AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychological Review

ISSN: 0033-295X

2021, Vol. 128, No. 6, 1197–1205
https://doi.org/10.1037/rev0000327

THEORETICAL NOTE

Serial Memory: Putting Chains and Position Codes in Context

Gordon D. Logan[1] and Gregory E. Cox[2]
[1] Department of Psychology, Vanderbilt University
[2] Department of Psychology, University at Albany

From the beginning of research on serial memory, chaining theories and position coding theories have been pitted against each other. The central question is whether items are associated with each other or with a set of position codes that are independent of the items. Around the turn of this century, the debate focused on serial recall tasks and patterns of error data that chaining models could not accommodate. Consequently, theories based on other ideas flourished and position coding models became prominent. We present an analysis of a retrieved context model that integrates chains and position codes. Under some parameter values, it produces classic chains. Under most parameter values, it produces context representations that contain information sufficient to specify the position codes in position coding theories. We suggest three ways to extract position codes from context representations and show the codes they produce are mathematically equivalent to the codes in position coding models. The extracted position codes can be substituted for the position codes in position coding models and run through their machinery to mimic their predictions exactly. We suggest that chains, position codes, and retrieved contexts may reflect different strategies for extracting desired information from a common set of memory representations, and we emphasize the value of considering item-dependent context representations that are made from fading traces of past items encoded or retrieved.

Keywords: serial recall, start–end model, context retrieval and updating, retrieved context model, mimicry

> The fact that the model may be indistinguishable scientifically from other quite different models need not be a source of unhappiness. In fact, it is possible to take comfort in such equivalences. If a particular model is equivalent to many other models, we can be more confident in its basic truth. Anderson (1978, p. 275).

The century and a half of research on serial memory has been an epic battle between two theories of how serial order is represented: *chains* versus *position codes*. Ebbinghaus (1885) thought serial learning involves associating each item with the item that follows it, forming chains of associations that represent the list. Ladd and Woodworth (1911) argued that items are associated with representations of the positions they occupied on the list and not with each other. Hull (1932, 1934) argued that maze learning involves building chains of associations backwards from the goal, while Tolman (1948) argued that maze learning involves associating goals and choice points with positions in a cognitive map. Studies of serial learning in the 1960s pitted chaining against position coding (Ebenholtz, 1963; Young, 1961; also see Solway et al., 2012). Recently, the battle shifted to serial recall tasks like the memory span test. Murdock (1982, 1993, 1995) and Lewandowsky (Lewandowsky & Li, 1994; Lewandowsky & Murdock, 1989) and Shiffrin and Cook (1978) proposed chaining models of serial recall. Henson et al. (1996) published a classic article reporting patterns of errors that challenged chaining models of serial recall. Unlike human participants, the models could not recover from errors, produce transpositions, or respond appropriately to manipulations of phonological similarity. Consequently, chaining models were largely dismissed. Instead, there was an explosion of models based on mechanisms other than chaining (Brown et al., 2000, 2007; Burgess & Hitch, 1999; Farrell & Lewandowsky, 2002; Page & Norris, 1998; for comprehensive reviews, see Hurlstone et al., 2014; Lewandowsky & Farrell, 2008). Of these models, position coding was very successful (Anderson & Matessa, 1997; Farrell, 2012; Henson, 1998; Lewandowsky & Farrell, 2008; Oberauer et al., 2012) and has become the dominant theory.

The length and vigor of this battle suggest that chaining and position coding are incompatible and mutually exclusive. The purpose of this article is to show they are not. We will show that both chains and position codes can be obtained from the representations in *retrieved context* models that were developed to explain free recall (Howard & Kahana, 2002; Lohnas et al., 2015; Polyn et al., 2009; Sederberg et al., 2008; Talmi et al., 2019) and applied to serial recall by Logan (2018, 2021; also see Botvinick & Plaut, 2006). In these models, items are associated with contexts that are built from fading traces of previous items, and items are retrieved by matching *stored contexts* to a *current context* that is built from fading traces of previously retrieved items. Retrieved context models are like chaining models in that items are associated with contexts made from other items, and the item that was just retrieved becomes part of the current context that retrieves the next item. Under some parameterizations, they are exactly equivalent. We will show that retrieved context models are also equivalent to position

This article was published Online First September 27, 2021.
Gordon D. Logan https://orcid.org/0000-0002-8301-7726
Gregory E. Cox https://orcid.org/0000-0002-0602-1545
Correspondence concerning this article should be addressed to Gordon D. Logan, Department of Psychology, Vanderbilt University, Nashville, TN 37204, United States. Email: gordon.logan@vanderbilt.edu

coding models in that the stored contexts contain information that supports the construction of position codes that are mathematically equivalent to position codes from established theories of serial order (Farrell, 2012; Henson, 1998; Lewandowsky & Farrell, 2008; Oberauer et al., 2012). When these position codes are run through the machinery of position coding models, they make exactly the same predictions. Thus, retrieved context models can mimic position coding models (Anderson, 1978).

Despite their success, position coding theories have left important questions about the genesis and succession of position codes unanswered. Position coding theories generally do not specify how position codes are generated or how the model progresses from one position code to the next (see Henson & Burgess, 1997 and Brown et al., 2000 for explanations in terms of neural oscillators). The disfavored chaining models provide coherent answers to these questions: Items are coded with respect to each other, so position in the chain represents position in the list, and items are retrieved successively by progressing through the chain, using the retrieved item as the cue for the next retrieval. This is the baby that was thrown out with the bathwater when chaining models were abandoned. Position coding theorists have not replaced the baby, leaving important phenomena of serial order unexplained. We will show that retrieved context models can provide a coherent explanation.

Our reasoning is based on the common idea that different memory tasks reflect different ways of accessing the same underlying memory structures (e.g., recognition and recall; Atkinson & Shiffrin, 1968; Cox et al., 2018; Gillund & Shiffrin, 1984; Humphreys et al., 1989). Chaining and position coding may result from different strategies for using the stored contexts generated in retrieved context theories. We will show that retrieved context theories afford three different strategies for position coding. We derive our results from the context retrieval and updating (*CRU*) theory (Logan, 2018, 2021; Logan et al., 2021), which is a special case of the temporal context model (*TCM*; Howard & Kahana, 2002), which is the basis of modern retrieved context models (Lohnas et al., 2015; Polyn et al., 2009; Sederberg et al., 2008; Talmi et al., 2019). CRU uses the same context updating process as TCM but simplifies the context representations and learning rules.[1] These simplifications allow us to focus more clearly on the context updating process, which is the core assumption about serial order in all retrieved context models, but our conclusions should generalize to TCM more broadly. Viewing CRU as a special case of TCM raises the possibility that a single model can accommodate both serial recall and free recall (Grenfell-Essam et al., 2017; Grenfell-Essam & Ward, 2012; Ward et al., 2010).

Position Codes in Prominent Position Coding Theories

Having convinced the field that chaining theories were untenable (Henson et al., 1996), Henson (1998) presented and tested the influential start–end model (*SEM*). It assumes that items are associated with position codes, $s(i)$ and $e(i)$, that define the position of item i with respect to the start and end of the list. Following Henson (1998),

$$s(i) = S_0 S^{i-1}, \tag{1}$$

$$e(i) = E_0 E^{N-i}, \tag{2}$$

where S_0 and E_0 are *start* and *end markers*, respectively, which represent the maximum values of the start and the end codes, and S and E are decay parameters, which determine how steeply the start and end codes decay across position. The start and end codes are elements of a vector $\boldsymbol{p}(i) = [s(i), e(i)]$ that is used to calculate the similarity or *overlap* between codes representing different positions. The probability of retrieval is a function of the similarity between position codes. Codes for nearby positions are more similar than codes for remote positions, so SEM predicts more confusions between nearby positions than distant ones.

Equations 1 and 2 define SEM's representation of serial order. Henson (1998) also assumes *response suppression* as a core property of SEM, which is necessary to prevent repeated responses. We are concerned with the representation of serial order, so we focus on position codes. If retrieved context theories can mimic the position codes that drive SEM, they can use the rest of SEM's machinery to predict exactly the same behavior.

Lewandowsky and Farrell (2008), among others, represent position codes as random vectors of equal length in which some random proportion of elements changes from one list position to the next (Estes, 1955; Murdock, 1997). This random evolution generates an exponential similarity structure among the position codes such that

$$\boldsymbol{p}_i \cdot \boldsymbol{p}_j = \psi^{|i-j|}, \tag{3}$$

where $\boldsymbol{p}_i \cdot \boldsymbol{p}_j$ is the dot product of position vectors $\boldsymbol{p}_i$ and $\boldsymbol{p}_j$ for positions i and j, which measures similarity, and ψ is the *context drift* parameter, which reflects the proportion of elements that remain the same. The probability of retrieval is a function of similarity, such that nearby positions are confused more than remote ones. Farrell (2012) assumed this exponential similarity structure without specifying the codes (the position vectors) that produce it.[2] If stored contexts in retrieved context theories have the same exponential similarity structure, they can be substituted for position codes in these other theories and predict the same behavior.

An important difference between these position codes and stored contexts is that the position codes do not contain representations of the items in the list but the stored contexts do. As a result, position codes themselves provide no means of progressing through the list and require a mechanism that is able to regenerate the same codes during test that were used during study (e.g., Brown et al., 2000). Instead of relying on a secondary process to generate codes, retrieved context theories add retrieved items to the current context. When retrieval is accurate, this recapitulates at test the way context drifted during study without needing an additional process to (re-)generate codes.

Position Codes in Retrieved Context Theories

Like all retrieved context theories, CRU assumes that retrieval depends on the match between a representation of the *current context* and representations of *stored contexts* in memory

[1] CRU is equivalent to a version of TCM in which the only learning that occurs is from contexts to items. In CRU, the same item always evokes the same context to be integrated into the ongoing temporal context, and so there is no learning from items to context. Instead, the evolving temporal context is associated with each item as it occurs; this gives rise to CRU's stored context vectors. These stored context vectors represent the columns of a matrix of associative weights from contexts to items.

[2] An exponential similarity gradient is also assumed by SIMPLE (Brown et al., 2007), but this gradient depends on the logarithm of the time since an item was encoded, rather than its position relative to the list.

(Howard & Kahana, 2002; Logan, 2018, 2021; Lohnas et al., 2015; Polyn et al., 2009; Sederberg et al., 2008; Talmi et al., 2019). Contexts are built by an updating process that occurs at encoding and retrieval. At encoding, the context initially contains an element that represents the list; later, this same element is used as a cue to initiate retrieval from the beginning of the list. In serial recall, the list representation distinguishes one list from another (Logan, 2021). In skilled typing, the list representation distinguishes one word from another (Logan, 2018). After each item is presented or retrieved, a copy of that item is added to the context following the updating rule:

$$\boldsymbol{c}_{N+1} = \beta \boldsymbol{r}_N + \rho \boldsymbol{c}_N, \tag{4}$$

where $\boldsymbol{c}_N$ is a vector representing the current context on trial N, $\boldsymbol{c}_{N+1}$ is a vector representing the updated context, $\boldsymbol{r}_N$ is a vector representing the Nth item that was presented or retrieved, and β and ρ are weights on new and old information, respectively. All vectors are normalized to length = 1. If the item and the context are orthogonal,

$$\rho = \sqrt{1 - \beta^2}. \tag{5}$$

Otherwise

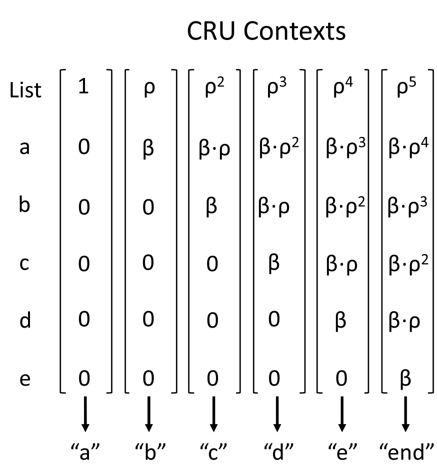
$$\rho = \sqrt{1 + \beta^2[(\boldsymbol{r}_N \cdot \boldsymbol{c}_N)^2 - 1]} - \beta(\boldsymbol{r}_N \cdot \boldsymbol{c}_N), \tag{6}$$

where $\boldsymbol{r}_N \cdot \boldsymbol{c}_N$ is the dot product of the vectors.

Figure 1 illustrates the stored contexts that result from encoding the list "abcde" and associating each item with its context. The list element and the items are represented with localist unit vectors with 1 in the element representing the item and 0 in all other elements. Consequently, the elements in the stored context vectors are values or products of β and ρ. The value for element (item) j in context $\boldsymbol{c}_i$ after the ith item is presented is

$$\boldsymbol{c}_i(j) = \beta\rho^{i-j}, \tag{7}$$

Figure 1
CRU Context Vectors



Note. Context vectors for the list "abcde" constructed from CRU's updating process (Equation 4). The elements in brackets represent context vectors. The first element represents the list and the other elements represent the items on the list. The arrow below each vector represents the association between the vector and the list item. CRU = context retrieval and updating.

if $i \geq j$ and 0 otherwise. The similarities among context vectors, expressed as dot products, are an exponential function of ρ (Logan, 2021):

$$\boldsymbol{c}_i \cdot \boldsymbol{c}_j = \rho^{|i-j|}. \tag{8}$$

The probability of retrieval depends on the similarity of the context vectors. Nearby contexts are confused more often than remote ones. The exponential terms in Equations 7 and 8 provide the basis for position codes in the stored context representations and CRU's ability to mimic position coding models. We considered three possibilities. The first two extract SEM position codes from stored contexts. The third is more general.

Position Codes From Sums of Contexts and Last Contexts

The first possibility builds on information that is available as the list is encoded. Information specifying the start code is available in the sum of the context vectors. The sum of element i over the N items in the list is

$$\beta \sum_{i=j}^{N} \rho^{N-j} = \frac{\beta}{1-\rho}(1 - \rho^{N-i+1}).$$

Rearranging the terms, taking the reciprocal of the above expression, and introducing a constant S_0 results in

$$s(i) = S_0 \frac{1-\rho}{\beta\rho^N} \rho^{i-1}, \tag{9}$$

Equation 9 has the same form as SEM's start code representation (Equation 1): A constant multiplied by an exponential decay, where $S_0 \frac{1-\rho}{\beta\rho^N}$ is the constant corresponding to SEM's start marker and ρ^{i-1} represents the decay of the start codes from the start of the list with ρ as SEM's decay parameter. Because the constant representing the start marker includes the term ρ^N in the denominator, the overall scale of the start codes will increase with list length, N. Whether this affects behavior depends on how the start codes are used in the decision process. Possibly, S_0 could change with list length to rescale the values. Equation 9 reflects the summed strength for item i at the end of the list (after the Nth item is presented), similar to summed strength in models of recognition (Nosofsky et al., 2014).

The last encoded context contains information that specifies an end code that is formally equivalent to SEM's. The last encoded context contains a representation of each list item weighted by an exponential function of ρ that decreases from the last to the first position, like SEM's end code (see Figure 1). To rewrite Equation 2,

$$e(i) = E_0\beta\rho^{N-i}. \tag{10}$$

The information that supports these position codes is readily available to the system online, as the list is presented. The end codes are contained in the most recently updated context regardless of list length, providing SEM with a way to generate end codes from lists of uncertain length (cf. Henson & Burgess, 1997). The sum across contexts can also be generated as the list is presented. Summing activation is a common process in computational models, so little additional machinery needs to be added. Associating the position codes with the items is more complex, as it requires accessing individual elements of the summed contexts and the final context. It should be possible to select the *j*th element of a vector by multiplying the vector with a unit vector with 1 in the *j*th position. This operation applied to the sum of the contexts and the last context would give the start and end components of SEM's position codes, and CRU could be run on the stored contexts to retrieve the items to be associated with the position codes. This would happen after the list was presented when all the required contexts would have been encoded. After the items have been associated with position codes, they can be processed by the rest of SEM's machinery and predict the same results as SEM's position codes (Equations 1 and 2).

Position to Codes From Similarities First and Last Contexts

The second possibility for deriving position codes from CRU is to use the similarities between contexts expressed in Equation 8. Start codes can be constructed from the dot products of the initial (starting) context and each of the contexts in the list, which decrease as an exponential function across the list. Thus,

$$s(i) = S_0 \rho^{i-1}. \tag{11}$$

End codes can be constructed from the dot products of the last context and each context in the list, which decrease by the same exponential function of distance:

$$e(i) = E_0 \rho^{N-i}. \tag{12}$$

These position codes would also have to be constructed after the context representations have been encoded. Once they are available and associated with the items, SEM's machinery can take over and make the same predictions as SEM's position codes.

This way of generating position codes assumes that start and end codes decrease at the same rate, so $S = E = \rho$. Henson (1998) assumed $S > E$. His simulations fixed $S = .80$ and $E = .48$ ($E/S = .60$). Henson accounted for data with simulations rather than model fits, so it is unclear whether his inequality is necessary. To find out, we fit SEM to serial recall data from Logan (2021) and allowed S and E to vary freely, and we compared the fit with a version of SEM in which S was constrained to equal E. We compared the S and E parameters in the unconstrained fits and compared the goodness of fit of the constrained and unconstrained models. We were interested in SEM's serial order mechanisms, so we did not implement its other components, like response suppression.

We fit the serial recall data from Logan's (2021) Experiments 1 and 2. Experiment 1 used lists of 5, 6, and 7 random letters and Experiment 2 used 6 letter lists and varied repetition. Half of the lists contained a repeated letter and half contained unique letters. There were 192 lists per participant and 24 participants in each experiment. Each participant was fitted separately and independently. The details of the fitting procedure are presented in the Appendix. The mean values of the best-fitting parameters and the measures of goodness of fit (Bayesian information criterion [BIC]) are presented in Table 1. We do not report the empirical predictions of the models. SEM's ability to fit serial recall data is not in question. The issue is which values of S and E will be necessary to fit the data, and the parameter values and goodness of fit statistics address that issue directly.

Table 1

Mean Parameter Values and Measures of Goodness of Fit for Fits of the Unconstrained ($S \neq E$) and Constrained ($S = E$) Versions of SEM to Serial Recall Data From Logan (2021)

Model	S_0	E_0	S	E	Noise *SD*	BIC
Experiment 1						
$S \neq E$	9.5214	3.9464	.8490	.7747	.1608	5073.95
$S = E$	9.8103	4.2162	***.8464***	***.8464***	.1414	5137.37
Experiment 2						
$S \neq E$	12.9178	5.8187	.8450	.8354	.0558	3294.36
$S = E$	12.0630	5.3774	***.8416***	***.8416***	.1293	3304.99

Note. S_0 = start marker; E_0 = end marker; S = start decay; E = end decay; Noise *SD* = standard deviation of Gaussian noise added to the overlap scores; SEM = start–end model; BIC = bayesian information criterion = −2 log likelihood + k log(N), where k is the number of parameters and N is the number of observations. Values of S and E in bold italics were constrained to be equal.

In the $S \neq E$ unconstrained fits in Experiment 1, S and E were similar. The mean difference was .0743 and the E/S ratio was .9106. The difference was smaller and the ratio was larger than Henson's, but the difference was positive for all 24 participants. The $S = E$ model fit worse than the $S \neq E$ model in all 24 participants, but the BIC difference was small (63.42) compared to the overall BIC.

In the $S \neq E$ unconstrained fits in Experiment 2, S and E were very close. The mean difference was .0096 and the E/S ratio was .9899. The difference was positive in 14 participants. The $S = E$ model fit worse than the $S \neq E$ model in 19 participants, but the BIC difference was very small (10.63). Together, the fits suggest that start and end decay are approximately equal, which is consistent with the position codes derived from CRU's stored contexts (Equations 11 and 12). Unequal decay rates may be explained by variability in the rate of context drift across the list (a mechanism that has been implemented in CRU; Logan, 2021; Logan et al., 2021).

Fits of CRU to the same data provided estimates of ρ for each participant (Logan, 2021, nondecrease models, in which β was constant across serial position). For Experiment 1, the mean $\rho = .8929$, which is similar to but larger than the estimates of S and E in the $S = E$ model in 22 participants. The correlation between ρ and $S = E$ was .3930, $t(22) = 2.0047$, $p = .0575$. For Experiment 2, the mean $\rho = .8687$, which is larger than the estimates of S and E in the $S = E$ model in 22 participants. The correlation between ρ and $S = E$ was .3855, $t(22) = 1.9599$, $p = .0628$. The differences may be attributable to ancillary assumptions about the processes that extract information from the similarity structure, select responses, and (in SEM) suppress responses.

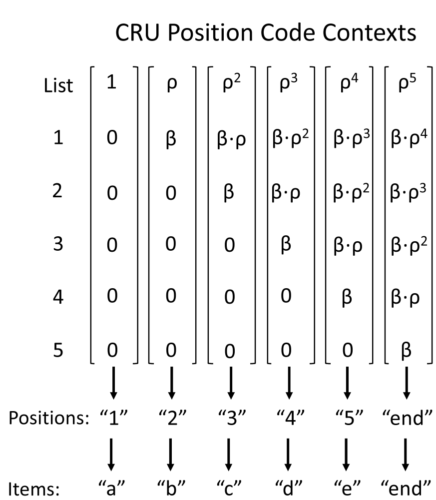
We are not claiming that fitting CRU and SEM (or serial order in a box [*SOB*]) to serial recall data will produce position codes with exactly the same values. The variables that determine the confusability of position codes interact with the decision processes (and the other assumptions), which differ between the models. For example, in CRU, the contextual drift parameter β trades off with the threshold parameter θ in the racing diffusion decision process, such that increases in β can be compensated for by lowering the threshold to yield the same accuracy—essentially a speed–accuracy tradeoff: response time (RT) = threshold/rate, approximately, so RT = θ/similarity = $\theta/\rho^{|i-j|}$. As β increases, ρ decreases, so θ must decrease to maintain the ratio. Because of this tradeoff, position codes calculated from β and ρ values will depend on θ, so we fix it at 200 in some applications (Logan, 2018, 2021). SEM and SOB have their own decision processes that produce similar tradeoffs. Consequently, the position codes estimated from CRU would not necessarily equal the position codes estimated from SEM or SOB. They are equivalent mathematically but estimated values may differ when they are estimated with different models.

Position Codes as Retrieved Contexts

The third possibility for deriving position codes from CRU is to use the updating process (Equation 4) to form a set of generic contexts that represent a sequence of positions, like "first, second, third, fourth, fifth." The generic contexts are associated with position codes, and then the position codes are associated with items on the list (see Figure 2). During recall, the system can step through contexts using CRU's updating mechanism, retrieve the position codes, and then retrieve the items associated with them. Unlike the other possibilities, this one assumes that position coding can occur online, during encoding, by associating items with a preexisting set of generic contexts that represent positions. This version may be more similar in spirit to current position coding theories because it presumes a set of position codes that can be associated with the items.

This model mimics position coding models that associate items with contexts that drift independently of the items (Farrell, 2012; Lewandowsky & Farrell, 2008; Oberauer et al., 2012). Independently drifting contexts show the same exponential decay of similarity across position as CRU's contexts (compare Equations 3 and 8), so one set of contexts can be substituted for the other and make the same predictions. This version explains how position codes are generated (through context updating at encoding) and how the system steps through them at retrieval (through context updating at retrieval). Thus, the model replaces the baby that was thrown out with the bathwater.

Figure 2
CRU Position Codes



Note. Context vectors and associations for position codes "12345" in CRU. The vectors are acquired prior to the experiment. At encoding, CRU steps through the vectors as each item is presented. CRU retrieves a position code from the stored contexts and associates the position code with the current item. At retrieval, CRU steps through the stored contexts, retrieves the position associated with each context, and retrieves the item associated with the position. CRU = context retrieval and updating.

Chaining in Retrieved Context Theories

Classical chaining theories assume that each item is associated only with the item that follows it (Ebbinghaus, 1885; Ebenholtz, 1963; Hull, 1932, 1934; Young, 1961). CRU assumes each item is associated with a context that contains fading traces of previous items, which could be construed as associating each item with all previous items (Murdock, 1995) instead of just the one before it. Thus, CRU is not a classic chaining theory. However, it is possible to configure CRU as a classic chaining theory by setting $\beta = 1$. Then, $\rho = 0$ and all the terms including ρ become 0 (see Figure 3). CRU's context updating and retrieval process (Equation 4) will still retrieve the items in order. The list cue will retrieve the first item, the first item will retrieve the second, and so on, like a classical chain. However, the context representations that support classical chaining will not support any of the three ways to derive position codes. The sum of elements across contexts produce a vector of 1's with no gradient across position. The last context only contains information about the last item. Element values do not decay slowly as more items are added (Equation 7) and similarity does not decrease gradually with distance (Equation 8). A chain of positions could be associated with items (adding the associations in Figure 2 to the chains in Figure 3), but the position codes would not have the required similarity structure (Equation 3). Thus, classical chains, as implemented in CRU, cannot be used to derive position codes. As theorists have argued throughout history, classical chains and position codes are incompatible. The new insight here is that classical chains and position codes can be derived from different parameterizations of CRU: $\beta = 1$ produces classical chains; $\beta < 1$ produces position codes.

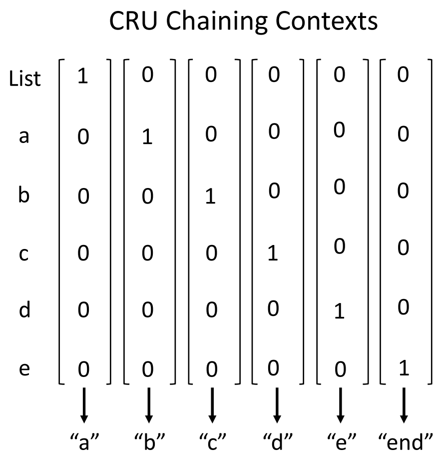
When $\beta < 1$, items are associated with a context that represents recent items more strongly than older items. This enables CRU to mimic particular kinds of compound chaining theories, such as those in which inter-item associations are the product of a limited capacity rehearsal buffer (Atkinson & Shiffrin, 1968). For example, the rehearsal buffer in search of associative memory (*SAM*) assumes that items are dropped from the buffer in proportion to how long they have stayed in the buffer (Phillips et al., 1967; Raaijmakers & Shiffrin, 1981), and the degree to which items are associated to one another is a function of the amount of time they are simultaneously present in the buffer. In combination, these operations can lead to an approximately exponential falloff in associative strength between items as a function of their difference in list positions, once that difference is larger than the size of the buffer (Howard & Kahana, 1999). The CRU context can thus be related to the probability that a prior item is still present in a rehearsal buffer of size one. When $\beta = 1$, associations between items and contexts represent only "isolated" pairs. When $\beta < 1$, associations between items and contexts place each item in the broader context of the rest of the list (cf. Caplan, 2005).

Implications

Mimicry

Our analyses suggest that information about items' positions is naturally encoded in the context representations in CRU and other

Figure 3
Classical Chaining in CRU



Note. CRU produces a classic chaining model if β = 1 and ρ = 0. CRU work as usual, reproducing the sequence. None of the three ways to extract position codes from stored contexts work in this case. All sums of elements across contexts equal 1 and the last context specifies only the last item. The dot product of the first context with the set of contexts equals 1 for the first context and 0 for all other contexts. The dot product of the last context equals 1 for the last context and 0 for all other contexts. More generally, the dot product for context i and context j is 0 if $i \neq j$, so similarity between contexts does not decrease as an exponential function of distance. CRU = context retrieval and updating.

retrieved context theories. Position codes extracted from CRU have the same mathematical form as position codes in SEM and other position coding theories. Thus, CRU's position codes can be substituted for SEM's and run through SEM's machinery to predict the same results, mimicking them exactly. This does not imply that the fit of the full CRU model to a data set will mimic the fit of the full SEM model or other position coding models. Instead, it implies that the difference in the fits is due to ancillary assumptions about response suppression, decision processes, and so on, and not to the assumptions about the representation of serial order. Differences between position coding and chaining models that have been attributed to *representations* may be due to the different *processes* invoked to explain how behavior arises from those representations.

Different models of phenomena like serial recall are usually viewed as incompatible and mutually exclusive, and the purpose of modeling is usually viewed as determining which one provides the best account. This approach focuses on differences among models instead of commonalities. Our focus on commonalities suggests that a model's representations may contain information beyond that required for the specific tasks being modeled. We found position codes buried in CRU's context representations. Who knows what might be unearthed in other models' representations?

Models or Strategies?

Our analyses suggest that it might be useful to think of models as representing different strategies for accessing information encoded in a common set of representations (Gillund & Shiffrin, 1984; Humphreys et al., 1989). CRU's context representations can support a variety of strategies for serial recall, from position coding to chaining. Different strategies may be appropriate for different conditions. People might use chains to represent overlearned lists like the alphabet, the series of numbers in counting, or PINs and passwords. They might use position codes to represent series in which ordinal position is important, like the series of presidents, Superbowl winners, or the colleagues in the offices down the hall. They might use associations with context to retrieve items when order is not important (e.g., in free recall; Howard & Kahana, 2002; Lohnas et al., 2015; Polyn et al., 2009; Sederberg et al., 2008; Talmi et al., 2019). Our results suggest that the memory representations underlying all of these strategies may be the stored contexts in CRU and other retrieved context models. We have shown they contain information that can support both chains and position codes. Recently, Caplan (2015) made a similar point.

We think that CRU might be the default encoding strategy. People remember what they attend to, so memory records the trajectory of attention through a task environment (Kirsner & Dunn, 1985; Landauer, 1975; Logan & Etherton, 1994). The item in the current focus of attention is encoded in the context in which it appears, applying β to the focal item and ρ to the context. Recently focused items become part of the context in which the currently focused item is encoded, producing CRU-like representations that contain information about serial order. We assume this encoding is an obligatory consequence of attention (Logan, 1988), so behaving in any environment will produce a set of CRU-like records of the experience (Kragel & Voss, 2021).

The obligatory nature of item–context encoding helps explain why responding in free recall tends to be serially ordered even when this is not required by the task (Klein et al., 2005), why recognition is facilitated when items are tested in the same order they were studied (Kachergis et al., 2013; Schwartz et al., 2005), and why memory for pairs reflects to an extent the order in which they were presented (Kato & Caplan, 2017; Yang et al., 2013). In addition, similar types of item–context associations have been invoked to explain language learning and comprehension (Elman, 1991; Howard et al., 2011), event perception (Reynolds et al., 2007), music cognition (Cox, 2010), and typewriting (Logan, 2018). The prominence of item–context representations across domains lends support to the idea that they are encoded by default and are therefore available to support a wide variety of cognitive abilities.

From this perspective, position coding is a strategic option that is deliberately chosen to support performance. It requires going beyond the default encoding process and it requires computations on CRU representations to extract position codes (Possibilities 1 and 2) or a choice to use generic contexts to construct position codes (Possibility 3). Of course, this is speculation and future research will be required to determine whether it is reasonable to think of different models as different strategies and to ask which strategies are defaults or options. An important goal for future research is to extend these ideas about strategies to achieve a theoretical rapprochement between serial recall and free recall, along with other memory phenomena (Grenfell-Essam et al., 2017; Grenfell-Essam & Ward, 2012; Ward et al., 2010).

A Broader View of Context

The idea that items are associated with contexts is pervasive in theories of memory (Cox & Shiffrin, 2017, in press;

Criss & Shiffrin, 2004; Dennis & Humphreys, 2001; Gillund & Shiffrin, 1984; Howard & Kahana, 2002; Humphreys et al., 1989; Lohnas et al., 2015; Murnane et al., 1999; Osth & Dennis, 2015; Polyn et al., 2009; Tulving & Thomson, 1973). Theories of serial recall also invoke contextual associations. Except for the primacy model, which represents order as differential activation across items (Page & Norris, 1998), all theories of serial recall assume items are associated with contexts. They differ primarily in their assumptions about the nature of the contexts. We view position codes as contexts. Position codes are represented as vectors, like other contexts (Lewandowsky & Farrell, 2008; Oberauer et al., 2012), which can be expanded to include hierarchical representations of position in grouped lists (Farrell, 2012; Henson, 1998).

By definition, position codes represent contexts that are independent of the items. We believe the impetus for that definition was to sharpen the distinction between position codes and classical chains, in which the only context is the preceding item. Retrieved context theories provide a different perspective, representing contexts as fading traces of prior items, so context is dependent on the items. We suggest that both item-dependent and item-independent contexts may be associated with items, reflecting the many dimensions that contribute to context at any given moment (Klein et al., 2007) as well as the flexibility with which different contextual features may be used to support retrieval (cf. Anderson & Pichert, 1978). We suspect that item-independent context forms the background, while item-dependent context forms the foreground. We attend to things that are relevant to the task at hand, and item-dependent contexts reflect recent foci of attention, which are often relevant to current processing. Outside of serial recall experiments, many cognitive activities exploit item-dependent contexts. In multistep tasks like cooking, current activities depend on past activities. In language, the current word is interpreted in the context of previous words and the ideas they expressed. Serial recall tasks require people to attend to the items and their order. We should expect the currently attended item to be encoded in the context of recently attended items. The kernel of truth in classical chaining theories is that previous items provide a context that supports retrieval of the next item. The item-dependent codes in retrieved context models provide a bridge between chains and the item-independent codes in position coding theories.

Conclusions

We have demonstrated mimicry between position coding and retrieved context representations of serial order. Both position codes and chains can be extracted from CRU's context representations. As Anderson (1978) suggested in the quote that began this note, our demonstration of mimicry brings us comfort instead of dismay. The mimicry reveals a basic truth that all models share: Similarity decreases exponentially with distance in a list. The simple exponential function $similarity = \psi^{|i-j|}$ is the key idea that explains serial memory no matter how it is modeled (cf. Shepard, 1987). Without formal modeling and careful distinction between representation and process, this deep commonality would not have been apparent.

References

Anderson, J. R. (1978). Arguments concerning representations for mental imagery. *Psychological Review*, *85*(4), 249–277. https://doi.org/10.1037/0033-295X.85.4.249

Anderson, J. R., & Matessa, M. (1997). A production system theory of serial memory. *Psychological Review*, *104*(4), 728–748. https://doi.org/10.1037/0033-295X.104.4.728

Anderson, R. C., & Pichert, J. W. (1978). Recall of previously unrecallable information following a shift in perspective. *Journal of Verbal Learning and Verbal Behavior*, *17*(1), 1–12. https://doi.org/10.1016/S0022-5371(78)90485-1

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2, pp. 89–195). Academic Press.

Botvinick, M. M., & Plaut, D. C. (2006). Short-term memory for serial order: A recurrent neural network model. *Psychological Review*, *113*(2), 201–233. https://doi.org/10.1037/0033-295X.113.2.201

Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review*, *114*(3), 539–576. https://doi.org/10.1037/0033-295X.114.3.539

Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review*, *107*(1), 127–181. https://doi.org/10.1037/0033-295X.107.1.127

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. https://doi.org/10.1037/0033-295X.106.3.551

Caplan, J. B. (2005). Associative isolation: Unifying associative and list memory. *Journal of Mathematical Psychology*, *49*(5), 383–402. https://doi.org/10.1016/j.jmp.2005.06.004

Caplan, J. B. (2015). Order-memory and association-memory. *Canadian Journal of Experimental Psychology*, *69*(3), 221–232. https://doi.org/10.1037/cep0000052

Cox, G. E. (2010). On the relationship between entropy and meaning in music: An exploration with recurrent neural networks. In S. Ohisson & R. Catrambone (Eds.), *Proceedings of the32nd annual conference of the Cognitive Science Society* (pp. 429–434). Cognitive Science Society.

Cox, G. E., Hemmer, P., Aue, W. R., & Criss, A. H. (2018). Information and processes underlying semantic and episodic memory across tasks, items, and individuals. *Journal of Experimental Psychology: General*, *147*(4), 545–590. https://doi.org/10.1037/xge0000407

Cox, G. E., & Shiffrin, R. M. (2017). A dynamic approach to recognition memory. *Psychological Review*, *124*(6), 795–860. https://doi.org/10.1037/rev0000076

Cox, G. E., & Shiffrin, R. M. (in press). Computational models of event memory. In M. J. Kahana & A. Wagner (Eds.), *Oxford handbook of human memory*. Oxford University Press.

Criss, A. H., & Shiffrin, R. M. (2004). Context noise and item noise jointly determine recognition memory: A comment on Dennis and Humphreys (2001). *Psychological Review*, *111*(3), 800–807. https://doi.org/10.1037/0033-295X.111.3.800

Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review*, *108*(2), 452–478. https://doi.org/10.1037/0033-295X.108.2.452

Ebbinghaus, H. (1885). *Über das gedachtnis: Untersuchungen zur Experimentellen Psychologie*. Dunker.

Ebenholtz, S. M. (1963). Serial learning: Position learning and sequential associations. *Journal of Experimental Psychology*, *66*(4), 353–362. https://doi.org/10.1037/h0048320

Elman, J. L. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning*, *7*(2), 195–225. https://doi.org/10.1007/BF00114844

Estes, W. K. (1955). Statistical theory of spontaneous recovery and regression. *Psychological Review*, *62*(3), 145–154. https://doi.org/10.1037/h0048509

Farrell, S. (2012). Temporal clustering and sequencing in short-term memory and episodic memory. *Psychological Review*, *119*(2), 223–271. https://doi.org/10.1037/a0027371

Farrell, S., & Lewandowsky, S. (2002). An endogenous distributed model of ordering in serial recall. *Psychonomic Bulletin & Review*, *9*(1), 59–79. https://doi.org/10.3758/BF03196257

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, *91*(1), 1–67. https://doi.org/10.1037/0033-295X.91.1.1

Grenfell-Essam, R., & Ward, G. (2012). Examining the relationship between free recall and immediate serial recall: The role of list length, strategy use, and test expectancy. *Journal of Memory and Language*, *67*(1), 106–148. https://doi.org/10.1016/j.jml.2012.04.004

Grenfell-Essam, R., Ward, G., & Tan, L. (2017). Common modality effects in immediate free recall and immediate serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(12), 1909–1933. https://doi.org/10.1037/xlm0000430

Henson, R. N. A. (1998). Short-term memory for serial order: The Start-End Model. *Cognitive Psychology*, *36*(2), 73–137. https://doi.org/10.1006/cogp.1998.0685

Henson, R. N. A., & Burgess, N. (1997). Representations of serial order. In J. A. Bullinaria, D. W. Glasspool, & G. Houghton (Eds.), *Perspectives in neural computing* (pp. 1–14). Springer.

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*, *49*(1), 80–115. https://doi.org/10.1080/713755612

Howard, M. W., & Kahana, M. J. (1999). Contextual variability and serial position effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(4), 923–941. https://doi.org/10.1037/0278-7393.25.4.923

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, *46*(3), 269–299. https://doi.org/10.1006/jmps.2001.1388

Howard, M. W., Shankar, K. H., & Jagadisan, U. K. (2011). Constructing semantic representations from a gradually-changing representation of temporal context. *Topics in Cognitive Science*, *3*(1), 48–73. https://doi.org/10.1111/j.1756-8765.2010.01112.x

Hull, C. L. (1932). The goal-gradient hypothesis and maze learning. *Psychological Review*, *39*(1), 25–43. https://doi.org/10.1037/h0072640

Hull, C. L. (1934). The concept of the habit-family hierarchy and maze learning: Part I. *Psychological Review*, *41*(1), 33–54. https://doi.org/10.1037/h0070758

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. https://doi.org/10.1037/0033-295X.96.2.208

Hurlstone, M. J., Hitch, G. J., & Baddeley, A. D. (2014). Memory for serial order across domains: An overview of the literature and directions for future research. *Psychological Bulletin*, *140*(2), 339–373. https://doi.org/10.1037/a0034221

Kachergis, G., Cox, G. E., & Shiffrin, R. M. (2013). The effects of repeated sequential context on recognition memory. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th annual conference of the Cognitive Science Society* (pp. 704–709). Cognitive Science Society.

Kato, K., & Caplan, J. B. (2017). Order of items within associations. *Journal of Memory and Language*, *97*, 81–102. https://doi.org/10.1016/j.jml.2017.07.001

Kirsner, K., & Dunn, J. C. (1985). The perceptual record: A common factor in repetition priming and attribute retention. In M. I. Posner & O. S. M. Marin (Eds.), *Attention and performance XI* (pp. 547–565). Routledge.

Klein, K. A., Addis, K. M., & Kahana, M. J. (2005). A comparative analysis of serial and free recall. *Memory & Cognition*, *33*(5), 833–839. https://doi.org/10.3758/BF03193078

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* (pp. 171–189). Psychology Press.

Kragel, J. E., & Voss, J. L. (2021). Temporal context guides visual exploration during scene recognition. *Journal of Experimental Psychology: General*, *150*(5), 873–889.

Ladd, G. T., & Woodworth, R. S. (1911). *Elements of physiological psychology: A treatise of the activities and nature of the mind from the physical and experimental point of view*. Charles Scribner's Sons.

Landauer, T. K. (1975). Memory without organization: Properties of a model with random storage and undirected retrieval. *Cognitive Psychology*, *7*(4), 495–531. https://doi.org/10.1016/0010-0285(75)90020-1

Lewandowsky, S., & Farrell, S. (2008). Short-term memory: New data and a model. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 49, pp. 1–48). Academic Press.

Lewandowsky, S., & Li, S.-C. (1994). Memory for serial order revisited. *Psychological Review*, *101*(3), 539–543. https://doi.org/10.1037/0033-295X.101.3.539

Lewandowsky, S., & Murdock, B. B., Jr. (1989). Memory for serial order. *Psychological Review*, *96*(1), 25–57. https://doi.org/10.1037/0033-295X.96.1.25

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*(4), 492–527. https://doi.org/10.1037/0033-295X.95.4.492

Logan, G. D. (2018). Automatic control: How experts act without thinking. *Psychological Review*, *125*(4), 453–485. https://doi.org/10.1037/rev0000100

Logan, G. D. (2021). Serial order in perception, memory, and action. *Psychological Review*, *128*(1), 1–44. https://doi.org/10.1037/rev0000253

Logan, G. D., Cox, G. E., Annis, J., & Lindsey, D. R. B. (2021). The episodic flanker effect: Memory retrieval as attention turned inward. *Psychological Review*, *128*(3), 397–445. https://doi.org/10.1037/rev0000272

Logan, G. D., & Etherton, J. L. (1994). What is learned during automatization? The role of attention in constructing an instance. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *20*(5), 1022–1050. https://doi.org/10.1037/0278-7393.20.5.1022

Lohnas, L. J., Polyn, S. M., & Kahana, M. J. (2015). Expanding the scope of memory search: Modeling intralist and interlist effects in free recall. *Psychological Review*, *122*(2), 337–363. https://doi.org/10.1037/a0039036

Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, *89*(6), 609–626. https://doi.org/10.1037/0033-295X.89.6.609

Murdock, B. B. (1993). TODAM2: A model for the storage and retrieval of item, associative, and serial-order information. *Psychological Review*, *100*(2), 183–203. https://doi.org/10.1037/0033-295X.100.2.183

Murdock, B. B. (1995). Developing TODAM: Three models for serial-order information. *Memory & Cognition*, *23*(5), 631–645. https://doi.org/10.3758/BF03197264

Murdock, B. B. (1997). Context and mediators in a theory of distributed associative memory (TODAM2). *Psychological Review*, *104*(4), 839–862. https://doi.org/10.1037/0033-295X.104.4.839

Murnane, K., Phelps, M. P., & Malmberg, K. (1999). Context-dependent recognition memory: The ICE theory. *Journal of Experimental Psychology: General*, *128*(4), 403–415. https://doi.org/10.1037/0096-3445.128.4.403

Nosofsky, R. M., Cox, G. E., Cao, R., & Shiffrin, R. M. (2014). An exemplar-familiarity model predicts short-term and long-term probe recognition across diverse forms of memory search. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(6), 1524–1539. https://doi.org/10.1037/xlm0000015

Oberauer, K., Lewandowsky, S., Farrell, S., Jarrold, C., & Greaves, M. (2012). Modeling working memory: An interference model of complex

span. *Psychonomic Bulletin & Review*, *19*(5), 779–819. https://doi.org/10.3758/s13423-012-0272-4
Osth, A. F., & Dennis, S. (2015). Sources of interference in item and associative recognition memory. *Psychological Review*, *122*(2), 260–311. https://doi.org/10.1037/a0038692
Page, M. P., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review*, *105*(4), 761–781. https://doi.org/10.1037/0033-295X.105.4.761-781
Phillips, J. L., Shiffrin, R. M., & Atkinson, R. C. (1967). Effects of list length on short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *6*(3), 303–311. https://doi.org/10.1016/S0022-5371(67)80117-8
Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review*, *116*(1), 129–156. https://doi.org/10.1037/a0014420
Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, *88*(2), 93–134. https://doi.org/10.1037/0033-295X.88.2.93
Reynolds, J. R., Zacks, J. M., & Braver, T. S. (2007). A computational model of event segmentation from perceptual prediction. *Cognitive Science*, *31*(4), 613–643. https://doi.org/10.1080/15326900701399913
Schwartz, G., Howard, M. W., Jing, B., & Kahana, M. J. (2005). Shadows of the past: Temporal retrieval effects in recognition memory. *Psychological Science*, *16*(11), 898–904. https://doi.org/10.1111/j.1467-9280.2005.01634.x
Sederberg, P. B., Howard, M. W., & Kahana, M. J. (2008). A context-based theory of recency and contiguity in free recall. *Psychological Review*, *115*(4), 893–912. https://doi.org/10.1037/a0013396
Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*(4820), 1317–1323. https://doi.org/10.1126/science.3629243
Shiffrin, R. M., & Cook, J. R. (1978). Short-term forgetting of item and order information. *Journal of Verbal Learning and Verbal Behavior*, *17*(2), 189–218. https://doi.org/10.1016/S0022-5371(78)90146-9
Solway, A., Murdock, B. B., & Kahana, M. J. (2012). Positional and temporal clustering in serial order memory. *Memory & Cognition*, *40*(2), 177–190. https://doi.org/10.3758/s13421-011-0142-8
Talmi, D., Lohnas, L. J., & Daw, N. D. (2019). A retrieved context model of the emotional modulation of memory. *Psychological Review*, *126*(4), 455–485. https://doi.org/10.1037/rev0000132
Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, *55*(4), 189–208. https://doi.org/10.1037/h0061626
Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*(5), 352–373. https://doi.org/10.1037/h0020071
Ward, G., Tan, L., & Grenfell-Essam, R. (2010). Examining the relationship between free recall and immediate serial recall: The effects of list length and output order. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(5), 1207–1241. https://doi.org/10.1037/a0020122
Yang, J., Zhao, P., Zhu, Z., Mecklinger, A., Fang, Z., & Li, H. (2013). Memory asymmetry of forward and backward associations in recognition tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(1), 253–269. https://doi.org/10.1037/a0028875
Young, R. K. (1961). The stimulus in serial verbal learning. *The American Journal of Psychology*, *74*(4), 517–528. https://doi.org/10.2307/1419662

Appendix

Model Fitting

We fit SEM to the serial recall data in Experiments 1 and 2 of Logan (2021). Experiment 1 presented list of 5, 6, and 7 random letters. The letters were displayed visually and presentation was simultaneous: The letters appeared as a string in the center of the screen. The letters were recalled by typing them on a standard QWERTY keyboard. No corrections were allowed and participants pressed the return key when they finished recalling a list. Experiment 2 presented lists of 6 random letters, half of which contained a repeated item with 0–3 items intervening and half of which contained unique letters. Each participant was presented with 192 lists. There were 24 participants in each experiment.

We fit SEM to the sequence of approximately 1,152 items each participant recalled. We calculated the likelihood for each item recalled by calculating the overlap between the position code of the recalled item and the other position codes on the list (Henson, 1998, Equation 2). Following Henson (1998), we added Gaussian noise to the overlap score and calculated the probability (likelihood) that the Gaussian representing the recalled item was larger than the rest (the maximum). We summed the negative log likelihood across all retrievals and used Matlab's SIMPLEX routine, fminsearch, to find values of S_0, E_0, S, E (see Equations 1 and 2) and the standard deviation of the Gaussian noise distribution that maximized the likelihood.

We fit two versions of SEM to each participant in each experiment. The versions varied in the constraints on the decay parameters S and E. In the $S \neq E$ models, the values of the decay parameters were allowed to vary independently. We wanted to see whether the best fits would be obtained when $S = E$ or they were close. In the $S = E$ models, the values of the decay parameters were constrained to be equal. We wanted to see whether constraining S and E to be equal would produce fits that were equivalent to the $S \neq E$ models.

In each fit, all parameters were constrained to take positive values. We ran 10 fits of each model for each participant using random starting values that ranged from 5–15 for S_0, 0–10 for E_0, and .5–1.0 for S and E. For the $S = E$ models, the starting range for single decay parameter ranged from .5–1.0. We chose the best fit of the 10 for each model for each participant. The majority of the 10 fits for each model and participant converged on the same best-fitting values. The mean parameters and BIC scores across participants are presented in Table 1 in the main text.

Received March 22, 2021
Revision received July 8, 2021
Accepted August 9, 2021 ■