# Computing the Meanings of Words in Reading:



Mark S. Seidenberg
University of Wisconsin–Madison

Psychological Review
2004, Vol. 111, No. 3, 662–720


0033-295X/04/$12.00 DOI: 10.1037/0033-295X.111.3.662

# Computing the Meanings of Words in Reading: Cooperative Division of Labor Between Visual and Phonological Processes

Michael W. Harm
Stanford University School of Medicine

Mark S. Seidenberg
University of Wisconsin—Madison

Are words read visually (by means of a direct mapping from orthography to semantics) or phonologically (by mapping from orthography to phonology to semantics)? The authors addressed this long-standing debate by examining how a large-scale computational model based on connectionist principles would solve the problem and comparing the model's performance to people's. In contrast to previous models, the present model uses an architecture in which meanings are jointly determined by the 2 components, with the division of labor between them affected by the nature of the mappings between codes. The model is consistent with a variety of behavioral phenomena, including the results of studies of homophones and pseudohomophones thought to support other theories, and illustrates how efficient processing can be achieved using multiple simultaneous constraints.

Although humans have been reading for several thousand years and studying reading for more than a century, the mechanisms governing the acquisition, use, and breakdown of this skill continue to be the subject of considerable interest and controversy (see Adams, 1990; National Institute of Child Health and Human Development, 2000; Rayner, Foorman, Perfetti, Pesetsky, & Seidenberg, 2001, for reviews). The present article focuses on a central aspect of reading, the processes involved in determining the meanings of words from print.

In principle, a skilled reader could determine the meaning (or meanings) of a word directly from knowledge of its spelling. However, alphabetic orthographies, in which written symbols represent sounds, afford another possibility: Spelling could be translated into a phonological representation that is then used in determining a word's meaning. These mechanisms have traditionally been termed *direct* (orthography to meaning) and *phonologically mediated* (orthography to phonology to meaning) lexical access. The extent to which one or the other mechanism is used is a classic issue in reading research, one whose importance is magnified by its relevance to concerns about how reading should be taught (Rayner et al., 2001).

This debate is very old, and contemporary views range from those that assign no useful role to phonological processing in the computation of meaning to the view that phonological recoding is obligatory. There is also a reconcilist position, which holds that both mechanisms are important but under different conditions (e.g., as a function of type of word, type of orthography, or skill level). The pendulum has swung between the extremes with considerable regularity (compare the overviews provided by Coltheart, 1978; Frost, 1998; McCusker, Hillinger, & Bias, 1981; Smith, 1973).

In this article we propose a resolution of this debate that emerged from considering the issues from a computational perspective. Theories of reading (and the design of behavioral experiments) have been closely tied to intuitions about how the process works derived from extensive personal experience. However, the phenomenon we are trying to understand is a process that is largely unconscious: People are aware of the outcome of this process—that words are understood—not the mental operations involved in achieving it. The computational approach used here represents an attempt to address the nature of underlying mechanisms at a level that intuition does not easily penetrate. We developed a model of the computation of word meaning from print based on general computational principles that have been explored in previous research on reading (Harm & Seidenberg, 1999; Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989) and other phenomena. Whereas our earlier reading models focused on the translation from print to sound, the present model addresses reading for meaning. Conversely, Hinton and Shallice (1991) and Plaut and Shallice (1993) addressed complementary issues concerning the computation from orthography to meaning in their work on acquired deep dyslexia, an unusual reading impairment

Some of this research was conducted while both authors were at the University of Southern California and while Michael W. Harm was a postdoctoral fellow at Carnegie Mellon University. This work was supported by National Institute of Mental Health (NIMH) Grant MH47566, National Institute of Child Health and Human Development Grant 29891, Research Scientist Development Award NIMH MH01188, and National Research Service Award DC00425 from the National Institute of Deafness and Other Communication Disorders.

We thank David Plaut, Maryellen MacDonald, Robert Thornton, Jason Zevin, Gerry Altmann, and Jelena Mirkovič for helpful comments on the article. This article was in process for several years, during which it was known as the "monster in a box." With this in mind, we regretfully note the passing of Spaulding Gray (1941–2004).

Correspondence concerning this article should be addressed to Michael W. Harm, Department of Information Resources and Technology, Stanford University School of Medicine, MSOB x300, 251 Campus Drive, Stanford, CA 94305-5412, or to Mark S. Seidenberg, Language and Cognitive Neuroscience Laboratory, Department of Psychology, University of Wisconsin—Madison, 1202 West Johnson Street, Madison, WI 53706. E-mail: mharm@stanford.edu or seidenberg@wisc.edu

observed following some types of brain injury. Our model builds on this work but differs from it insofar as it is the first large-scale model to address how meaning is computed in a system in which both visual (orth–sem) and phonologically mediated (orth–phon–sem) pathways are available. The implemented model was then assessed against a body of critical findings from behavioral studies.

As it turns out, the proposed model is consistent with many aspects of earlier accounts but differs from them in important respects because of specific properties of the computational mechanisms that are used. Within this framework, the meaning of a word is a pattern of activation over a set of semantic units that develops over time based on continuous input from both orth→sem and orth→phon→sem components of the "triangle" (see Figure 1). The main theoretical issue concerns the computational considerations that determine how the model (and by hypothesis, the reader) arrives at an efficient division of labor between these sources of input. Thus the concept of independent visual and phonological recognition routines, one of which (e.g., the fastest finishing) provides access to meaning (e.g., Caplan, 1992; Carr & Pollatsek, 1985; Frost, 1998; McCusker et al., 1981), is replaced by a cooperative computation in which semantic patterns reflect the joint effects of input from different sources. The manner in which the division of labor emerges in the model relates well to findings concerning the primacy of phonological codes in reading acquisition. The model is also consistent with and provides insight about a number of important empirical findings concerning the processing of homophones (e.g., BARE–BEAR)[1] and pseudohomophones (e.g., BAIR) that have figured prominently in previous accounts.

The structure of the article is as follows. We first review the pretheoretical arguments and critical empirical data that led to differing conclusions about the importance of direct versus phonologically mediated access. There are good arguments on both sides of the debate as the inconclusive state of current theorizing would predict. We then describe an approach to this issue based on general computational principles concerning knowledge representation, acquisition, and processing derived from the connectionist, or parallel distributed processing (PDP), approach (Rumelhart, McClelland, & the PDP Research Group, 1986). A computational model embodying these principles and other assumptions about critical characteristics of reading and the conditions under which children learn to read is introduced and analyzed, and its behavior is linked to empirical findings. In the GENERAL DISCUSSION section we summarize the important properties of the model and consider some limitations of the work, unresolved issues, and directions for future research.

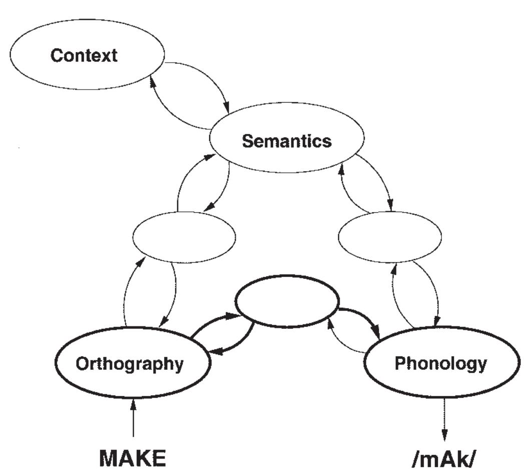


*Figure 1.* The "triangle" model of Seidenberg and McClelland (1989). The implemented model examined how phonological codes are computed from orthography. The present research examined processes involved in computing semantic codes from orthography, given the availability of both direct (orth→sem) and phonologically mediated (orth→phon→sem) pathways. From "A Distributed, Developmental Model of Word Recognition and Naming," by M. S. Seidenberg and J. L. McClelland, 1989, *Psychological Review, 96,* p. 526. Copyright 1989 by the American Psychological Association.

## INTUITIONS AND EVIDENCE

This section provides an overview of previous research on visual and phonological processes in reading. Before we proceed, a terminological issue needs to be addressed. Basic processes in reading are often discussed in terms of "models" that illustrate theoretical claims (e.g., Coltheart, Curtis, Atkins, & Haller, 1993; LaBerge & Samuels, 1974; Marshall & Newcombe, 1973; Morton, 1969; Seidenberg & McClelland, 1989). Models that incorporate both direct-visual and phonologically mediated computations from print to meaning are often termed "dual-route models" (see Frost, 1998, for a recent example and discussion of this use of the term). However, this usage is potentially confusing, because the term has also been extensively used in the reading literature in reference to a different issue, the mechanism(s) involved in generating pronunciations from print (e.g., Coltheart et al., 1993).

That there are both direct-visual and phonologically mediated mappings from print to meaning is not a theoretical claim specific to any particular model of reading. Rather, the basic design feature of alphabetic writing systems is that although strings of letters can be directly associated with meanings (as can other visual stimuli such as @), the letters also represent the sounds of words, which are in turn associated with one or more meanings. Where theories differ is with respect to how these lexical codes and relations between them are structured, how this knowledge is acquired and represented in memory, and what roles these types of information play in reading.[2]

[1] The following notational conventions are used in this article. The written form of a word is shown in small caps, the phonological form is coded in International Phonetic Alphabet notation between slashes, the semantic concept for the item is shown in braces, and the semantic features comprising that concept are denoted in brackets. Hence the visual form CAT corresponds to the phonological representation /kæt/ and the semantic concept {cat}, which consists of semantic features such as [feline], [has-fur], [living-thing], and so forth.

[2] All alphabets exhibit strong correspondences between spelling and sound, thus affording the phonologically based reading process for most words. Orthographies vary in the extent to which they admit exceptions to these central tendencies. English is relatively "deep" (i.e., spelling–sound correspondences are less consistent) compared to "shallower" alphabets such as the ones for Serbo–Croatian and Italian (Frost, Katz, & Bentin, 1987). In connectionist models (Seidenberg & McClelland, 1989), phonological codes can be correctly computed from orthography for all words (including "exceptions" such as PINT and HAVE), whereas dual-route models of naming assume that the exceptions require a separate mechanism.

Whereas the above sense of "dual-route model" refers to mechanisms for translating from print to meaning, the term also refers to a specific theoretical proposal, studied for many years by Coltheart and others (e.g., Paap & Noel, 1991), concerning mechanisms for translating from print to sound. According to this theory, pronouncing letter strings in English (words and pseudowords such as NUST) requires two mechanisms, one involving knowledge of whole words and one involving rules governing the correspondences between graphemes and phonemes. All theories of reading are *not* dual route in this sense; in particular, connectionist models dating from Seidenberg and McClelland (1989) have suggested that the functions achieved by the two mechanisms in dual-route models arise from a single connectionist mechanism (see also Glushko, 1979). These alternative theories are the subject of continuing research and debate but are not the focus of the present article.[3]

## Evidence for Direct Access

For many years, the standard view among reading researchers and educators was that direct-visual access is the efficient way to read for meaning. The basic argument was that phonological recoding is an extra computational step that skilled readers avoid. Three aspects of the English orthography were also thought to work against the use of phonology. First, English has a large number of homophones (phonological forms such as /pleyn/ that are associated with two or more spellings and meanings). Phonological recoding would therefore create ambiguities that could be avoided by computing directly from print to meaning. Second, using arguments from signal-detection theory, Smith (1973) concluded that a two-stage decoding process (orth→phon, phon→sem) would be too slow to support automatic, rapid reading and that skilled reading must rely on direct access. Finally, Smith (1971) and others have argued that even though the orthography is alphabetic, the correspondences between spelling and sound in English are extremely complex, given the inconsistencies illustrated by pairs such as MINT–PINT and GAVE–HAVE. Mastering such a complex set of pronunciation rules was thought to be a daunting task and thus not the path to skilled reading. As Smith (1983) asserted, "Reading by 'phonics' is demonstrably impossible. Ask any computer" (p. 5). He concluded that only the orth→sem mechanism is viable. These arguments have had enormous impact on educators responsible for formulating programs for teaching reading in schools; they provided a foundation for the *whole language* approach that discourages direct instruction in spelling–sound correspondences (Rayner et al., 2001).

Through the early 1980s, the evidence that phonological recoding plays a causal role in the access of meaning was equivocal (McCusker et al., 1981; Perfetti & McCutchen, 1982). It was very difficult to create conditions that showed not merely that readers activated phonological information but that they used this information in accessing meaning. Several models that emphasized visually based recognition procedures were proposed (e.g., Baron, 1973; McClelland & Rumelhart, 1981; Paap, Newsome, McDonald, & Schvanevelt, 1982). Coltheart (1978) also argued strongly for direct-visual access.

## Evidence for Phonological Mediation

Over the past 20 years the direct-visual-access view has been strongly called into question. The direct-access view has an air of paradox about it: The development of writing systems since about 2500 B.C.E. has been toward symbols that represent sounds rather than meanings (Hung & Tzeng, 1981). Why are there alphabetic writing systems if phonological information plays no useful role in reading? There is now strong evidence for the extensive use of phonology in reading for meaning in English and other languages (e.g., Perfetti, Bell, & Delaney, 1988; Van Orden, Johnston, & Hale, 1988), derived from behavioral studies of children and adults and from observations about differences between the mappings between spelling and sound versus spelling and meaning that affect learning. We summarize this evidence and related arguments briefly (for fuller discussion, see Frost, 1998; Rayner & Pollatsek, 1989; Van Orden, Pennington, & Stone, 1990).

Children have large spoken-word vocabularies by the time reading instruction begins. Reading, on this view, involves learning how written symbols relate to known spoken word forms. In alphabetic orthographies such as the one for English, written symbols represent sounds, specifically phonemic segments. Thus, successful reading acquisition requires developing segmental representations of speech and grasping the "alphabetic principle" concerning the mapping between letters (or combinations of letters) and phonemes (Gathercole & Baddeley, 1993; Liberman, Shankweiler, & Liberman, 1989).

Jorm and Share (1983) further observed that the ability to sound out words (either overtly or covertly) gives the child a self-teaching mechanism that facilitates learning to read: The child can

[3] Historically, the issue of direct versus phonologically mediated mechanisms for translating from print to meaning predates the issue of whether there are one or two mechanisms for translating from print to sound. Crowder (1982), for example, traced interest in the print-to-meaning issue to St. Augustine. In the modern era, important early studies included Rubenstein, Lewis, and Rubenstein (1971); Meyer, Schvaneveldt, and Ruddy (1974); LaBerge and Samuels (1974); and Baron (1973). Coltheart (1978) provided a review of studies to that date. Interest in the topic waned in the 1980s as many researchers turned their attention to the mechanisms that underlie overt pronunciation. It was in connection with this issue that Coltheart introduced the term *dual-route model,* which referred to proposed *lexical* and *sublexical* pronunciation procedures (see general introduction to Patterson, Marshall, & Coltheart, 1985, for an overview; see Marshall & Newcombe, 1973, for an earlier version of this account). However, others subsequently adopted the term in reference to the direct-visual and phonologically mediated computations from print to meaning, probably in part because it seemed more felicitous than other terms that had been used, such as the *dual-encoding hypothesis* (Meyer et al., 1974) or *parallel coding systems models* (Carr & Pollatsek, 1985). As recently as 2000 Coltheart used this term in reference to both the computation of meaning (direct vs. phonologically mediated) and the computation of phonology (lexical vs. sublexical procedures; Coltheart, 2000). We think this usage is confusing, however, for the following reason: Evidence that there are "dual" visual and phonologically mediated mappings to *meaning,* which is true of all alphabets, often registers as evidence for the dual-route model of *pronunciation* and the claim that there are two mechanisms for pronouncing letter strings. Because of this ambiguity, because *dual-route model* is used in different ways in different contexts, and because our model differs from the Coltheart pronunciation model with which the term is strongly associated, we avoid it in the remainder of this article.

sound out a word and determine whether it matches a known spoken word. Connectionist models provide a mechanistic interpretation of this type of learning. The comparison between the self-generated pronunciation and information about a word's sound can be seen as the basis for computing an error signal that allows adjustment of the weights on connections mediating the orth→phon mapping.

Van Orden and colleagues (Van Orden, 1987; Van Orden et al., 1988, 1990) have presented a somewhat different argument. They have observed that in English, orthography and semantics are largely uncorrelated, whereas orthography and phonology are highly correlated; thus, the former should be harder to learn than the latter. As Van Orden et al. (1990) stated, "We propose that the relatively invariant correspondence between orthographic representations and phonologic representations explains why word identification appears to be mediated by phonology" (p. 513). Flesch (1955) also made this argument with greater polemical fervor, asserting that teaching children to read by rote memorization of the associated meanings of word forms rather than logical deduction of the sounds of words "consists essentially of treating children as if they were dogs. . . . It's the most inhuman, mean, stupid way of foisting something on a child's mind" (p. 126).

Merging these arguments yields a theoretical stance in which orth→phon is easier to learn than orth→sem, and phon→sem is already known for many words. Hence, early reading relies on orth→phon→sem much more than orth→sem.

There is considerable evidence that children use phonological information in reading (Liberman & Shankweiler, 1985) and that the quality of phonological representations is strongly related to reading achievement (Snowling, 1991). The most compelling evidence derives from studies showing that prereaders' knowledge of phonological structure is predictive of reading achievement several years later (Bradley & Bryant, 1983; Lundberg, Olofsson, & Wall, 1980). There is also evidence that impairments in the representation of phonology are often observed in individuals with dyslexia (see Harm & Seidenberg, 1999, for a summary and a computational model of these effects). These developmental results find a natural interpretation within a theory that states that orthographic patterns activate phonological representations early in the process of reading words for meaning.

Given the extensive evidence for the use of phonological information in beginning reading, it has been often assumed that this strategy carries over to skilled adult reading. Frost (1998) termed this the *strong phonology* theory. Many studies of adult readers support this view; here, we review some critical findings that are relevant to the simulations reported below.

A classic study by Van Orden (1987) yielded direct evidence that phonological information has a causal role in the access of meaning. Participants performed a semantic decision task in which they had to decide if a target word was an exemplar of a specified category. For example, for the food category, the targets were either true exemplars (e.g., MEAT), homophonous foils (e.g., MEET), or nonhomophonous spelling controls (e.g., MOOT). Van Orden found that participants made a high number of false-positive responses on phonological foils relative to orthographic controls. Participants would not make false-positive responses unless they were activating phonological information and using it to access meaning. Later experiments (e.g., Van Orden et al., 1988) yielded the same effect for pseudohomophone stimuli (e.g., category: clothing; target: SUTE). These results were taken to indicate that word recognition progresses from spelling to sound to meaning, with homophones such as BEAR or PLANE disambiguated by a late "spelling check" procedure after meanings have been accessed.

Perfetti et al. (1988) demonstrated effects of the phonological form of words at a very early stage of processing. They found that when a word is presented very briefly and then masked by a homophonous word mask, identification of the target word is facilitated relative to a neutral mask. These results also suggest that the phonological form of a word is activated automatically at a very early stage in processing.

Lesch and Pollatsek (1993) and Lukatela and Turvey (1994a) extended these findings using homophones in a semantic priming paradigm. If the access of meaning is initially phonological, and homophones are disambiguated by a subsequent spelling check, there should be a point early in processing at which homophones activate multiple meanings; later, after the spelling check has occurred, only the appropriate meaning should be active. Lesch and Pollatsek and Lukatela and Turvey (1994a) used a masked priming paradigm to explore this hypothesis. In the critical conditions, a target such as FROG was preceded by a related prime such as TOAD or its semantically unrelated homophone TOWED. The prime word was presented for either a short (50 ms) or long (250 ms) duration and then masked by the target, which was to be named. Semantically related prime–target pairs (e.g., TOAD–FROG) produced facilitation compared to an unrelated control condition at both prime durations. Inappropriate primes (e.g., TOWED–FROG) produced facilitation only in the short condition. Thus the effects were consistent with Van Orden et al.'s (1990) account in which meaning is initially activated via phonology with homophones subsequently disambiguated by a spelling check. Masking the stimuli at an early stage in processing (50 ms) removes the orthographic information that normally supports the spelling check.

## RECONCILIST THEORIES

Although considerable attention has focused on direct-visual access and phonologically mediated access as competing alternatives, other theories have assumed that readers make use of both, with several factors determining which pathway will be dominant in a given situation (e.g., Baron & Strawson, 1976; Carr & Pollatsek, 1985; see Seidenberg, 1995, for discussion). The strong direct visual access position advocated by Smith (1971, 1973) cannot be correct; there are too many studies showing unambiguous phonological effects in reading for meaning. Moreover, Smith's (1973) argument about the difficulties involved in using phonological mediation rests on the assumption that spelling–sound correspondences are encoded by rules. Connectionist models such as Seidenberg and McClelland's (1989) subsequently provided an alternative in which the correspondences are encoded by weights on connections between units involved in the orthography–phonology mapping. Such systems can encode different degrees of consistency in the orth→phon mapping operating over many different orthographic and phonological subunits. Thus, the model instantiated a theory of how readers could efficiently activate phonological codes for all words, including ones that involve atypical mappings.

The strong version of the phonological mediation theory has also been questioned, however. Every normal individual can recognize and access conceptual information associated with objects without an intermediate phonological recoding step; why wouldn't this be possible when the objects in question happen to be familiar letter strings? Individuals who are profoundly deaf from birth and have not received speech training can determine the meanings of printed words, even when lacking access to phonological information. This observation suggests that meanings can be computed directly from print, but it leaves open the extent to which this process is used by individuals who also have access to phonology.

Other questions arise concerning the processing of homophones. The many homophones in English present a complication for a system in which meanings are exclusively activated through phonology; these words will have to be disambiguated every time they are read, which would seem to impose a considerable burden on the reading system, a burden that would be avoided if meanings were accessed directly from print. The solution that Van Orden (1987) proposed was a very rapid spelling check following the initial, phonologically driven activation of meaning, that is, comparing the activated meanings against the spelling of the word to determine which is correct. The spelling check idea seems to entail that the reader be able to compute the arbitrary association between a word's meaning and its spelling. If readers are able to compute from meaning to spelling, it is not clear why they would not be able to compute from spelling to meaning. Thus, a realistic implementation of the spelling check procedure seems to require mastering the kind of arbitrary mapping that is proscribed in strong phonology theories (Seidenberg, 1995).

There is also an empirical question: Jared and Seidenberg (1991) provided evidence that the extent to which phonology enters into the activation of meaning varies as a function of word frequency. They replicated the Van Orden (1987) results but also experimentally manipulated the frequencies of the exemplars (e.g., ROSE) and homophone foils (e.g., ROWS). In their studies, only homophones with two low-frequency meanings generated significant false positives. Higher frequency words did not yield significant false positives. Insofar as the presence of false-positive effects has provided the basis for diagnosing the use of phonological information, the absence of these effects could be taken as evidence that this information was not used.

The Jared and Seidenberg (1991) results have generated controversy, focused on the possibility that the failure to observe significant false positives in the higher frequency conditions was a Type II error. Lesch and Pollatsek (1993) did not explicitly manipulate prime frequency in their study, but they reported a post hoc analysis that revealed no effect of frequency on the magnitude of priming. Lukatela and Turvey (1994a) did manipulate prime frequency but found that it had no effect insofar as both high- and low-frequency conditions yielded evidence for phonologically based activation of meaning. However, as discussed below, the frequency manipulation in this study was quite weak, and other aspects of the stimuli and analysis raise questions about the results.

The processing of homophones and pseudohomophones is a major focus of the modeling described below. To foreshadow the results, we note that the model behaves somewhat differently than both Van Orden et al. (1990) and Jared and Seidenberg (1991) proposed and provides a reconciliation of their findings.

Logical and observational arguments about the relative ease of learning the orth→phon and orth→sem mappings also need to be examined carefully. The relationship between spelling and meaning is often said to be arbitrary and therefore difficult to learn because there is nothing about the spelling of a word such as DOG that demands that it, rather than some other spelling pattern, be associated with the concept {domestic canine}. However, English and some other alphabetic writing systems exhibit nonarbitrary form–meaning correspondences. For example, DOG makes similar semantic contributions to many related words (DOGS, DOGLEG, DOGHOUSE, etc.)—word-final *-s* often indicates plurality, word-final -ED usually indicates pastness, and so on. There are other correlations between sound (and hence spelling) and meaning, illustrated by words such as GLITTER, GLISTEN, GLEAM, GLINT, GLARE and SLIP, SLIDE, SLITHER (see Marchand, 1969, for many examples). Further, as Chomsky and Halle (1968) noted, English spelling preserves morphological information over phonological in many cases, such as SIGN–SIGNATURE and BOMB–BOMBARD. Shallow orthographies such as the one for Serbian sacrifice this morphological information in favor of preserving spelling–sound consistency. Seidenberg and Gonnerman (2000) discussed the role of such nonarbitrary form–meaning correspondences in the development of morphological representations. Although the mapping from spelling to meaning is less systematic than from spelling to sound in English, it is far from arbitrary (see also Kelly, 1992).

It is also clear that with sufficient training connectionist models can learn arbitrary mappings. Moreover, it should be noted that even words with highly unusual pronunciations are not wholly arbitrary and therefore partially overlap with other words. Higher frequency words may be encountered often enough for the orth→sem mapping to become established relatively quickly regardless of the degree of inconsistency in pronunciation. In general, conjectures about the relative ease of learning different types of mappings needs to be examined using explicit models of these computations.

Considerations such as these support a theory incorporating both direct-visual and phonologically mediated processes. Which pathway provides access to meaning for a given word is thought to depend on factors such as the relative speed of the two mechanisms, word frequency, orthographic–phonological regularity, and the depth of the orthography (Frost, Katz, & Bentin, 1987; Henderson, 1982; Seidenberg, 1995).

## SUMMARY

The literature to date has focused on empirical evidence and theoretical arguments concerning the relative prominence of the direct-visual and phonologically mediated mechanisms. Each alternative continues to have strong proponents: Researchers who mainly study issues concerning reading acquisition and dyslexia tend to emphasize the importance of phonological coding (e.g., Wagner & Torgesen, 1987), whereas many researchers who mainly study visual word recognition in adults have focused on the role of orthography (e.g., Grainger & Jacobs, 1996). The modeling work described below represents an attempt to end this impasse by treating the issue as a computational one. We did not build the model with a particular answer to the division of labor question in mind; rather, we asked, given a model of the computation of meaning based on the principles explored in our previous work, how would it solve the problem? In particular, what are the computational factors that determine the division of labor given an architecture in which both pathways can activate semantics? We then asked whether the model was consistent with facts about reading acquisition and skilled performance and whether it provided further insight about these phenomena.

The model that we describe has an affinity to the reconcilist models in the sense that both visual and phonological processes can activate lexical semantics; of importance, however, these components are not independent. Rather than parallel processing routes that develop independently and operate in parallel, with one or the other providing access to meaning, our model emphasizes the dependence between the two and the way they jointly and cooperatively achieve an efficient solution in the course of learning to master the task.[4]

Our model is closer in spirit to Van Orden et al.'s (1990) discussion of a lexical system in which all parts are operating simultaneously and therefore contributing to the activation of meaning. Our work differs from their account in some ways, however. Although Van Orden et al. (1990) discussed a "resonance" theory in which all components of the lexical system are continuously interacting, they also emphasized the primacy of the orth→phon→sem component and suggested that the role of orth→sem was minimal because of the arbitrariness of the mapping. In implementing a computational model, we found that it behaved in ways that suggest a somewhat different picture of the role of the orth→sem component. Our work also places greater emphasis on the mutual dependence of the two components: what each component contributes to the activation of semantics depends on what the other contributes. This division of labor develops in the course of learning to master the task, and we devote considerable attention to the factors that affect it and its relevance to reading behavior.

The remainder of the article is structured as follows: The DESIGN CONSTRAINTS section summarizes the principles and assumptions that guided the development of the model. We then describe the simulations, which were conducted in two phases. Phase 1 involved training the phonological and semantic attractors and the mappings between them; this was intended to approximate the kinds of lexical knowledge that children possess in advance of learning to read. Phase 2 involved introducing orthography. For both phases we provide details about the model's architecture and training, summarize overall performance, and then compare the model's performance to behavioral data. We then describe simulations of central behavioral phenomena. We conclude by discussing limitations of the model and future directions.

Our presentation necessarily goes into considerable detail concerning the motivation for the approach; the structure of the model, which incorporates some technical innovations; descriptions and analyses of the model's behavior; and comparisons to behavioral data. This material is in the service of addressing four central issues.

1. *Cooperative computation of meaning.* One principal goal was to examine the feasibility of a system in which semantic activity is determined by computations involving both orth→sem and orth→phon→sem and to explore the extent to which such a model is compatible with evidence concerning human performance.

2. *Transition from beginning to skilled reading.* The major feature of this transition is that whereas beginning reading relies heavily on phonological information, in skilled reading the role of the visual process increases greatly. The model addresses why this developmental sequence occurs.

3. *Processing of homophones and pseudohomophones.* Studies of these stimuli have provided critical evidence concerning the role of phonological information in word reading. Deciding that BEAR means {ursine animal} not {naked} requires using information about the relationship between spelling and meaning. Pseudohomophones such as BAIR provide a way to diagnose whether phonological information has been activated and raise questions about the role of orthography in determining that they are not actual words. The model addresses the nature of the computations involved in processing such stimuli.

4. *Differential effects of masking.* We used the model to study effects of the masking procedure used in many studies in this area. The model suggests that masking has somewhat different effects than standardly assumed in interpreting the results of such studies and invalidates some of the conclusions standardly drawn from such data.

## DESIGN CONSTRAINTS

The present research is part of the ongoing development of a theory of word reading. In studying the computation of meanings from print, we used the same research strategy as in our previous work on the computation of phonology (Harm & Seidenberg, 1999; Plaut et al., 1996; Seidenberg & McClelland, 1989). We work back and forth between a high-level theory of how people read and computational models that instantiate parts of the system. The theory is based on principles concerning knowledge representation, learning, and processing that are components of the PDP approach (Rumelhart, McClelland, & the PDP Research Group, 1986). These principles are general—thought to underlie many aspects of perception and cognition—rather than specific to reading. This is consistent with the observation that reading, a technology invented relatively recently in human history, makes use of capacities that did not evolve specifically for this purpose. The theory also incorporates considerations that are reading specific (e.g., concerning the conditions under which children learn to read). The computational model is an implementation of important aspects of the theory; it acts both as a test of the adequacy of proposed mechanisms and as a discovery procedure, that is, a source of additional insight about the behavior in question. The results of the modeling can lead to modifications or extensions of both the reading theory and the general computational approach.

In this section we discuss the factors that determined the form of the implemented model. These design constraints involved three kinds of considerations:

1. Computational considerations. The principles underlying PDP models and their rationale have been discussed elsewhere (e.g., McLeod, Plunkett, & Rolls, 1998; O'Reilly & Munakata, 2000; Rumelhart, McClelland, & the PDP Research Group, 1986); below we focus on

[4] Coltheart, Davelaar, Jonasson, and Besner (1977) discussed the possibility that visual and phonological mechanisms cooperatively activate entries in the mental lexicon, but they rejected this view in favor of direct-visual access because the phonological pathway was thought to operate too slowly to contribute significantly.

properties that played the most important roles in determining our model's behavior.

2. Facts about reading acquisition. The way the model was structured and trained reflected observations about the capacities that children bring to bear on learning to read and critical aspects of their early reading experience.

3. Practical and theoretical considerations that led us to focus on specific aspects of the task and make simplifying assumptions about others.

## Architectural Homogeneity

Standard dual-mechanism approaches (e.g., Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001) assume that there are separate mechanisms involving different types of knowledge and processes. The phonological mechanism is usually assumed to involve rules governing spelling–sound correspondences, whereas the direct-visual route involves lexical lookup or an interactive-activation procedure. The mechanisms behave differently because they are constructed out of different elements and governed by different principles. The system that we implemented (like other PDP models) is homogeneous in the sense that all computations involve the same kinds of structures (distributed representations of orthographic, phonological, and semantic codes) and computations (equations governing the spread of activation along weighted connections between units). This is a central tenet of the reading theory, one that distinguishes it from other approaches. The homogeneity assumption is motivated by two main considerations. First, we wanted the model's division of labor to emerge in the course of learning to perform the task, not as a consequence of built-in differences between the two mechanisms, because we think this is how children solve the problem. Second, we assume that the brain uses the same basic mechanisms to encode different lexical codes and the mappings between them. There is no independent evidence, for example, that the different brain structures that support orthography to phonology conversion and phonology to semantics conversion, respectively, have intrinsically different computational properties (e.g., temporal dynamics). These computations end up having different characteristics because they involve different types of information and because the codes relate to each other in different ways, not because they involve different types of computational or neural mechanisms.

## Distributed Representations

The model uses *distributed* representations, meaning that each code (orthography, phonology, semantics) is represented by a set of units and each unit participates in the representation of many words. This contrasts with *localist* systems in which individual units are used to represent the spelling, sound, and meaning of a word or the word's "lexical entry." Important advances have been made using both types of representation (e.g., localist: Dell, 1986; Joanisse & Seidenberg, 1999; McClelland & Rumelhart, 1981; distributed: Gaskell & Marslen-Wilson, 1997; Plaut & Booth, 2000).

Our use of distributed representations was motivated by several considerations. First, this type of representation is tied to other aspects of the computational framework we used including the use of multilayer networks that incorporate underlying, "hidden" units and the use of a weight-adjusting learning algorithm. Second, it was our desire to maintain continuity with our previous work, in which models that used such representations provided insight about other aspects of word reading. Third, there is evidence that the brain widely uses distributed representations (see, e.g., Andersen, 1999; Ishai, Ungerleider, Martin, & Haxby, 2000; Rolls, Critchley, & Treves, 1996). Although much simplified with respect to the underlying neural mechanisms, the use of these representations represents a step toward incorporating biologically motivated constraints on cognitive models. Fourth, the use of these representations figures in several of the reading phenomena that are the focus of the work (e.g., the effects of masking discussed in the HOMOPHONES section).

Thus, the use of distributed representations is part of the theory of word reading that is proposed here. There are no implemented localist models that address the behavioral phenomena discussed below with which to compare our approach; whether a localist model could exhibit the same behavior is not clear in advance of attempting to implement one. Any such model would treat the behavior as having a very different basis than ours does, however.[5]

Because it uses distributed representations, our model departs from the common metaphor of "accessing" the meaning of a word (see Seidenberg, 1987; Seidenberg & McClelland, 1989, for discussion). The lexical access idea arose in the context of early models in which a word was said to be recognized when its entry in lexical memory was contacted through an activation (e.g., Morton, 1969) or search (Forster, 1976) process, creating what Balota (1990) called the "magic moment" of lexical access. The lexical entry acted as an index for where to find associated types of information, including a word's spelling, sound, and meaning. The representations for different words were distinct from each other and therefore isolable, as in a dictionary.

Our model has a different character. Processing does not involve accessing the lexical representation for a word because there are none in the model to access. All weights on connections between units are used in processing all words. The hidden units that mediate these computations allow the model to encode complex relations between codes, but individual hidden units (or subsets of

[5] Two reviewers suggested that Page's (2000) "localist manifesto" raised questions about the use of distributed representations in models such as ours. Page did not argue against the use of distributed representations ("I will advocate a modeling approach that supplements the use of distributed representations (the existence of which, in some form, nobody could deny) with the additional use of localist representations," p. 446) and went so far as to say "No localist has ever denied the existence of distributed representations, especially, but not exclusively, if these are taken to include featural representations" (p. 447). We have no reason to deny that localist models can be useful, particularly in the early stages of investigating phenomena. In the present context, supplementing the model with additional localist units could not be justified on either practical or theoretical grounds. As detailed later in this article in the description of Phase 1, the implemented model learned 6,103 words, which would have required a large increase in network size and complexity. Finally, the functions usually ascribed to localist lexical representations (e.g., representing word frequencies) can be captured in other ways by networks using distributed representations (e.g., connection weights).

them) are not dedicated to individual words (they cannot be because there are many fewer hidden units than words in the model's vocabulary). The representation of a word is not isolable; thus, it could not be cut out of the network without affecting performance on all other words. Rather than attempting to access the stored lexical entry for a word, the model takes a spelling pattern as input and computes its semantic and phonological codes on demand. There is no magic moment; the model is a dynamical system that settles into a stable pattern of semantic activation over several time steps, based on continuous but time-varying input from orth→sem and orth→phon→sem (as detailed below). Thus, the weights in the model allow a meaning to be computed from an orthographic input pattern; meanings are not "accessed" in the standard sense.[6] Although the knowledge that permits the network to compute the meaning of each word is stored in the network, meanings are not themselves accessed in the standard sense.

## Differing Ease of the Mappings

Given the architectural homogeneity assumption and the use of distributed representations, the nature of the mappings between codes assumes great importance in determining the model's behavior. As many have observed, spelling and sound are more highly correlated than are spelling and meaning in English. Given the first consonant letter of a word, one has a strong clue as to how the pronunciation of the word begins but no hint as to its meaning. This makes the initial learning of orth→phon much easier than orth→sem. However, we have stressed the fact that there are exceptions to this generalization on both sides: regularities within orth→sem that arise primarily in connection with morphology and irregularities within orth→phon due to factors such as diachronic changes in pronunciation not accompanied by changes in spelling. Thus, the mappings between orth→phon and orth→sem differ in degree rather than in kind. The model picks up on the regularities inherent in the training corpus and encodes them in the weights. The differences between the mappings affect how the model learns given exposure to a large sample of words, but the same learning procedure applies to both orth→phon and orth→sem.

## Attractor Basins and Dynamical Systems

The model incorporates *attractor structures,* which have been used in previous models of lexical processing and other phenomena. Plaut and Shallice (1993) and Hinton and Shallice (1991) have made extensive use of semantic attractor networks in a model of deep dyslexia, a form of reading impairment observed after some types of brain injury. Harm and Seidenberg (1999) used phonological attractor networks to account for behavior observed in a phonological form of developmental dyslexia. In the present work, attractor structures were created by including feedback connections via a set of "cleanup units" (i.e., all semantic units connected to all cleanup units, which in turn are all connected back to the semantic units). A network has an *attractor basin* when it develops stable points in activation space and has the tendency to pull nearby points toward the stable attractor points. In this way, partial or degraded patterns of activity are driven toward more stable, familiar representations. Attractor basins are also important because they influence what is learned by the system that maps into them (Harm & Seidenberg, 1999). For example, given a phonological attractor system that is able to repair partial or noisy patterns, the connections from orthography to the attractor can be less precise than if there were no attractor.

The use of attractor basins and recurrence in the reading system adds a time-varying component to processing; the network can change its state in response to its own state, as well as external input. This architecture creates a dynamical system whose state varies in complex ways over time. Properties of the attractor basins have important effects on the reading system's dynamics. For example, Plaut and Shallice (1991) examined how the semantic dimension of abstractness–concreteness was related to the paraphasias of patients with deep dyslexia and the dynamics of a semantic attractor that encodes this type of information. Further constraints on the formation of semantic attractors are discussed in the PHASE 1: THE PHONOLOGY↔SEMANTICS MODEL section. Our modeling builds on this earlier research implicating attractor structures in the explanation of reading and other phenomena.

## Preexisting Knowledge

The model is concerned with the central task confronting a beginning reader: learning to compute meanings from print. In learning to read children make use of preexisting perceptual, learning, and memory capacities that are not reading specific, as well as preexisting knowledge (e.g., knowledge of spoken forms of words and their meanings and knowledge of the world). Our model is not a general account of perceptual, cognitive, or linguistic development; rather, it addresses the question, Given such preexisting capacities and types of knowledge, how is the task of learning to map from print to meaning accomplished? Thus, it focuses on what is novel about reading, the fact that it involves learning about orthography, and in particular how characteristics

[6] Whether PDP reading models can be said to have meanings or lexical entries that are accessed during processing has been debated since the concept of a lexical system without lexical entries was introduced by Seidenberg and McClelland (1989). Clearly, there is a broad sense in which all knowledge of words is "stored" in memory. However, there is a valid and useful distinction between accessing a stored representation and computing this representation on demand. A calculator computes the answer to a problem such as $3 \times 3$ rather than retrieving an answer stored in advance. Similarly, linguists standardly distinguish between forms that are generated by grammatical rules (e.g., the past tense of BAKE) and forms that are stored in lexical memory (e.g., the past tense of TAKE; see Anderson, 1988; Spencer, 1991). According to Pinker (2000), the stored and generated forms involve completely different mechanisms; although this claim is controversial (see Joanisse & Seidenberg, 1999; Patterson et al., 2001), the theoretical distinction is clear. Coltheart et al.'s (2001) model also illustrates the distinction: The pronunciations of nonwords are generated by grapheme–phoneme correspondence rules, whereas the pronunciations of words are stored as nodes in a phonological lexicon. The distinction is clear, but the point is that it does not apply to our model. Like inflectional and grapheme–phoneme correspondence rules, our model instantiates the idea of generating forms based on general rather than word-specific knowledge; however, the model does this using a wholly different type of mechanism built out of units, connections, and weights that also handles the forms previously thought to require a word-specific subsystem. Thus, this method of storing information is neither a wave nor a particle but captures elements of both. In short, the lexical access concept does not extend gracefully to a dynamical system using distributed representations.

of the relationships between the written, spoken, and semantic representations affect learning and skilled performance.

In general, the design of the model involved making minimal assumptions about the nature of the orthographic, phonological, and semantic codes, while incorporating strong assumptions about the relations between them. Consider first the model's phonological representations. Phonological information plays a critical role in learning to read (see Rayner et al., 2001, for an overview); the quality of prereaders' phonological representations is related to their success in learning to read and to some forms of dyslexia. Many issues concerning the role of phonology in reading acquisition and dyslexia were discussed in our previous work (Harm & Seidenberg, 1999). We assume that phonology develops as an underlying representation that mediates between the production and comprehension of spoken language, but we did not attempt to model this (see Plaut & Kello, 1999, however). Rather, we gave the model the capacity to encode phonetic features and then trained it on the mappings between the phonological and semantic patterns for many words. The pretrained phonology–semantics component was then in place when the model was introduced to orthography.

This design feature is important for our account of the behavioral phenomena addressed below. These phenomena concern the relative contributions of the orth→phon→sem and orth→sem pathways over the course of reading acquisition. The former (the phonologically mediated pathway) does involve an extra "step" compared to the direct (orth→sem) pathway as many have observed; however, the child's learning to use the mediated pathway is facilitated by the fact that part of it is already known. Hence it was important to recreate this condition in the modeling.

The phonological representation that we used does not capture all aspects of phonological knowledge, nor have we attempted to simulate the course of phonological acquisition, issues that are clearly beyond the scope of the present project. Leaving aside this pragmatic issue, the use of this representation is justifiable on independent grounds. The feature set we used was drawn from phonetic research, where such representations are often used despite their inherent limitations because they capture generalizations at a level that is appropriate for an important range of phenomena. Our use of this type of representation has a similar basis: It is pitched at a level that is useful and appropriate given the type and grain of the behavioral data that are addressed. The main limitations of this feature scheme arise in connection with facts about multisyllabic words (e.g., assignment of syllabic stress), but the present model is limited to monosyllables. Similarly, although phonological knowledge continues to develop through the early years of schooling (Locke, 1995; Vihman, 1996), much of the system is in place by about age 5. The additional learning that occurs again mainly involves more complex words than are used in the current model. Thus, phonological acquisition is similar to the acquisition of syntax insofar as both systems are largely in place by the start of schooling, although both continue to be refined with additional experience.[7]

In summary, the heuristic value of phonetic feature representations is clear from previous research. We assume with many others that the features are approximations that will eventually be explained in terms of more basic perceptual and articulatory–motor mechanisms that give rise to them (see, e.g., Browman & Goldstein, 1990).

The semantic features that were used had a similar rationale. The goal was not to address issues about the structure of concepts or the contributions of innate and experiential factors to their development. Nor would we claim that knowledge of word meanings is exclusively represented in terms of featural primitives or that such a feature scheme merely has to be scaled up in order to account for a broader range of semantic phenomena. Rather, the representation entailed making minimal assumptions about the beginning readers' knowledge of word meanings in order to examine a more basic issue, the effects of the differing mappings between codes on how the reading system develops. Thus, the model's semantic representations reflect the assumption that meanings are composed out of elements that recur in many words, that different meanings have different representations (e.g., the meanings of homophones such as PEAR–PARE–PAIR were distinct), and that meanings are computed over time rather than accessed at an instantaneous moment. In addition, the reading model was trained in a manner consistent with the observation that children know the meanings of many words from spoken language at the onset of reading instruction. Further detail about the properties of the semantic representations is provided below and in Harm (2002). Like the phonetic features, the semantic features also have heuristic value: They have been shown to provide a good approximation to the kinds of information that are initially activated when words are read, as indexed by measures such as semantic priming (McRae & Boisvert, 1998; McRae, de Sa, & Seidenberg, 1997; Plaut & Booth, 2000). These representations have also been used to understand selective patterns of semantic impairment following brain injury, the progressive loss of semantic information due to degenerative neuropathology, and the neural bases of semantics (Gainotti, 2000; Hinton & Shallice, 1991; Patterson & Hodges, 1992; Patterson, Lambon Ralph, Hodges, & McClelland, 2001). As in the case of phonetic features, we assume that the featural semantic representations are approximate; that semantic phenomena will ultimately be explained in terms of more basic biological and experiential factors; and that such a theory will explain the featuresque aspects of behavior identified in studies such as the aforementioned ones.

Finally, we gave the model the capacity to encode letter strings even though in reality children have only partially mastered this by the start of formal instruction. A proper treatment of the nature of letter recognition and how this skill is acquired goes far beyond the issues addressed here. We assume that this simplification had a similar impact on both the orth→sem and orth→phon→sem components of the system and therefore had little biasing effect on the results.

[7] Whereas the additional phonological development that occurs in children has little impact on learning to read monosyllabic words, the converse is not true: There is good evidence that learning an alphabetic writing system affects the structure of phonological knowledge (Bertelson & de Gelder, 1989), in particular, the development of phonemic-level representations. Spoken words are not sequences of discrete phonemes. Rather, phonemic representations—that is, the notion that the initial sound in PACK and the final sound in TAP are both exemplars of the phonemic category /p/—may be partially due to the fact that these sounds are spelled with the same letter. Knowledge of spelling thoroughly penetrates phonological representations in literate individuals (Seidenberg & Tanenhaus, 1979) and may contribute significantly to performance on "phonological awareness" measures (Harm & Seidenberg, 1999). See Harm and Seidenberg (1999) for discussion of this issue and some preliminary computational evidence concerning the effects of orthography on phonological representation.

In summary, we approximated some aspects of the child's knowledge and experience in order to explore a central issue in considerable detail. Every computational model necessarily involves such simplifications; for further discussion, see Seidenberg (1993). The particular simplifications we made were appropriate because more general properties of the task and network exert much greater influence on the target phenomena. Thus the grain of the simulation matches the grain of the behavioral phenomena to be explained.

## Learning

The model instantiates the idea that learning to read involves learning the mappings between lexical codes and that this is a statistical learning problem, solved using a statistical learning procedure. The correspondences between the codes differ in the degree to which they are correlated, and none of the correlations are perfect. The child has to learn that -AVE is always pronounced /ev/ except in the context of H-, whereas OUGH is pronounced differently in the contexts R-, C-, D-, PL-, THR-, and coda -T. Similarly, BEAK and BEAKS overlap in meaning whereas BEAT and BEAST do not. The relations between codes are probabilistic, and learning is statistical in the sense of being driven by the frequency and similarity of patterns. The weights reflect the aggregate effects of exposure to many patterns rather than learning a set of rules or exemplars. There is good evidence that people (including babies; Saffran, Aslin, & Newport, 1996) and other species engage in this type of learning, and its neurobiological bases are beginning to be understood (O'Reilly & Munakata, 2000).

As with other aspects of the model, we attempted to capture core components of this type of learning and made simplifying assumptions about others. Three aspects of learning need to be considered: the nature of the learning procedure itself, the nature of the input ("experience") from which the model learns, and the relationship between this training procedure and the child's experience. Early models such as Seidenberg and McClelland's (1989) used a supervised learning procedure called *backpropagation,* which is suitable for training strictly feedforward networks. In the present model we used a variant of backpropagation that is suitable for training attractor networks that settle into patterns over time. Details of the learning procedure are provided below. Here the important point is that learning involved presenting a letter pattern to the model; letting it compute semantic output; comparing the computed output to the correct, target pattern; and using the discrepancy to make small adjustments to the weights. Through many such experiences the weights gradually assume values that yield accurate performance.

The primary motivation for using backpropagation is its apparent relevance to the behavior in question. The demands of the reading task appear to exceed the limited computational capacities of networks trained using other principles (e.g., Hebbian or reinforcement learning). The network has to both learn the words in the training set and represent this knowledge in a way that supports generalization. The task therefore requires the computational power provided by multilayer networks trained using algorithms such as backpropagation. The fact that this algorithm is sufficiently powerful to learn the task and the fact that models trained using this procedure simulate detailed aspects of human performance are consistent with the conclusion that the principles by which people learn have similar properties. The brain may achieve this type of performance by using backpropagation or another learning principle or combination of principles that have similar effects, although this issue is unresolved (see O'Reilly & Munakata, 2000, for discussion).

A second computational consideration is that the backpropagation procedure results in cooperative learning across different parts of the system: Thus, the performance of each component is subject not only to its own intrinsic capabilities but also to the successes and failures of other components. In practice, this pressures the system to produce the correct output using whatever means are available. If one component of the system (e.g., orth→phon→sem or orth→sem) fails or is slow for a given item, this generates error. This error can arise from many sources: It may arise because the model has received insufficient training to have learned a mapping; because the mapping is a difficult one, such as spelling to meaning; or because there are ambiguities in the training set that limit performance (e.g., homophony in the mapping from sound to meaning). Given the nature of the learning procedure, the error that one component is slow or unable to reduce creates pressure for the system to make up the difference somewhere else. Hence, each component of the system is sensitive to the successes and failures of other components.

This type of learning contrasts with mechanisms that are correlative rather than driven by error, the classic example being Hebbian learning (Hebb, 1949). In such systems, learning of an item by one component (again, e.g., orth→sem) would be independent of the success or failure of orth→phon→sem for that item. However, it is shown in subsequent sections that the division of labor that results from using the error-correcting learning algorithm plays an important role in accounting for behavioral phenomena. We view the mutual dependence between different components of the system as a central property of the reading system that emerges in the course of learning.

In our model, then, the computation of meaning from orthography is a constraint satisfaction problem: The computed meaning is the output pattern that best satisfies the constraints encoded by the weights on connections in the network. In reading, the weights include those mediated by both the orth→sem and orth→phon→sem components. Learning involves finding a set of weights that yields the best performance possible given the capacity of the network and the structure of the input. See Rumelhart, McClelland, and the PDP Research Group (1986) for discussion of constraint satisfaction processes in PDP models, and see Seidenberg and MacDonald (1999) for an overview of the role of constraint satisfaction in several aspects of language use.

The fact that our model involves a cooperative division of labor using input from all parts of the system can be contrasted with other recent models. In Coltheart et al.'s (2001) dual route cascade (DRC) model, two procedures (one involving rules, the other a localist connectionist network) pass activation to a common set of phonological output units. This captures the idea that the computed output is determined by input from different sources, and it contrasts with earlier pronunciation models in which the routes operate in parallel with a race between them (for discussion, see Henderson, 1982; Paap & Noel, 1991). Aside from the fact that it is concerned with the computation of pronunciation rather than meaning, the Coltheart et al. (2001) model does not incorporate the idea that the contributions of different parts of the system are mutually dependent and emerge in the course of learning. In our model, what one set of weights contributes to the output depends on what other sets of weights contribute, as described above. In contrast, the contributions of the routes in DRC are independently determined by the intrinsic computational characteristics they are assigned. These intrinsic characteristics include the fact that the rules are formulated so that they generate correct pronunciations for only some words (e.g., MINT and LINT but not PINT) and the route-specific parameters that determine their speeds. Coltheart et al.'s (2001) implementation of a system in which two pathways jointly determine output is a major step toward a constraint satisfaction system, but it does not incorporate the idea of mutual dependence between different components arising through a common learning mechanism.

More closely related to our model is the work by Plaut et al. (1996), which like DRC addressed mechanisms involved in generating pronunciations from print. Plaut et al. proposed that pronunciations are determined by input from both orth→phon and orth→sem→phon components of the lexical triangle (see Figure 1). Specifically, they assumed that the division of labor in pronunciation is such that the contribution from orth→sem→phon is greater for words with atypical pronunciations (such as PINT) than for words with more consistent spelling–sound correspondences, which were encoded by the orth→phon pathway. They implemented a model of the orth→phon computation and simulated the contribution of orth→sem→phon by means of an equation specifying that its input increases gradually over time and is stronger for higher frequency words. The model was then used to address issues concerning the pronunciation errors that occur in surface dyslexia, a type of reading impairment following brain injury.

Our model originated with some observations by Seidenberg (1992a) concerning the computation of meaning in different writing systems. Seidenberg (1992a) introduced the idea that semantics could be partially activated by both direct-visual and phonologically mediated processes within the triangle framework: "According to this theory, codes are not accessed, they are computed; semantic activation accrues over time, and there can be partial activation from both orthographic and phonological sources" (p. 105). Seidenberg (1992a) discussed properties of different writing systems that would affect what he termed the "equitable division of labor" that would emerge in such a system. The present model is an extended exploration of the feasibility and psychological plausibility of this idea. Unlike both the Plaut et al. (1996) and Coltheart et al. (2001) models, the division of labor between components developed through learning rather than external specification. Consistent with Plaut et al., orth→sem developed more slowly than orth→phon→sem in our model. However, Plaut et al.'s analysis of the division of labor was limited and left open a broad range of possibilities for how the system would solve the computation of meaning problem. It was not clear in advance, for example, whether the model would divide up the problem by assigning some words to orth→sem and others to orth→phon→sem (as in some preconnectionist accounts, e.g., Baron & Strawson, 1976) or on the basis of other structural characteristics. As discussed below, the division of labor to semantics was greatly affected by factors such as homophony and visual similarity (which were not relevant to earlier models of pronunciation), and the two pathways jointly determined the meanings of most words.

### *Bases for Deriving the Error Signal*

In backpropagation, learning depends on the specification of the correct target or "teacher" in order to generate an error measure. As in previous models, we merely provided the target on every trial rather than attempting to model the sources for it or other aspects of the child's experience. Several points should be noted in considering how this training procedure relates to the child's experience.

For tasks such as reading, for which there is explicit instruction, there often is an actual target provided by a literal teacher. In fact, children typically receive more types of explicit feedback than we used in training the model, including instruction about the pronunciation of letters, digraphs, onsets and rimes, and syllables. In this respect the model's "experience" is more impoverished than the child's, making the learning task more difficult.

In other cases the child can be thought of as using various strategies to derive a teaching signal rather than using an extrinsically provided one. For example, there may be pragmatic or contextual information providing evidence about the correct meanings of words on some occasions, to which children can compare their own computed meanings. The teaching signal may also be internally generated on some learning trials. For example, the child may generate a target by comparing the meaning computed on the basis of orthography to the one computed on the basis of saying the word to oneself (i.e., through the spoken word recognition pathway). This is a version of the self-teaching mechanism described by Jorm and Share (1983). The child will often remember the identity of a word from previous exposure to the text in which it occurs or be able to piece together the correct target by using a conjunction of visual and contextual clues.

Finally, the hippocampus is thought to provide an important internal source for the error signal. Briefly, there is evidence that there are two principal forms of learning in humans and some other species (McClelland, McNaughton, & O'Reilly, 1995). Cortical learning is thought to be gradual, require repeated experiences, and be sensitive to similarities among input patterns. Learning in the hippocampal formation is relatively rapid, requires few exposures (possibly only one), and is item specific. According to this theory, the representations of words encoded in the hippocampus act as teachers for the cortical system. That is, the hippocampal representation of a word may be played back to the cortex multiple times, providing the teaching signal for the gradual learning procedure. Again, rather than modeling this component of human learning and memory, we merely provided the teaching signal.[8]

On other trials the feedback to the child is incomplete or wholly absent. Sometimes the child may know that a computed meaning does not fit in a given context but may not know exactly what the discrepancy is; thus, the child receives positive or negative feedback for a response rather than the correct answer ("reinforcement learning"; Barto, 1985; Sutton, 1988). In cases in which there is no internal or external basis for the teaching signal, the child's own computed response may provide the basis for learning (e.g., in an

[8] Learning in the hippocampus may be the basis for the "fast-mapping" or "single-trial" learning observed in vocabulary acquisition (Carey, 1978) and other domains. See also Landauer and Dumais (1997), who suggested that learning new vocabulary items is rapid because structure in the child's semantic system prepares them for their occurrence.

unsupervised, Hebbian manner). In the near future it should be possible to implement a more realistic learning procedure in which the specificity and accuracy of the feedback varies across trials. Here, it should be noted that it cannot be assumed that providing full, explicit feedback on every trial necessarily yields faster learning or better asymptotic performance compared to the more variable situation characteristic of children's learning. There is some evidence that providing more variable, less precise feedback may lead to more robust performance than merely providing the correct target on every trial (Bishop, 1995). The use of variable types of feedback may discourage the development of overly word-specific representations in favor of representations that capture structure that is shared across words, improving generalization, but this issue needs to be investigated further.

In summary, the claim that learning the mappings between lexical codes is a statistical problem is central to the theory and differentiates it from theories in which learning involves rule induction or encoding exemplars. We used an error-correcting learning algorithm that is sensitive to differences in the correlations between codes and thus captures the relative difficulty of learning the orth→sem versus orth→phon mappings. It also creates cooperation between different components of the network, giving rise to the division of labor described below.

It should be clear from this presentation that the model attempts to capture much of what the child learns about relations between lexical codes without addressing detailed aspects of children's classroom experience. Children typically learn to read through explicit instruction, which rarely resembles a trial of backpropagation learning. Our model attempts to capture a form of statistical learning that is implicit in the sense of recent models of learning and memory (see Cleermans, 1997, for an overview). The ostensive goal of overt instruction is to promote explicit learning, which occurs in many domains and may contribute to the child's knowledge of the lexicon. Our model does not address this form of learning. However, it should also be noted that the relationship between the teacher's explicit instruction and how the child learns from it is complex and not fully understood. When a teacher explicitly draws a child's attention to the similarities among BAT, CAT, and SAT, the child's learning may be mediated by an implicit statistical mechanism like the one we have used. Similarly, whereas a teacher may think he or she is teaching a child a pronunciation rule, the effect of this experience may be to tune the representation of statistical regularities. There are important unresolved questions about how explicit instructional experiences translate into brain-based learning events that need to be addressed in future research. In the present context, we only intend to show that much of what the child knows about the relationships between lexical codes is statistical in nature and closely approximated by our model, including the learning procedure we use.

## Pressure to Compute Rapidly

The model incorporates the assumption that the reader's task is to compute meanings both quickly and accurately. Aside from the obvious practical importance of rapid reading, data from eye movement studies suggest that reading skill is more constrained by the efficiency of cognitive processes involved in comprehending words in texts than by the efficiency of oculomotor processes such as making saccades (see Rayner, 1998, for a review). Thus, we assumed that the model should be driven not only by the need to be asymptotically accurate but also by the need to recognize a word rapidly using whatever resources are available. This tenet results in a system that is "greedy": It demands activation from all available sources to the maximum degree. This assumption was operationalized by penalizing the network not only for producing incorrect responses but also for being slow; error was injected into the network early in processing to encourage the quick ramp up of activity.

The decision to emphasize both speed and accuracy in training the model was principally motivated by observations about reading behavior. However, as with aspects of the training regime discussed in the next section, a design decision that was based on behavioral considerations also contributed importantly to the model's capacity to perform the task and simulate human performance. Bullinaria (1996) implemented a model that, like ours, examined the division of labor between visual (orth→sem) and phonological (orth→phon→sem) components of the Seidenberg and McClelland (1989) triangle model. Bullinaria trained the model on a small vocabulary (300 words) in which semantic codes were represented by random bit patterns. Bullinaria's model learned to compute phonological codes from orthography and semantic codes via the orth→phon→sem pathway. However, almost no learning occurred within the orth→sem pathway. Bullinaria concluded from these results that reading proceeds by orth→phon→sem, with orth→sem contributing little. In pilot simulations we obtained very similar results (Harm, 1998).

Learning did not occur within the orth→sem pathway in Bullinaria's model (or in our pilot simulations) because there was no source of error that would force it to. These models were not trained with pressure to compute rapidly. The phon→sem pathway had been pretrained, leaving only orth→phon and orth→sem to be learned. Because orthography and phonology are correlated and orthography and semantics are not, the models learned to produce correct semantic output via orth→phon→sem. This was adequate because these simulations had virtually no homophones. In effect, orth→sem had nothing further to contribute, and so learning did not occur within this pathway.

The situation changes when the pressure to compute rapidly is introduced. Now the orth→sem pathway has a chance to learn because it is a shorter pathway than orth→phon→sem. As we detail below, this results in an elegant sharing of responsibility between the two pathways. This sharing is particularly relevant to disambiguating the many homophones in the language, which were included in the much larger training set used in the simulations described below.

In summary, the training procedure emphasized both speed and accuracy; this design feature was motivated by observations about the nature of skilled reading but also by preliminary simulations of Bullinaria (1996) and our own indicating that speed pressure promotes learning within the orth→sem pathway.

## Training Regime

Finally, we need to consider the way the model was trained and how this procedure relates to children's experience. Children learn to read in the context of other linguistic and nonlinguistic experiences. The various uses of language are interspersed: The child learns to both produce and comprehend language; learning to read is intermixed with using spoken language; and so on. The way the model was trained reflected this basic fact about the child's experience.

As detailed below, the first phase of the simulations involved training the model on the mapping from phonology to semantics (as in listening) and from semantics to phonology (as in speech production). During this phase, the model was also trained on tasks related to learning about the structure of phonology and semantics. Which task the model was trained on varied quasi-randomly from trial to trial. This procedure (which Hetherington & Seidenberg, 1989, termed *interleaving*) contrasts with *blocked* training procedures in which a single task (or set of patterns) is learned to some criterion, at which point training on that task ends and training begins on a second task (McCloskey & Cohen, 1989). The second phase of the modeling, in which orthography was introduced, followed the same logic, although the training procedure was somewhat different. The weights that resulted from the first phase were frozen, and the model was trained to map from orthography to semantics and phonology. As discussed below, freezing the weights has much the same effect as interleaving reading and spoken language tasks but requires much less computer time, which was a significant consideration given the size of the model.

The main reason for using this procedure was the observation that children's experience with language is not strictly blocked. Although we did not attempt to closely model the child's prereading experience, the Phase 1 training procedure was broadly consistent with the fact that prior to the onset of reading instruction, children have acquired considerable knowledge of phonological and semantic structure and the mappings between them, and that the different uses of language through which this knowledge is acquired are intermixed.

As with the pressure for speed discussed above, although the intermixing of trials was largely motivated by facts about children's experience, this design feature also had a beneficial effect on network performance: Using a blocked procedure can create the effect that McCloskey and Cohen (1989) termed *catastrophic interference.* In brief, McCloskey and Cohen found that training a simple feedforward network on one set of patterns (e.g., a random list of words), followed by training the network on a second set of patterns, resulted in unlearning of the first set. This effect was thought to be unlike human performance and to reflect a limitation on the capacity of this type of network. However, catastrophic interference is related to the strict blocking of trials, which occurs in some verbal learning paradigms but not in learning a language or learning to read. Hetherington and Seidenberg (1989) found that relaxing the strict blocking of training trials (e.g., providing occasional trials to refresh learning on the first set while training the second) eliminated the interference effect. Thus, the child's experience in learning language coincides with conditions that facilitate learning in connectionist networks.[9]

The final issue concerns the way in which words were presented to the model during training. As in previous models (Plaut et al., 1996; Seidenberg & McClelland, 1989), the model was trained on a large vocabulary of words, with the probability that a word would be presented being a function of its frequency as estimated by the Francis and Kučera (1982) norms. This ensured that words such as THE would be presented many times more often than words such as SIEVE. This procedure differs from children's experience; in learning to read, children start with a small number of simple words that occur with high frequency in speech, and the size of their reading vocabularies expands over time. We used the frequency-weighted sampling procedure mainly because it is easier to implement than a procedure in which the size of the training vocabulary grows over time. It is also difficult to obtain reliable independent information about when and how often children are exposed to different words, and there is likely to be considerable variability across children. In recent work we have begun to investigate whether ways of structuring the training regime have an impact on network behavior. First, we have trained some orth→phon models using data from Zeno (1995) concerning the frequencies of words in the texts that are read by children at different grade levels to determine which words are presented at different points in training and how often. We have also trained an orth→phon model using a procedure in which words are introduced in the order in which they occur in children's basal readers (Foorman, Perfetti, Seidenberg, Francis, & Harm, 2001). Finally, we have examined more specific ways of ordering the words in the training regime to determine whether there is a sequence that optimizes speed of learning (Harm, McCandliss, & Seidenberg, 2003). In general these different training regimes yield performance that does not differ greatly from what was obtained using the frequency-biased sampling procedure. Because the words are all represented in an alphabet, what is learned about one item carries over to other items with which it shares structure; this reduces the model's sensitivity to exactly when individual words are presented. Although we had initially thought that adhering more closely to the child's experience in learning words over time would improve the model's performance, we have not observed strong beneficial or interfering effects. Harm et al. found that whereas structuring the training corpus has little impact on normal performance, it did improve the performance of a model that was given an impaired capacity to represent phonological information. Thus, there may be ways of optimizing the training sequence for children with cognitive or perceptual deficits that interfere with normal learning; however, within broad limits (see Plaut et al., 1996; Seidenberg & McClelland, 1989, for discussion), different sampling procedures yield similar performance in nonimpaired models.

In summary, the sampling procedure does not literally correspond to the child's experience. However, because of the shared structure among words in an alphabetic writing system, the model is not highly sensitive to how the training trials are ordered. In reality, the exact sequence of training trials and other reading-relevant experience varies across children and would be expected to affect when specific words are learned by an individual. In addition, these factors may be relevant to designing interventions

[9] Catastrophic interference is also eliminated if the nature of the problem and the way it is represented in a model are such that what is learned from earlier trials carries over to later trials. The quintessential example of this is learning the pronunciations of letter strings: In this case what is learned about the earlier trained words carries over to later trained words because the system is an alphabet and different words share structure. See Zevin and Seidenberg (2002) for relevant simulation results and discussion. Thus, retroactive interference is not normally a problem for human learners both because task-relevant experience is not strictly blocked and because they can represent similarities across patterns (e.g., by using distributed representations).

for children who are not learning to read normally. However, these issues are not central to the present research.

We now describe the procedures used to train the model. We begin with the preliterate speaking–hearing model and continue with the full reading model.

## PHASE 1: THE PHONOLOGY↔SEMANTICS MODEL

We began by implementing a model of the computations between phonology and semantics. This phase was intended to approximate the knowledge of prereaders, who have acquired substantial spoken-word vocabularies and know a considerable amount about the phonological structure of their language and about semantic structure (e.g., that it contains objects, living things, animals, actions, and states). Learning to read builds on this existing knowledge. The phonology to semantics computation is relevant to how people comprehend speech, and the semantics to phonology computation, to production; however, these tasks were not addressed in detail in the present work.

### Network Dynamics

Many previous models have used a simple feedforward architecture consisting of a set of input units, a set of output units, and a set of hidden units mediating between them. On each trial, the $j$ input units $u_j$ are clamped to some desired value. The hidden units compute their values based on the input unit activity and the weights $w$ that map the input units to the hidden units. Each hidden unit $h_i$ for each of the $i$ hidden units computes its output value as $h_i = f(\Sigma_j w_{ij}u_j)$, where $f$ is a nonlinear squashing function. Similarly, each of the $k$ output units $o_k$ computes its output based on the hidden unit outputs: $o_k = f(\Sigma_i w_{ki}h_i)$. Weights are adjusted by propagating error backward through the network and moving each weight in a direction that minimizes the error (the backpropagation of error algorithm; Rumelhart, Hinton, & Williams, 1986).

Such networks adhere to a neural metaphor to the extent that the processing of each unit is driven by the local propagation of activity along weighted connections, rather than, for example, by a central processing executive. However, the metaphor stops there. Such networks are explicitly *stateless*, that is, there are no state transitions in the network, just the final computed state in which activity has propagated through the entire system. There is no time course of activation, no processing dynamics, and no sense in which the current state of the network modifies its subsequent states.

Recurrent networks using backpropagation of error through time (BPTT; Williams & Peng, 1990) address some of these limitations. In such networks, a notion of time is added, such that the output of a unit at time $t$ depends not on the activity of units in a previous layer, as in feedforward networks, but on that of all units at a previous time slice. This kind of network is a generalization of the feedforward network and allows for recurrent, or cyclic, connectivity patterns. The activity of a unit $u_i$ at time $t$, $u_i^t$ is defined as $u_i^t = f(\Sigma_j w_{ij}u_j^{t-1})$. A unit's activity at time $t$, then, is totally determined by the activity of all units connected to it at time $t - 1$. These networks form dynamical systems, exhibiting either stable fixed points or oscillating behaviors. Further, activity within a group of units can build up over time, with the units influencing each other's states.

However, the temporal dynamics of such networks are still quite simple. They operate in a lockstep fashion, where the output of the unit is the squashed sum of its input regardless of anything else. The output of units, then, tends to "jump"; activity does not ramp up or down gradually but instead can respond instantaneously. Hence, although the network will exhibit global temporal dynamics, each individual unit still has a very simple time course of activation.

Pearlmutter (1989, 1995) formalized a way to train networks with much more subtle time courses of activity. Continuous time networks such as those introduced by Pearlmutter (1989, 1995) add unit dynamics: A unit's output ramps up gradually as a function of its input, based on a leaky integrator equation:

$$\sigma \frac{\partial o_i}{\partial t} = (y_i - o_i) + b_i, \tag{1}$$

$$y_i = f(\sum_j w_{ij}o_j), \tag{2}$$

where $y_i$ is the squashed input to the unit (or what its output would be in a discrete time network), $o_i$ is the instantaneous output of the unit, and $b_i$ is a resting state of the unit. The parameter $\sigma$ controls the speed at which a unit ramps up or down. Essentially, the rate of increase of a unit's activity is proportional to the difference between its current activity $o_i$ and what its activity ought to be ($y_i$). In simulations, the continuous dynamics defined by Equation 1 are approximated by discrete samples. In this case, the output of the unit at time $t$ changes by the difference between the output at time $t - 1$ and its asymptotic output multiplied by $\sigma$. To take a concrete example, suppose the input to a given unit was strong enough to asymptotically drive the output to 1.0, the unit's output is initially zero, and one used $\sigma = 0.1$. On the first sample, the unit's output would move from 0.0 to 0.1 (increasing by $\sigma$ times the difference between actual and asymptotic output). On the next sample, it would move from 0.1 to 0.19 (again, increasing by $\sigma$ times the difference between actual output 0.1 and asymptotic output 1.0). On the third sample, it would increase to 0.271—that is, $0.19 + 0.1 \times (1.0 - 0.19)$. And so on.

Pearlmutter (1989, 1995) generalized the backpropagation of error equations to allow error gradients to be integrated up over time, the way that activity is integrated up over time. This allows one to train such networks with the full power of the backpropagation of error algorithm.

Plaut et al. (1996) introduced a subtle but important change to the Pearlmutter equations. The Pearlmutter (1989) formulation had the output of a unit ramping up over time in response to the instantaneous squashed input to that unit. Plaut et al. made the output of a unit the instantaneous squashed value of the input to a unit and caused the input to units to ramp up over time. Formally,

$$o_i = f(y_i) + b_i, \tag{3}$$

$$\sigma \frac{\partial y_i}{\partial t} = (x_i - y_i), \tag{4}$$

$$x_i = f(\sum_j w_{ij}o_j). \tag{5}$$

Although they are mathematically similar, there are important theoretical differences between these two processing dynamics. In the Pearlmutter (1989) formulation (which has been termed *time-averaged outputs;* TAO), the maximum output of a unit (typically 1) determines the maximum rate of climb of the unit. As such, if one unit receives an input of 10, its asymptotic output is 0.99999, and so it climbs to that value; if a second unit receives an input of 100, its asymptotic output is 0.999999, and it climbs to that value at almost exactly the same rate as the first unit. The error gradient equations reflect this: If a unit is ramping up as rapidly as it can, additional input does not help, and the error gradient for the additional input is zero. In contrast, with the time-averaged input (TAI) networks, if one unit gets an input of 10 and another gets 100, the second unit ramps up much more rapidly than the first. Equation 1 cannot evaluate to more than 1.0 (assuming $b_i$ is zero, as is typical), whereas Equation 4 is unbounded, because the summed input to a unit, $x_i$, is unbounded.

Pilot simulations using TAO networks failed because they implemented the wrong theory. A crucial design principle of this project is that summed activation causes more rapid rise times of units. It was found early on that if orth→phon→sem was driving semantic units as rapidly as they could be driven (i.e., with an output of 1.0), then there was no advantage to additional input from orth→sem; such input would not drive the semantic units any faster. It is a theoretical assumption of this work that greater activation produces faster responses and that the network is under pressure to rapidly compute the correct output. For these reasons the TAI networks are used throughout this work. Figure 2 shows the temporal processing dynamics for a unit in this network when activated with varying input strengths. The stronger the input, the faster it moves away from its resting value of 0.5 toward its asymptotic value, which is the squashed value of the input, $f(x)$.

Although the continuous time networks are considerably more sophisticated and interesting than feedforward networks, they are still quite simplified compared to what is known about actual neurons and the techniques of modeling their activity. However, the research is following a normal progression in which the range of phenomena to be modeled is expanding, and with it, the fidelity to actual biological systems is also increasing. Plaut and Shallice (1993) and Harm and Seidenberg (1999) used attractor dynamics in BPTT networks to explain patterns of impairment in deep and phonological dyslexia. Such studies revolve around the idea of attractors in state space and hence would not have been possible with simple feedforward networks. In a similar vein, the current study demands continuous time networks to fully implement the principles outlined above. Further advances in understanding the behavioral phenomena, the neurobiology of learning and processing, and the properties of these computational models will both enable and demand greater biological realism.

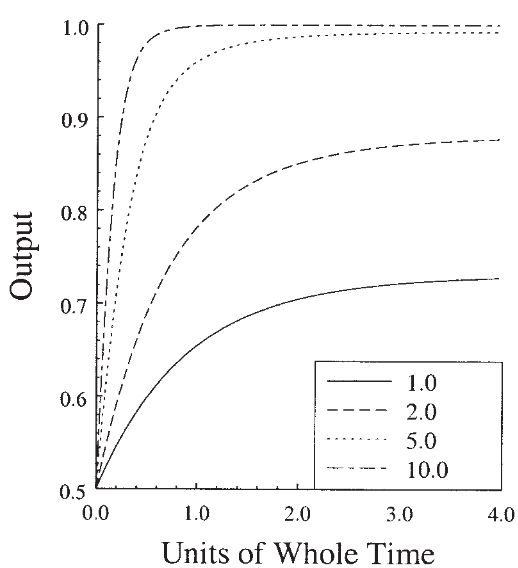


*Figure 2.* Temporal dynamics of a unit receiving input values of 1.0, 2.0, 5.0, and 10.0. Larger input to a unit produces larger asymptotic output but also more rapid rise times.

## Training Corpus and Representations

The training corpus included 6,103 monosyllabic words, consisting of all monosyllabic words and their most common inflection, for which semantic and phonological representations could be derived. There were 497 sets of homophones containing 1,047 words: 447 sets having two members, 47 sets having three members (e.g., THREW, THROUGH, THRU), and 3 sets having four members (e.g., AIR, ERE, ERR, HEIR). There were 39 words in which a single spelling was associated with two or more meanings (mainly words such as SHEEP, FISH, or HIT, whose plural or past tense morphological inflection involves no change from the stem).[10]

The frequency of each item was coded using a square-root compression of the *Wall Street Journal* (WSJ) corpus (Marcus, Santorini, & Marcinkiewicz, 1993) according to the formula

$$p_i = \frac{\sqrt{f_i}}{\sqrt{m}}, \tag{6}$$

where $f_i$ is the WSJ frequency of the $i$th item and $m$ is 30,000 (a reasonable cutoff frequency). Values over 1.0 were set to 1.0; those less than 0.05 were set to 0.05.

Semantic representations were derived in a quasi-algorithmic manner. A full description of the method of deriving semantic features and their properties was given in Harm (2002). The properties that are relevant to the present simulations are summarized here. Words were categorized for their part of speech based on the most frequent occurrence given in the Francis and Kučera (1982) corpus. For uninflected nouns and verbs, the WordNet (Miller, 1990) online semantic database was used to generate semantic features. WordNet is a hierarchically organized semantic database in which groups of words are linked with relations such as IS-A and HAS-PART. For each word, the set of features for that word was generated by climbing the IS-A tree and

[10] With the exception of homographs such as WIND, each word in the corpus was assigned one pronunciation. We did not attempt to capture the dialectical variation in how words are pronounced in English. Such variation may have a large impact on a word's pronunciation difficulty, however. For example, POOR rhymes with TOUR in some dialects and TORE in others. Thus, different neighborhoods are relevant to *poor* depending on how it is pronounced. This factor will affect the fit of the model to behavioral data, particularly if there is a mismatch between the model's dialect (roughly, Southern Californian) and the dialect of participants tested in other regions or countries.

following HAS-PART pointers. Hence, the representation for a word like DOG consisted of features such as [canine], [mammal], [has_part_tail], [has_part_snout], [living_thing], and so on. Inflected items such as plurals, past tenses, and third-person singulars were generated by taking the features for the base word and adding inflectional features such as [plural]. A total of 1,989 semantic features were generated to encode the 6,103 words. The representations were rather sparse, with the number of features used to encode a word ranging from 1 to 37 ($M = 7.6$, $SD = 4.3$, $Mdn = 7$, out of 1,989 features).

Eight phoneme slots were used to encode the CCCVCCCC words (where C is a consonant and V is a vowel), with vowel centering to minimize the "dispersion" problem (see Plaut et al., 1996). A set of 25 phonological features was used to describe each phoneme; these were derived from feature matrices in Chomsky and Halle (1968), with minor modifications. All features were binary, taking values of 0 or 1. The 25 features per phoneme over eight phoneme slots yielded a total of 200 features. The feature representations for phonology were considerably more dense than for semantics: Over the whole training set, the average semantic feature was on 0.38% of the time, whereas the average phonological feature was on 5.7% of the time. We did not set out to create representations with this asymmetry in sparseness, but this seems to accurately represent an important difference between the two domains. The structure of phonological space is highly constrained by articulatory and acoustic factors; thus, the number of possible segments is small and they can be described in terms of a small number of primitives, creating a large degree of overlap between segments. Semantic space is larger and more variable; this creates less overlap, on average, between the meanings of words compared to their sounds. It turns out that the difference in sparseness of semantics and phonology is relevant to explaining masking effects that are discussed below.

## Architecture

Figure 3 depicts the model used in the first phase. The semantic component consisted of the 1,989 semantic features described above. These units were all connected to 50 units in the semantic cleanup apparatus, which projected back onto the semantic features. This architecture, when trained properly, is capable of forming attractors in semantic space that repair noisy, partial, or degraded patterns and tend to pull the state of the semantic units into consistent patterns (Plaut & Shallice, 1993).

The phonological representation consisted of the 200 phonological units (eight slots of 25 units each), which projected onto a set of 50 phonological cleanup units. These cleanup units project back onto the phonological units. Here again an attractor network can be created that will repair partial or degraded phonological patterns. Harm and Seidenberg (1999) examined the role of this attractor, and damage to it, in learning orthographic–phonological correspondences.

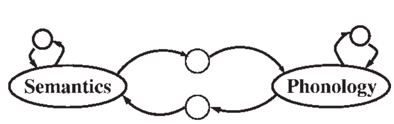


*Figure 3.* The phonology–semantics model. During this preliterate phase, the model developed structure within the semantic and phonological components and learned the mappings between them.

The semantic component mapped onto the phonological component via a set of 500 hidden units. There was feedback in both directions. The number 500 was chosen from pilot studies; it is a number large enough to perform the mapping without being too computationally burdensome.

## Training

Phase 1 involved training the model on the structure of phonology and semantics and on the mappings between them. The model was trained on four tasks: a phonological task (10% of the trials), a semantic task (10%), a phonology to semantics task ("comprehension"; 40%), and a semantics to phonology task ("production"; 40%). Training on the four tasks was intermixed. Once a word was selected for training, it was assigned to one of the four tasks. Online learning was used, with words selected for training according to their probability of presentation (see Equation 6). To model the continuous time dynamics defined by Equation 4, we used a discrete time approximation in which actual time defined by the integral was broken down into smaller units. In training the network, the network was run for 4.00 units of whole time, modeled by using 12 samples and an integration constant of 0.33.

### *Phonological Task*

The phonological task develops the phonological attractor and is intended to approximate the child's acquisition of knowledge about the structure of spoken words (see Jusczyk, 1997). The phonological task was similar to that used in Harm and Seidenberg (1999), except that it was modified slightly to accommodate continuous time networks. The phonological form of the target word was clamped on the phonological units for 2.66 units of time. Then a target signal was provided for the next 1.33 units of time, in which the network was required to retain the phonological pattern in the absence of external clamping. In Harm and Seidenberg (1999), auto-connections were used to give the units a tendency to retain their value but gradually decay. To accomplish the task, the network had to learn enough of the statistical regularities of the representations to prevent this decay. In the current simulations, the idea is the same, but because continuous time units were used, auto-connections were not necessary to provide the units with a tendency to gradually decay; this was part of the units' normal processing dynamics.

On the phonological task, only the weights from the phonological units to the phonological cleanups and back were modified. Figure 4a shows the connections in the model that were trained in this task. Harm and Seidenberg (1999) found that training on this task allowed the network to form attractors, which allowed it to reliably repair corrupted phonological patterns and gave rise to other interpretable behavior (e.g., categorical perception of consonants, phoneme restoration effects). Thus, the task causes the model to absorb basic information about the sound structure of English.

### *Semantic Task*

These trials were devoted to training the semantic attractor. This task was constructed to be analogous to the phonological task: The

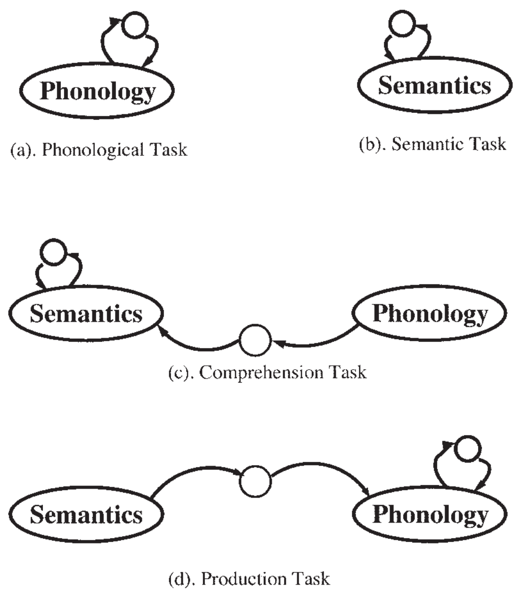


*Figure 4.* The tasks used in training the phonology–semantics model.

pattern of semantic units corresponding to the selected word was clamped onto the units for 2.66 units of time, and the network was allowed to cycle. Then the semantic units were unclamped, and the network's task was to maintain their activity in the face of the tendency of the units' activity to decay for 1.33 units of time. To accomplish the task, the network had to learn about the distributions of semantic features across words—specifically, the complex correlational structure that the representations exhibit. Encoding these systematic aspects of semantic structure allowed the attractor to maintain patterns in the face of decay. This task is more difficult than the phonological task because there are many more semantic units than phonological units and the correlations between units are generally lower. The connections used in training this task are shown in Figure 4b.

### *Production Task*

This task involved training the semantics to phonology pathway (sem→phon). It was loosely based on the task of producing an utterance, for example, naming an object or generating free speech. The task involved the production of the appropriate phonological form for a word given its semantic representation.

On a training trial, the semantic pattern of a word was clamped on the semantic units for the full 4 units of time and the task was to produce the correct phonology. The output of the phonological units for the final 1.0 units of time was compared with the target values; error was injected according to the standard back-propagation of error equations. The connections used in training this task are shown in Figure 4d. All weights were updated, except those leading back into semantics (because the values of the semantic units were clamped, no weight changes would have resulted). Note that the weights in the phonological attractor were trained as well as those involved in the computation from semantics to phonology.

### *Comprehension Task*

The final task, comprehension, was the complement of the production task. The connections used in training this task are shown in Figure 4c. The phonological form of a word was clamped on the phonological units for the full 4 units of time. During the final 1.0 units of time, the output of the semantic units was compared with their targets. The task was to produce the semantic pattern accurately.

In summary, the model was trained for 700,000 word presentations (approximately 280,000 production, 280,000 comprehension, 70,000 semantic, and 70,000 phonological trials). A learning rate of 0.2 was used for 500,000 word presentations, then lowered to 0.1 for the remaining 200,000 word presentations. Beginning with a high learning rate and then lowering it during training often results in faster convergence than either maintaining a high learning rate (which can lead to network oscillations) or starting with a lower one (which can dramatically slow initial learning).

### *Scoring Method*

The computed semantic output was considered correct if each semantic feature whose target was 1.0 was activated to at least 0.5 and each feature whose target was 0.0 was activated to less than 0.5; thus, the output for each feature had to be closer to the target than to its opposite. The computed phonological output was assessed as follows. For each slot in the phonological template, the euclidean distance between the representation in that slot and each of the veridical set of phonemes was calculated. If the output in each slot was closest to its corresponding target, the output was considered correct; otherwise, it was considered an error.

## Results of Training

Figure 5 summarizes the model's accuracy on the production (generating phonology from semantics) and comprehension (generating semantics from phonology) tasks over the course of training. At the end of training, the model correctly generated phonological codes for 90% of the words and correctly computed the semantics for 86% of the words that were not homophones. Although model performance could be improved with additional training, our goal was not to achieve perfect performance in this phase, on the view that the 5-year-old beginning reader does not have perfect knowledge of all 6,000 words in the corpus. The nonhomophones on which errors were made were generally limited to one or two incorrect semantic features (e.g., it recognized the item PRIM as having features such as [abstraction], [attribute], and [clean], but not [R4], which is the randomly generated feature that distinguishes PRIM from NEAT). The model was therefore scored as incorrectly computing the full semantics of PRIM by producing a representation that is identical to NEAT.

For the 1,125 homophones, the model produced the correct semantic pattern 26% of the time. For the other homophones, the model generally produced a mix of features from the alternative meanings. For example, ALE was interpreted as [beverage] at an activity level of 0.70 and as a state of being (as in AIL) with the [be] feature at an activity level of 0.61. This behavior is typical; the network's semantic units are not driven to extreme values for either interpretation. This reflects the inherent ambiguity of the

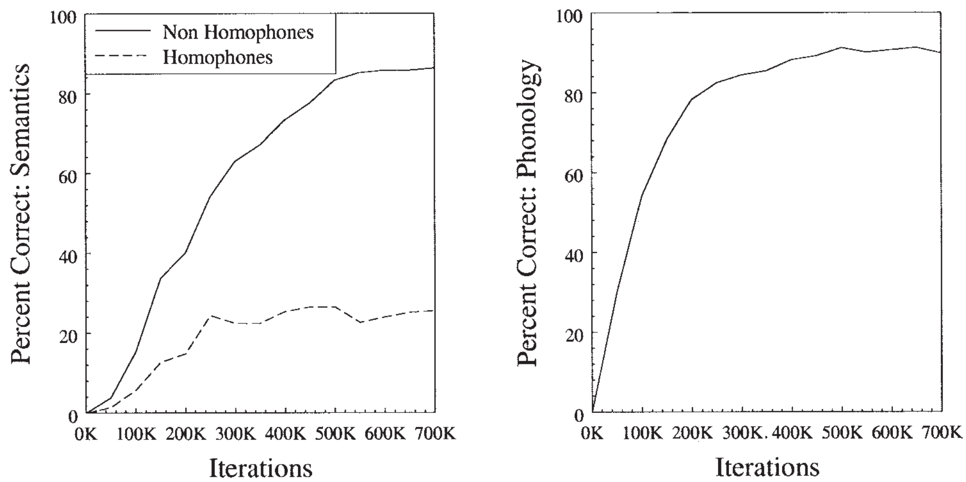


*Figure 5.* Development curves for the comprehension (left) and production (right) tasks. In this and all other figures "Iterations" refers to the number of randomly selected training trials, measured in thousands (K). In the comprehension task, the mapping from phonology to semantics is inherently ambiguous for homophones and therefore the model performs more poorly.

phonological form; the network is "on the fence" as to which interpretation is correct. Such words are normally disambiguated by contextual information.

### *Simulation 1: Homophones in the Phonology↔Semantics Model*

The model makes errors in producing the semantics for many homophones because their phonological forms are associated with multiple meanings. We conducted additional analyses to examine how such words were processed.

## *Method*

*Stimuli.* The 1,125 homophones in the training set included pairs such as BEAR–BARE and triplets such as PAIR–PARE–PEAR. Each pair of homophones was categorized as follows. If one word had a probability of presentation more than 1.5 times that of the other, the higher frequency item was considered dominant and the lower frequency one was considered subordinate. If the probabilities did not differ by this much, they were treated as balanced. This procedure yielded 404 dominant, 404 subordinate, and 317 balanced items.

*Procedure.* Three presentation conditions were used. In the no-context condition, the phonological form of the item was clamped onto the phonological features, and the trained network processed the item as usual. In the helpful-context condition, the phonological form was again clamped, and the most frequent semantic feature that distinguishes the word from its homophone was also clamped. For example, for the homophonous pair BEAR–BARE, the [entity] feature would be activated when BEAR was presented, and the [physical_property] feature would be activated for BARE. In the distracting-context condition, the procedure was the reverse; the semantic feature for the opposing member of the homophone pair was activated. The computed semantic representation was compared to the target representation in terms of hits, misses, false alarms, and correct rejections, and $d'$ was computed. In conditions in which a semantic feature was clamped, that feature was excluded from the $d'$ calculation.

## *Results*

Figure 6 summarizes the results. In the no-context condition there was a dropoff in $d'$ as a function of type of homophone. This result indicates that the model tended to default to the semantics of the dominant (higher frequency) sense. The helpful context yielded improved performance in all conditions, with the biggest gain in the subordinate condition. Thus, even a small amount of relevant semantic information was sufficient to push the semantic attractor to the less frequent member of a homophone pair. Finally, when the semantic context was unhelpful, performance declined relative to the no-context condition, most prominently for the dominant homophones. This is because the dominant homophones enjoy a frequency advantage over their subordinate item in the no-context condition; the unhelpful context pulls the representations from

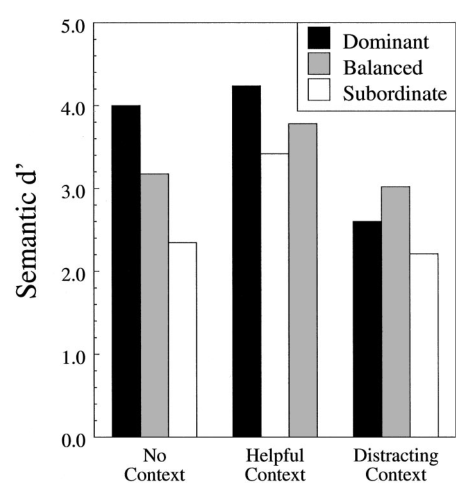


*Figure 6.* Semantic codes activated by homophones, measured in $d'$ units. In the absence of context, the model tends to produce the dominant (more frequent) meaning. Relevant ("helpful") contextual information causes the model to produce the correct meaning, regardless of dominance. Distracting contextual information (i.e., a bit of information related to an alternative meaning of the homophone) was most harmful to the dominant meanings, pulling their activations to the levels of subordinate and balanced meanings.

deep in the dominant interpretation and toward the subordinate one.

Thus, in the absence of biasing contextual information, the model is biased toward producing the semantics of the higher frequency member of a homophone pair. The effects due to the addition of a small amount of biasing information indicate that the model had formed attractors for the alternative meanings of homophones; this information pushes the model toward one of the attractors. Such information is typically provided by the syntactic, pragmatic, or discourse contexts in which words occur.

Although a detailed exploration of the use of context in lexical ambiguity resolution is beyond the scope of this work, this behavior of the model is promising. The model shows sensitivity to frequency differences between alternative meanings of homophones (examined further below) and also suggests a mechanism by which contextual information can affect the computation of meaning. It remains for future research to examine the behavior of the model with respect to the extensive literature on lexical ambiguity (e.g., Simpson, 1994), particularly the interaction between meaning dominance and contextual constraint (e.g., MacDonald, 1993; Rayner & Duffy, 1986).

## *Simulation 2: Morphological Regularities*

The semantic representation included features that are associated with number and tense inflections in English. Thus the model was trained that a plural form such as GOATS was associated with the semantic features for GOAT plus the plural feature; similarly, a past tense form such as BAKED was associated with the semantic features of BAKE plus the past feature. There were also words such as BAKES whose most common usage in the Francis and Kučera (1982) corpus is as a verb with a third-person-singular inflection. There are strong but imperfect correlations between these features and phonology, reflecting the quasi-regularity of the mappings. The plural feature is usually associated with the plural inflection that is spelled *s* and has three phonological allomorphs (as in LAKES, HANDS, BUSSES); however, there are irregular plurals such as MEN and MICE. Conversely, there are words that have the phonological forms of plurals but are not plural; these include pluralia tanta such as PANTS and TIGHTS and others such as LENS and PONS. The past tense behaves similarly: The past tense feature was usually associated with one of the allomorphs of the inflection spelled ED; however, there are many irregular past tenses such as GAVE and forms that sound like past tenses but are not (e.g., SCOLD, MELD).

The model learned to produce correct semantic output for the words on which it was trained; the additional question we addressed was whether this knowledge was represented in a way that supported generalization to novel, untrained forms. Given a nonword such as GOMES, would the model produce either the plural or third-person-singular semantic feature; given a nonword such as BLAKED, would it activate the past tense feature?

### *Method*

*Stimuli.* The stimuli were based on 86 nonwords from Glushko (1979). One list consisted of plural forms of these nonwords (e.g., GOME→GOMES). Five items for which the resulting plural was bisyllabic (e.g., COSE→COSES) were excluded because the phonological representation is limited to monosyllables. Past tenses were also generated from these items, resulting in 49 monosyllabic stimuli. The third list consisted of the uninflected nonwords themselves.

*Procedure.* The phonological forms of the nonword were presented to the trained model, which processed them using the normal parameters for integration constant and number of samples. The activities on the [plural], [third_person_singular], and [past_tense] features were recorded. For stimuli such as GOMES, both the plural and third-person singular are valid interpretations. As before, a semantic feature was considered active if its activity level was greater than or equal to 0.5, that is, if it was closer to the active state of 1.0 than the inactive state of 0.0.

### *Results*

Table 1 summarizes the results. For 90% of the items such as GOMES the model activated either the plural or the third-person-singular feature or both. The past tense feature was activated for 88% of the items such as GOMED. Uninflected items such as GOME activated the plural feature on 1.6% of the items and the past feature for no items. One of the uninflected items happened to be the pseudohomophone (DERE), which activated the [plural] feature because it phonologically overlaps with the word DEER. In general the model picked up on the regularities concerning the mapping between these features and their phonological realizations. The model's level of performance is plausible given that the correlations between phonology and these features are not perfect; the model treats most nonwords such as GOMES as inflected but does not treat all of them as inflected because some words with this ending are not inflected.

The model also generated some activation of semantic features in addition to the morphological features shown in Table 1. However, these features tend to be rather weakly activated, relative to the semantic activation that words produce. Plaut (1997) used a measure called *stress* to quantify the extent to which features were driven to extremal values. Plaut's method was symmetrical: A unit that was strongly driven to zero provided the same stress as one driven equally close to 1. However, this network has such strong negative biases on semantic features (owing to their sparseness) that including such negative stress results tends to wash out any variation in positive stress. Therefore, for this demonstration we examined only positive stress—the extent to which units were driven on. Formally, for units whose output was 0.5 or greater, stress was computed using the formula used in Plaut (1997):

$$s_j = o_j \log_2 o_j + (1 - o_j) \log_2 (1 - o_j) - \log_2 0.5. \quad (7)$$

Table 1
*Meanings Activated by Inflected Words and Their Stems (in Percentages)*

| | Feature | | | |
|---|---|---|---|---|
| Inflection | Plural | Third person | Plural and third person | Past tense |
| Plural | 70 | 15 | 5 | 2 |
| Past tense | 0 | 0 | 0 | 88 |
| Stem | 1.6 | 0 | 0 | 0 |

*Note.* Values do not add up to 100% because the model sometimes did not produce any of the inflectional features.

We computed the mean stress for the inflected nonwords and words, as well as the stress values for the three morphological features for the nonwords. Figure 7 shows the distribution of stress values for the words and nonwords and for either the plural, the third-person, or the past tense feature, whichever was greater.

The stress values for the words tend to be concentrated at the higher end of the scale, whereas the nonwords are much weaker. The mean stress for all semantic features for nonwords was 0.58, but the stress of morphological features for these items was reliably higher, at 0.84, $F(1, 110) = 94$, $p < .001$. In addition, the stress for words ($M = 0.87$) was reliably higher than for the nonwords, $F(1, 179) = 97$, $p < .001$. Overall, the model strongly activated morphological features for inflected nonwords, and semantic features for words, but the activation of other semantic features for nonwords was far lower.

In summary, the Phase 1 results show that the model learned to accurately map between phonology and semantics for a large number of words, subject to limitations imposed by the ambiguities inherent in homophones and nonwords such as GOMES. The model encoded some basic aspects of lexical knowledge that children possess before the onset of reading instruction. We now turn to the second phase, in which the task of learning to map orthographic patterns onto phonology and semantics was introduced.

## PHASE 2: THE READING MODEL

### Architecture

Figure 8 shows the architecture of the reading model. The top section is the Phase 1 model described above. A slot-based localist representation was used to represent the spelling of a word as in several previous models. The orthographic features were defined by creating 10 slots of 26 features corresponding to the letters of the alphabet. The slots were arranged in a vowel-centered template. The features were then pruned by removing features in slots that never occurred in the training set (e.g., only the letters C, P, S, and T occurred three positions before the vowel). This resulted in

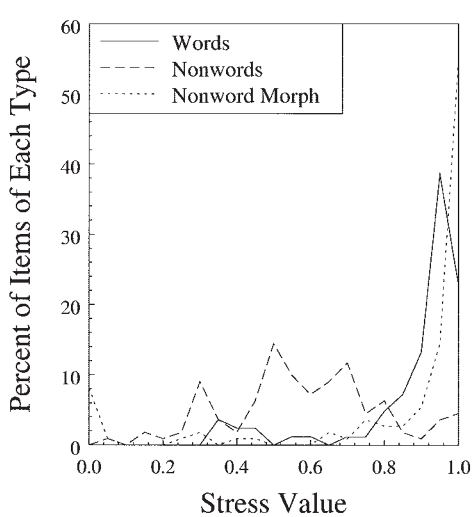


*Figure 7.* Semantic stress values for words, nonwords, and nonword morphological features (Morph) from Simulation 2.

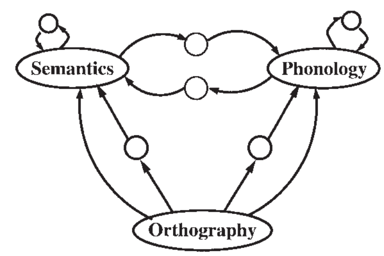


*Figure 8.* The implemented reading model. The semantics–phonology component was taken from the model trained in Phase 1.

111 orthographic units. One set of 500 hidden units mediated the mapping from these orthographic units to semantics, forming the orth→sem pathway. Similarly, a second set of 100 hidden units mediated the orth→phon pathway. The number of hidden units in the orth→phon pathway was the same as in previous models. More units were used in the orth→sem pathway because the mapping is more difficult. Varying the number of hidden units affects performance in ways that are interpretable in terms of individual differences among readers (Seidenberg & McClelland, 1989), but we did not examine this factor in the present work. The architecture of the phonology↔semantics component was identical to that used in the Phase 1 model. The integration constant and number of samples for the reading model were also the same as in the Phase 1 model.

The model also included a set of connections mapping orthographic units directly onto phonological units and another set mapping orthographic units onto semantic units. The former were added because they tend to improve generalization. The latter were added chiefly for symmetry. The inclusion of the direct connections from orthography to phonology was suggested by the work of Zorzi, Houghton, and Butterworth (1998), who explored a spelling to sound model that contained both these connections and the more usual orthography→hidden→phonology connections. They characterized their model as a dual-route model, with the direct connections corresponding to a sublexical route encoding regular, rule-governed mappings, and the hidden-unit pathway corresponding to a lexical route necessary for exceptions. When only direct connections were implemented, their model performed quite well reading nonwords (100% correct) and poorly on exceptions (14% correct). They then examined a model containing both direct connections and connections mediated by hidden units. When the hidden-unit-mediated pathway was selectively impaired, performance on regular words was spared (at or near 100% correct) but exceptions were impaired (approximately 45% correct; see Figure 12 in Zorzi et al., 1998). Putting these two pieces of information together, the model seemed to be a connectionist implementation of the dual-route model with separate mechanisms for regular/rule-governed words and exceptions.

In exploratory simulations we found that including direct connections between orthography and phonology improved performance, facilitating the learning of regularities that support nonword reading. We therefore included them in the model described below. However, we disagree with the further claim that the hidden-unit and direct-connections pathways become highly specialized for exceptions versus regulars, respectively. Zorzi et al.'s (1998) own model does not exhibit a high degree of specialization, and neither have the models we implemented. The direct-connections pathway in their model read regulars much better than exceptions; however, the hidden-unit-mediated pathway did not read exceptions well at all. Zorzi et al.'s Table 7 presented the model's performance for 10 representative stimuli; the model with the direct connections severed did not produce the correct pronunciations for any words, regular or exception. Reading exceptions correctly apparently required input from both pathways. This is probably because exception words share structure with many regular words (e.g., HAVE overlaps with HAT, HAS, HIM, HIVE, etc.); the direct connections tend to encode strong regularities such as the pronunciation of word initial *h*, which occurs in both regular and irregular forms. Thus the hidden-unit pathway in the Zorzi et al. model was not comparable to the lexical route in traditional dual-route models of naming because it does not produce the correct pronunciations for exceptions by itself.

Our model does not divide things up as Zorzi et al. (1998) described, either. We tested the model on a set of exceptions from Patterson and Hodges (1992) and nonwords from Glushko (1979). The intact model produced the correct pronunciations for 88.4% of the nonwords and 99.2% of the exceptions. Removing the hidden units mediating orthography and phonology yielded 74.4% accuracy on the nonwords and 40.3% on the exceptions. Thus, performance on nonwords was more impaired than in the Zorzi et al. simulation, whereas performance on exceptions was less impaired. The higher rate of accuracy on exception words in our model derives from the fact that there is a semantic pathway to phonology in contrast to the Zorzi et al. model. The semantic path takes responsibility for many of the exception words, which is unaffected by removing the hidden units between orthography and phonology. The lower rate of accuracy on nonwords indicates that the hidden-unit-mediated pathway encoded some regular though complex mappings from spelling to sound. This was facilitated by the use of a distributed phonological representation rather than the localist one used by Zorzi et al. In summary, the direct connections facilitate performance and there is no a priori reason to exclude them; however, the resulting model does not organize itself into the lexical and sublexical routes in traditional dual-route models.

## Training Regime

The weights that were obtained at the end of the Phase 1 model were frozen and embedded in the larger reading model. Thus, only the connections from orthography to other units were trained in Phase 2. Freezing the weights is not strictly necessary; earlier work (Harm & Seidenberg, 1997) used a process of *intermixing* in which comprehension trials were used along with reading trials. Weight freezing has the same effect but is simpler and less computationally burdensome to implement. Intermixing is effective and realistic but adds substantially to network training time.

Items were presented to the network according to the same online learning scheme as before with the same frequency distributions. Error signals were provided for both the phonological and semantic representations of a word.

To computationally instantiate the principle that the reading system is under pressure to perform rapidly as well as accurately, we injected error into the semantic and phonological representations early, from time samples 2 to 12. The network therefore received an error signal not only if it produced incorrect semantic or phonological codes but also if it did not produce them rapidly.

## Overall Results of Training

The network was trained for 1.5 million word presentations. At the conclusion of training, the network produced the correct semantic representations for 97.3% of the items. For the other 2.7% of the words, it activated an average of 1.6 spurious features and failed to activate an average of 0.8 features. The model produced correct phonological representations for 99.2% of the words. On the remaining 0.8% of the words, it produced an average of 1.1 incorrect phonemes. Figure 9 depicts semantic and phonological accuracy over the course of training.

The focus of this research is on behavioral phenomena concerning the activation of meaning. However, in order to establish continuity with previous research on the activation of phonology, we examined the model's performance on some benchmark phenomena: the interaction of frequency and spelling–sound consistency, nonword generalization, and morphological processing.

### *Simulation 3: Frequency by Regularity Interaction*

One well-established phenomenon in reading is the frequency by regularity interaction (Seidenberg, Waters, Barnes, & Tanenhaus, 1984; Taraban & McClelland, 1987). These studies examined exception words such as PINT and regular words such as MUST. The word PINT is an exception because -INT should be pronounced as in MINT and LINT. The word MUST is regular insofar as all monosyllabic words ending in -UST rhyme. The two factors interact: Lower frequency exceptions take longer to name than lower frequency regulars, but the two types of higher frequency items do not differ. The regular versus exception distinction was inherited from the dual-route model, which distinguishes between words pronounced by rule (regulars) and words that violate the rules (exceptions). Our models treat spelling–sound correspondences as a continuum: Spellings differ with respect to the degree of consistency in the mapping between spelling and sound. "Rule-

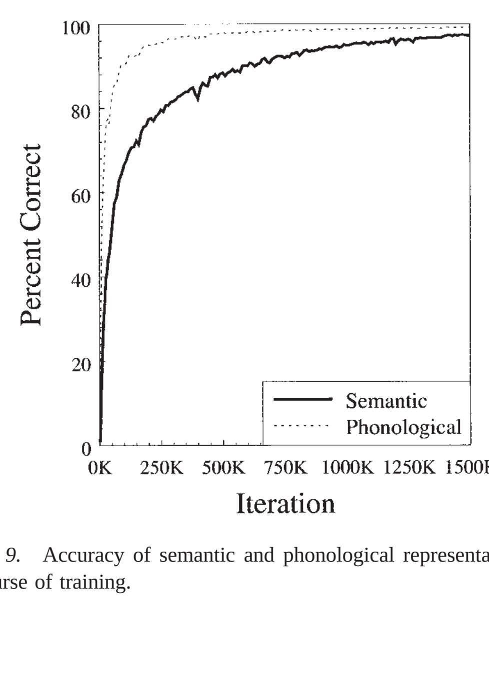


*Figure 9.* Accuracy of semantic and phonological representations over the course of training.

governed" forms and "exceptions" represent different points on this continuum; there are also intermediate cases such as MINT, which is rule governed but inconsistent because of the irregular neighbor PINT; see Jared, McRae, and Seidenberg (1990) for a summary of evidence that degree of consistency affects word naming.

Data from Taraban and McClelland (1987), Experiment 1A (from Table 2, p. 614), are plotted in Figure 10 (left). The conditions are labeled as in the original study. This result and others like it were replicated by the Seidenberg and McClelland (1989) model and analyzed by Plaut et al. (1996), who showed how the interaction of frequency and consistency arises from computational properties of simple connectionist networks.

### *Method*

The words from Taraban and McClelland (1987), Experiment 1A, were used. There are 96 words in four conditions that resulted from crossing frequency (high, low) and regularity (regular, exception).

Each item was presented to the trained network. In previous simulations of this effect (Plaut et al., 1996; Seidenberg & McClelland, 1989) the data concerned the mean summed squared error for the phonological code, which was computed in a single feedforward step. In the present model, the error computed at the end of processing was essentially zero for almost all items. This is because the model incorporates a phonological attractor, which tends to pull unit activities to their external values over time. In order to measure the difficulty the network had in reaching these states, we recorded the integral of the error over the course of processing the item from time step 4 to the final time step, 12 (the summation began with time step 4 because it takes four samples for information to flow to phonology from orthography via all routes).

### *Results*

The mean sum squared error is plotted in Figure 10 (right). There was a main effect of frequency, $F(1, 92) = 5.66$, $p < .02$, a main effect of regularity, $F(1, 92) = 4.19$, $p < .05$, and a marginally reliable interaction of the two, $F(1, 92) = 3.62$, $p < .06$. A post hoc test revealed an effect of regularity for the low-frequency items, $F(1, 46 = 4.04$, $p < .05$, but no such effect for high-frequency items ($F < 1.0$).

## *Simulation 4: Nonword Reading*

An important issue that arose regarding the Seidenberg and McClelland (1989) model concerned its relatively poor ability to generalize to novel forms (Besner, Twilley, McCann, & Seergobin, 1990; Coltheart et al., 1993), a limitation addressed in subsequent research (Harm & Seidenberg, 1999; Plaut et al., 1996; Seidenberg, Plaut, Petersen, McClelland, & McRae, 1994). It was therefore important to evaluate the new model's behavior on this task.

### *Method*

The model was tested on 86 nonwords from Glushko (1979), Experiment 1. This list consisted of 43 nonwords derived from consistent neighborhoods and 43 derived from inconsistent neighborhoods. Eighty nonpseudohomophone nonwords from McCann and Besner (1987) were also tested.

Each nonword was presented to the model, and the computed output was compared to the most common pronunciation (or, in some cases, the two most common pronunciations). For example, for the nonword GROOK, either /gruk/ (as in SPOOK) or /grʊk/ (as in CROOK) were considered correct. Seidenberg et al. (1994) found that the two most common pronunciations accounted for more than 90% of participants' responses to a large set of nonwords.

### *Results*

The model produced correct pronunciations for 93% of the nonwords derived from regular words and 84% of the ones derived from exception words. Corresponding results for the participants in the Glushko (1979) study were 93.8% and 78.3%, respectively. For the McCann and Besner (1987) stimuli, the model scored 83% correct, whereas human participants averaged 88.6%. The model performs slightly worse than people; this is mainly due to the fact that the exception nonwords include some spelling patterns that did not occur in the training corpus (e.g., the -JE in JINJE) and hence could not be fully represented in the orthographic units. This limitation could be overcome by using a non-slot-based representation (Plaut et al., 1996), by expanding the corpus to include multisyllabic words that contain the spelling patterns, or by mod-

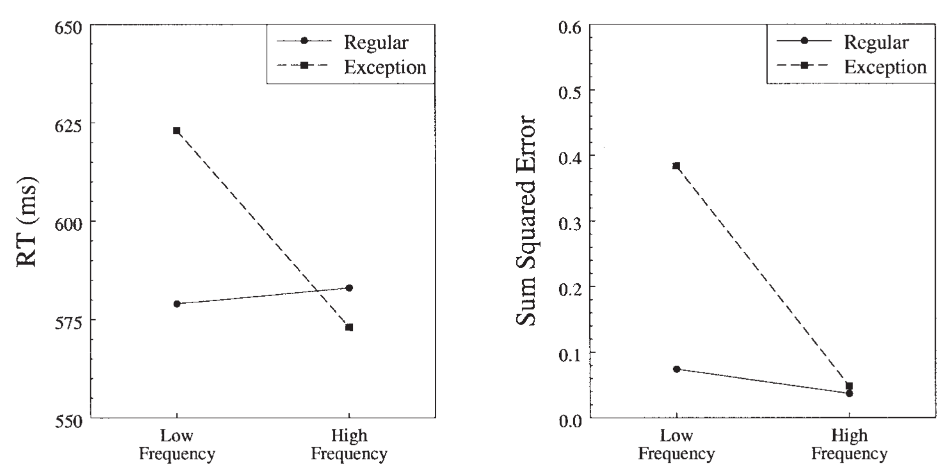


*Figure 10.* Frequency by regularity interaction. Data are from Taraban and McClelland (1987), Experiment 1A (left), and simulation results (right) of the integrated sum squared error (see text). RT = reaction time.

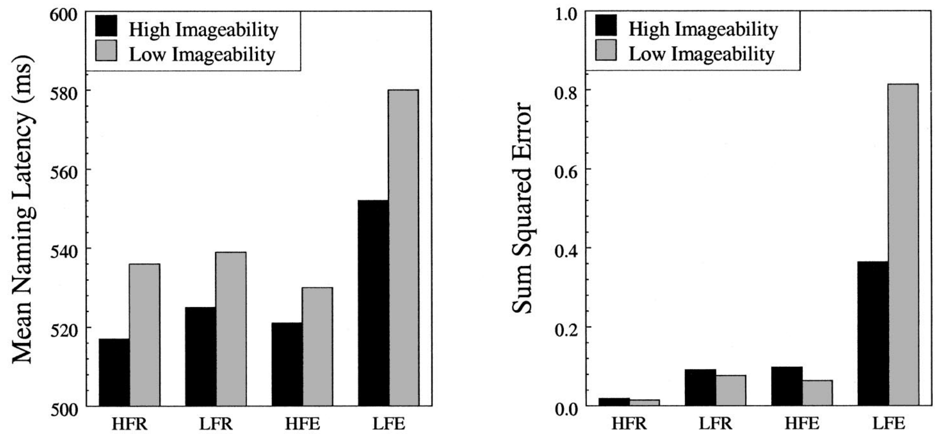


*Figure 11.* Data are from Strain et al. (1995; left) and Simulation 5 (right). Statistically reliable effects of imageability were only observed for lower frequency exception words in both experiment and simulation. Note that the stimuli in the experiment and simulation were not identical, as explained in the text. HFR = high-frequency regular; LFR = low-frequency regular; HFE = high-frequency exception; LFE = low-frequency exception.

eling additional strategies that participants may use in pronouncing difficult nonwords (e.g., pronounce JINJE by reference to INJURE).

## *Simulation 5: Imageability Effects*

As noted in the introduction, many studies have demonstrated effects of phonological variables on the computation of meaning. Here we consider the reciprocal effect, in which semantic properties of words affect naming. Such effects have been observed in brain-injured patients whose ability to compute from orthography to phonology has been compromised. Thus, there are semantic paraphasias in deep dyslexia (Coltheart, Patterson, & Marshall, 1980) and concreteness effects in phonological dyslexia (Patterson, Suzuki, & Wydell, 1996). However, semantic effects on naming have also been observed in unimpaired readers. Models such as Seidenberg and McClelland's (1989) suggest that most monosyllabic words can be read using the orth→phon pathway. The model performed most poorly on relatively low-frequency words with atypical spellings and pronunciations such as *angst* and *barre*. Thus, the model suggested that correctly reading such words requires additional input from orth→sem→phon (Plaut et al., 1996).

Strain, Patterson, and Seidenberg (1995) tested this prediction by examining effects of imageability, a semantic variable, on the naming performance of skilled adult readers. Their stimuli factorially varied imageability, frequency, and spelling–sound regularity. The prediction, then, was that there would be an effect of imageability (higher imageability words named faster than lower) only for low-frequency words with irregular spelling–sound correspondences. The main results from their study, shown in Figure 11 (left), exhibited this pattern. The Strain et al. result is important because it represents a nonobvious prediction concerning the involvement of orth→phon→sem in naming based on analyses of the capacities of orth→phon.[11] We therefore examined whether the present model would replicate this effect.

### *Method*

Many of the items used by Strain et al. (1995) were multisyllabic and could not be used in this simulation. A new stimulus set exhibiting the same properties was therefore constructed. We first performed a median split of all items in the training set along the frequency dimension. All words were then categorized as regular or exception. Finally, we used the imageability norms of the Medical Research Council Psycholinguistic Database (Coltheart, 1981) to code all items in the training set that were in the database and did a median split on these items, categorizing them as high or low in imageability. We then identified words that fit each of the categories formed by crossing frequency, regularity, and imageability.

The smallest number of items, 28, was obtained for the low-frequency, low-imageability irregular cell in the design. For each of the other cells in the design we randomly chose 28 of the qualifying words. All words were presented to the model, and its output was analyzed as in the simulation of frequency by consistency.

### *Results*

Figure 11 (right) shows the results. The three-way interaction of frequency, regularity, and imageability was reliable, $F(1, 216) = 3.97$, $p < .05$. The effect is clearly carried by the lower frequency exception words as in Strain et al. (1995). When the data were reanalyzed collapsing across the imageability factor, a reliable frequency by regularity interaction was observed, $F(1, 220) = 12.1$, $p < .001$, replicating the pattern observed in Simulation 3 using the Taraban and McClelland (1987) stimuli.

[11] Ellis and Monaghan (2002) questioned the reliability of the Strain et al. result, noting that the predicted interaction with imageability was not statistically significant if one irregular item, COUTH, was removed from the stimuli. Removing this item changes the significance level to .08 but does not otherwise affect the pattern of results. Moreover, the same interaction of frequency, regularity, and imageability was found by Strain and Herdman (1999), whose results also do not depend on including the word COUTH.

In summary, the model learned to accurately compute phonological and semantic codes from orthography, exhibited basic phenomena observed in participants and in earlier models, and generated plausible phonological codes for nonwords. The model demonstrates the feasibility of an approach in which semantics builds up based on input from both orth→sem and orth→phon→sem components.

## DIVISION OF LABOR

### Model Dynamics and Effects of "Lesioning"

We now consider the central issue addressed in this research, the model's division of labor in the computation of meaning and its relationship to human performance. We have seen that the model was able to compute the meanings of words accurately. The question is, how? Specifically, to what extent is the computation of meaning driven by the orth→sem versus orth→phon→sem components? As a first step, we report a simulation that provides information about how rapidly input arrives at the semantic layer from different sources. We then report analyses of how the model performed with one or the other pathway disabled ("lesioned").

### *Simulation 6: Dynamics of the Trained Reading Model*

The dynamics of the reading model are complex. The theoretical model assumes that activation spreads in continuous time, much like electricity in a circuit or water pressure in a plumbing network. Thus, in principle, activation to semantics arrives continuously from all sources and builds over time. In practice, a discrete time approximation is required. Time is sampled, and the behavior of the network is updated at each time sample. In training the network, 4 units of whole time were used, sampled over 12 discrete time slices; hence, each sample was 0.333 units of time in duration. The strength of activation from each pathway varies according to factors that we explore in the remainder of the article.

For each discrete sample, activity spreads from the orthographic representations to semantics and phonology along the direct connections, and to the hidden units along those pathways (see Figure 8), causing the activity in those units to begin to rise. On subsequent samples, as units increase in activity, their influence on subsequent units increases. As the influence of orthography on phonology increases, that in turn influences semantics, which is also influenced by orthography. As the semantic and phonological representations build up, they are influenced by their respective attractors, and they begin to influence each other as well. In the theoretical model, activation builds up throughout the network continuously; in practice, it is a close approximation to continuously. Activation of the semantic representation accumulates from both pathways in this fashion. However, the rate at which activity builds up along the various pathways is a function of the representational capacity of those pathways and of how tuned to aspects of the stimuli those pathways have become.

The purpose of this simulation was to examine the time course of activation along different pathways. The data concern the activation of semantics from orthography (from both the direct and hidden-unit-mediated pathways), the activation of phonology from orthography (again, from both direct and hidden-unit-mediated pathways), the activation of semantics from phonology, and the activation of semantics from the cleanup units.

#### *Method*

All words in the training set were presented to the trained reading model. To assess the time course of activity at a more fine grain, we ran the network for 4 units of whole time, as in training, but discretized over 48 samples, rather than 12, giving an integration constant of 0.083. The total input to target phonological units from the orth→phon path was summed at each sample. Similarly, the total input to target semantic units from orth→sem, from phon→sem, and from the semantic cleanup units was measured at each sample.

#### *Results*

As indicated in Figure 12, the input to semantics from orthography and the input to phonology from orthography rise at very similar rates, with orth → phon having a somewhat higher asymptote. Of interest, the contribution to semantics from phonology rises at a much slower rate. Activation from phonology to semantics cannot begin until significant activation builds up on the phonological units from orthography. Hence, the phon→sem line in Figure 12 rises at a rate proportional not to the constant input from orthography (unlike orth→sem and orth→phon) but rather at a rate proportional to the activity in phonology, indicated by the orth→phon line. Hence, although orthography directly activates semantics and phonology rapidly, the contribution to semantics via orth→phon→sem lags behind. The cleanup units are the weakest source of input to semantics; their activity is driven by activity in semantics itself and is limited by the very sparse nature of the semantic representations.

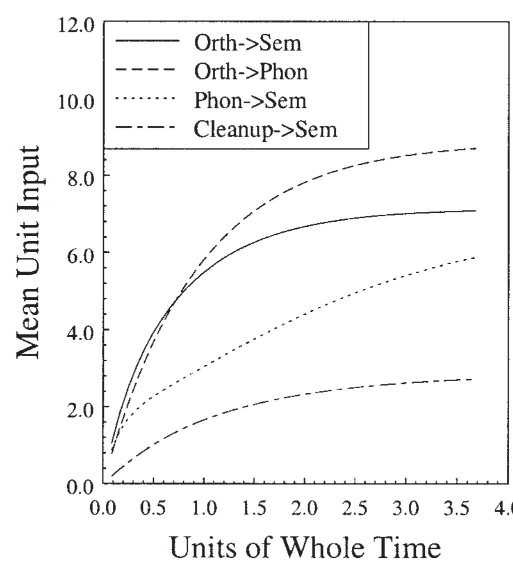


*Figure 12.* Input to phonological and semantic units over time. Activation rises most rapidly for the phonological and semantic units, which are closest to the orthographic input; however, the phonological units reach higher asymptotic levels, indicating somewhat better learning of this mapping. Activation of semantic units from phonology occurs more slowly because the phonological units must first be activated sufficiently by orthography. In this and subsequent figures, Orth = orthography, Sem = semantics, and Phon = phonology.

Figure 12 demonstrates two of the key properties of this model. First, the activation of semantic information is driven by input from multiple sources; there is no one pathway that is doing all of the work. Second, the strength of that input varies according to properties of the pathways. In the fully trained model activation arrives more rapidly from orth→sem than orth→phon→sem. It is equally important to note, however, that over most time steps there is significant input to semantics from both pathways. Moreover, this analysis ignores the interactivity between semantics and phonology that occurs in the intact model. As orthographic information begins activating semantics, that in turn activates phonology via the sem→phon pathway, which in turn can further activate semantics via the phon→sem pathway. This property also contributes to the involvement of both pathways in the activation of meaning. Finally, the contributions from the different pathways are modulated by word-specific properties such as frequency and homophony as described below.

Figure 13 shows how individual features for a typical item, BOOT, are activated over time by the orth→sem and orth→phon→sem pathways, and the total of the two. The [object], [artifact], [covering], and [footwear] features are shown. For most features, the orth→sem pathway dominated the computation. However, for the [artifact] feature, the orth→phon→sem pathway provided greater input toward the end of processing. For all four target features, both pathways are providing positive input; thus, the sum of their contribution is greater than either pathway's contribution alone.

### *Simulation 7: Development of the Division of Labor*

We next present a series of simulations that provide further information about the division of labor using a lesioning methodology. The first of these simulations examined the model's accuracy in computing semantics over the course of training under three testing conditions: the intact model, the model with input from orth→sem disabled (i.e., with the direct and hidden-unit-mediated orth→sem connections disabled), and the model with input from orth→phon→sem disabled. The model was tested on all items in the training corpus once every 10,000 trials with each configuration of the model. Thus, the intact model was trained throughout but was tested at regular intervals in the three ways described above.[12]

The model was tested on all words in the training corpus with performance scored as described previously. The results are summarized in Figure 14. The accuracy of the intact model rises rapidly, then flattens out, growing more slowly for the remainder of the training period. Initially, the accuracy of the intact model and that of the model with only orth→phon→sem parallel each other, indicating that the latter is doing most of the work. Quickly, however, the performance of the intact model surpasses that of the phonology-only model, whose performance reaches asymptote. After the orth→phon→sem pathway peaks, increases in the accuracy of the intact model are due to additional learning within orth→sem. Note also that orth→sem continues to improve even after learning in the intact model has slowed.

Figure 14 reveals an important result. Early in training, the phonological pathway is responsible for much of the accuracy of the intact model. This is because orth→phon is easier to learn than orth→sem, for reasons discussed previously. However, the orth→sem pathway continues to develop for two reasons. First, the model cannot read many homophones correctly via orth→phon→sem because of their inherent ambiguity; second, even when orth→phon→sem activates the correct semantics of a word, the orth → sem pathway continues to develop because of the pressure to respond quickly. The orth→phon→sem pathway must compute an intermediate representation (phonology) to activate semantics; this limits its speed. Thus, although the orth→sem pathway is more difficult to learn, it has the potential to activate semantics more rapidly than orth→phon→sem. Moreover, English monosyllables contain far more homophones than homographs, and thus the orth→sem pathway has much less intrinsic ambiguity than orth→phon→sem.

One thing that is not clear from Figure 14 is whether different words are being read by different pathways. It is possible, for example, that the model could partition the words such that some are largely read via orth→phon→sem and others by orth→sem. Words correctly read by the intact network were categorized into four disjoint subgroups: those that require both pathways to be read (cannot be read by either path in isolation), those that can be read by either pathway, those that can be read by orth→sem but not orth→phon→sem, and those that can be read by orth→phon→sem but not orth→sem. Figure 15 shows this breakdown over the course of development.

As expected, there is an initial burst of words that can be read only by the phonological pathway. This advantage begins to fall off by 500,000 training trials, at which point more words can be read by either route. Of interest, at that point about 15% of the items can only be read by the orth→sem route. This number grows to about 22%, where it flattens out. Asymptotically, about half of the words are redundant; they can be read accurately by either route. Fairly low percentages of items can be read only with input from both pathways, only by orth→phon, or only by orth→sem.

This behavior of the model is consistent with a central finding in the reading acquisition literature: the importance of phonological information in the early stages of learning to read (Adams, 1990; Bradley & Bryant, 1983; Liberman & Shankweiler, 1985). The system initially affords both orth→sem and orth→phon→sem possibilities. Development within the two subsystems is determined by their inherent computational properties: Orthography and phonology are correlated, phon→sem is known, and orth→sem is difficult to acquire but ultimately faster to compute. The system (and by hypothesis the child) does not choose an initial strategy or switch strategies as skill is acquired; rather it responds

[12] The lesioning methodology is informative about the capacities of different parts of the network, but it should be noted that the role of a given component (orth→sem or orth→phon→sem) in the intact model is not identical to its role in isolation. The semantic system is an attractor, meaning that the activation of units changes over time based on input from both pathways and feedback within the semantic attractor itself. In this highly interactive system, the extent to which, say, orth→sem contributes to the activation of meaning depends in part on how much input there has been from orth→phon→sem. The lesion method is informative about what activation each pathway delivers to semantics, not the subsequent interactivity within the semantic attractor. The lesion procedure also provides information about the capacity of the remainder of the system when one component is eliminated.

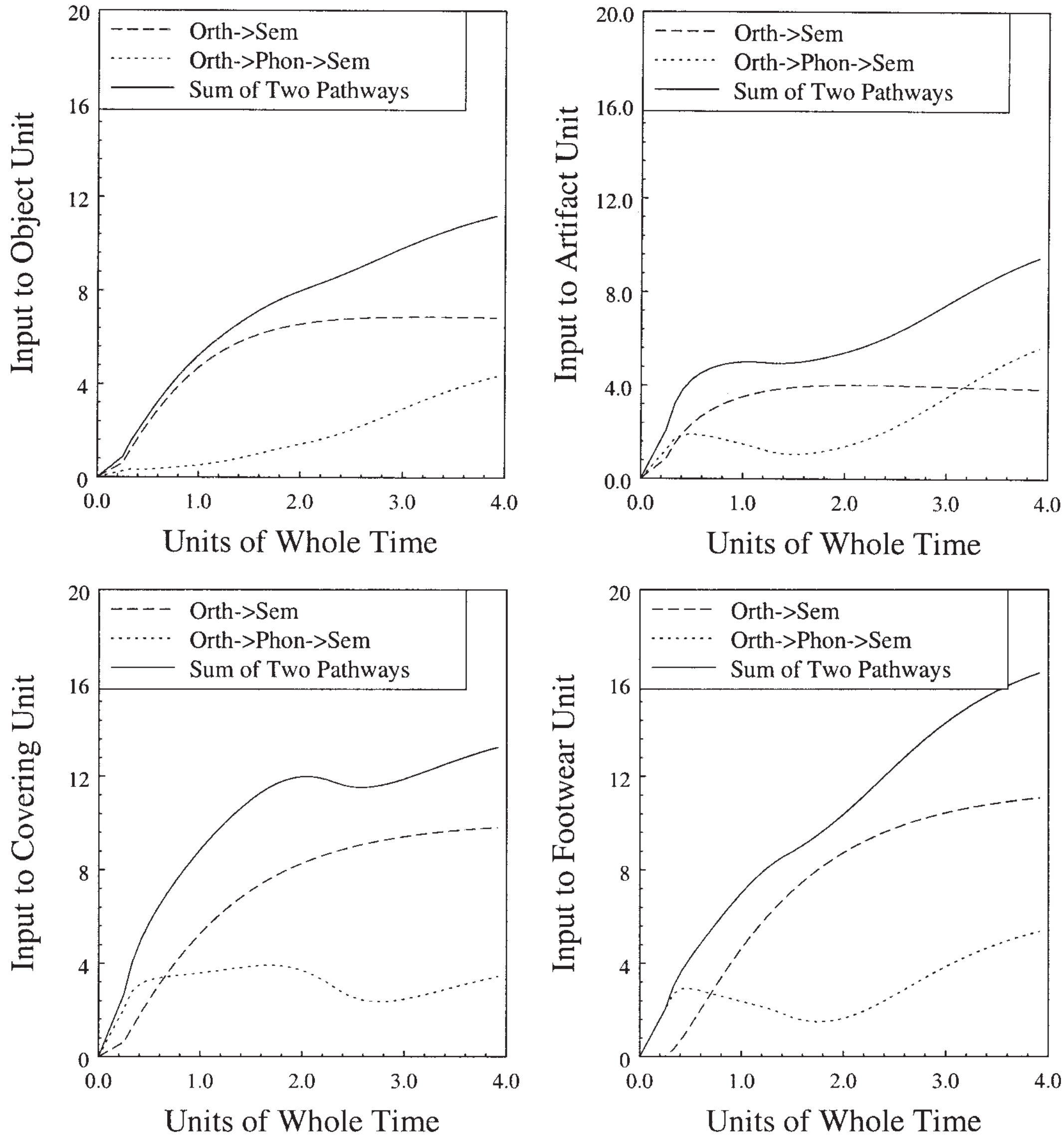


*Figure 13.* Sources of the activation of individual semantic features for a typical word, BOOT. All four types of features receive significant input from both direct (orth→sem) and phonological (orth→phon→sem) pathways; thus, the activation summed over the two sources of input is greater than for either pathway in isolation.

to the task it is assigned: computing the meaning of the word quickly and accurately, subject to intrinsic computational constraints, yielding the observed division of labor. The model also suggests that the division of labor gradually shifts as skill is acquired, with the orth→sem pathway becoming increasingly efficient over time.

These results need to be interpreted carefully, however. The analysis in Figure 14 provides information about the capacities of each component of the system. It is clear, for example, that the orth→phon→sem component develops more rapidly than orth→sem. However, as we have noted, in the intact model semantics receives activation from both parts of the system. The words in the by-either-path condition make this point most clearly. The fact that they can be read by either path in isolation means that both paths will be strongly activating semantics in the intact model. Similarly, there are words that can only be correctly read by orth→sem in isolation, but it would be incorrect to infer that these words only receive activation via this pathway in the intact model. Below we present additional analyses bearing on this point.

### *Simulation 8: Speed Effects*

The pressure to activate semantics rapidly is an important property of the model; it is what forces the orth→sem pathway to continue to develop even for words correctly recognized by the orth→phon→sem pathway. In this simulation we examined how the intact model and the two paths in isolation compare in terms of how rapidly semantics is activated.

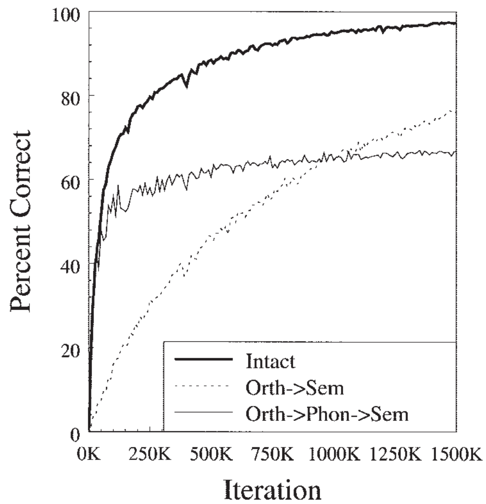


*Figure 14.* Division of labor assessed using a "lesioning" method. The data reflect the accuracy of the computed semantic representations in the intact model (input from both pathways) and with either the orth→sem or the orth→phon→sem component disabled. Early in training the intact model performs little better than the isolated phonological pathway. However, performance in the phonological pathway rapidly reaches asymptote, whereas performance in the orth→sem pathway continues to improve.

As before, all words in the training set were tested. The time course of semantic activation was assessed as follows. The network was run for 4 units of time, as before, but again a finer discretization was used to more precisely measure time. In this simulation, the 4 units of time were discretized over 48 samples, giving an integration constant of 0.083. An item was assumed to be recognized when all semantic features had settled—that is, their activation values did not change by more than 0.05 for 0.5 units of time (6 samples). Settling times were computed for all correct items and averaged. This measure was taken at various points in development as in the previous simulation.

The results are shown in Figure 16. Because the network was pressured to activate semantics quickly as well as accurately, latencies continued to decrease even after accuracy was high.

As noted in the previous section, a number of words can be read by either pathway in isolation. This fact masks a subtle but important point that is revealed by the latency analyses: The effect of the two components working together is different from the effect of each in isolation. The speed of the orth→phon→sem path eventually flattens out; its maximum is limited by the fact that it must compute a reasonably stable phonological representation to begin activating semantics. There is no such limitation on the orth→sem pathway, which continues to improve over time. As a result, the overall speed of the network also improves with training. Of importance, the speed of the network with both components operating is faster than the speed of either component in isolation. This arises because of the processing dynamics of the model; as shown in Figure 2, the rate at which a unit's activity increases is a function of the strength of its input activation. Thus, the network achieved greater efficiency using both components. This property stands in contrast to the "horse race" model of Paap and Noel (1991), in which the latency to recognize a word is chiefly determined by which of two independent routes finishes faster.

### *Simulation 9: Reading With Reduced Phonological Feedback*

As shown in Simulation 7, the orth→phon→sem pathway develops more rapidly than orth→sem. In this simulation we explored the effect of reducing the phonological feedback the network received, which forced the model to rely more on the orth→sem pathway.

#### *Method*

*Materials.* All items in the training set were used.

*Procedure.* The reading model described above was retrained with a change in procedure: Feedback about the accuracy of computed phonological codes was provided on only 1% of the training trials, whereas feedback about semantics was provided on all trials as before. The same orth→phon→sem model was used, and the model was again trained for 1.5 million trials.

#### *Results and Discussion*

At asymptote, the normal model computed the correct semantics for 97.3% of the items in the training set; the model with reduced feedback on phonology was correct on 91.8% of the items. The reduced phonology (RP) model also took much longer to reach this

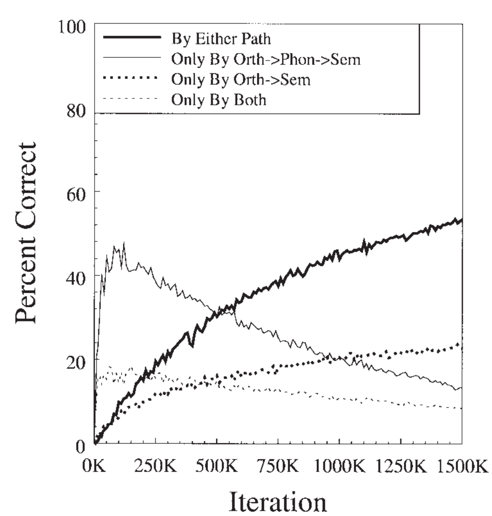


*Figure 15.* Accuracy of each component of the model in computing the semantic patterns for words. Early in training correct output is mainly produced by the phonological pathway, reflecting more rapid learning within orth→phon than orth→sem. This is consistent with the predominance of phonological recoding in children's early reading. With additional training, however, the largest class consists of words for which both pathways produce correct output ("By Either Path"). The relatively small class of words that require input from both pathways ("Only By Both") primarily consists of the subordinate meanings of homophones. These analyses provide information about what has been learned in each pathway; however, even if a word cannot be read by a given pathway in isolation, it may contribute significant partial activation in the intact model. In fact, almost all words receive some activation from both pathways.

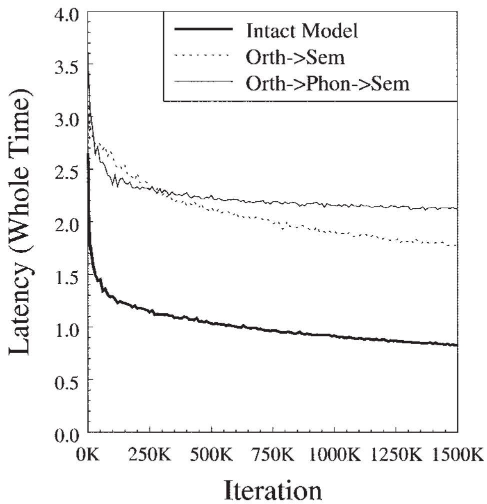


*Figure 16.* Semantic latencies for words processed by individual pathways and by both together over the course of training. The main finding is that the two pathways acting together produce output more rapidly than either in isolation. Units are measures of whole time, as defined by Equation 4.

lower level of asymptotic performance. Figure 17 shows the accuracy of the normal model, the intact RP model, and the component pathways of the RP model. The RP model exhibited less reliance on the orth→phon→sem pathway throughout training compared to the normal model and a greater reliance on the orth→sem pathway. Throughout development, the RP model's performance lagged behind the intact model.

Figure 18 shows the latencies of the models over the course of development. The mean latency on correct items for the normal model was 0.82 units of time, whereas the mean latency on correct items for the reduced phonological feedback simulation was 1.08 units of time. This effect of simulation condition, measured over items that were correct in both simulations, was reliable, $F(1, 5521) = 182.5$, $p < .001$.

The asymptotic differences in latency and accuracy between the RP model and the normal model were not very large. However, there were pronounced developmental differences. Reducing feedback on the sounds of word forms significantly reduced the rate at which the meanings of words can be learned and the speed at which this computation can be performed.

This simulation makes two points. First, it provides further support for the observation that the model performs most efficiently (in terms of speed, accuracy, and rate of learning) using input from both components. Second, the simulation has some suggestive implications regarding methods for teaching reading. One of the main controversies in reading education concerns whether or not instruction should emphasize the correspondences between the spoken and written forms of language. "Whole language" methods tend to discourage this type of instruction, focusing instead on developing efficient procedures for computing meanings directly from print. The present simulation suggests that failing to provide feedback about spelling–sound relations may make the task of learning to compute meanings more difficult. The simulation can only be taken as suggestive because we have not examined all of the factors that can play a role in learning to read words; whole language methods, for example, often emphasize the use of linguistic and nonlinguistic textual information and guessing strategies in place of phonological recoding. Moreover, the reduction in phonological feedback in the simulation was severe and so represents an extreme case. Other factors being equal, however, feedback about both the meanings and sounds of written words will yield more rapid acquisition and better performance than meaning alone.

## *Simulation 10: Modulation of Division of Labor by Frequency*

We now examine several lexical factors that have been widely studied in behavioral experiments that influence the division of labor. One issue raised by behavioral studies is whether the relative contributions of the different pathways depend on word frequency, with more input from orth→sem for higher frequency words. This simulation examined how frequency affected division of labor in the model.

### *Method*

*Stimuli.* Items for testing were selected as follows. The training set items were sorted according to frequency, and 500 items from the top one third were selected randomly; these were the high-frequency items used for testing. Another 500 items were selected randomly from the bottom third; these were the low-frequency items used for testing. This yielded a very strong frequency manipulation, $t(1016) = 29.15$, $p < .001$, where the mean high-frequency item had a probability of presentation of .42 and the mean low-frequency item had a probability of presentation of .05.

*Procedure.* The network was tested on each item at the conclusion of training, and accuracy over the semantic units was recorded for the model with no orth→sem pathway and for the model with no orth→phon→sem pathway.

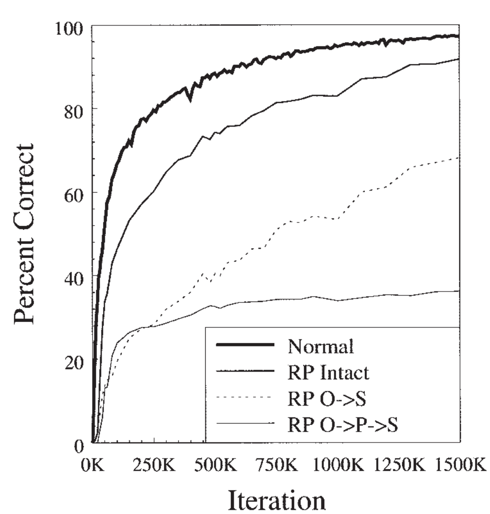


*Figure 17.* Accuracy by pathway for the normal model and for the model with reduced phonological feedback (RP) over the course of training. The RP model learns more slowly, with the biggest decrement in the O→P→S pathway. O = orthography; S = semantics; P = phonology.

### Results and Discussion

Figure 19 summarizes the main results at asymptote. As expected, high-frequency items were read more accurately than low-frequency items; however, frequency interacted with pathway. For high-frequency items, the orth→sem pathway performed more accurately than the orth→phon→sem pathway. For low-frequency items, the difference was much smaller. Considering the high- and low-frequency items over the orth→sem and orth→phon→sem pathways, the interaction was reliable, $\chi^2(1, N = 900) = 5.94$, $p < .015$. The accuracies for the intact model were 99% for the high-frequency items and 95% for the low-frequency items.

Recall that the model is pressured to produce the semantics of the word as rapidly as possible by creating error for each sample of time that the model has not yet settled to the correct semantic representation for that word. Over the course of training, this error affects the network weights; hence, the network is pressured to reduce the running error over all words on which it is trained. Words are presented probabilistically; early in training, an error on a frequent word such as THE affects the network much more than an error generated on presentation of a much lower frequency word such as YULE. Minimizing the total error is therefore best accomplished by primarily optimizing the high-frequency items over the low-frequency items. Thus, although all items are pressured to be read as quickly as possible, the more rapid orth→sem pathway receives greater pressure from the high-frequency items than the lower frequency ones.

As an example, the frequency of presentation of the word THE is 20 times that of BRIM, meaning that the error due to slowness in processing items is 20 times greater for THE than BRIM. Thus, all other things being equal, the network resources allocated to rapidly processing THE will far outpace those allocated to processing BRIM.

This behavior of the model strongly contradicts Smith's (1971, 1973) conjectures about the efficiency of different decoding strategies. Smith (1971) argued that reading is accomplished too rapidly to accommodate phonological recoding. However, Zipf's (1935) law states that there is a constant relationship between the number of words at a given frequency range and the square of that frequency range; that is, the frequency histogram for any language follows a curve $y = k/x^2$, for some constant $k$. Only the most highly frequent items tend to violate this relationship. What this means is that there are a very small number of words that occur very frequently, and a very large number of words that are much more infrequent. Even if strong reliance on orth→sem is limited to these highest frequency words, they account for a large proportion of the tokens a person reads.

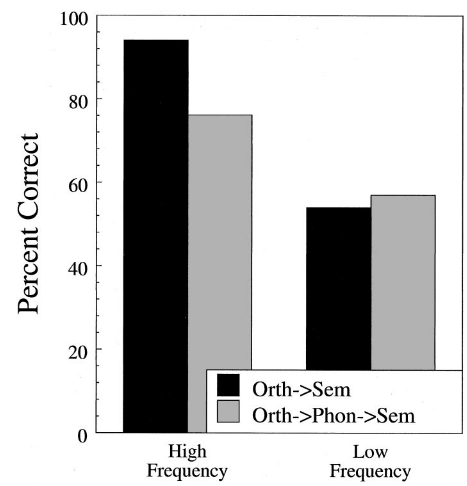


*Figure 19.* Division of labor in the computation of semantics: Effects of word frequency. The data are for each pathway in isolation. For higher frequency words, the orth→sem pathway is more accurate; for lower frequency words, both pathways are equally accurate.

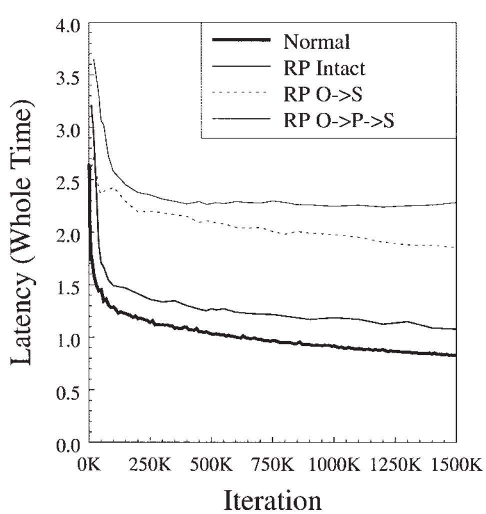


*Figure 18.* Latencies in the normal and reduced phonological feedback (RP) models over the course of training. The RP model computes semantic codes more slowly, with the biggest decrement again in the O→P→S pathway. O = orthography; S = semantics; P = phonology.

Figure 20 shows the latencies (calculated as described previously) for these items by path and frequency. As in Figure 16, the intact model is faster than either the orth→sem or orth→phon→sem paths alone. The high-frequency items are computed more rapidly by the orth→sem path, whereas the low-frequency items are computed about equally fast by both. The interaction of frequency and pathway (orth→sem vs. orth→phon→sem) was reliable, $F(1, 998) = 73.47$, $p < .001$. The intact model also showed a main effect of frequency in its latencies, $F(1, 998) = 252$, $p < .001$. We matched the items used in this test with items from a large-scale study of reading times (Seidenberg & Waters, 1989). There were 351 high-frequency items and 122 low-frequency items present in both lists; in the latencies reported by Seidenberg and Waters these items showed a strong frequency effect, $F(1, 471) = 29.6$, $p < .001$. The subset of items in both lists also show a strong frequency effect in the model, $F(1, 471) = 210.1$, $p < .001$.

The intact model also showed a main effect of frequency in its latencies, $F(1, 998) = 252$, $p < .001$. Matching these test items with those used in a large-scale collection of reading times (Seidenberg & Waters, 1989) revealed that participants also exhibit a reliable advantage for these high-frequency items ($p < .001$).

### Simulation 11: Interaction of Frequency and Consistency

The above analysis considered the effects of frequency on the division of labor in computing meaning. We next examined

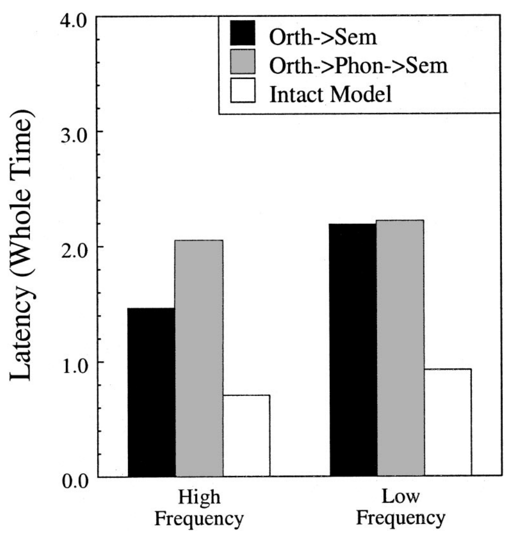


*Figure 20.* Effects of frequency on latencies to compute semantics by individual pathways and in the intact model.

whether these effects are modulated by another lexical factor, spelling–sound consistency (Seidenberg & McClelland, 1989). Consistency affects the difficulty of computing phonological codes, especially for lower frequency words (Seidenberg, Waters, Barnes, & Tanenhaus, 1984; Taraban & McClelland, 1987). This factor should therefore slow the activation of semantics via orth→phon→sem, creating greater dependence on orth→sem (see also Strain et al., 1995).

## *Method*

Words in the training set were categorized according to their consistency, which, as in previous studies (Jared et al., 1990), was defined in terms of orthographic rimes (e.g., -INT in MINT).[13] All items with a word's orthographic rime that have the same phonological pronunciation are counted as *friends* of that word (e.g., LINT, TINT). Words with that rime but a different pronunciation (e.g., PINT) are *enemies*. If a word has more enemies than friends, it was categorized as inconsistent. The inconsistent items also included strange words such as YACHT, which have neither close friends nor enemies.

Frequency was coded as high or low using a procedure similar to that used in Simulation 10, except that a median split of the items into low and high frequency was used in order to have a larger set of items per cell. This yielded four conditions: high-frequency consistent, high-frequency inconsistent, low-frequency consistent, and low-frequency inconsistent items. A total of 225 items were sampled randomly from each cell and used for analysis. The network parameters and presentation method were the same as in Simulation 10.

## *Results and Discussion*

Figure 21 shows the effects of frequency and consistency along the direct and phonological pathways. A log-linear analysis of the data (which is essentially a chi-square test for data with more than two dimensions) revealed a reliable three-way interaction of frequency, consistency, and pathway, $\chi^2(4, N = 900) = 13.68$, $p < .01$. The relative accuracy of the two pathways is clearly mediated by frequency and consistency. For the higher frequency words, the orth→sem pathway is mostly unaffected by spelling–sound consistency and so performs quite well on both consistent and inconsistent items. The orth→phon→sem pathway, in contrast, performs more poorly on inconsistent words. For lower frequency words, an interesting pattern appears. Orth→sem and orth→phon→sem perform about equally well on consistent items, but orth→phon→sem performs more poorly on inconsistent items. The interaction of pathway and consistency was reliable for the low-frequency items, $\chi^2(1, N = 900) = 5.99$, $p < .015$, and marginally reliable for the high-frequency items, $\chi^2(1, N = 900) = 3.23$, $p < .073$.

Thus, frequency and consistency jointly affect the division of labor. Consistency shows a strong effect on the orth→phon→sem pathway for both low- and high-frequency items. The magnitude of the consistency effect, collapsing across frequency, was greater for the orth→phon→sem pathway than the orth→sem pathway, $\chi^2(1, N = 900) = 9.06$, $p < .003$, whereas the magnitude of the frequency effect, collapsing across consistency, was marginally greater for the orth→sem pathway than the orth→phon→sem pathway, $\chi^2(1, N = 900) = 2.71$, $p < .1$.

Figure 22 shows the latencies for these items by pathway, including the intact model. With regard to the orth→sem and orth→phon→sem pathways, the three-way interaction of pathway, frequency, and consistency was not reliable. There was a reliable interaction of frequency and pathway, $F(1, 896) = 173.06$, $p < .001$, and consistency and pathway, $F(1, 896) = 40.9$, $p < .001$.

With regard to the intact model, there was an interaction between frequency and consistency, $F(1, 896) = 4.4$, $p < .05$; the effect of consistency was not reliable for high-frequency items (*M*s = 0.756 vs. 0.730) $F(1, 448) = 1.7$, $p > .150$, but it was for the low-frequency items (*M*s = 1.08 vs. 0.96), $F(1, 448) = 9.26$, $p < .01$. This is particularly important, because numerous studies have shown that in standard word recognition tasks, consistency effects are not found for high-frequency items.[14] Inspection of the lesioned models suggests why this may be the case. For the high-frequency items, the orth→phon→sem pathway has reliably lower latencies for consistent items than inconsistent (*M*s = 2.22 vs. 2.49), $F(1, 448) = 17.6$, $p < .001$, whereas the orth→sem pathway has reliably lower latencies for inconsistent items than consistent (*M*s = 1.47 vs. 1.69), $F(1.448) = 13.3$, $p < .001$. This pattern of results illustrates two important properties of the model: First, the success or failure of one path drives the success or failure of another path, and second, although the operation of the intact

[13] We followed this procedure because many studies have shown that consistency defined in terms of this unit has a significant impact on processing and because it is the subword unit that has the biggest impact in our models (see Jared et al., 1990). Statistical regularities involving other parts of words can also affect performance, but not as strongly.

[14] Jared (1997) reported a consistency effect for higher frequency words in contrast to previous studies in which consistency (or regularity) had no effect in the higher frequency range. Jared's "high-frequency" items were much lower in frequency than in studies such as Taraban and McClelland's (1987). For example, the mean Kučera and Francis (1967) frequency for the high-frequency inconsistent items in Jared's Experiment 1 was 127; in Taraban and McClelland's research the mean frequency for the comparable items was 952. In our models, as the frequency of the word itself decreases, the effects of neighboring words increase. Thus, as Jared noted, the Seidenberg and McClelland (1989) model simulated her results quite closely.

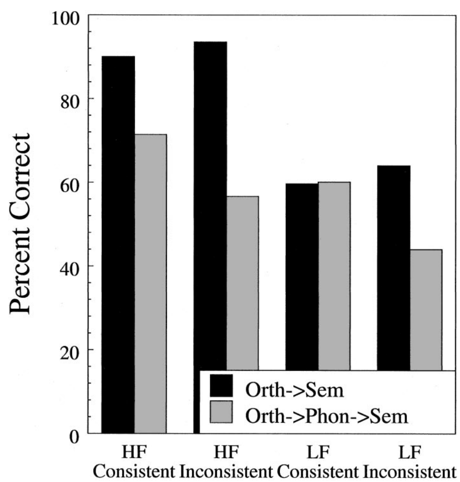


*Figure 21.* Division of labor in the computation of semantics: Effects of frequency and spelling–sound consistency in each pathway. In this and subsequent figures, HF = high frequency and LF = low frequency.

system may reveal no effect of a given stimulus condition in some contexts, this may in fact arise from robust (but opposite) effects in the component pathways.

## *Simulation 12: Morphological Regularities*

We next provide data concerning the model's knowledge of morphological regularities. This simulation is a replication of Simulation 2, which addressed the same issue in the phonology–semantics model. As we have noted, inflectional morphology involves nonarbitrary mappings between form and meaning. The Phase 1 model learned that certain phonological forms tend to be associated with features such as plural and past tense. Here, we examined whether the reading model also learned about these regularities. In particular, would the orth→sem component learn that the spelling -ED is strongly associated with pastness and that the spelling -*s* is associated with plural and third-person singular?

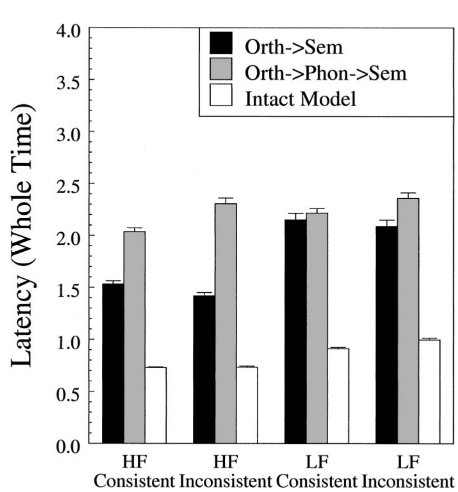


*Figure 22.* Effects of frequency and consistency on latencies to compute semantics by individual pathways and in the intact model. Error bars represent standard errors of the mean.

### *Method*

The stimuli were the orthographic forms of the nonwords used in Simulation 2. These included uninflected nonwords such as GOME, nonwords with the past tense morpheme -ED such as GOMED, and ones with the plural/third-person singular morpheme such as GOMES. These items were tested using three versions of the fully trained model: intact, severed orth→phon, and severed orth→sem. The computed semantic and phonological codes were recorded, and the activation of the plural, past tense, and third-person-singular features were examined.

### *Results*

As indicated in Table 2, the intact model produced a plausible inflection for 82% of the plural inflected items and 100% of the past tense items. Of interest, the orth→sem pathway in isolation was almost as accurate as the intact model, whereas the isolated orth→phon→sem pathway was less accurate. The data indicate that the orth→sem pathway encoded the fact that -ED and -S are associated with particular semantic features. Thus, there was learning of sublexical quasi-regularities within the orth→sem component.

The model was more accurate in determining the inflection of past tense nonwords than plurals. This is because in the training set, -ED at the end of the coda is very strongly predictive of past tense; no items with a coda ending in -ED are not past tense. However, many items end in -S that are not semantically plural or third-person singular (BUS, PLUS, NEWS). Hence, -ED is a much better morphological cue than -S; this is reflected in the model's performance.

Although our model was not designed to address the many issues that have arisen concerning inflectional morphology, principally the past tense in English, its behavior is relevant to one of the major controversies. The model was trained on words that are inflected for tense (e.g., BAKED) or number (e.g., DOGS) because many of them are included in the training set. It learned to produce the correct semantics of such words from their phonological rep-

Table 2
*Morphological Effects in Reading Model (in Percentages)*

| | Feature | | |
|---|---|---|---|
| Simulation | Plural | Third person | Past tense |
| Intact model | | | |
| Plural | 62 | 20 | 0 |
| Past tense | 0 | 0 | 100 |
| Stem | 0 | 0 | 0 |
| Orth→phon→sem | | | |
| Plural | 55 | 1.2 | 1.2 |
| Past tense | 4.1 | 2.0 | 37.5 |
| Stem | 1.6 | 0 | 1.6 |
| Orth→sem | | | |
| Plural | 61.7 | 16 | 0 |
| Past tense | 0 | 0 | 100 |
| Stem | 1.6 | 0 | 3.3 |

*Note.* In this and subsequent tables, Orth = orthography, phon = phonology, and sem = semantics.

resentations, including both rule-governed forms (such as the above mentioned) and irregular forms such as TOOK and MEN. Moreover, this knowledge generalized to novel forms (Simulation 2), limited only by the intrinsic ambiguity of stimuli such as GOMES, where the final /z/ could indicate a plural (as in HOMES) or not (as in LENS). In the present simulation, the model generated the correct semantics for inflected forms from print, and again generalized in a principled way. These findings address a concern raised by Pinker and Ullman (2003) concerning the capacity of connectionist networks to capture facts about inflectional morphology. Pinker has long argued for a dual-mechanism theory of the past tense, similar to the dual-route model of pronunciation (see, e.g., Pinker, 1991). In both domains there are both rule-governed forms and exceptions, which are thought to involve separate mechanisms (a set of rules, a lexicon) governed by different principles. Pinker has repeatedly argued that generating the past tenses of irregular past tenses such as TOOK or irregular plurals such as MEN requires accessing lexical representations for these words (see, e.g., Pinker, 2000; Pinker & Prince, 1988). We have argued that, like words with regular and irregular pronunciations, words with regular and irregular past inflections are generated by a single processing system (Daugherty & Seidenberg, 1992; Joanisse & Seidenberg, 1999). Pinker and Ullman questioned the adequacy of the Joanisse and Seidenberg (1999) model of the past tense because it happened to use localist representations of the semantics of words, which according to Pinker and Ullman corresponded to lexical entries for individual words. However, the present model shows that the use of nodes corresponding to individual words is not required. The model correctly generates the phonological forms of both regularly inflected words and "exceptions" from semantic input; it also generates correct semantic representations from either orthographic or phonological input (again subject only to limitations imposed by the intrinsic ambiguity of some forms).

## Division of Labor: Summary

We have described a model in which direct-visual and phonologically mediated pathways jointly determine the semantics of words. The relative contributions of the two pathways are influenced by factors including the skill level of the model and lexical properties such as frequency and spelling–sound consistency. In the next two sections we examine the model's performance in processing homophones and pseudohomophones, stimuli that have played an important role in theorizing about the role of phonological information in reading.

# HOMOPHONES

## Disambiguating Word (BEAR) and Nonword (SUTE) Homophones

As noted earlier, spelling and phonology are highly correlated in English because the orthography is alphabetic; in contrast, the correspondences between spelling and meaning are more arbitrary, although as the previous simulation showed, the orth→sem pathway can learn morphological regularities such as number and tense morphology. We have seen how these characteristics of the mappings affect the development of the orth→sem and orth→phon→sem pathways. Homophones present an important test case because orth→phon→sem computation is ambiguous; ROSE and ROWS activate the same phonological code, which is associated with two distinct meanings. In this section we first characterize how homophones are processed in the model and then present simulations of three representative behavioral studies (Jared & Seidenberg, 1991; Lesch & Pollatsek, 1993; Van Orden, 1987).

### *Simulation 13: Homophones*

The division of labor for homophones over the course of training was examined using the lesioning methodology. Effects of the relative frequencies of the alternative senses of the homophones, a factor that previous studies have shown affects performance (Rayner & Duffy, 1986; Simpson, 1994), were also assessed.

### *Method*

There were 497 pairs of homophones in the training set. All homophones whose probability of presentation was at least 1.5 times greater than the other member of its pair were categorized as dominant and the alternative as subordinate. All other pairs were coded as being approximately balanced in frequency. This yielded 324 high-frequency homophones, 324 low-frequency ones, and 346 balanced ones. The division of labor analysis used in previous simulations was repeated using these items. The method of presenting items to the network and lesioning pathways was identical to that used in the preceding simulations.

### *Results and Discussion*

The results for all homophones, collapsed across frequency, are shown in Figure 23. The data reflect the fact that the orth→phon→sem pathway has a limited capacity to read homophones because of their inherent ambiguity, whereas the orth→sem pathway is only limited by the amount of training. Figure 24 presents

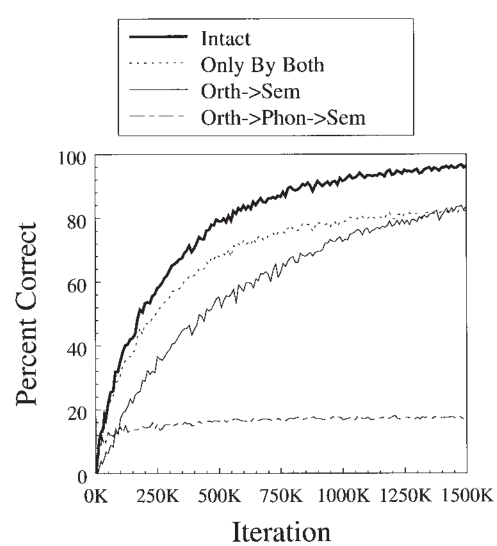


*Figure 23.* Division of labor in computing the semantics of homophones over the course of training. The model learns to produce the correct meanings using information from both pathways. Most can be computed correctly only using input from both pathways; a small and nearly fixed proportion can be read by orth→phon→sem alone (these are dominant, high-frequency meanings).

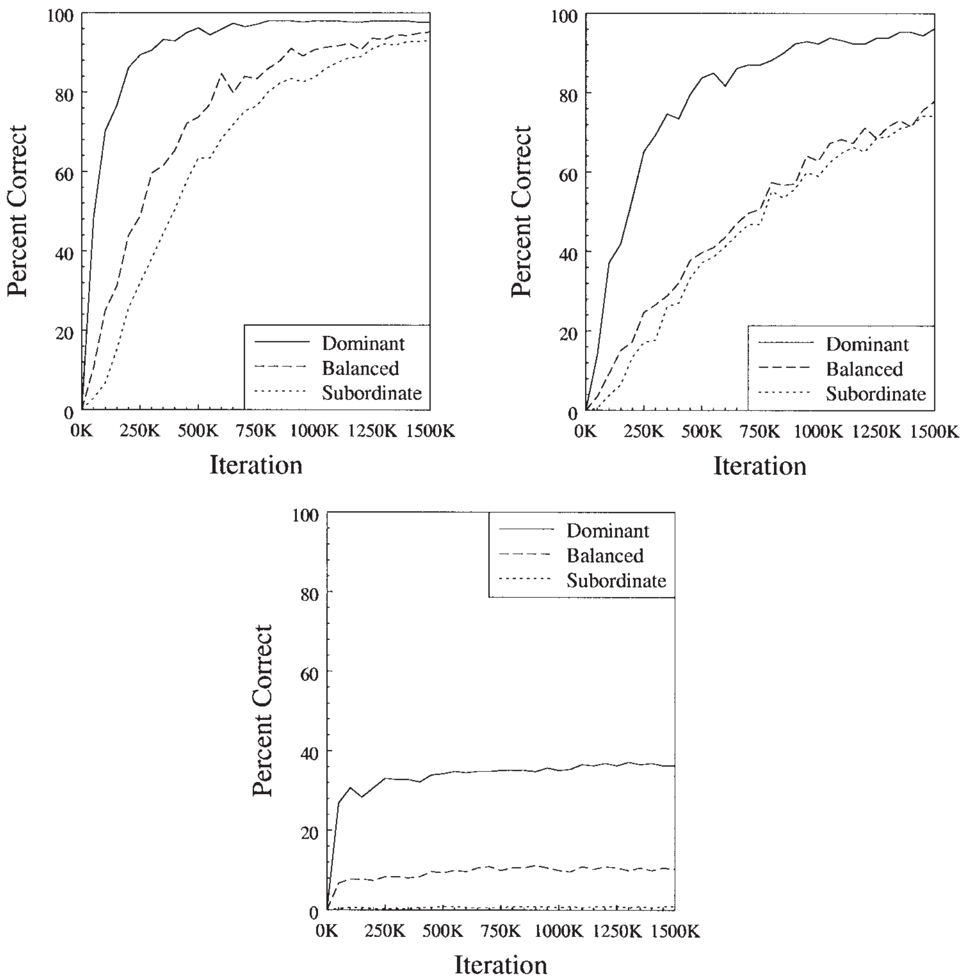


*Figure 24.* Homophone accuracy over the course of training: intact model (top left), by orth→sem (top right), and by orth→phon→sem (bottom).

the effects of relative frequency in the intact model and the two isolated pathways. At asymptote the intact model is able to compute the correct meanings for almost all homophones regardless of relative frequency; the dominant items are acquired first, with the balanced and subordinate homophones learned more slowly. The orth→sem pathway (top right) learns the dominant homophones more slowly than the intact model but asymptotes at nearly the same accuracy level. This pathway performs much less well than the intact model on balanced and subordinate homophones. The orth→phon→sem pathway (bottom) can read some of the dominant homophones, fewer of the balanced ones, and almost none of the subordinates. All of the homophones are inherently ambiguous, but this pathway gets some dominant and balanced items correct because it defaults to one meaning that turns out to be correct.

The data in Figure 24 show that the intact model performs better than either of the isolated pathways for all types of homophones throughout training. This finding provides additional evidence that the two pathways jointly determine meaning in the intact model. The most direct evidence is provided by the balanced and subordinate items, for which the intact model's accuracy is greater than the sum of the accuracies of the two independent pathways. This result is also seen in Table 3, which summarizes the results at the end of training. The dominant items do not show this effect because the model does so well on them; the isolated orth→sem gets most of them correct, and orth→phon→sem also gets more than 30% of them. Thus, for dominant, higher frequency homophones, both pathways contribute because they become tuned to these items (orth→sem more so than orth→phon→sem), whereas for balanced and subordinate homophones, both pathways contribute because they are jointly needed to compute semantics accurately.

At the end of training, all homophones were read more accurately by the semantic path than the phonological one. In fact, essentially none of the homophones could be read only by orth→phon→sem and not by orth→sem. This is not surprising given that the orth→phon→sem pathway is fundamentally ambiguous for homophones. Of interest, the orth→phon→sem pathway was almost totally unable to read the subordinate homophones (e.g., EWES vs. USE); the bulk of the subordinate homophones could be read either by the orth→sem path in isolation or by the two paths together. The orth→phon→sem pathway was much more successful in reading the balanced members and still more suc-

Table 3
*Asymptotic Performance on Homophones: Percentage Correct*

| Model | Dominant | Balanced | Subordinate |
|---|---|---|---|
| | Homophone type | | |
| Intact | 97 | 95 | 93 |
| Orth→sem | 96 | 78 | 74 |
| Orth→phon→sem | 36 | 10 | 1 |

cessful at reading the dominant members. The reason the orth→phon→sem pathway was more successful at reading the dominant member of a homophone pair than the subordinate members is in part because the phon→sem pathway was better at reading such items.

Figure 23 also reveals an interesting developmental effect. The only-by-both condition consists of items that could not be read by either pathway in isolation but could be read by the conjoined efforts of the two pathways. This is of particular interest, in that the orth→sem pathway was not able to read these items by itself but could provide enough information to disambiguate the phonological form of the word. Recall Simulation 1, in which a small amount of semantic context had a dramatic effect on the ability of the network to disambiguate homophonous phonological patterns. The sharp initial rise in the only-by-both condition early in training shows that the orth→sem pathway was not providing enough information to produce the correct semantics by itself but was providing enough to disambiguate many homophones.

For all three types of homophones, this condition reached a peak in the early stages of training and then dropped off as the model continued to develop. The orth→sem pathway became better able to read homophones in isolation as training progressed. The broad implication of this simulation is that the extent to which homophones require input from orth→phon→sem, orth→sem, or both depends on the relative dominance of the homophone and on the overall degree of reading skill.

The semantic feature $d'$ was computed for the three classes of homophones for the three simulation conditions (intact, by phonology, and by semantics) for the fully trained model. For each item to be presented, the semantic representation was recorded and compared with the target representation. Hits, misses, false alarms, and correct rejections were used to compute the value of $d'$. In addition, for each homophone pair, the $d'$ for the generated semantics and the targets for the other member of the pair was also computed. Thus, for example, for the homophone pair EWES–USE, EWES is a subordinate member; when it was presented to the network, the semantic representation it produced was compared to the targets for EWES and USE. These two $d'$ values are shown in Table 4. Of interest, there is some information available to the semantic system in all conditions; the $d'$ is never zero. The reliability and completeness of this information are what vary according to pathway and relative frequency of the homophone. Further, for the subordinate homophones being read by orth→phon→sem, the $d'$ for the opposing member of the homophone pair is higher. This indicates that the presentation of EWES results in more USE-like information being generated along the orth→phon→sem path.

## *Simulation 14: Van Orden (1987)*

The above analyses of homophone processing are consistent with previous analyses based on the entire corpus. For most words, including homophones, semantic patterns are determined by input from both pathways. The question, then, is whether people perform this way as well. Many behavioral studies have been taken as evidence for a different view: that spelling patterns are recoded into a phonological code that is then used to access meanings. Because meanings are accessed via phonology, homophones will activate multiple senses, with the inappropriate ones suppressed by a subsequent procedure that checks the activated meanings against the spelling of the word (Lesch & Pollatsek, 1993; Lukatela & Turvey, 1994a, 1994b; Van Orden, 1987; Van Orden et al., 1988, 1990). The primary evidence derived from studies of homophones and pseudohomophones. We next present simulations of three representative studies of homophones, followed by two simulations of pseudohomophones.

We first consider the influential studies by Van Orden (1987). The basic methodology involved presenting a question such as “Is it a flower?” and then a word that is a homophone of a category exemplar (e.g., ROWS). Homophones were coded as either visually similar (e.g., BEECH–BEACH) or dissimilar (e.g., DOUGH–DOE). The data concerned the false-positive rates in these conditions compared to controls (e.g., ROBS, the control for ROWS).

In Experiment 1, participants were presented with the target item for 500 ms, and then it was replaced by a pattern mask. Participants made significantly more false-positive errors on homophone trials than spelling controls. The error rate for the similarly spelled homophones was higher than for the dissimilarly spelled homophones.

In Experiment 2, stimuli were presented for either a very short or longer duration, then masked. Van Orden (1987) showed the

Table 4
*Semantic Feature $d'$ for Homophones*

| Model | Dominant: Correct | Dominant: Alternative | Balanced: Correct | Balanced: Alternative | Subordinate: Correct | Subordinate: Alternative |
|---|---|---|---|---|---|---|
| | Homophone type | | | | | |
| Intact | Undefined | 1.2 | 7.3 | 2.0 | 7.3 | 1.3 |
| Orth→sem | 6.4 | 1.3 | 5.4 | 1.6 | 5.4 | 1.2 |
| Orth→phon→sem | 2.2 | 1.8 | 2.0 | 2.0 | 1.8 | 2.2 |

*Note.* Undefined = $d'$ is undefined in this condition because there were no misses or false alarms.

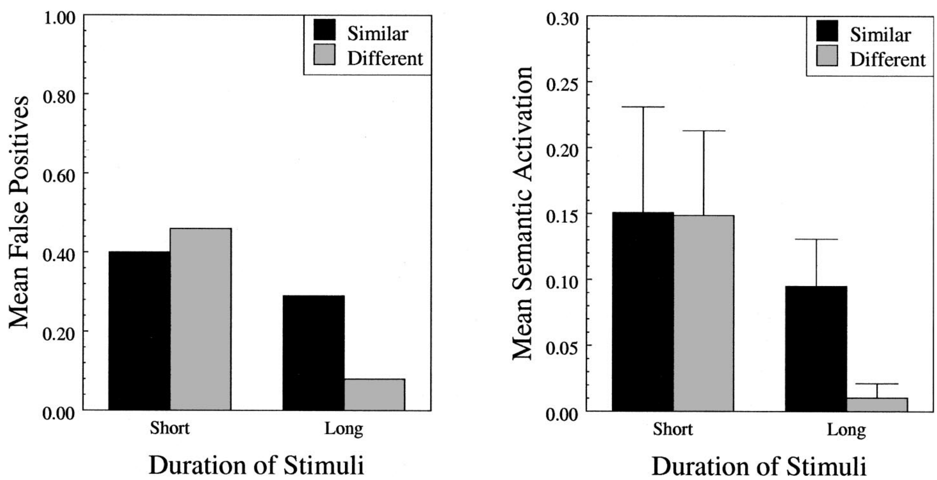


*Figure 25.* Experiments 1 and 2 from Van Orden (1987; left), and simulation results (right). Data from Van Orden show the difference in false positives between foils and controls. Data from the model show the difference in semantic activation between foils and controls. Error bars represent standard errors of the mean.

percentage of false positives for foil items above and beyond those for the control items. The results are summarized in Figure 25 (left). There were more false positives in the short-duration condition than long, relative to controls. In addition, there was no effect of visual similarity in the short condition but a significantly lower error rate for dissimilar homophones in the long condition.

The results were interpreted as indicating that meanings are activated via phonology, with orthographic information subsequently used to disambiguate via a spelling check. Participants would produce false positives only if the homophones had been phonologically recoded, activating incorrect meanings. The effect of presenting the stimuli for a short duration before masking was to remove the information necessary to perform the spelling check, yielding false positives for both visually similar and dissimilar homophones. With longer stimulus durations, only the visually similar items produce a large false-positive effect, the spelling check having successfully disambiguated most of the dissimilar items.

Our model differs from this account insofar as the orth→sem and orth→phon→sem pathways jointly determine the meanings of homophones and other words. Moreover, the implemented model did not include the connections from semantics to orthography that would be required in order to perform the hypothetical spelling check. Hence, we sought to determine whether the model would exhibit the pattern observed in Van Orden's (1987) study.

## Method

The simulation used the items in the Van Orden (1987) experiment (excluding four items: three multisyllabic words that could not be represented in the current model and one item, BORE, that was absent from our training set). An additional four items were added to equalize the number of items per cell with Van Orden's study. Semantic features that correspond broadly to the kinds of semantic questions that Van Orden asked of his participants were identified. For example, for the homophone pair MEAT–MEET, we examined the semantic feature [foodstuff], which would only be on for MEAT; for the pair WEIGHT–WAIT, we examined the semantic feature [physical property], which applies to WEIGHT but not WAIT. Table 5 shows the exemplars, foils, controls, and semantic features used in this experiment.

We presented exemplars, homophones, and foils to the network for short and long durations and examined whether the model activated the critical semantic feature for the homophone distractor (e.g., for MEET, the activation of the semantic feature [foodstuff]). The network was run for 8 units of time in both the short- and long-presentation conditions. For the short condition, the orthographic input was removed after 2 units of time and the network continued to cycle for 6 time units.[15] For the long condition, the orthographic input was removed at 7.33 units of time and the network continued to cycle for 0.67 units of time. The activity of the inappropriate semantic feature ([foodstuff]) was recorded throughout processing.

The activity of the relevant semantic feature was integrated over the course of processing for the foil and controls. Following the method of Van Orden (1987), we measured the extent to which the foil inappropriately activated the relevant semantic features above and beyond the control. Concretely, we measured the integrated semantic activity for the foils and subtracted from that the integrated semantic activity for the controls.

## Results and Discussion

Figure 25 (right) shows the results.[16] In the short-presentation condition, there was no reliable effect of visual similarity. In the long-presentation condition, there was a reliable effect of visual

[15] In the interactive activation model of McClelland and Rumelhart (1981), units corresponding to segments of letters activated localist letter representations, which in turn activated word representations. The weights had been chosen so that when all segments of a letter position were activated, the letter nodes were suppressed. We have not implemented a letter segment representation but assume, following McClelland and Rumelhart, that the effect of a pattern mask is to obliterate activity in letter representations. There is a grain issue insofar as the model sometimes makes more specific predictions than can be observed in behavioral studies (see, e.g., Simulation 16 later in this article).

[16] A note about comparing the modeling data to human data. One difference between modeling data and behavioral data is that the former involve no measurement error. In comparing the two we emphasize the extent to which they exhibit similar, theoretically relevant patterns. In several cases (e.g., see Figure 25), the simulation data appear to be slightly cleaner versions of the results than exhibited by the participants.

similarity, $F(1, 18) = 5.08$, $p < .05$. Thus, the Van Orden (1987) results appear in a model that incorporates very different mechanisms concerning the activation of meaning. The present model has no explicit spelling check mechanism; rather, the correct meanings of homophones are computed on the basis of input from both orth→sem and orth→phon→sem pathways.[17]

To see why these results obtain, consider the data in Figure 26. The data are the sum squared error for both the semantic and phonological representations measured at each time sample of the time course of processing in the short-duration condition. The sum squared error is the square of the difference, over each feature, from its actual output at that moment in time and the target value, summed over each feature. When the orthographic input was removed at time step 2, the error associated with the semantic representation grew much more rapidly than the error associated with phonology. Thus, removing orthographic input (by masking or, in the model, simply turning it off) has different effects on the computation of semantics and phonology: Semantic representations decay much more rapidly than do phonological ones.

This behavior of the model is related to the fact that phonology and semantics are both represented as attractor structures in which activation continues to propagate after the initiating stimulus (the orthographic pattern) is removed. The phonological representations are more dense and intercorrelated than the semantic representations; this intercorrelation allows the phonological attractor to retain and repair partial patterns of activity more efficiently than does the semantic attractor.

These findings have important implications concerning the interpretation of data from masking studies. The use of this procedure was motivated by the assumption that it provided a way to halt processing, yielding a snapshot of what information had been activated to a given point in time. Thus, an effect of homophony on false positives for stimuli masked after 40 ms was interpreted as evidence that phonological information only took this long to be activated (see, e.g., Perfetti & Bell, 1991). In the model, however, although masking eliminates further input from orthography, *it does not halt processing*. What the model computes after the input has been removed depends on the characteristics of the attractor structures, which differ for phonology and semantics. The model suggests that masking interferes with the orth→sem computation much more than orth→phon and, thus, orth→phon→sem. With brief stimulus presentation, sufficient activation does not pass from orthography to semantics to compute the correct meaning. The semantic attractor cannot complete the pattern because of its relative sparseness. Thus, the effect of masking is to eliminate activation from orthography to semantics that normally contributes to homophone disambiguation. There is no effect of visual similarity because input from orthography to semantics has been disabled. The situation with phonology is different. With even a brief stimulus presentation, sufficient activation passes to phonology to permit a stable phonological representation to be computed, resulting in the activation of multiple meanings.

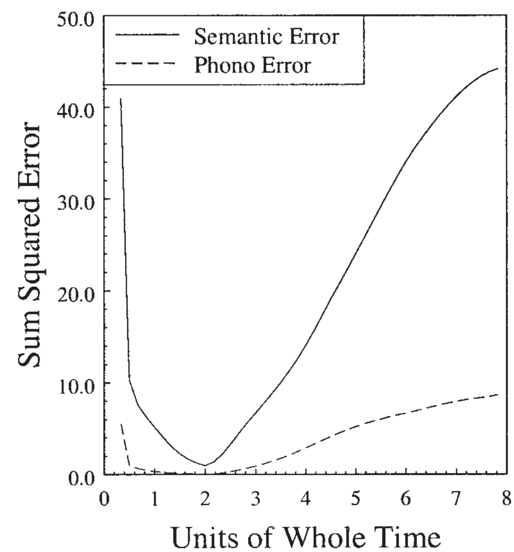


*Figure 26.* Sum squared error of semantic and phonological (Phono) representations when orthographic input is masked at time 2.

Table 5
*Stimuli Used in Simulation 14*

| Exemplar | Foil | Control | Semantic feature |
|---|---|---|---|
| | Similarly spelled items | | |
| Beach | Beech | Bench | Geological formation |
| Creek | Creak | Cheek | Brook |
| Team | Teem | Term | Unit |
| Seam | Seem | Slam | Joint |
| Rein | Rain | Ran | Implement |
| Peak | Peek | Peck | Indefinite quantity |
| Meat | Meet | Melt | Foodstuff |
| Bowl | Boll | Boil | Vessel |
| Arc[a] | Ark | Are | Container |
| Poll[a] | Pole | Pale | Analyze |
| | Less similar items | | |
| Doe | Dough | Doubt | Animal |
| Nose | Knows | Snobs | Organ |
| Suite | Sweet | Sheet | Musical composition |
| Maid | Made | Maim | Life form |
| Nun | None | Noon | Life form |
| Lute | Loot | Lost | Material |
| Rose | Rows | Robs | Rise |
| Weight | Wait | Writ | Physical property |
| Neigh[a] | Nay | Bay | Horse |
| Hawk[a] | Hock | Bock | Has part wing |

[a] Substituted items.

[17] We examined one alternative interpretation of the results: that the effect of visual similarity in the long presentation condition was due to an unintentional frequency bias in the items; that is, that the visually similar items were actually dominant items and hence the results were due to the phon→sem pathway's defaulting to the dominant meanings. To assess this possibility, we ran the simulation again with the orth→sem pathway deleted. There was no effect of visual similarity, and there was much stronger activation of the semantic feature. Hence it is clear that the orth→sem pathway is doing considerable work in disambiguating these items.

In summary, the model behaves quite differently under normal and masked conditions. Masking creates a condition in which the orth→phon→sem pathway assumes primacy. This behavior is different than that which occurs in the unmasked case, in which orth→sem contributes significantly to semantic activation and homophone disambiguation. The implication of these findings concerning the interpretation of masking experiments should be clear: It cannot be assumed that what occurs in the masked condition also occurs when the input is not masked. Thus, the apparent primacy of orth→phon→sem observed in these experiments is in part due to the use of an experimental technique that differentially disrupts processing within orth→sem versus orth→phon→sem. We return to this issue below in connection with simulations of another study using the masking procedure.

### *Simulation 15: Jared and Seidenberg (1991)—Homophones*

We now turn to the study by Jared and Seidenberg (1991) that provided evidence concerning the effects of homophone frequency on false positives. As in Van Orden (1987), participants performed a semantic decision task (e.g., "Is it an object?"), and target items were either exemplars (MEAT), a homophonous foil (MEET), or a spelling control (MEAN). Words were not masked but rather were presented until the participant responded. The homophone foils varied in terms of their frequencies (high vs. low) and the frequencies of the matched exemplar (high vs. low) in a factorial design. The principal data concern the number of false positives in each foil condition compared to those on spelling controls.

Figure 27 shows the net effects (percentage of false positives in a foil condition minus the spelling control condition). The only condition in which presentation of the foil yielded a significant number of false positives was the one in which both the homophone foil and its corresponding homophone exemplar are low in frequency. High-frequency foils and low-frequency foils with high-frequency exemplars did not yield statistically reliable false-positive effects.

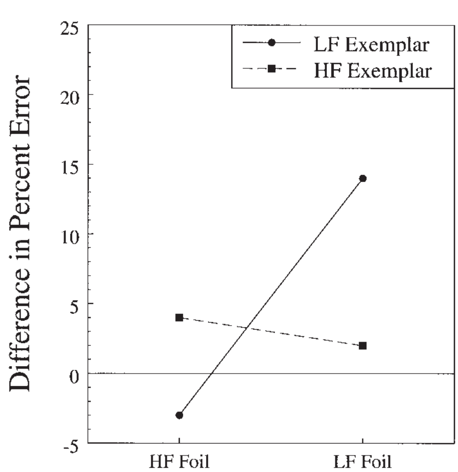


*Figure 27.* The Jared and Seidenberg (1991) homophone results. False positives occurred only when a target was a low-frequency foil and the relevant exemplar was also low in frequency.

These results are a bit puzzling. It is easy to see from the simulations presented previously why low-frequency foils, but not high-frequency ones, would produce false positives. High-frequency items are more likely to benefit from the direct orth→sem route than low-frequency ones; the orth→sem route is not "fooled" by homophony the way the orth→phon→sem route is. As more orthographic information is available to the semantic system, the probability of a false positive for a homophone decreases. What is puzzling is the effect of the exemplar frequency on the tendency of homophone foils to produce false positives. Why should the frequency of MEAT modulate the probability of a false positive for MEET? Jared and Seidenberg (1991) were not able to provide a definitive answer, instead emphasizing the lack of a false-positive effect for high-frequency words, which seemed to contradict the strong position that orth→sem does not influence the initial computation of meaning. If the false-positive effect is taken as evidence for phonologically activated access of meaning, then the absence of the effect in some conditions implied that meaning was not accessed via phonology. We conducted a replication of the Jared and Seidenberg study using the model with the goal of clarifying these effects.

## *Method*

*Stimuli.* Stimuli were selected as follows. All items in the training set were divided into the categories of object, living thing, or other, based on the presence or absence of the semantic features [object] and [life_form]. Items that are objects or living things were candidates for exemplars. Items that are not objects were candidates to be a foil or spelling control for object exemplars. Those that are not living things were candidates to be foils or spelling controls for living thing exemplars.

For each candidate exemplar, we determined whether the item had a corresponding homophone foil. To create spelling controls, we identified an item with the same number of letters as the exemplar, the same initial letter, and a spelling that differed by at most one letter from the exemplar. All foils and exemplars with a probability of presentation of .21 or greater were coded as high frequency; those with a probability of .05 or less were coded as low frequency. Table 6 shows a sample set of items; a total of 397 foils and matched controls resulted.

*Procedure.* Jared and Seidenberg's (1991) procedure was simulated by presenting the foils and spelling controls to the intact model and observing the activation on the semantic feature for the exemplar. For example, if CAUGHT was presented to the model, the [object] feature would be monitored, because the exemplar (COT) is an object. Activity for the inappropriate semantic feature for the foil was recorded, as was the activity of that feature for the spelling control. These values were integrated in the same fashion as in the previous simulation. Following Jared and Seidenberg, we plotted the difference between the false positives for the foil and the control.

## *Results and Discussion*

The results are shown in Figure 28. The data broadly match those of Jared and Seidenberg (1991). The items that yielded the strongest activation of the critical exemplar feature were the ones for which both the exemplar and the foil were low in frequency. As in Jared and Seidenberg, the only condition that produced a difference between foil and control that was reliably greater than zero was the low-frequency exemplar, low-frequency foil condition, $t(98) = 2.18$, $p < .05$. The inhibition in the high-frequency foil,

Table 6
*Sample Stimuli for Jared and Seidenberg (1991) Replication*

| Exemplar | Exemplar frequency | Foil | Foil frequency | Spelling control |
|---|---|---|---|---|
| Ales | LF | Ails | LF | Aids |
| Cot | LF | Caught | HF | Taught |
| Road | HF | Rode | LF | Bode |
| Son | HF | Sun | HF | Bun |

*Note.* LF = lower frequency; HF = higher frequency.

high-frequency exemplar condition approached significance, $t(96) = 1.79$, $.05 < p < .10$; this was the only condition that produced a numerical inhibition effect in the Jared and Seidenberg study, although it too was not significant.

Why do these effects obtain? The earlier analysis of frequency effects demonstrated that high-frequency items are better able to be read via orth→sem than are low-frequency ones, so the finding that high-frequency foils do not result in false positives is simple to explain. However, it is less clear why low-frequency foils of high-frequency exemplars do not also show false positives. The participant, and the model, does not see the exemplar in the trial; hence, why should its frequency matter?

To analyze the time course of processing words from the four conditions, we took four illustrative foil items from the Jared and Seidenberg (1991) homophone simulation (one for each condition). The exemplar–homophone pairs were TUX–TUCKS (low-frequency exemplar, low-frequency foil), LOAD–LODE (high-frequency exemplar, low-frequency foil), LYE–LIE (low-frequency exemplar, high-frequency foil), and SON–SUN (high-frequency exemplar, high-frequency foil). We plotted the aggregate input to the inappropriate semantic feature (either [object] or [life form]) over the course of presentation of the foil, breaking the input out into the contribution from orth→sem and phon→sem as in Figure 12. The results for the four conditions are shown in Figure 29.

1. *High-frequency foil, high-frequency exemplar* (see Figure 29a). Jared and Seidenberg (1991) noted an absence in their study of false positives in conditions in which the exemplar was high in frequency. They took this to be evidence that orth→phon→sem was not strongly activating the inappropriate semantic feature. However, consistent with the claims of Van Orden and colleagues (Van Orden, 1987; Van Orden et al., 1988, 1990), the model's orth→phon→sem pathway did indeed activate the exemplar's semantic feature. This is consistent with the behavior of the model shown in Table 4, in which the orth→phon→sem pathway produced a $d'$ of 2.0 for both members of balanced homophone pairs, indicating some weak activation of both meanings of the words. However, contrary to claims of Van Orden and colleagues, the orth→sem pathway was able to suppress the activation of the inappropriate semantic feature, resulting in no reliable false positives in this condition.

In this condition, the model's orth→sem pathway learned to suppress this inappropriate activation from orth→phon→sem for two reasons. The first is that the training in the model was error driven; when one pathway produced incorrect activation, the other pathway was pressured to overcome that error. Hence, orth→sem was repairing the error produced by orth→phon→sem. The second reason is that in this condition, the foil itself was high in frequency. Recall from Simulation 10 that the orth→sem pathway was very sensitive to the item's frequency because of the speed pressure to which the model was subjected. Thus, the orth→sem pathway was particularly good at recognizing high-frequency foils and, hence, suppressing inappropriate semantic features.

2. *Low-frequency foil, high-frequency exemplar* (see Figure 29b). In this condition, the orth→phon→sem pathway was also activating the inappropriate semantic feature, more strongly than in Figure 29a. This is consistent with the data from Table 4, in which the orth→phon→sem pathway produced the semantics of a dominant homophone (the exemplar, in this case) much more so than the subordinate homophone (here, the foil). As above, when the foil was presented, the orth→sem pathway had to extinguish this inappropriate activation, and hence a strong negative input to the inappropriate semantic feature developed. This resulted in no reliable false positives in this condition.

3. *High-frequency foil, low-frequency exemplar* (see Figure 29c). Here, both pathways were inhibiting the inappropriate semantic feature. The orth→phon→sem pathway did so because the semantics of the dominant homophone (here, the foil) were activated, and the semantics of the subordinate homophone were suppressed. Thus, there was very little error produced by orth→phon→sem for the orth→sem pathway to correct. However, the foil was high in frequency, and consistent with the results of Simulation 10, orth→sem developed the ability to quickly recognize the item and, hence, suppress inappropriate semantic information.

4. *Low-frequency foil, low-frequency exemplar* (see Figure 29d). This was the condition that produced reliable false positives, both in the empirical study by Jared and Seidenberg (1991) and in this simulation. Here, the homophones are balanced and low in frequency; therefore, the orth→phon→sem pathway produces rather ambivalent activation of the exemplar's semantic feature, particu-

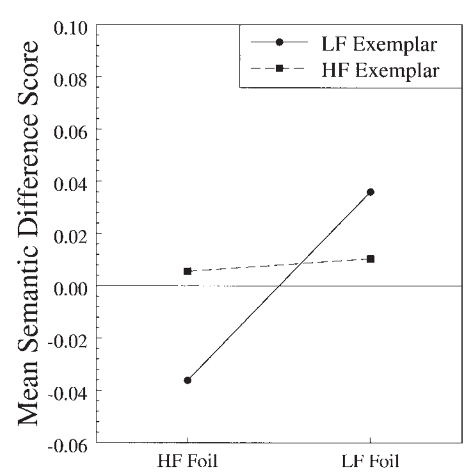


*Figure 28.* Simulation of the Jared and Seidenberg (1991) homophone results. As in Figure 27, only low-frequency foils of low-frequency exemplars yielded false positives.

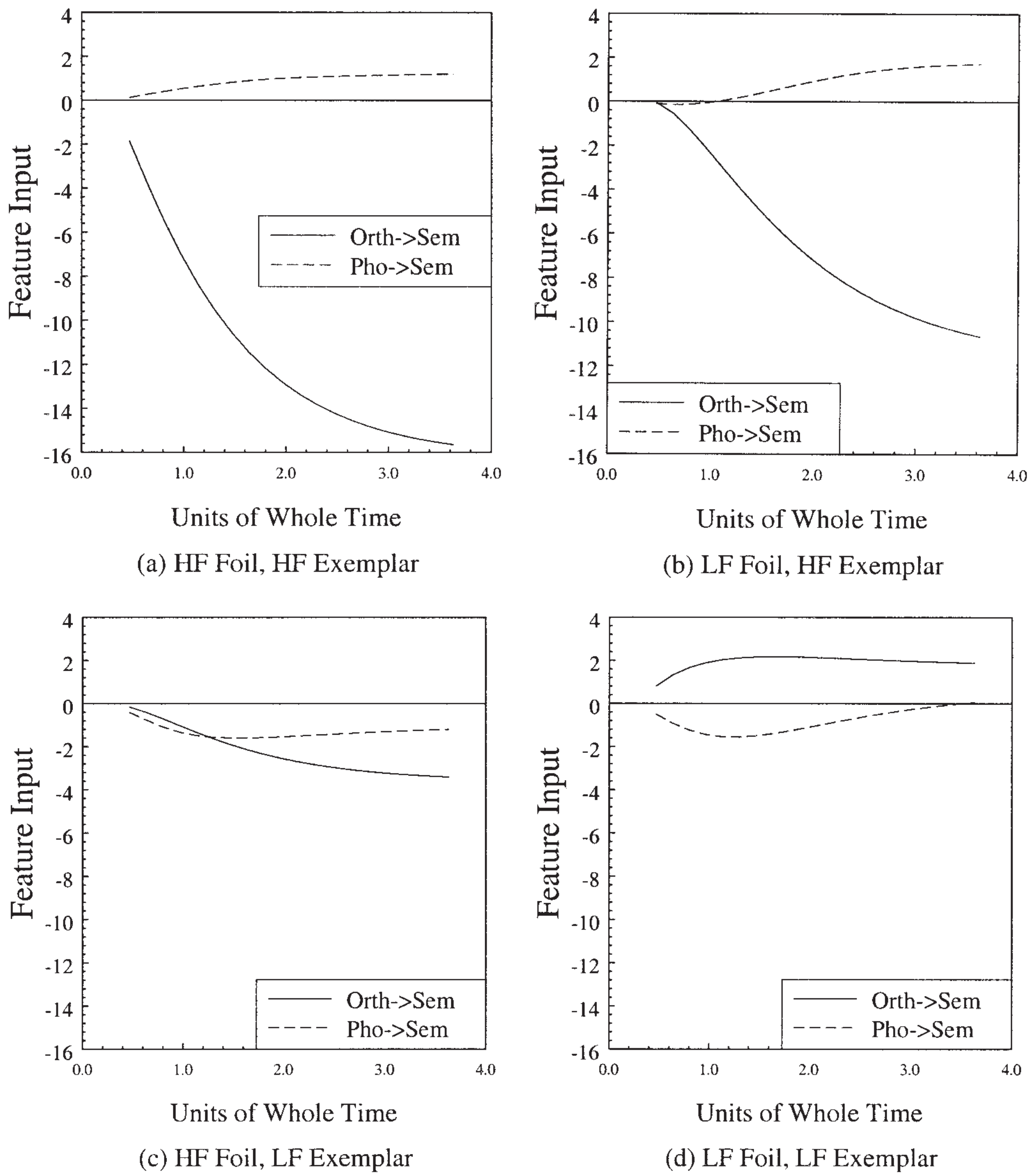


*Figure 29.* Input to distractor semantic feature for four foil conditions. a: High-frequency foil, high-frequency exemplar. b: Low-frequency foil, high-frequency exemplar. c: High-frequency foil, low-frequency exemplar. d: Low-frequency foil, low-frequency exemplar. Pho = phonology.

larly at the end of processing.[18] When the foil was processed by the orth→sem pathway, it did not have to suppress strong erroneous responses generated by orth→phon→sem as in cases in which the exemplar was high in frequency. The foil itself was also low in frequency, and hence the ability of the orth→sem pathway to process it was limited relative to high-frequency foils. Hence, spurious false positives resulted.

The results of this simulation provide a reconciliation of the views of Van Orden and colleagues (Van Orden, 1987; Van Orden et al., 1990) and Jared and Seidenberg (1991). Consistent with Van Orden et al.'s (1990) interpretation (and contrary to Jared & Seidenberg, 1991), the orth→phon→sem pathway produces some semantic activation for high-frequency homophones. However, consistent with Jared and Seidenberg (and contrary to Van Orden et al., 1990), high frequency foils suppress inappropriate activation of their paired homophone via the orth→sem route in parallel with the processing of the orth→phon→sem pathway rather than as a result of a postlexical spelling check operation. This novel account of the Jared and Seidenberg study arises from core computational

[18] Recall that Figure 29 shows the input to semantic units. The activation function used in this model will produce a positive output (0.5) when given an input of zero. Hence, some weak positive activation results from the orth→phon→sem pathway in this condition.

principles of the model: (a) cooperative computation to reduce error and (b) the pressure for the model to respond rapidly.

### *Simulation 16: Lesch and Pollatsek (1993)*

Important additional evidence concerning the role of phonology in word reading has been obtained from studies using a different methodology, semantic priming. Lesch and Pollatsek (1993) created triplets of words consisting of an exemplar such as TOAD, a homophone such as TOWED, and a target that is semantically related to the exemplar such as FROG. Participants were presented with a prime that was either the exemplar, the homophone, or an unrelated control and then with the target, which was named aloud, with naming latency as the dependent measure. The study used two presentation conditions: short (prime presented for 50 ms then pattern masked for 200 ms) and long (prime presented for 200 ms then masked for 50 ms). The critical question was whether homophones such as TOWED would prime targets such as FROG. The data are summarized in Figure 30. In the short condition, both exemplars (such as TOAD) and homophones (such as TOWED) yielded significant priming (e.g., target FROG) compared to the unrelated prime condition. In the long-prime-duration condition, only the exemplar produced significant priming. These results closely resemble earlier findings in the lexical ambiguity literature (e.g., Swinney, 1979; Tanenhaus, Leiman, & Seidenberg, 1979).

These data were consistent with the Van Orden et al. (1990) account. On this view, the visual form of a word is phonologically recoded, and this phonological code activates an associated semantic representation or representations. Thus both TOAD and TOWED activate the meaning related to FROG. The short prime presentation prevents the spelling check from occurring; hence, both meanings are active when the target FROG is presented, yielding facilitation. With longer prime presentation, the spelling check proceeds, and the inappropriate meaning is suppressed and is no longer available to facilitate the processing of FROG.

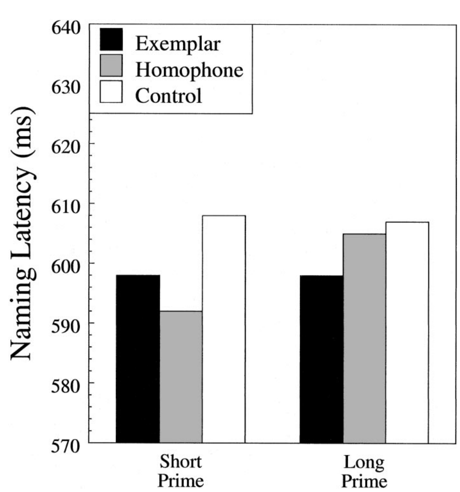


*Figure 30.* Data are from Lesch and Pollatsek (1993). Exemplar = homophone prime related to target (e.g., TOAD–FROG); Homophone = homophone prime unrelated to target (e.g., TOWED–FROG). With brief masked presentation of the prime, both exemplar and homophone primes produced significant facilitation compared to unrelated controls. At the longer prime duration, only the exemplar produced significant priming.

Given the findings concerning the effect of masking in Simulation 14, we repeated the Lesch and Pollatsek (1993) study using the model. The prediction was that the model would replicate their results even though it does not incorporate the spelling check procedure because they reflect the effects of masking on the input from orthography to semantics.

#### *Method*

*Stimuli.* A list of homophonic word pairs was created algorithmically by scanning the training corpus for words with different spellings but identical phonological representations. A second list of semantic associates was created by scanning the semantic representations of all uninflected words and finding all pairs in which the semantic representations differed by no more than one feature. From these two lists we found a set of triplets consisting of an exemplar, a homophone, and a target semantically related to the exemplar. A control item was selected for each triplet that differed from the homophone by at most two letters. Both the homophone and the control item had to differ from the target by at least eight semantic features. A further constraint was imposed such that for approximately half of the items, the exemplar had to be higher in probability of presentation to the model than the balanced homophone by at least a factor of two; for the other half, the homophone had to dominate by at least a factor of two. The homophones in both sets were matched on their overall mean semantic difference from the target. A set of 53 quadruples resulted, consisting of an exemplar, a homophone, a control, and a target (e.g., CREEK, CREAK, BLEAK, STREAM). There were 28 biasing the exemplar and 25 biasing the homophone.

*Procedure.* To simulate the short priming condition, primes were presented for 2 units of time. Then the model was allowed to continue processing for an additional 5 units of time. In the long condition, the prime was presented for 5 units of time and the model continued processing for an additional 2 units of time. Over the course of processing, the semantic and phonological error for all primes was recorded, as was their semantic distance from the target item. At the end of 7 units of time the state of the semantic units was also recorded. We assumed that the amount of priming would be a function of the amount of semantic overlap between the prime and target as shown in previous studies by McRae et al. (1997) and Plaut and Booth (2000).

#### *Results*

Figure 31 shows the semantic distance at 7 units of time as a function of prime type and duration. The results replicated Lesch and Pollatsek's (1993) finding that both the exemplar and the homophone produced priming at the short duration compared to an unrelated control; in the long-duration condition there was strong priming for only the exemplar. This pattern is reflected in a significant interaction between prime type and duration, $F(2, 312) = 10.9$, $p < .001$. There was a small residual priming effect for the homophone in the long-duration condition in the simulation, an effect size that would be difficult to detect in a behavioral experiment.

#### *Discussion*

To understand why these effects obtain, consider the data in Figure 32, which are the sum squared error of the model's phonological and semantic representation over the course of prime presentation for the short condition. As was shown in the simulation of the Van Orden (1987) data, phonology is much more resilient to the effect of the mask than is semantics. When the

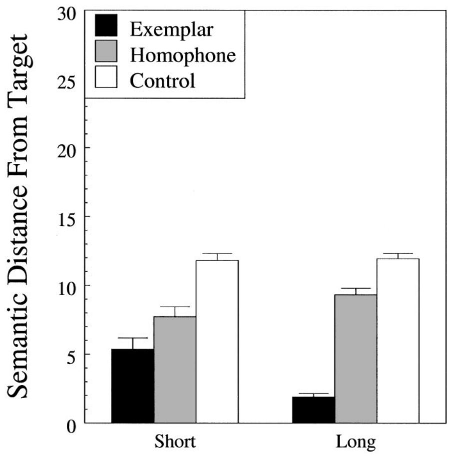


*Figure 31.* Simulation of Lesch and Pollatsek (1993). As in Figure 30, both exemplar and homophone primes produced significant priming at the short delay, whereas at the longer delay the exemplar produced larger facilitation than the homophone. Error bars represent standard errors of the mean.

visual input is masked, phonology tends to remain at a low level of error, whereas the semantic representation drifts from that associated with the orthographic form.

Figure 33 shows the average distance of the model's semantic representation from the target over time, for all three prime types. For the short prime condition, the representation of the exemplar and the homophone begin to converge. At the point at which the target would be presented, both are much closer, in semantic space, to the target than the controls. For the long condition, the visual stimuli drives the exemplar close to the target and the homophone and control away (and toward their own semantic representation). When the visual stimulus is removed, the homophone (but not the control) begins to be influenced by the phonological but not the visual information, and it drifts toward that of the target. However, the interstimulus interval (ISI) is shorter in the long condition, and thus it does not have as much time to move nearer to the target. Thus the main effect found by Lesch and Pollatsek (1993) occurs: Homophones prime much more effectively at short presentation durations and long ISI than the reverse. This effect in the model is not due to an initial activation of phonology and a subsequent spelling check but rather reflects the differential effect of the mask on semantic and phonological information.

### *Relative Frequencies of Homophones*

As described earlier, the computation of meaning for homophones along the phon→sem pathway is sensitive to the relative frequencies of the homophones. Such results are consistent with other studies manipulating this factor. The phon→sem pathway will activate the semantics of a dominant homophone most strongly, a subordinate one most weakly, and a balanced one to an intermediate degree. The model therefore predicts an effect of dominance on the degree of homophone priming with short stimulus presentation. For example, if an exemplar–homophone–target triple consisted of a strongly dominant exemplar (e.g., USE–EWES→MAKE), we would expect the auditory form of the subordinate homophone EWES to strongly activate semantics for MAKE–USE, and hence considerable priming would occur. Similarly, for a triple in which the homophone was strongly dominant (e.g., EWES–USE→SHEEP), we would not expect the homophone USE to activate SHEEP semantics very strongly, and hence much less (though perhaps more than zero) priming should occur. We reanalyzed the simulation output by grouping the stimuli into the two sets: *supportive* trials, in which the exemplar is the dominant member of the homophone pair (and thus the auditory form of both the exemplar and the homophone support the meaning of the target, e.g., USE–EWES priming MAKE), and *unsupportive,* in which the homophone is the dominant member and thus the auditory form of the exemplar and the homophone do not support the meaning of the target (e.g., EWES–USE priming SHEEP).

Figure 34 shows the results from Figure 33, broken down by supportiveness. When the stimulus is present, the semantic representation for the supportive homophone moves toward that of the target more rapidly than the unsupportive one. Similarly, when the stimulus is removed, the semantics for the unsupportive exemplar moves away from the target more rapidly than the supportive case. Crucially, even the unsupportive homophones are closer to the target than the matched controls when the visual input is removed, even though they are equidistant when the visual pattern is present. As predicted, there was a reliable effect of supportiveness on semantic distance from the target at Time 7 in the short prime condition, $F(1, 153) = 5.2$, $p < .03$, and the long prime condition, $F(1, 153) = 17.9$, $p < .001$.

Although these results are consistent with those of other studies manipulating the relative frequencies of homophones, there are two prominent failures to observe effects of relative frequency in the homophone processing literature: Lesch and Pollatsek (1993) and Lukatela and Turvey (1994a).

Lesch and Pollatsek (1993) did not manipulate the relative frequencies of homophone pairs but reported a post hoc test. The stimuli were divided into two sublists: one in which the exemplar was higher in frequency than its paired homophone (what we term

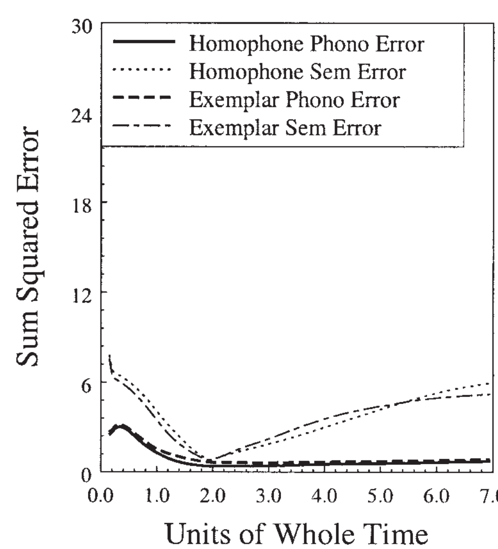


*Figure 32.* Semantic (Sem) and phonological (Phono) error for homophones and exemplars in the short prime condition.

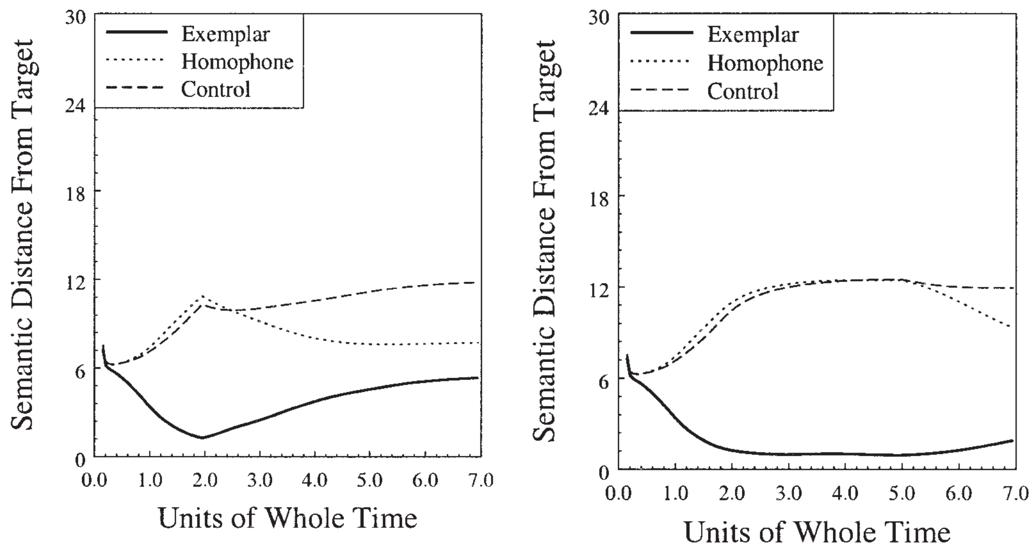


*Figure 33.* Semantic distance from target item as a function of prime condition and prime duration. For brief primes (input masked at time 2.0; left), homophones become drawn toward the semantic representation of the exemplar. At longer durations and shorter interstimulus intervals (mask at time 6.0; right), there is less time for the semantic representation of the homophone to become influenced by the sound pattern.

a *supportive* condition), and one in which the homophone was higher in frequency (*unsupportive*). They did not find a reliable effect of sublist (supportiveness) on priming effects. They therefore concluded that both high- and low-frequency homophones are processed via phonology in contrast to the Jared and Seidenberg (1991) results. The differing results appear to be related to differences in the size of the frequency manipulations in the two studies.

Inspection of the individual items from Lesch and Pollatsek (1993) indicated that the median difference between their high- and low-frequency matched items was 29. There were 8 paired items out of 32 for which the frequency difference was equal to or less than 10. These numbers should be considered in light of the known insensitivity of the Kučera and Francis (1967) norms at the lower end of the frequency distribution (Gernsbacher, 1984).

In contrast, the Jared and Seidenberg (1991) high- and low-frequency paired items had a median frequency difference of 50, and no items had a difference less than or equal to 10. Thus, the difference between conditions was larger and more consistent across items. It is not surprising that the frequency manipulation was stronger in the Jared and Seidenberg study; it was built into the design of the study rather than tested post hoc. In short, the Lesch and Pollatsek (1993) materials exhibited smaller, less con-

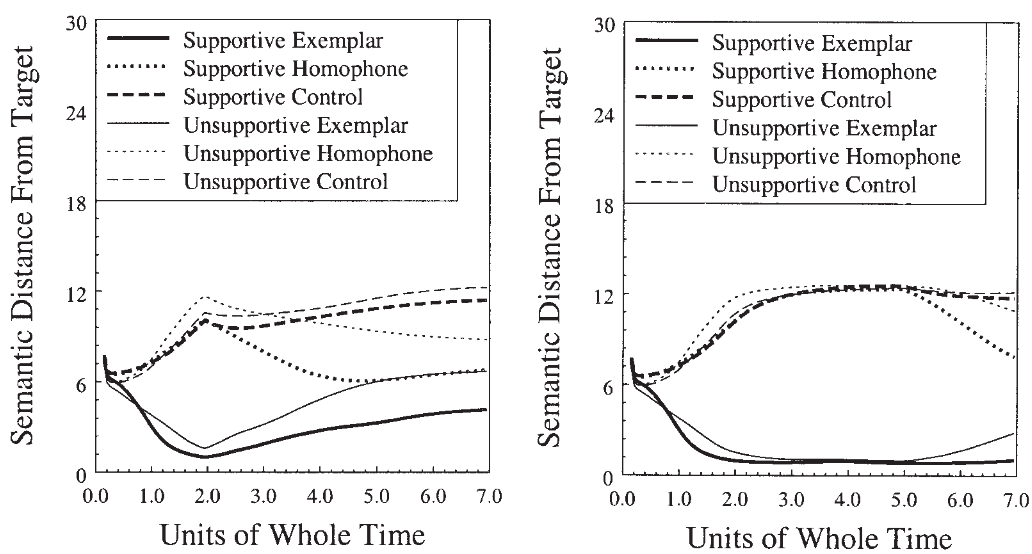


*Figure 34.* Closeness in semantic space to the target for prime types over the course of processing by supportiveness (short prime duration, left; long prime duration, right).

sistent differences between high- and low-frequency items, and the higher frequency words were relatively low in frequency, in the range in which the Kučera and Francis (1967) norms are less reliable. The failure to obtain a frequency effect in this study compared to Jared and Seidenberg's seems likely to be related to these properties of the stimuli.

Lukatela and Turvey (1994a) presented several additional studies using the priming methodology. Like Lesch and Pollatsek (1993), they found priming by exemplar and homophone at short durations, but not at long ones. They also manipulated what we have termed *supportiveness* and found no effect, leading to the conclusion that access to meaning is initially phonological regardless of homophone frequency. There are three problems with these studies that cloud the interpretation of the results. First, the Lukatela and Turvey (1994a) data are somewhat ambiguous, because the pattern of results differs depending on which of two control conditions is used to assess the magnitude of priming effects.[19] Second, there is a problem with the stimuli in the unrelated control condition used as a baseline. The words in these conditions contained much more unusual spelling patterns than those in the other conditions, which may have had the effect of producing larger estimates of the priming effects.[20] Finally, as in the Lesch and Pollatsek study, the manipulation of relative frequency was weak. The mean frequency of supportive primes was 145, whereas the unsupportive primes had a mean of 15. This seems like a strong frequency manipulation, but the difference is due to a few very high frequency outliers. Considering the median values, the supportive primes had a median frequency of 44, compared with a median frequency of 5 for the unsupportive items—a much smaller frequency differential than obtained by comparing the means. The median frequency difference between paired supportive and unsupportive items was only 28.5, and there were 23 pairs (out of 84) for which the frequency differential was less than or equal to 10. Given the small size of the priming effect, it should not be surprising that a weak frequency manipulation yielded a null effect of supportiveness. Given these methodological issues and the similarity of the results to those of Lesch and Pollatsek, we did not attempt to simulate the Lukatela and Turvey experiments.

### Homophone Reading: Summary

The model accurately computes the meanings of homophones using input from both pathways. The orth→phon→sem pathway learns quickly, but its role is limited by homophones' intrinsic ambiguity, which can be resolved using input from orth→sem. The conjunction of information from the two primary sources provides a highly effective way to achieve disambiguation. Thus, the orth→sem pathway begins to assume some of the processing burden in response to both the demand for speed and the need to disambiguate homophones. The simulations demonstrate that disambiguation does not require a spelling check that occurs after meanings have been activated via phonology. The orth→sem and orth→phon→sem components of the model are constructed out of the same elements and governed by the same principles. Under these conditions, the orth→sem pathway is observed to develop the capacity to contribute significantly to the processing of homophones and other words. The simulations of behavioral studies are consistent with the conclusion that people process these words in a similar manner.

The other major finding from the simulations concerned the effects of masking on the course of lexical processing. The simulations suggest that masking has different effects on processing within the orth→sem and orth→phon→sem pathways; it mainly eliminates normal input from orth→sem. Under the masked condition, participants can only process homophones via orth→phon→sem, yielding a significant number of false-positive responses on the semantic decision task. However, it does not follow from this demonstration that orth→sem also makes no contribution with normal stimulus presentation.

Finally, the simulation of the Jared and Seidenberg (1991) study provided further evidence concerning the dependencies between the two main pathways in generating behavior. The simulation captured the main features of the human data, but the processes that gave rise to these effects were different than either Van Orden et al. (1988, 1990) or Jared and Seidenberg had surmised. We found these results surprising but also sobering insofar as they suggest that behavioral data can be consistent with unanticipated underlying mechanisms that are only recognized by using a computational model.

## PSEUDOHOMOPHONES

### Do Pseudohomophones Activate Meaning?

The final phenomena to be addressed concern the processing of pseudohomophones such as SUTE. These stimuli have been widely studied because of the leverage they provide with respect to diagnosing the use of phonological information in reading. Pseudohomophones are novel stimuli that happen to sound like actual words. A participant will not have encountered such stimuli before; hence, he or she will not have formed any associations between their spellings and specific meanings. A false-positive response on a trial such as "Is it an article of clothing?: SUTE" would result if the participant phonologically recoded the stimulus,

[19] In the crucial experiments from Lukatela and Turvey (1994a, Experiments 5 and 6), a prime word was presented at a short or long duration, followed by a target. The conditions included trials in which the prime was a semantically related exemplar or a homophone prime (e.g., TOAD and TOWED, respectively, for the target FROG). There were also two control conditions: a visual control condition (TOLD—FROG) and an unrelated condition (PLASM–TOAD). The results of the study are unclear because they differ depending on whether the visual control or the unrelated control is taken as the baseline for calculating net priming effects. When the unrelated control is used, the results are similar to Lesch and Pollatsek's (1993): priming for both TOAD and TOWED at the short duration but only for TOAD at the longer duration. When the visual control is used as baseline, priming effects in the long duration condition are similar across all conditions, ranging from 3 to 6 ms.

[20] We calculated the mean bigram frequencies for the stimuli from an electronic version of the Carroll, Davies, and Richman (1971) corpus provided by J. B. Carroll. The mean summed bigram frequencies of the exemplar, homophone, and spelling control primes were higher than for the unrelated controls (33,368 vs. 24,838, respectively) $F(1, 502) = 19.40$, $p < .001$. These differences were also observed when comparing each condition to its matched unrelated control: exemplars (33,693) versus unrelated (25,549); homophones (34,855) versus unrelated (25,253); and spelling controls (31,556) versus unrelated (23,713). All of these differences are statistically reliable, $F(1, 166) \geq 6.00$, $p < .05$.

which activated the meaning associated with SUIT. The fact that the participant does not know in advance whether the target is a word or a pseudohomophone implies that phonological recoding occurs in reading words as well.

Our model is consistent with the observation that pseudohomophones can activate semantics via phonology; in general, an orthographic pattern such as SUTE activates a phonological code that is very similar to that produced by SUIT, which in turn activates SUIT-semantics, providing the basis for a false positive. However, the model provides additional information that raises questions about pseudohomophone processing and its relation to normal reading. The standard view that false positives for pseudohomophones are due to phonologically mediated activation of semantics assumes that they cannot activate meaning directly from orthography. This assumption is worth examining more closely. Some pseudohomophones overlap considerably with the words from which they are derived, for example BOXX or GHOAST. As we have seen, in the model many familiar words activate semantic information directly from orthography. Although participants will not have learned to associate a meaning with a novel pattern such as BOXX, it may overlap sufficiently with BOX to produce significant semantic activation. If this is correct, false positives for such stimuli would not necessarily implicate phonological recoding. Simulation 17 addresses this possibility.

A related issue concerns how participants correctly reject pseudohomophones on most trials. It has been assumed that deciding that SUTE is not an article of clothing requires a spelling check—assessing the semantic pattern computed via orth→phon→sem against the input orthographic pattern (Van Orden et al., 1988). This process was also assumed to apply to homophones such as BEAR. As we have seen, homophones can be disambiguated via orth→sem in the model, suggesting that the spelling check is not required. Some pseudohomophones (e.g., ones like BOXX) may also activate semantics from orthography, but unlike homophones, this would only increase the likelihood of a false-positive response. One possibility is that, unlike homophones, pseudohomophones do require a spelling check. There may be other bases for making this decision, however. For example, pseudohomophones could differ systematically from words in terms of the quality of the phonological or semantic codes they activate. Like lexical decision (deciding if a stimulus is a word or not), semantic decision (deciding if a stimulus is a member of a designated category) is a judgment task in which participants must establish reliable criteria for making accurate responses (Balota & Chumbley, 1984; Seidenberg, Waters, Sanders, & Langer, 1984). Simulation 18 examined how pseudohomophones are processed in the model in order to address these possibilities.

## *Simulation 17: Reading of Pseudohomophones by Orth→Sem*

This simulation addressed whether pseudohomophones activate semantics via the orth→sem pathway. The distribution of words in the space of possible orthographic patterns is nonrandom: For example, there are dense clusters of words (e.g., ones containing -AT) and there are words that have no close neighbors (so-called strange or hermit words such as YACHT), as well as intermediate cases. The "receptive fields" of units in the orth→sem pathway will vary in response to these distributional facts. Words such as CAT have so many close neighbors that the weights must be narrowly tuned to that particular word or errors will result. In contrast, a word like GHOST has few neighbors, and so the network can have a broader attractor for that word without generating errors. This analysis predicts that two factors should jointly influence what the orth→sem pathway activates for a pseudohomophone: (a) the similarity of the pseudohomophone to the base word and (b) the neighborhoods of the base word and pseudohomophone. For example, the pseudohomophone KAT is unlikely to produce semantic activation for CAT via orth→sem because both the word and the pseudohomophone are from very dense orthographic neighborhoods; if the units that detect CAT were insensitive to the first letter, for example, they would draw false positives from HAT, RAT, MAT, and so on. Pseudohomophones such as GHOAST, however, may activate GHOST-like semantics; GHOST has few neighbors, and so the correct semantics may be activated even with partial information about the input. In effect, the receptive field for GHOST may include a pseudohomophone such as GHOAST, whereas the receptive field for CAT does not include KAT. The prediction, then, is that the ability of the orth→sem pathway to activate semantics for pseudohomophones will be jointly determined by neighborhood density and closeness to the base word.

### *Method*

A set of word–pseudohomophone pairs was generated by algorithmically identifying onsets and rimes that have multiple possible spellings and then creating pseudohomophones that have the same pronunciation as a corresponding word. These items were split along three dimensions: visual similarity of the pseudohomophone and corresponding word, word neighborhood density, and pseudohomophone neighborhood density. Words were considered visually similar to their pseudohomophone if they differed by one letter, and dissimilar otherwise. Neighborhood density was assessed using the Coltheart $N$ (Coltheart, Davelaar, Jonasson, & Besner, 1977) measure (which equals the number of words that can be derived from a letter string by changing one letter at a time). Dense neighborhoods were defined as $N \geq 10$, and sparse as $N \leq 1$. Eight hundred eighty-nine pairs were generated. Table 7 shows a sample of typical items in the eight conditions with their paired homophonous word.

The orth→phon pathway in the trained network was disconnected in order to examine the capacity of the orth→sem pathway to activate semantics. Pseudohomophones were presented to the network in the standard way, and for each trial the resulting semantic features were recorded and compared to the targets for the paired word. For example, for the pseudohomophone TOSE, the semantic targets for the homophonous word

Table 7
*Sample Items Used in Simulation 17*

| Neighborhood | Word similarity: High | Word similarity: Low |
|---|---|---|
| Dense word | | |
| Dense pseudohomophone | PASE–PACE | TOSE–TOES |
| Sparse pseudohomophone | WROOT–ROOT | DAWL–DOLL |
| Sparse word | | |
| Dense pseudohomophone | NAT–GNAT | NOX–KNOCKS |
| Sparse pseudohomophone | TWEAD–TWEED | URLS–EARLS |

*Note.* The first member in each pair is the pseudohomophone; the second member is the corresponding word.

TOES were compared with the semantic output for the input pseudohomophone TOSE. As before, we considered a semantic feature to be on if its value was above 0.5 and to be off otherwise. The $d'$ was then computed based on the hits, misses, false alarms, and correct rejections with respect to the activated semantic features compared with the veridical semantic representation of the target word. For example, if the word TOES contained the semantic features [digit, extremity, body-part, foot, entity], and the pseudohomophone TOSE activated [digit, extremity, entity] and also [animal], then there would be three hits, two misses, one false alarm (from [animal]), and correct rejections for all other semantic features.

### *Results and Discussion*

Figure 35 shows the results. In the analysis of variance, there were main effects of visual similarity of the pseudohomophone to the base word, $F(1, 887) = 56.0$, $p < .001$; word neighborhood density, $F(1, 887) = 6.2$, $p < .02$; and pseudohomophone neighborhood density, $F(1, 887) = 10.7$, $p < .001$. The three-way interaction of these factors was also significant, $F(1, 881) = 5.85$, $p < .02$. Pseudohomophones that were visually dissimilar to their source words did not activate the source words' semantics; hence, the $d'$ values are small. Pseudohomophones that were visually similar to their source words activated semantic patterns that strongly overlapped with the source words' semantics, yielding large $d'$ values. However, the latter effect was modulated by neighborhood density. If both the pseudohomophone and source word were from dense neighborhoods, the $d'$ was very small. Thus, the fact that a pseudohomophone such as KAR is visually similar to the homophonous word CAR had little impact because it is also close to many other words. When either the word or the pseudohomophone was from a sparse neighborhood, the semantic activation effect was much stronger.

The model suggests that some pseudohomophones activate semantic information directly from orthography. These findings are relevant to previous behavioral studies of pseudohomophones, which included items that produced semantic activation via orthography in the model. Many of the pseudohomophones in the Van Orden et al. (1988) study, for example, were visually similar to the source words. In addition, the pseudohomophones and visual controls differed in terms of relevant neighborhood characteristics. Van Orden et al. (1988) carefully equated the visual similarity of the control nonwords and the pseudohomophones to the exemplar using a measure of orthographic distance. However, 5 of the 10 pseudohomophones used in their first experiment were created by changing the spelling of the vowel of the source word (e.g., SHEAP–SHEEP) while retaining the onset and coda, whereas 1 of the 10 control nonwords involved this minimal change. Many of the control nonwords were also closer orthographically to other words than to the exemplar (e.g., PARRIT, the control for CARROT–KARRET, is visually closer to PARROT; HERT, the control for HEAT–HEET, is closer to, and homophonous with HURT). The net result is that the stimuli varied in terms of the neighborhood properties that affected semantic activation via orthography in this simulation.

We tested the pseudohomophones, the matched base words, and the matched control items used by Van Orden et al. (1988; excluding one set, containing KARRIT, because it was bisyllabic) and computed the semantic activations for these items in the intact model, the model with the orth→phon→sem pathway deleted, and the model with the orth→sem pathway deleted. The $d'$ of the resulting semantic representation to the veridical semantic representations of the base words was calculated as before. Table 8 shows the results.

Consistent with the above observations regarding differences between the pseudohomophone stimuli and nonword controls, the model produces more accurate semantic representations of the target words via the orth→sem pathway than the control nonwords, although this difference is rather small (a $d'$ of 2.0 vs. 1.7). The disparity between pseudohomophones and control nonwords is much greater for the orth→phon→sem pathway, indicating that the bulk of the activation of semantics is done via the phonological pathway, as would be expected. It should be noted, however, that the intact model produces even stronger semantic activation than

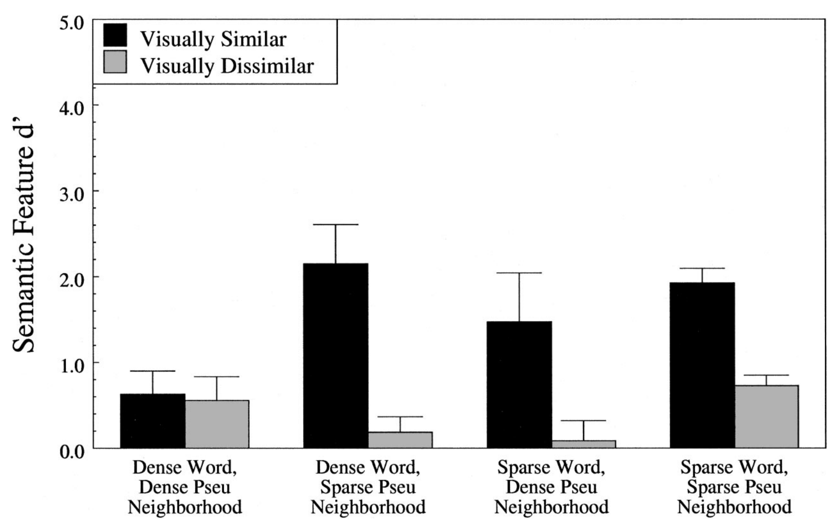


*Figure 35.* Effects of visual similarity, word neighborhood density, and pseudohomophone neighborhood density for the computation of semantic features along orth→sem. Error bars represent standard errors of the mean. Pseu = pseudohomophone.

Table 8
*Semantic d′ for the Van Orden, Johnston, and Hale (1988) Stimuli*

| Model | Stimuli: Word | Pseudohomophone | Control |
|---|---|---|---|
| Intact | Undefined | 4.3 | 1.9 |
| Orth→sem | 4.9 | 2.0 | 1.7 |
| Orth→phon→sem | 4.4 | 3.8 | 1.9 |

*Note.* Undefined = *d′* is undefined in this condition because there were no misses or false alarms.

the orth→phon→sem pathway alone. As with words, meaning is jointly determined by input from both pathways. Thus, the model is consistent with Van Orden et al.'s (1988) conclusion that these stimuli activate meaning via phonological recoding but suggests that semantics is also partially activated via orth→sem, contributing to the occurrence of false-positive responses.

## Processing Pseudohomophones With and Without a Spelling Check

In the final simulations we used the model to examine possible bases for participants' decisions that pseudohomophones are not words. Pseudohomophones activate semantics via phonology, and, as has just been seen, some may activate semantics directly from orthography as well. In the simulations to be reported we examined the patterns of activation produced by pseudohomophones and asked whether they differed systematically from those produced by words, providing a basis for correct rejections.

### *Simulation 18: Jared and Seidenberg (1991)—Pseudohomophones*

For this simulation we used the stimuli from Jared and Seidenberg (1991). Their studies included both homophones (the results of which were discussed above) and pseudohomophones. As with the homophones, for the pseudohomophones they manipulated the frequency of the homophonous exemplar (e.g., high frequency, DAWG→DOG; low frequency, CAUD→COD). Only the pseudohomophones of low-frequency exemplars (e.g., CAUD) produced a statistically reliable number of false positives.

The account of the Jared and Seidenberg (1991) homophone data presented earlier suggested that the orth→sem pathway provided disambiguating information that allowed participants to avoid false positives on most trials. It is not clear whether this account can also accommodate the pseudohomophone results. The stimuli in their experiment were visually dissimilar pseudohomophones, which, we have observed, do not produce very much activity along orth→sem in the model, and so this pathway would not provide the disambiguating information. Van Orden et al. (1988) suggested that participants use a spelling check. The simulation examined whether pseudohomophones provide any other basis for making correct decisions.

#### *Method*

The stimuli were constructed algorithmically by selecting sets of four words consisting of two pairs of words that rhyme but have different orthographic rimes. Pseudohomophones were algorithmically generated by swapping the orthographic word rimes (e.g., WAX, CRACKS→WACKS, CRAX). Nonwords were generated by changing the onsets. This method generated a large set of words, nonwords, and pseudohomophones in which each set of 12 items was perfectly matched for distribution of onsets and rimes (see Table 9 for an example). A set of 158 pseudohomophones resulted, 28 derived from high-frequency exemplars and 130 from low-frequency exemplars.[21] Here, as in all sets of 12, the onsets (e.g., F, CH, D, and CL) appear once and only once in each of the three columns, as do the rimes (ACT, ACKED, IDE, and IED). The visual similarity between pseudohomophones and their yoked words was generally low.

Rows whose word exemplars were [objects] or [living things] were extracted, and these were split into groups with a high-frequency exemplar and a low-frequency one in the same manner as described in the previous section. The presentation procedure was identical to the simulation of the Jared and Seidenberg (1991) homophone conditions; the intact model was used. As before, we tracked the activation levels of the distinguishing semantic features for the object and living thing concepts.

#### *Results*

Table 10 presents summary data concerning semantic activity for the pseudohomophones of low- and high-frequency exemplars. Pseudohomophones produced high amounts of activation on the critical semantic features, much more than seen in the simulations of Van Orden (1987) or the word effects in Jared and Seidenberg (1991) described previously. This degree of semantic activation is consistent with producing a larger false-positive rate than that observed in the behavioral study. Further, the effect is in the opposite direction: The pseudohomophones of high-frequency exemplars produced reliably more false positives than the low-frequency ones, $F(1, 156) = 14.8$, $p < .001$, whereas they produced fewer false positives.

These results follow from properties of the model we have discussed previously. The orth→sem pathway does not generate significant activation for pseudohomophones and nonwords that are "loners" (i.e., a pseudohomophone that is visually dissimilar to its source word or nonpseudohomophone that has few neighbors). The only source of semantic activation is via phon→sem, via which pseudohomophones reliably activate semantic features of the source word. Further, given that the phon→sem component is frequency sensitive, pseudohomophones of high-frequency exemplars activate semantics more strongly than low-frequency exemplars. Something else is clearly needed, however, to account for the fact that participants' false-positive rates are typically low, with pseudohomophones of high-frequency exemplars generating fewer false positives than those of low-frequency exemplars.

One possibility is that there are other sources of information relevant to making the decision available within the existing model. As mentioned earlier, Plaut (1997) used the statistic *stress* (see Equation 7) to measure how strongly units were activated. Plaut found that words tended to produce higher stress than nonwords, and this was posited as a basis for making lexical decisions. We found in Simulation 2 that words produced greater stress than nonwords. Therefore, we followed the method used in Simulation 2 and computed the stress for items in this simulation.

[21] There are far more pseudohomophones of low-frequency exemplars than high, because there are far more low-frequency words than high, and so the pool of candidate words is much larger.

Table 9
*Sample Pseudohomophone Stimuli*

| Word | Pseudohomophone | Nonword |
|---|---|---|
| Fact | Facked | Dact |
| Chide | Chied | Blide |
| Died | Dide | Fied |
| Clacked | Clact | Flacked |

Unfortunately, like the semantic activation, the semantic stress measure showed the opposite pattern from the behavioral data (see Table 10). The pseudohomophones of high-frequency exemplars produced higher stress, which means there is less of a reason to reject the item as a nonword. However, pseudohomophones derived from higher frequency words are easier for participants to reject as words than ones derived from lower frequency words (Jared & Seidenberg, 1991). The pseudohomophones of high-frequency exemplars produce higher stress for the same reason that they produce more activation of the inappropriate semantic feature: The phon→sem pathway is frequency sensitive, and high-frequency phonological forms can more powerfully activate semantics.

In the present context, the important question is whether the model we have described is compatible with the facts about how participants process pseudohomophones. Our general view is that making a semantic decision ("Is it a member of a category?"), like making a lexical decision ("Is it a word?") is a judgment task of considerable complexity (see Seidenberg, 1985, for discussion). The task demands that the participant establish criteria for reliably making accurate decisions. The model tells us something about the kinds of information that become available when words and nonwords are processed. This information is then used in performing various tasks, such as simply computing the meaning of a letter string, naming it aloud, or making semantic or lexical decisions about it. Tasks such as lexical and semantic decision involve additional processes related to making such judgments. We know that words and pseudowords produce different activation patterns in the model. For example, SUIT is a more familiar spelling pattern than SUTE, which could be detected if orthography were, like phonology and semantics in the implemented model, treated as an attractor system. Similarly, SUTE and SUIT do not produce identical semantic patterns. How these differences translate into decision criteria requires a theory of how such tasks are performed that is beyond the scope of the current work.

Table 10
*First Replication of Jared and Seidenberg (1991) Pseudohomophone Experiment*

| Condition | Pseudohomophone LF exemplar | Pseudohomophone HF exemplar |
|---|---|---|
| Jared and Seidenberg (1991) | | |
| False positives (%) | 10 | 6 |
| Simulation | | |
| Semantic activity | 0.19 | 1.53 |
| Semantic stress | 0.60 | 0.74 |

*Note.* LF = low-frequency; HF = high-frequency.

Of course, there is another possibility: a spelling check. Although the spelling check procedure is not necessary for disambiguating homophones (as discussed above), it may be required for pseudohomophones and other very wordlike nonwords. This makes intuitive sense: For familiar, learned words, the orth→sem pathway provides disambiguating information; for novel, unlearned words, orth→sem provides no useful information and so the model–reader must check to see whether a meaning is associated with a particular spelling (i.e., generate the orthographic code from semantics). The model we have been discussing cannot perform this computation; for simplicity we did not implement the semantics to orthography connections, which would have added significantly to the already considerable time required to train the model. Seidenberg and McClelland (1989) conducted preliminary research along these lines, however. Their model of the orth→phon computation included a feedback loop from the orthographic input to itself via an intermediate set of hidden units. This permitted the calculation of the discrepancy between the veridical input pattern and the one that was recreated on the input units via this feedback loop. This score reflected how wordlike a letter string was relative to the entire training corpus and provided a basis for making some word–nonword decisions. The following simulation extends this idea by considering the discrepancy between the orthographic input and one computed by means of the sem→orth pathway. We implemented a simple form of the semantics to orthography computation in order to determine whether it would provide a sufficient basis for detecting that pseudohomophones are not words as suggested by Van Orden et al. (1988).

### *Simulation 19: Jared and Seidenberg (1991)—Pseudohomophones Revisited*

#### *Method*

The basic method involved adding a semantics to orthography pathway to the model, illustrated in Figure 36. After training of this component was complete, the spelling check procedure was operationalized as follows. A pseudohomophone was presented as input, and semantics was activated via the intact orth→sem and orth→phon→sem pathways. In addition, the activated semantic pattern was used to compute an orthographic representation via sem→orth. The spelling check was based on the orthographic pattern computed on the backward pass from semantics.

This method of implementing the semantics–orthography computation is a simplification insofar as it uses a duplicate set of orthographic units and then a comparison between them to determine how wordlike a letter string is. As noted above, this was done for computational feasibility. Ideally, the semantic units would have feedback connections to the same orthographic units used to

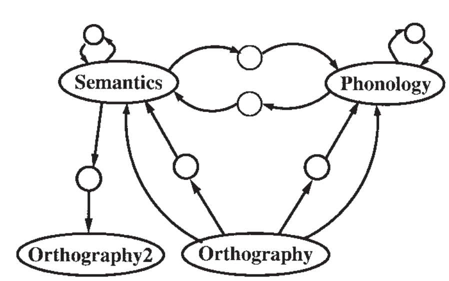


*Figure 36.* The revised model, with sem→orth pathway implemented.

Table 11
*Second Replication of Jared and Seidenberg (1991) Pseudohomophone Experiment*

| | Words | | | | Pseudohomophones | | | | Nonwords | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | HF | | LF | | HF | | LF | | HF | | LF | |
| Measure | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| Semantic stress | 0.999 | 0.01 | 0.972 | 0.07 | 0.736 | 0.20 | 0.596 | 0.23 | 0.611 | 0.24 | 0.635 | 0.22 |
| Orthographic stress | 0.995 | 0.01 | 0.922 | 0.07 | 0.919 | 0.07 | 0.860 | 0.07 | 0.872 | 0.07 | 0.858 | 0.07 |
| Orthographic distance | 0.068 | 0.06 | 1.610 | 1.90 | 6.510 | 3.40 | 8.570 | 3.90 | 9.410 | 3.90 | 7.680 | 3.00 |

*Note.* HF = high frequency; LF = low frequency.

input a word. The spelling check would then be performed by determining how well the model recreates the input pattern through the feedback connections. Seidenberg and McClelland (1989) implemented this procedure in their much simpler model and used it to compute what they termed an *orthographic error score*, which provided an index of how worldlike a letter string is. Seidenberg and McClelland provided evidence that this computation of orthographic familiarity plays a role in making lexical decisions. The present model implemented the same idea using a somewhat simpler technique, necessitated by the complexity of training the much larger model.

The stimuli for this experiment were the same as in the previous simulation. The sem→orth component was trained in the same fashion as the other simulations. It was trained for 800,000 word presentations using the entire training corpus at which point training had asymptoted at 99% accuracy. The sem→orth model was then attached to the existing model as shown in Figure 36.

We measured three variables: the disparity between the orthographic input and the orthographic representation re-created from semantics, the stress on those re-created orthographic representations, and the semantic stress. The spelling check was operationalized as a comparison between the input orthographic pattern and the pattern recomputed on the backward pass from semantics. If the input is a correctly spelled word in the model's vocabulary, the two patterns will closely match. If the input is not the correct spelling of a word, there will be a discrepancy between the two orthographic codes. Thus, SUTE will activate the semantics of SUIT via orth→phon→sem, but this semantic pattern will activate the spelling SUIT via sem→orth. The decision to reject the stimulus will depend on the degree of discrepancy and the model's confidence about the word's spelling pattern, which was reflected in the stress measure over the orthographic units.

## Results and Discussion

Table 11 depicts the results for the words, pseudohomophones, and nonwords. The effect of exemplar frequency was reliable for the pseudohomophones for the orthographic stress measure, $F(1, 156) = 33$, $p < .001$; the semantic stress measure, $F(1, 156) = 4.1$, $p < .05$; and the orthographic distance measure, $F(1, 156) = 6.5$, $p < .01$. As before, the semantic stress measure produced effects in the opposite direction to that of the empirical data: higher stress for the high-frequency items, which would make it more difficult to reject such items as nonwords.

However, the orthographic stress measure and the orthographic distance measure each patterned in the correct direction. This was because the semantic representations were activated more weakly by the phonological form of the low-frequency exemplars and, hence, re-created a more noisy orthographic representation, resulting in greater orthographic distance and a greater basis for rejecting the item as not being a word. Further, the orthographic stress for the pseudohomophones of low-frequency exemplars was lower than for those of high-frequency exemplars. Thus, the network had greater evidence for rejecting the pseudohomophones of high-frequency exemplars, on the basis of error and confidence in spelling, than the low-frequency exemplars.

In summary, the simulations indicate that some pseudohomophones (ones that are visually similar to their homophonous word and from orthographically sparse neighborhoods) activate semantic information via orth→sem. Visually dissimilar pseudohomophones yield little activation along this pathway. The effect of exemplar frequency observed by Jared and Seidenberg (1991) can be accounted for by including feedback from the semantic system to the orthographic component. This is a simple version of the spelling check proposed by Van Orden and colleagues (e.g., Van Orden et al., 1990), but without the additional assumption that orthography does not activate semantics directly.

As noted earlier, a formal simulation of lexical decision is beyond the scope of this work. Such a simulation would include a detailed account of the processes involved in making both "yes" and "no" decisions, and it would have to account for a mass of published results showing that lexical decision results are affected by experiment-specific factors that affect participants' response strategies (see Seidenberg, 1995, for discussion). However, Table 11 shows some of the sources of information that could plausibly be used in the lexical decision task. The semantic stress measure differs very strongly between words and nonwords, $F(1, 157) = 681$, $p < .001$, as does the orthographic distance, $F(1, 157) = 752$, $p < .001$. Figure 37 shows the distribution of stress values for the words, nonwords, and pseudohomophones in this experiment. Orthographic stress is a less strong discriminator but still differs reliably for words and nonwords, $F(1, 157) = 129$, $p < .001$. These variables are most likely not the only ones that could be involved in performing lexical decision (for instance, one could plausibly judge that XPMK is not a word simply by noting that it does not contain a vowel). Nonetheless, these variables produce results that provide a basis on which lexical decisions could be made.[22]

[22] Ratcliff, Gomez, and McKoon (2004) presented a diffusion model of the lexical decision task. Their model incorporates the idea that lexical decisions are made by establishing a decision criterion based on orthographic differences between words and nonwords (Seidenberg & McClelland, 1989). Ratcliff et al. fit a diffusion model with eight parameters to

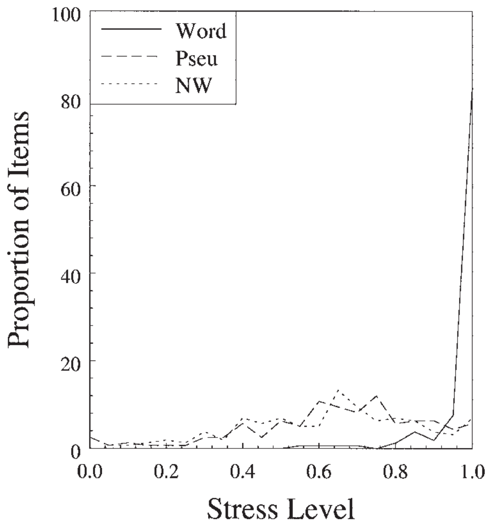


*Figure 37.* The distribution of stress values for items used in Simulation 19. Pseu = pseudohomophones; NW = nonwords.

## GENERAL DISCUSSION

As noted at the outset, there has been considerable debate concerning the mechanisms involved in computing the meanings of words from print. Although positions on the issue vary, most discussions have presupposed that there are independent direct-visual and phonologically mediated pathways and that for any given word, one of these mechanisms provides access to meaning. Some theories assume the direct route is used, some that phonological mediation is dominant, and others that both routes are used but for different words or writing systems. The Van Orden et al. (1990) article was a departure insofar as it emphasized the interactivity between different components of the system; however, it also identified factors that were thought to cause processing to proceed primarily via orth→phon→sem, with additional feedback from semantics to orthography for homophones.

Our view is that the existence of the direct-visual and phonologically mediated pathways must be considered separately from computational properties such as the kinds of representations they operate over, how they are learned, and whether they are independent. Every model of word reading will have to incorporate some version of the two procedures because they are licensed by the nature of the orthographic, phonological, and semantic codes and the relationships among them. Our model differs from previous accounts in a critical way: Meanings are determined by both pathways simultaneously. The model also differs from other proposals with respect to how the mechanisms work. For example, the traditional idea that the visual pathway involves activating atomic entries in a mental lexicon differs in essential ways from the idea that a pattern of activation develops over the semantic units based on orthographic input. Similarly, most previous theories have assumed that phonological codes are generated by applying grapheme–phoneme correspondence rules, but our model involves a statistical learning procedure. These differences between theories matter; they represent different claims about how knowledge is represented, acquired, and used and ultimately about how it is represented in the brain.

We have described the basic operation of the model in detail and have shown that it is consistent with various types of behavioral data. The dynamics of the implemented model are complex, but the principles that govern its behavior are much simpler. We built and trained a model consistent with the theoretical framework outlined in the introduction, which includes explicit claims about the nature of the word reading problem and how the task is performed by humans. The behaviors of the model that we have described followed as empirical consequences. We then observed that the model was consistent with various behavioral phenomena. The model also provided novel insights about many phenomena, insofar as they arise from somewhat different mechanisms than had been proposed in other theories.

In concluding this article we summarize the essential properties of the model. We also discuss issues that should be addressed in future research, including limitations of the current implementation.

### Summary of the Model's Basic Properties

#### *Activation of Meaning From Multiple Sources*

The activation of semantics builds up over time based on continuous input from all available sources, principally orth→sem and orth→phon→sem but also the semantic cleanup circuit. This characteristic derives from the architecture of the model, particularly the fact that it settles into a distributed semantic pattern over time rather than instantaneously accessing a stored definition-like meaning. In connectionist terminology, the computation of meaning is a *constraint satisfaction problem:* The computed meaning is that which satisfies the multiple constraints represented by the weights on connections between units in different parts of the network.

#### *Course of Learning*

Learning occurs within both the orth→sem and orth→phon→sem components throughout the course of training as indicated by the lesion experiments (see Figures 14–15). With sufficient training both pathways become highly accurate for many words, and thus both make significant contributions in the intact model.

#### *Precedence of the Phonological Pathway in Acquisition*

Because of the nature of the mappings, the orth→sem component takes longer to develop than orth→phon→sem, and so phonological mediation assumes primacy early on, as in beginning readers.

data from experiments examining effects of frequency and type of nonword on decision latencies. Although this model is an important step, the range of phenomena to which it was applied is limited. Studies of other phenomena such as pseudohomophone effects suggest that under some conditions lexical decisions are influenced by phonological and/or semantic information as well as orthographic information. These conditions involve a more complex weighing of different types of information than captured by Ratcliff et al.'s model.

### *Development of the Orth→Sem Pathway*

The orth→sem pathway has an advantage over orth→phon→sem because in the latter case semantic activation cannot occur until the pattern over the phonological units has become sufficiently clear. The orth→sem pathway has an intrinsic speed advantage because it involves fewer intermediate steps. This property, taken with a training procedure that emphasizes producing semantic patterns rapidly as well as accurately, leads to continued development in orth→sem even when orth→phon→sem produces correct output. The other factor that promotes learning in the orth→sem pathway is its role in disambiguating the many homophones in the language.

### *Capacity of the Orth→Sem Pathway*

Although it has an intrinsic speed advantage, the orth→sem pathway takes longer to learn, which limits its role initially. The capacity of orth→sem is also limited by the fact that some of the work is done by orth→phon→sem and the semantic cleanup apparatus. Thus, the orth→sem pathway is not forced to deliver the entire semantic pattern for a word by itself within the first few time steps. In reading a low-frequency homophone such as EWES, for example, orth→sem only has to activate sufficient information to suppress the incorrect meaning being activated through orth→phon→sem.

### *Cooperative Computation of Meaning*

Given the dynamics of the system and the computational properties of the components, the net result is that semantics receives significant input from both orth→sem and orth→phon→sem for almost all words. Moreover, the model with both pathways intact computes meanings more efficiently than the paths do independently. The division of labor between the two is affected by lexical properties including frequency and spelling–sound consistency as well as the amount of training.

### *Processing of Homophones*

Under normal presentation conditions, homophones are disambiguated through the use of both orth→sem and orth→phon→sem. The isolated orth→phon→sem pathway can produce correct patterns for higher frequency, dominant homophones. In the intact model, however, orth→sem also delivers relevant activation quickly, particularly for higher frequency words. The role of orth→sem is shaped by the fact that the orth→phon→sem pathway cannot accurately compute both meanings of a homophone pair. The latter pathway eventually becomes more tuned to the higher frequency member of a pair because it is trained more often; however, orth→sem also processes these words effectively and so contributes significantly. The analysis shown in Figure 29 demonstrates that the orth→sem pathway becomes very effective at suppressing features associated with the alternative meaning that are activated through phonology.

### *Use of Orth→Sem and Sem→Orth*

In the implemented model, homophones are disambiguated using information from orth→sem rather than a spelling check (sem→orth). This aspect of the model demonstrates that there is no computational reason why orth→sem cannot contribute to semantic activation, and the model's behavior in disambiguating homophones was consistent with that seen in human research participants. Although we did not include it in this implementation, there is no reason to prohibit feedback from semantics to orthography, which may also play a role in human performance. The contribution from orthography to semantics is more direct, however, and thus can be made use of more rapidly.

### *Effects of Masking*

The simulations suggest that the false-positive responses observed in studies such as Van Orden et al.'s (1988) arise because the normal input from orth→sem is terminated by presentation of a mask. This contrasts with the standard interpretation that the mask removes the orthographic pattern used in making a postaccess spelling check. Masking has less of an effect on activation within orth→phon→sem; the phonological system is a highly structured attractor that allows pattern completion to occur even in the absence of continued orthographic input. Although the semantic system is also an attractor, it is more sparse and therefore highly dependent on input from other sources (either orthography or phonology). The priming effects observed in studies such as Lesch and Pollatsek's (1993) arise in a similar manner.

## Future Directions

The model we have described is a partial realization of a broader theory. The implementational step was not trivial; it involved significant challenges concerning developing the phonological and semantic representations, training both components of the model simultaneously, analyzing the model's behavior, and relating it to behavioral evidence. Although the model has considerable scope, there are many other phenomena that can be explored using this version of it. Our decision to limit the discussion of the model to the results presented above was motivated by practical considerations (the need to keep the article to a manageable length; the desire to get the theoretical framework into the literature so that others could use it) rather than by our having exhausted the range of phenomena to which the model can be applied. Below we describe some of the issues that can be pursued using the existing model. However, the model can also be seen as instantiating a computational framework or tool kit for generating and testing hypotheses about many aspects of reading by varying how it is configured and trained. Such explorations may shed light on additional reading phenomena and also help in identifying limitations of the framework and the current implementation, which can be addressed in future models. We take this exploratory function of the model to be as important as showing that this particular implementation can account for additional facts. Below we briefly summarize some of the prominent directions for future research.

### *Robustness of the Implementation*

The general form of the model was closely tied to theoretical concerns, but many details of the implementation were not. Implementing the model requires making decisions about details such as the number of hidden units in a pathway, the setting of the parameter that determines how rapidly activation ramps up, and the way words are sampled during training. It will be necessary to determine whether these aspects of the implementation contribute in significant ways to its behavior, which can be done by comparing variants of the basic models. We think the model's behavior is likely to be robust because of the way it was developed, which did not involve trying a large number of possibilities and then finding the ones that produced the best results. We made implementational decisions based on previous experience and our understanding of network behavior and then observed the consequences. This strongly contrasts with the approach of Coltheart et al. (2001), whose methodology explicitly involves fitting models to data rather than deriving results from more general principles. Some parameters of our model are expected to affect performance but in theoretically interpretable ways. For example, Seidenberg and McClelland (1989) found that reducing the number of hidden units in the orth→phon pathway affected their model's capacity to learn less common spelling–sound mappings; this parameter may be related to individual differences among readers. Other parameters that were chosen for pragmatic reasons (e.g., to keep network running time within the limits set by our computers) can also be varied (e.g., using faster computers). These kinds of parameters should not have a large impact on core aspects of the model (e.g., the fact that meanings are jointly determined by input from both pathways), but this needs to be determined empirically.

### *Generating and Testing New Predictions*

One question often raised in connection with simulation models is whether it is possible to go beyond merely accounting for the results of existing studies to generating testable novel predictions. This question is of particular concern with respect to models that are developed by fitting particular behavioral data (Seidenberg, Zevin, & Harm, 2002), but our model was not developed in this way as we have emphasized throughout this article. Two questions do need to be addressed, however: (a) Does our model account for phenomena other than the ones we have described? And (b) does the model generate novel predictions that can be tested in new behavioral experiments?

The model is a device that generates phonological and semantic codes for words. The researcher then generates hypotheses (based on human or model performance) and tests them by running appropriate simulation and behavioral experiments. Our experience with previous models (Harm & Seidenberg, 1999; Plaut et al., 1996; Seidenberg & McClelland, 1989) is that researchers have thought of many hypotheses that can be tested using our models. Thus, we have provided model-generated data that have been used in studies such as those of Spieler and Balota (1997); Jared (1997); Treiman, Kessler, and Bick (2003); and others. The current model generates many predictions that can be tested immediately; for example, on the basis of the model's performance, we could design an experiment that would be an advance on the Van Orden paradigm insofar as it made specific predictions about which homophones or pseudohomophones activate semantics and thus are likely to generate false positives. The semantic representations in the model provide a basis for generating predictions about how semantic structure affects performance on tasks such as semantic priming, category decision, similarity judgment, and many others. McRae et al. (1997) showed that the magnitude of semantic priming effects could be predicted by measures of featural overlap between prime and target; our model can also be used to generate predictions about the magnitude and time course of such effects, using masked and unmasked stimuli.

A much broader range of phenomena could be addressed by extending the model to incorporate an explicit theory linking measures of network performance to response latencies (see below). Finally, the model makes some predictions that are very explicit but challenging to test using existing methodologies. This situation, in which a theory makes predictions that await the development of methods for testing them, is not uncommon in many sciences. For example, the model maps out the time course of activation along different pathways, but this is difficult to assess in a behavioral study. As an illustration of the problem, there are methods for detecting the use of phonological information in activating meaning, but there is not a comparably direct method for detecting when meaning has been activated directly from print. A false positive for "Is it a flower?: ROWS" provides strong evidence for phonologically mediated activation of meaning, but the absence of a false positive cannot be taken as evidence that phonological mediation did not occur (it could be that phonological mediation occurred but the participant was able to avoid a false positive using other information, e.g., orth→sem). It may be that neuroimaging techniques will soon be able to provide evidence about the time course of processing in brain regions that underlie direct and phonologically mediated mechanisms, particularly ones such as magnetoencephalography that yield dynamic rather than static information. Coupling the model with such techniques would facilitate testing the model and would also facilitate interpreting such neuroimaging data.

### *Other Phenomena*

Our focus has been on issues concerning the division of labor in the computation of meaning. However, the model can be used to address additional issues.

#### *Division of Labor in Pronunciation*

Issues concerning the pronunciation of words and nonwords have been the focus of considerable previous modeling research within the triangle framework and in Coltheart et al.'s (1993, 2001) DRC model. One issue is whether the model we have proposed can account for the naming phenomena (e.g., frequency and consistency effects) that have been the focus of ongoing debate about the adequacy of the two approaches. A second issue concerns the role of semantic information in naming aloud. We have extensively discussed how the orth→sem and orth→phon→sem pathways jointly determine meanings. The complementary issue with respect to pronunciation concerns the contributions of the orth→phon and orth→sem→phon pathways in pronunciation. The computation of phonology is constrained by the same principles that we have discussed with respect to the computation of meaning. The phonological code for a word will be jointly determined by input from both pathways; however, the resulting division of labor may have a different character than we have observed for the computation of meaning. In the case of meaning, both pathways contribute significantly; the trade-offs between the two pathways with respect to computational efficiency mean that neither dominates in skilled performance. The direct pathway has an advantage because it involves fewer steps but has a disadvantage because the mapping is largely arbitrary. In the computation of phonology, however, the direct pathway also involves the more consistent mapping; hence, it should dominate to a considerable degree. There is some evidence that semantic information plays a role in naming for some types of words, particularly ones for which the computation from orthography to phonology is very difficult (e.g., because they involve highly atypical spelling–sound mappings; Strain et al., 1995), but these effects may be relatively rare, at least in English.

### *Other Writing Systems*

One of the main factors that determined the division of labor in the present model was the nature of the mapping between orthography and phonology, which is quasiregular (Seidenberg & McClelland, 1989). Other alphabetic writing systems (such as the ones for Italian, Spanish, and Serbo–Croatian) adhere more closely to the principle that individual letters or combinations of letters correspond to a single phoneme (Hung & Tzeng, 1981; Seidenberg, 1992b). The model was trained on English, but with minor changes in the input representation and the development of suitable training corpora it can be trained on other writing systems. The model could then be used to make cross-orthography predictions and simulate results of behavioral studies.

How the division of labor is achieved in different writing systems is likely to be a complex issue involving interactions among several properties of the writing systems and the languages they represent. To date, most discussion has focused on one design feature, orthographic depth—that is, the consistency of the mapping between graphemes and phonemes. Other factors being equal, this factor will certainly affect the division of labor between visual and phonological pathways. However, the effects of numerous other factors need to be considered. Consider the dual Cyrillic and Roman writing systems for Serbo–Croatian, which have been extensively studied (e.g., Lukatela, Turvey, Feldman, Carello, & Katz, 1989). Both alphabets are shallow and therefore lack minimal pairs such as MINT–PINT in English. However, these writing systems do not represent syllabic stress, and Serbo–Croatian has many minimal pairs consisting of words with the same spelling but different pronunciations and meanings, due to differences in stress or intonation contour. For example, LUK has two distinct meanings (arch, onion) depending on whether the vowel is short and rising or long and falling. Thus, the Serbian and Croatian orthographies exhibit considerable ambiguity in the mapping between spelling and sound despite being shallow at the level of graphemes and phonemes. Moreover, these ambiguities also exist in the mapping from spelling to meaning. Resolving the ambiguities may therefore require using contextual information (as required for English homographs such as WIND and noun–verb alternations such as *CONtrast* vs. *conTRAST*). Similarly, Hebrew is a shallow orthography when its vowels are represented, but typically they are not. Removing the vowels shifts the orthography to deep, again creating dependence on contextual information for ambiguity resolution. Although we have drawn diagrams of our modeling framework with context units, we have not explored their use. Context seems particularly relevant, however, to understanding ambiguities that arise in writing systems for reasons other than transparency of grapheme–phoneme correspondences.

Although most research has focused on alphabetic writing systems, there is considerable data concerning the nonalphabetic writing systems for Chinese and Japanese. An important recent corpus analysis of Chinese (Shu, Chen, Anderson, Wu, & Xuan, 2003) showed that a large percentage of Chinese words consist of phonological and semantic components that jointly provide cues to the word's meaning. Thus, the visual and phonological processes in the model are realized by components of the words themselves. Reading Chinese words is a classic constraint satisfaction problem: Whereas the components in isolation may be ambiguous, the conjunction of the components is highly constraining. Shu et al.'s analyses suggest that Chinese has much in common with English with respect to the nature of the mappings between the written, spoken, and semantic codes for words; the fact that irregular mappings tend to occur in higher frequency words; the existence of quasiregular neighborhoods of related words; and so on. These facts suggest that there may be more similarities between the processing of English and the nonalphabetic Chinese writing system than between English and a shallow alphabetic writing system, but this remains to be explored in detail. It would not require major technical innovation to be able to represent Chinese characters as the "orthographic" input in our model. With a suitable training corpus, the model could then be used to examine where the statistical regularities in the writing system lie, how the different components of words jointly determine meaning, and how the resulting division of labor compares to that for English and other writing systems.

### *Acquisition*

Most of the findings discussed in this article concern skilled performance. Reading acquisition was considered only with respect to computational properties that yield initial dominance of the orth→phon→sem pathway. In ongoing research we are examining developmental issues in more detail. One goal is to use a training regime that adheres more closely to the child's classroom experience. In learning to read, children are initially exposed to a small vocabulary of words that expands over time. Instructional programs structure this experience in different ways. Our models use a frequency-biased sampling procedure that does not build much structure into the sequence of learning events. In current work we are examining how performance is affected by different ways of structuring this sequence (Foorman et al., 2001), especially whether there are ways to optimize efficiency of learning. A related issue concerns the nature of the feedback provided to the child or model in the course of learning. We used an idealized procedure in which the model was provided with feedback about the correct semantic and phonological codes for words. Children receive more variable feedback; explicit feedback from a teacher or listener is sometimes provided, but more often children provide their own feedback (e.g., by listening to what they have said and by using background knowledge or illustrations to infer intended meanings of words). This feedback can be partial or even incorrect. Our general view is that the learning that occurs under these conditions follows the same principles as we have explored but may be less efficient. On the other hand, children receive additional instruction that focuses on parts of words (e.g., the pronunciations of letters or rimes), which can also be incorporated in the training regime and may improve efficiency. In general, the model provides a powerful tool for examining assumptions about how to teach word reading.

### *Acquired Dyslexia*

Data concerning the partial loss of reading ability following brain injury have provided important evidence concerning basic mechanisms in reading and their brain bases. Different types of acquired dyslexia have been addressed using connectionist models of specific components of the triangle (see Hinton & Shallice, 1991, and Plaut & Shallice, 1993, for applications to deep dyslexia; Patterson, Seidenberg, & McClelland, 1989, and Plaut et al., 1996, for surface dyslexia; and Harm & Seidenberg, 2001, for phonological dyslexia). It would be a clear advance to determine whether all of these types of acquired dyslexia can be handled within a single, unified model.

## *Extensions to the Existing Model*

The range of phenomena the model can address is limited by various aspects of the implementation. At the time we began the research it seemed important to limit its scope somewhat in order to make progress in understanding basic computational mechanisms and in assessing the potential relevance of the framework to division of labor questions. Given what has been learned from the present work, as well as additional insights about computational mechanisms that have been achieved since we began several years ago, it should be possible to address many of these limitations in next-generation models.

### *Orthographic Representation*

Whereas we have spent considerable effort examining properties of semantic and phonological representations and processes, the nature of orthographic knowledge has not been addressed to the same degree. This asymmetry reflects a broader pattern within the study of reading: Important aspects of orthographic processing have been neglected. Although much is known about eye movements in reading (Rayner, 1998), our concerns focus on two areas: letter recognition and the encoding of sequential orthographic structure. Letter recognition is a complex categorization task in which the perceiver must abstract away from variation in size, color, font, and other properties. Models of reading have focused on the fact that letters represent sounds; clearly a child who has difficulty identifying letters will also experience greater difficulty in learning how they map onto sounds. However, letter processing interacts with phonological knowledge in other important ways. One is via the fact that letters have names (e.g., *D* is "dee"). Children's knowledge of letter names is strongly related to early reading ability (Treiman, Tincoff, Rodriguez, Mouzaki, & Francis, 1998). This may be due in part to the fact that letter names provide a basis for categorizing visual letters. That is, one cue that the varying exemplars of the letter *D* are members of the same category is the fact that they are all given the name "dee." Letter names may be particularly relevant to the formation of categories for letters such as *A*, *D*, and *E*, whose written forms exhibit a high degree of variability (e.g., because they have different upper- and lowercase forms). An impairment in the capacity to represent phonological information, as assumed by the phonological deficit account of dyslexia (Snowling, 1991), would affect the representation of letter names and, by hypothesis, the formation of letter categories.

Conversely, it is also possible that impairments that interfere with the formation of categories of visual letters could affect the development of phonological representations. We have already noted that the representation of speech in terms of phonemes (e.g., the three segments in BAT) seems to be a function of exposure to an alphabet, rather than a function of the demands of spoken language production or comprehension. A failure to develop appropriate categories for letters would then be expected to have downstream effects on phonological structure. According to this hypothesis, the phonological deficits so often observed in dyslexic children are due (wholly or in part) to impairments that have a nonphonological origin. In a highly interactive system, an impairment that affected the capacity to develop appropriate letter categories would affect the development of phonological representations, which would in turn feed back on the development of letter categories, via letter names. On this view, deficits in "phonological awareness," as measured by tasks that tap a segmental level of representation, are consequences of being a poor reader, not necessarily proximal causes. These conjectures suggest a need for additional research on letter processing and its role in the development of phonological representations.

Our models also ignore the development of knowledge concerning the sequential structure of written language, that is, orthographic redundancy. Skilled readers have expert knowledge of orthographic structure: They know that written language exhibits a highly constrained statistical structure. One obvious direction for future research would be to implement orthography in a manner analogous to what we have done with semantics and phonology, using distributed representations of orthographic features and an attractor structure capable of encoding a variety of cross-dependences among letters. We would expect this component of the model to exhibit properties associated with the "visual word form" area, a left inferior temporal region (the fusiform gyrus) involved in the processing of letter strings (e.g., Polk & Farah, 2002). Readers' expert knowledge of the structure of written words may be analogous to other types of visual expertise (e.g., faces, types of birds or vehicles) and have a similar brain basis (Gauthier & Tarr, 2002).

### *Modeling Response Latencies*

There are unresolved issues about the modeling of response latencies in connectionist and other types of computational models. Our models compute semantic or phonological patterns; there have to be additional assumptions that link the behavior of the model to the performance of tasks such as naming or semantic decision and to the response measures that are collected (e.g., naming or decision latencies and errors). We have not as yet attempted to model response times in a rigorous way. In previous research we found that general measures of the model's performance (e.g., mean summed squared error, settling times) related closely to general measures of human performance (e.g., mean latencies by condition). These measures do less well at accounting for more detailed aspects of performance such as response latencies for individual words (Spieler & Balota, 1997). The relatively poorer fit at this more refined grain reflects limitations of both the models and the human data, which contain considerable measurement error (Seidenberg & Plaut, 1998). Nonetheless, it is clear that much more could be done in terms of modeling response latencies. Settling times are easy to calculate (they simply reflect when activation stops changing significantly in an attractor net), and they capture some aspects of relative difficulty, but they need to be replaced by a measure with better theoretical motivation. Settling times reflect how long it takes the model to complete a pattern, whereas many tasks that participants perform can be initiated before the processing of the entire stimulus has been completed. Naming latencies, for example, reflect the time to initiate a spoken response, which may occur well before the participant has compiled an articulatory motor program for the entire word (Kawamoto, Kello, Higareda, & Vu, 1999; Kawamoto, Kello, Jones, & Barne, 1998). Thus, what is needed in the model is a measure related to how long it takes for enough of the pronunciation to have been computed to initiate a response, not the amount of time it takes the entire pattern to settle. Settling times for the onset phoneme(s) or onset and vowel may provide a closer account of naming latencies. The same issue arises with respect to performing tasks that involve meaning. A participant may be able to decide that SUIT is an object and not a living thing well before the entire semantic pattern has been computed. In this case, the settling times for features that identify SUIT as an object may provide a better fit to decision latencies. These are unresolved issues, however.

### *Multisyllabic Words*

The model was limited to monosyllabic words as in previous research (e.g., Seidenberg & McClelland, 1989). Multisyllabic words introduce many additional issues, for example, concerning the assignment of syllabic stress in pronunciation and the development of morphological representations (Seidenberg & Gonnerman, 2000). Expanding the scope of the model to include multisyllabic words will entail a larger model that takes longer to train and generates more complex behavior. The labor involved in developing, training, and testing a model of this scope is considerable, of course. Leaving this practical issue aside, the main obstacle is theoretical, not computational. How are multisyllabic words read? Complex words could be processed as wholes (as in our current model) or in parts (as seems to occur when words are fixated more than once; Rayner, 1998). The parts could be syllables or morphemes or clumps of adjacent letters that sometimes cross structural boundaries. These issues have not been resolved by behavioral research. If there were better information about how complex words are processed, it could be used to guide the development of a model. However, considerable additional work is needed here on both computational and behavioral fronts.

### *Connections to the Brain*

Our model was based on computational and behavioral considerations; it makes use of some design principles thought to reflect general properties of how the brain learns, processes, and represents information but is not closely tied to facts about the brain. In the period since we began this research, a growing body of information about lexical processing, particularly in reading, has emerged from the use of neuroimaging methodologies. Given the specificity of the computational theory and the increasing specificity of neuroimaging methodologies concerning both brain circuitry and the time course of processing, it should be possible to establish closer links between the two. Three types of questions can be addressed.

First, are basic properties of the model consistent with evidence concerning how reading is accomplished by the brain? Although we cannot yet closely link the model to the brain, there are some encouraging preliminary results. On the basis of functional magnetic resonance imaging studies of individuals with and without dyslexia, Pugh et al. (2000) argued that there are two major circuits involved in normal reading. One, termed the *dorsal parietotemporal system,* involves the angular gyrus, supramarginal gyrus, and posterior portions of the superior temporal gyrus. The other circuit, termed the *ventral occipitotemporal system,* involves portions of middle temporal gyrus and middle-occipital gyrus. Pugh et al. noted several differences between the two systems: The dorsal system develops earlier in reading acquisition than the ventral system, the dorsal system is more strongly implicated in phonological processing, and the dorsal system operates more slowly in skilled readers. There are some striking correspondences between the properties of these two systems and the major components of our model. The dorsal system seems to exhibit characteristics of the orth→phon→sem component of the model: It develops more rapidly and is responsible for phonological coding but ultimately activates semantics more slowly. The ventral system, like the orth→sem pathway in the model, develops more slowly, is not associated with phonological processing, and ultimately activates semantics more efficiently. Thus, there are isomorphisms between the brain circuits and model at least at a general level. These suggestive results raise many questions that can be addressed in future research. We do not know if the two circuits that Pugh et al. have identified solve the word reading problem in the same way as our model. For example, in our model, the two pathways cooperatively activate semantics; the Pugh et al. data do not address this issue and so are also consistent with an independent pathways account.

A second type of question is, How can neuroimaging data be incorporated in the models to make them more biologically realistic? As an example, there is a growing body of evidence concerning the brain's representation of different types of semantic information (see Martin, 2002, for a review). There is considerable evidence concerning the representation of different semantic categories (e.g., animals, tools, body parts) and different types of semantic information (e.g., sensory, motoric, affective, factual, etc.). The principles governing the organization of semantic memory in the brain, including many of the basic topographic facts, are still unknown. Still, it is clear that semantic memory is not the unordered vector of units in our model. It is a reasonable goal for a future-generation model to incorporate information about the organization of semantic representations as it becomes available. We expect future models to incorporate an increasing number of such neurobiological constraints.

Finally, the third type of question is whether our models can inform the investigation of the brain bases of reading (and other aspects of cognition) using neuroimaging. As we have already suggested, the model makes specific predictions about the time course of processing for different types of words, which suggests an important direction for neuroimaging techniques, such as magnetoencephalography, which can provide time course information. Similarly, understanding how reading is accomplished in the computational model may help in interpreting the results of neuroimaging studies, for example, by suggesting what functions different circuits are performing. This would take such studies beyond localization questions to issues of how the brain accomplishes a task such as reading.

Thus, we envision a productive feedback loop between model development and neuroimaging, where each can constrain the other and ultimately converge on an integrated computational–neurobiological model that captures facts about overt behavior.

## Conclusion

We have described a general theory of the computation of meaning from print based on motivated principles, and we have presented an implemented model that instantiates the theory and relates well to behavioral data. To our knowledge, this is the first large-scale implemented model that addresses how meanings are computed in a multicomponent processing system. The results of this work are quite promising and suggest a wide range of future directions for behavioral, neuroimaging, and modeling research on reading.

In implementing the model we attempted to address some controversies about basic mechanisms in reading at a more explicit computational level than in previous theorizing. The model is not likely to be correct in every detail, and, of course, the goal is to replace it with something better. The model serves an important function by raising the bar in terms of the theoretical and mechanistic levels at which these behavioral phenomena can be engaged and by clarifying the inferences that can be validly drawn from the behavioral studies that have provided the main data to be explained.

The model was constructed from theoretical components such as distributed representations and statistical learning procedures that are general rather than specific to reading and have already been applied to a broad range of phenomena. The novel aspects of the model concern the emergence of the division of labor in a multicomponent system, a concept that is also beginning to be applied in other domains (Gordon & Dell, 2003). Thus, the way that people achieve an efficient solution to the computation of meaning problem may exemplify how many complex tasks are mastered.

## References


Adams, M. (1990). *Beginning to read.* Cambridge, MA: MIT Press.

Andersen, R. (1999). Multimodal integration for the representation of space in the posterior parietal cortex. In N. Burgess & K. Jeffery (Eds.), *The hippocampal and parietal foundations of spatial cognition* (pp. 90–103). New York: Oxford University Press.

Anderson, S. (1988). Morphological theory. In F. Newmeyer (Ed.), *Linguistics: The Cambridge survey: Vol. 1. Linguistic theory: Foundations* (pp. 146–191). Cambridge, England: Cambridge University Press.

Balota, D. A. (1990). The role of meaning in word recognition. In D. A. Balota, G. B. Flores d'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 9–32). Hillsdale, NJ: Erlbaum.

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 340–357.

Baron, J. (1973). Phonemic stage not necessary for reading. *Quarterly Journal of Experimental Psychology, 25,* 241–246.

Baron, J., & Strawson, C. (1976). Use of orthographic and word-specific knowledge in reading words aloud. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 207–214.

Barto, A. G. (1985). Learning by statistical cooperation of self-interested neuron-like computing elements. *Human Neurobiology, 4,* 229–256.

Bertelson, P., & de Gelder, B. (1989). Learning about reading from illiterates. In A. M. Galaburda (Ed.), *From reading to neurons* (pp. 1–23). Cambridge, MA: MIT Press.

Besner, D., Twilley, L., McCann, R., & Seergobin, K. (1990). On the connection between connectionism and data: Are a few words necessary? *Psychological Review, 97,* 432–446.

Bishop, C. (1995). Training with noise is equivalent to Tikhonov regularization. *Neural Computation, 7,* 108–116.

Bradley, L., & Bryant, P. (1983). Categorizing sounds and learning to read—A causal connection. *Nature, 301,* 419–421.

Browman, C., & Goldstein, L. (1990). Representation and reality: Physical systems and phonological structure. *Journal of Phonetics, 18,* 411–424.

Bullinaria, J. (1996). Connectionist models of reading: Incorporating semantics. In *Proceedings of the First European Workshop on Cognitive Modeling* (pp. 224–229). Berlin, Germany: Technische Universitat Berlin.

Caplan, D. (Ed.). (1992). *Language: Structure, processing, and disorders.* Cambridge, MA: MIT Press.

Carey, S. (1978). The child as word-learner. In M. Halle, J. Bresnan, & G. Miller (Eds.), *Linguistic theory and psychological reality* (pp. 264–293). Cambridge, MA: MIT Press.

Carr, T. H., & Pollatsek, A. (1985). Recognizing printed words: A look at current models. In D. Besner, T. G. Waller, & G. E. MacKinnon (Eds.), *Reading research: Advances in theory and practice* (Vol. 5, pp. 2–82). New York: Academic Press.

Carroll, J. B., Davies, P., & Richman, B. (1971). *American Heritage word frequency book.* New York: Houghton Mifflin.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English.* New York: Harper & Row.

Cleermans, A. (1997). Principles for implicit learning. In D. Berry (Ed.), *How implicit is implicit learning?* (pp. 195–234). Oxford, England: Oxford University Press.

Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Ed.), *Strategies of information processing* (pp. 151–216). New York: Academic Press.

Coltheart, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 33*(A), 497–505.

Coltheart, M. (2000). Dual routes from print to speech and dual routes from print to meaning: Some theoretical issues. In A. Kennedy, R. Radach, D. Heller, & J. Pynte (Eds.), *Reading as a perceptual process* (pp. 475–490). Oxford, England: Elsevier.

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.

Coltheart, M., Davelaar, E., Jonasson, K., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention & performance VI* (pp. 135–155). Hillsdale, NJ: Erlbaum.

Coltheart, M., Patterson, K. E., & Marshall, J. C. (Eds.). (1980). *Deep dyslexia.* London: Routledge & Kegan Paul.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108,* 204–256.

Crowder, R. (1982). *The psychology of reading.* New York: Oxford University Press.

Daugherty, K., & Seidenberg, M. S. (1992). Rules or connections? The past tense revisited. In *Proceedings of the 14th Annual Meeting of the Cognitive Science Society* (pp. 259–264). Hillsdale, NJ: Erlbaum.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93,* 283–321.

Ellis, A. W., & Monaghan, J. (2002). Reply to Strain, Patterson, and Seidenberg (2002). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 215–220.

Flesch, R. (1955). *Why Johnny can't read.* New York: Harper.

Foorman, B. R., Perfetti, C., Seidenberg, M., Francis, D., & Harm, M. (2001, April). *What kind of text is a decodable text? And what kind of text is an authentic text?* Paper presented at the meeting of the American Education Research Association, Seattle, WA.

Forster, K. I. (1976). Accessing the mental lexicon. In R. J. Wales & E. Walker (Eds.), *New approaches to language mechanisms* (pp. 257–287). Amsterdam: North-Holland.

Francis, W. N., & Kučera, H. (1982). *Frequency analysis of English usage.* Boston: Houghton Mifflin.

Frost, R. (1998). Toward a strong phonological theory of visual word recognition: True issues and false trials. *Psychological Bulletin, 123,* 71–99.

Frost, R., Katz, L., & Bentin, S. (1987). Strategies for visual word recognition and orthographic depth: A multilingual comparison. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 104–115.

Gainotti, G. (2000). What the locus of brain lesion tells us about the nature of the cognitive deficit underlying category-specific disorders: A review. *Cortex, 36,* 539–559.

Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes, 12,* 613–656.

Gathercole, S., & Baddeley, A. (1993). Phonological working memory: A critical building block for reading development and vocabulary acquisition. *European Journal of Psychology of Education, 8,* 259–272.

Gauthier, I., & Tarr, M. J. (2002). Unraveling mechanisms for expert object recognition: Bridging brain activity and behavior. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 431–446.

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General, 113,* 256–281.

Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance, 5,* 674–691.

Gordon, J., & Dell, G. (2003). Learning to divide the labor: An account of deficits in light and heavy verb production. *Cognitive Science, 27,* 1–40.

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review, 103,* 518–565.

Harm, M. W. (1998). *Division of labor in a computational model of visual word recognition.* Unpublished doctoral dissertation, University of Southern California, Los Angeles.

Harm, M. W. (2002). *Building large scale distributed semantic feature sets with WordNet* (Tech. Rep. No. PDP.CNS.02.01). Pittsburgh, PA: Carnegie Mellon University, Center for the Neural Basis of Cognition.

Harm, M. W., McCandliss, B. D., & Seidenberg, M. S. (2003). Modeling the successes and failures of interventions for disabled readers. *Scientific Studies of Reading, 7,* 155–182.

Harm, M. W., & Seidenberg, M. S. (1997, August). *The role of phonology in reading: A connectionist investigation.* Paper presented at the Computational Psycholinguistics Conference, Berkeley, CA.

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review, 106,* 491–528.

Harm, M. W., & Seidenberg, M. S. (2001). Are there orthographic impairments in phonological dyslexia? *Cognitive Neuropsychology, 18,* 71–92.

Hebb, D. O. (1949). *The organization of behavior.* New York: Wiley.

Henderson, L. (1982). *Orthography and word recognition in reading.* London: Academic Press.

Hetherington, P., & Seidenberg, M. S. (1989). Is there "catastrophic interference" in connectionist networks? In *Proceedings of the 11th Annual Conference of the Cognitive Science Society* (pp. 26–33). Hillsdale, NJ: Erlbaum.

Hinton, G. E., & Shallice, T. (1991). Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological Review, 98,* 74–95.

Hung, D., & Tzeng, O. (1981). Orthographic variations and visual information processing. *Psychological Bulletin, 90,* 377–414.

Ishai, A., Ungerleider, L., Martin, A., & Haxby, J. (2000). The representation of objects in the human occipital and temporal cortex. *Journal of Cognitive Neuroscience, 12*(Suppl. 2), 35–51.

Jared, D. (1997). Spelling–sound consistency affects the naming of high-frequency words. *Journal of Memory and Language, 36,* 505–529.

Jared, D., McRae, K., & Seidenberg, M. S. (1990). The basis of consistency effects in word naming. *Journal of Memory and Language, 29,* 687–715.

Jared, D., & Seidenberg, M. S. (1991). Does word identification proceed from spelling to sound to meaning? *Journal of Experimental Psychology: General, 120,* 358–394.

Joanisse, M. F., & Seidenberg, M. S. (1999). Impairments in verb morphology after brain injury: A connectionist model. *Proceedings of the National Academy of Sciences, USA, 96,* 7592–7597.

Jorm, A. F., & Share, D. L. (1983). Phonological recoding and reading acquisition. *Applied Psycholinguistics, 4,* 103–147.

Jusczyk, P. W. (1997). *The discovery of spoken language.* Cambridge, MA: MIT Press.

Kawamoto, A., Kello, C., Higareda, I., & Vu, J. (1999). Parallel processing and initial phoneme criterion in naming words: Evidence from frequency effects on onset and rime duration. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 362–381.

Kawamoto, A. H., Kello, C. T., Jones, R., & Barne, K. (1998). Initial phoneme versus whole-word criterion to initiate pronunciation: Evidence based on response latency and initial phoneme duration. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 862–885.

Kelly, M. H. (1992). Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review, 99,* 349–364.

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

LaBerge, D. L., & Samuels, J. (1974). Toward a theory of automatic word processing in reading. *Cognitive Psychology, 6,* 293–323.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104,* 211–240.

Lesch, M. F., & Pollatsek, A. (1993). Automatic access of semantic information by phonological codes in visual word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 285–294.

Liberman, I. Y., & Shankweiler, D. (1985). Phonology and the problems of learning to read and write. *Remedial and Special Education, 6*(6), 8–17.

Liberman, I. Y., Shankweiler, D., & Liberman, A. M. (1989). The alphabetic principle and learning to read. In D. Shankweiler & I. Y. Liberman (Eds.), *Phonology and reading disability: Solving the reading puzzle* (pp. 1–33). Ann Arbor: University of Michigan Press.

Locke, J. L. (1995). Development of the capacity for spoken language. In P. Fletcher & B. MacWhinney (Eds.), *The handbook of child language* (pp. 278–302). Oxford, England: Blackwell.

Lukatela, G., & Turvey, M. T. (1994a). Visual lexical access is initially phonological: I. Evidence from associative priming by words, homophones, and pseudohomophones. *Journal of Experimental Psychology: General, 123,* 107–128.

Lukatela, G., & Turvey, M. T. (1994b). Visual lexical access is initially phonological: II. Evidence from phonological priming by homophones and pseudohomophones. *Journal of Experimental Psychology: General, 123,* 331–353.

Lukatela, G., Turvey, M., Feldman, L., Carello, C., & Katz, L. (1989). Alphabetic priming in bi-alphabetic word perception. *Journal of Memory and Language, 28,* 237–254.

Lundberg, I., Olofsson, A., & Wall, S. (1980). Reading and spelling skills in the first school years predicted from phonemic awareness skills in kindergarten. *Scandinavian Journal of Psychology, 21,* 159–173.

MacDonald, M. C. (1993). The interaction of lexical and syntactic ambiguity. *Journal of Memory and Language, 32,* 692–715.

Marchand, H. (1969). *The categories and types of present-day English word-formation: A synchronic–diachronic approach* (2nd ed.). Munich, Germany: Beck.

Marcus, M., Santorini, B., & Marcinkiewicz, M. A. (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics, 19,* 313–330.

Marshall, J. C., & Newcombe, F. (1973). Patterns of paralexia: A psycholinguistic approach. *Journal of Psycholinguistic Research, 2,* 175–199.

Martin, A. (2002). Functional neuroimaging of semantic memory. In R. Cabeza & A. Kingstone (Eds.), *Handbook of functional neuroimaging of cognition* (pp. 153–186). Cambridge, MA: MIT Press.

McCann, R. S., & Besner, D. (1987). Reading pseudohomophones: Implications for models of pronunciation assembly and the locus of word-frequency effects in naming. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 14–24.

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,* 419–457.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review, 88,* 375–407.

McCloskey, M., & Cohen, N. J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 23, pp. 109–164). New York: Academic Press.

McCusker, L., Hillinger, M., & Bias, R. (1981). Phonological recoding and reading. *Psychological Bulletin, 89,* 217–245.

McLeod, P., Plunkett, K., & Rolls, E. T. (1998). *Introduction to connectionist modelling of cognitive processes.* Oxford, England: Oxford University Press.

McRae, K., & Boisvert, S. (1998). Automatic semantic similarity priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 558–572.

McRae, K., de Sa, V. R., & Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General, 126,* 99–130.

Meyer, D. E., Schvaneveldt, R. W., & Ruddy, M. G. (1974). Functions of graphemic and phonemic codes in visual word recognition. *Memory & Cognition, 2,* 309–321.

Miller, G. A. (1990). WordNet: An on-line lexical database. *International Journal of Lexicography, 3,* 235–312.

Morton, J. (1969). The interaction of information in word recognition. *Psychological Review, 76,* 165–178.

National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction.* Retrieved from http://www.nicdh.nih.gov/publications/nrp/smallbook.htm

O'Reilly, R. C., & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience: Understanding the mind by simulating the brain.* Cambridge, MA: MIT Press.

Paap, K., Newsome, S., McDonald, J., & Schvanevelt, R. W. (1982). An activation verification model for letter and word recognition—The word-superiority effect. *Psychological Review, 89,* 573–594.

Paap, K. R., & Noel, R. W. (1991). Dual route models of print to sound: Still a good horse race. *Psychological Research, 53,* 13–24.

Page, M. (2000). Connectionist modelling in psychology: A localist manifest. *Behavioral and Brain Sciences, 23,* 443–512.

Patterson, K., & Hodges, J. R. (1992). Deterioration of word meaning: Implications for reading. *Neuropsychologia, 30,* 1025–1040.

Patterson, K., Lambon Ralph, M. A., Hodges, J. R., & McClelland, J. L. (2001). Deficits in irregular past-tense verb morphology associated with degraded semantic knowledge. *Neuropsychologia, 39,* 709–724.

Patterson, K. E., Marshall, J. C., & Coltheart, M. (Eds.). (1985). *Surface dyslexia: Neuropsychological and cognitive studies of phonological reading.* London: Erlbaum.

Patterson, K. E., Seidenberg, M. S., & McClelland, J. L. (1989). Connections and disconnections: Dyslexia in a computational model of reading. In P. Morris (Ed.), *Parallel distributed processing: Implications for psychology and neuroscience.* (pp. 131–181). Oxford, England: Oxford University Press.

Patterson, K., Suzuki, T., & Wydell, T. N. (1996). Interpreting a case of Japanese phonological alexia: The key is in phonology. *Cognitive Neuropsychology, 13,* 803–822.

Pearlmutter, B. A. (1989). Learning state space trajectories in recurrent neural networks. *Neural Computation, 1,* 263–269.

Pearlmutter, B. A. (1995). Gradient calculations for dynamic recurrent neural networks: A survey. *IEEE Transactions on Neural Networks, 6,* 1212–1228.

Perfetti, C. A., & Bell, L. (1991). Phonemic activation during the first 40 ms of word identification: Evidence from backward masking and priming. *Journal of Memory and Language, 30,* 473–485.

Perfetti, C. A., Bell, L., & Delaney, S. (1988). Automatic phonetic activation in silent word reading: Evidence from backward masking. *Journal of Memory and Language, 27,* 59–70.

Perfetti, C., & McCutchen, D. (1982). Speech processes in reading. In N. Lass (Ed.), *Speech and Language: Advances in basic research and practice* (Vol. 7, pp. 237–269). New York: Academic Press.

Pinker, S. (1991, August 2). Rules of language. *Science, 253,* 530–535.

Pinker, S. (2000). *Words and rules: The ingredients of language.* New York: HarperCollins.

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition, 28,* 73–193.

Pinker, S., & Ullman, M. (2003). The past and future of the past tense. *Trends in Cognitive Sciences, 6,* 456–463.

Plaut, D. C. (1997). Structure and function in the lexical system: Insights from distributed models of word reading and lexical decision. *Language and Cognitive Processes, 12,* 765–805.

Plaut, D. C., & Booth, J. R. (2000). Individual and developmental differences in semantic priming: Empirical and computational support for a single-mechanism account of lexical processing. *Psychological Review, 107,* 786–823.

Plaut, D. C., & Kello, C. T. (1999). The interplay of speech comprehension and production in phonological development: A forward modeling approach. In B. MacWhinney (Ed.), *The emergence of language* (pp. 381–415). Mahwah, NJ: Erlbaum.

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

Plaut, D. C., & Shallice, T. (1991). Effects of word abstractness in a connectionist model of deep dyslexia. In *Proceedings of the Thirteenth*

*Annual Conference of the Cognitive Science Society* (pp. 73–78). Hillsdale, NJ: Erlbaum.

Plaut, D. C., & Shallice, T. (1993). Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology, 10,* 377–500.

Polk, T. A., & Farah, M. (2002). Functional MRI evidence for an abstract, non-perceptual word-form area. *Journal of Experimental Psychology: General, 131,* 65–72.

Pugh, K., Mencl, W., Jenner, A., Katz, L., Lee, J., Shaywitz, S., & Shaywitz, B. (2000). Functional neuroimaging studies of reading and reading disability (developmental dyslexia). *Mental Retardation and Developmental Disabilities Review, 6,* 207–213.

Ratcliff, R., Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological Review, 111,* 159–182.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* 372–422.

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition, 14,* 191–201.

Rayner, K., Foorman, B., Perfetti, C., Pesetsky, D., & Seidenberg, M. (2001). How psychological science informs the teaching of reading. *Psychological Science in the Public Interest, 2*(2), 31–74.

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading.* Englewood Cliffs, NJ: Prentice Hall.

Rolls, E., Critchley, H., & Treves, A. (1996). Representation of olfactory information in the primate orbitofrontal cortex. *Journal of Neurophysiology, 75,* 1982–1996.

Rubenstein, H., Lewis, S. S., & Rubenstein, M. A. (1971). Evidence for phonemic recoding in visual word recognition. *Journal of Verbal Learning and Verbal Behavior, 10,* 645–657.

Rumelhart, D. E., Hinton, G., & Williams, R. (1986). Learning internal representations by error propagation. In D. E. Rumelhart, J. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 1: Foundations* (pp. 318–362). Cambridge, MA: MIT Press.

Rumelhart, D. E., McClelland, J. L., & the PDP Research Group (Eds.). (1986). *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 1: Foundations.* Cambridge, MA: MIT Press.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996, December 13). Statistical learning by 8-month-old infants. *Science, 274,* 5294.

Seidenberg, M. S. (1985). The time course of information activation and utilization in visual word recognition. In D. Besner, T. G. Waller, & E. M. MacKinnon (Eds.), *Reading research: Advances in theory and practice* (pp. 199–252). New York: Academic Press.

Seidenberg, M. S. (1987). Sublexical structures in visual word recognition: Access units or orthographic redundancy. In M. Coltheart (Ed.), *Attention & performance XII: The psychology of reading* (pp. 245–263). Hillsdale, NJ: Erlbaum.

Seidenberg, M. S. (1992a). Beyond orthographic depth: Equitable division of labor. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology and meaning* (pp. 83–114). Amsterdam: North-Holland.

Seidenberg, M. S. (1992b). Dyslexia in a computational model of word recognition in reading. In P. Gough, L. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 243–274). Hillsdale, NJ: Erlbaum.

Seidenberg, M. S. (1993). Connectionist models and cognitive theory. *Psychological Science, 4,* 228–235.

Seidenberg, M. S. (1995). Visual word recognition: An overview. In P. Eimas & J. L. Miller (Eds.), *Handbook of perception and cognition: Language* (pp. 137–179). New York: Academic Press.

Seidenberg, M. S., & Gonnerman, L. M. (2000). Explaining derivational morphology as the convergence of codes. *Trends in Cognitive Science, 4,* 353–361.

Seidenberg, M., & MacDonald, M. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive Science, 23,* 569–588.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96,* 523–568.

Seidenberg, M. S., & Plaut, D. C. (1998). Evaluating word reading models at the item level: Matching the grain of theory and data. *Psychological Science, 9,* 234–237.

Seidenberg, M. S., Plaut, D. C., Petersen, A. S., McClelland, J. L., & McRae, K. (1994). Nonword pronunciation and models of word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 1177–1196.

Seidenberg, M. S., & Tanenhaus, M. K. (1979). Orthographic effects on rhyme monitoring. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 546–554.

Seidenberg, M. S., & Waters, G. S. (1989). Word recognition and naming: A mega study [Abstract 30]. *Bulletin of the Psychonomic Society, 27,* 489.

Seidenberg, M. S., Waters, G. S., Barnes, M. A., & Tanenhaus, M. K. (1984). When does irregular spelling or pronunciation influence word recognition? *Journal of Verbal Learning and Verbal Behavior, 23,* 383–404.

Seidenberg, M. S., Waters, G. S., Sanders, M., & Langer, P. (1984). Pre- and postlexical loci of contextual effects on word recognition. *Memory & Cognition, 12,* 315–328.

Seidenberg, M., Zevin, J., & Harm, M. (2002, November). *DRC doesn't read correctly.* Paper presented at the meeting of the Psychonomic Society of America, Kansas City, MO.

Shu, H., Chen, X., Anderson, R. C., Wu, N., & Xuan, Y. (2003). Properties of school Chinese: Implications for learning to read. *Child Development, 74,* 27–47.

Simpson, G. B. (1994). Context and the processing of ambiguous words. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 359–374). San Diego, CA: Academic Press.

Smith, F. (1971). *Understanding reading.* New York: Holt, Rinehart & Winston.

Smith, F. (1973). *Psycholinguistics and reading.* New York: Holt, Rinehart & Winston.

Smith, F. (1983). *Essays into literacy.* Exeter, NH: Heinemann Educational Books.

Snowling, M. J. (1991). Developmental reading disorders. *Journal of Child Psychology and Psychiatry, 32,* 49–77.

Spencer, A. (Ed.). (1991). *Morphological theory.* London: Blackwell.

Spieler, D. H., & Balota, D. A. (1997). Bringing computational models of word naming down to the item level. *Psychological Science, 8,* 411–416.

Strain, E., & Herdman, C. M. (1999). Imageability effects in word naming: An individual differences analysis. *Canadian Journal of Experimental Psychology, 53,* 347–359.

Strain, E., Patterson, K., & Seidenberg, M. S. (1995). Semantic effects in single-word naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1140–1154.

Sutton, R. S. (1988). Learning to predict by the method of temporal differences. *Machine Learning, 3,* 9–44.

Swinney, D. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior, 18,* 645–660.

Tanenhaus, M., Leiman, J., & Seidenberg, M. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Learning and Verbal Behavior, 18,* 427–440.

Taraban, R., & McClelland, J. L. (1987). Conspiracy effects in word pronunciation. *Journal of Memory and Language, 26,* 608–631.

Treiman, R., Kessler, B., & Bick, S. (2003). Influence of consonantal context on the pronunciation of vowels: A comparison of human readers and computational models. *Cognition, 88,* 49–78.

Treiman, R., Tincoff, R., Rodriguez, K., Mouzaki, A., & Francis, D. J. (1998). The foundations of literacy: Learning the sounds of letters. *Child Development, 69,* 1524–1540.

Van Orden, G. C. (1987). A ROWS is a ROSE: Spelling, sound and reading. *Memory & Cognition, 15,* 181–198.

Van Orden, G. C., Johnston, J. C., & Hale, B. L. (1988). Word identification in reading proceeds from the spelling to sound to meaning. *Journal of Experimental Psychology: Language, Memory, and Cognition, 14,* 371–386.

Van Orden, G. C., Pennington, B. F., & Stone, G. O. (1990). Word identification in reading and the promise of subsymbolic psycholinguistics. *Psychological Review, 97,* 488–522.

Vihman, M. M. (1996). *Phonological development: The origins of language in the child.* Oxford, England: Blackwell.

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin, 101,* 192–212.

Williams, R. J., & Peng, J. (1990). An efficient gradient-based algorithm for on-line training of recurrent network trajectories. *Neural Computation, 2,* 490–501.

Zeno, S. (Ed.). (1995). *The educator's word frequency guide.* Brewster, NJ: Touchstone.

Zevin, J. D., & Seidenberg, M. S. (2002). Age of acquisition effects in reading and other tasks. *Journal of Memory and Language, 47,* 1–29.

Zipf, G. K. (1935). *The psycho-biology of language: An introduction to dynamic philogy.* Boston: Houghton Mifflin.

Zorzi, M., Houghton, G., & Butterworth, B. (1998). Two routes or one in reading aloud? A connectionist dual-process model. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1131–1161.

Received May 2, 2001
Revision received June 6, 2003
Accepted August 8, 2003 ■