



Published in final edited form as:

*Cognition*. 2010 June ; 115(3): 466–490. doi:10.1016/j.cognition.2010.03.013.

## Representation of Letter Position in Spelling: Evidence from Acquired Dysgraphia

Simon Fischer-Baum<sup>\*</sup>, Michael McCloskey, and Brenda Rapp

Department of Cognitive Science, Johns Hopkins University

### Abstract

The graphemic representations that underlie spelling performance must encode not only the identities of the letters in a word, but also the positions of the letters. This study investigates how letter position information is represented. We present evidence from two dysgraphic individuals, CM and LSS, who perseverate letters when spelling: that is, letters from previous spelling responses intrude into subsequent responses. The perseverated letters appear more often than expected by chance in the same position in the previous and subsequent responses. We used these errors to address the question of how letter position is represented in spelling. In a series of analyses we determined how often the perseveration errors produced maintain position as defined by a number of alternative theories of letter position encoding proposed in the literature. The analyses provide strong evidence that the grapheme representations used in spelling encode letter position such that position is represented in a graded manner based on distance from both edges of the word.

### Keywords

letter position coding; dysgraphia; spelling; letter perseveration errors; orthographic processing

---

Many cognitive functions require the ability to represent and process sequences of items or events. Sequence information is essential, for example, in recalling a telephone number, reasoning about causes and effects, navigating a route through an environment, or producing a sentence. As Karl Lashley pointed out more than 50 years ago in *The problem of serial order in behavior* (Lashley, 1951), the question of how the brain represents and processes ordered sequences is far from trivial; and this question remains a central concern for research in a variety of cognitive domains (e.g., working memory: Henson, 1998; motor control: Bullock, 2004; reading: Grainger & Whitney, 2004; music performance: Palmer, 2005; spoken language production: Dell, Burger & Svec, 1997).

This article addresses the serial order issue in the context of spelling. Spelling a word requires not only information about the identities of the letters in the word, but also information about the ordering of those letters. This ordering information could be encoded in a variety of ways. In the word PENCIL, for example, the letter E could be represented as the second letter in the word, the letter five positions from the end of the word, the letter in the nucleus of the first (orthographic) syllable, or the letter that follows P and precedes N. In each case the E's position

---

<sup>\*</sup>Corresponding author: Department of Cognitive Science, Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218, Tel: 410-516-7625, Fax: 410-516-8020, fischerbaum@cogsci.jhu.edu.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

is specified according to a different representational scheme. If we say that the E is the second letter in the word, we implicitly adopt a left-edge based scheme, in which a letter's position is defined in terms of distance in letters from the left edge of the word. By this positional scheme, P is the first letter, E is the second, and so forth. Alternatively, if we say that E is the letter following P and preceding N, we are using a letter-context scheme, in which a letter's position is specified with respect to the surrounding letters.

The goal of present study was to identify the scheme for representing letter position in the graphemic representations that underlie spelling performance. Several researchers have offered hypotheses about the representation of letter position in spelling (e.g., Brown & Loosemore, 1994; Caramazza & Hillis, 1990; Glasspool, 1998; Glasspool & Houghton, 2005; Houghton, Glasspool & Shallice, 1994; Houghton & Zorzi, 2006). However, the relevant empirical evidence is sparse, and no studies have directly compared the alternative proposals. In the present study we examine a broad range of positional schemes in light of data from two individuals with acquired dysgraphia, LSS and CM. In spelling tasks LSS and CM made frequent letter perseveration errors, in which letters from prior responses intruded into subsequent responses. We argue that these letter perseveration spelling errors motivate strong conclusions about the representational scheme used for specifying letter position in spelling

Perseveration errors – both from impaired and unimpaired individuals – have been used in a variety of domains to infer how the positions of elements in a sequence are represented (e.g., Boomer & Laver, 1968; Cohen & Dehaene, 1990; Henson, 1999). In the present study extensive testing of LSS and CM provided large sets of letter perseveration errors that allowed us to contrast alternative hypotheses of letter position representation in spelling. Additional aspects of the participants' spelling performance localized their perseveration errors to the level of abstract letter representation – what we call the level of graphemic spelling representation. This localization places some constraints on the levels of processing at which the implicated letter position representations may be active.

Patterns of performance in individuals with dysgraphia acquired as a result of neural insult (e.g., stroke) have been used extensively as a basis for conclusions about the cognitive representations and processes that support spelling in the intact brain (e.g., Caramazza & Hillis, 1990; Caramazza & Miceli, 1990; McCloskey et al., 1994; McCloskey, Macaruso & Rapp, 2006; Rapp et al., 1997; Tainturier & Caramazza, 1996, see McCloskey, 2003 for discussion). The logic by which impaired performance can be used to draw inferences about normal cognition has been discussed at length elsewhere, and we refer the interested reader to those sources (e.g., Caramazza, 1984, 1986, 1992; Caramazza & Coltheart, 2006; McCloskey, 1993, 2001, 2003; McCloskey & Caramazza, 1988). In accord with this logic we assume that the brain damage suffered by LSS and CM has caused their previously-normal spelling processes to malfunction (leading to perseverations and other errors), but has not resulted in creation of novel representational schemes for specifying letter position. Given this assumption we can use the perseveration errors to draw conclusions about representation of letter position in the normal spelling system. An important advantage of studying impaired performance is that, as in the present study, one can often accumulate large corpora of errors that arise from a single level of representation, and are highly informative about the nature of the representations at that level.

The question of how the position of an element in a sequence is represented is critical for all domains that rely on sequence processing. In the past few years, this question has received a great deal of attention in the domain of reading (e.g., Davis, 1999; Davis & Bowers, 2004, 2006; Grainger et al., 2006; Grainger & van Heuven, 2003; Gomez, Perea & Ratcliff, 2008; Kinoshita & Norris, 2009; Perea & Lupker, 2003, 2004; Schoonbaert & Grainger, 2004; Van Assche & Grainger, 2006; Whitney, 2001). The research on reading provides a source of

hypotheses regarding position representation in orthographic processing generally. However, it is important to emphasize that our results and conclusions are specific to spelling. Because we do not know whether reading and spelling use the same scheme for representing letter position (and because we did not study LSS's or CM's reading in detail), we make no claims about position representation in reading.

As a framework for subsequent discussion, we begin by sketching a theory of the cognitive mechanisms involved in spelling, and then lay out a variety of hypotheses concerning the encoding of letter position in graphemic spelling representations. Next we offer case histories for CM and LSS, and characterize their spelling deficits. Following this introductory material we present results demonstrating that both participants often perseverate letters from spelling responses into subsequent responses. We then report an extensive series of analyses that use the letter perseveration phenomenon as a tool for probing the representation of letter position in graphemic spelling representations. Finally, we conclude with a brief discussion of issues arising from our results and conclusions.

## A Cognitive Spelling Theory

Most of the data we report come from a writing to dictation task, in which a word or nonword is dictated, and the participant produces a written spelling response. Consequently, we describe the cognitive spelling theory in the context of this task (see Tainturier & Rapp, 2001, and Miceli & Capasso, 2006).

The theory assumes that when a familiar word (e.g., “table”) is dictated, the corresponding phonological lexeme is activated in a phonological lexicon (see Figure 1). This lexeme then activates a lexical-semantic representation, which in turn activates an orthographic lexeme in an orthographic lexicon. Some authors have also proposed a direct connection between the phonological and orthographic lexemes (e.g., Patterson, 1986). Next, the orthographic lexeme activates a graphemic representation that specifies the identities and ordering of the letters in the word (e.g., T-A-B-L-E). These graphemic representations are assumed to abstract away from information about how the letter is to be produced; TABLE is represented with the same T grapheme whether the word is to be typed, handwritten or spelled aloud. Note that we use the terms *grapheme* and *graphemic* simply to refer to abstract letter representations, and not specifically to letter representations corresponding to single phonemes (e.g., Rapp & Caramazza, 1997). In the present study our focus is on the representation of letter order, and we leave open the question of whether digraphs (letter pairs associated with a single phoneme, such as the SH in FISH) are represented by one unit or two at the graphemic level (Houghton & Zorzi, 2003; Tainturier & Rapp, 2004). The analyses we report treated digraphs as two separate letters, but the pattern of results is the same if digraphs are treated as single units.

When a nonword (e.g., /floup/) is dictated, no phonological lexeme or lexical-semantic representation is available to be activated. For nonword stimuli, the theory assumes that graphemic representations (e.g., F-L-O-P-E) are generated by a phoneme-to-grapheme conversion process that uses sublexical sound-to-spelling correspondences (e.g., /f/ is usually spelled with the letter F).

The abstract graphemic representations generated from a dictated word or nonword provide the basis for activating specific letter-shape representations (e.g., representations of lower-case print letters). In an oral spelling task the graphemic representation would instead activate letter-name representations, such as /ti/ for the letter T. During the selection and processing of letter-shape or -name representations, the graphemic representation is assumed to be processed within a working memory system termed the graphemic buffer (e.g., Caramazza, Miceli & Villa, 1987).

## Representation of Letter Position

The letter perseveration errors produced by CM and LSS provide an opportunity to explore how letter position is encoded at the level of graphemic representations. In this article we address two specific questions. First, relative to what reference point (or points) is the position of a letter defined? Second, are letter position representations discrete, such that each position has a representation that is entirely distinct from that of every other position; or are position representations graded and overlapping, such that representations specifying closer positions are more similar than those specifying more distant positions?

### Reference Point(s)

We illustrate the different hypotheses to be examined using a representational scheme first proposed in the interactive activation reading model (McClelland and Rumelhart, 1981). Each position in a word is represented by a distinct pool of 26 letter units. For example, Figure 2A illustrates a representation of the word FACE in a system that uses six pools of letter nodes to represent words of up to six letters. The example assumes that positions are defined relative to the left edge of the word, and so FACE is represented by activating the F node in the L+1 pool, the A node in the L+2 pool, and so forth.

A variety of reference points have been posited in proposed schemes for encoding letter position in the graphemic representations that underlie reading and/or spelling. The schemes can be divided into three broad classes. In *content-independent* schemes, a letter's position is determined by distance (in number of letters) and direction from one or more reference points that do not themselves depend on the identities of the letters in the word. In *letter-context* schemes, a letter's position is defined relative to its surrounding letters. In *syllabic* schemes, a letter's position is defined relative to the syllable in which it appears, and its role within that syllable.

**Content-Independent Representational Schemes**—We consider five content-independent schemes that differ with respect to the reference point(s) used in defining letter position: left-edge, right-edge, center, closer-edge, and both-edges.

**Left-Edge:** The left-edge scheme, illustrated in Figure 2A, defines position relative to the beginning of the word, such that the first letter in any word occupies position L+1, the second letter position L+2, and so forth. Left-edge representational schemes have been adopted in several computational models of reading (Coltheart, Curtis, Atkins & Haller, 1993; Coltheart, Rastle, Perry, Langdon & Ziegler, 2001; Ellis, Flude & Young, 1987; Seymour, 1979).

**Right-Edge:** The right-edge scheme, illustrated in Figure 2B, defines position relative to the end of the word, such that the last letter in any word occupies position R-1, the next-to-last letter position R-2, and so forth. The right- and left-edge schemes differ with respect to which letters in words of different lengths are assigned to the same position. For example, in the left-edge scheme the first letter in every word occupies the same position (L+1), but the last letters in words of different lengths will differ in their position assignments (e.g., L+4 for the E in FACE vs. L+6 for the T in SILENT). For the right-edge scheme, in contrast, all last letters occupy the same position (R-1), but first letters vary in position depending on length (e.g., R-4 for the F in FACE vs. R-6 for the S in SILENT). To our knowledge, the right edge has never been proposed as the sole reference point for representing letter position in reading or spelling, but we include this scheme for the sake of completeness.

**Center:** A center-based representational scheme defines position relative to the center of the word, as illustrated in Figure 2C for the word FACE. In this example the letter F is assigned

to position C-2 (two positions left of center), the A to position C-1, the C to position C+1 (one position right of center), and the E to position C+2.<sup>1</sup> On the basis of errors produced by an individual with neglect dyslexia and dysgraphia, Caramazza and Hillis (1990) proposed a center positional scheme for the graphemic representations underlying reading and spelling.

**Closer-Edge:** Jacobs et al. (1998) proposed that the graphemic representations implicated in reading instantiate what we will refer to as a closer-edge scheme. In this scheme, letter position is encoded relative to the left or right edge of the word, whichever is closer (see Figure 2D). For the word FACE the F and A are represented by units in pools L+1 and L+2, respectively, because these letters are closer to the left than to the right edge of the word. However, the C and E are represented by units in the R-2 and R-1 pools, respectively, because these letters are closer to the right edge of the word.<sup>2</sup>

**Both-Edges:** An alternative to the closer-edge scheme is a scheme that assigns position on the basis of distance from both edges of the word simultaneously, as illustrated in Figure 2E. The F in FACE is represented by a unit in the L+1 pool and also by a unit in the R-4 pool, because the F is both the first letter from the left edge of the word, and the fourth letter from the right edge. Glasspool and colleagues posit a both-edges scheme in their competitive queuing models of spelling (Glasspool, 1998; Glasspool & Houghton, 2005; Houghton, Glasspool & Shallice, 1994).

**Letter-Context Schemes**—In letter-context schemes a letter's position is represented relative to other letters in the word; that is, the reference points for representing a letter's position are provided by the other letters in the word. Various theories of word recognition (Dehaene, Cohen, Sigman & Vincker, 2005; Grainger & Van Heuven, 2003; Mozer, 1987; Seidenberg & McClelland, 1989; Whitney, 2001) and spelling (Brown & Loosemore, 1994) posit some form of letter-context coding. Figure 3A illustrates a simple letter-context scheme, in which the position of a letter is defined by the immediately preceding letter. Each pool of letter units corresponds to a preceding letter context. For example, the A\_ pool is used to represent any letter that is preceded by an A. The C in FACE is therefore represented by activating the C unit in this pool. The #\_ pool represents the position preceded by the left word boundary, and hence the F in FACE is represented by a unit in this pool.

This preceding-letter scheme is a bigram representation, because the word is represented as a set of bigrams (e.g., #F, FA, AC, CE, E# for FACE). Alternatives to this simple bigram scheme have been proposed in both spelling and reading. In trigram schemes (Seidenberg & McClelland, 1989; Brown & Loosemore, 1994) the position of a letter is defined jointly by the immediately preceding and immediately following letters, so that the C in FACE, for example, could be represented by a unit in an A\_E pool (i.e., a pool representing letters preceded by an A and followed by an E). In open bigram schemes (Dehaene et al., 2005; Grainger & Van Heuven, 2003; Whitney, 2001) the position of a letter is defined not only by the letter that comes immediately before and/or after it, but also by other letters. For example, representing the C in FACE might involve activating the C unit in an F\_ pool as well as in an A\_ pool, to indicate that the C was preceded by both an F and an A.

<sup>1</sup>Three versions of the center representational scheme can be formulated, depending upon how a position is assigned to the central letter in words with an odd number of letters. Using the A in STAMP for purposes of illustration, this letter could be assigned to (1) position C, a position that does not occur for words of even length; (2) position C+1 or (3) position C-1. We considered all three of these variants in the analyses reported below. We will report results from the second of these center schemes, as it performed the best relative to the other two.

<sup>2</sup>Two variants of the closer edge scheme may be defined, differing in the position assigned to the central letter in words with an odd number of letters. In these variants the A in STAMP is assigned either to position L+3 or position R-3. Both variants were considered in our analyses, though we report results from the closer-edge scheme in which the central letter was assigned a position relative to the right edge of the word, as it was the better performing of the two.



**Syllabic Schemes**—In the third major class of representational scheme the letters of a word are organized into orthographic syllables, analogous to phonological syllables (Caramazza & Miceli, 1990; Badecker, 1996). As in phonology, each syllable has an internal structure, consisting of a nucleus, an optional onset, and an optional coda. In syllabic schemes a letter's position is represented in terms of the syllable in which it appears and its role within that syllable.

Figure 3B shows a simplified syllabic position representation for the word FACE, which is bisyllabic orthographically (FA-CE) although monosyllabic phonologically. The F is represented by a unit in a Syllable 1 Onset pool, and the A by a unit in a Syllable 1 Nucleus pool. Similarly, the C and E are represented by units in Syllable 2 Onset and Nucleus pools, respectively. Orthographic onsets, nuclei and codas may contain multiple letters (as, in BREATH); as a consequence, multiple onset, nucleus, and coda pools are needed for an adequate syllabic positional scheme (e.g., Onset 1, Onset 2, Onset 3).

Many researchers have argued that graphemic representations include information about syllabic structure (e.g., Badecker, 1996; Caramazza & Miceli, 1990; Olson & Caramazza, 2004; Spoehr & Smith, 1973; Prinzmetal, Treiman & Rho, 1986, Rapp, 1992; Taft, 1979). This claim does not necessarily imply that syllabic information plays a role in representing the letter order. Nevertheless, it is possible that letter position could be represented through a syllabic scheme, as assumed in several simulation models of reading aloud (Harm & Seidenberg, 1999; Plaut, McClelland, Seidenberg & Patterson, 1996; Zorzi, Houghton & Butterworth, 1998) and spelling-to-dictation (Houghton & Zorzi, 2003).

### Discrete vs. Graded Position Representations

Schemes for representing letter position differ not only with respect to reference points, but also with respect to whether the posited representations are discrete or graded. In a discrete scheme two position representations are either the same (e.g., L+1 and L+1) or different (e.g., L+1 and L+2), such that L+1 and L+2 are no more similar to one another than L+1 and L+5. In a graded scheme, on the other hand, position representations exhibit a similarity structure, such that the closer two positions are, the more similar their representations.

Figure 4 illustrates one possible form of a graded representation for a positional scheme with a left-edge reference point. Each letter in a word is represented by activation in more than one pool of units, and each pool of units contributes to representing multiple letters in a word (see Gomez, Perea & Ratcliff, 2008 for a discussion of graded position representation in reading). For example, the A is represented by activating the A unit in the L+2 pool strongly, the A unit in the L+1 and L+3 pools somewhat less strongly, and so forth. Thus, as shown in Figure 4, the representation of each letter is distributed across several pools and each pool is associated with a gradient of activation across letters. For example, the L+1 pool shows a gradient of activation in which the F unit is most strongly activated, the A unit somewhat less strongly activated, and so forth (e.g., Houghton et al., 1994). Thus, the F and A have more similar position representations than the F and the E. Discrete representations do not have this property.

The issue of graded versus discrete position representation applies to syllabic as well as content-independent schemes. Some syllabic schemes (e.g., Harm & Seidenberg, 1999; Plaut et al., 1996) assume graded representations in which positions within a syllabic role (e.g., first vs. second onset position) have more similar representations than positions for different syllabic roles (e.g., first onset position vs. first coda position). Other syllabic theories (e.g., Zorzi, Houghton & Butterworth, 1998) assume discrete position representations with no similarity structure.

We turn now to LSS and CM, the two dysgraphic individuals whose letter perseveration errors provide a basis for distinguish among the alternative hypotheses of letter position representation in spelling. We first present brief case histories and summaries of neuropsychological assessment results, focusing on the participants' spelling deficits. We then report analyses establishing that both participants produce letter perseveration errors that arise at an abstract graphemic level of representation. Finally, we present an extensive series of analyses that contrast the competing position representational schemes.

## Case Histories & Neuropsychological Assessments

CM was a right-handed man with a Ph.D. in electrical engineering. He worked as a university professor until suffering a stroke in September 1986 at age 59. LSS was a left-handed man with a Master's degree in psychology. He worked as a regional sales representative for a health services company until a stroke in July 2003 at age 54. For both participants CT scan showed extensive cortical and subcortical damage in the distribution of the left middle cerebral artery. Both reported excellent premorbid spelling, confirmed in CM's case by a sample of handwritten lecture notes. A standardized battery of neuropsychological tests was administered to each participant: the Boston Diagnostic Aphasia Examination (BDAE: Goodglass & Kaplan, 1983) for CM, and the Western Aphasia Battery (WAB: Kertesz, 1982) for LSS. Appendix 1 provides a detailed discussion of the results. CM performed well on tests of auditory comprehension and reading comprehension, but showed impairment in both spoken and written language production tasks. LSS performed well in auditory comprehension of single words and short, simple sentences, but was impaired in comprehension of longer, more complex sentences. He was relatively intact on several spoken language production tasks, but performed more poorly on measures of fluency. For additional information and findings concerning CM see McCloskey, Macaruso and Rapp (2006).

### Auditory Comprehension of Words

Additional tests of auditory word comprehension were administered because single-word auditory comprehension is required for the writing to dictation task used to probe the participants' spelling. On a picture-word matching task in which a spoken word was paired with a matching picture, a semantic foil, or a phonological foil, CM correctly indicated whether the word and picture matched on 98% (425/432) of the trials. In a similar task LSS was 97% correct (253/260). In two administrations of the Peabody Picture Vocabulary Test–Revised (Dunn & Dunn, 1981), a more challenging spoken word-picture matching task, CM's raw scores were 147/175 and 161/175, placing him within the Average range for normal adults. LSS scored 135/175 and 128/175, placing him within the Moderately Low range. However, most of LSS's errors involved words (e.g., “fettered”) far lower in frequency than the words presented for spelling in the present study. We conclude that single-word auditory comprehension for both participants was sufficient to support comprehension of the dictated stimuli in the writing to dictation task.

### Spelling

Both participants' spelling abilities were tested extensively. Over five years, CM spelled 3797 words to dictation, with 55% word accuracy (2101/3797). Over 16 months, LSS spelled 1636 words to dictation, with an accuracy of 20% (329 /1636). Performance remained stable over the testing period for both participants. Table 1 presents examples of their spelling errors. For both CM and LSS, errors were orthographically related to the correct spelling. The errors included letter substitutions (e.g., CM: “absence” → ABSONCE), deletions (e.g., LSS: “future” → FUTUE), insertions (e.g., LSS: “could” → COUNLD) and transpositions (e.g., CM: “riot” → ROIT). For each participant, some errors were correct spellings of words similar to the target (e.g., CM: “belief” → BEING; see McCloskey et al., 2006, for detailed discussion

of these errors in the case of CM). Virtually all errors were phonologically implausible spellings, and very few were semantically or morphologically related to the target.

Given that CM and LSS performed adequately in spoken word comprehension tasks, their spelling errors cannot be attributed to impaired comprehension of the dictated stimulus words. To assist in identifying the functional locus at which the spelling errors arose, both participants were tested with the Johns Hopkins Dysgraphia Battery (Goodman & Caramazza, 1985). Detailed results are provided in Appendix 2; here we summarize the principal findings.

Both CM and LSS performed very poorly in spelling nonwords, and their spelling accuracy for words was unaffected by phoneme-to-grapheme probability (i.e., spelling regularity, or the likelihood that a word could be spelled correctly via the sublexical phoneme-grapheme conversion route). These results imply that the sublexical route is severely impaired in both participants.

When CM and LSS directly copied word and nonword stimuli, transcoding from upper to lower case, or vice versa, they performed well, indicating that their spelling errors did not result from peripheral deficits in processing letter shape representations. However, both were impaired at a delayed copy transcoding task. These findings, in conjunction with the observed error types (predominantly letter substitutions, deletions, insertions, and transpositions), implicate a deficit at the level of abstract grapheme representations.

Both CM and LSS spelled high-frequency words more accurately than low-frequency words, and LSS also spelled concrete words more accurately than abstract words. These effects of lexical variables point to impairment affecting some aspect of the lexical spelling process. Given that single-word auditory comprehension was intact, we conclude that both participants could activate phonological lexemes and lexical-semantic representations from the dictated words. Hence, the lexical effects in spelling presumably reflect impairment at the level of orthographic lexemes. Neither CM nor LSS showed the effect of word length on letter accuracy that is typical of deficits affecting the graphemic buffer or working memory. Taken together, therefore, these results indicate that the errors made by both participants in spelling words stem from disruption in processes that activate abstract grapheme representations from orthographic lexemes and maintain appropriate levels of grapheme activation during production of a spelling response.

## Letter Perseveration Errors

Most of the spelling errors produced by LSS and CM contained intruded letters—that is, letters that did not appear in the correct spelling of the target word. Letter intrusions may take the form of substitutions or insertions. For example, in the substitution error “head” → HEAT, the T is an intruded letter, as is the I in the insertion error “spend” → SPIEND.

For both participants, intruded letters were often present in one or more of the preceding responses, raising the possibility that at least some of the intruded letters were perseverations from prior responses. For example in CM's error “edge” → ERGE the R is an intruded letter. Table 2 illustrates the trial on which the error occurred, labeled E, and the five immediately preceding trials (E-1, E-2, etc.). As shown in the table, the intruded letter was present in the E-1 response (FRENCE) as well as the response on trial E-2 (MORTCH). Conceivably, then, the intruded R in ERGE was a perseveration from FRENCE and/or MORTCH. However, one would expect sometimes to find an intruded letter in a prior response simply by chance. To establish whether CM and/or LSS were perseverating letters from preceding responses, we evaluated whether the intruded letters appeared in preceding responses more often than expected by chance.



The letter perseveration analyses evaluated whether intruded letters were more likely than expected by chance to be present in the responses on each of the five trials immediately preceding an error (trials E-1 to E-5). For each participant and for each intruded letter, a computer program first tabulated whether or not the letter was present in the response on trial E-1. For example, the intruded R in CM's error "edge" → ERGE was present in the E-1 response, FRENCE. From this tabulation the program calculated the proportion of intruded letters that were present in the corresponding E-1 responses.

Next the program computed the proportion of intruded letters that did not appear in the E-1 response but did appear in the response on trial E-2. For example, in LSS's error "kitchen" → KITCHEM, the intruded M was not present in the response on trial E-1 (BELEAFE), but did appear in the E-2 response (SYSTEM). The program then computed the proportion of intruded letters that did not appear in the E-1 or E-2 response, but did appear in the response on trial E-3; and so forth through trial E-5.

The results are shown in the solid curves of Figure 5A (CM) and 5B (LSS). The .76 value plotted for trial E-1 in Figure 5B reflects the fact that 1748 of LSS's 2301 intruded letters (76%) were present in the response immediately prior to the error. The .396 value for trial E-2 indicates that 39.6% of the intruded letters that did not appear in the E-1 response were present in the response on trial E-2.

The perseveration analysis program also estimated for each participant the likelihood of an intruded letter being present by chance in the E-1 through E-5 responses. The chance estimates were based on the following rationale: If the occurrence of a letter intrusion was unrelated to the presence of the intruded letter in the immediately preceding responses, then we should be just as likely to find the intruded letter in responses occurring on trials distant from the intrusion trial.

We illustrate our method for evaluating chance by considering the trial E-1 chance analysis for CM. A total of 2704 intrusion errors were included in the analysis, 4 of which are shown in Figure 6. For each actual E-1 response (e.g., FRENCE), regardless of whether the E-1 response contained the letter intruded on trial E, a computer program created a pool of control responses consisting of all responses in the participant's spelling corpus having the same length as the actual E-1 response. These were drawn randomly from the participants' responses that were produced outside the window of the five trials either preceding or following the intrusion error. As illustrated in Figure 6, the control pool for the actual E-1 response FRENCE included the responses BALLON, WINDOW and SHINGE (among many others).

A Monte Carlo analysis used the control response pools to estimate the proportion of E-1 responses expected by chance to contain the intruded letter. For each of the 2704 letter intrusions the program randomly sampled a response from the control pool, and tabulated whether that response included the intruded letter. The program then computed the proportion of sampled control responses that contained the intruded letter. This process—randomly sampling an E-1 control response for each letter intrusion, and computing the proportion that contained the intruded letter—was carried out 10,000 times for the complete set of letter intrusions for each participant. On the first run SHINGE was selected as the control for the actual E-1 response FRENCE, HORN was selected as the control for KIGH, and so forth (see *Control Sample 1* column in Figure 6). Across one random control run, 709 of the 2704 E-1 control responses (a proportion of .262) contained the corresponding intruded letter. On a second random sampling run CUPRET and NOSE were selected as controls for FRENCE and KIGH respectively, and for this run the proportion of controls containing the intruded letter was .254. Each of the random control proportions constitutes an estimate of the proportion of actual E-1 responses expected by chance to contain the intruded letter.

The mean of the 10,000 control proportions was .268, indicating that for CM we would expect an intruded letter to be present in the E-1 response by chance about 27% of the time. This is in sharp contrast to the finding that intruded letters were present in the actual E-1 responses far more often, about 45% of the time. In fact all of the 10,000 individual chance proportions were lower than the observed proportion of .453, indicating that, for CM, intruded letters were present in the response immediately preceding the intrusion significantly more often than expected by chance ( $p < .0001$ ).

This type of Monte Carlo chance analysis was carried out for trials E-1 through E-5 for both CM and LSS. For CM the observed proportion of preceding responses containing the intruded letter was reliably higher than the proportion expected by chance for trials E-1 through E-4 ( $p < .0001$  for E-1 through E-3,  $p < .01$  for E-4). At E-5, however, the observed proportion did not differ from the chance proportion ( $p > .1$ ). For LSS, observed proportions were reliably higher than chance for trials E-1 and E-2 ( $p < .0001$ ) but not for trials E-3 through E-5 ( $p > .5$ ).

In addition to demonstrating that at least some of LSS's and CM's letter intrusion errors were true perseveration errors, these results allow us to define for each participant the *window of perseveration*, the range of preceding trials from which letters may perseverate into the current response. For CM, intruded letters were present at above-chance rates in the responses on trials E-1 through E-4, but not in trial E-5 responses. This finding indicates that letters from as many as four trials back could occur as perseverations in CM's responses. Accordingly, we define the window of perseveration for CM to include the four responses prior to an error. For LSS, intruded letters were present at above-chance rates only in the E-1 and E-2 responses, and so the perseveration window for LSS includes only the two trials preceding an error.<sup>3</sup>

### Functional Locus of Letter Perseverations

Having established that CM and LSS commit letter perseveration errors, we may ask at what level of representation these errors arise. Several considerations lead to the conclusion that for both participants the perseverations occur at a level of abstract grapheme representations. Earlier, we argued on the basis of the language and spelling assessments that the spelling errors made by CM and LSS result from impairment in processes that activate grapheme representations, and/or maintain the activation of these representations during production of spelling responses. McCloskey et al. (2006) presented evidence from additional tasks with CM that further support this conclusion. We administered similar tasks to LSS with the same results. Briefly, the critical findings were as follows: (1) Both LSS and CM showed significant perseveration effects when spelling in a modality other than writing (oral spelling for LSS and typing for CM). (2) In writing to dictation both participants showed significant cross-case perseveration effects (e.g., intruding a lower-case e after writing an upper-case E in the immediately preceding response). These results indicate that both participants' letter perseverations arose prior to the activation of form-specific letter representations (e.g., letter shape representations). (3) Other results establish that perseverations arose from grapheme and not phoneme representations. LSS and CM perseverated the letter F only when preceding targets included Fs (e.g., AFRAID) and not when preceding targets included an /f/ phoneme

<sup>3</sup>The perseveration analyses presented here differed in several respects from those reported for CM in McCloskey et al. (2006). (1) McCloskey et al. tabulated the presence of the intruded letter in each of the E-1 through E-5 responses, regardless of whether the letter appeared in a response closer to the trial E response. The present analysis is more suited to determining the window of perseveration. (2) McCloskey et al. counted as intrusions letters that appeared in the correct spelling of a word but occurred too many times in the response (e.g., E in "sheriff" → SHREET); however, the present analyses excluded these letters. (3) McCloskey et al. created the pool of control responses only from responses CM made in spelling the word list in which the intrusion error occurred, whereas the present analyses used the participant's entire spelling corpus. The rationale for generating chance estimates applied here and in McCloskey et al. (2006) is similar to that of Cohen and Dehaene (1998). One notable difference is that our procedure ensures that control responses are matched in length to the corresponding actual responses, and may therefore yield more accurate estimates of chance rates.

not spelled with F (e.g., SPHERE). These results converge on the conclusion that the representations implicated in CM's and LSS's letter perseverations are abstract grapheme representations (Figure 1).

## Letter Position Analyses

In this section we present analyses that use CM's and LSS's letter perseveration errors as a basis for inferences about the positional scheme used in graphemic spelling representations. Before presenting these analyses, we first need to introduce some terminology. We refer to a response with a perseverated letter as a *perseveration response*, and the prior response from which the letter was perseverated as the *source response*. For example, assuming that the R in CM's error “edge” ? ERGE was a perseveration from the immediately preceding response FRENCE, then ERGE is a perseveration response and FRENCE is the corresponding source response. Also, we refer to the to-be-perseverated letter in a source as the source letter; hence the R in FRENCE is the source letter for the perseverated R in ERGE.

The rationale for the position analyses was as follows: If the grapheme representations activated in the course of spelling a word are associated with a particular position, then perseverated letters may non-accidentally maintain source position—that is, the perseveration position may match the source position for reasons other than chance. Consider, for example, the left-edge representation for FACE shown in Figure 2A. Suppose that after FACE is written, the E grapheme remains in an abnormally active state, or that the grapheme in position L+4 of the next stimulus word—for example, the G in PROGRAM—is activated only weakly. Under these circumstances, the E from FACE may perseverate into the response for PROGRAM, and more specifically may perseverate into position L+4, yielding the error PROERAM. In this example the E maintains the same left-edge position when perseverating from the source response FACE to the perseveration response PROERAM.

Suppose, however, that position is defined by a right-edge scheme, as illustrated in Figure 2B. Under this scheme the E in FACE is represented as occurring in position R-1. Consequently, if the E perseverated into the response for PROGRAM, we might expect the perseveration to occur at position R-1, yielding PROGRAE. As these examples illustrate, the various representational schemes will sometimes differ with respect to the position in a source response that matches the position of a perseverated letter in a perseveration response. These differences between representational schemes provide the basis for the position analyses reported below.

The first set of analyses considers discrete content-independent positional schemes (i.e., the schemes illustrated in Figure 2). The second set of analyses asks whether the relationship between perseveration and source positions is more successfully captured by a discrete or gradient form of content-independent representations. The third set of analyses compares the best content-independent scheme with letter-context schemes, and finally the fourth analysis considers syllabic schemes. All analyses were conducted separately for CM and LSS.

### Position Analysis 1: Discrete Content-Independent Schemes

In evaluating the discrete content-independent positional schemes we began by asking, for each scheme, whether CM's and LSS's perseveration errors maintained source position more often than expected by chance.

**Analysis 1a: Observed vs. Chance Position Matches**—Analysis 1a was carried out over all potential perseveration-source pairs. A potential perseveration was defined as a letter intrusion error in which the intruded letter appeared in one or more of the preceding responses within the perseveration window (4 prior responses for CM and 2 for LSS). The preceding responses that contained the intruded letter were considered potential source responses. For

example, Table 3 shows CM's error "blunt" → BLANT, as well as the preceding responses in the four-trial perseveration window (FARM, NOSE, ORANGE, and PIANO). BLANT is a potential perseveration response, because the intruded letter A appears in at least one (and in fact three) of the four responses within the perseveration window (FARM, ORANGE, PIANO). For CM, therefore, the set of potential perseveration-source pairs included BLANT-FARM, BLANT-ORANGE, and BLANT-PIANO. The total number of these pairs was 3265 for CM and 1993 for LSS.

The set of potential perseveration-source pairs for each participant unavoidably includes not only *true* perseveration-source pairs (i.e., pairs associating a true perseveration with its true source), but also what may be called *pseudo* perseveration-source pairs. In a pseudo pair the intruded letter is either not a perseveration from a prior response, or is not paired with its true source. Suppose, for example, that CM intruded an A into the response BLANT for reasons having nothing to do with the presence of As in prior responses. In this case BLANT-FARM, BLANT-ORANGE, and BLANT-PIANO would all be pseudo perseveration-source pairs. Suppose, alternatively, that the A in BLANT was a true perseveration, and that the only source for the perseveration was the immediately preceding response FARM. The pair BLANT-FARM would then be a true perseveration-source pair, but the pairs BLANT-ORANGE and BLANT-PIANO would be pseudo pairs.

Because we had no way of knowing which particular pairs were true perseveration-source pairs and which were pseudo pairs, we included all potential perseveration-source pairs in the analysis. Fortunately, however, we can determine from this mixture of true and pseudo pairs whether true perseverations maintain the position of the perseverated letter in the true source more often than expected by chance. For pseudo pairs, the letter intrusion in the potential perseveration response has nothing to do with the presence of the letter in the potential source, and therefore the intruded letter should maintain source position no more often than expected by chance. Consequently, if we find that position is maintained at an above-chance rate in the full set of potential perseveration-source pairs, we can conclude that true perseverations maintain true source position more often than expected by chance.

For each participant and for each discrete content-independent scheme a computer program calculated the proportion of potential perseveration-source pairs in which the intruded letter was in the same position in the perseveration and source responses. For example, Table 4 shows how the left- and right-edge representational schemes assign position to the A in the perseveration response BLANT and the source response FARM. When position is defined by the right-edge scheme, the A is in the same position in the perseveration and source responses. According to the left-edge scheme, however, the position of the A does not match in the perseveration and source responses. Hence, for BLANT-FARM the analysis program tallied a position match for the right-edge scheme but not for the left-edge scheme.

The program also estimated for each positional scheme the proportion of position matches expected by chance. The chance likelihood of an intruded letter having the same position in the perseveration response as in the source response can be estimated by examining source control responses. A source control response was defined as a response that had the same number of letters as the source response and contained the intruded letter, but was not produced in close proximity to the perseveration response. (As in the perseveration analysis, close proximity was defined to encompass the five preceding through five following responses.) For each perseveration-source pair, the analysis program generated a pool of source control responses. A Monte Carlo analysis used the source control responses to estimate for each representational scheme the proportion of position matches expected by chance. For each potential perseveration-source pair, the program randomly sampled a response from the source control pool, and tabulated for each positional scheme whether the perseverated letter was in

the same position in that response as in the perseveration response. For example, on one run of the analysis CAPERT was randomly chosen as the control for the source response ORANGE in the BLANT-ORANGE pair. According to both left-edge and right-edge schemes the A in the perseveration response (BLANT) and the A in the source control response (CAPERT) are in different positions (L+3 vs. L+2, respectively, for the left-edge scheme, and R-3 vs. R-5 for the right-edge scheme). The result of each run of the Monte Carlo analysis is a single estimate of proportion of perseverated letters expected by chance to match position in the perseveration and source responses. The entire process was carried out 10,000 times, yielding a distribution of chance proportions for each of the schemes.

The results are presented in Figure 7A (CM) and 7B (LSS). The black bars show the proportion of perseveration-source pairs in which the position of the intruded letter in the perseveration response matched the position of the letter in the actual source response. For example, the .51 proportion for the left-edge representational scheme observed for LSS reflects the finding that for 1013 of the 1993 perseveration-source pairs (51%), the intruded letter appeared in the same left-edge position in the perseveration and source responses. The white bars in the figure show the proportions expected by chance (more specifically, the mean chance proportion across the 10,000 runs of the Monte Carlo analysis).

For both CM and LSS, the observed proportion of positional matches between perseveration and source responses was greater than each of the 10,000 chance estimates for all five content-independent representation schemes ( $p < .0001$  for both participants for all schemes). These results demonstrate, for both participants, that letters were more likely than expected by chance to maintain the same content-independent position when perseverating from a source response to a perseveration response.

The results of Analysis 1a are robust over variation in criteria for inclusion of potential perseveration-source pairs. In the reported analyses we excluded potential perseveration responses containing more than one instance of an intruded letter, because of potential ambiguities concerning which instance(s) should be treated as intrusions. For example, in CM's error "umbrella" → UBLERRER the response includes two extra Rs, but it is not clear which of the Rs should be considered intruded letters, leading in turn to uncertainty about the positions in which intruded letters appeared. For similar reasons, source and control responses were limited to responses containing only one instance of the intruded letter. However, analyses that included responses with multiple instances of the intruded letter yielded the same pattern of results. We also conducted analyses in which each potential perseveration response was paired with only a single potential source (the most recent response in the perseveration window that contained the intruded letter). Once again the pattern of results was the same.

The finding that position was maintained at above-chance rates for all five content-independent schemes does not necessarily mean that all of these schemes contribute to the representation of letter position. The content-independent schemes agree with one another regarding which position in a source response matches the position of an intruded letter in a perseveration response for any pair in which the potential perseveration response has the same number of letters as the potential source response. Consider the pair RINT-REST, where RINT has an intruded R. The position of the R in RINT differs by scheme (e.g., L+1, C-2, R-4), but all of the content-independent schemes agree that the R occupies the same position in RINT (potential perseveration response) and REST (potential source response). Hence, if letter position were encoded according to any of the content-independent schemes one would expect all of the schemes to show above-chance proportions of positional matches. Additional analyses were therefore conducted to identify which of the content-independent positional schemes best accounted for the pattern of responses. Analysis 1b defined a measure of success



for candidate schemes, and Analysis 1c applied this measure in evaluating the various content-independent schemes.

**Analysis 1b: Proportion of Position Matches for True Perseveration-Source Pairs**—The observed proportions of position matches reported in Analysis 1a were calculated over all potential perseveration-source pairs. As previously noted, however, the pool of potential perseveration-source pairs includes not only true perseveration-source pairs (i.e., pairs consisting of a true perseveration and its true source), but also pseudo perseveration-source pairs (i.e., pairs in which the intruded letter is present in the potential source by chance).

In Analysis 1b we estimated the proportion of position matches for the true perseveration-source pairs alone. For each positional scheme the estimated proportion of position matches for true perseveration-source pairs (henceforth *pPMforTrPS*) provides a measure of how well the scheme performs in using the position of a perseverated letter in a perseveration response to predict the position of the corresponding source letter in the true source response. For example, a *pPMforTrPS* of .60 would indicate that in 60% of the true perseveration-source pairs the scheme in question correctly predicted the position of the source letter from the position of the perseverated letter. As we shall see, a scheme's *pPMforTrPS* value provides a basis for assessing that scheme's contribution to letter position encoding.

To estimate *pPMforTrPS* values, we first estimate the numbers of true and pseudo perseveration-source pairs in the pool of potential perseveration-source pairs. Next, we estimate for each scheme, the number of position matches for true perseveration-source pairs, and then convert these values to proportions.

The perseveration analysis provided three relevant values to estimate the number of true perseveration-source pairs: (1) *IP*, the total number of intrusion-prior response pairs, (2) *PotentialPS*, the number of potential perseveration-source pairs (i.e., the number of intrusion-prior response pairs in which the prior response contained the intruded letter), and (3) *pChanceILinP*, the estimated probability of the intruded letter appearing in the prior response by chance.<sup>4</sup>

*PotentialPS* is the sum of *TruePS* and *PseudoPS*, the numbers of true and pseudo perseveration-source pairs, respectively:

$$PotentialPS = TruePS + PseudoPS \quad (1)$$

*TruePS* and *PseudoPS* are both unknowns. However, we can substitute another expression for *PseudoPS* and thereby eliminate this unknown. A pseudo perseveration-source pair is any intrusion-prior response pair that is not a true-perseveration-source pair, but includes the intruded letter in the prior response by chance. Accordingly, the expected number of pseudo pairs is the number of intrusion-prior response pairs that are not true perseveration-source pairs (*IP* - *TruePS*) times the probability of the intruded letter being present in a prior response by chance:

---

<sup>4</sup>This perseveration analysis was the same as the analysis reported in the Letter Perseveration Analysis section, except that (a) for each intruded letter, the analysis considered all and only the prior responses within the perseveration window (4 preceding responses for CM and 2 for LSS), and (b) responses in which the intruded letter appeared more than once were excluded as intrusion, source and source control responses.

$$PseudoPS = (IP - TruePS) \times pChanceILinP \quad (2)$$

Substituting this expression for PseudoPS into equation (1) we obtain

$$PotentialPS = TruePS + (IP - TruePS) \times pChanceILinP \quad (3)$$

Inserting LSS's values for PotentialPS, IP, and ChanceILinP yields

$$1993 = TruePS + (3375 - TruePS) \times .224$$

Solving for TruePS gives us 1595 as the estimated number of true perseveration-source pairs. From this value we straightforwardly obtain 398 as the estimate for PseudoPS, the number of pseudo perseveration-source pairs. Performing the same calculations for CM yields estimates of 1392 for TruePS and 1873 for PseudoPS.

Using these values we can estimate the number of position matches for true perseverations ( $PsnMtchforTruePS$ ) for any given positional scheme. We first express the number of observed position matches for the full set of potential perseveration-source pairs in Analysis 1a ( $PsnMtchforPotentialPS$ ) as a sum of position matches for true perseveration-source pairs and position matches for pseudo perseveration-source pairs:

$$PsnMtchforPotentialPS = PsnMtchforTruePS + PsnMtchforPseudoPS \quad (4)$$

For pseudo perseveration-source pairs position matches will occur only by chance. Hence, the number of position matches for the pseudo pairs should equal the number of pseudo pairs ( $PseudoPS$ ) times the probability of a position match by chance ( $pChancePsnMtch$ ):

$$PsnMtchforPseudoPS = PseudoPS \times pChancePsnMtch \quad (5)$$

Substituting this expression into equation (4) and solving for  $PsnMtchforTruePS$  yields

$$PsnMtchforTruePS = PsnMtchforPotentialPS - PseudoPS \times pChancePsnMtch \quad (6)$$

We have values for all of the variables on the right side of the equation. The  $PsnMtchforPotentialPS$  and  $pChancePsnMtch$  values for each discrete content-independent position scheme are available from Analysis 1a; For any given scheme we have calculated the observed number of position matches ( $PsnMtchforPotentialPS$ ) for that scheme, and the chance proportion of a position match ( $pChancePsnMtch$ ). The value for PseudoPS does not depend upon the positional scheme.

Consider, for example, LSS's results for the left-edge scheme. The observed number of position matches for this scheme across the full set of potential perseveration-source pairs was 1013, and the chance proportion of a position match was .252. Given the estimate of 398 for the

number of pseudo perseveration-source pairs, we can estimate the number of position matches for true perseveration-source pairs as follows:

$$PsnMtchforTruePS=1013 - 398 \times .252=913$$

Finally, we can convert the number of position matches for true perseverations to a proportion by dividing by the estimated number of true perseverations:

$$pPMforTrPS=PsnMtchforTruePS/TruePS \quad (7)$$

For LSS the estimated number of true perseverations (TruePS) was 1595, and so his estimated proportion of position matches for true perseverations was 913/1595, or .57. In other words we estimate that in 57% of LSS's true perseverations the perseverated letter occupied the same left-edge position in the perseveration response as in the true source response.

Table 5 presents for each participant the estimated pPMforTrPS value for each of the discrete content-independent schemes. For both CM and LSS pPMforTrPS was highest for the both-edges scheme. In more than three-fourths of the true perseveration-source pairs (78% for CM and 87% for LSS) this scheme was successful in using the position of the perseverated letter in the perseveration response to predict the position of the source letter in the source response.

Results from the Monte Carlo procedure in Analysis 1a provide a basis for assessing the statistical reliability of differences among schemes in pPMforTrPS. For each positional scheme each Monte Carlo run provided a different estimate of the chance probability of a position match (pChancePsnMtch). Using these pChancePsnMtch estimates, we estimated pPMforTrPS for each scheme on each of the 10,000 Monte Carlo runs. For both CM and LSS, the estimated pPMforTrPS was highest for the both-edges scheme on all 10,000 runs; hence, for each participant the difference between the both-edges scheme and each of the other schemes was significant at  $p < .0001$ .

These results would seem to indicate that within the class of discrete content-independent schemes, the both-edges scheme most closely approximates the actual scheme for representing letter position in graphemic spelling representations. However, before drawing this conclusion we must address an issue arising from the nature of the both-edges scheme.

**Analysis 1c: The Both-Edges Scheme and its Components**—The left-edge, right-edge, center, and closer-edge schemes are simple schemes; each assigns only position representation to each letter in a word. However, the both-edges scheme is a combination of left-edge and right-edge schemes, and so assigns both left- and right-edge position representations to each letter (e.g., L+1 and R-4 for the F in FACE). We refer to such schemes as compound schemes. In evaluating compound schemes one must ask whether each component scheme (e.g., the left-edge and right-edge components for the both-edges scheme) plays a role in the success of the overall scheme.

For each potential perseveration-source pair, a simple scheme predicts at most one position in the source response where the intruded letter should be found.<sup>5</sup> Consider our example pair ERGE-FRENCE. According to the left-edge scheme, the intruded R occupies position L+2 in ERGE, and therefore the scheme predicts that the R in the source response will be found at position L+2. For the source response FRENCE the R does in fact occupy position L+2, and so a position match would be tallied for the left-edge scheme.

Compound schemes can make multiple predictions about what positions in a source response count as position matches. For the ERGE-FRENCE pair, the both-edges scheme assigns position representations L+2 and R-3 to the intruded R in ERGE. As a consequence, the scheme predicts that the R will be found in either position L+2 or position R-3 of the source response FRENCE, and a position match would be tallied if the R occupied either of these source positions.

On average the number of predicted source positions is greater for the both-edges scheme than for either the left- or right-edge scheme. As a consequence, the both-edges scheme has more opportunities for tallying position matches—that is, more opportunities for finding the intruded letter at a predicted source position. This point has implications for interpreting the finding of higher pPMforTrPS values for the both-edges scheme than for the component left- and right-edge schemes. Consider the comparison between both-edges and right-edge schemes. For all potential perseveration-source pairs the both-edges scheme predicts the same source position as the right-edge scheme (e.g., the R-3 position for ERGE-FRENCE). However, for many pairs the both-edges scheme also predicts an additional source position on the basis of left-edge coding (e.g., the L+2 position in FRENCE). Now suppose that the left-edge scheme is not implicated in letter position encoding, and that as a consequence the additional predicted left-edge positions are completely uncorrelated with the source positions predicted by the scheme that actually encodes letter position in graphemic spelling representations. In other words, suppose that the additional source positions predicted by the left-edge scheme are random relative to the positions predicted by the actual scheme for letter position representation. Even a random guess will occasionally be correct, and so we expect the additional left-edge predictions to yield occasional position matches for true perseveration-source pairs (assuming that the right-edge scheme does not always predict source position correctly for true perseveration-source pairs). As a consequence we expect the proportion of position matches for true perseveration-source pairs to be higher for the both-edges scheme than for the right-edge scheme, even if the left-edge scheme makes no systematic (i.e., non-random) contribution to predicting the source positions of perseverated letters. Hence, the finding that pPMforTrPS was higher for the both-edges scheme than for the right-edge scheme does not necessarily mean that the left-edge component of the both-edges scheme makes any systematic contribution over and above that of the right-edge scheme in predicting the source position of perseverated letters.

To determine whether the left-edge component of the both-edges scheme makes a systematic contribution beyond that of the right-edge component we compare the additional predictions made by the left-edge scheme with an equal number of additional predictions generated randomly. That is, we can compare a right-edge+left-edge (i.e., both-edges) scheme to a right-edge+random scheme that makes the same number of source position predictions, and so has the same number of opportunities for position matches. If the proportion of position matches for true perseverations (pPMforTrPS) is significantly higher for the right-edge+left-edge scheme than for the right-edge+random scheme, we can conclude that the left-edge component of the both-edges scheme makes a systematic contribution over and above that of the right-edge component in predicting the source position of perseverated letters. Similarly, we can assess the contribution of the right-edge component by comparing the both-edges scheme to a left-edge+random scheme.

The methods for these analyses may be explained by considering the both-edges vs. right-edge +random comparison. For potential perseveration-source pairs in which the perseveration and

---

<sup>5</sup>We say *at most* one position because occasionally the position of the intruded letter in the perseveration response has no counterpart in the source response. In the pair DIAN-FUN, the intruded N in DIAN occupies the left-edge position L+4, but FUN does not include a position L+4. Therefore, according to the left-edge scheme, no position in the source response is the same as the position of the intruded letter in the perseveration response.

source responses are of the same length, the right-edge and both-edges schemes do not differ in number of predicted source positions. For these pairs, no random source predictions were generated for the right-edge+random scheme. However, when the perseveration and source responses have different numbers of letters, the both-edges scheme usually predicts two source positions, whereas the right-edge scheme usually predicts only one. For every pair in which the both-edges scheme predicted an additional source position, the source predictions of the right-edge+random scheme consisted of the right-edge prediction plus an additional source position selected at random by the analysis program. For example, in the case of the ERGE-FRENCE pair, the right-edge scheme designates one predicted position in the source FRENCE (i.e., the R-3 position, occupied by the letter N). However, the both-edges scheme designates not only this position but also the L+2 position (occupied by the R in FRENCE). Accordingly, an additional position was selected randomly for the right-edge plus random scheme; the chosen position could be the same as, or different from, the left-edge position predicted by the both-edges scheme. Through this method the number of predicted source positions, and hence the number of opportunities for position matches, were equated on a pair-by-pair basis across the both-edges and right-edge+random schemes. On each of 10,000 Monte Carlo runs, a new set of random source predictions was generated.

Using the procedures described for Analyses 1a and 1b, proportions of position matches for true perseverations were estimated on each Monte Carlo run for the both-edges and right-edge+random schemes. The results were straightforward: For both CM and LSS the pPMforTrPS estimate was higher for the both-edges scheme than for the right-edge+random scheme on all 10,000 runs of the Monte Carlo analysis. For LSS the mean pPMforTrPS was .87 for both-edges but only .74 for right-edge plus random; for CM the values were .78 and .66 ( $p < .0001$  for both participants). These results demonstrate that the left-edge component of the both-edges scheme made a systematic contribution to the scheme's success, over and above the contribution of the right-edge component.

A similar analysis demonstrated that the converse was also true: The right-edge component made a systematic contribution beyond that of the left-edge component to the success of the both-edges scheme. This analysis found that that pPMforTrPS values for both CM and LSS were significantly higher for the both-edges scheme than for a left-edge plus random scheme (see Table 6),  $ps < .0001$ .

The both-edges scheme can also be decomposed into closer-edge and farther-edge schemes. As shown in Table 6 the pPMforTrPS values were significantly higher for the both-edges scheme than for a closer-edge+random scheme,  $ps < .0001$ , indicating that farther-edge coding (i.e., left-edge coding for letters in the right half of a word and vice versa) contributed systematically to the success of the both-edges scheme. Not surprisingly, a both-edges vs. farther-edge+random comparison revealed that the closer-edge component also played a significant role over and above that of the farther-edge component,  $ps < .0001$ .

These analyses demonstrate that the success of the compound both-edges scheme relative to its component left-, right-, closer- and farther-edge schemes cannot be explained by assuming that one of the scheme's components (e.g., the right-edge component) does all of the systematic work in predicting source positions of perseverated letters, with the complementary component (e.g., the left-edge component) contributing only in the sense of making essentially random source position predictions that occasionally happen to yield position matches for true perseverations. Rather, each component (left- and right-edge components given one decomposition of the scheme, and closer- and farther-edge components given an alternative decomposition) plays a systematic and significant role in the success of the both-edges scheme.



## Position Analysis 2: Graded vs. Discrete Position Representations

Analysis 1 revealed that for most of the true perseverations made by each participant (78% CM and 87% for LSS) the perseverated letter appeared at a source location that exactly matched the left-edge or right-edge position of the letter in the perseveration response. However, for the remaining true perseverations (22% for CM and 13% for LSS) the discrete both-edges scheme did not correctly predict the source location of the perseverated letter. Consider the pair ZEROBA-FLOWER, and assume that this pair is a true perseveration-source pair in which the intruded O in ZEROBA (a misspelling of ZEBRA) is a perseveration from the source response FLOWER. In ZEROBA the O occupies positions L+4 and R-3, and so the discrete both-edges scheme predicts that the O will appear in position L+4 or R-3 in the source FLOWER. In fact, the O in FLOWER occupies positions L+3 and R-4. Why did the perseveration and source positions not match?

**Analysis 2a: Graded vs. Discrete Both-Edges Schemes**—One possibility is that the both-edges representations are graded rather than discrete. Considering first the left-edge component of the both-edges scheme, the O in FLOWER might be represented by activating the O unit strongly in the L+4 pool, moderately in the L+3 and L+5 pools, and perhaps more weakly in more distant positions. As a consequence, the O would be most likely to perseverate into position L+4 of a subsequent response, but might also perseverate into an adjacent position (L+3 or L+5) or, rarely, into a more distant position. The same point applies to the right-edge component of the both-edges scheme: The O in FLOWER would be represented by activating the O unit not only in the R-3 pool, but also (less strongly) in the R-2 and R-4 pools, and so forth. Accordingly, the O would be most likely to perseverate into position R-3 of a subsequent response, but might also perseverate into positions R-2 or R-4 (or perhaps even a more distant position).

By the same logic, when we use the position of a perseverated letter in a perseveration response to predict the position of the source letter in the source response (as in our analyses), we expect sometimes to find the letter not at an exactly matching source position, but instead at a nearby position. Consider once again the pair ZEROBA-FLOWER. Given a graded both-edges scheme, we expect most often to find the O at one of the exactly matching source positions (i.e., L+4 or R-3); however, we also expect sometimes to find the O instead at one of the nearby source positions (e.g., L+3, L+5, R-2, or R-4).

In the present analysis we evaluated a version of graded both-edges scheme that defined a predicted source position to be either an exactly matching position, or an immediately adjacent position. For the ZEROBA-FLOWER pair, the graded scheme predicted that the O would be found at position L+3, L+4, L+5, R-2, R-3, or R-4. In this example, the left-edge and right-edge positions coincide, and therefore the graded both-edges scheme identifies three predicted source positions (those occupied by the O, W, and E). Because the O appears at one of these positions, the graded scheme scores a position match for the ZEROBA-FLOWER pair.

The graded both-edges scheme may be viewed as a combination of the discrete both-edges scheme (which predicts an exact match between perseveration and source positions) and what we will call an adjacent both-edges scheme (which predicts that the source position will be adjacent to an exact-match position). Therefore, the question we need to ask is whether the addition of the adjacent both-edges component makes a systematic contribution over and above that of the discrete component in predicting the source positions of perseverated letters.

To answer this question we compared the graded both-edges scheme to a discrete both-edges +random scheme that was matched to the graded scheme in number of opportunities for position matches. The results reveal that for both CM and LSS the proportion of position matches for true perseverations (pPMforTrPS) was significantly higher for the graded scheme

than for the discrete+random scheme ( $p < .001$  for both participants). This result demonstrates that the adjacent component of the graded both-edges scheme was significantly better than a random position selection process at predicting the source positions of perseverated letters, and hence that the adjacent component contributed systematically to the success of the graded scheme. Additional analyses confirmed that even within the context of a graded positional scheme, the various subcomponents of the both-edges scheme contributed systematically to its success: pPMforTrPS was significantly higher for the graded both-edges scheme than for graded left-edge plus random, graded right-edge plus random, graded closer-edge plus random, and graded farther-edge plus random schemes ( $p < .0001$  for all comparisons for each participant).

For both participants the pPMforTrPS for the graded scheme was remarkably high: .94 for CM and .97 for LSS. In other words, in almost all of each participant's true perseverations the perseverated letter appeared at a predicted position in the source response. The graded both-edges scheme did treat a large number of positions as predicted positions: a mean of 3.2 positions per word for CM and 3.1 for LSS, constituting about 60% (CM) and 55% (LSS) of the positions in a word. Nevertheless, it is impressive that the actual source position for nearly 100% of the true perseverations fell within the 55-60% of source positions picked out by the graded both-edges scheme.

**Analysis 2b: The Center Scheme**—The final issue to be addressed in our analyses of content-independent positional schemes concerns the center scheme. Analysis 1b showed much lower pPMforTrPS values for the discrete center scheme than for the discrete both-edges scheme. A similar analysis contrasting graded both-edges and graded center schemes yielded the same result: pPMforTrPS was significantly higher for the both-edges than for the center scheme ( $p < .0001$  for both participants). In conjunction with the evidence that each component of the both-edges scheme contributes to its success, these results indicate that the center scheme does not merit consideration as an alternative to the both-edges scheme. However, one can nevertheless ask whether the center scheme might conceivably play a role *in addition* to that of the both-edges scheme in coding letter position in graphemic spelling representations.

We addressed this question by comparing a graded both-edges+center scheme to a graded both-edges+random scheme. For both participants the pPMforTrPS value for the both-edges+center scheme was not significantly different from that of the both-edges+random scheme ( $p > .3$  for both participants), indicating that the graded center scheme made no systematic contribution beyond that of the graded both-edges scheme in predicting the source positions of perseverated letters. Given that the graded both-edges scheme alone predicted nearly all of the source positions, this result is entirely unsurprising.

We conclude that within the class of content-independent positional schemes, the graded both-edges scheme most closely approximates the scheme for encoding letter position in graphemic spelling representations. In Analyses 3 and 4 we examine letter-context and syllabic schemes, asking whether these schemes mediate letter-position coding instead of, or in addition to, the both-edges scheme.

### Position Analysis 3: Letter-Context Schemes

Letter-context schemes represent the position of a letter relative to other letters in the word. Analysis 3 considered three letter-context schemes. In the preceding-letter scheme the position of each letter is defined by the immediately-preceding letter. According to this scheme, the A in FACE occupies the F\_ position (i.e., the immediately-preceded-by-an-F position), and therefore could be represented by activating the A unit in an F\_ pool (see Figure 3A). The following-letter scheme similarly represents each letter relative to the immediately following letter. In this scheme the A in FACE occupies the \_C position. Finally, in the trigram scheme

the position of a letter is defined jointly by the immediately-preceding and immediately-following letters. In this scheme the A in FACE occupies the F\_C position.

Applying methods from previous analyses, Analysis 3 evaluated the letter-context schemes and compared these schemes with the graded both-edges scheme. The set of potential perseveration-source pairs was limited in these analyses in the following way: For each letter-context scheme we excluded an intruded letter from the position analysis if a context letter in the perseveration response was also an intruded letter. For example, in CM's error STUB → SLAB, we excluded the intruded A from the preceding-letter analysis because the context letter L was also an intruded letter. For the following-letter scheme we excluded intruded letters when the following letter was also an intruded letter, and for the trigram analyses we excluded an intruded letter when either the preceding or following letter was also an intrusion. Including these intrusions would lead to overestimation of the pPMforTrPS value. When two letters intrude into adjacent positions, these letters may often be perseveration from adjacent positions in the source. For example, CM wrote PLAY correctly on the trial immediately preceding the STUB → SLAB error, and hence the intruded L and A in SLAB may both have been perseverated from PLAY. If two adjacent intrusions are both perseverations from adjacent positions in the source, the second intrusion will *always* maintain preceding letter position (the A is in the position L\_ in both SLAB and PLAY), even if the preceding letter scheme is completely unrelated to the brain's scheme for representing letter position in graphemic spelling representations. To obtain unbiased estimates of pPMforTrPS for the preceding letter scheme we must exclude all but the first intruded letter in adjacent-letter intrusions<sup>6</sup>. By the same logic, in analyses of the following-letter scheme, we must exclude all but the last intruded letter in adjacent-letter intrusions, and in trigram-scheme analyses we must exclude all of the letters in adjacent-letter intrusions. Given this restriction, for CM and LSS, respectively, the number of potential perseveration-source pairs was: 1679 and 699 for preceding letter, 925 and 455 for following letter, and 600 and 207 for the trigram scheme.

**Analysis 3a: pPMforTrPS for Letter-Context and Both-Edges Schemes**—In this analysis we estimated pPMforTrPS values for each of the letter-context schemes, and compared these values to values obtained by applying the graded both-edges scheme to the same data sets. For example, results for the preceding-letter scheme were compared with both-edges results obtained from the set of potential perseveration-source pairs used in the preceding-letter analysis.

The results are presented in Table 7. The pPMforTrPS values for the letter-context schemes were generally quite low, and in all cases were far lower than the values for the graded both-edges scheme ( $p < .0001$  for all comparisons). These results demonstrate clearly that none of the letter-context schemes is a viable alternative to the graded both-edges scheme. Nevertheless, letter-context schemes might conceivably make some contribution, in addition to that of the both-edges scheme, to coding of letter position in graphemic spelling representations. Analysis 3b explored this possibility.

**Analysis 3b: Both-Edges+Letter-Context Schemes**—This analysis applied the method used to evaluate potential contributions of the center scheme in Analysis 2b. We defined three graded both-edges+letter-context schemes, one for each of the three letter-context schemes (e.g., graded both-edges+preceding-letter), and compared each combined scheme to a graded both-edges+random scheme that made the same number of source predictions. The

---

<sup>6</sup>We cannot resolve this problem by excluding only the cases in which the adjacent intruded letters are present in adjacent source positions (as in SLAB-PLAY), because this strategy could lead to *underestimation* of pPMforTrPS values. Briefly, if the context letters in adjacent-to-adjacent intrusions (e.g., the L in SLAB) are not always true perseverations, the strategy would exclude some position matches that should properly be credited to the preceding-letter scheme, without also excluding any position mismatches.

results, presented in Table 8, are extremely clear: The pPMforTrPS values for the graded both-edges plus letter-context schemes were in all cases indistinguishable from the values for the matched graded both-edges plus random schemes ( $p > .25$  for all comparisons). This finding demonstrates that none of the letter-context schemes made a systematic contribution beyond that of the graded both-edges scheme in predicting source locations of perseverated letters. In fact the pPMforTrPS values in Table 8 were little if at all higher than the values shown in Table 7 for the both-edges scheme alone. This result reflects the fact that the context schemes made very few source predictions that were not also made by the both-edges scheme, and therefore were unlikely to achieve additional position matches even by chance.

The results from Analysis 3 make a strong case that simple bigram and trigram schemes (e.g., Brown & Loosemore, 1994; Seidenberg & McClelland, 1989) do not play a role in coding letter position in graphemic spelling representations. Furthermore, the results argue against open bigram schemes (e.g., Grainger & van Heuven, 2003; Whitney, 2001) for graphemic spelling representations. Any plausible open bigram scheme includes at least one of the simple bigram schemes as a key subcomponent. For example, an open-bigram preceding-letter scheme (in which, for example, the C in FACE has position representations F\_ and A\_) includes the simple preceding-letter scheme. Hence, if an open-bigram scheme played a role representing letter position in spelling representations, we should have observed systematic contributions of one or both of the simple bigram schemes. No hint of such contributions was observed.

#### Position Analysis 4: Syllabic Representations

In this analysis we evaluated syllabic positional schemes, and compared these schemes with the graded both-edges scheme. A computer program syllabified each participant's responses according to the principle of orthographic onset maximization, and assigned each letter a position defined jointly by the position of the syllable in which the letter appeared, and the role of the letter within that syllable. The position of a syllable within a word was defined by a both-edges scheme in which the syllable's position was encoded in terms of its distance (in syllables) from both the left and the right edges of the word. In the word HYDRANT, for example, DRANT is the second syllable from the left (Syllable L+2) and the first syllable from the right (Syllable R-1). Other methods for specifying the position of the syllable in the word were also considered – left-edge, right-edge, closer-edge, and center. For both CM and LSS, the best-performing syllabic schemes were those using both-edges syllable position coding, and hence we report results for these schemes.

Within each syllable the position of each letter was defined by its role as an onset, nucleus, or coda letter. In the syllable DRANT, for instance, D and R make up the onset, A the nucleus, and N and T the coda. Finally, letters were differentiated by their positions within onsets, nuclei, and codas. The possible within-role positions were first, second, and third onset positions (Onset 1 - Onset 3), first and second nucleus positions (Nucleus 1 and Nucleus 2), and first through fourth coda positions (Coda 1 - Coda 4). Onset and nucleus letters were assigned to positions in a left-justified manner whereas coda letters were right-justified. In DRANT the D and R appear in the Onset 1 and Onset 2 positions, respectively, the A is in the Nucleus 1 position, and the N and T occupy, respectively, the Coda 3 and Coda 4 positions. The syllabic scheme as a whole therefore assigns the following two position representations to the D in DRANT: Syllable L+2: Onset 1 and Syllable R-1: Onset 1.

We considered both discrete and graded syllabic schemes. In the discrete scheme the coding of position within syllable role (e.g., Onset 1, Onset 2) was assumed to be exact, and hence a position match was tallied only when the position of the intruded letter in the perseveration response exactly matched the position of that letter in the source response. Consider the pair LAMP-BLOCK, in which the perseveration response LAMP is a misspelling of DAMP and so contains an intruded L. In LAMP the L is in an Onset 1 position, and so has syllabic position

representations Syllable L+1: Onset 1 and Syllable R-1: Onset 1. In the potential source response BLOCK, however, the L occupies an Onset 2 position, and so is coded as Syllable L+1: Onset 2 and Syllable R-1: Onset 2. Onset 1 and Onset 2 positions are not the same, and consequently no position match would be tallied for this pair.

For the graded syllabic scheme we assumed that coding of position within syllable role was graded. In this scheme, the L in BLOCK might be represented by activating the L unit strongly in the Onset 2 pool for the appropriate syllable, and somewhat less strongly in the Onset 1 and Onset 3 pools. Accordingly, we tallied a position match for the graded syllabic scheme whenever the position of the intruded letter in the perseveration response matched its position in the source response with respect to syllable (e.g., Syllable L+1) and syllable role (e.g., Onset), regardless of whether the position-within-role matched exactly. Therefore, the graded syllabic scheme, unlike the discrete scheme, scored a position match for LAMP-BLOCK, even though the two responses do not exactly match with respect to the L's position within syllable role.

We excluded from Analysis 4 all responses that could not be parsed into orthographically acceptable syllables (e.g., DRNDY). For CM and LSS, respectively, the syllabic analyses included 2868 and 1641 potential perseveration-source pairs, with 1191 and 1309 of these estimated to be true perseveration-source pairs.

**Analysis 4a: Discrete vs. Graded Syllabic Schemes**—In Analysis 4a we first estimated the proportion of position matches for true perseverations for the discrete and graded syllabic schemes. As shown in the first two rows of Table 9, the graded scheme yielded substantially higher pPMforTrPS values for both CM and LSS ( $p < .0001$ ). However, no firm conclusions can be drawn from this result, because the discrete syllabic scheme is a subcomponent of the graded scheme. To choose between the schemes we need to know whether the additional source predictions made by the graded scheme contribute systematically to its success. Accordingly, we compared the graded syllabic scheme with a discrete syllabic +random scheme that made the same number of source predictions. The results for this scheme are presented in the third row of Table 9. For both CM and LSS the estimated pPMforTrPS value was significantly higher for the graded scheme than for the discrete+random scheme,  $p < .0001$ . These results demonstrate that extending source predictions from exactly matching within-syllable-role position (e.g., Onset 1) to the entire syllable role (e.g., the entire onset) systematically improves the ability of a syllabic scheme to predict the source locations of perseverated letters.

The high pPMforTrPS values for the graded syllabic scheme are perhaps not too surprising, given that this scheme is well-correlated with the graded both-edges scheme (e.g., onsets are usually near the left edge of a word, and codas are usually near the right edge). In Analysis 4b we attempted to tease apart these two schemes.

**Analysis 4b: Graded Syllabic vs. Graded Both-Edges Schemes**—We first compared the pPMforTrPS values for the graded syllabic scheme with values obtained by applying the graded both-edges scheme to the data set used in the syllabic-scheme analyses. As shown in Table 10 the estimated pPMforTrPS for both participants was higher for the graded both-edges scheme: This scheme successfully predicted the source location of the perseverated letter in 94% and 98% of the true perseveration-source pairs for CM and LSS, respectively, whereas the graded syllabic scheme succeeded only 91% of the time for each participant. The differences between both-edges and syllabic schemes were highly reliable ( $p < .0001$  for both participants). Given that the graded syllabic scheme is not a subcomponent of the graded both-edges scheme, this result implies that the latter is a closer approximation than the former to the brain's scheme for coding letter position in graphemic spelling representations.



However, the graded syllabic scheme might nevertheless play some role in addition to that of the graded both-edges scheme in coding letter position. To evaluate this possibility we defined a graded both-edges+graded syllabic scheme, and compared this scheme to a graded both-edges+ random scheme. The results are presented in Table 11. For LSS the pPMforTrPS for the graded both-edges plus graded syllabic scheme (.97) was indistinguishable from that for the graded both-edges plus random scheme (.98), and in fact from that of the graded both-edges scheme alone (.98),  $ps > .25$ . These results indicate that the graded syllabic scheme made no systematic contribution beyond that of the graded both-edges scheme to the success of the graded both-edges+graded syllabic scheme.

For CM the results were not quite as clear. Adding the graded syllabic scheme to the graded both-edges scheme increased the pPMforTrPS from .94 to .97, whereas adding random source predictions yielded a slightly smaller increase (from .94 to .95). The difference between the .97 value for the both-edges plus syllabic scheme and the .95 value for the both-edges plus random scheme was significant only at the .05 level (as opposed to the .0001 level observed for almost all of our significant effects). Thus, for CM (but not for LSS) the graded syllabic scheme may make a small systematic contribution to the success of the graded both-edges +graded syllabic scheme.

What conclusions should we draw from these results? Given that both CM and LSS showed higher pPMforTrPS values for the graded both-edges scheme than for the graded syllabic scheme, the both-edges scheme should clearly be preferred to the syllabic scheme as a candidate for the scheme that underlies encoding of letter position in graphemic spelling representations. Furthermore, given that (a) the syllabic scheme clearly made no contribution beyond that of the both-edges scheme for LSS, (b) the syllabic scheme's contribution was small and uncertain for CM, and (c) with the sole exception of the both-edges plus syllabic versus both-edges plus random comparison CM and LSS presented with identical patterns of significant effects across all of our analyses, the most parsimonious conclusion is that syllabic position representations do not contribute to letter position coding in graphemic spelling representations. This conclusion should, however, be considered tentative.

## General Discussion

The aim of the present study was to identify the cognitive scheme for encoding letter position in the orthographic representations that underlie the ability to spell. We considered more than ten candidate schemes falling into three categories: content-independent, letter-context, and syllabic. The candidate schemes also varied regarding whether the postulated position representations were discrete or graded. We evaluated the candidate positional schemes through analyses of letter perseveration errors produced by two dysgraphic participants, CM and LSS. Results from four sets of analyses strongly supported a content-independent, both-edges scheme that represents letter position in terms of distance and direction from both the left and right edges of a word. For example, according to this scheme the position of the N in PENCIL is represented jointly as L+3 (3 letter positions to the right of the left edge of the word) and R-4 (four positions to the left of the right edge). The analyses also provided evidence that the both-edges position representations are graded rather than discrete. In accounting for the results from both CM and LSS, the graded both-edges scheme clearly outperformed the alternative schemes, including the other content-independent schemes, the letter context schemes (including simple bigram, open bigram and trigram schemes), and the syllabic schemes. Furthermore, analyses examining subcomponents of the both-edges scheme (e.g., left-edge and right-edge, closer-edge and farther-edge) demonstrated that each subcomponent contributed substantively to the success of the scheme. Finally, compound schemes that combined alternative schemes with the graded both-edges scheme were not systematically better than the graded both-edges scheme alone. Across all of the analyses CM and LSS showed

remarkably similar patterns of results, and this concordance between participants strengthens our conclusion in favor of the graded both-edges scheme.

### Levels of Representation

We have argued that at a level of graphemic spelling representation, position is represented by a graded-both edges scheme. We motivated this claim by showing that the perseveration errors produced by our participants arose at a graphemic level of representation, and maintained position defined by the graded both-edges scheme.

According to the spelling theory outlined in the Introduction, the level of graphemic spelling representations receives input from both the lexical and the sublexical routes (Folk & Rapp, 2004; Rapp, Epstein & Tainturier, 2002). Lexical input comes from an orthographic output lexicon that contains a node for each orthographic lexeme. The lexeme nodes are connected to nodes corresponding to letters at a graphemic level of representation. By this account, knowledge about the position of the C in the word CAT is encoded by connections between the node that represents [CAT] at the level of the orthographic lexicon, and nodes representing the letter C in positions [L+1] and [R-3] at the level of graphemic spelling representations. The nodes representing C in positions [L+1] and [R-3] could also be activated via the sublexical route, by processes that determine the likely sequence of graphemes given a sequence of phonemes. The graphemic level of representation serves as input to a serial order system that produces letters, one at a time, in the correct order (Caramazza, Miceli & Villa, 1987).

Within this framework, our interpretation of LSS's and CM's deficit is straightforward. We assume that on some trials grapheme representations from preceding responses retained sufficient activation to compete successfully with graphemes activated for the current target. This state of affairs could have arisen because graphemes from preceding trials sometimes persisted abnormally in an activated state and/or because the activation of the current target's graphemes was sometimes abnormally weak, perhaps due to sub-normal activation of the target orthographic lexeme (see Cohen & Dehaene, 1998, for further discussion). The grapheme representations are position specific, with the consequence that the position of a perseverated letter in a source response is maintained, at least approximately, in the perseveration response. Because LSS's and CM's perseveration errors maintained position relative to both the left and the right edge of the word but in an "approximate" manner, we conclude that graphemic spelling representations use a graded both-edges scheme to encode position.

Our account assumes a single positional scheme in the graphemic representations underlying our ability to spell. Conceivably, however, there may be multiple levels of graphemic spelling representation that each use a different position representation scheme (e.g., Houghton & Zorzi, 2003, see also Grainger & Holcomb, 2009, and Perry, Ziegler & Zorzi, 2007, for a similar claim in reading). For example, some researchers make a distinction between long-term and working memory graphemic spelling representations (e.g., Houghton & Zorzi, 2003) and could potentially posit different schemes for representing letter position at these different levels. Therefore, while we clearly can conclude that position is represented by a graded both-edges scheme at some level of graphemic spelling representation, we cannot rule out the possibility that other schemes are also involved in the spelling process.

### Prior Proposals Concerning Position Coding in Graphemic Spelling Representations

Graded both-edges position representations are posited in some competitive queuing (CQ) simulations of spelling (e.g., Glasspool & Houghton, 2005; Houghton, Glasspool & Shallice, 1994), and this choice of positional scheme plays a significant role in the performance of the simulations (e.g., in generating certain types of spelling error). For example, the posited both-edges position representation allows for successful simulation of the phenomenon, observed

in both normal (e.g., Wing & Baddeley, 1980) and impaired (e.g. Caramazza & Miceli, 1990) spellers, that letters near the beginning or end of a word are more likely to be produced correctly than letters near the middle. Whether the patterns of perseveration errors observed in CM and LSS can be captured within a competitive queuing framework remains to be seen; current CQ spelling simulations generate each spelling response without any influence of prior responses, and so cannot simulate perseverations. Nevertheless, our results support the CQ assumption of graded both-edges position coding.

The CQ both-edges representations have an interesting property: Left-edge coding plays a more important role for letters close to the left edge of a word than for letters farther to the right, whereas right-edge coding is more important for letters close to the right edge than for those farther to the left. This property may be elucidated by reference to our illustrative representational format. (see Figure 2). Suppose for the word FACE that the F unit is activated strongly in the L+1 pool, the A unit somewhat less strongly in the L+2 pool, the C unit rather weakly in the L+3 pool, and so on. Suppose further that the opposite pattern holds for right-edge representations, such that the E unit is activated strongly in the R-1 pool, whereas the F unit is activated weakly in the R-4 pool. Given this pattern of activation, the left-edge representations would play a dominant role in selection of the to-be-produced letter for positions close to the left edge, whereas right-edge representations would predominate for letters toward the right edge. We can use the perseverations made by CM and LSS to assess whether graphemic spelling representations do in fact have this property. If left-edge positions are more important than right-edge positions for the left half of a word, then perseverations into the left half of a response should be more likely to maintain left-edge than right-edge position. Similarly, if right-edge position is more important than left-edge position for the right half of a word, then perseverations into the right half of a response should be more likely to maintain right-edge than left-edge position. Results from both CM and LSS showed exactly this pattern. For errors in the left half of a response pPMforTrPS values were significantly higher for the left-edge scheme than for the right-edge scheme (.93 vs. .70, respectively, for CM; .93 vs. .83 for LSS),  $ps < .0001$ . For errors in the right half of a response, however, pPMforTrPS was significantly higher for the right- than for the left-edge scheme (.84 vs. .72, respectively, for CM; .95 vs. .86 for LSS),  $ps < .0001$ .

These findings may help to resolve an apparent conflict between our results and those reported by Caramazza and Hillis (1990) in an elegant study of NG, a woman with right-side hemispatial neglect. NG made spelling errors almost exclusively in the right halves of words, and the likelihood of error increased with (rightward) distance from the center of the word. On the basis of these results, and similar findings from NG's reading, Caramazza and Hillis concluded that letter position is coded relative to the center of the word in graphemic representations.

We suggest, however, that NG's performance is equally consistent with our conclusion in favor of both-edges positional coding, given the assumption that left-edge coding is more important the closer a letter is to the left word edge, whereas right-edge coding is more important the closer a letter to the right edge. Suppose that NG's right-side neglect affects her ability to represent the positions of letters relative to the right edge of the word, but not relative to the left edge. Suppose in particular that for each right-edge position the activation of the correct letter is abnormally weak, so that random noise often causes incorrect letters to be activated as strongly as, or more strongly than, the correct letter. In that case, for letters in the right half of a word, the impaired right-edge representations are likely to cause spelling errors. Consider, for example, production of the final letter in JURY. In an intact spelling system the Y unit would be strongly activated in the R-1 pool, and weakly activated in the L+4 pool. For NG, however, an incorrect unit (e.g., D) may be more strongly activated than the correct unit in the R-1 pool. Although the Y unit will be activated normally in the L+4 pool, the activation in this pool is weak, and may be insufficient to prevent selection of an incorrect letter on the basis of

activation in the R-1 pool. As a consequence, a spelling error (e.g., JURD) may well occur. Furthermore, likelihood of an error on a letter should increase with rightward distance from the center of the word, because the predominance of the (impaired) right-edge representations in letter selection should increase with rightward distance from the center. For letters in the left half of a word, however, the impaired right-edge representations should have little impact. Consider the first letter in JURY. In a normal spelling system the J unit would be strongly activated in the L+1 pool, and weakly activated in the R-4 pool. For NG an incorrect unit may have more activation than J in the R-4 pool, but this erroneous activation pattern should be outweighed by the strong activation of the J in the L+1 pool, and the correct letter should be produced. Given this account of NG's deficit, there is no conflict between our results and those reported by Caramazza and Hillis (1990). The findings from NG, as well as those from CM and LSS, are consistent with both-edges position representation, and do not require postulation of a center-based scheme.

Syllabic position representations have also been proposed in the context of spelling (Houghton & Zorzi, 2003). Our results certainly do not rule out the possibility that graphemic representations include syllabic information, but do suggest that letter position is not encoded by a syllabic scheme (see Perea & Carreiras, 2006 for a similar finding in reading). Finally, our results argue against trigram position representations, which were incorporated by Brown and Loosemore (1994) in their connectionist simulation of spelling. The trigram scheme was apparently chosen without specific empirical motivation, and the consequences of the choice for the simulation's performance are not entirely clear. In our analyses the trigram scheme performed very poorly, indicating that this scheme should not be adopted in future theories or simulations of spelling.

### Position Representation in Reading

The present study focused on spelling representations, and consequently the results do not have strong implications for the question of how letter position is represented in reading. For example, our arguments for a graded both-edges scheme and against bigram schemes in spelling would be problematic for bigram theories of letter position encoding in reading (e.g., Binary Open Bigram: Grainger & van Heuven, 2003; Overlapping Open Bigram: Grainger et al., 2006; SERIOL: Whitney, 2001, 2008) only given evidence that the same form of graphemic representation underlies both reading and spelling.

Caramazza and Hillis (1990) proposed a level of graphemic representation shared between reading and spelling (see also Caramazza, Capasso & Miceli., 1996; Tainturier & Rapp, 2003; Tsapkini & Rapp, in press; and Hillis & Rapp, 2004, for a review) based on the finding that the error pattern for their patient NG was very similar in spelling and reading. However, the issue remains far from settled, and consequently we make no claims about position representations in reading. In this context we also note that it remains to be seen whether the graded both-edges scheme we proposed for graphemic spelling representations could account for the wide range of experimental results that constrain theories of position representation in reading.

### Position Representations in Other Cognitive Domains

The question of whether the same representation scheme is used across domains can be asked more broadly as well. As we have noted, the representation of serial position is crucial not only in spelling and reading, but also in a broad range of other cognitive functions, including speaking, recalling sequences of items or events, processing numerical information, and navigating routes through an environment. This observation raises an important question: Does the graded both-edges scheme mediate position representation only in spelling, or might this

scheme be applied more generally across many, or even all, cognitive domains that implicate encoding of serial position?

The available evidence is sparse. In verbal working memory the Start-End model (Henson, 1998) posits graded both-edges representations in which the position of an item in a to-be-recalled list is represented relative to both the beginning and end of the list. The model successfully interprets a variety of results from studies of serial word list recall in neurologically-intact adults, including results involving list-to-list perseveration errors (e.g., Conrad, 1960; Henson, 1999 but cf. Farrell & Lelievre, 2009). Studies comparing the full range of positional schemes we considered in this article have not yet been carried out in research on list recall. Nevertheless, much of the available evidence is at the very least consistent with Henson's proposal that position information is represented by a graded both-edges scheme in the serial short-term memory domain.

Progress in identifying the positional scheme(s) that underlie performance in various cognitive domains may shed light on the issue of domain-general versus domain-specific serial order processing mechanisms. Results showing that different positional schemes mediate performance in different domains would suggest that serial-order mechanisms are domain-specific, whereas evidence that the same scheme is implicated in many cognitive domains would enhance the plausibility of a domain-general hypothesis. Although the available data are insufficient to warrant firm conclusions, the evidence pointing to graded both-edges representations in the disparate domains of spelling and serial short-term memory suggest domain-general serial order mechanisms or processing principles. The picture should become clearer as representation of serial position is studied systematically in other cognitive domains.

## Acknowledgments

This research was supported by NIH grants NS22201 and DC 006740. We thank Paul Smolensky, Manny Vindiola, Ariel Goldberg, Özge Gürcanlı and members of the JHU CogNeuro Lab for feedback and suggestions, Donna Aliminosa, Tony Pastor, Sumin Lee, Jenna Rowen and Julia Thorn for help with data analysis and testing, as well as two anonymous reviewers for their helpful suggestions. We especially thank LSS and CM for their cheerful participation; working with them was truly a pleasure.

## Appendix 1: General Language Assessment Results for CM and LSS

CM was administered the Boston Diagnostic Aphasia Examination (BDAE: Goodglass & Kaplan, 1983). On the auditory comprehension tasks he performed well: He made no errors in word discrimination or body part identification, and scored 14/15 in following spoken commands and 10/12 in responding to questions involving complex ideational material. CM also performed well in reading comprehension, obtaining high scores in symbol discrimination, word recognition, and word/picture matching. His comprehension score for reading sentences and paragraphs was 7/10.

In spoken language production CM showed impairment. He was able to recite automatized sequences and repeat single words, but scored only 2/16 in repeating phrases. The speech pathologist conducting the examination rated his speech at 3 (of a possible 7) for melodic line, and 4 for articulatory agility. Phrase length was two words. Confrontation and responsive naming were poor, and CM generated only 1 animal name in 90 seconds. He read aloud 3 of 10 words and no sentences.

LSS was administered the Western Aphasia Battery (WAB: Kertesz, 1982). He scored 53/60 on the auditory word recognition task, 54/60 in answering yes/no questions, and 48/80 in following sequential commands. LSS performed well on repeating single words and sentences, scoring 98/100. He was also relatively intact in spoken picture naming, correctly naming 17/20



pictures. He scored 8/10 on sentence completion and 8/10 on responsive speech. However, in sixty seconds he was unable to generate a single animal name. The speech pathologist rated his spontaneous speech 8 (of a possible 10) for information content and 6 (of a possible 10) for fluency, grammatical competence and paraphasias.

## Appendix 2: Performance of LSS and CM on the JHU Dysgraphia Battery

Task	# of Items	CM % Correct	LSS % Correct
<i>Writing to Dictation</i>			
<i>Lexical Status</i>			
Words	326	128 (39%)	63 (19%)
Nonwords	34	4 (12%)	0 (0%)
<i>Word Frequency</i>			
High Frequency	146	64 (44%)	36 (25%)
Low Frequency	146	48 (33%)	13 (9%)
<i>Grammatical Class</i>			
Nouns	28	10 (36%)	8 (29%)
Verbs	28	8 (29%)	1 (4%)
Adjectives	28	4 (14%)	3 (11%)
Functors	20	11 (55%)	8 (40%)
<i>Concreteness</i>			
Concrete	21	8 (38%)	11 (52%)
Abstract	21	6 (29%)	2 (10%)
<i>Phoneme-to-Grapheme Probability</i>			
High Probability	30	15 (50%)	5 (17%)
Low Probability	80	42 (53%)	15 (18%)
<i>Length</i>			
4-letters	14	6 (43%)	2 (14%)
5-letters	14	3 (21%)	3 (21%)
6-letters	14	6 (43%)	2 (14%)
7-letters	14	5 (36%)	1 (7%)
8-letters	14	3 (21%)	2 (14%)
<i>Written Picture Naming</i>			
All Words	51	13 (25%)	14 (27%)
<i>Copy Transcoding</i>			
<i>Direct</i>			
Words	84	84 (100%)	74 (88%)
Nonwords	40	40 (100%)	38 (95%)
<i>Delayed</i>			
Words	84	59 (70%)	50 (60%)
Nonwords	40	23 (58%)	19 (48%)

## References

Badecker W. Representational properties common to phonological and orthographic output systems. *Lingua* 1996;99:55–83.

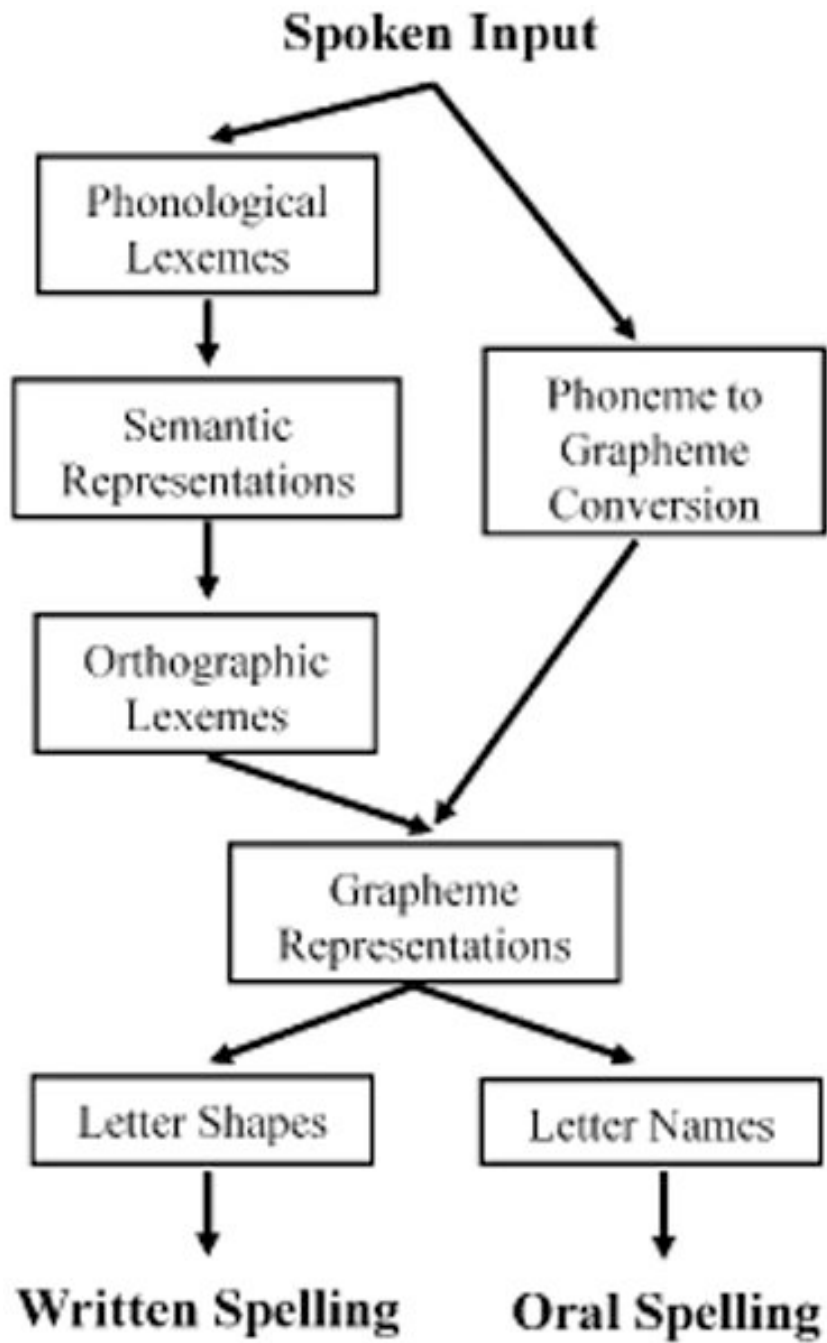
- Boomer DS, Laver JD. Slips of the tongue. *British Journal of Disordered Communication* 1968;3:2–12.
- Brown, GDA.; Loosemore, RPW. Computational approaches to normal and impaired spelling. In: Brown, GDA.; Ellis, NC., editors. *Handbook of spelling: Theory, process and application*. Chichester: John Wiley & Sons; 1994.
- Bullock D. Adaptive neural models of queuing and timing in fluent action. *Trends in Cognitive Science* 2004;8:426–33.
- Caramazza A. The logic of neuropsychological research and the problem of patient classification in aphasia. *Brain & Language* 1984;21:9–20. [PubMed: 6697172]
- Caramazza A. On drawing inferences about the structure of normal cognitive systems from the analysis of patterns of impaired performance: The case for single-patient studies. *Brain & Cognition* 1986;5:41–66. [PubMed: 3954906]
- Caramazza A. Is cognitive neuropsychology possible? *Journal of Cognitive Neuroscience* 1992;4:80–95.
- Caramazza A, Capasso R, Miceli G. The role of the graphemic buffer in reading. *Cognitive Neuropsychology* 1996;13:673–698.
- Caramazza A, Coltheart M. Cognitive Neuropsychology twenty years on. *Cognitive Neuropsychology* 2006;23:3–12.
- Caramazza A, Hillis AE. Levels of representation, co-ordinate frames, and unilateral neglect. *Cognitive Neuropsychology* 1990;7:391–445.
- Caramazza A, Miceli G. The structure of graphemic representations. *Cognition* 1990;37:243–297. [PubMed: 2282774]
- Caramazza A, Miceli G, Villa G. The role of the Graphemic Buffer in spelling: Evidence from a case of acquired dysgraphia. *Cognition* 1987;26:59–85. [PubMed: 3608396]
- Cohen L, Dehaene S. Neglect dyslexia for numbers? A case report. *Cognitive Neuropsychology* 1990;8:39–58.
- Cohen L, Dehaene S. Competition between past and present: Assessment and interpretation of verbal perseverations. *Brain* 1998;121:1641–1659. [PubMed: 9762954]
- Coltheart M, Curtis B, Atkins P, Haller M. Models of reading aloud: Dual-route and parallel-distributed processing approaches. *Psychological Review* 1993;100:589–608.
- Coltheart M, Rastle K, Perry C, Langdon R, Ziegler J. DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychological Review* 2001;108:204–256. [PubMed: 11212628]
- Conrad R. Serial order intrusions in immediate memory. *British Journal of Psychology* 1960;51:45–48. [PubMed: 13811577]
- Davis, CJ. Unpublished doctoral dissertation. University of New South Wales; Australia: 1999. The Self-Organising Lexical Acquisition and Recognition (SOLAR) model of visual word recognition.
- Davis CJ, Bowers JS. What do letter migration errors reveal about letter position coding in visual word recognition? *Journal of Experimental Psychology: Human Perception & Performance* 2004;30:923–941. [PubMed: 15462631]
- Davis CJ, Bowers JS. Contrasting five theories of letter position coding. *Journal of Experimental Psychology: Human Perception & Performance* 2006;32:668–687. [PubMed: 16822131]
- Dehaene S, Cohen L, Sigman M, Vinckier F. The neural code for written words: a proposal. *Trends in Cognitive Science* 2005;9:335–341.
- Dell GS, Svec W, Burger L. Language Production and Serial Order: A Functional Analysis and a Model. *Psychological Review* 1997;104:123–147. [PubMed: 9009882]
- Dunn, LM.; Dunn, LM. *Peabody Picture Vocabulary Test – Revised*. Circle Pines, MN: American Guidance Service; 1981.
- Ellis AW, Flude BM, Young AW. “Neglect Dyslexia” and early visual processing of letters in words and nonwords. *Cognitive Neuropsychology* 1987;4:439–464.
- Farrell S, Lelièvre A. End anchoring in short-term order memory. *Journal of Memory and Language* 2009;60:209–227.
- Folk J, Rapp B. Interaction of lexical and sublexical information in spelling : Evidence from nonword priming. *Applied Psycholinguistics* 2004;25:565–585.
- Glasspool, DW. Doctoral Thesis. University College London; 1998. *Modelling serial order in behaviour: Studies of spelling*.

- Glasspool DW, Houghton G. Serial order and consonant-vowel structure in a graphemic output buffer model. *Brain and Language* 2005;94:304–330. [PubMed: 16098380]
- Gomez P, Ratcliff R, Perea M. The overlap model: a model of letter position coding. *Psychological Review* 2008;115:577–600. [PubMed: 18729592]
- Goodglass, H.; Kaplan, E. *The Boston Diagnostic Aphasia Exam*. Philadelphia: Lea & Febiger; 1983.
- Goodman, RA.; Caramazza, A. *The Johns Hopkins University Dysgraphia Battery*. Baltimore, MD: Johns Hopkins University; 1985.
- Grainger J, Granier JP, Farioli F, Van Assche E, van Heuven W. Letter position information and printed word perception: The relative-position priming constraint. *Journal of Experimental Psychology: Human Perception and Performance* 2006;32:865–884. [PubMed: 16846285]
- Grainger J, Holcomb PJ. Watching the word go by: On the time-course of component processing in visual word recognition. *Language and Linguistics Compass* 2009;3:128–156. [PubMed: 19750025]
- Grainger, J.; van Heuven, W. Modeling Letter Position Coding in Printed Word Perception. In: Bonin, P., editor. *Mental lexicon: “Some words to talk about words”*. New York: Nova Science Publishers; 2003.
- Grainger J, Whitney C. Does the huamn mnid raed wrods as a wlohe? *Trends in Cognitive Science* 2004;8:58–59.
- Harm MW, Seidenberg MS. Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review* 1999;106:491–528. [PubMed: 10467896]
- Henson RNA. Short-term memory for serial order: the start-end model. *Cognitive Psychology* 1998;36:73–137. [PubMed: 9721198]
- Henson RNA. Positional information in short-term memory: relative or absolute? *Memory and Cognition* 1999;27:915–927.
- Hillis, A.; Rapp, B. Cognitive and neural substrates of written language comprehension and production. In: Gazzaniga, M., editor. *The Cognitive Neurosciences III*. Cambridge, MA: MIT Press; 2004.
- Houghton, G.; Glasspool, D.; Shallice, T. Spelling and serial recall: Insights from a competitive queueing model. In: Brown, GDA.; Ellis, NC., editors. *Handbook of spelling: Theory, process and intervention*. Chichester: John Wiley & Sons; 1994.
- Houghton G, Zorzi M. Normal and Impaired Spelling in a Connectionist Dual-Route Architecture. *Cognitive Neuropsychology* 2003;20:115–162.
- Jacobs, A.; Rey, A.; Ziegler, J.; Grainger, J. MROM-p. An Interactive Activation, Multiple Readout Model of Orthographic and Phonological Processes in Visual Word Recognition. In: Grainger, J.; Jacobs, AM., editors. *Localist Connectionist Approaches to Human Cognition*. Hillsdale, N.J.: Lawrence Erlbaum Associates Ltd; 1998.
- Kinoshita S, Norris D. Transposed-letter priming of prelexical orthographic representations. *Journal of Experimental Psychology: Learning, Memory and Cognition* 2009;35:1–18.
- Kertesz, A. *Western Aphasia Battery*. San Antonio, TX: The Psychological Corporation; 1982.
- Lashley, KS. The problem of serial order in behavior. In: Jeffress, LA., editor. *Cerebral mechanisms in behavior*. New York: Wiley; 1951.
- McClelland JL, Rumelhart DE. An interactive activation model of context effects in letter perception, Part I: An account of basic findings. *Psychological Review* 1981;88:375–407.
- McCloskey M. Theory and evidence in cognitive neuropsychology: A “radical” response to Robertson, Knight, Rafal, and Shimamura. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 1993;19:718–734.
- McCloskey, M. The future of cognitive neuropsychology. In: Rapp, B., editor. *The Handbook of Cognitive Neuropsychology: What Deficits Reveal About the Human Mind*. Philadelphia: Psychology Press; 2001.
- McCloskey M. Beyond task dissociation logic: A richer conception of cognitive neuropsychology. *Cortex* 2003;39:196–202. [PubMed: 12627779]
- McCloskey M, Badecker W, Goodman-Schulman RA, Aliminosa D. The structure of graphemic representations in spelling: Evidence from a case of acquired dysgraphia. *Cognitive Neuropsychology* 1994;11:341–392.

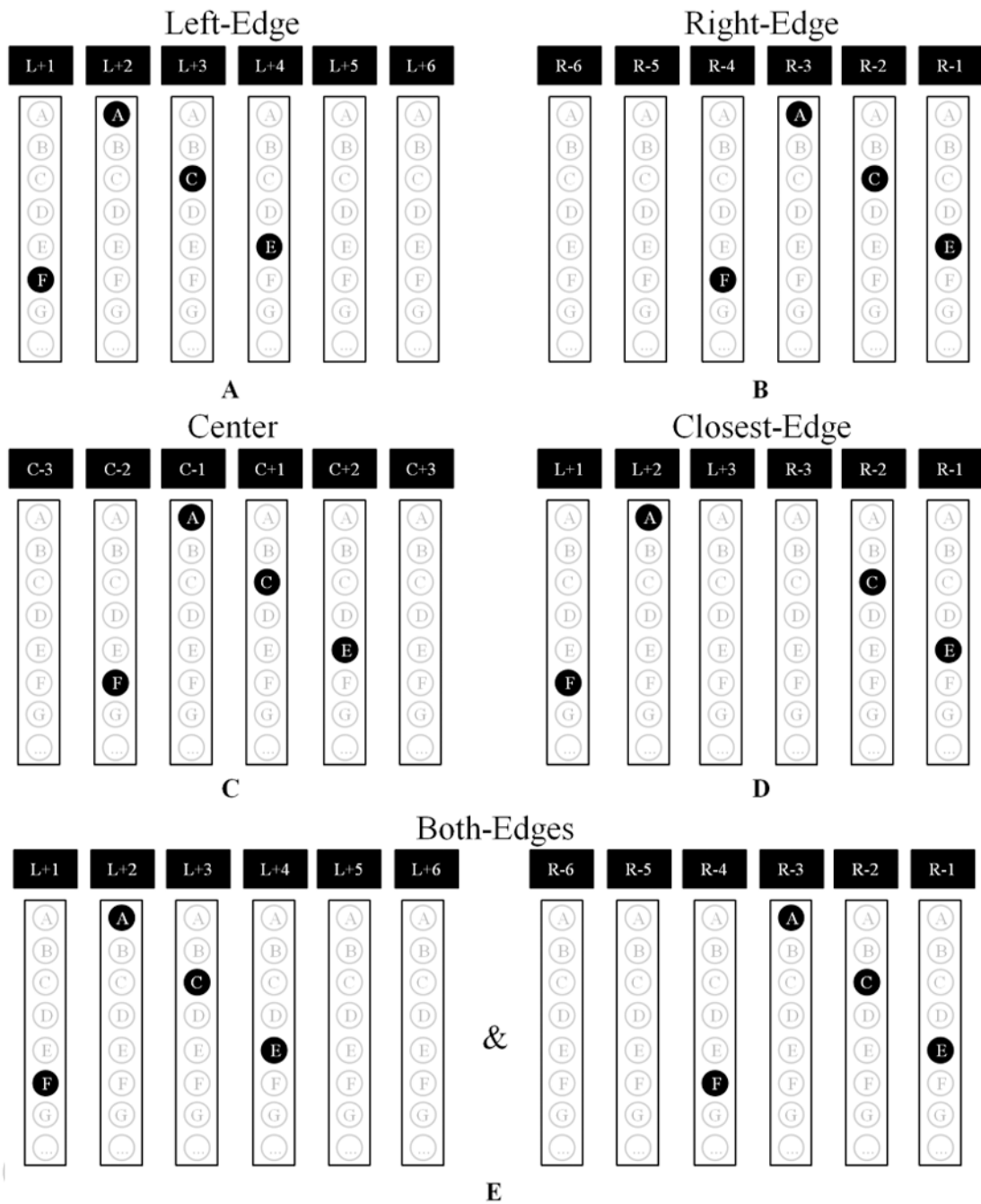
- McCloskey M, Caramazza A. Theory and methodology in cognitive neuropsychology: A response to our critics. *Cognitive Neuropsychology* 1988;5:583–623.
- McCloskey M, Macaruso P, Rapp B. Grapheme-to-lexeme feedback in the spelling system: Evidence from a dysgraphic patient. *Cognitive Neuropsychology* 2006;23:278–307.
- Miceli G, Capasso R. Spelling and dysgraphia. *Cognitive Neuropsychology* 2006;23:110–134.
- Mozer, M. Early parallel processing in reading: A connectionist approach. In: Coltheart, M., editor. *Attention and Performance XII: The psychology of reading*. Hove, UK: Lawrence Erlbaum Associates Ltd; 1987.
- Olson A, Caramazza A. Orthographic structure and deaf spelling errors: syllables, letter frequency, and speech. *Quarterly Journal of Experimental Psychology A* 2004;57:385–417.
- Palmer C. Sequence memory in music performance. *Current Directions in Psychological Science* 2005;14:247–250.
- Patterson K. Lexical but non-semantic spelling? *Cognitive Neuropsychology* 1986;3:341–367.
- Perea M, Carreiras M. Do transposed-letter similarity effects occur at a syllable level? *Experimental Psychology* 2006;53:308–315. [PubMed: 17176663]
- Perea, M.; Lupker, SJ. Transposed-letter confusability effects in masked form priming. In: Kinoshita, S.; Lupker, SJ., editors. *Masked priming: State of the art*. Hove, UK: Psychology Press; 2003. p. 97-120.
- Perea M, Lupker SJ. Can CANISO activate CASINO? Transposed-letter similarity effects with nonadjacent letter positions. *Journal of Memory and Language* 2004;51:231–246.
- Perry C, Ziegler JC, Zorzi M. Nested incremental modeling in the development of computational theories: the CDP+ model of reading aloud. *Psychological Review* 2007;114:273–315. [PubMed: 17500628]
- Plaut DC, McClelland JL, Seidenberg MS, Patterson K. Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review* 1996;103:56–115. [PubMed: 8650300]
- Prinzmetal W, Treiman R, Rho SH. How to see a reading unit. *Journal of Memory and Language* 1986;25:461–475.
- Rapp B. The nature of sublexical information: The bigram trough hypothesis examined. *Journal of Memory and Language* 1992;31:33–53.
- Rapp B, Benzing L, Caramazza A. The autonomy of lexical orthography. *Cognitive Neuropsychology* 1997;14:71–104.
- Rapp B, Caramazza A. From graphemes to abstract letter shapes: Levels of representation in written spelling. *Journal of Experimental Psychology: Human Perception and Performance* 1997;23:1130–1152. [PubMed: 9269731]
- Rapp B, Epstein C, Tainturier MJ. The integration of information across lexical and sublexical processes in spelling. *Cognitive Neuropsychology* 2002;19:1–29.
- Schoonbaert S, Grainger J. Letter position coding in printed word perception: Effects of repeated and transposed letters. *Language and Cognitive Processes* 2004;19:333–367.
- Seidenberg MS, McClelland JL. A distributed, developmental model of word recognition and naming. *Psychological Review* 1989;96:532–568.
- Seymour, PHK. *Human Visual Cognition*. New York: St. Martin's Press; 1979.
- Spoehr K, Smith E. The Role of Syllables in Perceptual Processing. *Cognitive Psychology* 1973;5:71–89.
- Taft M. Lexical access via an orthographic code: The Basic Orthographic Syllabic Structure (BOSS). *Journal of Verbal Learning and Verbal Behavior* 1979;18:21–39.
- Tainturier MJ, Caramazza A. The status of double letters in graphemic representations. *Journal of Memory and Language* 1996;35:53–73.
- Tainturier, MJ.; Rapp, B. The Spelling Process. In: Rapp, B., editor. *The Handbook of Cognitive Neuropsychology: What Deficits Reveal About the Human Mind*. Philadelphia: Psychology Press; 2001.
- Tainturier MJ, Rapp B. Is a single graphemic buffer used in reading and spelling? *Aphasiology* 2003;17:537–562.
- Tainturier MJ, Rapp B. Complex graphemes as functional spelling units: Evidence from acquired dysgraphia. *Neurocase* 2004;10:122–131. [PubMed: 15788251]

- Tsapkini K, Rapp B. The orthography-specific functions of the left fusiform gyrus: Evidence of modality and category specificity. *Cortex*. in press.
- Van Assche E, Grainger J. A study of relative-position priming with superset primes. *Journal of Experimental Psychology: Learning, Memory and Cognition* 2006;32:399–415.
- Whitney C. How the brain encodes the order of letters in a printed word: The SERIOL model and selective literature review. *Psychonomic Bulletin & Review* 2001;8:221–243. [PubMed: 11495111]
- Wing, AM.; Baddeley, AD. Speech errors in handwriting: A corpus and a distributional analysis. In: Frith, U., editor. *Cognitive Processes in Spelling*. London: Academic Press; 1980.
- Zorzi M, Houghton G, Butterworth B. Two routes or one in reading aloud? A connectionist dual-process model. *Journal of Experimental Psychology: Human Perception and Performance* 1998;24:1131–1161.

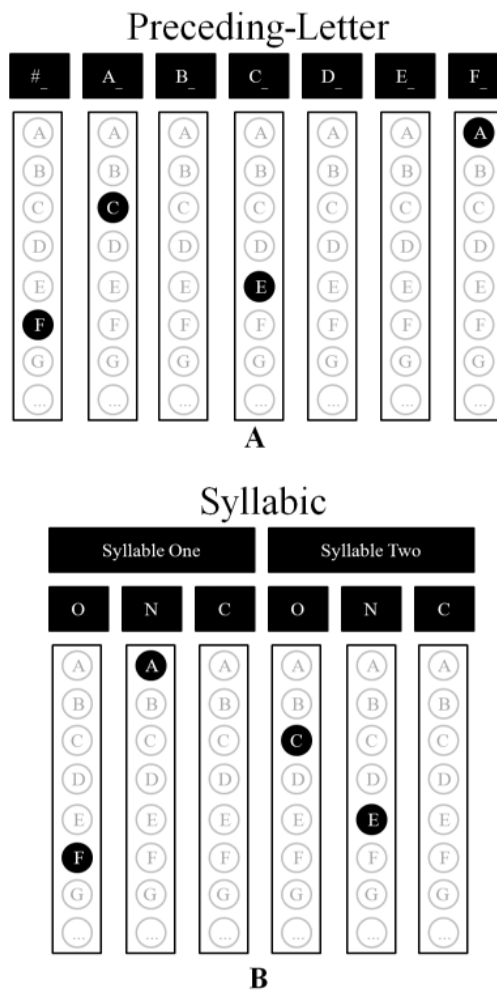




**Figure 1.**  
Schematic depiction of the cognitive spelling mechanisms.

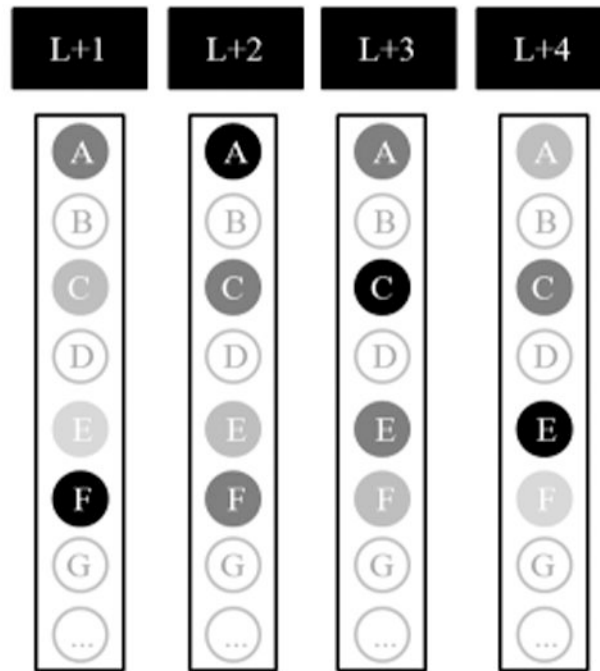


**Figure 2.** Left-edge (A), right-edge (B), center (C), closer-edge (D), and both-edges (E) representations for the word FACE.

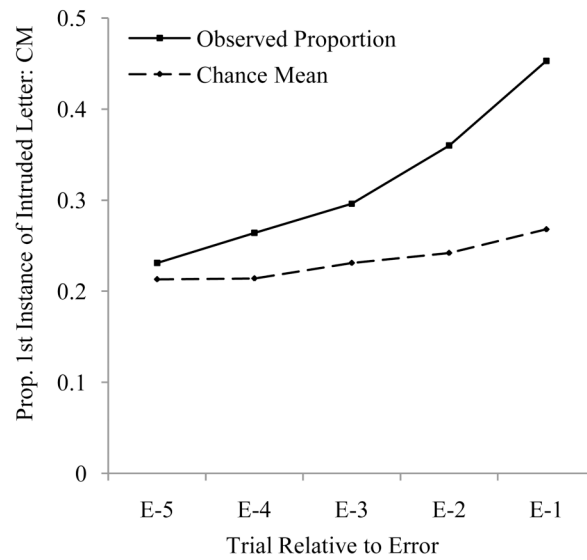


**Figure 3.** Letter-context (A) and syllabic (B) representations for the word FACE. The letter-context representation is a preceding-letter scheme, in which the position of a letter is defined by the immediately preceding letter (# indicates a word boundary). The syllabic positional scheme defines each letter's position by its role as an orthographic syllable onset (O), nucleus (N), or coda (C) letter.

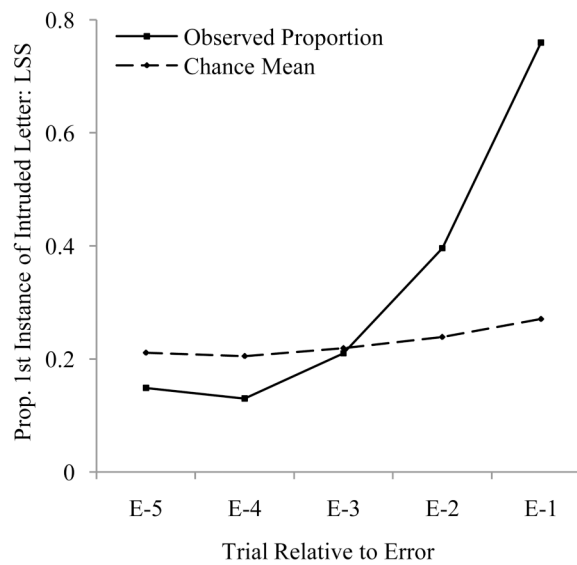
# Gradient Left-Edge



**Figure 4.**  
A graded left-edge representation of the word FACE.



A



B

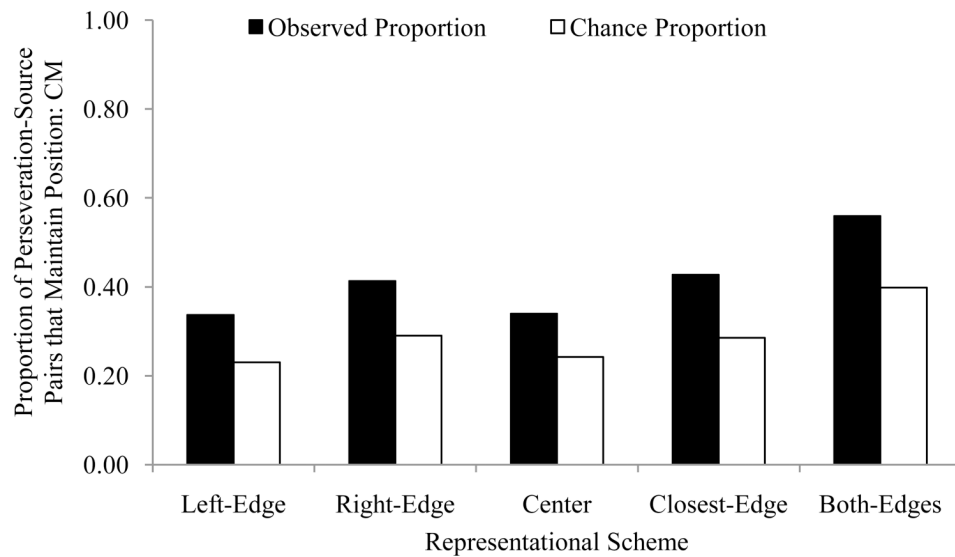
**Figure 5.**

The proportion of intruded letters produced by CM (A) and LSS (B) on trial E, observed most recently in preceding response (E-1 to E-5). The solid line is the proportion observed, while the dotted line is the mean proportion expected by chance obtained from 10,000 chance analysis runs.

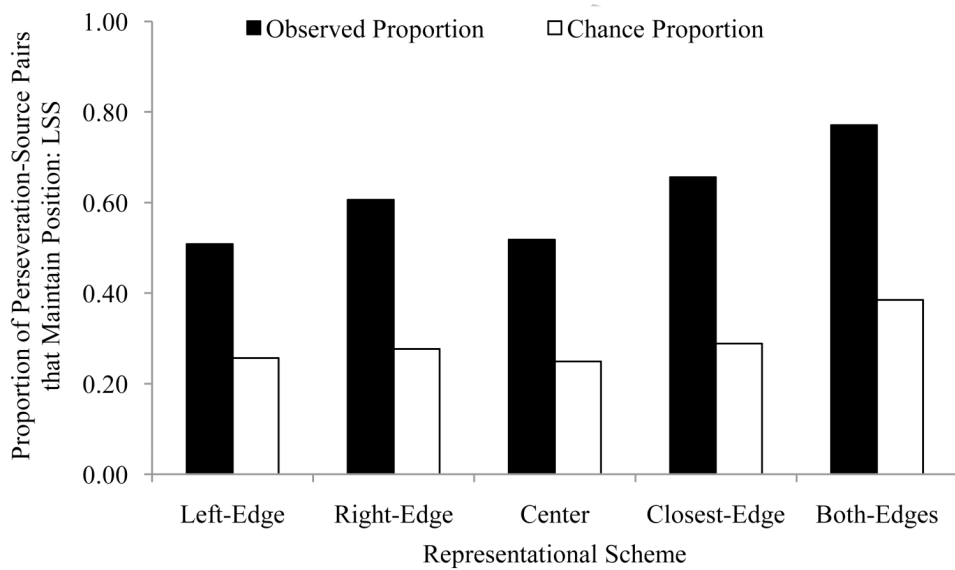


TRIAL E				TRIAL E - 1						
Intrusion Number	Stimulus	Response	Intruded Letter	Actual E-1 Response	Control Response Pool	Control Sample 1	Control Sample 2	Control Sample 3	...	Control Sample 10,000
1	"edge"	ER <u>G</u> E	R	F <u>R</u> ENCE	BALLON, WINDOW, SHINGE, PEBBEL, CUP <u>R</u> ET, COFFEE ...	SHINGE	CUP <u>R</u> ET	PEBBEL	...	WINDOW
2	"crank"	CL <u>A</u> NK	L	K <u>I</u> GH	EDIT, ROIT, NOSE, LAEP, HORN, IDGO, CLEW, TUBE ...	HORN	NOSE	CL <u>E</u> W	...	NOSE
3	"obvious"	OBVIO <u>S</u> E	E	S <u>H</u> E	HET, COL, ASK, E <u>F</u> T, TOP, HUS, SIT, NAT, ICE, DIG, TOB ...	E <u>F</u> T	NAT	HET	...	TOP
...	...	...	...	...	...	...	...	...	...	...
2704	"slip"	SP <u>R</u> IL	R	C <u>R</u> OW	EDIT, ROIT, NOSE, LAEP, HORN, IDGO, CLEW, TUBE ...	LAEP	HOSP	IDGO	...	HORN
<i>Proportion of E-1 Responses (Actual or Control) Containing Intruded Letter</i>				.453		.262	.254	.278	...	.269
				<b>Actual</b>		<b>Control (<math>\bar{x} = .268</math>)</b>				

**Figure 6.** Examples of intrusion errors, prior responses, and control responses from CM's corpus, illustrating method for estimating the likelihood of an intruded letter being present in a prior response by chance.



A



B

**Figure 7.** Proportion of perseveration-source response pairs in which an intruded letter in the perseveration response appear in the same position in the source response, as defined by various content-independent positional schemes (black bars) compared to the proportion expected by chance (white bars), for CM (A) and LSS (B).

**Table 1**  
**Example spelling errors for LSS and CM**

<b>Target</b>	<b>CM Response</b>	<b>LSS Response</b>
absence	ABSONCE	ABERE
basis	BOSIC	BACIC
belief	BEING	BALCE
could	COUTH	COUNLD
future	FURANCE	FUTUE
instinct	INSTICT	INSPECT
iron	INLY	INSTON
leopard	LEOPORD	OPROD
riot	ROIT	ROID
yawn	YALW	GAMT

**Table 2**  
**LSS's error EDGE → ERGE, and the five preceding target-response pairs**

<b>Trial</b>	<b>Target</b>	<b>Response</b>
E-5	CULT	CUNPH
E-4	HINT	HINT
E-3	WISH	WISH
E-2	MORTAL	MORTCH
E-1	FRENCH	FRENCE
E	EDGE	ERGE

**Table 3**  
**CM's Error BLUNT ? BLANT, and the Four Preceding Responses within the Perseveration Window**

<b>Trial</b>	<b>Target</b>	<b>Response</b>
E - 4	PIANO	PI <u>A</u> NO
E - 3	ORANGE	OR <u>A</u> NGE
E - 2	NOSE	NOSE
E - 1	FARM	F <u>A</u> RM
E	BLUNT	BL <u>A</u> NT



**Table 4**  
**Position of A in Source and Perseveration Responses According to Left-Edge and Right-Edge Positional Schemes**

	Response	Positional Scheme	
		Left-Edge	Right-Edge
Source Response	F <u>A</u> RM	L2	R3
Perseveration Response	BL <u>A</u> NT	L3	R3
<i>Same Position?</i>		<i>No</i>	<i>Yes</i>

**Table 5**  
**Estimated Proportions of Position Matches for True Perseveration-Source Pairs**  
**(pPMforTrPS) for Discrete Content-Independent Position Schemes**

<i>Position Scheme</i>	<i>Participant</i>	
	CM	LSS
Left Edge	.48*	.57*
Right Edge	.58*	.69*
Center	.48*	.60*
Closer Edge	.62*	.75*
Both Edges	.78	.87

\* p<.0001 when compared to Both Edges

**Table 6**  
**Estimated Proportions of Position Matches for True Perseveration-Source Pairs**  
**(pPMforTrPS) for the Discrete Both-Edges Scheme and Component Schemes Plus Random**  
**Source Predictions**

<i>Position Scheme</i>	<i>Participant</i>	
	CM	LSS
Left-Edge + Random	.58*	.65*
Right-Edge + Random	.66*	.74*
Closer-Edge + Random	.70*	.80*
Farther-Edge + Random	.51*	.58*
Both Edges	.78	.87

\*p<.0001 when compared to Both Edges

**Table 7**  
**Estimated Proportions of Position Matches for True Perseveration-Source Pairs**  
**(pPMforTrPS) for the Letter-Context Schemes and the Graded Both-Edges Scheme**

<i>Position Scheme</i>	<i>Participant</i>	
	CM	LSS
Preceding-Letter	.25*	.38*
Graded Both-Edges	.93	.97
Following-Letter	.29*	.52*
Graded Both-Edges	.95	.94
Trigram	.08*	.17*
Graded Both-Edges	.94	.91

\* $p < .0001$  when compared to Both Edges

**Table 8**  
**Estimated pPMforTrPS for Graded Both-Edges Plus Letter-Context Schemes and Matched**  
**Graded Both-Edges Plus Random Schemes**

<i>Position Scheme</i>	<i>Participant</i>	
	CM	LSS
Graded Both-Edges + Preceding-Letter	.94	.98
Graded Both-Edges + Random	.94	.97
Graded Both-Edges + Following-Letter	.97	.94
Graded Both-Edges + Random	.96	.94
Graded Both-Edges + Trigram	.94	.91
Graded Both-Edges + Random	.94	.91

**Table 9**  
**Estimated pPMforTrPS Values for Discrete Syllabic, Graded Syllabic, and Discrete Syllabic Plus Random Schemes**

<i>Position Scheme</i>	<i>Participant</i>	
	CM	LSS
Discrete Syllabic	.78*	.85*
Graded Syllabic	.91	.91
Discrete Syllabic + Random	.83*	.87*

\* $p < .0001$  when compared to Graded Syllabic

**Table 10**  
**Estimated pPMforTrPS Values for Graded Syllabic and Graded Both-Edges Schemes**

<i>Position Scheme</i>	<i>Participant</i>	
	CM	LSS
Graded Syllabic	.91*	.91*
Graded Both-Edges	.94	.98

\* $p < .0001$  when compared to Graded Both Edges



**Table 11**  
**Estimated pPMforTrPS Values for Graded Both-Edges, Graded Both-Edges Plus Graded Syllabic, and Graded Both-Edges Plus Random Schemes**

<i>Position Scheme</i>	<i>Participant</i>	
	CM	LSS
Graded Both-Edges	.94	.98
Graded Both-Edges + Graded Syllabic	.97	.97
Graded Both-Edges + Random	.95	.98