

Language Production and Serial Order: A Functional Analysis and a Model

Gary S. Dell and Lisa K. Burger
University of Illinois at Urbana-Champaign

William R. Svec
Indiana University Bloomington

In speech production, previously spoken and upcoming words can impinge on the word currently being said, resulting in perseverations (e.g., “beef needle soup”) and anticipations (e.g., “cuff of coffee”). These errors reveal the extent to which the language-production system is focused on the past, the present, and the future and therefore are informative about how the system deals with serial order. This article offers a functional analysis of serial order in language and develops a general formal model. The centerpiece of the model is a prediction that the fraction of serial-order errors that are anticipatory, as opposed to perseveratory, can be closely predicted by overall error rate. The lower the error rate, the more anticipatory the errors are, regardless of the factors influencing error rate. The model is successfully applied to experimental and natural error data dealing with the effects of practice, speech rate, individual differences, age, and brain damage.

More than 45 years ago, Lashley (1951) directed the attention of psychologists to the problem of serial order. Ordered behavior, he wrote, cannot simply be the product of associations between elementary responses. Instead, there must be a hierarchically organized plan or schema that is separate from the responses and yet determines their order. Although it took some time for Lashley’s paper to be appreciated (Bruce, 1994), his insights are now acknowledged to be central to theory in memory (e.g., Murdock, 1974), psycholinguistics (e.g., Levelt, 1989; MacKay, 1987), and motor control (e.g., Rosenbaum, 1990).

This article examines the nature of order schemata in that behavior in which the serial-ordering problems are perhaps most acute, in language. We begin by reviewing evidence from serial-order errors in speech, particularly errors in which sounds or words are either *anticipated*, spoken ahead of their time, or *perseverated*, produced later than they should be. This evidence suggests that the past and future often impinge on the present while we are speaking and that the extent to which the language-production system is focused on the past or the future depends on a number of factors.

We then present an experimental study of one of these factors,

the degree to which a spoken phrase is familiar. This study shows that, as a phrase gains familiarity, the pattern of order errors moves from one in which perseverations dominate to one in which anticipations are more common. This is termed the *anticipatory practice effect*. The ability of various serial-order theories to explain this effect is reviewed, and a simple quantitative model is developed from a class of existing activation-based theories (e.g., Dell, 1986; Estes, 1972; Houghton, 1990; MacKay, 1987). This model makes predictions about the relationship between speech rate, practice, overall error rate, and the extent to which errors are anticipatory or perseveratory. Most importantly, it states that the proportion of errors that are anticipatory should be predictable from the overall error rate, regardless of the combination of factors that led to that particular rate. This claim, called the *general anticipatory effect*, is tested in another experiment.

Ultimately, we argue that a theory of serial order in speech must satisfy a set of functional requirements: The system must activate the present, deactivate the past, and prepare to activate the future. Furthermore, the order schema that accomplishes these functions must be capable of doing so for both stored and novel sequences. We attempt to show that the general anticipatory effect follows from these functions, given certain assumptions. In short, we claim that when the language-production system is working well, it looks to the future and does not dwell on the past.

Serial-Order Errors in Speech

Everyday speech errors, or slips of the tongue, occur once or twice every 1,000 words on average for normal adults (Garnham, Shillcock, Brown, Mill, & Cutler, 1981; Hotopf, 1983). Young children appear to have a higher error rate (e.g., seven times the adult rate, for 2–3-year-olds; Wijnen, 1992), and aphasic patients may have error rates in their spontaneous speech that are hundreds of times greater than normal (e.g., Schwartz, Saffran, Bloch, & Dell, 1994).

Speech errors can be categorized along two dimensions: the size of the linguistic units involved and the nature of the distur-

Gary S. Dell and Lisa K. Burger, Beckman Institute, University of Illinois at Urbana-Champaign; William R. Svec, Department of Psychology, Indiana University Bloomington.

This research was supported by National Science Foundation Grant SBR 93-19368 and National Institutes of Health Grants DC-00191 and HD 21011 and grew out of a project involving Gary S. Dell, Myrna Schwartz, Eleanor Saffran, and Nadine Martin.

We are grateful for the advice of David Adams, Howard Berenbaum, Gary Bradshaw, Kay Bock, Cooper Cutting, Vic Ferreira, Zenzi Griffin, Susan Garnsey, Anita Govindjee, Prahlad Gupta, John Huitema, Gordon Logan, Gregory Murphy, Neal Pearlmuter, Brian Ross, Christine Sevald, and Ulrich Schade and for the contributions of Linda May, Julie Owles, and the Beckman Institute.

Correspondence concerning this article should be addressed to Gary S. Dell, Beckman Institute, University of Illinois, 405 N. Mathews Avenue, Urbana, Illinois 61801. Electronic mail may be sent via Internet to gdell@s.psych.uiuc.edu.

bance. Linguistic units of all sizes can slip, from articulatory gestures to whole clauses (Fromkin, 1971; Mowrey & MacKay, 1990). The most commonly collected errors, at least in the languages that have been extensively studied, involve units that roughly correspond to phonemes, as in Errors 1–3 below, or to words (see Error 4) or morphemes (see Error 5). The nature of the disturbance refers to whether errors involve the intrusion of linguistic material from outside the utterance, called *noncontextual errors*, or from within the utterance, called *contextual* or *movement errors*. Anticipations and perseverations are contextual errors. Error 1 below can be categorized as the anticipation of the phoneme /f/. The /f/ was spoken too early and replaced the correct sound /p/.¹ In Error 2 the vowel sound /i/ is perseverated. When an anticipation and perseveration occur on corresponding units, the error is called an *exchange*; Error 3 is an exchange of phonemes, and Error 4 is an exchange of words.

1. cup of coffee → cuff of coffee (Fromkin, 1971)
2. beef noodle → beef needle (Fromkin, 1971)
3. left hemisphere → heft lemisphere (Fromkin, 1971)
4. fill the car up with gas → fill the gas up with car (Dell & Reich, 1981)
5. thinly sliced → slicely thinned (Stemberger, 1982)

To a considerable extent, error categorization is a theory-laden decision, both with respect to the size of the disrupted unit and the nature of the disruption. In our studies, we distinguish between errors that involve whole meaningful units (words, stem morphemes), which we call *word errors*, and those that involve smaller pieces of the speech stream, called *sound errors*. We further assume that many errors have contextual sources and specifically adopt the categories of anticipation, perseveration, and exchange to label these errors. These categories are widely accepted (e.g., Cutler, 1981; Fromkin, 1971; Garrett, 1975, 1980; MacKay, 1970, 1974; Shattuck-Hufnagel, 1979, 1983; Stemberger, 1985), although it is also acknowledged that it is not easy to tell whether an error is a word or sound error or if it is contextual or noncontextual and, if contextual, whether it is an anticipation, perseveration, or exchange.

Anticipations, perseverations, and exchanges are often informally explained by the need for several linguistic units to coexist in a buffer (e.g., Fromkin, 1971; MacKay, 1970; Reich, 1977). Buffers serve many functions in information-processing systems, two of which are particularly important for language. First, the nature of language, both its abstract linguistic structure and its articulatory manifestation in speech, dictates that the form of a particular unit may depend on nonadjacent units, and moreover, on units from both the past and future. Having the units co-occur in a buffer helps the system compute these dependencies.

Second, buffers allow the processing levels of a system to work separately (e.g., Reich, 1977). A buffer can store the products of an earlier level, thereby allowing later levels that use these products to work at their own rate. Language does appear to be produced by a multileveled system, each level being associated with its own representation of the utterance. These representations, seen as organized sets of symbols, serve as buffers. Much of the research on language production has been concerned with identifying the number and character of these representations (e.g., Garrett, 1975). Current theory (e.g., Bock & Levelt, 1994) distinguishes between *grammatical en-*

coding, *phonological encoding*, and storage in an *articulatory buffer*. Grammatical encoding builds a syntactic representation of the utterance. Words are chosen but not specified in terms of their sounds; these abstract word symbols are assigned to grammatical functions such as subject and object and are placed in slots in a syntactic frame that dictates their order. Phonological encoding comprises the assembly of the sounds of words and the generation of intonation. The articulatory buffer is capable of storing motor representations of syllable strings in advance of overt articulation (Sternberg, Monsell, Knoll, & Wright, 1978). Contextual speech errors occur during the manipulation of the grammatical and phonological representations. Specifically, sound errors are associated with the phonological representation, and most word errors are associated with the grammatical representation (Fromkin, 1971; Garrett, 1975; MacKay, 1982; Stemberger, 1985).

The association of speech errors with distinctly linguistic representational levels comes from a consideration of the variables that influence the errors. Here we offer a brief review, focusing on contextual speech error types. (See Berg, 1988; Dell, 1986; Fromkin, 1973, 1980; Garrett, 1975; MacKay, 1970; Shattuck-Hufnagel, 1979; Stemberger, 1985, for complete discussions.) A serial-order error can be characterized by a *target linguistic unit* and an *error unit*. In anticipations and perseverations, a single target is replaced by an error unit from within the utterance. In an exchange, each of the exchanging units is both a target and an error.²

Previous discussions of anticipations, perseverations, and exchanges hypothesize a close relationship among the types (Dell, 1986; MacKay, 1970; Shattuck-Hufnagel, 1979). An anticipatory error (“left hemisphere” → “heft . . .”) can evolve into an exchange (“heft lemisphere”) because the replaced unit (e.g., /l/) seeks expression in the next available location in which it is allowed. Alternately, it can emerge as a simple anticipation (“heft hemisphere”), or the speaker can stop, thereby producing an incomplete anticipation (“heft . . . I mean, left hemisphere”). Perseverations and exchanges are related in that both involve the production of a unit in a position later than the intended position. The three error types are further related in that they are influenced by the same variables. Anticipations, perseverations, and exchanges at the phonological level usually involve similar sounds from similar word and syllable positions, and the resulting errors are typically phonologically well formed (see, e.g., Errors 1–3, above). At the word level, the three error types usually involve target and error words of the same grammatical class (e.g., Error 4). Word and sound exchanges, however, differ somewhat from anticipations and perseverations

¹ Many phoneme errors such as this one have an interpretation as phonological feature slips. If the voiceless labial stop acquires a continuant feature, the closest English sound is /ff/.

² We are not considering the categories of shift and contextual addition and deletion to be separate from serial-order errors involving replacement. Rather a perseveratory addition such as “black boxes” is simply viewed as a perseveration, with the target /b/ replaced by the error /bl/. Or the shift “back bloxes” is viewed as the exchange of /b/ and /bl/. This treatment of the error categories may not be justified generally (see, particularly, Garrett, 1975 and Stemberger, 1991), but for the purposes of elucidating anticipatory and perseveratory influences in speech it is a useful simplification.

in that the interacting units in an exchange are more constrained by similarity and tend to be closer together than those in anticipations and perseverations (Garrett, 1980).

Anticipations, Perseverations, and the “Good–Bad” Dimension

Serial-order errors that have a clear directionality—anticipations and perseverations—can be used to determine when behavior is focused on the past or the future. In particular, one can compare the extent to which errors are anticipatory versus perseveratory in certain situations or people. Such comparisons have been made with respect to four independent variables: the presence of aphasia, the speaker’s age, the rate of speech, and the speaker’s familiarity with the material uttered. These are discussed in turn.

Schwartz et al. (1994) noted that perseverations are a feature of many aphasic patients’ speech and went on to show specifically that F.L., classified as a jargon-aphasic patient, made more perseverations than anticipations in his spontaneous speech. His anticipatory proportion (*AP*), which we define as the proportion of anticipation and perseveration errors that are anticipations, was .32. The same analysis performed on spontaneous speech errors from nonaphasic speakers, culled by Garnham et al. (1981) from a tape-recorded and transcribed source (the London–Lund corpus, Svartvik & Quirk, 1980), showed that nonaphasic speakers make more anticipations than perseverations; the *AP* was .75. The difference between the aphasic patient and the nonaphasic speakers’ *AP* was highly reliable and, moreover, the error-pattern differences occurred for both sound and word errors.

Another variable that appears to influence the anticipatory proportion is age. Stemberger (1989) collected speech errors from adults and from two children with most of the observations coming at ages 2 and 3. He found that the adults’ slips were predominately anticipatory (*AP* = .60), while the children tended to perseverate more (*AP* = .47), particularly when they were young (for age 2, *AP* = .41).

There is some evidence that the proportion of errors that are anticipatory decreases as speech rate increases. Dell (1990) had experimental participants say two- or three-word phrases within either a short or long deadline and found that anticipations and incomplete anticipations were unaffected by the deadline, but perseverations and exchanges were less likely at the long deadline. Hence, the anticipatory proportion increased as the available time for speaking increased.

Finally, Schwartz et al. (1994) investigated the effect of practice on saying difficult phrases, such as “chef’s sooty shoe soles.” Phrases were produced by normal adults in time with a metronome. There were eight practice blocks during which participants said each of the nine phrases in the study twice. Practice both reduced errors and had a large effect on the anticipatory proportion. For the first practice block, perseverations were more likely (*AP* = .38), but by the eighth block, anticipations dominated (*AP* = .70). This change in *AP* with practice is the anticipatory practice effect.

In summary, the anticipatory proportion is lower in the spontaneous speech of at least some aphasic patients and young children compared to normal adults and when speakers are producing unfamiliar compared to practiced phrases. There is also

some evidence that a faster speech rate promotes a lower *AP*. According to Schwartz et al. (1994), the generalization from these data is that when the error rate is higher, errors tend to be more perseveratory. They distinguished between a “good” error pattern, in which errors were less likely and mostly anticipatory, and a “bad” error pattern characterized by more errors overall and by a high proportion of perseverations. Another feature of the good–bad dimension concerned the extent to which errors produced words over nonwords. The good pattern was associated with more slips with word outcomes.

Our goal is to examine the relation between the anticipatory proportion and other variables that affect overall error rate and to explain it by appealing to principles of serial ordering. But before we turn to these principles, we first replicate the anticipatory practice effect found by Schwartz et al. (1994). It is our view that practice is a particularly informative variable. In most theories of the production of verbal sequences, practice is assumed to increase the strength of the connections among units representing concepts, phrases, words, and sounds (e.g., Dell, Juliano, & Govindjee, 1993; Estes, 1972; Houghton, 1990; Jordan, 1986; MacKay, 1982, 1987; Rumelhart & Norman, 1982). Furthermore, variation in connection strength can be hypothesized to underlie the contrast between nonaphasic adults on the one hand and young children and aphasic patients on the other.

Experiment 1: Anticipatory Practice Effect

The goal of the first experiment is to replicate the anticipatory practice effect in a larger study than that of Schwartz et al. (1994). With a large data set, we can evaluate alternative accounts of the effect and see whether it generalizes across items.³

Method

Materials. The 16 phrases shown in Appendix A provided the materials that were practiced. Each phrase is a novel noun phrase consisting of four content words. Nine of these were also used by Schwartz et al. (1994). The onsets of each of the four content words in a phrase were chosen to make the phrases be tongue twisters. Specifically, each phrase made use of two similar word onsets (e.g., /s/ and /ʃ/ in “chef’s sooty shoe soles,” or /b/ and /br/ in “Brad’s burned bran buns”). The goal was to create phrases that are novel, meaningful, and short enough to be within memory span, but nonetheless are difficult to say at a normal speaking rate.

Participants. Forty-one students randomly sampled from the University of Illinois psychology participant pool received course credit for participation. All but one were native speakers of English, and the nonnative speaker’s data were not examined, resulting in 40 participants.

Procedure. There were eight trial blocks, during each of which participants produced each of the 16 phrases twice. For the first trial block, a single trial went as follows: A prompt on a computer screen (Dell System 200, Dell Computer Corp., Houston, TX) informed the participant to press *y* on the keyboard, which displayed a phrase. The experimenter turned on an audiovisual metronome (Franz electric metronome [Model LM FB5], Franz Manufacturing Co., New Haven, CT) at a rate of 0.8 beats/s and repeated the phrase at this rate, aligning the stressed syllable of each word with a beat. The participant then repeated the

³ In all of our experiments, we treat items as random effects and pool over participants. Because we have many fewer items than participants and item variance is larger than participant variance, this is the conservative approach.

phrase at the 0.8 rate. This is a very slow rate, and errors were extremely rare. The purpose of the participants' slow repetition was to verify their pronunciation of the phrase and to aid in memorization. The phrase then disappeared from the screen and was replaced by the prompt. The experimenter increased the metronome speed to 2.1 beats/s, which is a normal speech rate. When the participant pressed *y*, the phrase was redisplayed for 5 s for study. The phrase's disappearance after this interval was the signal for the participant to repeat the phrase from memory in time with the metronome. Participants were instructed to say the phrase twice with a four-beat pause between repetitions. This procedure was repeated for all 16 phrases to complete the first block. The second through eighth blocks were the same as the first with one exception. The participant did not repeat the phrase at the 0.8 rate; only the experimenter did. The phrases were presented in random order in each block.

Instructions consisted of running the participants through two trials on phrases that were not part of the experiment.

Error coding. Tape recordings made of the experimental sessions (Marantz cassette recorder PMD201, Marantz Company, Inc., Chatsworth, CA) were transcribed by a trained assistant who was unaware of the research hypotheses. Transcriptions indicated phonemic content but not information about timing.

Error coding consisted of the following stages: (a) The transcriber identified utterances that contained at least one error; (b) transcriptions of these utterances were listed without information identifying what trial block they came from; error utterances that occurred several times in the experiment (e.g., "Bonnie's brown bread box" spoken as "Bronnie's brown bread box" occurred 18 times) were listed only once; and (c) a single coder (Gary S. Dell) identified and categorized the errors in each error utterance in the list. Hence, errors were categorized in ignorance of trial block, and identical error utterances were coded in the same way. The error categories were sound perseveration, sound anticipation, sound exchange, sound anticipation-perseveration, word perseveration, word anticipation, word exchange, word substitution, and other. The other category included noncontextual sound errors, affix errors, and word deletions. The sound anticipation-perseveration category was for sound errors that are ambiguous between anticipation and perseveration, specifically where the intruding sound occurs both before and after the target location and where these are equidistant (in words)

from the location. (e.g., "chef's sooty *sue* soles" is sound anticipation-perseveration because there is an /s/-onset in both "sooty" and "soles"). The categorization rules included those of Schwartz et al. (1994), along with extra specifications that were needed to increase reliability. These are given in Appendix B.

The reliability of the categorization is good, but far from perfect in practice. A second coder was taught Categorization Rules 1-4 (see Appendix B) and categorized error utterances that corresponded to 227 errors according to the primary coder. Agreement was 82%. Upon inspection of the disagreements, it became clear that the primary coder was using additional heuristics based on his knowledge of speech error patterns. These are specified in Rules 5-6. In a further study of reliability, the primary coder recoded error utterances corresponding to 482 errors a year after the original coding. The agreement with the original coding was 94%. Disagreements were due to ambiguity from multiple errors and from mistakes made during the second coding. No data analyses were altered because of disagreements; that is, the initial coding was always retained.

Results and Discussion

There were 1,120 errors obtained making for an overall rate of 2.7 errors for every 100 words produced. Practice reduced errors considerably, from 5.9 errors/100 words in the first trial block to 1.5 errors/100 words in the eighth block. Table 1 gives the category totals for each block.

The relation between the log(number of errors) and log(trial block) was linear ($r = -.99$; see Figure 1) and is therefore consistent with other data showing performance to be a power function of practice (e.g., Anderson & Schooler, 1991; Logan, 1988; Newell & Rosenbloom, 1981). Our concern is less with the correctness of the power law, however, than with the fact that the obtained log/log slopes can be used to index learning and, particularly, that slope differences between error types indicate changes in the relative proportions of types with practice. These slopes and their corresponding correlations are given for the various error categories in Table 1.

Table 1
Error Totals, Log-Log Slopes, Anticipatory Proportions (APs),
and Correlations for Experiment 1

Category	Trial								Total	Slope	<i>r</i>
	1	2	3	4	5	6	7	8			
Sound	206	126	104	85	68	63	49	46	747	-.714	-.99
SA	56	38	37	22	25	23	23	20	244	-.487	-.95
SP	88	57	35	37	24	24	11	15	291	-.914	-.95
SE	26	9	11	11	10	10	10	6	93	-.462	-.80
SAP	36	22	21	15	9	6	5	5	119	-1.028	-.95
Word	87	72	53	35	38	20	19	29	353	-.726	-.92
WA	21	15	22	10	15	6	6	8	103	-.585	-.79
WP	47	30	18	6	11	6	5	8	131	-1.073	-.92
WE	16	22	7	16	7	5	8	6	87	-.566	-.73
WS	3	5	6	3	5	3	0	7	32		
Other	9	2	3	3	0	1	1	1	20		
Total	302	200	160	123	106	84	69	76	1,120	-.721	-.99
Anticipatory proportion	.36	.38	.53	.43	.53	.49	.64	.55			

Note. For categories with zero entries, no slope or correlation is computed. SA = sound anticipation; SP = sound perseveration; SE = sound exchange; SAP = sound anticipation-perseveration; WA = word anticipation; WP = word perseveration; WE = word exchange; WS = word substitution.

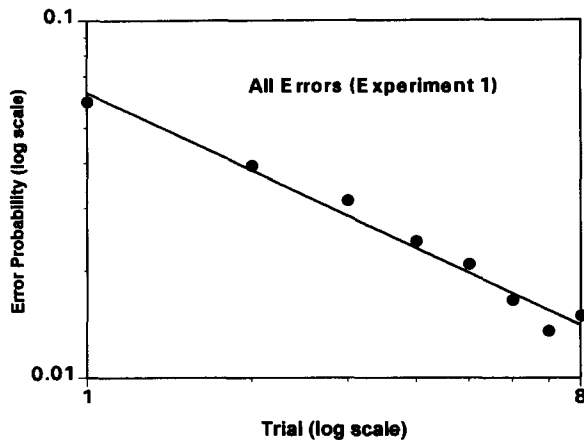


Figure 1. The effect of practice on overall error rate for Experiment 1.

The main result of the experiment was a clear anticipatory practice effect. The anticipatory proportion favored perseverations early in learning and favored anticipations after practice. AP was .37 pooled over the first two blocks and .59 over the last two blocks. This trend can be indexed by the difference in slopes for sound and word anticipations and perseverations. The slopes are nearly twice as great for perseverations. Furthermore, the slope differences were consistent across phrases. To show this, we computed log/log slopes for anticipations and perseverations for each phrase (combining word and sound errors). Because of the small number of errors in some phrases, particularly in the late practice blocks, some blocks had no errors in them for a particular phrase and error type, complicating the computation of logs. We dealt with this by randomly pooling each phrase with a zero entry for one or more blocks with another phrase with a zero entry somewhere doing this recursively until all phrases had nonzero entries for anticipations and perseverations for all eight blocks. This resulted in seven "superphrases." Each superphrase had a greater slope for perseverations than anticipations, with the mean of the differences being .513, $t(6) = 4.91$, $p < .002$.

In the following analyses, we show that this slope difference is robust in several respects. First, because of the potential ambiguities in coding utterances with multiple errors, we recomputed the slopes for anticipations and perseverations with all errors from multiple-error utterances removed. This reduced 769 anticipations and perseverations to 613 errors. The slopes, pooled over word and sound errors, was $-.477$ for anticipations and $-.923$ for perseverations, little changed from the original $-.510$ and $-.958$, respectively.

Next the slopes were recomputed with another subclass of errors removed, those sound anticipations and perseverations where the error unit occurred both before and after the target, but where one unit was closer to the target than the other and, hence, determined classification. Removing these cases resulted in 647 anticipations and perseverations, and the relevant slopes were $-.441$ and $-.927$, which are again very similar to the original slopes.

One feature of anticipations is that they tend to occur early in the phrase, whereas perseverations tend to occur later. For

example, word perseverations cannot possibly occur on the first word and word anticipations cannot occur on the last. This raises the question of whether the anticipatory practice effect is a product of an interaction between serial position and learning or whether it is independent of any serial position effects. To examine this, we broke down word anticipations, word perseverations, sound anticipations, and sound perseverations by serial position and examined their slopes. The following error-type/position combinations had enough errors to compute slopes: sound-anticipation/1, word-anticipation/1, sound-perseveration/2, word-anticipation/2, sound-anticipation/3, word-perseveration/3, sound-perseveration/4, word-perseveration/4.⁴ To eliminate zero entries, the eight trial blocks were reduced to four by combining the first with the second, the third with the fourth, and so on. The slopes from steepest to shallowest were -2.09 (sound-perseveration/2), -1.43 (word-perseveration/4), -1.26 (word-perseveration/3), -1.17 (sound-perseveration/4), -0.85 (word-anticipation/1), -0.81 (sound-anticipation/3), -0.52 (sound-anticipation/1), -0.44 (word-anticipation/2). The four perseveration slopes are steeper than the four anticipation slopes, and there is clearly no trend associated with serial position. Therefore the anticipatory practice effect is not a byproduct of differential learning across serial positions.

Finally, we consider the possibility that the anticipatory practice effect results from an increased tendency for participants to stop speaking after realizing that they have erred. With more practice, they may pause and correct themselves after making an initial anticipatory slip, for example, "sef's . . . chef's sooty shoe soles," whereas with less practice, this error might have ended up as an exchange, "sef's shooty shoe soles." Because incomplete anticipations such as "sef's . . . chef's sooty shoe soles" are counted as anticipations, an increased anticipatory proportion could conceivably result from an increased tendency to stop in the middle of exchanges. We therefore eliminated all errors from utterances in which speakers stopped before finishing. As expected, this eliminated more anticipations (84) than perseverations (36), but the difference in slopes for the remaining errors between anticipations ($-.400$) and perseverations ($-.966$) was actually a bit larger than originally. Clearly, the anticipatory practice effect is not due to an increasing relative proportion of incomplete anticipations.

In summary, the experiment found a strong anticipatory practice effect, one that is present on sound and word errors, generalizes across items and serial positions, and is robust over different conceptions of what constitutes an anticipation.

Before turning to the theoretical section of the article, there is one more empirical issue that needs to be briefly addressed. We have been assuming that practicing a phrase does something fairly specific to that phrase, for example, changes the connection weights among the linguistic units that represent the phrase. Alternately, the improvement with practice could be a general effect, such that practicing some phrases makes other phrases

⁴ Sound-perseveration/3 and sound-anticipation/2 had very few errors because of the patterning of onsets in the phrases—either ABBA or ABAB—and the rules about coding sound anticipation-perseverations. A contextual onset error in Position 2 was either sound-perseveration or sound anticipation-perseveration, while such an error in Position 3 was either sound-anticipation or sound anticipation-perseveration. Word-anticipation/3 had only eight errors.

easy to say. Experiment 2 examined this issue by testing the production of a new group of phrases after eight practice blocks with the original group of phrases. If practice produces only specific effects, the error rate for the new group should be similar to that of the first block of the original group.

Experiment 2: Are Practice Effects Specific or General?

Method

The 16 phrases used in Experiment 1 were divided into two groups of 8. Using the same procedure as in Experiment 1, participants practiced one of the phrase groups for eight trial blocks and then on the ninth, or transfer, block recited the phrases for the other group. Twenty-two participants from the same population as Experiment 1 were tested, counter-balancing the assignment of phrase groups to Blocks 1–8 and the transfer block. Two participants were eliminated because of failure to keep time with the metronome, and one was eliminated because he was not a native speaker of English.

Results and Discussion

Audiotapes were transcribed by William R. Svec, and errors were coded as before by Gary S. Dell. As in Experiment 1, the error totals diminished with practice: 121, 64, 52, 34, 30, 29, 30, 29 for Blocks 1–8 respectively (log/log line, $r = -.97$, slope = $-.717$). On the transfer block, however, there were 98 errors, a total that is not significantly different from that of block 1 ($p > .10$). We conclude that practice effects are largely specific to the practiced phrases, although we cannot rule out the existence of a very small general effect. Because this experiment had fewer than one-quarter of the observations of Experiment 1, it lacked the power to test for the anticipatory practice effect. The expected pattern emerged, but the number of relevant errors is too small. The anticipatory proportion increased with practice ($AP = .53$ for Blocks 1 and 2, 77 errors; $AP = .61$ for Blocks 7 and 8, 23 errors) and, most importantly, went back down for the new items in the transfer block ($AP = .42$, 62 errors).

The results of Experiments 1 and 2 show that practice changes the error probability and error pattern associated with the practiced phrases. The next section considers these effects in the light of theories of serial order.

Serial Order and the Anticipatory Practice Effect

Why does practice create an error pattern that is more oriented to the future than the past? One can address this question in two ways, functionally and mechanistically. That is, one can ask why the system should work this way, given its purposes, and one can ask how it does so. We first provide a functional analysis of serial order and then consider the mechanisms involved in several activation-based theories that have been applied to language, asking how they achieve these functions. From these theories, we develop a formal model and show that it accounts for the anticipatory practice effect and makes the set of predictions that we termed the general anticipatory effect.

What Must a Serial-Order Mechanism Do?

The production of a sequence involves the ordered activation of representations of its elements. We use the designation *pres-*

ent for the representational unit corresponding to the element that should currently be produced, *past* for the unit for the immediately preceding element, and *future* for the unit representing the element after the *present* element. The principal functions of a serial-order mechanism are as follows:

1. *Turn-On Function.* The system must activate the *present*. It must have some way of determining which unit is the present one and cause it to adopt an activated state.

2. *Turn-Off Function.* The system must deactivate the *past*. Most activation-based theories have assumed that an activated unit is associated with a large positive number and that this high level of activation does not decay immediately. Thus, when the present unit is not the initial one of a sequence, the system must counter the activation of previous units. The turn-off function serves this purpose.

In principle, any system that identifies and activates the *present* and deactivates the *past* is an effective serial-order mechanism. Practically speaking, though, a third function is required:

3. *Prime Function.* The system must prepare to activate the *future*. One factor that motivates the prime function can be called the *throw-away principle*: It is easier to throw away what you have than to find what you do not have. Applying this principle to serial order means that the turn-off function is easy: you can deactivate the past element because the system currently has access to it by virtue of its activated state. As we shall see, many models turn off the past through a *self-inhibition mechanism*: After each element has been fully activated, it throws itself away. The turn-on function, in contrast, is hard: the present element is something that you do not yet have. The prime function counters the throw-away principle by making the *present* something that you do, at least partially, have. It does this by initiating the process of activating units after the present one. As we shall see, the serial-order theories that have been applied to language implement the prime function by means of the activation of a *plan representation*, a unit or set of units that can influence fairly directly the activation of elements of the sequence. By virtue of the activation of the plan, each element is a step away from activation.

While some theories restrict priming to just activation of the plan, most assert that the activation of the plan causes anticipatory activation of units for upcoming elements. Actual activation of future elements is important for language because the appropriateness and form of the *present* often depends on what the *future* will be. Every language offers countless examples of these dependencies at the grammatical, phonological, and articulatory levels. For instance, in English whether you say *a* or *an* is governed by the form of the next word. Or, in German, the identity of the perfect auxiliary, *sein* (to be) or *haben* (have), depends on a verb that could be several words in the future.

Examples such as *a-an* are true future-oriented dependencies because one really must have a representation of the *future* to determine the *present*. You do not first decide whether to say *a* or *an* and then choose a next word that fits. The earlier depends on the later, rather than the reverse. Many dependencies in language, though, are past-oriented; a later item depends on an earlier one, for example, English subject-verb number agreement when constituents are in their normal order. *Mary is*, not *Mary are*. With past-oriented dependencies one can imagine two possible production mechanisms. One is a look-to-the-past mechanism. When deciding whether to say *is* or *are*, look back

to what was already encoded for the subject of the sentence. The alternative is to program the *future* with the *present*. When *Mary* is the *present*, prepare the *future* verb that is to agree with it. To the extent that past-oriented dependencies are processed by this latter mechanism, they are much like future-oriented dependencies in that there is a need for co-activation of the *present* and *future*.

The tendency for a system to produce anticipations and perseverations should be related to the relative activation of the *past*, *present*, and *future*. To the extent that the turn-on and turn-off functions are working well, one would not expect perseverations. The balance between the turn-on and priming functions should affect the probability of anticipations. How effectively these functions are achieved and how this might vary with practice requires an analysis of the mechanisms behind them in serial-order models, to which we now turn.

Theories of Serial Order in Language

Hierarchies with forward lateral inhibition. These models (e.g., Estes, 1972; Rumelhart & Norman, 1982) use a plan (or "chunk" or "control") unit that connects by excitatory links to all of the elementary units in the sequence. Hence, when the plan is activated, it tends to activate the elements. Plan units can themselves act as elements in higher-order plans creating the multileveled hierarchies that appear to be necessary to account for data on sequence generation (e.g., Gordon & Meyer, 1987; Rosenbaum, Kenny, & Derr, 1983) and language production (e.g., Bock & Loebell, 1990; Ferreira, 1993; Fowler, Treiman, & Gross, 1993; MacKay, 1973; Meyer, 1994). Furthermore, each elementary unit inhibits all other units after it with the result that, initially, only the first element has considerable activation. After the first element has reached a threshold level of activation, it is deemed to have been produced, and the unit undergoes self-inhibition: It turns itself off. As a consequence of the inactivity of the first unit, it no longer inhibits later units and, hence, the second unit gains activation and so on until the end of the sequence. The top part of Figure 2 shows the connections involved in a forward-lateral-inhibition model representing the sounds of the word *cat*.

In forward-lateral-inhibition models, the prime function happens through the activation of the plan. Whether there is any anticipatory activation of the *future*, though, depends on the strength of the inhibitory connections and the exact nature of the activation function, but generally, the inhibition should dampen every unit except the *present*. The turn-off function is achieved by the self-inhibition of the *past*. Finally, the entire system effectively turns on the *present* because the prime function works with the turn-off function and the lateral inhibition so that the turn-off of the *past* automatically turns on the *present*.

Competitive-queuing model. Forward-lateral-inhibition models store a sequence by exciting its elements and using forward lateral inhibition among them to sort out the order. However, if the same response units are used for the same elements when they appear in different sequences, effective serial order is not always achieved because of interference among the sequences. For example, if response units for words are something like phonemes, words that are phonological anagrams, such as *cat*, *act*, and *tack*, cannot all be represented (see, e.g., Dell & O'Seaghda, 1994; Houghton, 1990; and MacKay, 1987). A

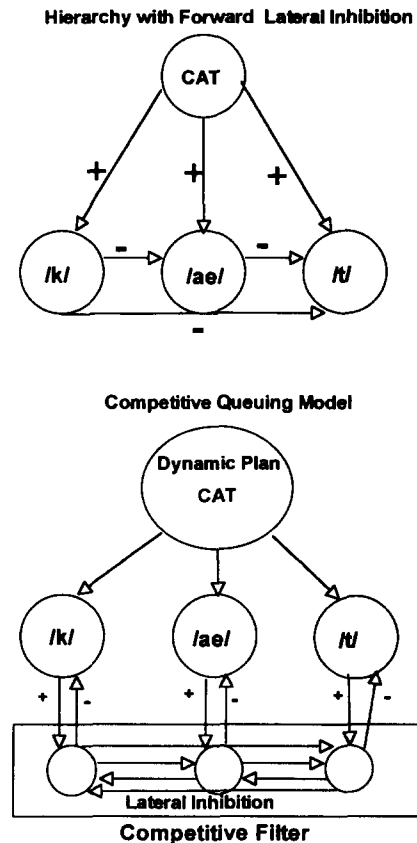


Figure 2. The mechanisms of forward-lateral-inhibition and competitive-queuing models illustrated with the sounds of the word *cat*. Excitatory connections are indicated by plus signs and inhibitory connections by minus signs. Self-inhibition is not shown.

related problem concerns sequences with repeated elements such as the word *did*. How can the lateral inhibition between /I/ and /d/ sort out this sequence? The competitive-queuing model developed by Houghton (1990) provides a good alternative (see also Burgess & Hitch, 1992; Gupta & MacWhinney, in press; and Hartley & Houghton, 1996). This model, based on interactive activation, learning, and serial-order principles pioneered by Grossberg (1978, 1982), has three layers of units: plan units, response units, and a *competitive filter*. Consider the sequence of phonemes in *cat*, as illustrated in the bottom of Figure 2. Plan units for *cat* connect to the response units /k/, /æ/, and /t/, such that there is, initially, a gradient of activation with /k/ being most activated, then /æ/, and then /t/. The gradient results from a time-varying signal in the plan that evolves from a state that can be interpreted as *beginning of cat* to one representing *end of cat* and from excitatory connection weights between plan and response units that are set by a supervised Hebbian learning algorithm. The activation levels of the response units are then copied into the competitive filter, which contains one unit for each response unit. However, the units in the competitive filter strongly inhibit one another with the result that only the most highly activated unit, the one for /k/, remains active. This constitutes the selection of the initial element. The competitive filter unit for /k/ then inhibits the response unit for /k/ in the middle

layer, leading to a pattern in which /æ/ is most active, then /t/, and then /k/. The competitive filter is then reloaded and selects /æ/; /æ/ is then inhibited, leaving /t/ as the most active, and so on. Probably the most important difference between the competitive-queuing model and forward-lateral-inhibition models is that the competitive-queuing model separates the activation gradient in the middle layer from the competitive interactions that take place in the third layer. This enables the model to retain the anticipatory activation of future units (in the middle layer) while it is throwing all of the activation to the present unit (in the competitive filter layer). Moreover, anagrams and repeated elements are not a problem. The plans for *cat*, *act*, and *tack* can create different gradients among the response units /k/, /æ/, and /t/ because the gradients are determined by connections between the plans and the response units, not by lateral inhibition among the response units. The variation in each plan itself over time handles the repeated element problem (Houghton, 1990).

In sum, the competitive-queuing model effectively turns on the *present* by making it the most activated response unit, given the signal from the plan, and by inhibiting all non-*present* units in the competitive filter. It turns off the past by the architectural assumption of inhibition from each competitive filter unit to each response unit. Finally, it prepares the *future* by excitatory connections from plan units to response units.

Novel serial orders and frame-based models. The models discussed thus far can store a sequence and reproduce it when called upon. So, a phrase such as *the big dog* can be produced if it was previously encountered and a suitable plan was stored. However, the language-production system must be able to order novel sequences during grammatical encoding and, to a lesser extent, during phonological encoding. We can assemble a *tiny aardvark* even though we have never heard or said it before because we know the categories of the words (determiner, adjective, noun) and we know how those categories should be ordered. Similarly, given the phonemes /ŋ/, /æ/, and /m/, English speakers know that there is only one way to order them, even if they have never encountered the string *mang* before.

Clearly, serial-order plans need to be extendable to novel sequences. Psycholinguistic models of production have addressed this need by including rule-governed *frames* as components to plans (Bock, 1982; Dell, 1986; Garrett, 1975; Levelt, 1989; MacKay, 1972; Shattuck-Hufnagel, 1979; Stemberger, 1991). A frame is a sequence (or hierarchy) of categorically specified slots. A syntactic frame would contain syntactic categories (e.g., Determiner-Adjective-Noun), and a phonological frame would have phonological categories (e.g., Onset-Nucleus-Coda). The serial order of a sequence is determined when elements are inserted into the slots. We should make clear that frames may be necessary, but are not sufficient for generating novel serial orders. The frame contains knowledge about how to order a novel set of elements; it does not uniquely determine the elements.

A good example of an activation-based model that includes frames is the node structure theory of MacKay (1982, 1987). The theory has been developed for both the order of words in sentences, the *phrasal system*, and for the order of sounds within words, the *phonological system*. Figure 3 illustrates the theory. Consider the sequence of words *the cat*. The relevant content

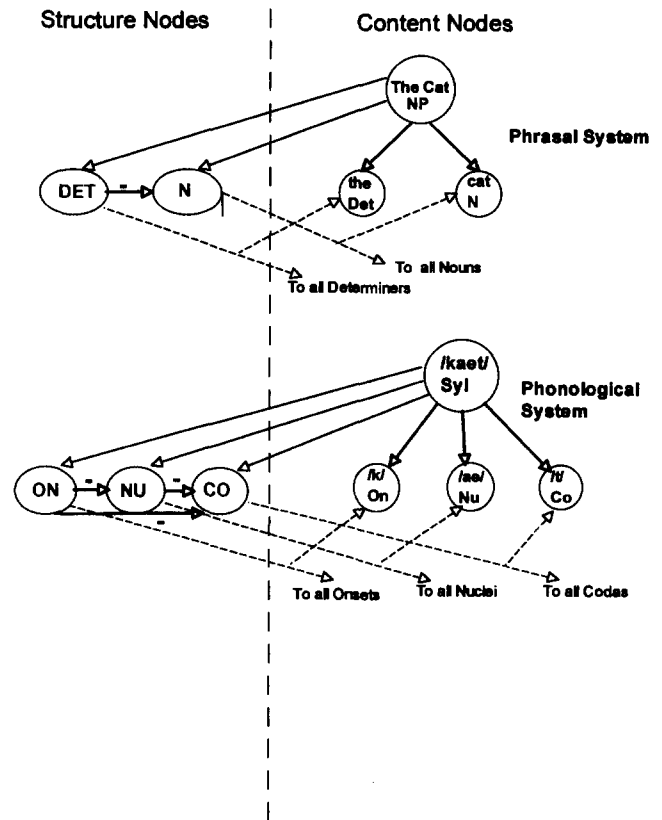


Figure 3. The phrasal and phonological systems in the node structure theory. Inhibitory connections are labeled by minus signs and excitatory connections are unlabeled. Content nodes are circles, and structure nodes are ovals. Self-inhibition is not shown. NP = noun phrase; DET = determiner; N = noun; Syl = syllable; ON = onsets; NU = nuclei; CO = codas.

nodes consist of nodes for the words *the* and *cat* and a phrasal node that we informally label *the cat* (for the complex concept of *cat-sing-def*). There are structural nodes for *Det* (for determiner) and *N* (for noun). The phrasal node has excitatory connections to *the*, *cat*, *Det*, and *N*; *Det* has positive connections to all determiners, including *the*; and *N* has positive connections to all nouns, including *cat*. An important feature of the model is that order is stored only in the structural nodes. Specifically, *Det* inhibits *N*, representing the rule that determiners come before nouns in noun phrases. These structural nodes and their connections constitute the frame. Note that the node structure theory is similar to the forward-lateral-inhibition models in that order is represented by lateral inhibition. However, the node structure theory differs by separating structure, where order is stored, from content, which derives its order from the structure.

More specifically, the node structure theory works like this: The production of *the cat* starts with the activation of *the cat*. Although many such phrasal nodes will already be stored in memory, one can assume that they can be created by the conjunction of existing features. The phrasal node then delivers some activation to *the*, *cat*, *Det*, and *N*. The structural nodes also receive input from a separate set of *timing nodes* (not shown in the figure), which control the rate of production. This

results in a large buildup of activation in *Det* and *N*, and because *Det* inhibits *N*, only *Det* reaches a special "activation threshold," causing it to sustain its activation at a high level. *Det* then sends a large amount of activation to all determiners. The content node for *the* thus obtains activation both from *the cat* and from *Det*, sufficient for it to pass the activation threshold. This constitutes the filling of the *Det* slot with the word *the*. After a node has reached activation threshold and has had its activation level sustained for a period of time, it enters a period of self-inhibition. When *Det* becomes self-inhibited, the next structure node *N* then reaches threshold. The *N* sends activation to all nouns and so activation converges on *cat* causing it to reach threshold, effectively filling the *N*-slot with *cat*.

The ordering of the phonemes of a word such as *cat* uses the same kind of frame-based mechanism in the node structure theory. Here the content nodes might include the syllable *kæt*, and the phonemes /*k*-onset, /*æ*-nucleus, and /*t*-coda, and the structural nodes would be *stressed syllable*, *onset*, *nucleus*, and *coda*. As in the phrasal system, order is represented by forward lateral inhibition between the structural nodes. By activating the syllable /*kæt*/, eventually the sequence /*k*-onset, /*æ*-nucleus, and /*t*-coda would be produced.⁵

The node structure theory and other activation-based models with frames (e.g., Berg, 1988; Dell, 1986, 1988/1989; Eikmeyer & Schade, 1991; Gasser, 1988; Gupta & MacWhinney, in press; Hartley & Houghton, 1996; Levelt, 1989; Meijer, 1994; Roelofs, 1992; Schade, 1992; Stemberger, 1985, 1990, 1991) have two advantages over the models without frames. First, as we explained before, they are consistent with our abilities to order novel sets of elements. Second, there is a fairly glaring empirical difficulty with the models without frames as theories of phonological encoding (Hartley & Houghton, 1996; MacKay, 1987). In the forward-lateral-inhibition and competitive-queuing models, the predicted pattern of misorderings of phonemes involves rearrangements of the phonemes within a word. So, /*kæt*/ would be predicted to slip to /*tkæ*/, /*ækt*/ or /*ktæ*/. Unfortunately, these kinds of errors never happen. Most commonly, phonological errors on a syllable like /*kæt*/ involve the substitution of sounds from nearby syllables, where the substituting sound emerges in the same syllabic position and where the resulting string of sounds is phonologically acceptable (Boomer & Laver, 1968; MacKay, 1970). For example "the cat is chasing . . ." results in "the chat . . ." or "the cass . . ." and Error Examples 1–3 discussed previously. Categorically specified frames derived from the phonological patterns in the language, together with a mechanism that inserts sounds from particular categories into frame slots, provide for exactly the right kind of phonological errors (Dell, 1986; MacKay, 1982; Shattuck-Hufnagel, 1979; Stemberger, 1991).

Frame-based, spreading-activation approaches to serial order in language, such as the node structure theory, are numerous and diverse. Consequently, we will not review them all other than to note some dimensions of variation. The principal way that these theories vary concerns their assumptions about particular levels of processing. For example, the node structure theory has content nodes for phrases, words, morphemes, syllables, syllabic constituents, phonemes, and features. Other theories (e.g., Gupta & MacWhinney, in press; Gupta, 1995; Roelofs, 1992) have fewer levels among content items. Theories also differ greatly in how the frames work. Schade (1992), for exam-

ple, uses excitatory forward connections from frame element *i* to element *i* + 1 to store the order of categories, whereas the node structure theory uses forward lateral inhibition. Another dimension of difference concerns how many different frames are needed for particular processing levels. The theories of Dell (1986), Gupta (1995), and Hartley and Houghton (1996) used a single frame at the phonological level, one that was sufficiently flexible to handle all possible syllables. Others (e.g., Dell, 1988/1989; Stemberger, 1991) hypothesize the existence of several word-shape frames, one frame for each possible patterning of consonants and vowels in a word. Finally, frame-based models have differed on whether elements send activation to plans, as well as plans to elements. For example, Dell (1986), Harley (1993), Stemberger (1985), as well as the node structure theory, allow for two-way interaction between word plans and phonological elements. Roelofs (1995) does not.

In summary, frame-based approaches are both varied and complex. In fact, the added complexity of separate frames must be considered their main weakness, when compared to the models lacking frames. Individual models, of course, have strengths and weaknesses related to particular linguistic and psycholinguistic assumptions. For our purposes, though, frame-based, spreading-activation approaches share important general principles: The activation of the *present* is achieved by input from both a structural frame node and a higher level content node (e.g., in the node structure theory the phrasal node for word sequences or a word or syllable node for phoneme sequences). The *future* is prepared because input from the higher level content node is sent to all elements dominated by that node. Finally, the *past* is deactivated by a period of self-inhibition after a node has reached activation threshold.

PDP recurrent network models. For the sake of completeness, we mention the parallel-distributed-processing (PDP) recurrent network models of serial order (e.g., Cleeremans & McClelland, 1991; Dell et al., 1993; Elman, 1990; Jordan, 1986). These models map between a plan representation and a sequence of outputs by means of a changing contextual representation that is derived from the network's previous state of activation and a set of hidden units that computes a nonlinear combination of the plan and the context. The connection strengths among the units are learned by back-propagation or related supervised algorithms. However, because the configuration of excitation and inhibition is determined by the particular sequences that are learned, rather than by architectural assumptions as in the previous models, it is not so easy to characterize. We can say, though, that if the learned sequences do not often exhibit immediate repetition of units (which is true for word or phoneme sequences), then the network should develop a general mechanism for turning off each *past* representation.

The PDP approaches have been used successfully to account for coarticulation (Jordan, 1986) and some facts about phonological speech errors (Dell et al., 1993). It is unclear, though, whether these models, insofar as they lack explicit frames, can account for the speech error patterns that have motivated the frame-based models (see Dell et al., 1993; Hartley & Houghton, 1996).

⁵ To simplify the presentation, we are ignoring the node structure theory's use of rime constituents and phonological features.

A General Model

All of the models discussed share four characteristics: (a) Learning involves the creation and strengthening of excitatory connections from plans to elements; (b) the abilities to activate the *present* and to prepare the *future* are achieved by these connections: Consequently, the activation of the *present* and the preparation of the *future* are linked; (c) the turn-off function is effective: That is, regardless of how activated the *past* is, it can be quickly brought down to an absolutely low level; and (d) the turn-off function does not require new learning.⁶ These characteristics reflect what we earlier called the throw-away principle: Getting rid of what is already activated is easier than activating what is needed. People do not need to learn how to turn off the *past* the way they need to learn to activate the *present* and prepare the *future*.

These generalizations from the models allow us to account for the anticipatory practice effect. Compare Figure 4, which shows the hypothetical activations of the *past*, *present*, and *future* under good circumstances, that is, when a sequence is well learned, and under bad circumstances, when it is not. The idea is that practice increases the activation of both the *present* and *future*, and does little to the *past*. Learning a sequence in all of the models involves the strengthening of connections between plans and their elements. This increases the potential to activate the *present* and the *future*. Practice does not, however, have much effect on the deactivation of the *past*, because this function is not achieved through the learning of the plan. Given that anticipations reflect the relative activation of the *present* and *future*, and perseverations, that of the *present* and *past*, the anticipatory proportion should increase with practice.

We now work out the empirical consequences of this account of the anticipatory practice effect. Rather than derive predictions from each model, though, we offer a new model that exhibits many of their common characteristics and is sufficiently simple in its size, activation function, and selection decisions to allow for formal derivations.

The model contains four nodes, one for the plan, and one each for *past*, *present*, and *future*. The plan connects to each element with a long-term positive weight, w , the value of which

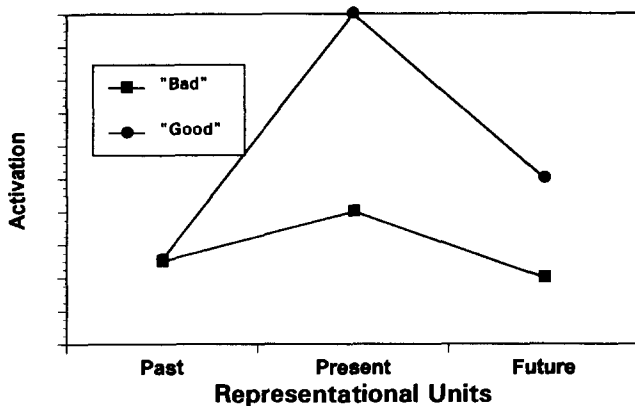


Figure 4. A mechanism for the anticipatory practice effect. The difference between good and bad circumstances occurs in the amount of activation of the *present* and *future*, not in the *past*.

is a product of learning. These weights are temporarily altered during sequence production, the alteration depending on when the element in question is scheduled. The weight from plan to *past* becomes 0; the weight to the *present* remains at w ; and the weight to the *future* becomes bw , where b is a positive fraction representing anticipatory activation. One can view these modifications as inputs to *past*, *present*, and *future* of 0, 1, and b , respectively, that combine multiplicatively with the inputs of w from the plan.

We adopt two perspectives on these short-term weight alterations. First, we view them as corresponding to various mechanisms in the particular serial-order models. In the forward-lateral-inhibition models and the node structure theory, the multiplicative input of 0 to the *past* is part of the self-inhibitory mechanism. In the competitive-queuing model, it is part of the competitive-filter-to-response-units inhibition. Parameter b corresponds to the joint effects of plan-to-element excitation and forward-lateral inhibition in the forward-lateral-inhibition models, to the extent of the activation gradient set up by plan-to-response-unit connection weights in the competitive-queuing model, and to the combination of priming from the higher level content node and upcoming frame nodes in the node structure theory.

Second, we have our own perspective on the weight modifications in the general model. For reasons mentioned before, we are persuaded that structural frames are needed for serial order in language. The order of a set of linguistic units must come, at least in part, from knowledge of how their categories are ordered. Let us therefore assume that *past*, *present*, and *future* are content nodes that belong to the categories X, Y, and Z, respectively. The frame consists of structural nodes for X, Y, and Z, specified for that order. We propose that the frame works by sending a signal of 0 to all members of the "past" category (here, X), a signal of 1.0 to all members of the "present" category (Y), and signal of b to all members of the "future" category (Z). These signals are multiplicative with other inputs to the content nodes resulting in net temporary weights of 0 to *past*, w to *present*, and bw to *future*. Notice that this perspective does not say how the order of categories within the frame itself is stored, but it is easy enough to build the competitive queuing, recurrent net, or the forward-lateral-inhibition mechanism (as MacKay does in the node structure theory) into the frame itself, so that the frame signals to the content nodes are ordered appropriately. We assume the existence of such a mechanism. Figure 5 illustrates the model, under the interpretation that the weight alterations derive from the frame.

The model uses assumptions about the initial activation of the nodes and a spreading activation rule to compute activations of each element. The relative activations of *past*, *present*, and *future* are then used to determine the probabilities of error. Specifically, an input of 1.0 is given to the plan element, and the *past* is assumed to have its activation initialized to some small positive value, c , reflecting its previous activation and the result

⁶ This is not strictly true in the recurrent net PDP models. But it is largely so because learning a new linguistic sequence would make use of previous learning involving its elements. The connections responsible for turning off a unit after it has been produced would be mostly in place prior to the learning of a new sequence because linguistic units are almost never immediately repeated.

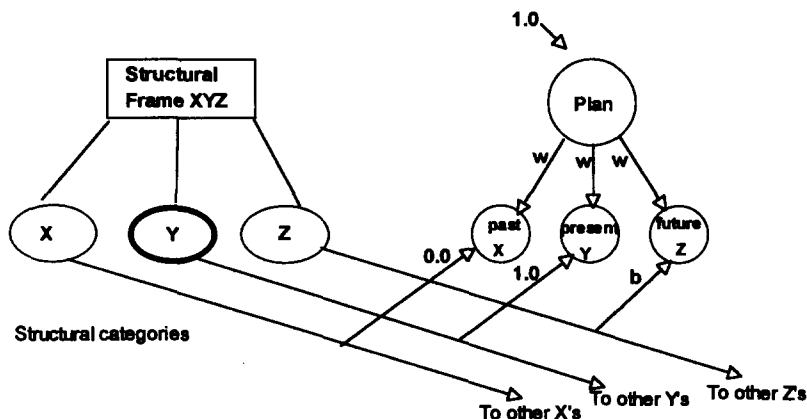


Figure 5. A general model of serial order in language. Content nodes (right) for *past*, *present*, and *future* receive inputs from the plan along connections of weight w and inputs from structural nodes (left). The figure indicates that structural node Y is the present category.

of an effective turn-off mechanism. Activation spreads for n time steps according to a linear rule:

$$a(j, t) = a(j, t - 1)(1 - d) + \text{Input}(j, t). \quad (1)$$

Where $a(j, t)$ is the activation of node j at time step t , d is the decay rate, and $\text{Input}(j, t)$ is the net input to node j given by

$$\text{Input}(j, t) = \sum_i w(i, j)a(i, t - 1), \quad (2)$$

where $w(i, j)$ is the weight from node i to node j .

In computing the activation of the elements, we consider two kinds of input to the plan, a *pulsed* and a *clamped* input. The pulsed input assumes that, for each element in the sequence, the plan's activation is set to 1.0. However, the plan is subject to decay like any node and, hence, must get a new pulse for each element. The clamped input assumes that the plan's activation is held at 1.0 throughout the production of the sequence. Our consideration of these two cases arises because both kinds of input have been proposed in serial-order models (e.g., Dell, 1986, is pulsed; most others are explicitly or functionally clamped). As we will show, though, both kinds of input lead to the same principal predictions.

The absolute activation levels after n time steps for the *past*, *present*, and *future* after pulsed and clamped input to the plan are given by Equations 3–5 and Equations 6–8, respectively. (Proofs, by induction, are shown in Appendix C.)

$$\text{(pulsed)} \quad a(\text{present}) = nw(1 - d)^{(n-1)} \quad (3)$$

$$\text{(pulsed)} \quad a(\text{past}) = c(1 - d)^n \quad (4)$$

$$\text{(pulsed)} \quad a(\text{future}) = nbw(1 - d)^{(n-1)} \quad (5)$$

$$\text{(clamped)} \quad a(\text{present}) = w(1/d)[1 - (1 - d)^n] \quad (6)$$

$$\text{(clamped)} \quad a(\text{past}) = c(1 - d)^n \quad (7)$$

$$\text{(clamped)} \quad a(\text{future}) = wb(1/d)[1 - (1 - d)^n] \quad (8)$$

What is important for determining order is the relative activation of the nodes. In this respect it is useful to define the perseverative

tendency, P , as the ratio of $a(\text{past})$ to $a(\text{present})$ and the anticipatory tendency, A , as that of $a(\text{future})$ to $a(\text{present})$. These are given by Equations 9–11.

$$P = \frac{c(1 - d)}{nw} \quad \text{(pulsed)} \quad (9)$$

$$P = \frac{cd}{w[(1 - d)^{-n} - 1]} \quad \text{(clamped)} \quad (10)$$

$$A = b \quad \text{(both pulsed and clamped)} \quad (11)$$

We further assume that the element that is output after n time steps is determined by a stochastic decision rule (e.g., McClelland & Rumelhart, 1981) based on a transformation of the relative activation levels. The relative activation, R , of an element is the proportion of the total activation in the elements that that element has. These are given by Equations 12–14.

$$R(\text{past}) = \frac{P}{P + A + 1} \quad (12)$$

$$R(\text{present}) = \frac{1}{P + A + 1} \quad (13)$$

$$R(\text{future}) = \frac{A}{P + A + 1} \quad (14)$$

The probability of selection of a particular element is then given by Equation 15.

$$\text{probability of selection for element } i = \frac{S(i)}{\sum_j S(j)}, \quad (15)$$

where $S(i) = e^{\mu R(i)}$, and μ is a parameter that determines how rapidly response strength grows with relative activation. The probability of a correct selection is therefore given by Equation 15 when i is the *present*. Anticipations and perseverations are analogous where i is the *future* and *past*, respectively.

Model Characteristics

Anticipatory practice effect. Practicing a sequence is assumed to increase w . The anticipatory practice effect then arises because the perseverative tendency, P , decreases as w increases (Equations 9 and 10), while the anticipatory tendency, A , (Equation 11) is independent of w . Consequently, the proportion of errors that are anticipatory increases as connection weights get larger. Figure 6 shows an example set of parameters. Increasing w makes both perseverative and anticipatory errors less likely, but the effect on perseverations is much stronger, resulting in a higher anticipatory proportion for larger values of w . To put it less formally, the activation of the *present* and *future* are both linked to practice and so increase together. The activation of the *past* is not linked to practice and so its relative influence diminishes with practice.

The effect of speech rate. The model makes two specific claims about speech rate, which is represented by n , the number of time steps passing during the retrieval of each element. The first of these is unsurprising: As n increases (as speech slows), errors become less likely. This claim is most definitely true empirically and gives the model a general basis for speed-accuracy trade-offs in production (e.g., Dell, 1986; MacKay, 1987). The second is more interesting: As n increases, the anticipatory proportion increases. This will be called the *anticipatory speech-rate effect*. The basis for the prediction can be seen by noting that n is similar to w in that P decreases as n increases, while A is independent of n . The actual relation between n and P differs in the clamped and pulsed cases and is not simple (see Equations 9 and 10). Nonetheless, any increase in time makes the perseverative tendency smaller, while leaving the anticipatory tendency unaffected. Consequently, the anticipatory proportion increases as speech slows. Earlier, we suggested that there was some evidence for this relation from a deadline task (Dell, 1990), and one of the purposes of Experiment 3 will be to gather further data on this question:

The general anticipatory effect. One of the most striking characteristics of the model is that the anticipatory proportion

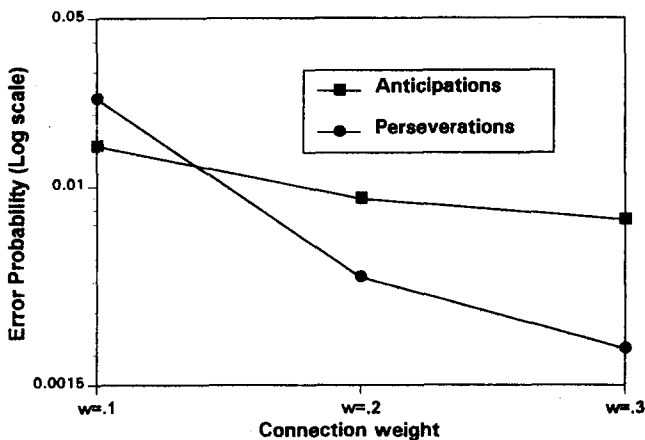


Figure 6. Anticipation and perseveration probabilities as a function of connection weight (w) in the general model when number of time steps (n) = 4, decay (d) = .5, residual activation rate (c) = .3, anticipatory activation rate (b) = .3, and response parameter (μ) = 10.

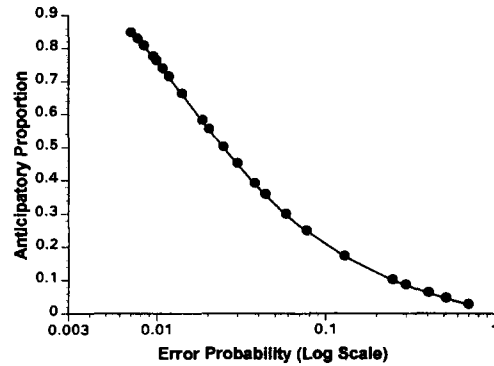


Figure 7. The general anticipatory effect in the model. The graphed points reflect a variety of values of connection weight (w) (.1-.3), number of time steps (n) (1-4), residual activation rate (c) (.2-.4), and decay (d) (.4-.6) with anticipatory activation rate (b) (.3) and response parameter (μ) (10) held constant.

can be predicted solely from overall error rate, provided that parameter b does not vary. That is, you do not need to know w , n , c , or d to predict the extent to which errors are anticipatory; you only need to know the overall error rate. This is illustrated in Figure 7, which shows the relation between the $\log(\text{overall error probability})$ and the anticipatory proportion for various combinations of values of w , n , c , and d . Regardless of the combination, the predicted anticipatory proportion falls on the line. The actual place on the line is determined by P , which is a computed combination of w , n , c , and d .

The relation between overall correctness, regardless of parameter values, and anticipatory proportion was called the general anticipatory effect. It is, we believe, a far-reaching prediction from the model, because there are many variables that can affect overall error rate. Our claim is that any error-prone circumstance will be associated mostly with perseverative errors, while relatively error-free cases will have relatively more anticipatory errors. Specifically, any variation in overall error rates that reflects the knowledge of individual sequences (w), time (n), activation decay (d), or residual activation in the *past* resulting from its turn-off (c), will be associated with a predictable anticipatory proportion. This relation does not, however, hold for variation that can be attributed to parameter b , the signal sent from the frame to the *future* that mediates anticipatory activation. Varying b actually shifts the line relating error rate to anticipatory proportion that is defined by variation among the other four parameters, rather than leading to movement along the line. Given our conception of b as a feature of the structural frame, we see its magnitude as reflecting the grammatical or phonological structure of the intended utterance (e.g., the kind or size of phrases) or even the nature of the language itself (e.g., the location of the heads of phrases in the language). One should note, as well, that the general anticipatory effect also does not hold for variation in parameter μ , which governs the decision process.

Experiment 3 varies both practice and speech rate in an attempt to test the predictions about speech rate and the general anticipatory effect. We assume that practice effects w and speech rate affects n . It is predicted that increased practice and a slower speech rate will lead to fewer errors and to a greater anticipatory

proportion. Most importantly, the anticipatory proportion should be predictable from overall error rate, regardless of the combination of practice and speech rate that led to that error rate. Specifically, the model expects an approximately linear relation between log(error rate) and anticipatory proportion, provided that the anticipatory proportion is not extreme (see Figure 7).

Experiment 3: General Anticipatory Effect

Method

Materials. The same 16 phrases were used. On the basis of a pilot experiment using a variety of speech rates, the phrases were placed in two groups of 8, such that each group of phrases was approximately equally error prone.

Participants. Forty-three students from the Illinois participant pool were tested. Two were eliminated because of failure to keep up with the metronome and one because of equipment failure, resulting in 40 official participants. Half of these produced the 8 phrases of Group 1 at a slow rate (1.73 beats/s) and the 8 phrases of Group 2 at a fast rate

(2.83 beats/s). The other 20 students had the assignment of phrase groups to speech rates reversed.

Procedure. Each phrase was practiced for eight trial blocks, using much the same procedure as in Experiment 1. There were two differences. First, there was within-subject manipulation of speech rate. Each trial block involved practice at a fast rate for some phrases and a slow rate for others. Second, participants were asked to leave "at least two beats" between the two repetitions of each phrase. Experiment 1 had asked them to leave exactly four beats, which may have caused the speakers to covertly produce the digits "one, two, three, four," an undesirable source of phonological interference.

Results and Discussion

Tape recordings were transcribed by Lisa K. Burger, and errors were coded as before by Gary S. Dell. Table 2 shows the number of errors in each category as a function of speech rate and practice block. Both rate and practice had powerful effects on the overall error rate. The fast and slow rates were associated with 1,551 and 849 errors, respectively. Practice reduced errors

Table 2
Error Totals, Log-Log Slopes, Anticipatory Proportions (APs),
and Correlations for Experiment 3

Category	Trial								Total	Slope	r
	1	2	3	4	5	6	7	8			
Slow rate											
Sound	154	110	80	69	68	50	36	35	602	-.710	-.97
SA	40	33	21	26	25	20	12	12	189	-.539	-.88
SP	70	42	33	25	25	15	11	15	236	-.834	-.96
SE	24	16	11	11	10	7	7	3	89	-.799	-.91
SAP	20	19	15	7	8	8	6	5	88	-.706	-.92
Word	64	45	25	19	17	16	13	9	208	-.903	-.98
WA	21	18	8	7	8	3	7	5	77	-.778	-.86
WP	22	17	13	7	8	11	3	4	85	-.842	-.86
WE	15	4	4	2	1	2	2	0	30		
WS	6	6	0	3	0	0	1	0	16		
Other	17	6	4	3	4	2	0	3	39		
Total	235	161	109	91	89	68	49	47	849	-.774	-.98
Anticipatory proportion	.40	.46	.39	.51	.50	.47	.56	.47			
Fast rate											
Sound	265	197	171	136	142	110	87	101	1209	-.507	-.97
SA	66	36	49	33	41	30	30	36	321	-.300	-.78
SP	108	85	76	53	47	48	29	40	486	-.574	-.94
SE	38	41	22	25	23	17	17	14	197	-.496	-.91
SAP	53	35	24	25	31	15	11	11	205	-.723	-.90
Word	81	39	43	28	18	19	26	21	275	-.659	-.91
WA	24	6	12	9	5	6	8	5	75	-.572	-.75
WP	33	25	19	9	9	10	16	11	132	-.572	-.81
WE	12	5	8	9	3	2	1	3	43	-.920	-.77
WS	12	3	4	1	1	1	1	2	25	-1.092	-.94
Other	20	16	5	7	8	4	4	3	67	-.880	-.91
Total	366	252	219	171	168	133	117	125	1551	-.549	-.99
Anticipatory proportion	.39	.28	.39	.40	.45	.38	.46	.45			

Note. For categories with zero entries, no slopes or correlations were computed. SA = sound anticipation; SP = sound perseveration; SE = sound exchange; SAP = sound anticipation-perseveration; WA = word anticipation; WP = word perseveration; WE = word exchange; WS = word substitution.

from 14.3 per 100 words in the first block to 4.9 per 100 words for the eighth block for the fast rate. For the slow rate, these rates were 9.2 per 100 and 1.8 per 100, respectively. Figure 8 shows that the effect of practice was well captured by a linear log/log relation. Interestingly, the slope was marginally steeper for the slow rate than the fast rate, $t(15) = 1.94, p < .10$.

Given that the two independent variables had expected and strong effects on error rate, we now turn to aspects of the data that impact the model. First, the expectation of an anticipatory practice effect was examined. The number of anticipations and perseverations for each trial block was determined separately for each phrase (pooled across speech rate and word-versus-sound errors), and the data from phrases with zero entries were combined as before, with the result that there were nine "super-phrases." The average slope for the perseverations was significantly steeper than that for the anticipations, with the mean difference being .307, $t(8) = 3.00, p < .025$. This is a smaller difference than that found in Experiment 1 (.513), but it adequately replicates the anticipatory practice effect.

Next, we tested the prediction of an anticipatory speech-rate effect, that the anticipatory proportion will be greater with a slower speech rate. The AP was determined for each phrase for each rate, pooling across practice block and error categories. In support of the prediction, the AP was significantly greater for the slow rate (.45) than the fast rate (.39), $t(15) = 2.78, p < .025$.

The most important test of the model concerns the general anticipatory effect. Can the AP for each of the 16 conditions of the experiment, 2 (rates) \times 8 (practice blocks), be predicted solely from overall error rate, that is, without regard for the combination of variables that led to that error rate? Figure 9 plots AP as a function of the log(10) of error rate per word and gives the resulting regression line. As predicted by the model, AP decreases linearly with the log of the overall error rate; slope = $-.193$, intercept = $.185$, $r = -.73, p < .001$. Each condition results in an AP that is close to the line, and deviations appear to be nonsystematic. Another way to illustrate the finding is to note that if the 8 points from the fast conditions and those from slow conditions are fit separately, the regression lines are very

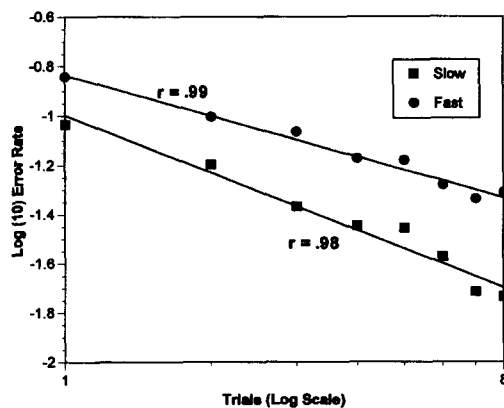


Figure 8. Overall error rate in Experiment 3 as a function of practice and speech rate.

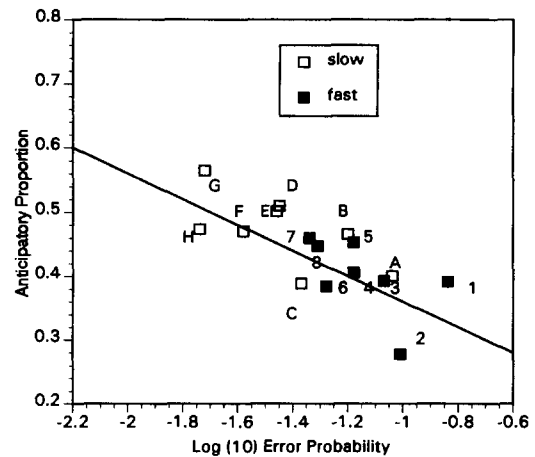


Figure 9. The general anticipatory effect illustrated for variation over speech rate and practice in Experiment 3. The numbers 1–8 and the letters A–H indicate practice Blocks 1–8 for the fast and the slow rates respectively.

similar (fast conditions: slope = $-.154$, intercept = $.248$; slow conditions: slope = $-.193$, intercept = $.178$).

The data presented in Figure 9 support the general anticipatory effect with respect to two manipulations—practice and speech rate. These manipulations are tied to specific parameters in the model, namely w and n , which work together to determine both how error-prone the system is and the extent to which errors are anticipatory.

Another factor affecting errors is individual speakers. There were large individual differences in error probability, particularly with the slow speaking rate, where error totals pooled over blocks ranged from 3 to 59 errors per speaker. This variation allows for another test of the general anticipatory effect. According to the model, if individuals differ in connection weights (w), residual activation in the past (c), decay rate (d), or a combination of any of these, and if this variation is large compared to individual variation in parameter b , the strength of the signal from the frame to the future, then the more error-prone speakers should have relatively more perseverations. The general anticipatory effect gives us this prediction without us having to know exactly which parameters are associated with individual differences—only that the differences are primarily associated with the cluster of parameters in Equations 9 and 10.

The prediction about individual differences was tested by examining each speaker's overall error rate and his or her AP. Because there was more individual variation in the slow speech-rate condition than in the fast one, error totals were kept separate for fast and slow rates for each speaker. Three of the resulting 80 speaker-condition cases were removed because they lacked anticipations and perseverations, making for 77 data points. So that the points only reflect individual variation, each point was adjusted to remove the known correlation between AP and speech rate. If this covariation is not removed, the predicted relation between AP and error rate across speakers could emerge for the simple reason that faster rates have more errors and lower APs. The resulting scatter plot on the adjusted data is shown in Figure 10. In support of the general anticipatory effect, the more error-prone participants had lower APs, $r = -.27, p$

< .05. The points scatter themselves around a regression line that is similar to the line derived by variation in practice and speech rate (slope = $-.188$, intercept = $.214$). The correlation, though, is much lower because each point is based on many fewer errors and, hence, noise is more dominant.

In summary, Experiment 3 replicated the anticipatory practice effect, found an anticipatory speech-rate effect, and most importantly, provided evidence for a general anticipatory effect. Variation in practice and speech rate, and individual differences, all led to the same relation between the log of overall error rate and the anticipatory proportion. These results provide good support for the principles behind the proposed serial-order model.

In the remainder of the article, we consider the scope of the general anticipatory effect, both with respect to other variables' influence on error rate and with respect to the universality of the functional relation found in Figure 9.

Is the General Anticipatory Effect General?

The relationship between overall error rate and the anticipatory proportion found in Experiment 3 should apply to other data sets as well. First, we consider whether the line presented in Figure 9 also fits the data from Experiments 1 and 2. Then we examine data sets from outside our experiments that used the same error categories that we did. Finally, we consider the potential for testing the general anticipatory effect in other domains, both linguistic and nonlinguistic.

Other Data Sets

Experiments 1 and 2. The line relating AP to log(overall error rate) derived from Experiment 3 should apply to the data from the earlier experiments. Experiment 3 found that variation due to practice, speech rate, and individual differences generates points along this line. According to the model, this is to be expected for any variation in the key parameters, the ones in Equations 9 and 10. Speech rate and practice are identified with parameters n and w , respectively. Individual differences are harder to identify with a single parameter. Nonetheless, the

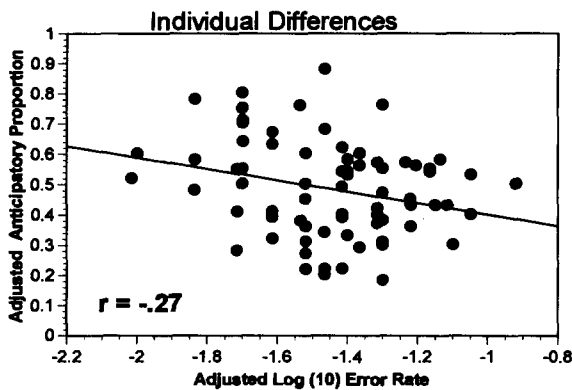


Figure 10. The general anticipatory effect illustrated for individual variation in Experiment 3. The error rates and anticipatory proportions (APs) are adjusted to remove the existing correlation between speech rate and AP. Each point represents a single participant for a particular speech rate.

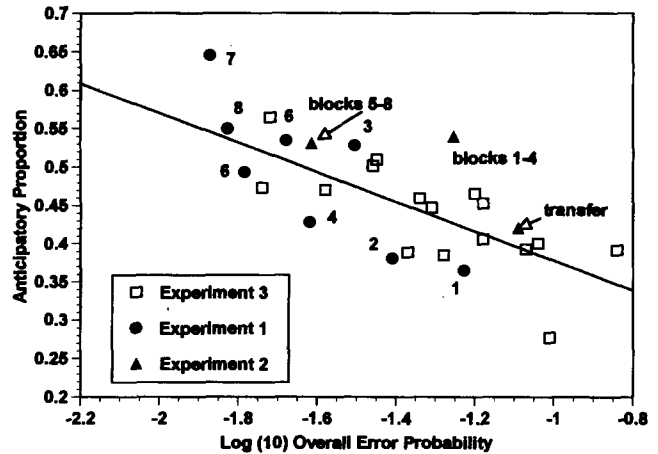


Figure 11. The general anticipatory effect for the conditions of Experiments 1–3. The numbers beside the filled circles label the practice blocks of Experiment 1. The regression line shown is based on the data from Experiment 3.

fact that individual AP was, to some extent, predictable from individual error rate suggests, within the context of the model, that some proportion of individual variation can be associated with the key parameters. Given these findings, the line in Figure 9 should fit the other experiments because the main differences among the experiments is their participants and speech rates. Importantly, the same phrases were used and the error-categorization scheme was the same. Any differences in overall error due to the particular speakers, speech rates, and interactions of these factors with practice, should still lead to a predictable AP.

Figure 11 takes the line derived from Experiment 3 and adds to it the relevant points for each practice block in Experiment 1. Experiment 2 is also included, but because of the many fewer errors in this experiment, Blocks 1–4 and Blocks 5–8 are pooled; the transfer block is kept separate. The points contributed by Experiments 1 and 2, by themselves, are associated with a line that is reliably different from zero ($r = -.65$, $p < .05$) and which is quite similar to the one from Experiment 3 (slope = $-.210$ compared to $-.193$; intercept = $.168$ compared to $.185$). At least within the confines of the materials and methods of these experiments, overall error rate, by itself, accounts for the bulk of the variation in AP, providing good support for the general anticipatory effect.

Natural speech samples. How far can we push the support for a general anticipatory effect? The strongest possible claim is that the particular line relating AP to error rate shown in Figures 9 and 11 applies to any speech sample from any persons or circumstances. Of course, the same definitions of error categories must be employed, and the sample should be based on a sufficient number of errors. There are two analyses of tape-recorded spontaneous speech samples that used the same error categories and coding that were used in the experiments, the London–Lund corpus of normal speech and the speech of the jargon-aphasic patient F.L., as analyzed by Schwartz et al. (1994). Because these data sets are complete samples, rather than just recordings of errors, they allow for the computation of overall error rate per word. Hence, it can be determined whether their observed AP is close to what would be predicted

from the error rate. Figure 12 shows the line derived from Experiment 3 expanded to accommodate the range of the error rates from the London–Lund corpus (.0007 per word) and patient F.L. (.067 per word). The obtained APs of .75 and .32, respectively, allow these points to be added to the figure.

Figure 12 also shows points from six other spontaneous speech samples. With one exception, these are based on large speech–error collections (not tape-recorded) and, hence, are not associated with any direct estimate of error rate, only with the relative proportions of error types. The collections of Shattuck-Hufnagel and Klatt (1979), Nooteboom (1969), Meringer (1908, as analyzed by Nooteboom, 1980), and Stemberger's (1989) adult corpus are all derived from nonaphasic adult speakers. Their error rates would be expected to be similar to that of the London–Lund corpus and, hence, these are graphed as if they have the same rate. Their APs are, respectively, .75, .81, .70, and .60.⁷ Stemberger's (1989) child corpus is derived from two children, ages 1–5, with the bulk of the observations coming at ages 2 and 3. Wijnen's (1992) analysis of recordings of two children, ages 2–3, found that the children's overall error rate (true speech errors as opposed to systematic errors) was around .005 per word, about seven times greater than that of adults, and that the AP was .55. If we assume that the overall error rate for Stemberger's children is the same as that of Wijnen's, then the AP associated with them (.47) can also be placed on the figure.

The natural error corpora provide good support for the general anticipatory effect. The spontaneous speech of normal adults exhibits a low error rate and anticipatory proportions that are reasonably close to the predicted value. Patient F.L.'s high error rate is expected to be associated with a low AP, and it is. For children, their accuracy is somewhat less than that of adults, as is their AP. The two points associated with Stemberger (1989), however, are a little lower than predicted. Stemberger noted that he found a few more perseverations than other researchers,

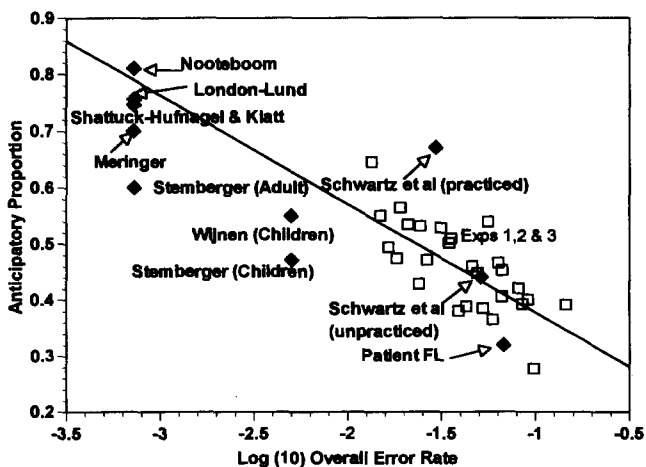


Figure 12. Extension of the general anticipatory effect to other error sets. Open squares represent points from Experiments 1–3 (see Figure 11). The labeled error collections are described in the text. The “practiced” and “unpracticed” conditions from Schwartz et al. (1994) correspond to Blocks 1–4 and 5–8, respectively. The regression line is based on the data from Experiment 3.

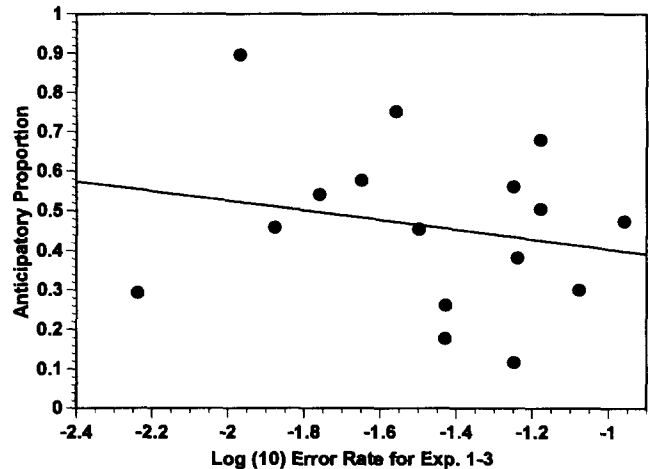


Figure 13. The general anticipatory effect for particular phrases. Each point represents the error rate and anticipatory proportion (AP) for one of the 16 phrases, pooled over Experiments 1–3. The regression line shown is based on these 16 points.

suggesting that error classification criteria may have differed among the corpora. All things considered, though, the AP of each corpus is well predicted by its overall error rate, using the same functional relation as was found in the experiments.

Variation due to individual phrases. The fact that the line generated from the experiments, which made use of a particular set of phrases, also fits error corpora from spontaneous speech samples raises the question of the influence of linguistic materials. Would error data from a very restricted class of utterances, for example, from a single phrase, also fall on the line? By pooling the data from all three experiments, it is possible to get enough data per phrase to address the question. Figure 13 shows the scatterplot relating log of overall error rate and the AP for each of the 16 phrases. The trend is as predicted—harder phrases tend to have lower APs—but the effect is weak and not reliable ($r = -.21$, $t < 1.0$). Is the effect small because of noise due to small N s (as was likely the case for individual participant variation)? On the contrary, the variation in both error rate (from .11 to .006 per word) and AP (from .11 to .89) reflects stable properties of particular phrases, rather than noise. The log (error rates) and APs for each phrase from Experiment 3 correlate highly with their counterparts from Experiment 1, $r = .74$ for error rate, $r = .69$ for AP. Moreover, for all but four of the phrases, the AP is based on at least 50 anticipations and perseverations. Hence, we cannot blame the low correlation between error rate and AP entirely on noise. Rather, speech samples from single phrases, although they may on average follow the general anticipatory effect, have idiosyncratic properties.

⁷ Because our experimental work, that of Schwartz et al. (1994), and that of Nooteboom (1969) classified incomplete anticipations along with anticipations, the incomplete anticipation categories of Shattuck-Hufnagel and Klatt (1979) and Stemberger (1989) have been included in their anticipation numbers. We do not claim that all incomplete anticipations are anticipations as opposed to exchanges. Our aim is simply to be consistent in classification.

Does the model, in fact, say that individual phrases' error rates and *AP* should fall on the line? It depends. The model predicts that the line applies universally to any speech sample provided that the sample's value of parameter *b* does not deviate much from that of the samples used in estimating the line. In the model, *b* is a property of the structural frames. When a signal of 1.0 is sent from the present frame category to all of its members, a smaller anticipatory signal of *b* is sent to all members of a future frame category. Because *b* emanates from the frame, it might vary with the linguistic structure of the utterance. For example, *b* could be influenced by the need to compute agreement or by argument structure or the location of phrasal heads, in short, by those aspects of language in which the appropriateness and form of a word may be governed by upcoming words. If particular utterances are associated with widely different values of *b*, variation in their error rates and *AP*s would not fall on a single line.

Nonetheless, there are ways that particular utterances can vary that are consistent with the general anticipatory effect. For example, utterances and their parts vary in familiarity, which is assumed to correspond to connection weights (*w*) among the relevant units. If differences in error rate among particular utterances are due to *w*, then one would expect *AP* to be predictable from error rate. The same point can be made for parameter *c*, the activation level assigned to the *past*. Any variation in *c* across utterances large enough to affect error rates, should also move *AP* along the line.

Thus the model provides some reason to expect error variation across phrases to follow the general anticipatory effect, but also allows that this effect might be weak. Aside from the model, there is another reason to expect particular phrases to deviate from the line. Consider the phrase "toy boat." The most likely error is a perseveration of the glide from "toy" to "boat" creating something like "toy boyt." If the phrase is "boat toy," one would likely get the same error "boyt," only this time it would be an anticipation. Another way to say this is that "boat" in the context of "toy" affords the production of the error "boyt." Whether the error is an anticipation or perseveration does not contribute as much to overall error rate for the phrase as the peculiarities of the sounds involved. In this case, the error-proneness of the phrase is a result of the similarity of the diphthongs /ow/ and /oy/, and what causes the replacement of /ow/ with /oy/ rather than vice versa is likely due to a property of this pair of sounds. Stemberger (1991) found that replacement asymmetries such as this arise when one sound is more marked than another, which is arguably true in this case.⁸ If a particular error's likelihood is strongly related to specific factors like these, whether the slip is anticipatory or perseveratory becomes less important. Consequently, any error sample based on a small set of materials, such as a single phrase, may have an unusual *AP* because of the relative location of similar sounds that exhibit replacement asymmetries. Thus, "toy boat" might have an *AP* near 0, and "boat toy" might be near 1, while their overall error rates might not differ much.

It appears that idiosyncratic properties of phrases, such as their structural properties and the relative locations of similar sounds that exhibit replacement asymmetries, limit our ability to test the general anticipatory effect with error data from very small sets of phrases. In general, we suggest the following conditions on tests of the actual function presented in Figure 12: (a)

The error sample should be large enough for the overall error rate and *AP* to be accurately estimated; (b) errors need to be categorized as we have done; in particular, incomplete anticipations are counted as anticipations; (c) for experimental data, there must be a range in the materials; as a rule of thumb, at least 12 phrases might be needed to reduce the likelihood that any sound replacement asymmetries are confounded with anticipation versus perseveration locations; and (d) the speech sample should be English. With regard to this last condition, we note that two points on Figure 12, those associated with Nootboom (1969) and Wijnen (1992), are based on Dutch, and one point, Meringer, is from German. Dutch and German, like English, are Germanic languages and are therefore relatively similar to English and each other. We also have data (not plotted on Figure 12) for a single French corpus based on Rossi and Peter-Defare's (1995) analysis. Here the *AP* is .65, which is within the range of that of corpora for Germanic languages. In general, though, languages differ greatly in their grammatical and phonological structures and, hence, would be expected to differ in the average value of *b*. Samples from a single language, if they are sufficiently large, should fall on a line, but each language or language group may have a different line.⁹

In summary, we feel we have sufficient evidence for the particular line shown in Figure 12 to predict that any speech sample adhering to the four conditions mentioned in the previous paragraph should produce a point on the line. Manipulation of either speaking conditions or speakers should result in an *AP* that is predictable from overall error rate.

Clinical populations. Our analysis can be applied to clinical syndromes associated with speech-production errors, for example, aphasia, Alzheimer's dementia, and some forms of schizophrenia. Aphasic speech has long been characterized as both errorful and perseverative (e.g., Buckingham, Whitaker, & Whitaker, 1979; Sandson & Albert, 1984). The aphasic patient analyzed by Schwartz et al. (1994), F.L., makes about 100 times more errors than nonaphasic adults and makes more perseverations than anticipations. We predict that corresponding analyses with other patients will show that they have lower *AP*s than nonaphasic speakers and, in particular, that *AP* will be related to severity, as dictated by the line in Figure 12. There is indirect evidence that this prediction will be verified: Talo (1980) examined the spontaneous speech errors of aphasic patients and reported that anticipations were "rare" (p. 84). Helm-Estabrooks, Bayles, and Bryant (1994) found that perseverations made by aphasic patients in description and confrontation naming tasks were more likely in moderately impaired patients than in mildly impaired ones, where impairment was assessed by an aphasic diagnostic profile. Furthermore, Alzheimer's patients produced a number of perseverations intermediate between the aphasics and same-aged controls, providing additional evidence that perseveration is related to overall speech error rate. However, because anticipations were not recorded and, indeed, on confrontation naming tasks are not really possible, these data

⁸ The diphthong /oy/ is more marked than /ow/ which, in Stemberger's account should lead to /oy/ replacing /ow/.

⁹ The phonological errors in Spanish reported by Del Viso, Igoa, and Garcia-Albert (1991) appear to have more perseverations than anticipations, suggesting that Spanish may have a different line than the other languages studied.

do not tell us whether the *AP* measure itself decreases as severity increases.

Schizophrenic speech has also been associated with perseveration (e.g., Andreasen, 1986; Chaika, 1982; Maher, 1983). In addition to perseverations of words and sounds that appear to be quite similar to those of nonaphasic and aphasic speakers, schizophrenic patients may also perseverate on topics (e.g., persistently speaking about towns in Iowa; Andreasen, 1986). Hence, it's hard to tell whether a deviation is a slip or was intended. To our knowledge, there is no quantitative analysis of schizophrenic speech that allows for *AP* to be computed and for this to be related to overall speech error rate. If such analyses could be performed on patient speech samples, our prediction is that the samples would exhibit the general anticipatory effect. However, because of the difficulty in determining the patients' intended utterances, it may be hard to get an estimate of overall error rate that can be compared to that of others.

What Is Causing the Variation in Error Rate and AP?

The general anticipatory effect allows for error variation due to any number of parameters to be associated with a predictable *AP*. For variation due to speech rate, there is a clear link to the time parameter, n . For the other factors that we have considered, in particular practice and differences among individuals due to age or the presence of brain damage, it is not so obvious which parameters are responsible. Here we argue that connection weight, w , is likely the most important parameter contributing to the effect.

Practice. We follow the serial-order models and attribute the effect of practice to increases in connection weights between the plan and its elements. Specifically, practice enhances the activation of the *present* and *future* at the expense of the *past*. So, as performance gets better, perseverations become relatively less common. There is a plausible alternative to this view. Perhaps practicing a phrase allows, instead, for more effective turn-off of the *past*. Such an explanation accords with theories of processing that emphasize the flexibility of inhibitory processes (e.g., see papers in Dagenbach & Carr, 1994).

The question of whether practice works on the turn-on and priming functions or on the turn-off of the *past* can be addressed with our experimental data. The model assumes that there is a competition between the *past*, *present*, and *future*. In reality, the competition includes linguistic units that are not a part of the sequence, and an intrusion of one of these creates a noncontextual error. Although these errors, represented by the "word substitution" and "other" categories, are relatively unlikely in the experiments, they occurred sufficiently often in the three experiments ($n = 259$) to determine how their probability changes with practice. If practice works by enhancing the *present* and the *future*, noncontextual errors would be expected to go away with practice at roughly the same rate as perseverations do. More specifically, if a node standing for a noncontextual error is introduced into the model and assigned an activation level, it is functionally like the *past* node, which was assigned a particular level of c . Consequently, if practice increases w , which causes the *present* and *future* to gain more activation, the probability of selecting the *past* or the noncontextual node becomes more reduced than that of selecting the *future*. The data should then show that the slope for noncontextual errors is more like

that of perseverations than that of anticipations. If practice works, instead, by improving the turn-off of the *past*, a noncontextual node will be more like the *future* in that its activation will not change much relative to the *present*. Only the *past* will be greatly reduced. On this view, the slope for noncontextual errors will be more similar to anticipations than perseverations.

The number of anticipations, perseverations, and noncontextual errors as a function of practice block was determined, pooling across all experiments and word versus sound errors. The log-log slopes were $-.501$ for anticipations, $-.751$ for perseverations, and $-.756$ for noncontextual errors. A test of the null hypothesis that anticipations and noncontextual errors have the same slope can be rejected assuming that the alternative hypothesis is that the noncontextual errors are steeper, $t(6) = 2.02$, $p < .05$. It appears that practice eliminates noncontextual errors, as well as perseverations, more effectively than anticipations. This supports the claim that practice builds up the *present* and the *future*, which, in turn, is consistent with the assumption that practice affects w .

Developmental effects. The finding that young children make more errors and have higher *AP*s than adults (Stemberger, 1989; Wijnen, 1992) is, in our view, explained most easily by proposing that w is lower in children. As was the case for practice, though, there are alternative accounts. For example, Berg and Schade (1990) attributed the finding of more perseverations with children to a difficulty in turning off the *past*. Their model (e.g., Berg, 1988; Schade, 1992) is an example of a frame-based activation model that uses self-inhibition to deactivate structure and content units after they have been fully activated. Less effective self-inhibition would correspond, in our model, to a higher value of c . Stemberger (1989) suggested that children may have a slower (smaller) decay of activation (our d) than adults. Each of these proposals, a decrease in w or d , or an increase in c , leads to more errors and a lower *AP* in the model. We favor the explanation for the difference between adults and children in terms of w because assigning the difference to w accords with the undeniable fact that children do, in fact, have less knowledge of phoneme and word sequences than adults do. Connection weight is supposed to represent this knowledge and, hence, the proposal that w is lower in children is motivated in a way that the alternatives are not. However, we acknowledge that the data on child speech errors are not sufficiently extensive to support one view over another.

The general anticipatory effect should apply to error increases associated with aging as well as those associated with children. To the extent that older adults make more speech errors, they should have a lower *AP*. This prediction applies whether the underlying deficit is in processing speed (e.g., Salthouse, 1985), information loss (Myerson, Hale, Wagstaff, Poon, & Smith, 1990), information transmission (Burke, MacKay, Worthley, & Wade, 1991), or inhibitory processes (Zacks & Hasher, 1994). Although it would not be easy to relate the particular theories of aging to particular parameters and processes in our model, one may be able to say that the theories as a group assign aging effects to the group of parameters associated with the general anticipatory effect, namely w , n , d , and c . These parameters regulate the model's ability to retrieve what it should be retrieving as a sequence is produced. To our knowledge, there are no extensive quantitative analyses of older adults' slips, and so the

statement that the general anticipatory effect should apply to aging remains only a prediction.

Brain damage. Accounts of the effect of brain damage on the language production system are numerous and varied (see e.g., Bates, Wulfeck, & MacWhinney, 1991; Caplan & Hildebrandt, 1988; Caramazza, 1984; Harley, 1993). The focus can be on either where the damage lies, that is, on what subsystem of the language processor is affected, or on what the deficit is. Our concern is with the latter issue. Given that most aphasic patients make excessive numbers of phonological, and word slips, one can ask how these errors' greater prevalence can be best characterized in activation-based theories. Schwartz et al. (1994) and Harley and MacAndrew (1992) suggested that jargon aphasia could be the result of diminished connection strengths in a lexical network such as that proposed in the various frame-based activation models. In addition, aphasic errors have been attributed to a slower buildup of activation (e.g., Prather, Shapiro, Zurif, & Swinney, 1991, Haarmann & Kolk, 1991), diminished memory capacity (e.g., Haarmann, Just, & Carpenter, 1994) or loss of units in a highly distributed system (e.g., Farah, 1994). It turns out to be hard to discriminate among such proposals because functions such as processing speed, memory capacity, and so on can be viewed as basic parameters in some models or they can be derived from more elementary parameters in others. Similarly, loss of units in a network with distributed representations can act like reduced connection weights in a network with local representations. Consequently, any claims about the nature of the damage must be made in reference to a particular model.

There is one class of models of aphasic deficits that we believe our analysis makes unlikely. These are spreading activation models that propose that the aphasia is associated entirely with more activation noise in the system (e.g., Laine & Juhola, 1991). Although such models may successfully capture many features of the data, they may not be able to explain the apparent decrease in *AP* associated with aphasia. In our model, making the activation levels of the *past*, *present*, and *future* more noisy or making the decision process more noisy, while keeping all other aspects of the system the same, causes an increase in overall error rate without the lawful decrease in *AP*. If aphasic deficits are associated with a big drop in *AP*, as Schwartz et al. (1994) suggested, a model would have to concentrate the increased noise on the *past* more than on the *future* to actually flip an anticipation-dominant pattern to a perseveration-dominant one.

A related model of aphasia, that of Martin, Dell, Saffran, & Schwartz (1994; see also Martin & Saffran, 1992; Schwartz et al., 1994), attributes fluent aphasic production deficits to a relative increase in activation noise, but makes this happen by assuming that brain damage causes the activation levels in a lexical network to get sufficiently small that intrinsic noise creates many errors. By itself, this proposal has the same problem as that of simply adding noise—reducing activation will not change an anticipatory pattern into a perseveratory one. However, in Martin et al.'s (1994) approach, activation can be made small in two ways, reducing connection weights or increasing decay rate. To the extent that low activation is caused by decreased weights, there should be a relative increase in perseverations. According to the various serial-order models presented here, reducing w should decrease *AP*. However, Martin et al.'s other mechanism, increased decay rate, does not naturally de-

crease *AP*. On the contrary, in our analysis increasing d actually increases *AP*, particularly in the clamped case (see Equations 9 and 10). Thus, Martin et al. would predict that some patients, those with decreased connection strength, should exhibit a decreased *AP* in proportion to the severity of their deficit, while those with increased decay would have an *AP* that is higher than what would be expected from their error rate. In other words, patients of the former type, which they characterize as jargon-type patients, should fall on the line relating *AP* to error rate, while patients of the latter type, deep dysphasics, should be above the line.

Applications to Nonlinguistic Domains?

Thus far, the model has been applied to spoken language production. Our functional analysis, however, suggests that other behavioral systems should exhibit the kinds of phenomena that we have been studying. If the ordered behavior appears to have rules that allow for novel sequences, and particularly rules in which *present* elements depend on *future* ones, there is a need for structural frames and anticipatory priming. In short, to the extent that the behavior is languagelike, our model should apply. The model would not, for example, be expected to apply directly to simple behavioral perseveration, such as that shown by infants in search tasks (Piaget, 1954), because the relevant responses—looking for an object in one or another location—are not being generated as part of a system with languagelike dependencies. Moreover, there is no clear sense of what would constitute an anticipation. Nonetheless, the question of whether this kind of perseveration results from failure to represent the present or failure to eliminate the past is common to all discussions of perseveration and, hence, the model may provide a framework for formalization.

The closest domains to spoken language are writing and typing. However, because production in these domains is parasitic on the linguistic system down to the phonological level (Hotopf, 1983), successful tests of the model's predictions will not support a claim of generalization to nonlinguistic systems. The production of music may, instead, be a more fruitful domain. Music production is nonlinguistic, yet shares many features with language production (Palmer & van de Sande, 1993, 1995). Many performance errors can be classified as anticipatory or perseveratory, and it appears that there is lawful variation among error proportions. Palmer and Drake (1995) found that children at intermediate levels of piano skill had an *AP* of .74, while those with less skill had a significantly lower *AP* of .57. This finding is consistent with an anticipatory practice effect because the more advanced group had more years of piano training. However, because the more advanced group played more difficult pieces, it is difficult to interpret the results in the light of the model. In general, though, the model's predictions should be testable in nonlinguistic domains, such as music, in which complex behavioral sequences are produced.

Summary and Conclusions

Activation-based theories of serial order, when applied to language production, must satisfy four functional requirements: (a) the *present* must be activated (b) the *past* must be deactivated, (c) the *future* must be prepared, and (d) the system must be

capable of assembling the order of novel sequences. We have reviewed activation-based theories and found that they tend to achieve the first three of these functions in similar ways. Each has a plan that is separated from the sequence elements, and the plan has excitatory connections to the elements, the strength of which is a product of learning. Furthermore, each has either a self-inhibitory turn-off mechanism or its functional equivalent, and this mechanism is part of the architecture of the system and, hence, does not have to be learned for each sequence. Finally, the activation of the *present* and the preparation of the *future* are achieved by plan-to-element connections with the result that the *present* and *future*'s activation are linked to the same parameters of the system. These features enable the models to explain the anticipatory practice effect.

With regard to the fourth function, the ability to assemble novel sequences, only the frame-based theories of production appear to be sufficient. These theories propose that order is regulated, in part, by structures that represent patterns among categories of linguistic items, rather than items themselves. Such structures nicely account for the form of sound and word errors (Dell, 1986; Garrett, 1975; Hartley & Houghton, 1996; Levelt, 1989, 1992; MacKay, 1982, 1987; Roelofs, 1992; Shattuck-Hufnagel, 1979; Stemberger, 1990) and no model without frames has been shown to do so.¹⁰ Moreover, experimental studies of language production have shown that structural frames have reality in the sense that they can be primed. Bock and Loebell (1990) found that syntactic structures persist across utterances in the absence of lexical, semantic, thematic, or prosodic overlap (see also, Bock, Loebell, & Morey, 1992; Bock, 1986). There are analogous effects for phonological structures (Meijer, 1994; Romani, 1992; Sevald, Dell, & Cole, 1995; Stemberger, 1990).

We offered a model that reflects the common characteristics of activation-based models with regard to the first three functional requirements and that incorporates structural frames. This model was able to account for the anticipatory practice effect and predicted the existence of an anticipatory speech-rate effect and a general anticipatory effect. Three experiments and analyses of error corpora from several populations provided support for the model. The centerpiece of this support, shown in Figures 8–13, was the finding that the anticipatory proportion can be predicted solely from overall error rate. Several factors that affect error rate—practice, speech rate, individual differences, age, or brain damage—appear to induce variation in anticipatory proportion along a single line. The findings are sufficiently compelling for us to hypothesize that any large sample of English speech should yield a point on the line and that corresponding analyses with other languages and even with complex nonlinguistic domains may discover similar relations between correctness and anticipatory proportion.

In sum, the data suggest that an error-prone language-production system is inherently perseveratory, while a relatively error-free system tends to anticipate. Because of the functional requirements of serial order and the nature of language, the system should look to the future and not dwell on the past.

¹⁰ For example, PDP recurrent network models, which lack separate frame structures (e.g., Dell et al., 1993), are not capable of explaining the existence of exchange errors and various frame-priming effects, although these models can produce a variety of structural and rulelike effects.

References

- Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science*, 2, 396–408.
- Andreasen, N. (1986). Scale for the assessment of thought, language and communication (TLC). *Schizophrenia Bulletin*, 12, 473–482.
- Bates, E., Wulfeck, B., & MacWhinney, B. (1991). Cross-linguistic research in aphasia: An overview. *Brain and Language*, 41, 123–148.
- Berg, T. (1988). *Die Abbildung des Sprachproduktionprozess in einem Aktivationsflussmodell* [A spreading activation model of language production]. Tuebingen, Germany: Max Niemeyer.
- Berg, T., & Schade, U. (1990). *Activation and inhibition in emergent processing systems*. Unpublished manuscript.
- Bock, J. K. (1982). Towards a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review*, 89, 1–47.
- Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, 18, 355–387.
- Bock, J. K., & Levelt, W. J. M. (1994). Language production: Grammatical encoding. In M. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 945–984). San Diego: Academic Press.
- Bock, J. K., & Loebell, H. (1990). Framing sentences. *Cognition*, 35, 1–39.
- Bock, J. K., Loebell, H., & Morey, R. (1992). From conceptual roles to structural relations: Bridging the syntactic cleft. *Psychological Review*, 99, 150–171.
- Boomer, D. S., & Laver, J. D. M. (1968). Slips of the tongue. *British Journal of Disorders of Communication*, 3, 1–12.
- Bruce, D. (1994). Lashley and the problem of serial order. *American Psychologist*, 49, 95–103.
- Buckingham, H. W., Whitaker, H., & Whitaker, H. A. (1979). On linguistic perseveration. In H. Whitaker & H. A. Whitaker (Eds.), *Studies in Neurolinguistics* (Vol. 4, pp. 329–352). New York: Academic Press.
- Burgess, N., & Hitch, G. J. (1992). Toward a network model of the articulatory loop. *Journal of Memory and Language*, 31, 429–460.
- Burke, D. M., MacKay, D. G., Worthley, J. S., & Wade, E. (1991). On the tip of the tongue: What causes word finding failures in young and older adults? *Journal of Memory and Language*, 30, 542–579.
- Caplan, D., & Hildebrandt, N. (1988). *Disorders of syntactic comprehension*. Cambridge, MA: MIT Press.
- Caramazza, A. (1984). The logic of neuropsychological research and the problem of patient classification in aphasia. *Brain and Language*, 21, 9–20.
- Chaika, E. (1982). A unified explanation for the diverse structural deviations reported for adult schizophrenics with disrupted speech. *Journal of Communication Disorders*, 15, 167–189.
- Cleeremans, A., & McClelland, J. L. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology: General*, 120, 235–253.
- Cutler, A. (1981). The reliability of speech error data. *Linguistics*, 19, 561–582.
- Dagenbach, D., & Carr, T. H. (1994). *Inhibitory processes in attention, memory, and language*. San Diego, CA: Academic Press.
- Dell, G. S. (1986). A spreading activation theory of retrieval in language production. *Psychological Review*, 93, 283–321.
- Dell, G. S. (1989). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. In W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 136–165). Cambridge, MA: MIT Press. (Reprinted from *Journal of Memory and Language*, Vol. 27, pp. 124–142, 1988)
- Dell, G. S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes*, 5, 313–349.
- Dell, G. S., Juliano, C., & Govindjee, A. (1993). Structure and content in language production: A theory of frame constraints in phonological speech errors. *Cognitive Science*, 17, 149–195.

- Dell, G. S., & O'Seaghda, P. G. (1994). Inhibition in interactive activation models of linguistic selection and sequencing. In D. Dagenbach & T. H. Carr (Eds.), *Inhibitory processes in attention, memory, and language* (pp. 409–453). San Diego, CA: Academic Press.
- Dell, G. S., & Reich, P. A. (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior*, 20, 611–629.
- Del Viso, S., Igoa, J. M., & Garcia-Albert, J. E. (1991). On the autonomy of phonological encoding: Evidence from slips of the tongue in Spanish. *Journal of Psycholinguistic Research*, 20, 161–185.
- Eikmeyer, H.-J., & Schade, U. (1991). Sequentialization in connectionist language-production models. *Cognitive Systems*, 3, 128–138.
- Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14, 213–252.
- Estes, W. K. (1972). An associative basis for coding and organization in memory. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory* (pp. 161–190). Washington, DC: Winston.
- Farah, M. J. (1994). Neuropsychological inference with an interactive brain: A critique of the "locality" assumption. *Behavioral and Brain Sciences*, 17, 43–104.
- Ferreira, F. (1993). Creation of prosody during sentence production. *Psychological Review*, 100, 233–253.
- Fowler, C. A., Treiman, R., & Gross, J. (1993). The structure of English syllables and polysyllables. *Journal of Memory and Language*, 32, 115–140.
- Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language*, 47, 27–52.
- Fromkin, V. A. (Ed.). (1973). *Speech errors as linguistic evidence* (pp. 11–45). The Hague, The Netherlands: Mouton.
- Fromkin, V. A. (1980). *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand*. New York: Academic Press.
- Garnham, A., Shillcock, R. C., Brown, G. D. A., Mill, A. I. D., & Cutler, A. (1981). Slips of the tongue in the London-Lund corpus of spontaneous conversation. *Linguistics*, 19, 805–817.
- Garrett, M. F. (1975). The analysis of sentence production. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 133–175). San Diego, CA: Academic Press.
- Garrett, M. F. (1980). Levels of processing in sentence production. In B. Butterworth (Ed.), *Language production* (Vol. 1, pp. 177–210). London: Academic Press.
- Gasser, M. E. (1988). A connectionist model of sentence generation in a first and second language (Tech. Rep. No. UCLA-AI-88-13). Los Angeles: University of California, Los Angeles, Computer Science Department.
- Gordon, P. C., & Meyer, D. E. (1987). Control of serial order in rapidly spoken syllable sequences. *Journal of Memory and Language*, 26, 300–321.
- Grossberg, S. (1978). A theory of human memory: Self-organization and performance of sensory-motor codes, maps, and plans. In R. Rosen & F. Snell (Eds.), *Progress in theoretical biology* (Vol. 5, pp. 233–374). New York: Academic Press.
- Grossberg, S. (1982). *Studies of mind and brain: Neural principles of learning, perception, development, cognition, and motor control*. Boston: Reidel.
- Gupta, P. (1995). *Word learning and immediate serial recall: Toward an integrated account*. Unpublished doctoral dissertation, Carnegie Mellon University, Pittsburgh, PA.
- Gupta, P., & MacWhinney, B. (in press). Vocabulary acquisition and verbal short-term memory: Computational and neural bases. *Brain and Language*.
- Haarmann, H. J., Just, M. A., & Carpenter, P. A. (1994). Computational modelling of normal and aphasic sentence comprehension. *Brain and Language*, 47, 389–391.
- Haarmann, H. J., & Kolk, H. H. J. (1991). A computer model of the temporal course of agrammatic sentence understanding: The effects of variation in severity and sentence complexity. *Cognitive Science*, 15, 49–87.
- Harley, T. A. (1993). Connectionist approaches to language disorders. *Aphasiology*, 7, 221–249.
- Harley, T. A., & MacAndrew, S. B. G. (1992). Modelling paraphasias in normal and aphasic speech. In J. K. Kruschke (Ed.), *Proceedings of the 14th Annual Conference of the Cognitive Science Society* (pp. 378–383). Hillsdale, NJ: Erlbaum.
- Hartley, T. A., & Houghton, G. (1996). A linguistically constrained model of short-term memory for nonwords. *Journal of Memory and Language*, 35, 1–31.
- Helm-Estabrooks, N., Bayles, K., & Bryant, S. (1994). Four forms of perseveration in dementia and aphasia patients and normal elders. *Brain and Language*, 47, 457–460.
- Hotopf, W. H. N. (1983). Lexical slips of the pen and tongue. In B. Butterworth (Ed.), *Language production* (Vol. 2, pp. 145–199). San Diego, CA: Academic Press.
- Houghton, G. (1990). The problem of serial order: A neural network model of sequence learning and recall. In R. Dale, C. Mellish, & M. Zock (Eds.), *Current research in natural language generation* (pp. 287–319). London: Academic Press.
- Jordan, M. I. (1986). Attractor dynamics and parallelism in a connectionist sequential machine. *Proceedings of the Eighth Annual Conference of the Cognitive Science Society* (pp. 531–546). Hillsdale, NJ: Erlbaum.
- Laine, M., & Juhola, M. (1991). *Modelling aphasic naming disorders by computer simulation*. Paper presented at the Fourth Finnish Conference on Neurolinguistics, Turku, Finland.
- Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior* (pp. 112–136). New York: Wiley.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Levelt, W. J. M. (1992). Accessing words in speech production: Stages, processes, and representations. *Cognition*, 42, 1–22.
- Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, 95, 492–527.
- MacKay, D. G. (1970). Spoonerisms: The structure of errors in the serial order of speech. *Neuropsychologia*, 8, 323–350.
- MacKay, D. G. (1972). The structure of words and syllables: Evidence from errors in speech. *Cognitive Psychology*, 3, 210–227.
- MacKay, D. G. (1973). Complexity in output systems: Evidence from behavioral hybrids. *American Journal of Psychology*, 86, 785–806.
- MacKay, D. G. (1974). Aspects of the syntax of speech: Syllable structure and speech ratio. *Quarterly Journal of Experimental Psychology*, 26, 642–657.
- MacKay, D. G. (1982). The problems of flexibility, fluency, and speed-accuracy trade-off in skilled behaviors. *Psychological Review*, 89, 483–506.
- MacKay, D. G. (1987). *The organization of perception and action: A theory for language and other cognitive skills*. New York: Springer-Verlag.
- Maher, B. A. (1983). A tentative theory of schizophrenic utterances. *Progress in Personality Research*, 12, 1–52.
- Martin, N., Dell, G. S., Saffran, E. M., & Schwartz, M. F. (1994). Origins of paraphasias in deep dysphasia: Testing the consequences of a decay impairment to an interactive spreading activation model of lexical retrieval. *Brain and Language*, 47, 609–660.
- Martin, N., & Saffran, E. M. (1992). A computational account of deep dysphasia: Evidence from a single case study. *Brain and Language*, 43, 240–274.
- McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part I. An account of basic findings. *Psychological Review*, 88, 375–407.
- Meijer, P. J. A. (1994). *Phonological encoding: The role of supraseg-*

- mental structures*. Unpublished doctoral dissertation, Nijmegen University, Nijmegen, The Netherlands.
- Meringer, R. (1908). *Aus dem Leben der Sprache* [On the life of language]. Berlin, Germany: V. Behr's Verlag.
- Meyer, A. S. (1994). Timing in sentence production. *Journal of Memory and Language*, 33, 471–492.
- Mowrey, R. A., & MacKay, I. R. A. (1990). Phonological primitives: Electromyographic speech error evidence. *Journal of the Acoustical Society of America*, 88, 1299–1312.
- Murdock, B. B., Jr. (1974). *Human memory: Theory and data*. Hillsdale, NJ: Erlbaum.
- Myerson, J., Hale, S., Wagstaff, D., Poon, L. W., & Smith, G. A. (1990). The information-loss model: A mathematical theory of age-related cognitive slowing. *Psychological Review*, 97, 475–487.
- Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1–55). Hillsdale, NJ: Erlbaum.
- Nooteboom, S. G. (1969). The tongue slips into patterns. In A. G. Sciarone, A. J. van Essen, & A. A. Van Raad (Eds.), *Leyden studies in linguistics and phonetics* (pp. 114–132). The Hague, The Netherlands: Mouton.
- Nooteboom, S. G. (1980). Speaking and unspeaking: Detection and correction of phonological and lexical errors in spontaneous speech. In V. A. Fromkin (Ed.), *Errors in linguistic performance* (pp. 87–95). New York: Academic Press.
- Palmer, C., & Drake, C. (1995). *Musical skill acquisition: Better monitoring or planning?* Unpublished manuscript.
- Palmer, C., & van de Sande, C. (1993). Units of knowledge in music performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 457–470.
- Palmer, C., & van de Sande, C. (1995). Range of planning in music performance. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 947–962.
- Piaget, J. (1954). *The construction of reality in the child*. New York: Basic.
- Prather, P., Shapiro, L., Zurif, E., & Swinney, D. (1991). Real-time examination of lexical processing in aphasics. *Journal of Psycholinguistic Research*, 20, 271–281.
- Reich, P. A. (1977). Evidence for a stratal boundary from slips of the tongue. *Forum Linguisticum*, 2, 211–217.
- Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, 42, 107–142.
- Roelofs, A. (1995). *A model of word-form encoding in speaking: Generating phonetic plans in context*. Unpublished manuscript.
- Romani, C. A. (1992). *The representation of prosodic and syllabic structure in speech production*. Unpublished doctoral dissertation, Johns Hopkins University, Baltimore, MD.
- Rosenbaum, D. A. (1990). *Human motor control*. San Diego, CA: Academic Press.
- Rosenbaum, D. A., Kenny, S. B., & Derr, M. A. (1983). Hierarchical control of rapid movement sequences. *Journal of Experimental Psychology: Human Perception and Performance*, 9, 86–102.
- Rossi, M., & Peter-Defare, E. (1995). Lapsus linguae: Word errors or phonological errors? *International Journal of Psycholinguistics*, 11, 5–38.
- Rumelhart, D. E., & Norman, D. A. (1982). Simulating a skilled typist: A study of skilled cognitive motor performance. *Cognitive Science*, 6, 1–36.
- Salthouse, T. A. (1985). *A theory of cognitive aging*. Amsterdam: North-Holland.
- Sandson, J., & Albert, M. L. (1984). Varieties of perseveration. *Neuropsychologia*, 22, 715–732.
- Schade, U. (1992). *Konnektionismus-Zur Modellierung der Sprachproduktion* [Connectionism: Modelling of language production]. Opladen, Germany: Westdeutscher Verlag.
- Schwartz, M. F., Saffran, E. M., Bloch, D. E., & Dell, G. S. (1994). Disordered speech production in aphasic and normal speakers. *Brain and Language*, 47, 52–88.
- Sevold, C. A., Dell, G. S., & Cole, J. (1995). Syllable structure in speech production: Are syllables chunks or schemas? *Journal of Memory and Language*, 34, 807–820.
- Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial-order mechanism in sentence production. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 295–342). Hillsdale, NJ: Erlbaum.
- Shattuck-Hufnagel, S. (1983). Sublexical units and suprasegmental structure in speech production planning. In P. F. MacNeilage (Ed.), *The production of speech* (pp. 109–136). New York: Springer.
- Shattuck-Hufnagel, S., & Klatt, D. (1979). The limited use of distinctive features and markedness in speech production: Evidence from speech error data. *Journal of Verbal Learning and Verbal Behavior*, 18, 41–55.
- Stemberger, J. P. (1982). *The lexicon in a model of language production*. Unpublished doctoral dissertation, University of California, San Diego.
- Stemberger, J. P. (1985). An interactive activation model of language production. In W. W. Ellis (Ed.), *Progress in the psychology of language* (Vol. 1, pp. 143–186). Hillsdale, NJ: Erlbaum.
- Stemberger, J. P. (1989). Speech errors in early child language production. *Journal of Memory and Language*, 28, 164–188.
- Stemberger, J. P. (1990). Wordshape errors in language production. *Cognition*, 35, 123–157.
- Stemberger, J. P. (1991). Apparent anti-frequency effects in language production: The addition bias and phonological underspecification. *Journal of Memory and Language*, 30, 161–185.
- Sternberg, S., Monsell, S., Knoll, R. L., & Wright, C. E. (1978). The latency and duration of rapid movement sequences: Comparisons of speech and typing. In G. E. Stelmach (Ed.), *Information processing in motor control and learning* (pp. 117–152). New York: Academic Press.
- Svartvik, J., & Quirk, R. (1980). *A corpus of English conversation*. Lund, Sweden: CWK Gleerup.
- Talo, E. S. (1980). Slips of the tongue in normal and pathological speech. In V. A. Fromkin (Ed.), *Errors in linguistic performance* (pp. 81–86). New York: Academic Press.
- Wijnen, F. (1992). Incidental word and sound errors in young speakers. *Journal of Memory and Language*, 31, 734–755.
- Zacks, R. T., & Hasher, H. (1994). Directed ignoring: Inhibitory regulation of working memory. In D. Dagenbach & T. H. Carr (Eds.), *Inhibitory processes in attention, memory, and language* (pp. 241–264). San Diego, CA: Academic Press.

Appendix A

Phrases Used in Experiments 1–3

Bonnie's brown bread box	Freida's fabulous freaky fabric
Brad's burned bran buns	Gloria's Greek green gloves
Brief beastly beach breezes	Pam's plain plaid pan
Chef's sooty shoe soles	Plastic potted pansy plants
Danny's dripping dish drain	Sappy shiny shop signs
Fine fresh free fish	Simple slender silver slippers
Five frantic fat frogs	Thirty-three throbbing thumbs
Floyd's fourth floor fort	Tike's tricky trike tire

Appendix B

Error Categorization

Categories

Word anticipation (WA), word perseveration (WP), word exchange (WE), word substitution (WS), sound anticipation (SA), sound perseveration (SP), sound exchange (SE), sound anticipation–perseveration (SAP).

1. *Word versus sound.* An error is a word error if the error string is a word or word stem from elsewhere in the utterance (contextual word error), or if it is word from outside the utterance that could not have arisen from the movement of sounds within the utterance (WS category). Otherwise, the error is a sound error, provided that it doesn't fall in the O category as specified in (2).

Examples: "Tike's tricky trike tire" spoken as "trike's tricky trike tire" is a word error because *trike* is a word from the phrase, even though the change is small. "Sappy shiny shop sign" spoken as "sappy shiny *stop* sign" is a WS. "Chef's sooty shoe soles" spoken as "chef's sooty *sue* soles" is a sound error, even though *sue* is a word. If the sound /s/ replaces the /ʃ/ in *shoe*, the result is *sue*.

2. *The "other" category.* Errors involving grammatical affixes ("chef's sooty *shoes* *sole*" for "chef's sooty shoe soles"), deletion of a word ("chef's sooty shoe"), or the noncontextual substitution of a sound ("chef's sooty *foo* soles") were placed in the O category.

3. *Serial-order errors.* All word and sound errors outside of the WS and O categories are counted as serial-order errors, either as anticipations, perseverations, exchanges, or anticipation–perseverations. For word errors, the error was categorized as WA, WP, or WE depending on the location of the intruding word (for WA vs. WP) and whether there was a corresponding substitution in the same utterance (WE).

Examples: "Brad's burned *brad* buns" is WP. "Chef's sooty *sole* soles" is WA. "Brad's *bran* burned buns" is WE.

For sound errors, the target and intruding sounds are first defined. These can be single phonemes or contiguous groups of phonemes smaller than a word stem. Errors in which singleton consonants interacted with clusters (e.g., "plastic potted" is spoken as "*pastic* potted," "plastic *plotted*," or "*pastic plotted*") are deemed to involve the replacement of one syllabic constituent by another (e.g., replacement of /p/ by /pl/) rather than as addition, deletion, or shift errors. This allows these errors to be categorized as anticipations, perseverations, and exchanges

in the same way that single phoneme errors would be. A sound error was classified as SA if all intended occurrences of the intruding constituent were after the target location, and as SP if all intended occurrences were before. An error was SE if there was a corresponding substitution in the same utterance (e.g., "*pastic plotted* . . ." for "plastic potted . . ." is SE). For errors in which the intruding constituent occurs both before and after the target, the error was classified as SAP if the before and after constituents were equidistant in words from the target. For example, "brad's burned *ban* buns" is SAP. If the two intruding constituents were not equidistant, the error was classified on the basis of the closest potential source. For example, "plastic *plotted* pansy plants" is SP, and "plastic potted *plansy* plants" is SA.

4. *Multiple errors.* More than one error in an utterance creates some ambiguity of classification. The following heuristics were used to resolve ambiguity: (a) try to classify as many of the errors as possible as exchanges as opposed to anticipations and perseverations; (b) prefer exchanges that are closer together than further apart; and (c) try to account for the utterance in terms of the fewest error categories. For example, for "Pam's plain plaid pan" prefer "Plam's *pain* (SE) plaid *plan* (SP)" to "Plam's (SA) *pain* (SP) plaid *plan* (SP)" or to "Plam's (SA) *pain* *plaid plan* (distant SE)."

Additional Rules

The following were added after the first reliability check.

5. *Onset bias.* Assume that sound errors in a word's onset have a source in the onsets of nearby words. Hence, "Freida's fabulous freaky fabric" spoken as "Freida's fabulous freaky *frabric*" is SP, where the target onset /f/ in *fabric* is replaced by the onset /fr/ (from *freaky* and *Freida*). This onset bias overrides a classification of SA from the nononset /r/ in *frabric*.

6. *Look for exchanges first.* Before determining whether errors are word or sound errors, look for possible exchanges. For example, for the phrase "tike's tricky trike tire," the error "trike's *ticky* trike tire" would be classified as "trike's (WA) *ticky* (SP) trike tire" if the determination of whether an error is a word or sound is made before looking for exchanges. Instead, exchanges should be looked for first, resulting in classification in this case as SE (involving /t/ and /tr/).

(Appendixes continue on next page)

An Extended Example

This complex case illustrates several classification principles (as well as the difficulty of classifying multiple errors).

“Tike’s tricky trike tire” spoken as “trike’s tricky tike trire.”

The rule of looking for exchanges first finds two possibilities, the WE of *trike* and *tike*, and the SE of the onsets of *tike* and *trire*. These characterizations are contradictory because they each use the same error

string *tike*, one analysis saying that it is the final part of a WE, and one saying that it is the initial part of an SE. If the WE option is chosen then the complete classification is “Trike’s tricky *tike* (WE) *trire* (SP).” (Note that *trire* is SP because of the onset bias.) If the SE option is chosen, the complete classification is “*trike*’s (WA) tricky *tike trire* (SE).” Because both analyses have the same number of errors (2) and both follow the rule of looking for exchanges, the only basis for a tie breaker is the exchange distance. Hence, the WA–SE analysis was chosen.

Appendix C

Proofs for Equations 3–8

The expressions for the absolute activations of the *past*, *present*, and *future* units after n time steps for clamped and pulsed input of 1.0 to the *plan* unit (Equations 3–8) can be proved by induction.

Initial activations:

$$A(\text{plan}, n = 0) = 1.0$$

$$A(\text{past}, n = 0) = c$$

$$A(\text{present}, n = 0) = 0$$

$$A(\text{future}, n = 0) = 0$$

Clamped Case— $A(\text{plan})$ Remains at 1.0

I. Past

To prove:

$$A(\text{past}) = c(1 - d)^n.$$

Because w to *past* is zero, the net input to *past* is zero and, hence, from Equation 2,

$$A(\text{past}, n = 1) = c(1 - d),$$

which proves the case for $n = 1$. Assume:

$$A(\text{past}, n = k) = c(1 - d)^k.$$

Prove for $n = k + 1$.

By Equation 2 and assumption,

$$A(\text{past}, n = k + 1) = c(1 - d)^k(1 - d) = c(1 - d)^{(k+1)},$$

which proves the case for $n = k + 1$ and, hence, proves for all n .

II. Present

To prove:

$$A(\text{present}) = w \frac{1 - (1 - d)^n}{d},$$

prove for $n = 1$.

Because the weight from *plan* to *present* is w and the initial activation of *present* is 0, then by Equations 1 and 2

$$A(\text{present}, n = 1) = 1w + 0(1 - d) = w = w \frac{1 - (1 - d)^1}{d},$$

which proves the case for $n = 1$. Assume:

$$A(\text{present}, n = k) = w \frac{1 - (1 - d)^k}{d}.$$

Prove for $n = k + 1$.

By Equations 1 and 2, and assumption

$$\begin{aligned} A(\text{present}, n = k + 1) &= w \left[\frac{1 - (1 - d)^k}{d} \right] (1 - d) + 1w \\ &= w \left[\frac{(1 - (1 - d)^k)(1 - d)}{d} + 1 \right] \\ &= w \left[\frac{(1 - d) - (1 - d)^{(k+1)}}{d} + 1 \right] \\ &= w \frac{(1 - d) + d - (1 - d)^{(k+1)}}{d} \\ &= w \frac{1 - (1 - d)^{(k+1)}}{d}, \end{aligned}$$

which proves the case for $n = k + 1$, and therefore proves for all n .

III. Future

To prove:

$$A(\text{future}) = bw \frac{1 - (1 - d)^n}{d}$$

The proof is structured the same as for the *present*, replacing w with bw because the effective connection weight from *plan* to *future* is bw .

Pulsed Case: *Plan* Undergoes Decay

Note:

$$A(\text{plan}) = (1 - d)^n$$

(Proof is analogous to that of the *past* shown above, with 1.0 replacing c .)

I. Past

This is the same as the clamped case because the *plan*’s activation is irrelevant, given that the effective weight from *plan* to *past* is zero.

II. Present

To prove:

$$A(\text{present}) = nw(1 - d)^{(n-1)},$$

prove for $n = 1$.

By Equations 1 and 2 and the note above,

$$A(\text{present}, n = 1) = w(1 - d)^0 + 0(1 - d) = w = 1w(1 - d)^0,$$

which proves the case for $n = 1$. Assume:

$$A(\text{present}, n = k) = kw(1 - d)^{(k-1)}.$$

to prove:

$$A(\text{present}, n = k + 1) = (k + 1)w(1 - d)^k.$$

By Equations 1 and 2, the note above, and assumption

$$\begin{aligned} A(\text{present}, n = k + 1) &= kw(1 - d)^{(k-1)}(1 - d) + w(1 - d)^k \\ &= kw(1 - d)^k + w(1 - d)^k \\ &= (k + 1)w(1 - d)^k, \end{aligned}$$

which proves the case for $n = k + 1$, and hence for all n .

III. Future

To prove:

$$A(\text{future}) = nbw(1 - d)^{(n-1)}.$$

The proof is analogous to the one for *present*, replacing w with bw .

Received July 17, 1995
Revision received May 13, 1996
Accepted May 15, 1996 ■

Statement of Ownership, Management, and Circulation
(Required by 39 USC 2082)

1. Publication Title: **Psychological Review**

2. Issue Frequency: **Quarterly**

3. Issue Date for Circulation Data Below: **October 1996**

4. Issue Period for Circulation Data Below: **October 1996**

5. Annual Subscription Price: **\$45/Year, \$97/Ind**

6. Annual Institutional Price: **\$167/Inst**

7. Complete Mailing Address of Headquarters or General Business Office of Publisher (Not printer):
750 First Street NE, Washington, DC 20002-4242

8. Complete Mailing Address of Publisher (Not printer):
750 First Street NE, Washington, DC 20002-4242

9. Full Names and Complete Mailing Addresses of Publisher, Editor, and Managing Editor (Do not leave blank):
Publisher: **American Psychological Association, 750 First Street NE, Washington, DC 20002-4242**
Editor: **Robert A. Stoff, Ph.D., Dept of Psychology, Univ of California, 405 Hilgard Ave, San Diego, CA 92024-1561**
Managing Editor: **Susan Knapp, American Psychological Association, 750 First Street NE, Washington, DC 20002-4242**

10. Owner (Do not leave blank. If the publication is owned by a corporation, give the name and address of the corporation immediately followed by the names and addresses of all stockholders owning or holding 1 percent or more of the total amount of stock. If not owned by a corporation, give the names and addresses of all individual owners. If owned by a partnership or other unincorporated firm, give its name and address as well as those of each individual owner. If the publication is published by a nonprofit organization, give its name and address.)
American Psychological Association, 750 First Street, NE, Washington, DC 20002-4242

11. Known Bondholders, Mortgagees, and Other Security Holders Owning or Holding 1 Percent or More of Total Amount of Bonds, Mortgages, or Other Securities. If none, check box.
 None

12. Tax Status (For completion by corporate organizations authorized to mail at special rates. Check one)
 This journal, magazine, or newspaper is published at the special rate for federal income tax purposes.
 This journal, magazine, or newspaper is published at the regular rate.
 This journal, magazine, or newspaper is published at the regular rate with the agreement.

PS Form 3526, September 1995

12. Publication Title: **Psychological Review**

13. Issue Date for Circulation Data Below: **July 1996**

14. Extent and Nature of Circulation	Average No. Copies Each Issue During Preceding 12 Months	Actual No. Copies of Single Issue Published Nearest to Filing Date
a. Total Number of Copies (Net press run)	8,652	8,555
b. Paid and/or Requested Circulation: (1) Sales Through Dealers and Carriers, Street Vendors, and Counter Sales (Net press run)	---	---
(2) Paid or Requested Mail Subscriptions (Include advertiser's proof copies and exchange copies)	6,689	6,218
c. Total Paid and/or Requested Circulation (Sum of 14b(1) and 14b(2))	6,689	6,218
d. Free Distribution by Mail (Employee, complimentary, and other mail)	151	147
e. Free Distribution Outside the Mail (Carriers or other means)	---	---
f. Total Free Distribution (Sum of 14d and 14e)	151	147
g. Total Distribution (Sum of 14c and 14f)	6,840	6,365
h. Copies not Distributed (1) Office Use, Leftovers, Spoiled (2) Publisher's Free Copies	1,812	2,190
i. Total (Sum of 14g, 14h(1), and 14h(2))	8,652	8,555
Percent Paid and/or Requested Circulation (14c / 14g * 100)	67.8	97.7

16. Publication of Statement of Ownership: Publication required, will be printed in the **January 1997** issue of this publication.

17. Signature and Title of Editor, Publisher, Business Manager, or Owner: **Robert A. Stoff, Dr. Publishing Services** Date: **revised**

18. I certify that all information furnished on this form is true and complete. I understand that anyone who furnishes false or misleading information on this form or who omits material or information requested on the form may be subject to criminal sanctions (including fines and imprisonment) and/or civil sanctions (including multiple damages and civil penalties).

Instructions to Publishers:
1. Complete and file one copy of this form with your postmaster annually on or before October 1. Keep a copy of the completed form for your records.
2. In cases where the publisher is a security printer in a nation, include in items 10 and 11 the name of the printer or printers for whom the matter is being printed. Also include the names and addresses of individuals who are stockholders who own or hold 1 percent or more of the total amount of bonds, mortgages, or other securities of the publishing enterprise. In item 11, if none, check the box. Use blank sheets if more space is required.
3. Be sure to furnish all circulation information called for in item 15. Free circulation must be shown in items 14d & 14e.
4. If the publication has received clear subscription or a general or regular publication, this Statement of Ownership, Management, and Circulation must be published, it must be printed in any issue of October or if the publication is not published during October, the last issue printed after October.
5. In item 16, indicate the date of the issue in which this Statement of Ownership will be published.
6. Item 17 must be signed.
7. Return to file or publish a statement of ownership only after a postmaster's authorization.

PS Form 3526, September 1995 (Rev. 10/95)