Episodic memory reflected in printed word naming

STEPHEN D. GOLDINGER and TAMIKO AZUMA
Arizona State University, Tempe, Arizona

Although memory is typically measured by recall or recognition, it is also expressed by fluent or stylized task performance. In this experiment, 12 volunteers (called *speakers*) completed four experimental stages over a 2-week period. They read printed words aloud in two sessions, before and after exposure to auditory training tokens. They later completed a recognition memory test, discriminating old from new words. Groups of *perceptual judges* assessed the speakers' vocal imitation by comparing utterances recorded before and after training and deciding which sounded like "better imitations" of the training tokens. The data showed clear evidence of postexposure imitation, with systematic effects that preclude strategic explanations. The contents of episodic memory were reflected by participants' speaking style while they were reading aloud. Together, the imitation and recognition data suggest that memory preserves detailed traces of spoken words; those traces were apparently activated when participants later read the same words in the same context.

In a classic article, Oldfield (1966) described the *mental lexicon*, a collection of words in long-term memory that has been a focus of extensive investigation and theorizing. Within that enterprise, questions have arisen regarding the nature of lexical representation, and about the relationship of word recognition to either episodic or semantic memory. Generally there are two main views on lexical representation: *Abstractionist* theories view the lexicon as a set of ideal units (e.g., Bowers & Michita, 1998; Caramazza, Laudanna, & Romani, 1988; Humphreys, Evett, & Quinlan, 1990; Marslen-Wilson & Warren, 1994; Morton, 1979), whereas *episodic* theories assume that groups of detailed memory traces collectively represent individual words (Goldinger, 1998; Jacoby, 1983; see Bowers, 2000, and Tenpenny, 1995, for reviews). Naturally, given such poles, there are intermediate views, including those that inform mixed and distributed models (Brown & Carr, 1993; Stark & McClelland, 2000).

In some regards, theorizing about lexical representations appears rather schizophrenic, owing to the near-universal application of words as experimental stimuli. In psycholinguistics, words are fairly magical entities, representing the psychological level at which 26 meaningless letters coalesce into thousands of meaningful units. Many researchers study word recognition itself; others follow linguistic pathways higher, studying how words are integrated into discourse. In either circumstance, words are typically treated in a manner consistent with linguistic theory (Halle, 1985),

as abstract, canonical units. Word recognition is appreciated for its stability despite variations and is theoretically likened to finding entries in a computer (Forster, 1979) or activating nodes (or patterns) in networks (Plaut, McClelland, Seidenberg, & Patterson, 1996).

In studies of general cognition, words are not the investigative focus but are used as a methodological convenience. Words are familiar, they can be manipulated along various dimensions (perceptual and conceptual), they have "built-in" manipulations (e.g., frequency variations), and they are plentiful. Words, presented in list format, are commonly used in memory research. Perhaps owing to these different empirical orientations toward words, different theoretical views have evolved across fields of research. In psycholinguistics, the abstract view is prominent; perception entails *information reduction*, matching tokens to types. This view is reasonable, since people are remarkably robust across perceptual variations in spoken- and printed-word perception. Moreover, word perception is rarely an end in itself, because people typically integrate words into discourse. Given these facts, it seems evident that some form of abstraction takes place in the mental lexicon (Bowers, 2000) or in memory at large.

Nevertheless, memory researchers have advanced theories with more dynamic portrayals of memory traces, including lexical representations. Models of long-term memory must allow for abstraction and generalization, but must also explain memory for specific episodes. To achieve this end, many models posit episodic representations, collections of instances in memory. Depending on the model, these instances may be stored separately or may become blended over time, allowing a continuum of theories from episodic to more abstract. For example, some global memory (Hintzman, 1986; Shiffrin & Steyvers, 1997) and categorization models (Nosofsky & Palmeri, 1997) are strongly episodic, with "abstraction" occurring only during re-

trieval. Holographic (Eich, 1982) and distributed (Mc-Clelland & Rumelhart, 1985) models also assume episodic traces, but blending occurs in information storage and retrieval. By allowing stored episodes to influence perceptual processes, these models achieve the stability of abstract representations, while allowing idiosyncratic effects of specific episodes.

Memory for Word Details

Word perception is a special case of perceptual classification, operating as a content-addressable memory system (Kohonen, Lehtiö, & Oja, 1989). This memory access may proceed differently, given differences in representations. In abstractionist theories, idiosyncratic features of words are *normalized* with respect to canonical representations (Jackson & Morton, 1984). In its strongest form (not often posed), normalization suggests that surface features of words never reach long-term memory. Conversely, episodic theories contend that such features are encoded and should be retrievable. Naturally, weaker forms of both views are possible (Brown & Carr, 1993), predicting that surface features of words are incidentally encoded in perception and may be later recalled if conditions are favorable.

With respect to spoken words, many data suggest that voice details are stored during lexical access, affecting both direct and indirect memory tests (Goldinger, Azuma, Kleider, & Holmes, 2003). Repetition effects are generally enhanced when voices remain constant across presentations of any given word. Similar experiments have assessed font (typography) effects in the visual domain, but the accumulated evidence is not entirely compelling. As Goldinger (1998) noted, voices have ecological value, communicating information about a speaker's gender, age, emotional state, and so forth. Moreover, voice inflections are often used paralinguistically, as in sarcasm. Fonts, however, communicate little information beyond the identities of their letters. Although fonts may be used to convey emphasis (e.g., boldface or italics), they generally have no meaningful variation across words, and the initial choice of font has little communicative value. They are, by definition, superficial details of printed words.

Despite their trivial nature, font details occasionally remain in episodic memory following printed word perception. The literature contains a mixture of positive and null effects: In general, repetition priming is robust for *all* repeated words, regardless of font changes (see Goldinger et al., 2003; Tenpenny, 1995). For example, priming persists even with dramatic changes between study and test presentations, such as switching Japanese words from Kanji to Hiragana (Bowers & Michita, 1998). Clearly, priming reflects more than simple, bottom-up activation. A basic tenet of cognition is that people elaborate upon stimuli, lending them meaning and purpose (e.g., if responses must be selected). Such top-down processes convey the potential for priming.

Nevertheless, bottom-up matches nearly always *increase* priming, relative to pure top-down matches. In printed-word perception, font-specificity effects are commonly observed (see Goldinger et al., 2003). These effects are often small, but they appear generally reliable. Thus, physical and procedural (bottom-up and top-down) matches produce extra priming, relative to conditions in which the study–test connection is purely conceptual. As Bowers (2000) wrote, "there has been a consistent trend for specificity effects across studies, and if a simple sign test were to be carried out on all these studies, the effect would be significant" (p. 87). Although this trend supports the episodic view (Tenpenny, 1995), the evidence is certainly not overwhelming. Moreover, reliable font-specificity effects are often "qualified" in some manner (e.g., only arising with unusual fonts; Jacoby & Hayman, 1987).

For purposes of theoretical advancement, font-specificity research may be nearly exhausted (Brown & Carr, 1993). To move forward, converging evidence is needed. In the present study, we tested the role of specific episodes in word reading, but our approach differed from that of most prior research. Although implicit memory is typically indexed by perceptual changes (e.g., fragment identification), people also express memory by idiosyncratic or stylized behavior. A common example is *linguistic convergence*, wherein people from diverse social backgrounds unintentionally modify their speech, coming to some "group average" in grammar and style (Giles, Coupland, & Coupland, 1991). In this investigation, we assessed the role of episodic traces in printed-word naming, using *spontaneous vocal imitation* to measure implicit memory.

This study extended work by Goldinger (1998), who observed imitation when people shadowed *spoken* words. Imitation appeared to reflect processes of word recognition, as it was influenced systematically by factors such as word frequency. In the present study, we extended this method to printed-word naming. Logically, naming may proceed by simple print-to-sound conversion (Coltheart, Curtis, Atkins, & Haller, 1993). As such, it provides a strong test for the hypothesis that lexical access taps episodic memory. Participants read words aloud in two sessions, before and after exposure to auditory training tokens. These training tokens were *never* spoken by participants; they merely listened to them. We tested episodic effects in reading by assessing postexposure imitation of the training tokens, in the absence of auditory stimuli. Although speech production is highly practiced, it can show sensitivity to recent experience (Dell, Reed, Adams, & Meyer, 2000). The naming task was followed 1 week later by a recognition memory test, allowing some comparison between direct and indirect memory. Of greater importance, the recognition test helped rule out strategic accounts of the naming data.

METHOD

Participants

The experiment was conducted with two groups, *speakers* and *listeners*. The speakers were 12 Arizona State University students (6 men, 6 women) who completed four experimental stages over a 2-week period. Each received $50 for participation. The listeners were 300 introductory psychology students who participated for course

credit. Listeners were divided into 12 groups, with 25 assigned to each speaker. All speakers and listeners were native speakers of English with (self-reported) normal hearing and normal or corrected-to-normal vision.

Materials

The materials were 160 English words, which were presented to the participants either in printed or spoken form, depending on conditions. All words were bisyllabic, and they were evenly divided into four classes according to frequency of occurrence, with 40 high-, 40 medium high-, 40 medium low-, and 40 low-frequency words. High-frequency words were indexed at >300 occurrences per million, medium high-frequency words ranged from 150 to 250 occurrences per million, medium low-frequency words ranged from 50 to 100 occurrences per million, and low-frequency words were indexed at <5 occurrences per million (Kučera & Francis, 1967). Although naming times were not collected, all frequency classes were balanced with respect to classes of word-initial phonemes.

For auditory training, tokens of all words were recorded by 4 volunteers (2 men, 2 women), with voices that covered a wide perceptual range (Goldinger, 1998). These tokens were recorded in a sound-attenuated booth (IAC Corporation) with a PC, a Beyer-Dynamics microphone, and a Marantz DAT recorder. Words were shown on the computer, and volunteers were asked to say each word twice and to avoid lapsing into a monotone. The tapes were low-pass filtered at 4.8 kHz, digitized at 10 kHz (with a 16-bit A/D processor), and the subjectively clearer token of each word was stored in a digital file. Four groups of 10 volunteers (1 group per speaker) listened to the tokens; all were identified at or above 90%.

Procedure

Speakers. For the speakers, the experiment consisted of four stages, distributed over a 2-week period (Figure 1). In all stages, speakers were tested individually in a sound-attenuated booth equipped with a PC, microphone, and cassette recorder with an external filter and amplifier. In the *baseline* stage, each speaker recorded tokens of all 160 words. Printed words were shown in random order, and the participants were asked to "speak each word clearly, to make a good recording." After recording, the baseline tokens were digitized (CSL, Kay Elemetrics) and were stored for later testing.

For each speaker, the *training* stage was conducted 1 day after baseline recording. The training consisted of listening to spoken versions of the same words and identifying each. The original 160 words were divided into four training sets of 40 words (10 per frequency class). For each training set, the speaker first saw a grid of 40 cells (8×5) on the computer, each containing one word. Spoken words were presented over headphones, with 10 words randomly assigned to each of the four training voices. Upon hearing each word, the speaker searched the grid, using the mouse to click on the word as quickly as possible.

In addition to word frequency, numbers of *repetitions* were manipulated: One (potential) training set of 40 words was never presented, another was presented twice, another was presented 6 times, and another was presented 12 times. Altogether, training entailed 20 blocks, totaling 800 trials (with breaks after every four blocks). The assignment of words to training sets and voices was counterbalanced across participants. Grids were randomly redrawn after each pass through a training set, but all repetitions of each word were in a consistent voice.

In the *re-recording* stage (conducted 1 week after the baseline stage), the speakers again recorded test tokens. As in the baseline stage, they simply read and recorded words from the computer screen. Test tokens were again digitized and stored. Finally, in the *recognition* stage (conducted 1 week after re-recording), each speaker received a visual recognition memory test, on which they sought to discriminate the original 160 words from 160 frequency-matched foils. Accuracy was emphasized, but decision times were also recorded.

Listeners. To obtain an imitation measure, the listeners performed *AXB classification* tasks: Each speaker's baseline (*A*) and test (*B*) tokens were juxtaposed with the training (*X*) tokens that the speakers
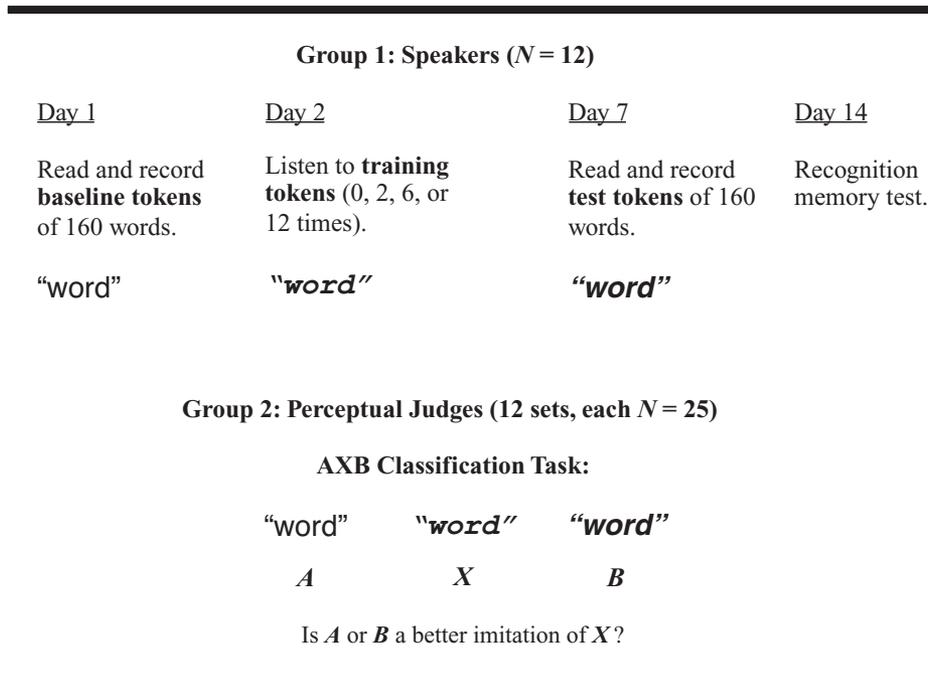

Group 1: Speakers (*N* = 12)

Day 1	Day 2	Day 7	Day 14
Read and record **baseline tokens** of 160 words.	Listen to **training tokens** (0, 2, 6, or 12 times).	Read and record **test tokens** of 160 words.	Recognition memory test.
"word"	*"word"*	*"word"*	

Group 2: Perceptual Judges (12 sets, each *N* = 25)

AXB Classification Task:

"word"	*"word"*	*"word"*
A	*X*	*B*

Is *A* or *B* a better imitation of *X*?

Figure 1. Schematic overview of all experimental stages.

heard in training (see Figure 1). The listeners heard all three words successively and judged which utterance, first or third, was a "better imitation of the middle word."[1] Test tokens were presented equally often in the first and third positions. Testing was conducted in a sound-attenuated room, with 4–8 listeners per group. Each trial began with a 500-msec visual warning (∗∗∗), followed by two response boxes on the computer screen, labeled "first" and "third." After 500 msec, three words were played at approximately 70 dB (SPL), separated by 500-msec intervals. The listeners then indicated whether *A* or *B* sounded more like *X*, by mouse-clicking either box. The likelihood of selecting the test token, rather than the baseline token, was our measure of imitation.[2]

RESULTS AND DISCUSSION

Recognition accuracy and imitation judgments are shown in the upper and lower panels of Figure 2, respectively. The data were analyzed in mixed-model, repeated measures analyses of variance. In recognition, clear effects of word frequency and repetition were observed, along with a strong interaction [all three $Fs(1,11) > 100.0, p < .0001$].[3] Recognition was better for lower frequency words and steadily improved with increased training repetitions. The word frequency effect diminished over repetitions, as all words approached asymptote.

In imitation, main effects of frequency and repetition were again observed, with another interaction [all three $Fs(1,11) > 60.0, p < .0001$]. Imitation was most evident among lower frequency words: Given two repetitions, imitation was reliable [$F(1,11) = 16.3, p < .01$] for LF words and was marginal [$F(1,11) = 4.6, p = .055$] for medium low-frequency words. As Figure 2 shows, imitation generally increased with more training repetitions. But, unlike recognition, the frequency effect *increased* with repetitions, primarily due to rising imitation among lower frequency words. Because the interactions in recognition and imitation were in opposite directions, a strong three-way interaction was observed [$F(1,11) = 159.8, p < .0001$].

The imitation data suggest that episodic traces of spoken words were activated while participants later read those words aloud. The auditory training tokens affected participants' speaking style, despite the passage of 6 days between training and re-reading. In the present *AXB* experiments, no baseline or test tokens were "shadowed." When Goldinger (1998) conducted similar tests, the baseline (*A*) tokens were screen-read, and comparison (*X*) tokens were those heard in training. The present study was similar in both regards; the key difference concerned the test (*B*) tokens. In Goldinger's study, these were shadowing responses: Participants heard training tokens and repeated them aloud. Given this procedure, the observed imitation was perhaps not surprising and could be explained by several theories (e.g., Liberman & Mattingly, 1985). By contrast, the present participants simply read words aloud, before and after exposure to auditory tokens of those same nominal words. Despite this major change, imitation was again observed. This suggests that reading aloud involves more than print-to-sound conversion; it also taps memory for prior perceptual episodes (Jacoby, 1983).
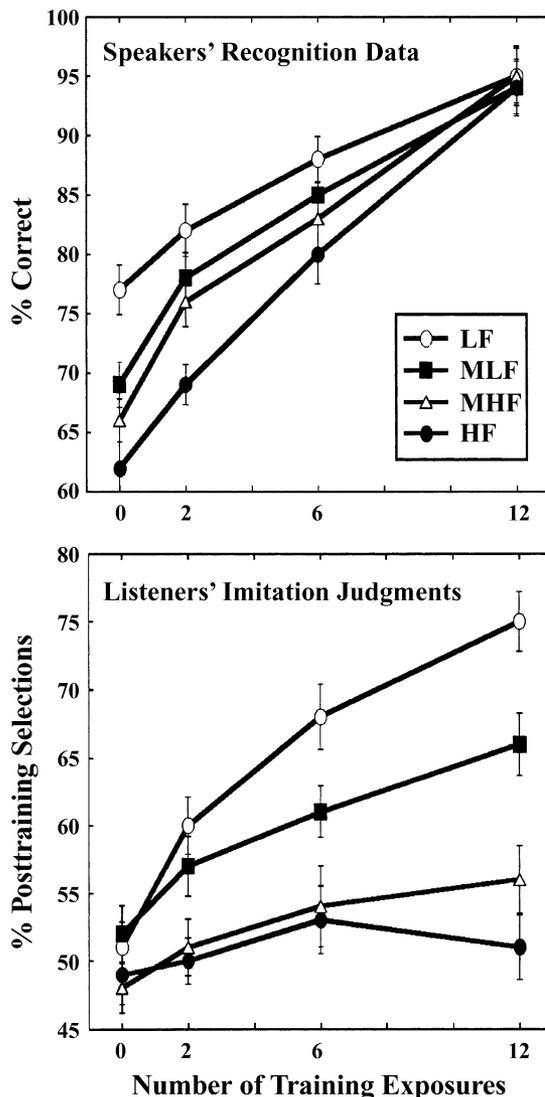
Figure 2. Recognition memory data (upper panel) and *AXB* discrimination data (lower panel). Means (±*SEM*) are shown as a function of training exposures and word frequency. LF = low-frequency words, MLF = medium low-frequency words, MHF = medium high-frequency words, HF = high-frequency words. In each panel, chance performance equals 50%. Note—In recognition, the false-alarm rate to new words was 19.4%.

The strength of imitation was systematically affected by word frequency, which is critical to interpretation. For theoretical relevance, imitation must be a *spontaneous* response, rather than a strategic tendency. Because word frequency is an abstract (nonperceptual) variable, its modulating effect helps rule out demand characteristics or other trivial accounts of imitation. Along similar lines, a dissociation emerged between imitation and recognition. As Figure 2 shows, many words that produced recognition hits failed to evoke appreciable imitation during re-reading. Notably, re-reading occurred a week earlier, when explicit memory would have been stronger. We also compared the

perceived imitation levels of words that eventually produced recognition hits and misses (Figure 3). Although imitation was slightly (2.25%) higher for words that evoked hits [$F(1,11) = 6.6, p < .05$], the frequency and repetition effects were identical across sets [$Fs(1,11) < 2.0$ for both possible interactions]. The similarity of patterns across panels suggests that imitation was not likely driven by explicit recall. Although such dissociations may implicate separate memory systems (Tulving & Schacter, 1990), the present data may simply suggest that imitation and recognition are driven by a common set of episodic traces, tapped in different ways across tasks (Roediger, Weldon, & Challis, 1989).

We have applied a formal model to these data, replicating the qualitative patterns in Figure 2.[4] MINERVA 2 (Hintzman, 1986) is a strongly episodic model which assumes that all experiences create separate, detailed memory traces. For every known word, a collection of prior traces resides in memory. When a new word (or *probe*) is presented, it is compared with all traces in parallel, activating each in proportion to their mutual similarity. Activated traces combine with the probe, forming an *echo*, a "conscious experience" of word recognition. By contacting prior traces, echoes contain information not present in the probe (e.g., conceptual knowledge), thus associating the stimulus to past experience.

Simulations of MINERVA 2 can separately address recognition and imitation because echoes have two distinguishable properties. Because probes activate stored traces to different degrees, echo *content* reflects the unique combination of the probe with a weighted average of relevant, activated traces. Closely related is echo *intensity*, which reflects the total activity created in memory by a probe: Intensity increases with closer matches, or with greater numbers of partial matches. Thus, intensity is partly dependent on content; it reflects stimulus familiarity and can simulate recognition judgments (Hintzman, 1988). In the present case, words were represented as vectors with elements that encoded names, physical details (e.g., fonts and voices), and context. Successful recognition requires word perception, coupled with above-threshold intensity among the context elements—this indicates that the model "knows" a word and has previously experienced that word in the experimental context. Although high-frequency words are well identified by MINERVA 2, the presence of many "preexperiment" traces (stored with random context elements) tends to suppress the association of test words with the experimental context. Thus, lower frequency words have a recognition advantage, but as study traces accrue in memory, all words approach asymptotic performance.

With regard to imitation, printed words constitute probes that contact long-term memory (the essence of perception) and are assembled into responses in working memory. But context again comes into play: When events are encoded in episodic memory, they create multidimensional traces affected by physical characteristics, state of mind, and other factors (Underwood, 1969). In this study, after training, printed words activate traces of their spoken counterparts heard in the same experimental context. Although the subjective experience is that one simply reads aloud, those activated traces seemingly shape the assembled representation in working memory, the presumed basis for task performance. Thus, previously heard words contribute to reading aloud, producing spontaneous imitation.[5]

As in recognition, the population statistics of stored memory traces create a modulating influence of word frequency on imitation. Higher frequency words are represented by many traces. Adding new traces, even context-specific
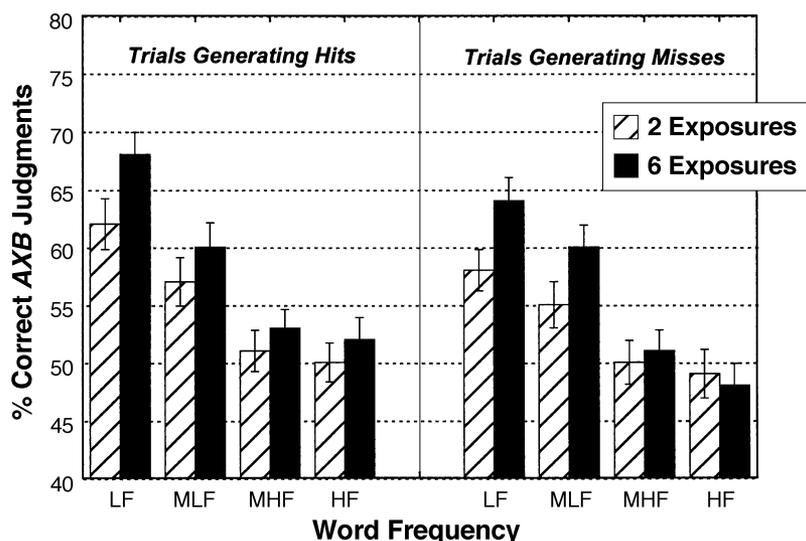

Figure 3. *AXB* discrimination data for words that led to recognition hits (left panel) and misses (right panel). Means (and standard errors) are shown as a function of word frequency and number of training exposures.

matches, has little effect on echo content. Echoes for lower frequency words are more affected by experimental traces, simultaneously creating more imitation and better recognition (the frequency $\times$ repetition interaction is easily derived; Goldinger, 1998). Finally, the dissociation of recognition and imitation derives from different mathematical evaluation of the echoes: "Old" recognition responses are recorded when the summed activity across all context elements exceeds a threshold. By contrast, degrees of imitation are estimated by dot products of "physical elements" between the echoes and training voices. Those dot products tend to increase slowly, relative to the summed activity among context elements.

Other memory models may predict these data, including instance theory (Logan, 1988) and REM (Shiffrin & Steyvers, 1997). However, none (including MINERVA 2) are truly models of word perception. In this regard, a mixed model by Ans, Carbonnel, and Valdois (1998) is promising. This model combines a connectionist network (capable of word naming) with a system of episodic traces. The network embodies statistical relations abstracted from a lifetime of reading experience. The episodic system allows recent traces to affect task performance. In similar fashion, a hybrid connectionist framework by McClelland, McNaughton, and O'Reilly (1995) entails a fast-learning "hippocampal" network and a slow-learning "cortical" network; behavior reflects the interactions of these complementary systems. Finally, a recent Hebbian-learning model by Stark and McClelland (2000) can encode regularities across words, while still keeping perceptual episodes somewhat separate. Such connectionist theories (either hybrid or widely distributed frameworks) can combine stable word perception with episodic influences.

REFERENCES

Ans, B., Carbonnel, S., & Valdois, S. (1998). A connectionist multiple-trace memory model for polysyllabic word reading. *Psychological Review*, **105**, 678-723.

Bowers, J. S. (2000). In defense of abstractionist theories of repetition priming and word identification. *Psychonomic Bulletin & Review*, **7**, 83-99.

Bowers, J. S., & Michita, Y. (1998). An investigation into the structure and acquisition of orthographic knowledge: Evidence from cross-script Kanji–Hiragana priming. *Psychonomic Bulletin & Review*, **5**, 259-264.

Brown, J., & Carr, T. (1993). Limits on perceptual abstraction in reading: Asymmetric transfer between surface forms differing in typicality. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 1277-1296.

Caramazza, A., Laudanna, A., & Romani, C. (1988). Lexical access and inflectional morphology. *Cognition*, **28**, 297-332.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review*, **100**, 589-608.

Dell, G. S., Reed, K., Adams, D., & Meyer, A. S. (2000). Speech errors, phonotactic constraints, and implicit learning: A study of the role of experience in language production. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 1355-1367.

Eich, J. M. (1982). A composite holographic associative recall model. *Psychological Review*, **89**, 627-661.

Forster, K. I. (1979). Levels of processing and the structure of the language processor. In W. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 27-86). Hillsdale, NJ: Erlbaum.

Giles, H., Coupland, N., & Coupland, J. (1991). Accommodation theory: Communication, context, and consequence. In H. Giles, J. Coupland, & N. Coupland (Eds.), *Contexts of accommodation: Developments in applied sociolinguistics* (pp. 1-68). Cambridge: Cambridge University Press.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, **105**, 251-279.

Goldinger, S. D., Azuma, T., Kleider, H. M., & Holmes, V. (2003). Font-specific memory: More than meets the eye? In J. S. Bowers & C. J. Marsolek (Eds.), *Rethinking implicit memory* (pp. 157-196). Oxford: Oxford University Press.

Halle, M. (1985). Speculation about the representation of words in memory. In V. Fromkin (Ed.), *Phonetic linguistics* (pp. 101-114). New York: Academic Press.

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

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, **95**, 528-551.

Humphreys, G., Evett, L., & Quinlan, P. (1990). Orthographic processing in visual word identification. *Cognitive Psychology*, **22**, 517-560.

Jackson, A., & Morton, J. (1984). Facilitation of auditory word recognition. *Memory & Cognition*, **12**, 568-574.

Jacoby, L. L. (1983). Remembering the data: Analyzing interactive processes in reading. *Journal of Verbal Learning & Verbal Behavior*, **22**, 485-508.

Jacoby, L. L., & Hayman, C. A. G. (1987). Specific visual transfer in word identification. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **13**, 456-463.

Kohonen, T., Lehtiö, P., & Oja, E. (1989). Storage and processing of information in distributed associative memory systems. In G. E. Hinton & J. A. Anderson (Eds.), *Parallel models of associative memory* (pp. 129-167). Hillsdale, NJ: Erlbaum.

Kučera, H., & Francis, W. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, **21**, 1-36.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, **95**, 492-527.

Marslen-Wilson, W., & Warren, P. (1994). Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review*, **101**, 653-675.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, **102**, 419-437.

McClelland, J. L., & Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General*, **114**, 159-188.

Morton, J. (1979). Word recognition. In J. Morton & J. C. Marshall (Eds.), *Structures and processes* (pp. 109-156). Cambridge, MA: MIT Press.

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, **104**, 266-300.

Oldfield, R. C. (1966). Things, words, and the brain. *Quarterly Journal of Experimental Psychology*, **18**, 340-353.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. E. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, **103**, 56-115.

Roediger, H. L., III, Weldon, M. S., & Challis, B. H. (1989). Explaining dissociations between implicit and explicit measures of retention: A processing account. In H. L. Roediger III & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in Honour of Endel Tulving* (pp. 3-41). Hillsdale, NJ: Erlbaum.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM–retrieving effectively from memory. *Psychonomic Bulletin & Review*, **4**, 145-166.

STARK, C. E. L., & MCCLELLAND, J. L. (2000). Repetition priming of words, pseudowords, and nonwords. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 945-972.

TENPENNY, P. L. (1995). Abstractionist versus episodic theories of repetition priming and word identification. *Psychonomic Bulletin & Review*, **2**, 339-363.

TULVING, E., & SCHACTER, D. L. (1990). Priming and human memory systems. *Science*, **247**, 301-306.

UNDERWOOD, B. J. (1969). Attributes of memory. *Psychological Review*, **76**, 559-573.

NOTES

1. Prior to conducting the *AXB* tests, all tokens from each speaker were screened to ensure that no words had qualitative phonological changes across recording sessions. Three words (out of 1,920) were eliminated by the first author; later testing with 10 volunteers validated these decisions.

2. With respect to imitation, two points of clarification are important. First, although imitation is detectable with *AXB* classification, the associated speech changes are quite subtle. For example, merely increasing the delay between tokens reduces listeners' sensitivity to the effect. Moreover, when informed of the experimental hypothesis in debriefing, none of the speakers believed their "reading voices" had been affected by the training tokens. Second, although the *AXB* classification data confirm that detectable imitation occurred, they do not reveal its perceptual basis. We explored this in other *AXB* experiments, using tokens that were modified by signal-processing software. By systematically removing potential cues to imitation, we found that duration and prosody were the best predictors, followed by small effects of amplitude and spectral characteristics.

3. In all reported ANOVAs, the data were grouped corresponding to the individual speakers. This was done to keep degrees of freedom constant across the recognition and imitation data, allowing easy statistical comparison. Grouping data in other manners (e.g., by items or groups of listeners) had no bearing on the patterns of reliable effects.

4. Notably, one finding was not well fit by the model. As can be seen in Figure 1, the perceived imitation of high-frequency words increased with 2 and 6 training exposures, relative to 0. However, imitation levels then decreased after 12 exposures, contrary to prediction.

5. It is often hypothesized that activated knowledge in long-term memory imposes itself in concurrent task performance, as in a Stroop task. In the present case, we hypothesize that activated traces of spoken words manifest themselves as subtle changes in reading aloud. An analogous, more salient phenomenon is "word insertion" while speaking. For example, while reading the newspaper over breakfast, you may ask someone to "pass the budget."