$, similar in sign but still non-significant for the Order-Ignore group, $r(17) = 0.39,\ p > 0.1$, and the difference between these correlations (Williams' Test) was not significant, $z = 0.29,\ p > 0.5$. The corresponding analysis for final associative recognition produced large and significant correlations for both groups, Order-Attend: $r(18) = 0.86,\ p < 0.0001$, Order-Ignore: $r(15) = 0.80,\ p < 0.001$, but these correlations were still not significantly different between groups; Williams' Test, $z = 0.57,\ p > 0.5$. Thus, this provides no more support for the idea that the Order-Attend group might have studied fundamentally differently in a way that might have changed the relationship between association- and order-memory. Second, comparing final recognition tasks, cued-recall accuracy was significantly less correlated with d' in final order-recognition than in final associative recognition (Figure C3A vs. C3G; $z = 2.11,\ p < 0.05$) for Order-Attend participants, and was in the same direction, but the difference only approached significance, for Order-Ignore participants (Figure C3D vs. C3J, $z = 1.77,\ p = 0.077$), suggesting order-memory may be somewhat uncoupled from memory for the association.

Discussion

In sum, participants could judge the order of constituent items within associations, better than chance, both initially and at the end of the testing session. This ability was not as high as might be expected if associations were always stored with a definite order, even when participants apparently had very good memory for the association itself, when cued recall had succeeded twice for a particular pair (Figure 6). These findings raise the possibility that constituent-order is not as hard-coded within the association as might be suggested by current matrix models.

Also noteworthy, symmetry of cued recall (null effects of Direction) and associative symmetry (high Q_{DIFF}) are consistent with many prior results reviewed in the introduction (e.g., Murdock, 1962; Kahana, 2002), but we have extended the boundary conditions for these results to a situation in which participants expected to be tested on constituent order (the Order-Attend group). The null main effect of Group on cued-recall accuracy suggests that the need to remember constituent-order did not come at an overall cost to association-memory.

However, one finding complicated the interpretation of the results: backward cued-recall disrupted subsequent order-recognition, but did not disrupt subsequent associative recognition (Figure 5). Although we do not know the cause of this testing effect, consider that when the participant answered a cued-recall question, the cue item would have preceded the target item in time. The participant's own response appeared underneath the cue item. Thus, participants may have processed this visual information as though it were a new presentation of the pair— a form of output-encoding. In backward cued recall, seeing the right item as a cue and remembering the left item as a target would result in output-encoding of the pair in the opposite order than at initial study, and lead to an incorrect order judgement later on. By the same logic, forward cued-recall probes might facilitate order-recognition, because the effective temporal order of items during cued recall would now be congruent with study-order. This may explain why, in contrast, associative recognition was not differentially influenced by cued-recall direction: if the participants studied AB, an additional output-encoding of BA could be just as effective as an additional trace of AB in providing the participant evidence that AB had, indeed, been studied— or, for example, to use recall-to-reject to rule out a rearranged probe, AD. In any case, the presence of these cue-direction testing effects raises the possibility that the previous result, showing that for order-recognition, $d' < d'_{max}$ even for CC pairs, may be largely due to confusion introduced by backward cued recall, and leaves open the possibility that associations are, indeed, retrieved in order. Experiment 2 will address this testing effect by leaving half the pairs untested until final recognition.

Experiment 2

To address possible testing effects, half of the studied pairs were not tested in both cued recall and initial recognition in the second experiment, but all pairs were tested in final recognition. To partly offset the reduced sensitivity expected due to withholding half the pairs from initial tests, and the additional within-subjects factor (tested versus untested), we sought a greater sample size than in Experiment 1.

Methods. Participants were 233 undergraduate students enrolled in an introductory psychology course at the University of Alberta, who participated in exchanged for partial course credit. No participants from Experiment 1 participated in this experiment. Data from five participants were excluded from analyses due to low cued-recall accuracy (<10%).

All materials and procedures were as in Experiment 1, including the group manipulation (Table 1), except that: (1) only half of the studied pairs were tested with cued recall and initial recognition, leaving the other half untested until final recognition, and (2) the number of the blocks was increased from six in the first experiment to ten in the second, making use of freed-up session time due to fewer cued-recall and initial-recognition tests. The increase in number of blocks also compensated for some of the reduction in power for cued recall and initial recognition measures, and added power to the final-recognition measures.

Results

As in Experiment 1, we first report comprehensive analyses, and then revisit the central results in the Discussion.

Cued recall. A three-way, mixed ANOVA on cued-recall accuracy (Figure 3), with design Group (Order-Attend, Order-Ignore) × Test (1, 2) × Cue Direction (forward, backward) produced a non-significant main effect of Group, $F(1,226) = 0.003$, $MSE = 0.18$, $p{>}0.5$, $\eta_p^2 < 0.0001$, but significant main effects of Direction, $F(1,226) = 16.4$, $MSE = 0.01$, $p < 0.001$, $\eta_p^2 = 0.068$, and Test, $F(1,226) = 139$, $MSE = 0.003$, $p < 0.0001$, $\eta_p^2 = 0.38$. Test×Group was the only significant interaction, $F(1,226) = 4.26$, $MSE = 0.003$, $p < 0.05$, $\eta_p^2 = 0.019$. Some interactions, while non-significant, had $F > 1$, so we checked the null main effect and interactions involving Group with a Bayesian ANOVA, with the same design. The main effects of Test and Cue Direction both had $BF_{inclusion} > 1000$, The interaction Test × Cue Direction was not conclusive, $BF_{inclusion} = 0.944$. For the main effect and all interactions involving Group, the null was favoured ($BF_{inclusion} < 0.3$), but was close to threshold for Group × Direction, $BF_{inclusion} = 0.297$. The middle panels in Figure 3 show that forward recall was slightly superior to backward recall, although this asymmetry is quite small. The results leave room for the possibility that there is a small tendency for the forward-probe advantage to be smaller for the Order-Attend than the Order-Ignore participants.

As in the first experiment, the correlations between successive cued-recall tests were not different between the two groups (Figure 4). A 2 × 2 (Group × Cue Combination) mixed ANOVA showed that the main effect of Group and the interaction were not significant, $F(1,226) = 0.10$, $MSE = 1.47$ $p > 0.5$, $\eta_p^2 < 0.0001$ and $F(1,226) = 1.26$, $MSE = 0.9$, $p > 0.1$, $\eta_p^2 = 0.0055$, respectively. The main effect of Cue Combination (SAME vs. DIFF) was significant, $F(1,226) = 462.5$, $MSE = 0.9$, $p < 0.0001$, $\eta_p^2 = 0.67$. A Bayesian t-test, comparing Q_{DIFF} between groups favoured the null, $BF_{10} = 0.21$. Thus, as in Experiment 1, intentionally studying for order did not seem to disrupt cued-recall accuracy or associative symmetry.

Initial recognition. Performance in initial recognition was almost the same as in the first experiment (Table C1). Given the number of pairs, the maximum obtainable value of d' for both order- and associative recognition was d'_{max}=3.92. For order-recognition, $d' = 1.38$, was significantly greater than zero, $t(113) = 19.10$, $p < 0.0001$, $d = 1.79$, and significantly lower than d'_{max}, $t(113) = 35.07$, $p < 0.0001$, $d = 3.28$. Similarly, for associative recognition, $d' = 2.25$, significantly greater than zero, $t(113) = 25.36$, $p < 0.0001$, $d = 2.38$ and significantly lower than d'_{max}, $t(113) = 18.90$, $p < 0.0001$, $d = 1.77$. Most pertinent to models, in an ANOVA on order-recognition d', with design Cued-Recall-Accuracy[CC,II]×Task[Order,Associative recognition], the main effects of Cued-Recall-Accuracy and Task were both significant, $F(1,216) = 339.0$, $MSE = 0.44$, $p < 0.0001$, $\eta_p^2 = 0.61$ and $F(1,216) = 78.0$, $MSE = 0.75$, $p < 0.0001$, $\eta_p^2 = 0.27$, respectively, as was the interaction, $F(1,216) = 31.6$, $MSE = 0.44$, $p < 0.0001$, $\eta_p^2 = 0.13$ (Figure 6). As in Experiment 1, the interaction was explained by d' for CC minus d' for II pairs being significantly greater for associative recognition than for order-recognition, $t(216) = -5.61,\ p < 0.0001$. For initial order-recognition when both cued-recall tests were correct (CC, Figure 6) d' was significantly lower than the maximum d', $t(111) = 14.2,\ p < 0.0001,\ d = 1.46$.

As in Experiment 1, testing effects were observed (Figure 5). When both tests had been in the forward direction (FF), d' for order-recognition was nearly as high as for associative recognition; the difference fell just short of significance, $t(226) = -1.90,\ p = 0.059,\ d = 0.25$. However, as in Experiment 1, the Bayesian analysis was inconclusive, $BF_{10} = 0.787$.

The Pearson correlation between cued-recall accuracy (log-odds ratio) and initial order-recognition (d'), $r(113) = 0.48,\ p < 0.0001$, was lower than in Experiment 1 (0.71), whereas the correlation between the cued-recall and the initial associative-recognition test, $r(113) = 0.81,\ p < 0.0001$, was similar to Experiment 1 (0.84) (Figure C2), and this difference was now significant, Williams' test, $z = 4.54,\ p < 0.0001$.

Final recognition. Recall that the major modification in Experiment 2 was to leave half the pairs initially untested, to check for possible confounding testing effects. Thus, the analyses of final recognition, which tested all studied pairs, will be broken down by whether pairs were previously tested or untested. The values of d' in final order-recognition (Figure 7), for both groups, both final recognition tasks (order- and associative recognition), and for both tested and untested pairs, were all significantly greater than zero and significantly less than their maximum possible values (d'_{max}), all $p < 0.0001$.

A $2 \times 2 \times 2$ mixed ANOVA, with design Group (Order-Attend, Order-Ignore) $\times$ Test Type (order-recognition, associative recognition) $\times$ Testedness (tested, untested probes) on d', found the main effects of Test Type and Testedness, and the interaction between the two was significant, $F(1, 224) = 69.46,\ MSE = 1.13, p < 0.0001,\ \eta_p^2 = 0.24,\ F(1, 224) = 86.06,\ MSE = 0.24,\ p < 0.0001,\ \eta_p^2 = 0.28$, and $F(1, 224) = 39.71,\ MSE = 0.24,\ p < 0.0001,\ \eta_p^2 = 0.15$, respectively. The main effect of Group, $F(1, 224) = 0.35,\ MSE = 1.13,\ p > 0.5,\ \eta_p^2 = 0.0019$, and other interactions were not significant. However, for some interactions, $F > 1$, so to check the null main effect and interactions involving Group, we ran a Bayesian ANOVA with the same design. The main effects of Test Type and Testedness received clear support, both $BF_{inclusion} > 1000$, as did their interaction, $BF_{inclusion} > 1000$. For all other effects, the null was favoured ($BF_{inclusion} < 0.3$) except the interaction Test Type $\times$ Testedness $\times$ Group, which was inconclusive ($BF_{inclusion} = 0.32$); if it were to be better supported, visual inspection of Figure 7 confirms that the form of this interaction is inconsistent with the idea that the Order-Attend group had a specific order-recognition advantage. Thus, consistent with Experiment 1, intentional effort to study for order (Order-Attend group) did not substantially improve performance on order judgments, even for pairs that were not tested in any way until final recognition.

Analyzing the relationship between cued recall and final recognition across participants (Figure C3), for tested pairs, Pearson correlations between log-odds ratio of the cued-recall accuracy and d' in final order-recognition were $r(58) = 0.42,\ p < 0.005$ for the Order-Attend group and $r(57) = 0.35,\ p < 0.01$ for the Order-Ignore group. These were not significantly different (Williams' Test, $z = 0.37,\ p > 0.5$). For final associative recognition, the correlations were higher: $r(54) = 0.76,\ p < 0.0001$ for the Order-Attend group and $r(55) = 0.78,\ p < 0.0001$ for the Order-Ignore group. These did not differ significantly between groups (Williams' Test, $z = 0.28,\ p > 0.5$). These correlations did significantly differ between final test types (order- vs. associative recognition) for both groups, $z = 2.84,\ p < 0.01$ (Order-Attend) and $z = 3.49,\ p < 0.001$ (Order-Ignore), respectively.

However, our main interest in Experiment 2 is in final recognition of the untested pairs. Correlating cued recall with final order-recognition, $r(58) = 0.50,\ p < 0.0001$ (Order-Attend) and $r(57) = 0.64,\ p < 0.0001$ (Order-Ignore), with a non-significant difference ($z = 1.15,\ p > 0.1$), although because $z > 1$, this result should be viewed with caution. Correlating cued recall with final associative recognition, $r(54) = 0.77,\ p < 0.0001$ (Order-Attend) and $r(55) = 0.70,\ p < 0.0001$ (Order-Ignore), with a non-significant difference ($z = 0.84,\ p > 0.1$). The difference in correlations between order- and associative recognition was significant for the Order-Attend group, z=2.53, p<0.05, but not significant for the Order-Ignore group, z=0.55, p>0.5. Thus, the testing effect did seem to reduce the apparent coupling between cued recall and order-recognition. However, correlations still did not rise to the level of those between cued recall and associative recognition. For the Order-Attend group, this was a statistically reliable difference, which hints at the possibility that Order-Attend participants were doing *something* differently during study, that slightly decoupled association- from order-memory.

Discussion

Experiment 2 replicated the main results of Experiment 1, even for pairs that are not subject to testing effects (Figure 7).

The results clarify the testing effects found in Experiment 1. For tested pairs, in both experiments, order-recognition was worst for pairs that had been tested twice backward in cued recall, and best for pairs that had been tested twice forward in cued recall (Figure 5). This could be due to either backward cued recall disrupting order-memory, or forward cued-recall facilitating order-memory, or a combination of both. Although there was a main effect of testedness on final recognition, the difference in d' between tested and untested pairs was far smaller (Figure 7A, middle panel) than the difference in d' caused by testing effects (compare with the difference between FF and BB conditions in Figure 5, order-recognition). This suggests that both forward cued recall facilitated, and backward cued recall disrupted, order-memory. Taking this into consideration, this suggests that the effects of prior cued-recall on order-recognition roughly cancel, and reinforces the idea that order-recognition took on a moderate value within the plausible range; hence, order-memory is present, but imperfect.

A model that assumes associations are stored along with order may predict better performance on order-recognition and more coupling between order-recognition and cued recall than we observed. In the General Discussion, we consider what kind of modification such a model might need to accommodate these results.

Experiment 3

In both experiments, the Order-Attend group did not show superior performance on final order-recognition than the Order-Ignore group, leading us to tentatively conclude that order memory cannot be easily, voluntarily strengthened. However, one factor complicates this interpretation. After studying a list of pairs, participants were first tested with cued recall— not only cued recall, but two sets of cued-recall tests (successive-testing was included to assess associative symmetry). It is possible that participants in the Order-Attend group did not make any effort to modify their study strategy to optimize for order-memory, not because they could not, but because they were far more concerned about the impending test of association-memory (cued recall). We reasoned that, if we removed cued recall from the initial tests, which we did in Experiment 3, we might reveal the effects of a

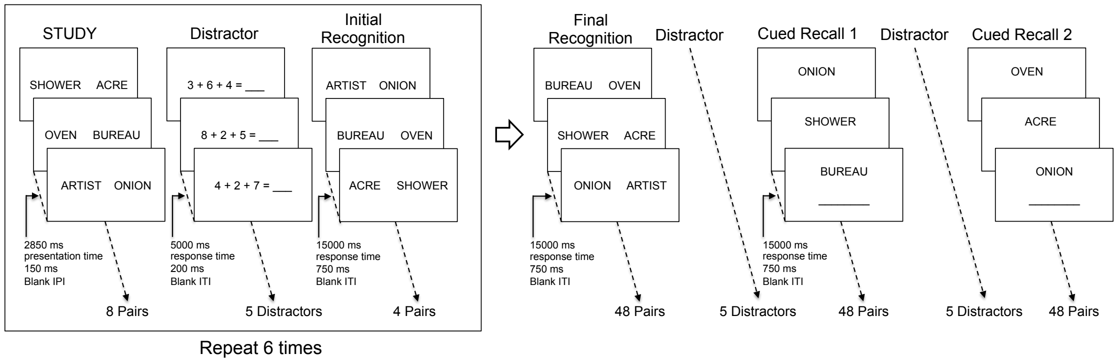


Figure 8. Schematic depiction of the experimental procedure in Experiment 3. For each list the participant studied eight word pairs and completed a distractor task followed by an initial recognition test (order- or associative recognition, for Order-Attend and Order-Ignore, respectively). After six cycles of this procedure, the participant had a final recognition test of all studied pairs ($48 = 8$ pairs $\times$ 6 lists), either order- or associative recognition, depending on group, and two full sets of cued recall.

strategy-difference between groups: Order-Attend participants might then favour order-memory at the expense of association-memory, and Order-Ignore participants might favour association-memory at the expense of order-memory. We still tested all pairs with cued recall, with successive testing, to retain the ability to measure symmetry of mean accuracy and associative symmetry, but this was placed at the end of the testing session.

Methods

Participants were 138 undergraduate students enrolled in an introductory psychology course at the University of Alberta, who participated in exchange for partial course credit. Data from thirteen participants were excluded from analyses due to low cued-recall accuracy (<5%). No participants from Experiments 1 or 2 participated in this experiment.

All materials and procedures were the same as those in the first two experiments as well as group manipulation (Table 1) except the relocation of the cued recall test. Half the studied pairs were tested in with initial recognition, leaving the other half untested until final recognition. Finally, six cycles (lists) were included, as in Experiment 1, prior to the final cued-recall tests.

Results

Initial recognition. Interestingly, without the intervening cued recall, d' was nearly equivalent for initial order-recognition and initial associative recognition, $t(122) = 0.36$, $p > 0.5$, $d = 0.065$ (compare with Experiments 1 and 2, Table C1 and Figure C1). That is, d' was greater in Experiment 3 than the previous experiments for order-recognition, but lower for associative recognition. This supports our suspicion that participants in the earlier experiments may have applied a study strategy that was optimized for cued recall, regardless of group. The Order-Attend group may thus have found a strategy that was more effective for order-memory. In addition, cued recall in the previous experiments may have benefited associative recognition, by providing additional study opportunities, which might explain the reduced d' for associative recognition here compared previous experiments, despite the study–test interval being shorter.

Final recognition. A 2×2 ANOVA with design Group (Order-Attend, Order-Ignore) × Testedness (tested, untested pairs) on final order-recognition d' produced no significant effects (all $p > 0.1$). Not all effects had $F > 1$, but a Bayesian ANOVA produced $BF_{inclusion} < 0.3$ for all effects, favouring null effects. Despite potentially having optimized their study strategy for order-memory, Figure 7 shows that the Order-Attend group did not perform better in final order-recognition than the Order-Ignore group, for both tested, $t(61) = 0.93,\ p > 0.1,\ d = 0.23$, although this was inconclusive according to a Bayesian t test, $BF_{10} = 0.37$, and untested pairs, $t(61) = 0.56,\ p > 0.5,\ d = 0.14$ (null favoured by Bayesian t test, $BF_{10} = 0.29$). In contrast, for final associative recognition, the Order-Ignore group performed better than the Order-Attend group, $t(59) = 2.37,\ p < 0.05,\ d = 0.61$.

Final cued recall. A three-way, mixed, repeated-measures ANOVA on cued-recall accuracy (Figure 3), with design Group (Order-Attend, Order-Ignore) × Test (1, 2) × Cue Direction (forward, backward), found a significant main effect of Group, $F(1, 122) = 5.53,\ p < 0.05,\ \eta_p^2 = 0.043$; thus, following on the heels of the final associative-recognition advantage for the Order-Ignore group, the Order-Ignore group was also superior at final cued recall. The main effect of Test approached significance, $F(1, 122) = 3.84,\ p = 0.052,\ \eta_p^2 = 0.030$. All other effects were not significant ($p > 0.1$), but because not all those $F > 1$, we conducted a Bayesian version of this ANOVA. All $BF_{inclusion} < 0.3$ except for the main effect of Group, which was inconclusive, $BF_{inclusion} = 1.05$; thus, the main effect of Group was evidently too small in magnitude to be supported in the Bayesian ANOVA.

Turning to Q_{DIFF}, our measure of associative symmetry (Figure 4), a two-way, mixed, repeated-measures ANOVA on the log-odds-transformed Q values, with design Group (Order-Attend, Order-Ignore) × Cue Combination (SAME, DIFF), yielded a significant main effect of Cue Combination, $F(1, 122) = 101,\ MSE = 1.33,\ p < 0.0001,\ \eta_p^2 = 0.45$, but neither the main effect of Group, $F(1, 122) = 0.158,\ MSE = 1.71,\ \eta_p^2 = 0.0013$, nor the interaction, $F(1, 122) = 0.322,\ MSE = 1.33,\ \eta_p^2 = 0.0026$, were significant (p>0.5). A Bayesian t-test of Q_{DIFF} between Groups also favoured the null hypothesis, $BF = 0.23$. Thus, whatever strategy Order-Attend participants applied, despite reducing mean associative-recognition and cued-recall accuracy, did not alter associative symmetry.

Discussion

Experiment 3 showed that there is a way in which participants may tune their study processes in anticipation of an order-memory test, which may be overridden when, as in Experiments 1 and 2, association-memory also needs to be maintained. However, sadly, this order-oriented study process does not seem enhance order-recognition, but rather, decrease performance on subsequent tests of association-memory (associative recognition and cued recall).

General Discussion

Memory for the order of constituent items of associations has been largely overlooked in verbal memory research. When it has been measured, it has tended to be either a side-point of the study, or measured in the absence of memory for associations themselves (Greene & Tussing, 2001; Kounios et al., 2001; Rehani & Caplan, 2011). However, as we showed in the Introduction, current models of association-memory include assumptions, often extreme, about the level of order-memory given that associations are learned (Rehani & Caplan, 2011). Thus, data on order-memory could provide powerful new evidence to test and select current models of association-memory, and to provide clues as to how those models should be developed in the future. Here we presented three experiments with the goal of characterizing human order-memory ability, and testing how this might relate to association-memory, itself.

Participants possess some ability to judge constituent-order. In all experiments, d' for order-recognition was significantly greater than zero, ruling out models that rely only exclusively on convolution to store associations. This does not rule out the use of convolution altogether, but suggests that convolution-based models require something additional. One intriguing possibility is to use a non-commutative form of convolution, like that used in BEAGLE (Jones & Mewhort, 2007; Kelly, Blostein, & Mewhort, 2013). However, the risk with this approach is that it may be susceptible to the same challenges faced by matrix models (see below). Alternatively, symmetric convolution models could be supplemented with an additional storage term from which relative order can be inferred. With this approach, some care must be taken to ensure that association- and order-memory are not completely independent, but this could presumably be easily achieved with a free parameter that controls the correlation between order- and association-encoding strengths across pairs.

Order-recognition performance is high but not as high as possible. For order-recognition, both initial and final, d' was significantly less than its maximal possible value. In both Experiments 1 and 2, testing effects modified order-recognition. For every backward cued-recall test, d' was reduced, but the similarity of d' on average for tested and untested pairs (Experiment 2) suggests that for every forward cued-recall test, d' was also facilitated. Apparently, participants re-encoded the association in reverse-order following backward cued recall and in presentation-order following forward cued recall. Thus, testing effects roughly cancelled out, leaving the previous conclusion unchanged: d' took on a moderate value. Most pertinent to models, like current matrix models, in which order is an intrinsic property of associations, cued-recall correctness influenced order-recognition, but far less than it influenced associative recognition, again, suggesting that order is moderately decoupled from association-memory (Figure 6).

Oddly, if order-memory were not near-maximal (given memory for the association itself), one would expect that it could be improved when participants have order-memory as their explicit goal. However, in Experiments 1 and 2, in which cued recall was arguably the dominant task, participants tested initially on order (Order-Attend group) performed equivalently to those tested initially without order (Order-Ignore group), both on final order-recognition, and on cued recall. In Experiment 3, cued recall was dropped from the initial tests, so that for Order-Attend participants, order-recognition was their primary task. However, even under those conditions, the Order-Attend group did not improve their order-memory performance. This suggests that participants were unable to come up with a strategy that selectively enhanced order-memory, beyond the level offered by spontaneously adopted study strategies for association-memory. The implication is that, despite not being perfectly coupled, constituent-order is not something that participants can necessarily improve without also improving association-memory. This is reminiscent of the hierarchical relationship between item- and association-memory, wherein item-memory may be retained at the expense of association-memory but not vice versa (Hockley & Cristi, 1996a, 1996b).

If order were embedded within the association itself— and consequently, driven by the same variability in memory as association-memory— we would predict that order-memory performance would covary positively with cued-recall performance across participants. In contrast, if order were inferred from other sources of evidence in memory (such as item-context associations), we would expect that order-recognition should be relatively decoupled from cued recall, and the correlation would be weaker. The correlation between order-recognition and cued recall was, in many comparisons, smaller than the correlation between associative recognition and cued recall (Figure C3).

Finally, one might suspect that all participants ignored order during study in Experiments 1 and 2, given that cued recall dominated the initial test phase. However, in Experiment 3, the group manipulation did influence performance, notably, reducing cued recall and final associative recognition— but still did not result in superior final order-recognition, again, suggesting that participants cannot easily increase their order-memory without similarly increasing their association-memory.

Our order-recognition d' values were within the range that has been previously reported with procedures that have tested order-recognition in the absence of tests of association-memory (Kounios et al., 2001, 2003; Greene & Tussing, 2001; Yang et al., 2013). It remains to be tested whether any of the manipulations investigated in those studies, such as pre-existing semantic similarity and conceptual fusion, either increase or decrease the coupling of order-memory and association-memory.

Finally, our results are consistent with Rehani and Caplan (2011), who inferred that their participants learned associations with some moderate, but not maximal, level of constituent-order, because their participants could often disambiguate the forward from backward associate (having studied AB, BC, ...: given B, recall A rather than C or given B, recall C rather than A).

New constraints on models. Given our findings, a model of association-memory (even for tasks in which order is not relevant during study) needs to simultaneously produce 1) order-recognition greater than chance but less than optimal; 2) only moderate coupling of order-recognition to association-memory. Moreover, it must do this while not compromising features of association-memory that we replicated within the same data sets here: 3) near-symmetric mean accuracy in cued recall; and 4) a high correlation between forward and backward cued recall of word pairs, Q_{DIFF} (Caplan et al., 2006; Caplan, Rehani, & Andrews, 2014; Caplan, Boulton, & Gagné, 2014; Kahana, 2002; Madan et al., 2010; Rizzuto & Kahana, 2000, 2001; Rehani & Caplan, 2011; Sommer et al., 2008).

As already mentioned, models that encode associations using convolution are already symmetric (Metcalfe Eich, 1982; Murdock, 1982), such models provide no ability to judge constituent-order, which is inconsistent with the current results, but could be supplemented with separate terms that could provide order.

It may be just as profitable to start with models that are intrinsically directional, and modify them to reduce the coupling between order- and association-memory. Here we consider matrix models, although many of the conclusions may apply equally well to concatenation-based association-memory models, for which order is also an intrinsic property of the association (e.g., Hintzman, 1984, 1986; Rizzuto & Kahana, 2000, 2001; Shiffrin & Steyvers, 1997). Matrix models (Anderson, 1970; Humphreys et al., 1989; Pike, 1984; Willshaw et al., 1969), to our knowledge, have never been developed to implement order-recognition, but could very easily be adapted to do this task. Consider a memory that contains only a single association in one direction, $M = \mathbf{b}\mathbf{a}^T$. Given an order-recognition probe, probing with one item, $M\mathbf{a} \simeq \mathbf{b}$. A simple dot-product can convert this into a scalar strength that could be used to make a response decision; $(M\mathbf{a}) \cdot \mathbf{b}$, will yield a high value, given that the encoding strength was high. Probing with the other probe item, $M\mathbf{b} \simeq 0$; the dot product, $(M\mathbf{b}) \cdot \mathbf{a} \simeq 0$ as well. Thus, even probing by multiplying only from the right, with one probe, or both, sequentially or in parallel, could potentially produce very high accuracy and d' levels for order-recognition.

The next question is how to implement both forward and backward cued recall in a matrix model. The basic matrix model is probed by multiplying from the right. This supports cued recall only in one direction. Pike (1984) pointed out it would be straightforward to model backward cued recall by probing by multiplying with the transpose from the left, which is equivalent to transposing the matrix, M, and multiplying as usual, from the right. Pike (1984) noted that, if the probe direction were known, the model could be tested in the forward (multiplying from the right) or backward (multiplying with the transpose from the left), depending on probe direction. Indeed, this is the essence of the approach taken by Rizzuto and Kahana (2000, 2001), although adapted to a concatenation-based representation, which we discuss further below. This was appropriate, because in the experiment Rizzuto and Kahana were fitting, probe direction was explicit in the cue. In the three experiments reported here, cued-recall probes were always presented as a lone, centrally presented word, with the response line directly below it, giving no hint as to whether the probe was forward or backward.

Given the high value of Q_{DIFF}, one thought might be that the model stores both forward and backward associations together, $M = \gamma_F \mathbf{b}\mathbf{a}^T + \gamma_B \mathbf{a}\mathbf{b}^T$, where γ_F and γ_B are encoding strengths. To produce a high Q_{DIFF}, one would need to assume that γ_F andγ_B are highly correlated (Rizzuto & Kahana, 2000, 2001). However, if $E[\gamma_F] = E[\gamma_B]$ (where $E[]$ denotes the expectation; here, the mean), the model could not distinguish direction better than chance, which means the model would produce $d' = 0$ when tested with order-recognition. To be able to perform order-recognition better than chance, the model could store the original order ("forward" term) with a greater mean encoding strength, $E[\gamma_F] > E[\gamma_B]$. This would preserve the high Q_{DIFF}, but would then lead one to predict higher performance on forward than backward cued recall; asymmetries, when we found them, were small, and moreover, the forward–backward difference did not correlate (positively) significantly with d' in order-recognition (Experiment 1: $r(34) = -0.18$; Experiment 2: $r(112) = 0.06$; Experiment 3: $r(58) = -0.05$; all $p > 0.2$; Figure C4).

As an alternative approach that might be useful when cued-recall probes do not reveal probe direction, Pike (1984) proposed that the model could be probed in both directions. Having stored $M = \gamma \mathbf{b}\mathbf{a}^T$, where γ is an encoding strength, to perform cued recall, one could simultaneously probe M and M^T, where ^{T} denotes the transpose: $(M + M^T)\mathbf{a} = \gamma\mathbf{b} + \gamma(0)$ (assuming item vectors are orthonormal). In this way, cued recall can proceed blind to order. Probing in the backward direction (and assuming $\mathbf{a}$ and $\mathbf{b}$ are orthogonal), $(M + M^T)\mathbf{b} = \gamma(0) + \gamma\mathbf{a}$, thus retrieving the other associate with the same encoding strength, γ. If γ varied across pairs, as is normally assumed, this would produce a high value of Q_{DIFF}. As already shown, to perform order-recognition, the model could simply probe M with one or both of the probe items, or also probe M^T, and compare the two outcomes.

Thus far, the algebra suggests that such an augmented matrix model should over-predict order-recognition success, and over-predict the coupling between order-recognition and cued recall. To check this, we simulated a very simple version of the model (see Appendix for a full description of the model and simulations details). We added one feature to the model. We reasoned that even though the model must encode each association unambiguously in one order, it is plausible that the wrong order is initially encoded. This would lead to errors, and might also explain some of the decoupling we saw, for example, wherein order-recognition was less coupled to prior cued-recall success than associative recognition (Figure 6). We added a parameter, p_{rev}, which is the probability that the incorrect order is stored.

Figure A1 plots d' from the simulated data, as a function of p_{rev}, separately for pairs that were previously correct or incorrect in cued recall. Panel A shows that order-recognition is highly dependent on cued-recall success when order is accurately encoded ($p_{rev} = 0$, at the left edge of the plot). As p_{rev} increases, cued-recall success is less related to success in order-recognition. In contrast, p_{rev} has no effect on the dependence of associative recognition on cued-recall outcome (panel B). With moderate p_{rev} (∼0.2), this model can come quite close to the effects we observed in the data (cf. Figure 6). In this implementation of associative recognition, we probed with the left-hand probe item from both sides at once (multiplying $M + M^T$), similar to our implementation of cued recall. When we modified the model to probe only from one direction (panel C), associative recognition became less coupled to cued recall. This is because this implementation of associative recognition is directionally sensitive; if BA were stored instead of AB, then testing a probe by multiplying from the right with $\mathbf{a}$ would fail (retrieve a vector resembling noise) whether part of an intact or rearranged probe. Thus, intact probes would be more likely to be judged as rearranged. If participants behave like a mixture of these two strategies, the coupling between associative recognition and cued recall would presumably be midway between panels B and C.

Thus, we have proposed a straight-forward and plausible way in which a model that unambiguously stores associations in a particular order can nonetheless produce an apparently decoupling of order- from association-memory. But, this leads to the following prediction: If order were intrinsic to the association, but simply encoded incorrectly compared to the presentation order, one would expect participants to stick with their incorrect order-recognition response to a given pair from initial to final recognition. If we examine the relationship between initial and final order-recognition, initial-correct&final-correct should be most common, but if order errors are due to the wrong order being stored, then initial-incorrect&final-incorrect should also be common. The remaining cases, initial-correct&final-incorrect and initial-incorrect&final-correct, should be the most rare. Moreover, this pattern should be more prominent when the association was well learned, namely, cued-recall of a pair had been correct twice (CC pairs) than when cued-recall had been incorrect twice (II pairs). Alternatively, if errors in order-recognition are guesses, one would not expect initial-incorrect&final-incorrect cases to be particularly frequent. For II pairs, one could assume there is more guessing (i.e., no order is known) than for CC pairs. Figure C5 shows that the rate of the critical condition, initial-incorrect&final-incorrect, is not substantially greater than initial-incorrect&final-correct pairs, and is in fact lower than the frequency of initial-correct&final-incorrect pairs. Most pertinently, the pattern is quite similar when the association was likely known (CC condition) as when the association was likely not known (II condition).

To check this, for Experiment 1, a 2×3 ANOVA with cued-recall accuracy[CC,II] × initial/final outcome[ci,ic,ii] (the cc condition was left out because it is directly determined by the remaining three values), found a non-significant interaction both for Experiment 1, $F(2, 32) = 0.66$, $MSE = 0.019$, $p = 0.53$, $\eta_p^2 = 0.039$, and Experiment 2, $F(2, 116) = 0.67$, $MSE = 0.015$, $p = 0.51$, $\eta_p^2 = 0.015$. This pattern is inconsistent with the idea that, when they remember the association, participants judge order wrong because the encoded the wrong order. Rather, it is more consistent with the idea that even when the association can be retrieved, order can be uncertain, leading participants to guess. Also, if the decoupling of cued recall and order-recognition were due to storing the wrong order, one would also expect that participants in the Order-Attend group could have improved their order-memory, by taking extra care to ensure the presentation-order was correctly encoded, whereas the Order-Ignore group would have had no such incentive. The absence of any advantage of order-recognition for the Order-Attend groups may also speak against the "wrong-order" hypothesis.

In sum, it is possible to augment current models that provide order to explain moderate-level order-recognition performance, and to reduce the initially high degree of coupling between order-recognition and cued-recall performance. However, the finer structure of the data suggest this approach may be insufficient. Rather, a model may need to provide a way in which associations may be remembered, but constituent-order can be unknown or ambiguous.

Rizzuto and Kahana's auto-associative concatenation-based model. The model designed by Rizzuto and Kahana (2000, 2001) deserves special attention. Their Hopfield network encoded pairs as a matrix-autoassociation of the concatenation of the pair of item-vectors. Thus, if $\oplus$ denotes concatenation, the model stored $(\mathbf{a} \oplus \mathbf{b})(\mathbf{a} \oplus \mathbf{b})^T$. They further assumed that weights (matrix elements) were stored probabilistically, and included a parameter, ρ, that controlled the correlation between each weight in the "forward" quadrant of the matrix (corresponding to $\mathbf{ba}^T$) and its counterpart in the "backward" quadrant ($\mathbf{ab}^T$). With a value of ρ close to 1, the model could fit the high Q_{DIFF} as well as symmetry in mean accuracy of cued recall. To model forward cued recall, the model was probed with $\mathbf{a} \oplus \mathbf{k}$, where $\mathbf{k}$ is a vector containing noise. To probe in the backward direction, the model was probed with $\mathbf{k} \oplus \mathbf{b}$. Just as we had to do for the simple matrix model, to adapt the model to our task, where cued-recall probes do not reveal their direction, the model needs a small modification. Both forward and backward cued-recall could be attempted in sequence. Alternatively, this model could be probed simultaneously in both directions, probing with, for example, $\mathbf{a} \oplus \mathbf{a}$. Because of the concatenation-based representation of pairs, $\mathbf{a}$ in the non-stored position, without further assumptions, should behave very similarly to noise, $\mathbf{k}$. The same logic may be applied to other models of association-memory that rely on concatenation (e.g., Hintzman, 1984, 1986; Shiffrin & Steyvers, 1997). Thus, with the simple extension to be able to perform cued-recall simultaneously in both directions and to judge the order of a two-item probe, Rizzuto and Kahana's model may already be adequate to capture many of our results. However, such a model would presumably suffer from the same limitations as the simple matrix model; namely, order appears to sometimes be ambiguously encoded, even when the association (pairing) is effectively retrieved. That said, the Rizzuto and Kahana model, because it includes probabilistic storage of individual weights and an additional degree of freedom (ρ), can accommodate a slight decoupling of forward and backward association strengths, which gives it an easy way of explaining the substantial reduction of Q_{DIFF} relative to Q_{SAME} (Rizzuto & Kahana, 2000, 2001).

In sum, models of association-memory that are inherently symmetric under-predict order-recognition but with order information stored separately, and allowing the symmetric association term and order term to have moderately correlated encoding strengths, such an enriched model might provide an adequate account of our data. Alternatively, models that assume order is unambiguously encoded along with the association may over-predict order-recognition accuracy, particularly when the association is, itself, well remembered, and over-predict the degree of coupling between order- and association-memory. Such models may be enriched by devising a way in which order can be made to be ambiguous when the association is retrieved.

Relevance for Recursive Reminding

Our findings inform Hintzman's Recursive Reminding theory, which he proposed in an attempt to explain interesting findings suggesting that participants can very accurately judge the relative order of occurrence of repeated, or nearly repeated, items (Jacoby & Wahlheim, 2013; Wahlheim, Maddox, & Jacoby, 2014). Hintzman (2011) proposed that, when participants experience an event (e.g., an item or an association) that reminds them (by virtue of being repeated, or similar) of a previous event, they retrieve the prior event, and then encode an association between both events, in a manner that accurately preserves their relative order. Prior to the current results, associative symmetry, and current formulations of convolution-based memory models, would have seemed to challenge recursive reminding. That is, if associations are stored without order, recursive reminding would require a new mechanism of storing associations with accurate order. However, our findings suggest that no special mechanism is required; participants' typical association-learning strategy already preserves relative-order moderately well.

Relevance of models of serial-order memory

There may be insights to gain from research that has specifically targeted serial-order memory. Researchers interested in memory for serial-order have mostly focused on serial-recall and serial-anticipation procedures. So-called associative chaining models explicitly assume that ordered, serial lists are composed of associations between pairs of items (Ebbinghaus, 1885/1913), suggesting a continuity between memory for long, ordered lists and memory for sets of pairs (Caplan, 2015). However, positional-coding models were developed in an explicit effort to avoid modelling serial-order and association-memory together, working on the assumption that serial-recall and cued recall of pairs have nothing in common (Caplan, 2015). Even within an associative-chaining framework, Murdock and Franklin (1984) assumed distinct modes of operation of the model; in other words, they assumed that participants either only learned associations between items within pairs (A–B, C–D), or also learned "between-pair" associations (B–C). Because their associations were based on convolution (see also Lewandowsky & Murdock, 1989), they contained no order information. Thus, their model of serial-recall could reproduce a list in order by starting at the beginning of the list (A–B is unambiguous), but the association-memory model would presumably not be able to perform above chance on our order-recognition task. Caplan (2004, 2005) revived the idea that serial lists and lists of pairs might be based on a common inter-item association mechanism. He showed that apparent dissociations between memory for associations and memory for serial lists could be explained in models that treated associations and serial lists identically. The insight was that associations are relatively isolated from interference from other studied items, whereas items within a serial list are susceptible to more within-list competition. Consequently, Q_{DIFF} was modestly reduced for cued recall of subsets of serial lists, compared to cued recall of pairs. This argument was supported even in comparing pairs to the smallest possible serial list, triples (Caplan et al., 2006). Our findings further suggest that an association in memory that encodes both pairs and transitions within serial lists may embody associative symmetry, in the sense that there is a single term that can be used to retrieve the "forward" or "backward" association, but it is not entirely ambiguous with respect to order, as in convolution models. Rather, it includes moderately high-quality order information that can be used to disambiguate recall direction. This may also explain why Kahana and Caplan (2002) found that participants produced very accurate responses to cued recall of 19-word lists, in both the forward and backward direction, in exchange for only a small reduction in Q_{DIFF} (Caplan, 2005).

Finally, if, as suggested by Murdock and Franklin (1984) and designers of positional-coding models (e.g., Brown, Preece, & Hulme, 2000; Brown, Neath, & Chater, 2007; Burgess & Hitch, 1999; Henson, Norris, Page, & Baddeley, 1996; Lewandowsky & Farrell, 2000), lists of associations are learned entirely differently than serial lists (containing the same number of items), one might expect order-memory to be superior when derived from a serial list than from a list of pairs. In the judgements of relative order (commonly known as "judgements of relative recency," JOR) procedure, participants study a list of items (e.g., Hacker, 1980; Hockley, 1984; Klein, Shiffrin, & Criss, 2007; Muter, 1979; Yntema & Trask, 1963). Following a retention interval, participants are given a two-item probe and asked to judge which item came later, or which item came earlier (Chan, Ross, Earle, & Caplan, 2009; Liu, Chan, & Caplan, 2014). In the JOR procedure, participants are typically given all combinations of list-items. However, at a distance of 1 (nearest-neighbour) pairs, the task is formally equivalent to the order-recognition task investigated here. Thus, if constituent-order is poorer in lists of pairs than in serial lists, we would predict d' to be greater in the JOR procedure, for lists of the same length, than in our procedure here. The closest condition for comparison we could find was LL=10 nouns, derived from Liu et al. (2014), Experiment 1. These lists were comprised of similar materials as ours, but study sets were smaller (LL=10, compared to LL=16 here), which should bias the results toward an advantage for the JOR data. Moreover, Liu et al.'s lists were presented at a rate of one word every 1650 ms, resulting in 3300 ms per "pair," compared to 3000 ms/pair here, which should offer an additional slight advantage to the JOR data. Also note that one group of participants judged which item was the earlier and the other group judged which item was later, so we consider both groups separately. When we computed d' from session 1 of the JOR experiment[5], using only nearest-neighbour probes (distance=1), mean $d' \pm SD = 0.65 \pm 0.22$ and 0.24 ± 0.49 for the Earlier and Later instructions, respectively. These d' values are lower than the values obtained here. Even on the fifth session, JOR participants still performed lower, despite the longer list length and slower presentation rate, $d' = 0.34 \pm 0.58$ and 0.29 ± 0.40, respectively. The greater difficulty of the JOR task is most likely due to those participants needing to consider all possible pairings of list items, not just nearest-neighbour transitions. However, such a result is at odds with the idea that constituent-order memory is severely compromised in lists of pairs compared to serial lists.

Boundary conditions. Although our results are clear and consistent across the three experiments, future studies may identify important boundary conditions. Two such conditions worth noting are presentation rate and strategy. The presentation rate used here is typical of association-memory studies, but if sped up, participants might reveal a limited capacity to store order along with the association. Conversely, when the presentation rate is slowed, rich, elaborative study strategies, such as interactive imagery, become available to participants; these strategies may or may not preserve order-memory well, and it remains to be seen whether they do so in a way that compromises association-memory or not.

Conclusion. Human participants, when studying lists of pairs, appear to learn those pairs along with their order, to a moderate degree, even when constituent-order is not required. Convolution- and matrix-based models, in their current formulations, are insufficient to explain this order-memory ability without compromising other aspects of

[5]Thanks to Yang S. Liu for providing these calculations.

the data. Both classes of model need to be modified with care, and further tested, to accommodate the full set of findings. Finally, the finding that associations are stored with moderate order, without compromising associative symmetry, reinforces the idea that memory for associations closely related to memory for serial lists, and that a broad range of memory phenomena, spanning from conventional paired associate learning to conventional serial learning and beyond, may be modelled within a single theoretical framework.

References

Anderson, J. A. (1970). Two models for memory organization using interacting traces. *Mathematical Biosciences*, *8*, 137-160.

Asch, S. E., & Ebenholtz, S. M. (1962). The principle of associative symmetry. *Proceedings of the American Philosophical Society*, *106*(2), 135-163.

Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. (1975). *Discrete multivariate analysis: theory and practice.* Cambridge, MA: MIT Press.

Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review*, *114*(3), 539-576.

Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review*, *107*(1), 127-181.

Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review*, *106*(3), 551-581.

Caplan, J. B. (2004). Unifying models of paired associates and serial learning: insights from simulating a chaining model. *NeuroComputing*, *58-60*, 739-743.

Caplan, J. B. (2005). Associative isolation: unifying associative and order paradigms. *Journal of Mathematical Psychology*, *49*(5), 383-402.

Caplan, J. B. (2015). Order-memory and association-memory. *Canadian Journal of Experiemntal Psychology*, *69*(3), 221-232.

Caplan, J. B., Boulton, K. L., & Gagné, C. L. (2014). Associative asymmetry of compound words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(4), 1163-1171.

Caplan, J. B., Glaholt, M., & McIntosh, A. R. (2006). Linking associative and list memory: pairs versus triples. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(6), 1244-1265.

Caplan, J. B., Rehani, M., & Andrews, J. C. (2014). Associations compete directly in memory. *Quarterly Journal of Experimental Psychology*, *67*(5), 955-978.

Chan, M., Ross, B., Earle, G., & Caplan, J. B. (2009). Precise instructions determine participants' memory search strategy in judgments of relative order in short lists. *Psychonomic Bulletin & Review*, *16*(5), 945-951.

Criss, A. H., Aue, W. R., & Smith, L. (2011). The effects of word frequency and context variability in cued recall. *Journal of Memory and Language*, *64*(2), 119-132.

Dressler, W. U. (2006). Compound types. In G. Libben & G. Jarema (Eds.), *The representation and processing of compound words* (p. 23-44). Oxford University Press.

Ebbinghaus, H. (1885/1913). *Memory: A contribution to experimental psychology.* New York: Teachers College, Columbia University.

Friendly, M., Franklin, P. E., Hoffman, D., & Rubin, D. C. (1982). The Toronto Word Pool: Norms for imagery, concreteness, orthographic variables, and grammatical usage for 1,080 words. *Behavior Research Methods and Instrumentation*, *14*, 375-399.

Geller, A. S., Schleifer, I. K., Sederberg, P. B., Jacobs, J., & Kahana, M. J. (2007). PyEPL: a cross-platform experiment-programming library. *Behavior Research Methods*, *39*(4), 950-958.

Greene, R. L., & Tussing, A. A. (2001). Similarity and associative recognition. *Journal of Memory and Language*, *45*, 573-584.

Hacker, M. J. (1980). Speed and accuracy of recency judgments for events in short-term memory. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(6), 651-675.

Henson, R. N. A., Norris, D. G., Page, M. P. A., & Baddeley, A. D. (1996). Unchained memory: Error patterns rule out chaining models of immediate serial recall. *Quarterly Journal of Experimental Psychology*, *49A*(1), 80-115.

Hintzman, D. L. (1980). Simpson's paradox and the analysis of memory retrieval. *Psychological Review*, *87*(4), 398-410.

Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers*, *16*(2), 96-101.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, *93*(4), 411-428.

Hintzman, D. L. (2011). Research strategy in the study of memory: fads, fallacies, and the search for the "coordinates of truth". *Psychological Science*, *6*(3), 253-271.

Hockley, W. E. (1984). Retrieval of item frequency information in a continuous memory task. *Memory & Cognition*, *12*(3), 229-242.

Hockley, W. E., & Cristi, C. (1996a). Tests of encoding tradeoffs between item and associative information. *Memory & Cognition*, *24*, 202-216.

Hockley, W. E., & Cristi, C. (1996b). Tests of the separate retrieval of item and associative information using a frequency-judgment task. *Memory & Cognition*, *24*(6), 796-811.

Horowitz, L. M., Brown, Z. M., & Weissbluth, S. (1964). Availability and the direction of associations. *Journal of Experimental Psychology*, *68*(6), 541-549.

Humphreys, M. S., Bain, J. D., & Pike, R. (1989). Different ways to cue a coherent memory system: A theory for episodic, semantic, and procedural tasks. *Psychological Review*, *96*(2), 208-233.

Jacoby, L. L., & Wahlheim, C. N. (2013). On the importance of looking back: the role of recursive remindings in recency judgments and cued recall. *Memory & Cognition*, *41*(5), 625-637.

Jones, M. N., & Mewhort, D. J. K. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review*, *114*(1), 1-37.

Kahana, M. J. (2002). Associative symmetry and memory theory. *Memory & Cognition*, *30*(6), 823-840.

Kahana, M. J. (2012). *Foundations of human memory.* New York: Oxford University Press.

Kahana, M. J., & Caplan, J. B. (2002). Associative asymmetry in probed recall of serial lists. *Memory & Cognition*, *30*(6), 841-849.

Kelly, M. A., Blostein, D., & Mewhort, D. J. K. (2013). Encoding structure in holographic reduced representations. *Canadian Journal of Experimental Psychology*, *67*(2), 79-93.

Klein, K. A., Shiffrin, R. M., & Criss, A. H. (2007). Putting context in context. In J. S. Nairne (Ed.), *The foundations of remembering: essays in honor of Henry L. Roediger, III* (p. 171-189). New York: Psychology Press.

Köhler, W. (1947). *Gestalt psychology.* New York: Liveright.

Kounios, J., Bachman, P., Casasanto, D., Grossman, M., & Smith, W., Roderick W. Yang. (2003). Novel concepts mediate word retrieval from human episodic associative memory: evidence from event-related potentials. *Neuroscience Letters*, *345*, 157-160.

Kounios, J., Smith, R. W., Yang, W., Bachman, P., & D'Esposito, M. (2001). Cognitive association formation in human memory revealed by spatiotemporal brain imaging. *Neuron*, *29*, 297-306.

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior* (p. 112-146). New York, NY: John Wiley and Sons.

Lewandowsky, S., & Farrell, S. (2000). A redintegration account of the effects of speech rate, lexicality, and word frequency in immediate serial recall. *Psychological Research*, *63*, 163-173.

Lewandowsky, S., & Murdock, B. B. (1989). Memory for serial order. *Psychological Review*, *96*(1), 25-57.

Liu, Y. S., Chan, M., & Caplan, J. B. (2014). Generality of a congruity effect in judgements of relative order. *Memory & Cognition*, *42*(7), 1086-1105.

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, *1*(4), 476-490.

Longuet-Higgins, H. C. (1968). Holographic model of temporal recall. *Nature*, *217*, 104.

Macmillan, N. A., & Kaplan, H. L. (1985). Detection theory analysis of group data: estimating sensitivity from average hit and false-alarm rates. *Psychological Bulletin*, *98*(1), 185-199.
Madan, C. R., Glaholt, M. G., & Caplan, J. B. (2010). The influence of item properties on association-memory. *Journal of Memory and Language*, *63*(1), 46-63.
Mandler, G., Rabinowitz, J. C., & Simon, R. A. (1981). Coordinate organization: the holistic representation of word pairs. *American Journal of Psychology*, *94*(2), 209-222.
Metcalfe Eich, J. (1982). A composite holographic associative recall model. *Psychological Review*, *89*(6), 627-661.
Murdock, B. B. (1962). Direction of recall in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *1*(2), 119-124.
Murdock, B. B. (1974). *Human memory: theory and data*. New York: John Wiley & Sons.
Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, *89*(6), 609-626.
Murdock, B. B. (1995). Developing TODAM: three models for serial-order information. *Memory & Cognition*, *23*(5), 631-645.
Murdock, B. B., & Franklin, P. E. (1984). Associative and serial-order information: different modes of operation? *Memory & Cognition*, *12*(3), 243-249.
Muter, P. (1979). Response latencies in discriminations of recency. *Journal of Experimental Psychology: Human Learning and Memory*, *5*(2), 160-169.
Neath, I., & Surprenant, A. M. (2003). *Human memory: an introduction to research, data, and theory* (2nd ed.). Belmont, CA: Thomson/Wadsworth.
Pike, R. (1984). Comparison of convolution and matrix distributed memory systems for associative recall and recognition. *Psychological Review*, *91*(3), 281-294.
Plate, T. A. (1995). Holographic reduced representations. *IEEE Transactions on Neural Networks*, *6*(3), 623-641.
Primoff, E. (1938). Backward and forward associations as an organizing act in serial and in paired-associate learning. *Journal of Psychology*, *5*, 375-395.
Rehani, M., & Caplan, J. B. (2011). Interference and the representation of order within associations. *Quarterly Journal of Experimental Psychology*, *64*(7), 1409-1429.
Rizzuto, D. S., & Kahana, M. J. (2000). Associative symmetry vs. independent associations. *NeuroComputing*, *32-33*, 973-978.
Rizzuto, D. S., & Kahana, M. J. (2001). An autoassociative neural network model of paired-associate learning. *Neural Computation*, *13*, 2075-2092.
Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM— Retrieving Effectively From Memory. *Psychonomic Bulletin & Review*, *4*, 145-166.
Sommer, T., Schoell, E., & Büchel, C. (2008). Associative symmetry of the memory for object–location associations as revealed by the testing effect. *Acta Psychologica*, *128*, 238-248.
Wahlheim, C. N., Maddox, G. B., & Jacoby, L. L. (2014). The role of reminding in the effects of spaced repetitions on cued recall: sufficient but not necessary. *Memory & Cognition*, *40*(1), 94-105.
Willshaw, D. J., Buneman, O. P., & Longuet-Higgins, H. C. (1969). Non-holographic associative memory. *Nature*, *222*, 960-961.
Yang, J., Zhao, P., Zhu, Z., Mecklinger, A., Fang, Z., & Han, L. (2013). Memory asymmetry of forward and backward associations in recognition tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(1), 253-269.
Yntema, D. B., & Trask, F. P. (1963). Recall as a search process. *Journal of Verbal Learning and Verbal Behavior*, *2*, 65-74.

Appendix A

Simulation of a simple matrix model

Items were simulated as 100-dimensional vectors. A pool of N_{pool} items was initially generated as independent, identically distributed values drawn from $N(0, 1)$, which were then normalized to unit length. N_{pool} was set to 96, the number of words used in a session of

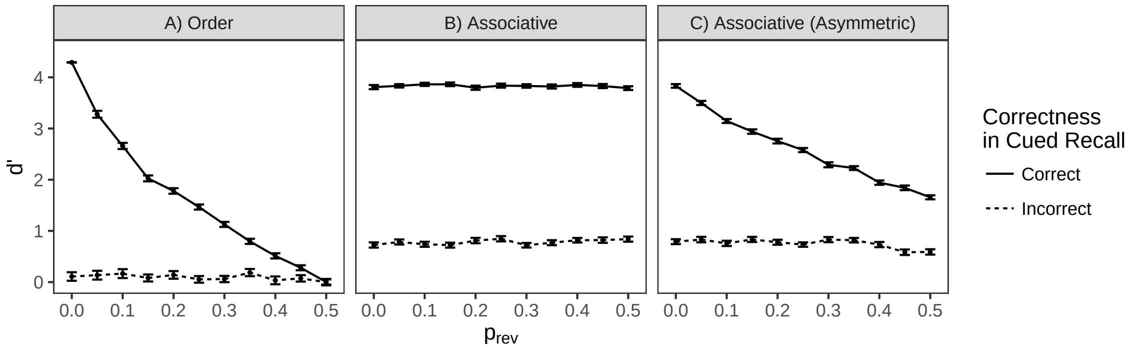


Figure A1. d' values from the simulated matrix model, as a function of cued-recall accuracy (separate lines) and the model parameter, p_{rev}, the probability that the wrong order is stored. **a**, Order recognition. **b**, Associative recognition. **c**, Associative recognition, probing only with the left-hand item in one direction. Note that the asymmetric variant of associative recognition (c) is sensitive to errors in order encoding, whereas the symmetric variant (b) is not. Error bars plot standard error of the mean across simulated participants.

Experiment 1. Memory for each list was stored as $\sum_{i=1}^{L} \gamma_i \mathbf{b}_i \mathbf{a}_i^T$, where $L = 8$, the list length, $\mathbf{a}_i$ and $\mathbf{b}_i$ are items, and $\gamma_i \sim N(\mu, \sigma)$ are the Gaussian-distributed encoding strengths, with $\mu = 1$ and $\sigma = 1$. To implement the idea that the wrong order could be encoded, with probability p_{rev}, $\mathbf{a}_i \mathbf{b}_i^T$ would be stored instead of $\mathbf{b}_i \mathbf{a}_i^T$

Cued recall was implemented by probing in both "directions" at once with probe, $\mathbf{f_x} : \mathbf{f}_r = (M + M^T)\mathbf{f}_x$. Similarity between the retrieved vector, $\mathbf{f}_x$, and each item in the word pool was computed with the dot product, but negative dot products were replaced with zero. A retrieval "strength" was computed as $s = \frac{\mathbf{f}_x \cdot \mathbf{f_t}}{\sum_{i=1}^{N_{pool}} \mathbf{f}_x \cdot \mathbf{f_i}}$. To avoid additional free parameters, we forced cued-recall accuracy to be 0.5 by considering cued recall to be correct for $s > \text{mean}(s)$ and incorrect otherwise.

For order-recognition (Figure A1a), given the probe $\mathbf{f}_L - \mathbf{f}_R$, a strength was computed, $s = (M\mathbf{f}_L) \cdot \mathbf{f}_R - (M\mathbf{f}_R) \cdot \mathbf{f}_L$. The response was "intact" if $s > 0$ and "reverse" otherwise.

For associative recognition (Figure A1b), rearranged probes were simulated by pairing the left-hand item of one pair with a randomly selected associate from a different pair from the same list. Each strength, s, was computed by probing both the original matrix and its transpose (as in cued recall) with the left-hand item and then comparing, with the dot product, to the right-hand probe-item, $s = ((M + M^T)\mathbf{f}_L) \cdot \mathbf{f}_R$. Threshold θ, was set to the mean of all s values for a given model-subject, and the model's response was "intact" when $s > \theta$ and "rearranged" otherwise. In a variant (Figure A1c), only the original matrix, M was probed, with the idea that this would make associative recognition somewhat more directionally dependent, and thus somewhat more decoupled from cued recall, for the same basic reason as for order-recognition.

Simulations were implemented in MATLAB (The Mathworks, Inc., Natick, MA, USA); the main model function and two ancillary functions are included in Appendix B. The simulation was run 100 times with the design 6 lists × 8 pairs/list (with memory, M, reset to zero at the start of each new list), simulating independent subjects.

Appendix B
Simulation code

```
% Simulate matrix model, to compare to Kato and Caplan's data
```

```
function [o_dp,a_dp,o_dp_cc,o_dp_ii,a_dp_cc,a_dp_ii]=matrix_with_order(prev,ARbothsides)

Nsubjects=100; % Number of subjects to simulate
n=100; % dimensionality of item-vectors
LL=8; % #pairs per list
nlists=6; % #lists per subject
nwords=nlists*LL*2; % stimulus pool is just the total used items
mu=1; % mean encoding strength
sigma=1; % SD of encoding strength (across pairs)
% probability the reverse order is encoded. Set to 0 to disable.
if(~exist('prev')), prev=0; end;
% ARbothsides=1: probe simultaneously from both sides
% ARbothsides=0: probe only by multiplying from the right
if(~exist('ARbothsides')), ARbothsides=1; end
rand('state',sum(100*clock)); % seed the random generator

for s=1:Nsubjects
    % Construct items and normalize
    I=randn(n,nwords); mags=sqrt(sum(I.^2)); I=I./(ones(n,1)*mags);
    for l=1:nlists     % Lists are item 1-item 2, item 3-item 4, etc.
        M{l}=zeros(n,n);
        for p=1:LL % encode each pair
            i=(l-1)*LL*2+(p-1)*2+1; % start of the next pair
            if(rand<prev) % sometimes store the reverse direction
                M{l}=M{l}+(mu+sigma*randn)*(I(:,i)*I(:,i+1)');
            else
                M{l}=M{l}+(mu+sigma*randn)*(I(:,i+1)*I(:,i)');
            end
        end
        % ### CUED RECALL ###
        % NOTE: Not modelling output encoding during cued recall,
        % nor successive testing
        % Without loss of generality in this very simple model:
        % only testing forward (but probing both directions at once)
        for p=1:LL % cued recall: probe each pair
            i=(l-1)*LL*2+(p-1)*2+1; % start of the next pair
            % 1 (TRUE, correct) if b.(Ma) > a.(Mb); else 0 (FALSE, incorrect)
            ret=(M{l}+M{l}')*I(:,i); % probe the matrix and its transpose
            % Luce Choice: match to target divided by sum of match to all items
            CR(l,p,s)=truncate(I(:,i+1)'*ret)/sum(truncate(I'*ret));
        end

        % ### "INITIAL" ORDER-RECOGNITION ###
        % To avoid a free parameter (response threshold for order-recognition),
        % we probe simultaneously in both directions and compare
        % strengths. Thus, no bias, and accuracy(intact)=accuracy(reverse)
        for p=1:LL % test each pair
            i=(l-1)*LL*2+(p-1)*2+1; % start of the next pair
```

```
            % Calculate a continuous strength
            orecStr(l,p,s) = ( (I(:,i+1)'*(M{l}*I(:,i))) - (I(:,i)'*(M{l}*I(:,i+1))) );
        end

        % ### "INITIAL"  ASSOCIATIVE RECOGNITION ###
        for p=1:LL % test each pair
            % First, test with an intact probe...
            i=(l-1)*LL*2+(p-1)*2+1;
            if(ARbothsides), MM=M{l}+M{l}'; else, MM=M{l}; end
            % Calculate a continuous strength, intact probe
            arecStrInt(l,p,s)=I(:,i+1)'*(MM*I(:,i));

            % ... then a rearranged probe
            % Choose a random other item as the associate
            availpairs=setdiff(1:LL,p); availpairs=availpairs(randperm(LL-1));
            j=availpairs(1); j=(l-1)*LL*2+(j-1)*2+1;
            % Calculate a continuous strength, rearranged probe
            arecStrRe(l,p,s)=I(:,j+1)'*(MM*I(:,i));
        end
    end

    % Cued Recall: Set threshold for retrievability = mean of CR strength. This
    % forces accuracy in CR to be 0.5.
    CRacc(:,:,s)=CR(:,:,s)>mean(mean(CR(:,:,s)));

    % Order recognition: use 0 as threshold (remember, this is the
    % strength for AB - strength for BA)
    orecAcc(:,:,s)=orecStr(:,:,s)>0;
    % NOTE: no explicit modelling of "reverse" probes because the
    % model is currently symmetric in that regard. So, FAs are 1-hits.
    o_dp(s)=calcdprime(orecAcc(:,:,s),1-orecAcc(:,:,s));
    O=orecAcc(:,:,s); C=CRacc(:,:,s);
    % Calculate d' contingent on CR accuracy
    hits=O(find(C==1)); o_dp_cc(s)=calcdprime(hits,1-hits);
    hits=O(find(C==0)); o_dp_ii(s)=calcdprime(hits,1-hits);

    % Associative Recognition: set threshold to the mean of all strengths
    ar_thresh=mean(mean([arecStrInt(:,:,s) arecStrRe(:,:,s)]));
    arecAccInt(:,:,s)=arecStrInt(:,:,s)>ar_thresh;
    arecAccRe(:,:,s)=arecStrRe(:,:,s)<ar_thresh;
    a_dp(s)=calcdprime(arecAccInt(:,:,s),1-arecAccRe(:,:,s));
    % Calculate d' contingent on CR accuracy
    AI=arecAccInt(:,:,s); AR=arecAccRe(:,:,s);
    hits=AI(find(C==1)); fas=1-AR(find(C==1)); a_dp_cc(s)=calcdprime(hits,fas);
    hits=AI(find(C==0)); fas=1-AR(find(C==0)); a_dp_ii(s)=calcdprime(hits,fas);
end % looping through subjects
```

Exp.	Test Type	Group	N	Target	Lure	d'	C
1	Order	Attend	36	0.850 (0.013)	0.581 (0.013)	1.36 (0.148)	-0.43 (0.039)
	Assoc	Ignore	34	0.892 (0.013)	0.847 (0.013)	2.57 (0.164)	-0.09 (0.054)
2	Order	Attend	114	0.848 (0.008)	0.588 (0.008)	1.42 (0.077)	-0.44 (0.026)
	Assoc	Ignore	114	0.859 (0.006)	0.821 (0.006)	2.34 (0.102)	-0.08 (0.026)
3	Order	Attend	60	0.924 (0.012)	0.671 (0.012)	1.93 (0.106)	-0.45 (0.040)
	Assoc	Ignore	64	0.850 (0.012)	0.777 (0.012)	1.99 (0.119)	-0.12 (0.034)

Table C1

Performance in initial recognition. Mean accuracy, d' and C (bias) are reported for each experiment and each group. Attend - Order-Attend, Ignore - Order-Ignore. N - sample size for each group. Test type: Order - order-recognition, Assoc - associative recognition. Target - intact probes, Lure - reverse (order-recognition) or rearranged (associative recognition) probes. Values in parentheses are the standard error of the mean.

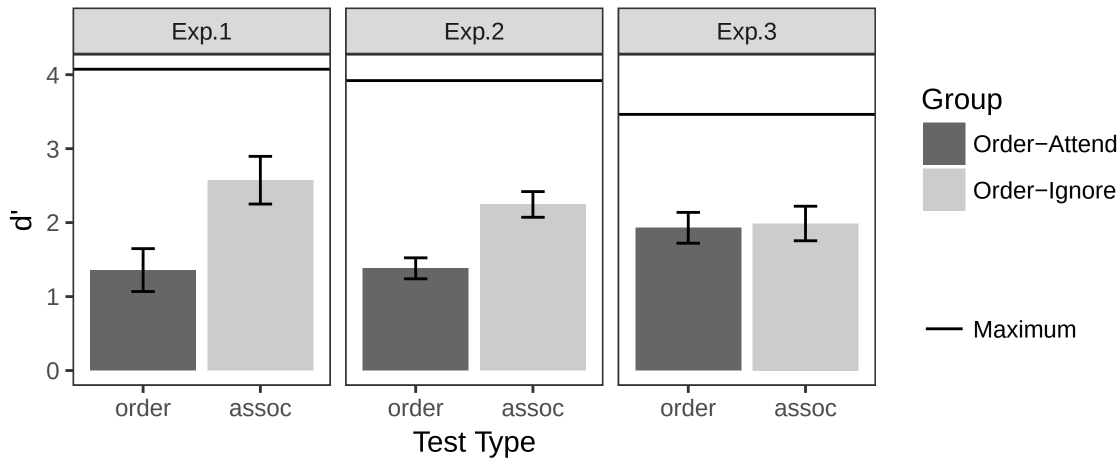


Figure C1. d' for initial recognition. "Maximum" denotes d'_{max}, the highest value possible, given the number of trials in each experiment. Error bars represent 95% confidence intervals based on standard error of the mean. order - order-recognition. assoc - associative recognition.

```
function vals=truncate(vals) % Change negative values to zero
vals(find(vals<0))=0;
```

```
function dp=calcdprime(hits,fas) % compute d'
H=mean(hits(:)); F=mean(fas(:)); corr=.5/length(hits(:)); % .5 observation
if(H==0), H=corr; elseif(H==1), H=1-corr; end;
if(F==0), F=corr; elseif(F==1), F=1-corr; end;
dp=norminv(H)-norminv(F);
```

Appendix C

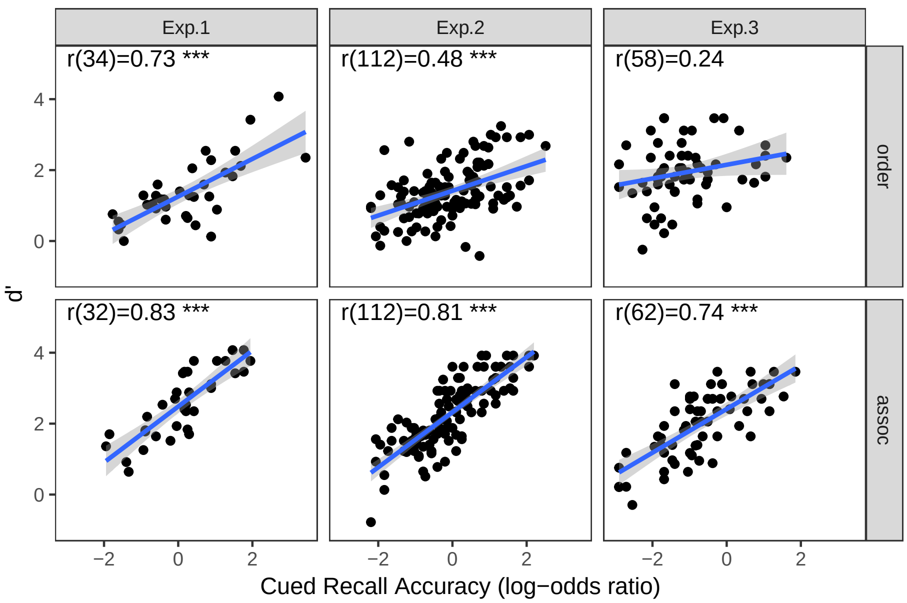
Supplementary tables and figures

The following are tables and figures supplementary to the manuscript.

Exp.	Test Type	Tested	Group	Target	Lure	d'	C
1	Order	Tested	Attend	0.767 (0.017)	0.480 (0.017)	0.748 (0.18)	-0.431 (0.055)
			Ignore	0.785 (0.031)	0.451 (0.031)	0.698 (0.15)	-0.537 (0.117)
	Assoc	Tested	Attend	0.855 (0.013)	0.897 (0.013)	2.589 (0.22)	0.092 (0.055)
			Ignore	0.800 (0.019)	0.893 (0.019)	2.368 (0.24)	0.219 (0.089)
2	Order	Tested	Attend	0.770 (0.014)	0.493 (0.018)	0.800 (0.043)	-0.425 (0.026)
			Ignore	0.811 (0.016)	0.490 (0.019)	0.944 (0.048)	-0.517 (0.026)
		Untested	Attend	0.750 (0.018)	0.509 (0.014)	0.797 (0.043)	-0.376 (0.026)
			Ignore	0.717 (0.019)	0.506 (0.017)	0.658 (0.048)	-0.322 (0.026)
	Assoc	Tested	Attend	0.813 (0.011)	0.815 (0.014)	2.073 (0.038)	0.031 (0.028)
			Ignore	0.769 (0.014)	0.832 (0.015)	1.909 (0.038)	0.134 (0.026)
		Untested	Attend	0.616 (0.019)	0.799 (0.016)	1.326 (0.053)	0.316 (0.028)
			Ignore	0.592 (0.016)	0.810 (0.015)	1.215 (0.038)	0.361 (0.026)
3	Order	Tested	Attend	0.827 (0.016)	0.516 (0.019)	1.065 (0.067)	-0.490 (0.034)
			Ignore	0.847 (0.015)	0.560 (0.029)	1.276 (0.088)	-0.450 (0.039)
		Untested	Attend	0.788 (0.020)	0.649 (0.022)	1.289 (0.067)	-0.224 (0.034)
			Ignore	0.778 (0.025)	0.605 (0.027)	1.167 (0.088)	-0.266 (0.039)
	Assoc	Mixed	Attend	0.675 (0.015)	0.738 (0.015)	1.159 (0.096)	0.096 (0.046)
			Ignore	0.731 (0.012)	0.793 (0.012)	1.672 (0.192)	0.121 (0.045)

Table C2

Performance in final recognition. Mean accuracy, d' and C (bias) are reported for each experiment and each group. Test type: Order - order-recognition, Assoc - associative recognition. Target - accuracy for intact probes, Lure - accuracy for reverse (order-recognition) or rearranged (associative recognition) probes. Tested - pairs that were tested in initial recognition, Untested - pairs previously untested, Mixed - mixed with tested and untested pairs. Attend - Order-Attend group, Ignore - Order-Ignore group. Values in parentheses are the standard error of the mean.



Williams' tests:

z=1.08 z=4.54 *** z=3.81 ***

Figure C2. Correlation between accuracy of cued recall and d' in initial recognition across the experiments. Each dot represents one participant. Shaded regions denote 95% confidence intervals on the linear regression. Williams' tests of the difference in correlation between groups are reported below each experiment. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

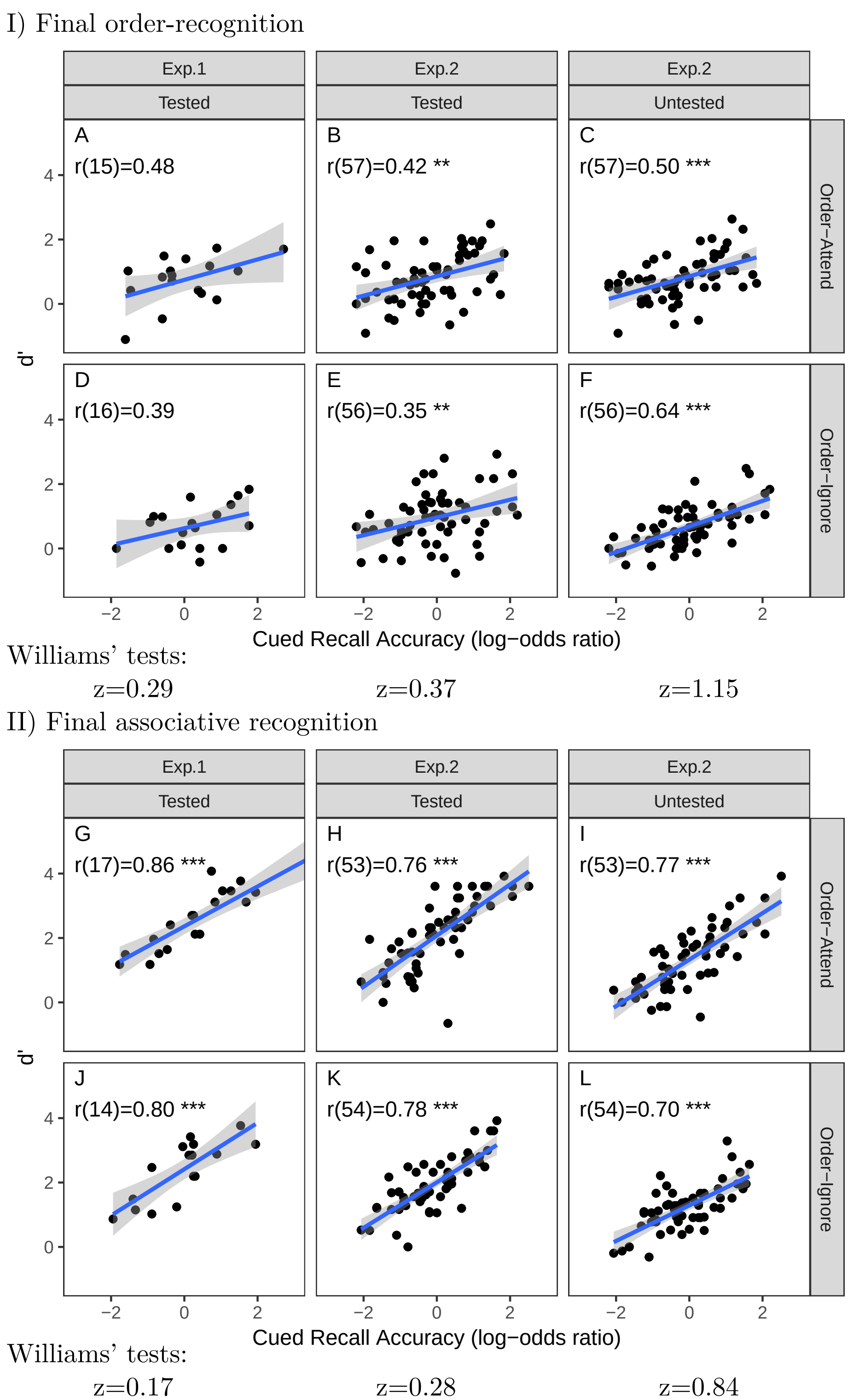


Figure C3. Correlation between accuracy in cued recall and d' in final recognition for each experiment. Each dot represents one participant. Shaded regions denote 95% CIs on the linear regression.

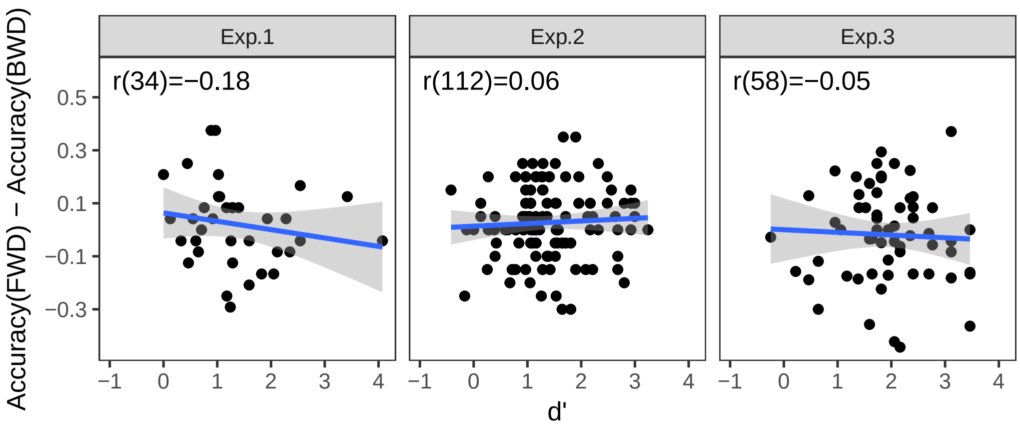


Figure C4. Scatter plots showing the lack of relationship between asymmetry in mean accuracy in cued recall, and d' in order-recognition, in all three experiments. FWD = forward cued recall; BWD = backward cued recall.

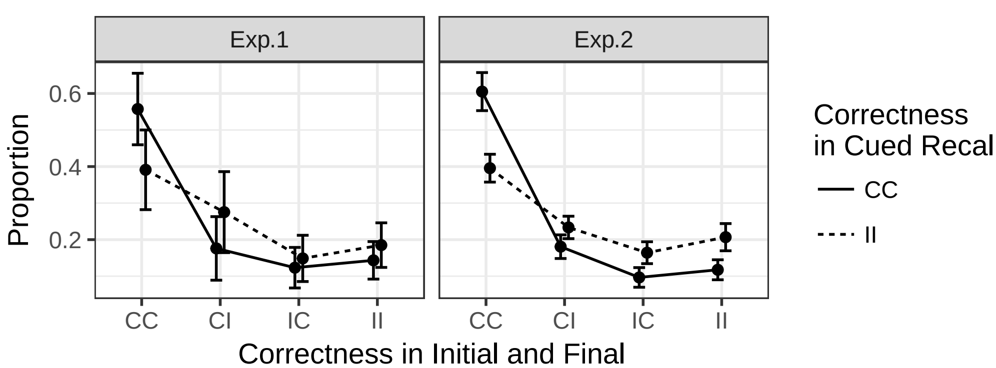


Figure C5. The relationship between accuracy on initial and final order-recognition of a pair, when prior cued recall had been correct twice (CC pairs, solid plot line) or incorrect twice (II pairs, dotted plot line). On the horizontal axis, CI refers to the case for which initial order-recognition was correct but final order-recognition was incorrect for the pair, etc. Error bars plot 95% confidence intervals based on standard error of the mean.