

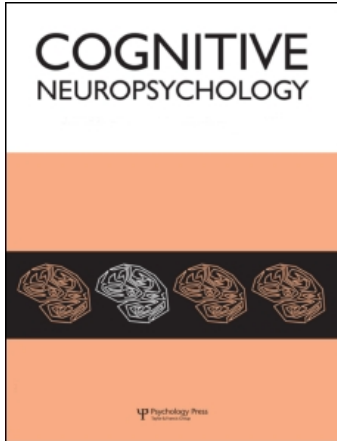
This article was downloaded by: [Harvard College]

On: 17 February 2011

Access details: Access Details: [subscription number 917861759]

Publisher Psychology Press

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Cognitive Neuropsychology

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title~content=t713659042>

Analysis and interpretation of serial position data

Andrew Olson^a; Cristina Romani^b; Alfonso Caramazza^{cd}

^a University of Birmingham, Birmingham, UK ^b University of Aston, Birmingham, UK ^c Cognitive Neuropsychology Laboratory, Harvard University, Cambridge, MA, USA ^d Center for Mind/Brain Sciences, University of Trento, Trento, Italy

First published on: 23 August 2010

To cite this Article Olson, Andrew , Romani, Cristina and Caramazza, Alfonso(2010) 'Analysis and interpretation of serial position data', Cognitive Neuropsychology, 27: 2, 134 – 151, First published on: 23 August 2010 (iFirst)

To link to this Article: DOI: 10.1080/02643294.2010.504580

URL: <http://dx.doi.org/10.1080/02643294.2010.504580>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Analysis and interpretation of serial position data

Andrew Olson

University of Birmingham, Birmingham, UK

Cristina Romani

University of Aston, Birmingham, UK

Alfonso Caramazza

Cognitive Neuropsychology Laboratory, Harvard University, Cambridge, MA, USA, and Center for Mind/Brain Sciences, University of Trento, Trento, Italy

The representation of serial position in sequences is an important topic in a variety of cognitive areas including the domains of language, memory, and motor control. In the neuropsychological literature, serial position data have often been normalized across different lengths, and an improved procedure for this has recently been reported by Machttynger and Shallice (2009). Effects of length and a U-shaped normalized serial position curve have been criteria for identifying working memory deficits. We present simulations and analyses to illustrate some of the issues that arise when relating serial position data to specific theories. We show that critical distinctions are often difficult to make based on normalized data. We suggest that curves for different lengths are best presented in their raw form and that binomial regression can be used to answer specific questions about the effects of length, position, and linear or nonlinear shape that are critical to making theoretical distinctions.

Keywords: Serial position; Short-term memory; Graphemic buffer; Dysgraphia; Spelling.

The organization of serial behaviour has been a topic of interest to psychologists since at least Lashley's seminal paper (Lashley, 1951). Organizing behaviour in time is important in many different domains, including speech (e.g., Acheson & MacDonald, 2009; Gupta, Lipinski, Abbs, & Lin, 2005; Page, Madge, Cumming, & Norris, 2007), spelling (e.g., Caramazza, Miceli, Villa, & Romani, 1987; Glasspool & Houghton, 2005; Goldberg & Rapp, 2008; Wing & Baddeley, 1980), short-term memory (e.g., Atkinson &

Shiffrin, 1971; Brown, Preece, & Hulme, 2000; Burgess & Hitch, 1999; Henson, 1998b; Lewandowsky, 1999; Lewandowsky & Murdock, 1989; Murdock, 1968; Page & Norris, 1998; and many others), perception (e.g., Mason, 1982; Tydgate & Grainger, 2009), motor control (Agam, Bullock, & Sekuler, 2005), and executive function (Schneider & Logan, 2005). Systematic changes in accuracy across position is an important kind of data reported from empirical studies in these areas, and computational models of the same data typically

Correspondence should be addressed to Andrew Olson, Department of Psychology, University of Birmingham, Birmingham B15 2TT, UK (E-mail: a.l.o.olson@bham.ac.uk).

fit serial position curves in the process of showing that they give an adequate account of the empirical results.

We are interested, here, in how data from serial position curves are summarized and analysed. Summary measures are critical when confronting theories and data in order to focus on critical differences in complex results. Summary measures inevitably trade off simplicity and loss of information. What is critical is that the summary measure preserves the information that is necessary for confronting theories. Our starting point is the method for summarizing data from serial position curves that were originally reported by Wing and Baddeley (1980) in their study of handwritten spelling errors. This type of analysis has been widely applied in studies of errors made by aphasic patients in spelling and in speech (e.g., Buchwald & Rapp, 2004, 2006; Caramazza, Papagno, & Ruml, 2000; Cipolotti, Bird, Glasspool, & Shallice, 2004; Cotelli, Abutalebi, Zorzi, & Cappa, 2003; Croisile & Hibert, 1998; Gagnon & Schwartz, 1997; Kay & Hanley, 1991; Neils, Roeltgen, & Greer, 1995; Papagno & Girelli, 2005; Schwartz, Wilshire, Gagnon, & Polansky, 2004; Ward & Romani, 1998). Recently, Machtynger and Shallice (2009) showed that there are some systematic distortions of the serial position curve that the Wing and Baddeley method can introduce (see also accompanying response Wing & Baddeley, 2009), and they suggested an alternative method that we have also used in a study of spelling errors made by deaf participants and speech errors made by aphasic patients (Olson, 1995; Olson & Caramazza, 1999).

We discuss dimensions that are important for confronting serial position data—that is, data that report accuracy at each position for items of different lengths—with theories that describe how position information is represented and maintained. Our goal is to illustrate some of the complexity involved in relating theoretical dimensions like *capacity* and *interference* to differences in length and serial position in empirical data. We start by defining dimensions that distinguish different theories of serial behaviour. We show that these dimensions do create differences in serial position data, but that the relationships

cannot be read directly from the raw data, and they would often be lost through normalization. Finally, we illustrate some alternative analyses that can be used to relate serial position data to theories.

Serial position data have been examined most closely for tasks thought to involve a working memory component. These include, in particular, studies of serial learning or recall (e.g., Healy, 1974; Henson, 1999; Nairne, 1991; Robinson & Brown, 1926; and many others), but also, in the neuropsychological literature, studies of the graphemic and phonological buffers (Caramazza, Miceli, & Villa, 1986; Caramazza et al., 1987; Schiller, Greenhall, Shelton, & Caramazza, 2001; Shallice, Rumiati, & Zadini, 2000; Ward & Romani, 1998). Theoretically, a working memory system should have capacity limitations. Empirically, it has repeatedly been found that initial and final positions are recalled better than medial positions in short-term memory tasks. These two observations have been combined, in the neuropsychological literature, to produce criteria considered diagnostic (among others) for an output buffer: There should be effects of length and a U-shaped function of accuracy across position (Caramazza et al., 1987).

The association between these effects and the characteristics of working memory is perfectly reasonable, but their reification as diagnostic criteria also poses certain problems: The connection between these effects and mechanisms of working memory is neither completely diagnostic nor simple. Nonetheless, these two dimensions remain fundamental to our discussion. We are interested in the source of effects of length because of the connection between length effects and capacity limits, and we are interested in differences in accuracy for different serial positions because of the connection between serial position effects and either interference or short- and long-term contributions to serial production.

Critical dimensions—length effects

Working memory should have capacity limitations. Capacity, in the everyday sense, is an absolute limit.

Items within the capacity of the system can be processed, but anything above the capacity limit will fail. However, in short-term memory experiments, it is the items from the middle of the list, not the most recent items, that are hard to remember. An absolute limit could still create this serial position function if initial items are well retained, and the most recent items are added to the end of a memory buffer, overwriting items from the middle of the list as they are added (Phillips, Shiffrin, & Atkinson, 1967).

A capacity limit, however, could be manifested in more than one way. It could also involve a reduction in processing efficiency for all items. Under this definition, addition of any items above some limit (which can be as low as a single item) makes all items in the buffer more difficult to process. An important prediction, which is common to both of these capacity limits, is that serial position curves for longer stimuli should be vertically displaced from the curves for shorter stimuli. Shorter stimuli should have an advantage across all (or nearly all) positions. This definition of capacity allows us to distinguish effects of capacity from effects of position in our analyses below.

We have said that length effects should be produced by capacity limits, but the reverse is not exclusively true. In other words, a significant effect of length in an analysis does not unambiguously indicate a capacity limit. This is because not only capacity limitations produce length effects. In fact, a constant probability of error at each position will produce length effects in the number of whole sequences correct (item = letter, sequence = whole words, in the buffer context, or item = word, sequence = whole lists, in the short-term memory context). The $p(\text{sequence correct}) = p(\text{item correct})^{\text{sequence length}}$, so that with $p(\text{item correct}) = .9$, $p(\text{sequence correct}) = .73, .66, .59, .53, .48$ for lengths 3–7. Thus, the number of sequences correct is not particularly diagnostic of capacity limitations. For a decline with length to indicate a capacity limit, the decline must exceed the decline predicted by a constant probability of error. We have previously called this a *super length effect* (Romani, Olson, Ward, & Ercolani, 2002).

A more sensitive measure of a length effect is to count the probability of error at each position (or the probability that items are preserved; see Olson, Romani, & Halloran, 2007). Even if we consider the probability that items are correct at each position, however, pure effects of position, which we would not associate with a capacity limitation, can produce length effects. By pure effects of position, we mean that the probability that an item is correct— $p(\text{item correct})$ —changes with position, but that the probability correct for any given position is not different for sequences of different lengths (see Figure 1a). Pure effects of position produce length effects when the probability correct declines with position because longer sequences have later positions where the probability correct continues to go down, giving a lower average $p(\text{correct})$ for the whole string. Pure effects of length, instead, will be found when the probability of error does not change with position (so there are no position effects), but longer sequences have higher rates of error at all positions (see Figure 1b). These are particularly diagnostic of capacity limitations, and we present an example of how a limitation in capacity in an implemented model produces exactly this kind of pure effect of length below.

Critical dimensions—nonlinear serial position effects

Better performance at the beginning and end of a word or list has been repeatedly observed in serial recall and also in patients with hypothesized phonological or graphemic buffer impairments (Caramazza et al., 1987; Healy, 1974; Murdock, 1968; Shallice et al., 2000). In the short-term memory literature the advantage for early items (the primacy effect) has been thought to occur because early items can be rehearsed often enough to enter long-term memory, where they are protected from decay. The advantage for recent items (the recency effect), instead, occurs because information in a short-term memory store decays over time (Atkinson & Shiffrin, 1971). Items in the middle of the list suffer more from decay than do the final items, but have not

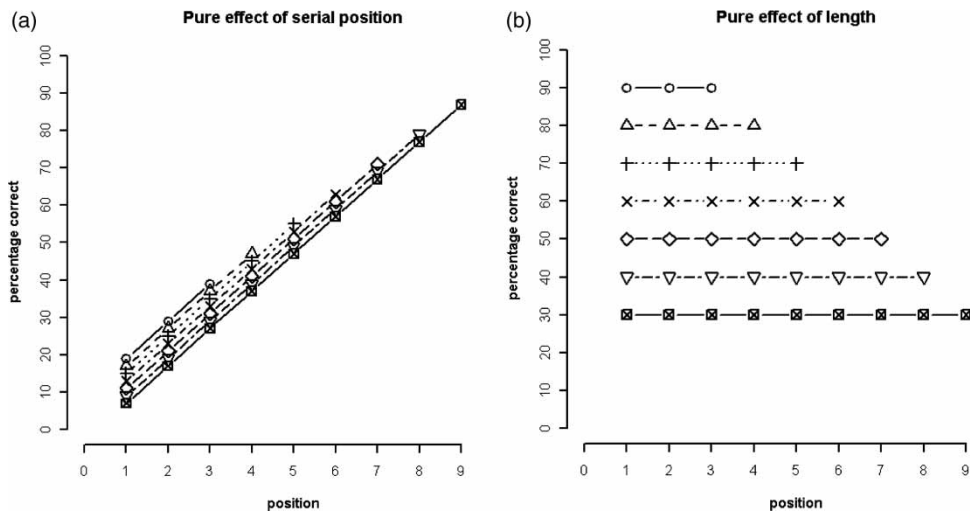


Figure 1. Serial position curves for pure effects of serial position (a) and pure effects of length (b). Curves for pure effects of serial position have been jittered from the position of complete overlap so curves for different lengths are visible.

been rehearsed often enough to enter long-term memory, producing a U-shaped function with position. This idea can be seen in a more recent form in the association between primacy and semantic abilities and recency and phonological abilities in aphasic patients (Martin & Saffran, 1997; see also Davelaar, Goshen-Gottstein, Ashkenazi, Haarmann, & Usher, 2005; Martin, Shelton, & Yaffee, 1994).

The influence of two gradients working in opposite directions appears, subsequently, in models where the gradients provide a two-dimensional code, rather than being associated with two different memory systems (Henson, 1998b; Houghton, 1990). In these models, the U-shaped function is partly the result of the two-dimensional code becoming less distinct for positions in the middle of the list, and partly it results from end effects (see discussion of end effects below).

U-shaped serial position curves have recently been explicitly attributed to *interference* rather than separate memory components (e.g., Lewandowsky, Oberauer, & Brown, 2008; Nairne, 2002), and in the Henson and Houghton models, in fact, it is interference that produces more errors in the middle of the list. By *interference*, we mean that

adjacent items impose a cost on each other. In the following discussion, interference usually occurs because the representation of position for near items is more similar than it is for far items. There is some probability that an item in position $X + 1$ or $X - 1$ will be retrieved instead of the item in position X because of this overlap in positional codes. Clearly, the strongest effects of interference should involve adjacent items, on this account, but interference could also involve items that are further away from the target position (e.g., $X \pm 2$). Errors in serial recall, in fact, show an effect of nearby positions very strongly (Henson, Norris, Page, & Baddeley, 1996).

It is important to note that first and last positions do not suffer from as much interference as internal positions because there are items on only one side of these positions. These end effects create the U-shaped function. The extent of the end effects gives a measure of the distance over which items interact. If only adjacent items interfere with each other, there will only be an advantage for the first and last items of a list (Figure 2). This creates a relatively shallow U-shape (Figure 2a). The U deepens as items interact over larger distances (Figure 2b; for details of the calculations used to create the curves, see the Appendix).

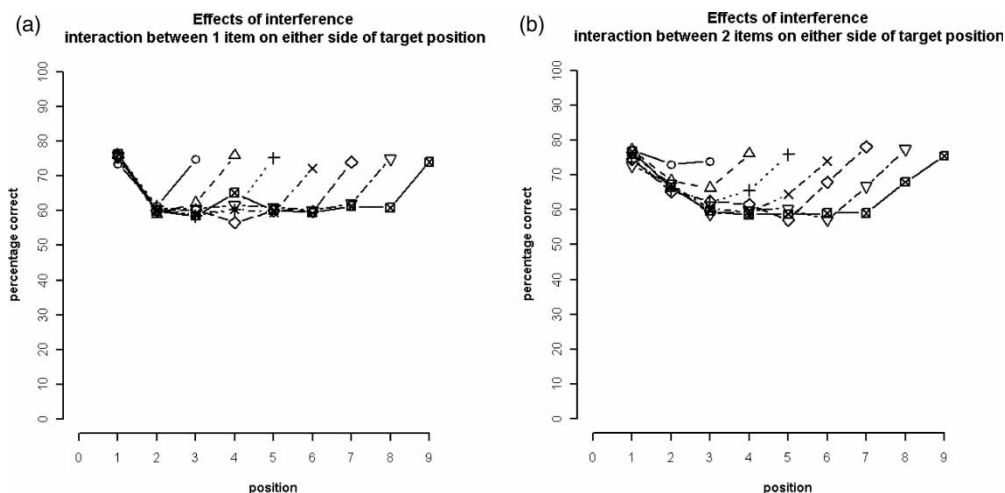


Figure 2. U-shaped serial position curves generated by interference. In (a), the probability of an item being reported in the correct position, X, is .4, and the probability of an item from position X - 1 or X + 1 being reported instead of X is .3. In (b), the probability of an item being reported in the correct position X, is .4. The probability of X + 1, X - 1, or X + 2, X - 2 being reported instead of X is .2 and .1, respectively.

As was the case for length, measuring the shape of the serial position curve in practice is more complicated than specifying the shape in theory. The shape that the curve assumes in empirical data is influenced by several factors, including any bias against producing items more than once (e.g., after an anticipation error), the tendency for anticipations to precipitate reciprocal perseverations, and the method used to score errors. We illustrate these factors below.

The dynamics of production that influence what happens after an error has been made can change the shape of the serial position function. Models making use of interference and end effects often assume, as an implementational detail, that once an item has been produced, it is inhibited and does not have the possibility to be produced again. If the Position 3 item is produced early in Position 2 it cannot be produced again in Position 3 (Brown et al., 2000; Burgess & Hitch, 1999; Farrell & Lewandowsky, 2002; Henson, 1998b; Page & Norris, 1998). This prohibition against repeating responses is justified based on data that show that participants in serial recall tasks are reluctant to repeat items, even when this is necessary for correct recall (Henson, 1998a).

The prohibition influences the form of the serial position curve because an item produced too early necessarily creates an error in two positions—the position where the item was produced early and the position where it should have been produced, but now can no longer occur. This effect of errors can accumulate. If Item 3 is produced too early, Items 2 and 4, both errors, may be the only competitors for Position 3. If Item 4 is produced in Position 3, this creates another error at Position 4. The important consequence for the shape of the serial position function is that it has a clear primacy gradient but the recency portion of the curve is reduced (Figure 3a), or eliminated (Figure 3b). When nonfinal responses produce the final position too early, this, by necessity, also creates an error in the final position. The amount by which the recency effect is reduced depends on how probable it is that Item 4 is produced, as opposed to Item 2, once Item 3 has been produced too early.

The recency effect reappears if, after an anticipation error, there is a high probability that a reciprocal perseveration error will create a swap. When this is the case, the final position is less likely to get produced as part of a set of related errors (i.e., errors

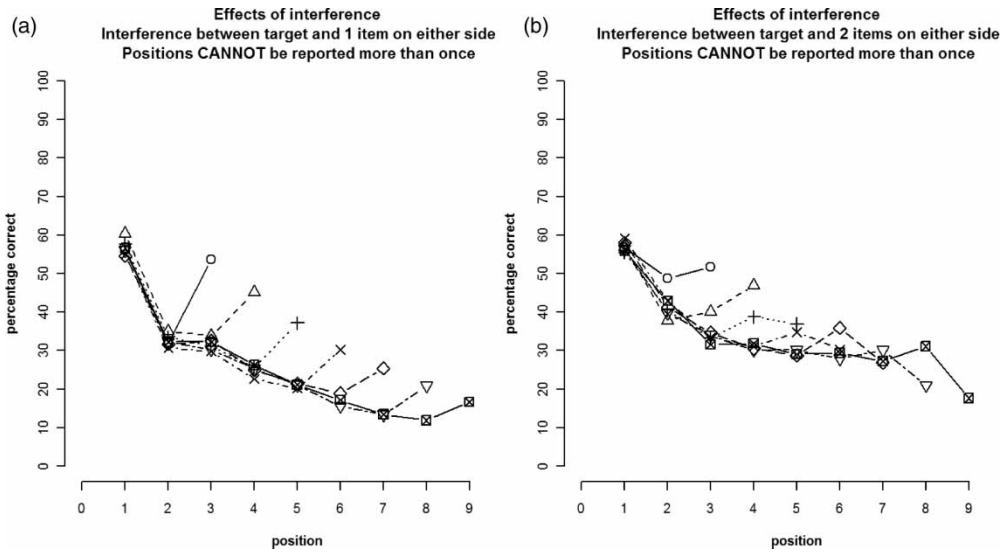


Figure 3. Serial position curves identical to those in Figure 2 except that once an item has been reported, it is suppressed and cannot be reported twice in the same response (e.g., if the sequence 12345 starts with the error 13, 3 cannot also be produced again in the correct position, making the error 13345 impossible).

stay local and do not accumulate), and so the final position shows a recency effect that mirrors the primacy effect. Swaps are common if there is a primacy gradient that makes earlier items stronger than later ones (as in Henson, 1998b; Houghton, 1990; Page & Norris, 1998; see Figure 4).

Clearly, the specific dynamics of production that generate and then follow errors are important to the shape of the serial position curve. These are important aspects of the production system itself. A factor external to the model that also affects the shape that the serial position curve assumes is the method used to score errors. In the discussion above, we categorized responses as errors according to what the model knows. If a 5-item list is produced as 1345X (where “X” is a “no response” error and occurs when the only unused item, number 2, is too weakly activated to be produced in Position 5), Items 3, 4 and 5 have all been produced too soon, and, according to what the model knows, these responses should all be counted as errors. From a point of view external to the model, however, it will appear simply that Item 2 has been deleted, and Items 3, 4, and 5 have been produced correctly.

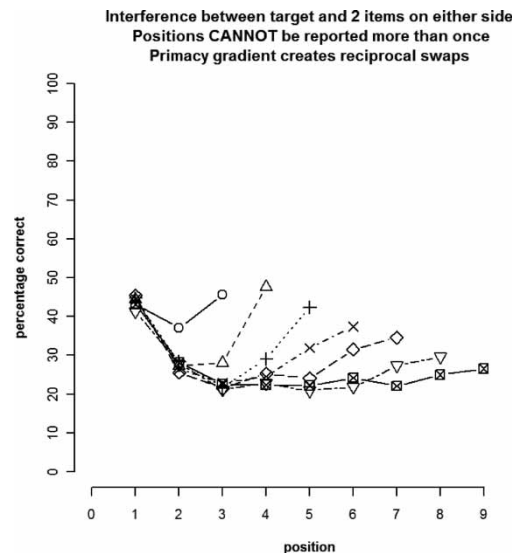


Figure 4. Recency effects created by a primacy gradient that encourages swaps. The probability of report across the positions $X - 2, X - 1, X$ (target position), $X + 1, X + 2$ was .3, .25, .2, .15, .1.

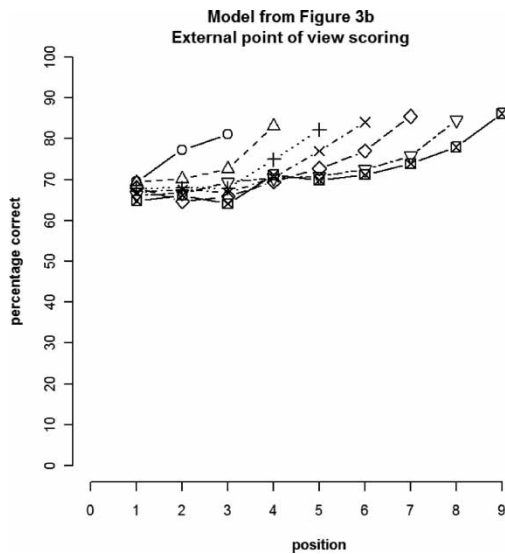


Figure 5. *The serial position curve that results when sequences from the model that produced Figure 3b are scored from the point of view of a naïve observer (rather than from a model-internal point of view). Scoring uses the longest increasing subsequence.*

A natural method of scoring assumes that the smallest number of changes possible created the error or that the largest possible number of items are in the correct position. Finding the longest increasing subsequence in a sequence of numbers implements this scoring procedure (see limits to this method in Tichy, 1984). If we rescore the sequences from Figure 3b from the point of view of a naïve observer using an algorithm for the longest increasing subsequence (Gusfield, 1997), there is no primacy portion of the curve and a strong recency portion (Figure 5). The reason for stronger recency effects is clear from the example. Items that the model produces too early will sometimes be counted as correct by a longest increasing substring algorithm. If we use a stricter criterion, however, and count items as correct only when they are in the correct numerical slot, the primacy

and recency portions of the curve will have the shape we plotted in Figure 3b. Which criterion is actually “correct” is not possible to determine from outside the model. If the omission of Position 2 was a true deletion, and the other items were produced in their correct positions, then the first scoring procedure reflects the actual set of errors. If Positions 3, 4, and 5 were all produced early, and then Position 2 could not be produced, then the second procedure reflects the actual set of errors. What we have shown is that any one scoring procedure does not necessarily produce the set of transformations that actually turned the target into the response and that the shape of the serial position curve depends, in part, on the scoring procedure.¹ For the purposes of theory testing, what is important is that simulation model data and empirical data are scored using the same procedure.

The effects we have outlined in this section are especially important when a substantial portion of errors are exchanges (as in serial recall; e.g., Henson, 1998b) and, to some extent, when items do not often appear more than once in a response. Understanding how the shape and position of serial position curves can be created by effects of capacity, edge effects, interference, and scoring is important, however, because factoring these effects is necessary to relate serial position curves to theories and because, as Farrell and Lewandowsky (2002) note, the factors we have described here can often be responsible for the shape of serial position curves rather than mechanisms that are more prominent in the models themselves (e.g., oscillations in the case of OSCAR; Brown et al., 2000).

In this section, we have seen that theoretically important factors like capacity limits, the presence and spread of interference, and the suppression of previous responses influence the position and shape of serial position curves. In the following

¹ This is a concrete example of a situation long recognized in the computer science literature devoted to matching text patterns (see algorithms for Levenshtein or edit distance, e.g., Gusfield, 1997). Reconstruction of the changes that produce a response from a target cannot be done with certainty. Since an infinite number of transformations are possible, any one can only be assigned a value that indicates its likelihood, and scoring errors is an optimization problem that involves picking the changes that are *most likely* to have occurred given the target and response.

section we examine what happens to critical information from serial responses when data are normalized. Initially, we do this using theoretical examples. We show, however, that our theoretical concerns also apply to implemented models and/or existing data. Finally, we discuss alternatives for analysing serial position data given the complexities we have outlined.

Normalized data and critical dimensions

Critical dimensions we have identified in the discussion above include the effect of length independent of position, and the form of nonlinear position effects determined by primacy and recency. Are these critical dimensions represented in normalized data?

Clearly, effects of length and position cannot be distinguished using normalized data. Normalization collapses data from different lengths and positions onto a single curve. The vertical displacement between curves that is critical for measuring length effects is eliminated.

Normalized data are more successful in characterizing the shape of the serial position curve over position, but here, too, there are factors to be aware of. Determining whether or not recency

effects are present and exactly how many items are advantaged cannot be guaranteed based on normalized data. Substantial recency effects that increase with length but are restricted to a single item produce a serial position curve (Figure 6a) that is very similar to the serial position curve produced when a decline in performance with position slows at later positions but there are no recency effects (Figure 6b). The similarity of the curves in Figures 6a and b are the consequence of compressing the number of positions in the longest sequences into fewer normalized positions, as has typically been the case in analyses that follow Wing and Baddeley's (1980) approach. In the case we illustrate here, up to nine positions were collapsed to five normalized positions.

Differentiating single-item recency effects from effects that extend over more items is also difficult using normalized data. Figure 7 illustrates normalized curves for recency effects that involve one and two items. The extent of the recency effect in the normalized curves does depend on the extent of the recency effect in the data, but it also depends on the number of positions that the data are normalized to. Normalizing to more positions extends the recency effect for the same unstandardized data.

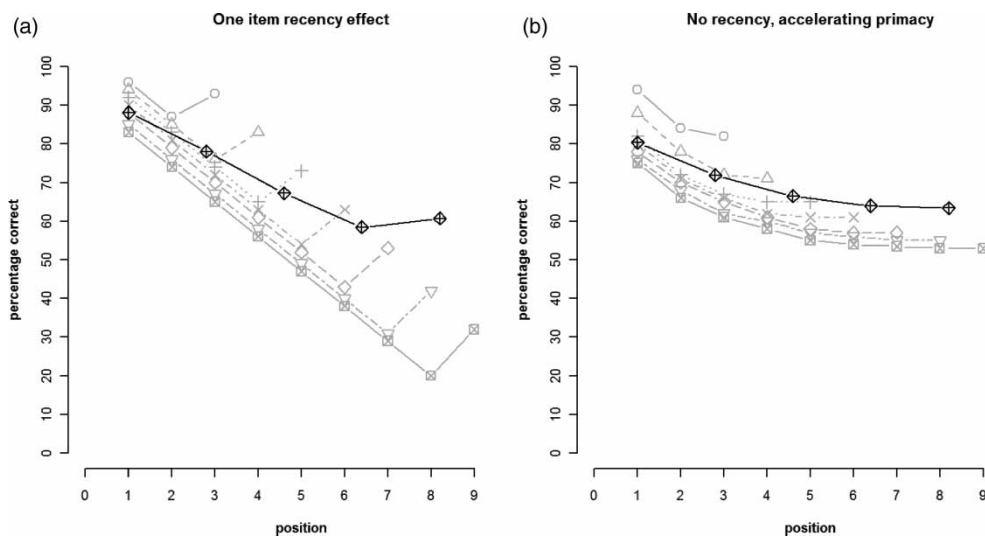


Figure 6. Normalized serial position curves (in black) for (a) one-item recency that increases with position and (b) no recency, but accelerating primacy effect. Unstandardized serial position curves are in grey.

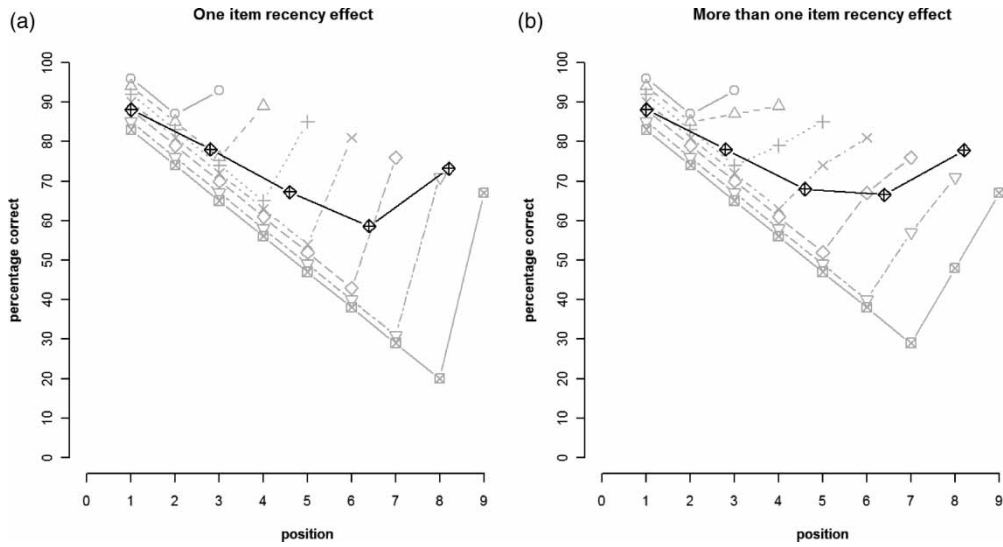


Figure 7. Normalized serial position curves (in black) for (a) one-item recency effects and (b) two-item recency effects.

Critical dimensions in implemented models and empirical data

These illustrations using hypothetical data call into question the utility of normalizing data when distinguishing critical factors in ordered production tasks, but are these factors important in actual models of ordered production? Below we present several examples using implemented models and/or empirical data to show that the dimensions that we have identified as critical really do show variation of the kind we have described above, and we suggest that using normalized data may not be the best way of confronting theories and data.

Capacity

Our first illustration involves two different ways of coding position that have both been used in the literature (Glasspool, Shallice, & Cipolotti, 2006; Page & Norris, 1998). One model shows clear capacity limits and the other clear position effects. We show that the capacity-limited model produces vertically displaced serial position curves, and the model with clear position effects but no capacity limit (in the range we explore) produces serial position effects without vertically

displaced curves for different lengths. Both models will produce effects of length and U-shaped serial position curves when the data are standardized, so the presence or absence of these effects does not distinguish them. Characteristics of their unstandardized serial position curves do, however, allow the important aspects of these models to be distinguished.

Our first model codes serial position with a series of Gaussian curves that are spread over a limited number of coding units (in our case 100). This approach is similar to the serial position units that are used by Glasspool et al. (2006) in their model of the graphemic buffer. The position units are part of a system that accomplishes letter production (in spelling) by using an associative memory to produce individual letters in a word in the proper order based on a whole word input and a changing set of position codes. In the short-term memory literature, Henson has called these models “positional theories,” and they have a variety of implementations (Brown et al., 2000; Burgess & Hitch, 1999; Conrad, 1965; Lee & Estes, 1977). In our particular implementation, when a smaller number of positions needs to be encoded, the Gaussian curves are broader, and they become increasingly narrow as more positions

Downloaded By: [Harvard College] At: 16:15 17 February 2011

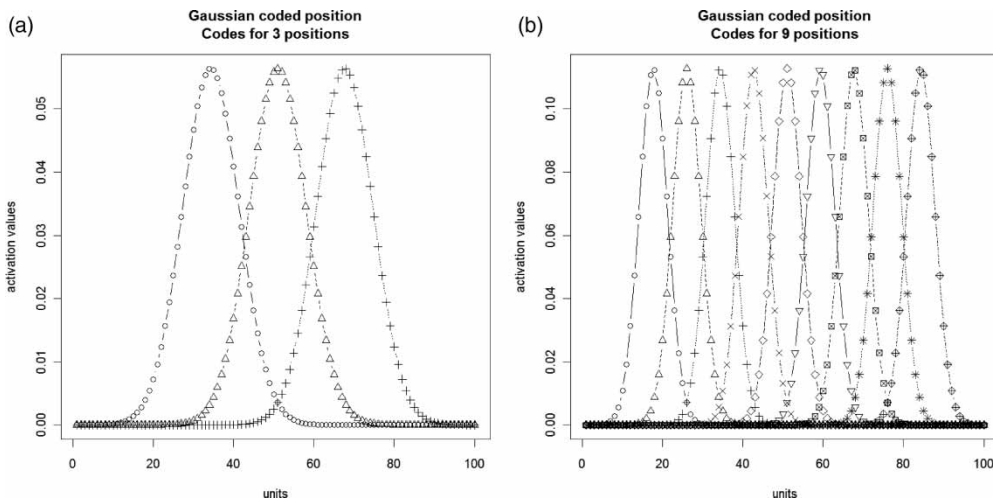


Figure 8. Gaussian position codes using 100 units to code (a) 3 positions and (b) 9 positions. Position codes are narrower when coding more positions over the same number of units.

need to be distinguished (Figure 8; this method was chosen to maximize the stability and redundancy of the codes for each length, but other coding schemes, e.g., Gaussian codes with a single width, produce the same critical outcomes). We assume that errors are made when noise shifts the position codes on the encoding units (i.e., noise shifts the Gaussian peaks left or right along the set of encoding units). The noisy position code produced by the model is compared with the noise-free codes for each position. This simulates the effect of passing a noisy position code to the associative memory we described above (as in Glasspool et al., 2006). The position that produces the largest dot product of noisy and noise-free codes is assumed to be the position reported by the associative memory.

The width of the noise distribution does not change as different length items are encoded by the model. For this reason, noise more strongly affects the positional codes for longer items, which are more crowded together than shorter items. In addition, the code for the initial position can only be confused with items to the right (and conversely, the final position with items to the left), making initial and final positions less prone to transposition errors. The data reported here assume that positions are not inhibited once

produced (i.e., they can be produced again), and they are scored according to what the model knows, since, for the moment, we are interested in how the model functions without other limitations and not how it compares to empirical data.

Results for 1,000 simulated trials at each length are shown in Figure 9. The noise distribution for

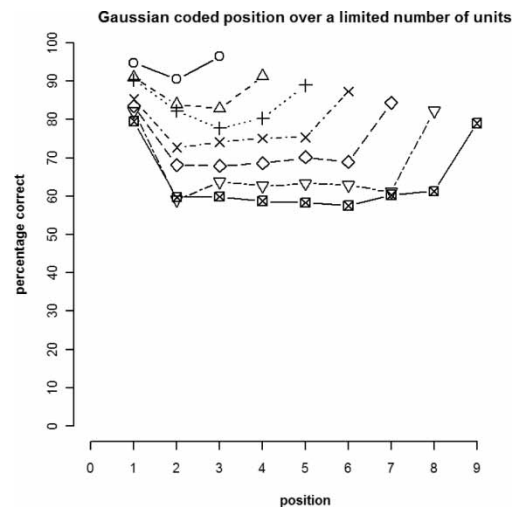


Figure 9. Results from 1,000 trials at each length coding position with noisy Gaussian position codes. Noise shifted the position of Gaussian distributions right or left. Noise had a mean of 0 and a standard deviation of 5 units.

these trials has a mean of 0 and a standard deviation of 5 units. There are evident end effects for the first and last positions, creating U-shaped position curves. Curves for shorter items are above those for longer items. This pure effect of length results from filling the capacity of the memory system. As positional codes for longer lengths become more crowded, they are more easily confused at every position. Aside from the end effects, there are no effects of serial position. Interior positions are equally susceptible to error.

A contrasting model that produces effects of serial position but not length is a simplified version of Page and Norris's (1998) Primacy Model. This model assumes there is a primacy gradient that orders items. Early items in a sequence are more highly activated than later items. At each point when an item needs to be produced, the model chooses the item that is most strongly active and then suppresses it so that it cannot be reactivated (following Page and Norris's assumptions). In addition, the overall level of activation slowly decays over time, so that the constant decrease of activation between items gets smaller as time goes on. Noise is added to the activation values to simulate transposition errors.

Figure 10 shows the results of 1,000 simulated trials at each length. Activation values started at 1 and decreased by 0.1 for each position. All activations decayed by $e^{-0.2}$ as each item was produced. Added noise had a mean of 0 and a standard deviation of 0.05. In this range, where production is relatively accurate for early positions, there is little or no indication of capacity effects, but accuracy decreases with position in accordance with the primacy gradient. The lack of capacity effects is not surprising. As long as the difference between items along the gradient remains relatively robust, capacity is not a limiting factor. If activation decay were stronger, bringing all items closer together, or if noise were greater, capacity would become a more evident factor, as would be appropriate for intuitions about how the model operates (i.e., it is not capacity limited until codes become increasingly confusable). In the data we present here, there is a recency effect limited to a single item, which occurs because transpositions tend to

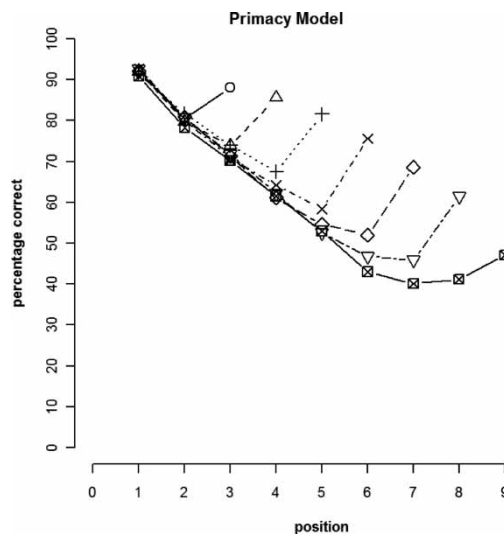


Figure 10. Results from 1,000 trials at each length from a simplified primacy model (based on Page & Norris, 1998).

be reciprocal as a result of the primacy gradient (see discussion above), and the last item can only exchange with the item to its left.

These two models, based on existing theories, illustrate that different ways of representing position can produce essentially pure effects of position and pure effects of length and that these differences in the models can be distinguished in the raw serial position curves. The important differences would be obscured, however, if we normalized the data prior to analysis. We now turn to problems related to measuring the presence of absence of nonlinearity (the U shape) in serial position data.

Nonlinearity

The presence of the U-shape in the serial position curve has been considered diagnostic of a response buffer in the neuropsychological literature and is a very common feature of short-term memory data (Murdock, 1968). There are also, however, neuropsychological patients who make errors involving single segments but do not show the improvement in performance for final items (Glasspool et al., 2006; Romani, Galluzzi, &

Olson, in press; Schiller et al., 2001; Ward & Romani, 1998). Ward and Romani attributed this pattern to a separate locus, involving weaker activation of temporary representations from the lexical level. They argued for this source based on stronger effects of frequency and imageability and based on a substantial number of lexical substitution and semantic errors. Glasspool et al. (2006) called this pattern a "Type B" graphemic buffer disorder, but, like Ward and Romani (1998), they attributed the pattern to degraded input to the buffer level.

Based on the similarity between the U-shaped function for slips of the pen (Wing & Baddeley, 1980) and the error function of patients, Schiller et al. (2001) argued that the U-shaped function is the result of noisy input to the buffer that exacerbates the normal pattern, and the linear decline represents damage to the buffer itself. Despite some differences in interpretation, what all of these authors agree on is that some patients show improvement with the final units of a sequence (the U-shaped serial position function), and others do not, and that this difference is theoretically important.

We have already noted that normalized serial position curves for data with substantial advantages for final items and curves for data without final-item advantages can be very similar. Several factors influence the presence and strength of non-linearity in the normalized curve.

One factor is the consistency of shape for short and long curves. If they do not share the same shape, the normalized data will be a mixture of the long and short curves. As we have seen above, if advantages for final items only emerge strongly in longer stimuli, and the normalized serial position curve is based on fewer positions than there are in the long stimuli, recency effects for longer items can be masked by shorter items (the converse would also hold, if the U-shape were present only for shorter items). This may be a particular worry for stimulus sets where shorter items outnumber longer items, as could naturally arise in data sampling different word lengths, but where word length is not specifically controlled. Another way in which different shapes could

emerge in word-based tasks is as a result of structural factors. If, for example, a U-shaped function operated over syllables rather than words, or if vowels and consonants had systematically different error rates, curves for long and short words would be expected to differ. If consonants were selectively preserved in responses, curves for short words, which in English are likely to have consonant initial and final portions, would produce a single U, while curves for longer words, which would be likely to have a consonant in the middle of the item, would produce a double U shape. The normalized data would have a single bowed shape, masking the informative heterogeneity of the underlying curves.

A second factor is the number of positions that data are normalized to. U-shaped data that are normalized to more positions will be more clearly nonlinear than data normalized to fewer positions. This is a simple consequence of the number of points available to describe the curve. The issue is important when nonlinearity is statistically tested based on normalized data. One way of testing the degree of curvature would be to fit an equation that has linear and quadratic components to the normalized curve and to use the significance of the quadratic component as a test of nonlinearity. Table 1 shows the significance of the quadratic component when the data displayed in Figure 6a are normalized to different lengths. As the number of points increases, the significance of the quadratic term increases.

The number of points that curves are normalized to is also important if the extent of the

Table 1. Significance of the quadratic component when the data from Figure 6a are normalized to different lengths

Number of normalized positions	<i>t</i> value for quadratic term	<i>p</i>
5	2.63	.12
6	2.87	.06
7	3.24	.03
8	3.56	.02
9	3.90	.008

Note: Larger numbers of normalized points result in a clearer quadratic component.

initial or final advantage is theoretically important. For example, the degree of interference determines the extent of the initial and final advantage when interference is the primary factor that produces a U-shape. Interference confined to a single item on either side of the intended position produces primacy/recency effects for only a single initial or final item. Interference that extends over more items allows the initial/final advantage to extend further from the ends of the curve (in a modest way). Interference, by itself, can only produce symmetrical end effects. Other sources of a U-shaped function, a primacy gradient (Page & Norris, 1998), opposing gradients (Henson, 1998b; Houghton, 1990), or structurally different memory systems (Atkinson & Shiffrin, 1971), can predict asymmetric and/or more extensive advantages for initial and/or final items. When data are normalized, however, the number of points that are advantaged at the beginning or end of a curve depends as much on the number of normalized positions as it does on the number of positions that are advantaged in the raw data. When the extent of initial or final effects is theoretically important, this is best measured using unstandardized data.

In this section we have seen that theoretically important aspects of serial position curves cannot always be recovered unambiguously from normalized data. The match between theory and data can be more easily judged based on the raw serial position curves for each length.

One major appeal for normalizing serial position is that it simplifies presentation and analysis. If we abandon normalization, can effects of length, position, and linearity be tested in a transparent and reliable, but reasonably straightforward, way? We present some options for analysis in the next section.

Analysis of serial position data

Analysis of raw serial position curves for different lengths can be done without too much difficulty using binomial regression. This is the good news. The bad news is that there is often no simple recipe that relates a particular serial position

shape or location to a theoretical model. We have seen, for example, that the method of scoring can substantially change the shape of serial position curves. What is important is that results from a theoretical model be developed under the same criteria as the actual data when model and data are compared. (We might want to score sequences by stopping at the first error, or by using the longest set of items in the right relative order. We would not want to score model results using model internal criteria, which in our example above would be closest to stopping at the first error, but then use the longest set of items in the right relative order for participant data.)

We illustrate binomial regression methods using the data from the simplified primacy model and the Gaussian coding model that we presented above (Figures 9 and 10). As we have shown, the primacy model produces position effects but not effects of length, and the Gaussian coding model produces length effects but not position effects. Both models have clear end effects.

When preparing data for analysis, an item's ordinal position, the stimulus length, and a binary code that indicates whether or not that position was preserved in the response must be coded. These data are predicted by a binomial regression model that has terms for length and position. Statistically evaluating length and position effects, however, can be complicated by end effects. The advantage for initial and final positions in a three-item sequence is much greater than the advantage in a nine-item sequence because initial and final positions make up two thirds of the data when length = 3 and two ninths of the data when length = 9. This can produce a length effect even if medial positions show no difference with length. For example, the data from the primacy model (Figure 9) produce both length and position effects if end effects are not accounted for (model: $correct = length + position$; length, $z = -3.46$, $p < .001$; position, $z = -15.0$, $p < .001$). One way to test for effects of length and position that are not artefacts of end effects is to use dummy variables that code 1 for initial position and 0 otherwise and 1 for final position and 0 otherwise and to include these terms in the model. This allows initial and

final positions to be fitted independently of medial positions. If we reanalyse data from the primacy and Gaussian coding models using the model $correct = length + position + initial + final$, the primacy model shows clear effects of position but not length (length, $z = 0.371$, $p = .71$; position, $z = -12.7$, $p < .001$), and the Gaussian coding model shows clear effects of length but not position (length, $z = -9.86$, $p < .001$; position, $z = 0.175$, $p = .861$).

The method to choose for evaluating nonlinearity depends on the kind of nonlinearity that needs to be evaluated. If the expectation is that there will be advantages for only initial and/or final positions, dummy coding, as we did above, can be used, and the significance of the terms for initial and final positions can be reported. For example, the effects of initial and final position in the Gaussian coding model are clear using this method (initial, $z = 6.77$, $p < .001$; final, $z = 6.70$, $p < .001$). If the initial and/or final advantage extends over more positions, comparing quadratic and linear models of the data may be more appropriate. If the quadratic term in a model of the general class $correct = length + position + position^2$ is significant, this indicates a reliable nonlinearity.

As is always the case, statistical significance is not a direct indicator of theoretical significance. The magnitude of the nonlinearity is important. A small, but significant, nonlinearity, may be less important than a more substantial nonlinearity that has the same level of significance. In general, it is worth paying attention to the value of the coefficients generated by the model. For example, the coefficients for initial and final positions in the Gaussian coding model are .965 and .952. This shows us that the initial and final advantages are symmetrical, which is theoretically significant.

What do the parameters .965 and .952 mean, however, in terms of percentage error, which is the measure we are interested in? The binomial regression model is based on a logistic function, so the change in probability of error as the parameters change is not constant over the range modelled. This makes coefficients from logistic models more complicated to interpret than the

parameters of linear models. An easy approximation to the rate of change in the probability of error for a unit change in a parameter value, however, is given by $B \times p \times (1 - p)$ where B is the coefficient we are interested in, and p is the average proportion of items correct (Agresti, 2002). Taking the initial parameter from above (.965) and the average proportion correct (about .75) leads us to expect a primacy advantage of about $.965 \times .75 \times .25 = .18$, or 18%, which is a bit of an overestimate (12% is the mean value from the data), but shows that the difference is substantial.

Here, the end effect is something that can be read fairly easily from the data (taking the average primacy advantage over all lengths). Evaluating the size of the end effect may be more critical and less clear when the end effect and some other effect, like a primacy gradient, overlap, as is the case in the primacy model. Binomial regression will help separate the general downward trend that affects all positions from any exaggerated decline that affects only the initial position. In this situation, the formula illustrated above is helpful. Separating end effects and other effects that depend on position is especially informative when the expectation is that end effects will be symmetrical for initial and final positions.

In general, we advocate presenting serial position data in their raw form. When the data have systematic structure they are not difficult to interpret, and when the structure is not systematic this should be a warning about the stability of any conclusions drawn from them. We suggest that specific theoretical questions can be statistically explored using binomial regression. In the neuropsychological context, theoretical development has progressed to the stage where simple classification of patients, for example, as buffer patients or not, can give way to an exploration of the more specific properties of a deficit that produces segmental errors, and analytic tools are available to support this enterprise. Specific mathematical models of short-term memory have been very useful, already, for directing empirical work in that area (Botvinick & Plaut, 2006; Brown et al., 2000;

Burgess & Hitch, 1999; Henson, 1998b; Murdock, 1993; Nairne, 1990; Page & Norris, 1998), and normalized data are less commonly reported for short-term memory studies. Here too, however, the links between the mechanisms responsible for effects and the effects themselves could sometimes be more transparent (as Farrell & Lewandowsky, 2002, note). Of course, with patients it is important to interpret serial position data in the context of the patient's more general pattern of performance (e.g., number of lexical or semantic errors, effects of frequency and imageability, etc.), and our focus here should not distract from that important point.

When specific quantifiable models are to be contrasted, perhaps the best method of approaching the problem is to formalize the models statistically and produce likelihood estimates for the data based on the models. Then, formal model selection procedures (see Burnham & Anderson, 2002) can be used to decide whether any one model gives a clearly superior account of the data. It is important to note that this approach does not produce a binary decision about a "winning" model. Instead the level of support for each model is quantified, which is appropriate, and signals when the data do not clearly distinguish between models. Describing this process, however, is beyond the scope of the present article.

DISCUSSION

Although our starting point has been to examine the effects of normalizing serial position data, our eventual aim has been broader. We have illustrated some of the complexity involved in relating serial position data to underlying theories. There are several important issues that we have highlighted. The differing theoretical roles of capacity and interference or capacity and short- and long-term contributions to memory mean that length and position effects need to be distinguished and evaluated. Capacity limitations produce vertical displacements between curves for different lengths, while position effects create overlapping curves for different lengths, but systematic changes with

position. Since normalizing collapses data from different lengths, length and position effects cannot be distinguished using normalized data. Normalized data often preserve the general shape of serial position curves, but critical information like the extent of the recency portion (if any) will not be preserved in a way that is independent of the parameters used to normalize the data. Likewise, statistical tests based on normalized data are also problematic. Both the shape of the curve and the significance of any nonlinearity depends on both the data and the number of positions used in the normalization procedure. When comparing data from quantitative models and empirical data it is important to match scoring procedures, since different scoring procedures (e.g., model internal vs. model external) can have substantial effects on the shape of the serial position curve. Finally, serial position information should not be used in isolation from the surrounding empirical context. In patient studies, for example, the types of error that patients make and the factors that influence their performance should be considered along with serial position data.

Does normalization ever have a role based on these considerations? Normalized data may be useful when what is needed is a compact summary of the serial position pattern, when the normalized pattern accurately reflects the underlying data and when a detailed match between specific theories and the data is not at issue.

If we need to make judgements that have specific theoretical consequences, like whether a memory buffer is involved in a pattern of errors, we would suggest theoretical and analytic developments allow us to go beyond the resolution that normalized data allow. Graphing serial position curves for different lengths is somewhat more complex than presenting a single normalized curve, but the general shape is usually recognizable, and raw data preserve detail that is theoretically important.

Binomial regression allows specific hypothesis about the influence of length, position, and shape to be tested. Specific quantitative models in this area often make predictions about the gross shape of serial position data that are similar and

will be hard to distinguish at a general level. These models may still, however, be distinguishable based on more detailed comparisons. When the likelihood of data given the models can be quantified, this presents a powerful way to contrast models. We are optimistic that, in the context of a dialogue between quantitative theories and empirical data, data from serial position curves will continue to be informative as they are applied to the questions that Lashley (1951) raised over 50 years ago.

Manuscript received day month year
 Revised manuscript received day month year
 Revised manuscript accepted 18 May 2010
 First published online 23 August 2010

REFERENCES

- Acheson, D. J., & MacDonald, M. C. (2009). Twisting tongues and memories: Explorations of the relationship between language production and verbal working memory. *Journal of Memory and Language*, *60*, 329–350.
- Agam, Y., Bullock, D., & Sekuler, R. (2005). Imitating unfamiliar sequences of connected linear motions. *Journal of Neurophysiology*, *94*, 2832–2843.
- Agresti, A. 2002. *Categorical data analysis*. Hoboken, NJ: John Wiley & Sons.
- Atkinson, R. C., & Shiffrin, R. M. (1971). Control of short-term memory. *Scientific American*, *225*, 82–90.
- Botvinick, M. M., & Plaut, D. C. (2006). Short-term memory for serial order: A recurrent neural network model. *Psychological Review*, *113*, 201–233.
- Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review*, *107*, 127–181.
- Buchwald, A., & Rapp, B. (2004). Rethinking the graphemic buffer? *Brain and Language*, *91*, 100–101.
- Buchwald, A., & Rapp, B. (2006). Consonants and vowels in orthographic representations. *Cognitive Neuropsychology*, *23*, 308–337.
- Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review*, *106*, 551–581.
- Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). New York: Springer.
- Caramazza, A., Miceli, G., & Villa, G. (1986). The role of the (output) phonological buffer in reading, writing, and repetition. *Cognitive Neuropsychology*, *3*, 37–76.
- Caramazza, A., Miceli, G., Villa, G., & Romani, C. (1987). The role of the graphemic buffer in spelling—evidence from a case of acquired dysgraphia. *Cognition*, *26*, 59–85.
- Caramazza, A., Papagno, C., & Rumel, W. (2000). The selective impairment of phonological processing in speech production. *Brain and Language*, *75*, 428–450.
- Cipolotti, L., Bird, C. M., Glasspool, D. W., & Shallice, T. (2004). The impact of deep dysgraphia on graphemic buffer disorders. *Neurocase*, *10*, 405–419.
- Conrad, R. (1965). Order error in immediate recall of sequences. *Journal of Verbal Learning and Verbal Behavior*, *4*, 161–169.
- Cotelli, M., Abutalebi, J., Zorzi, M., & Cappa, S. F. (2003). Vowels in the buffer: A case study of acquired dysgraphia with selective vowel substitutions. *Cognitive Neuropsychology*, *20*, 99–114.
- Croisile, B., & Hibert, O. (1998). Spatial or afferent agraphia without left-sided neglect. *Aphasiology*, *12*, 147–159.
- Davelaar, E. J., Goshen-Gottstein, Y., Ashkenazi, A., Haarmann, H. J., & Usher, M. (2005). The demise of short-term memory revisited: Empirical and computational investigations of recency effects. *Psychological Review*, *112*, 3–42.
- Farrell, S., & Lewandowsky, S. (2002). An endogenous distributed model of ordering in serial recall. *Psychonomic Bulletin & Review*, *9*, 59–79.
- Gagnon, D. A., & Schwartz, M. F. (1997). Serial position effects in aphasics' neologisms. *Brain and Language*, *60*, 87–89.
- Glasspool, D. W., & Houghton, G. (2005). Serial order and consonant-vowel structure in a graphemic output buffer model. *Brain and Language*, *94*, 304–330.
- Glasspool, D. W., Shallice, T., & Cipolotti, L. (2006). Towards a unified process model for graphemic buffer disorder and deep dysgraphia. *Cognitive Neuropsychology*, *23*, 479–512.
- Goldberg, A. M., & Rapp, B. (2008). Is compound chaining the serial-order mechanism of spelling? A simple recurrent network investigation. *Cognitive Neuropsychology*, *25*, 218–255.
- Gupta, P., Lipinski, J., Abbs, B., & Lin, P. H. (2005). Serial position effects in nonword repetition. *Journal of Memory and Language*, *53*, 141–162.

- Gusfield, D. (1997). *Algorithms on strings, trees, and sequences: Computer science and computational biology*. Cambridge, UK: Cambridge University Press.
- Healy, A. F. (1974). Separating item from order information in short-term-memory. *Journal of Verbal Learning and Verbal Behavior*, *13*, 644–655.
- Henson, R. N. A. (1998a). Item repetition in short-term memory: Ranschburg repeated. *Journal of Experimental Psychology: Learning Memory and Cognition*, *24*, 1162–1181.
- Henson, R. N. A. (1998b). Short-term memory for serial order: The start–end model. *Cognitive Psychology*, *36*, 73–137.
- Henson, R. N. A. (1999). Positional information in short-term memory: Relative or absolute? *Memory & Cognition*, *27*, 915–927.
- 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 Section A: Human Experimental Psychology*, *49*, 80–115.
- Houghton, G. (1990). The problem of serial order: A neural network model of sequence learning and recall. In R. Dale, C. Mellish, & M. Zock (Eds.), *Current research in natural language generation*. London: Academic Press.
- Kay, J., & Hanley, R. (1991). Simultaneous form perception and serial letter recognition in a case of letter-by-letter reading. *Cognitive Neuropsychology*, *8*, 249–273.
- Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior* (pp. 112–131). New York: John Wiley and Sons.
- Lee, C. L., & Estes, W. K. (1977). Order and position in primary memory for letter strings. *Journal of Verbal Learning and Verbal Behavior*, *16*, 395–418.
- Lewandowsky, S. (1999). Redintegration and response suppression in serial recall: A dynamic network model. *International Journal of Psychology*, *34*, 434–446.
- Lewandowsky, S., & Murdock, B. B. (1989). Memory for serial order. *Psychological Review*, *96*, 25–57.
- Lewandowsky, S., Oberauer, K., & Brown, G. D. A. (2008). No temporal decay in verbal short-term memory. *Trends in Cognitive Sciences*, *13*, 120–126.
- Machtynger, J., & Shallice, T. (2009). Normalizing serial position analyses: The Proportional Accountability algorithm. *Cognitive Neuropsychology*, *26*, 217–222.
- Martin, N., & Saffran, E. M. (1997). Language and auditory-verbal short-term memory impairments: Evidence for common underlying processes. *Cognitive Neuropsychology*, *14*, 641–682.
- Martin, R. C., Shelton, J. R., & Yaffee, L. S. (1994). Language processing and working-memory—neuropsychological evidence for separate phonological and semantic capacities. *Journal of Memory and Language*, *33*, 83–111.
- Mason, M. (1982). Recognition time for letters and non-letters—effects of serial position, array size, and processing order. *Journal of Experimental Psychology: Human Perception and Performance*, *8*, 724–738.
- Murdock, B. B. (1968). Serial order effects in short-term memory. *Journal of Experimental Psychology Monographs*, *76*(4, Pt. 2), 1–15.
- Murdock, B. B. (1993). Todam2—a model for the storage and retrieval of item, associative, and serial-order information. *Psychological Review*, *100*, 183–203.
- Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, *18*, 251–269.
- Nairne, J. S. (1991). Positional uncertainty in long-term-memory. *Memory & Cognition*, *19*, 332–340.
- Nairne, J. S. (2002). Remembering over the short term: The case against the standard model. *Annual Review of Psychology*, *53*, 53–81.
- Neils, J., Roeltgen, D. P., & Greer, A. (1995). Spelling and attention in early Alzheimers-disease—evidence for impairment of the graphemic buffer. *Brain and Language*, *49*, 241–262.
- Olson, A. C. (1995). *Syllables, letter frequency and sound: Orthographic structure in deaf reading and spelling*, Unpublished PhD thesis, The Johns Hopkins University, Baltimore, MD, USA.
- Olson, A. C., & Caramazza, A. (1999). Orthographic structure and deaf spelling errors: Syllables, letter frequency and speech. *Quarterly Journal of Experimental Psychology*, *57A*, 385–417.
- Olson, A. C., Romani, C., & Halloran, L. (2007). Localizing the deficit in a case of jargonaphasia. *Cognitive Neuropsychology*, *24*, 211–238.
- Page, M. P. A., Madge, A., Cumming, N., & Norris, D. G. (2007). Speech errors and the phonological similarity effect in short-term memory: Evidence suggesting a common locus. *Journal of Memory and Language*, *56*, 49–64.
- Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review*, *105*, 761–781.
- Papagno, C., & Girelli, L. (2005). Writing through the phonological buffer: A case of progressive writing disorder. *Neuropsychologia*, *43*, 1277–1287.

- 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, 303–311.
- Robinson, E. S., & Brown, M. A. (1926). Effect of serial position upon memorization. *American Journal of Psychology*, 37, 538–552.
- Romani, C., Galluzzi, C., & Olson, A. (in press). Phonological-lexical activation: A lexical component or an output buffer? Evidence from aphasic errors. *Cortex*.
- Romani, C., Olson, A., Ward, J., & Ercolani, M. G. (2002). Formal lexical paraphasias in a single case study: How “masterpiece” can become “misterpie-man” and “curiosity” “suretoy”. *Brain and Language*, 83, 300–334.
- Schiller, N. O., Greenhall, J. A., Shelton, J. R., & Caramazza, A. (2001). Serial order effects in spelling errors: Evidence from two dysgraphic patients. *Neurocase*, 7, 1–14.
- Schneider, D. W., & Logan, G. D. (2005). Modeling task switching without switching tasks: A short-term priming account of explicitly cued performance. *Journal of Experimental Psychology: General*, 134, 343–367.
- Schwartz, M. F., Wilshire, C. E., Gagnon, D. A., & Polansky, M. (2004). Origins of nonword phonological errors in aphasic picture naming. *Cognitive Neuropsychology*, 21, 159–186.
- Shallice, T., Rumiati, R. I., & Zadini, A. (2000). The selective impairment of the phonological output buffer. *Cognitive Neuropsychology*, 17, 517–546.
- Tichy, W. F. (1984). The string-to-string correction problem with block moves. *ACM Transactions on Computer Systems*, 2, 309–321.
- Tydgat, I., & Grainger, J. (2009). Serial position effects in the identification of letters, digits, and symbols. *Journal of Experimental Psychology: Human Perception and Performance*, 35, 480–498.
- Ward, J., & Romani, C. (1998). Serial position effects and lexical activation in spelling: Evidence from a single case study. *Neurocase*, 4, 189–206.
- Wing, A. M., & Baddeley, A. D. (1980). *Spelling errors in handwriting: A corpus and a distributional analysis*. London: Academic Press.
- Wing, A. M., & Baddeley, A. D. (2009). Righting errors in writing errors: The Wing and Baddeley (1980) spelling error corpus revisited. *Cognitive Neuropsychology*, 26, 223–226.

APPENDIX

Calculations used for the simple inference model

We assume that interference is created between adjacent positions because position codes are similar, but that this similarity decreases with distance. This means, for example, that when the second item in a sequence needs to be produced, there is some probability that the first or third item will be produced instead, but it is less likely that the fourth item will be produced. To put concrete numbers to these terms, when we assumed interference between adjacent items only, Position 2 had a probability of .4 of being produced correctly, and Positions 1 and 3 had a probability of .3 each of being produced instead. To implement this, we used a template of probabilities that we moved across the sequence. The template for adjacent-only interference was .3, **.4**, .3, and the template for interference extending across 2 positions was .1, **.2**, **.4**, **.2**, **.1** (the correct position is in bold). Where the template went

beyond the beginning or end of a list it was normalized. For example, in Position 1, an error to the left is not possible, so the probabilities of .4 for the first position and .3 for the second position were normalized to probabilities of $.4/.7 = .57$ and $.3/.7 = .43$. Using the probabilities generated by our templates, we “rolled the dice” and produced the item indicated by the template. We generated 1,000 Monte Carlo trials at each position for each sequence length to produce Figures 2a and 2b. Note that normalizing the probabilities in the way we did produces a shallower U-shaped function than would be produced if probabilities were simply “piled up” at the ends of the list so that the probability of producing Position 1 correctly was .7, and the probability of producing Position 2 was .3 when the template was .3, **.4**, **.3**. This is also a plausible account of end effects. The point is that both of these assumptions produce U-shaped functions with position, and the shape of the U depends on the distance over which position codes interfere due to similarity. To produce Figure 3, positions that had already been produced could not be produced again, and probabilities were normalized over positions that were still candidates for production.