Psychological Review
2000, Vol. 107, No. 4, 786–823

0033-295X/00/$5.00 DOI: 10.1037//0033-295X.107.4.786

Individual and Developmental Differences in Semantic Priming: Empirical and Computational Support for a Single-Mechanism Account of Lexical Processing

David C. Plaut
Carnegie Mellon University

James R. Booth
Northwestern University

Existing accounts of single-word semantic priming phenomena incorporate multiple mechanisms, such as spreading activation, expectancy-based processes, and postlexical semantic matching. The authors provide empirical and computational support for a single-mechanism distributed network account. Previous studies have found greater semantic priming for low- than for high-frequency target words as well as inhibition following unrelated primes only at long stimulus-onset asynchronies (SOAs). A series of experiments examined the modulation of these effects by individual differences in age or perceptual ability. Third-grade, 6th-grade, and college students performed a lexical-decision task on high- and low-frequency target words preceded by related, unrelated, and nonword primes. Greater priming for low-frequency targets was exhibited only by participants with high perceptual ability. Moreover, unlike the college students, the children showed no inhibition even at the long SOA. The authors provide an account of these results in terms of the properties of distributed network models and support this account with an explicit computational simulation.

It is well established that people are faster and more accurate to read a word (e.g., BUTTER) when it is preceded by a related word (e.g., BREAD) compared with when it is preceded by an unrelated word (e.g., DOCTOR; Meyer & Schvaneveldt, 1971). This priming effect occurs for word pairs that are either categorically related or associatively related (C. A. Becker, 1980) and in a variety of tasks, including both word naming and lexical decision (Meyer, Schvaneveldt, & Ruddy, 1975). The magnitude of priming effects is influenced not only by stimulus and experimental factors but also by individual differences in age and reading ability (Stanovich, 1980).

The robustness and generality of priming effects have led theorists to suggest that these effects reflect fundamental properties of lexical knowledge within the human cognitive system. Spreading-activation theories (e.g., Anderson, 1983; McNamara, 1992, 1994) assume that semantic memory consists of a network of interconnected nodes and that activation spreads along the connections in this network. This spread of activation is assumed to be fast and automatic, causing a related prime to facilitate the processing of a target word (Balota & Chumbley, 1984; Chumbley & Balota, 1984). Alternatively, compound-cue theories (e.g., Dosher & Rosedale, 1989; Ratcliff & McKoon, 1988) propose that, in processing a word, semantic memory is accessed using a cue consisting of the word conjoined with the context in which it occurs (i.e., the preceding word). Because related words co-occur more frequently than do unrelated words, their compound cues tend to have greater familiarity, resulting in faster retrieval (according to most general memory models; e.g., Gillund & Shiffrin, 1984; Hintzman, 1986; Murdock, 1982).

Distributed network models have also been proposed to account for semantic priming effects (e.g., S. Becker, Moscovitch, Behrmann, & Joordens, 1997; Cree, McRae, & McNorgan, 1999; Masson, 1995; McRae, Sa, & Seidenberg, 1997; Moss, Hare, Day, & Tyler, 1994; Plaut, 1995; Sharkey & Sharkey, 1992). The fundamental assumption in such models is that concepts are represented by distributed patterns of activity over a large number of interconnected processing units, such that related concepts are represented by similar (overlapping) patterns. Semantic priming arises because, in processing a target, the network starts from the

David C. Plaut, Department of Psychology and Computer Science and the Center for the Neural Basis of Cognition, Carnegie Mellon University; James R. Booth, Department of Communication Sciences and Disorders, Northwestern University.

The research was supported by National Institute of Mental Health (NIMH) FIRST Award MH55628, NIMH Training Grant 5T32MH19102, and National Institute of Child Health and Human Development Grant 80258.

The computational simulation was run using customized software written within the Xerion simulator (Version 3.1) developed by Drew van Camp, Tony Plate, and Geoff Hinton at the University of Toronto. We thank the students and administrators at Holton Arms School and Landon School for their participation in this study. We also thank Gregory Robison and Rwanda Latoya Jackson for their assistance in conducting the experiments and Marlene Behrmann, Chris Kello, William Hall, Chuck Perfetti, Jay McClelland, Brian MacWhinney, and the Carnegie Mellon University Parallel Distributed Processing Research Group for helpful comments and discussion.

Correspondence concerning this article should be addressed to either David C. Plaut, Mellon Institute 115, Carnegie Mellon University, 4400 Fifth Avenue, Pittsburgh, Pennsylvania 15213-2683, or James R. Booth, Department of Communication Sciences and Disorders, Frances Searle Building, Northwestern University, 2299 North Campus Drive, Evanston, Illinois 60208-3560. Electronic mail may be sent to either plaut@cmu.edu or j-booth@nwu.edu.

pattern produced by the prime, which is more similar to the representation of the target for a related prime compared to an unrelated prime. In some formulations (e.g., Moss et al., 1994; Plaut, 1995), associations among words are reflected by an increased probability of the transition from one concept to another during training, somewhat similar to the formation of cues in compound-cue theory. Associative priming thus arises because the network has learned to derive the representation of a target word more frequently, and hence more effectively, when starting from the representation of an associated prime word compared with a nonassociated prime.

All of these theoretical frameworks—spreading-activation theories, compound-cue theories, and distributed network theories—are challenged to varying degrees by the considerable body of research showing that priming effects are influenced by a variety of experimental factors, including target frequency, category dominance, relatedness proportion, stimulus quality, stimulus-onset asynchrony (SOA), and the task performed by participants (see Neely, 1991, for a review). The almost universal response to these challenges is to complicate theories of lexical processing by postulating additional mechanisms that collectively account for the range of findings, albeit in a post hoc manner.

A good example of a multiple-mechanism account is Neely and Keefe's (1989) hybrid three-process theory, in which spreading activation is augmented with expectancy-based processes and with retrospective semantic matching in attempting to explain all of the priming effects in lexical decision and naming. According to this theory, presentation of a prime first engages automatic spreading activation processes. In addition, individuals are assumed to use the prime to generate two expectancy sets of possible targets: a set of visually similar items and a set of semantically similar items. When the target is presented, individuals first search for the target among items in the semantic set in random order; if it is not found, they then search through the items in the visual set in order of their frequency. Performance is assumed to be facilitated if the target is found and inhibited if it is not. In addition, following lexical access of the target but before executing a response, individuals retrospectively compare the target with the prime to determine whether they are related; performance is further facilitated if they are and inhibited if they are not.

The central goal of the current work is to provide both empirical and computational support for a more parsimonious, single-mechanism account of semantic priming phenomena, and lexical processing more generally, in terms of the properties of distributed network models. In developing our account, we focus on two particular sets of empirical findings, concerning (a) the effects of target frequency on the magnitude of semantic priming and (b) the degree to which priming effects result from facilitation or inhibition (relative to a neutral prime baseline) as a function of SOA. Researchers have found greater priming for low- compared with high-frequency targets (C. A. Becker, 1979; Borowsky & Besner, 1993) and inhibition at long but not short SOAs (C. A. Becker, 1980; Heyer, Briand, & Smith, 1985; L. C. Smith, Briand, Klein, & Heyer, 1987). We chose to focus on these findings because they have played a fundamental role in the formulation of multiple-mechanism accounts of word recognition. In particular, accounting for them has required the addition of considerable complexity to spreading-activation theories, including the separation of frequency and semantic context effects (Borowsky & Besner, 1993) and the introduction of additional strategic, expectancy-based processes (C. A. Becker, 1980; Neely, 1977; Paap, Newsome, McDonald, & Schvaneveldt, 1982).[1] They have also placed important constraints on accounts of individual differences in the performance of young versus old and poor versus good readers (e.g., Perfetti & Hogaboam, 1975; Stanovich, 1980). We explicitly chose not to focus on other empirical findings that are often taken to support multiple-mechanism accounts, such as blocking and strategy effects (e.g., C. A. Becker, 1980; Groot, 1984; Neely, Keefe, & Ross, 1989; M. C. Smith, Besner, & Miyoshi, 1994), because, as we argue in the General Discussion, these findings may derive from characteristics of general decision processes outside the lexical system per se (see also Kello & Plaut, 2000).

We present the results of empirical studies that examined the extent to which frequency effects and patterns of facilitation versus inhibition in priming depend on individual differences among people in age and in perceptual ability. We found that greater priming for low-frequency targets was exhibited only by those with high perceptual ability and that this finding held across differences in age and SOA. We also replicated the finding of inhibition at a long but not short SOA for adults; we found no inhibition for children even at the long SOA. We provide an account of these results in terms of the properties of distributed network models and support this account by demonstrating that an implemented simulation that does not separate frequency and context effects and which lacks expectancy-based processes nonetheless reproduces the most important empirical findings. We consider the strengths and limitations of the approach, and how it might be extended to account for additional semantic priming phenomena, in the General Discussion.

Effects of Target Frequency on Semantic Priming

The finding of an interaction of target frequency and priming context in lexical decision tasks, such that priming effects are larger for low- compared with high-frequency targets, has important implications for theories of lexical processing. This Frequency × Context interaction has been demonstrated using both sentence contexts (e.g., Stanovich & West, 1981; Stanovich, West, & Feeman, 1981) and single-word primes (C. A. Becker, 1979; Borowsky & Besner, 1993). The traditional account of this interaction within a spreading-activation framework is that the resting activation level of a word unit is further from threshold for low-frequency words than for high-frequency words, resulting in a larger effect of priming context on the former than on the latter. A potential problem with this explanation, however, is that priming context also interacts with stimulus quality, but target frequency and stimulus quality do not interact with each other (Borowsky & Besner, 1993). This pattern of results makes it difficult—at least within an additive factors framework (Sternberg, 1969)—to locate context and frequency effects at the same stage of processing. This led Borowsky and Besner (1993) to postulate that, whereas priming effects arise within semantics, frequency effects are due to the

[1] We contrast our distributed network account primarily with spreading-activation accounts because such accounts have been elaborated in the greatest detail to address the issues under consideration. We believe, however, that similar issues arise with respect to compound-cue theories (e.g., Ratcliff & McKoon, 1988).

mapping between orthography and semantics. In their account, this mapping is stronger for high-frequency words because they have been encountered more often during reading than low-frequency words. Consequently, high-frequency target words generate more rapid activation of semantics from orthography and, therefore, are less affected by priming context, than are low-frequency targets.

Thus, spreading-activation theories can account for the interaction of priming context and target frequency, if it is assumed that these factors influence different stages of processing. However, certain other findings in the priming literature—the relative effects of facilitation and inhibition as a function of SOA in skilled readers, and individual differences in priming effects as a function of age or reading ability—have prompted the introduction of additional complexities into spreading-activation theories.

Effects of SOA on Facilitation and Inhibition in Priming

Much of the research on semantic priming in adults has focused on the time course of facilitation and inhibition in naming and lexical decision tasks. In these studies, *facilitation* is defined as a decrease in reaction time (RT) to a target word following a related priming context compared to a neutral context (e.g., a nonword or a string of Xs), whereas *inhibition* is defined as an increase in RT following an unrelated context versus a neutral context. In general, priming effects are smaller at short SOAs (i.e., <250 ms) compared with long SOAs (>800 ms). Furthermore, the effects at short SOAs are due only to facilitation for both categorical and associative priming. At long SOAs, associative priming effects still result primarily from facilitation, whereas categorical priming effects result from both facilitation and inhibition (see, e.g., C. A. Becker, 1980).

These findings have generally been interpreted in terms of a distinction between automatic and strategic processes (Posner & Snyder, 1975). Spreading activation is automatic and occurs without intention, whereas strategic processes require conscious attention and are of limited capacity. In an extension of this dichotomy, Neely (1977) suggested that facilitation in word recognition results from fast, automatic spreading activation, whereas inhibition results from conscious, expectancy-based processes. The latter are slow and strategic because they involve the explicit generation of a set of potential targets from the prime; processing is assumed to be facilitated if the set contains the actual target and inhibited if it does not (C. A. Becker, 1980; Neely & Keefe, 1989).

We refer to models that postulate separate spreading-activation and expectancy-based processes as *dual-mechanism* models. Such models account for the aforementioned findings concerning the relative time course of facilitation and inhibition in the following way. At a short SOA, the recognition of a target can be facilitated by a related prime as a result of fast, automatic spreading activation, but it cannot be inhibited by an unrelated prime because there is insufficient time for the slow, expectancy-based processes to operate. At long SOAs, expectancy-based processes (along with spreading activation, in some formulations) have time to influence word recognition, so the priming effects result from both facilitation and inhibition (see Neely, 1991).

Developmental and Individual Differences in Priming

Additional constraints on theories of semantic priming come from studies of developmental differences in the influence of priming context on word recognition. A number of studies have shown larger priming effects for younger and poor readers than for older and good readers when reading target words presented after a single-word or sentential priming context (e.g., Nation & Snowling, 1998a; Perfetti & Hogaboam, 1975; Schwantes, 1985; Simpson & Lorsbach, 1983; Stanovich et al., 1981).[2] Other studies have investigated contextual processes in children by examining their oral reading errors (Biemiller, 1970; Goldsmith-Phillips, 1989; Jackson & Biemiller, 1985; Wijnen, 1992). These studies show that the oral reading errors of older and good readers tend to be phonemically related to the text being read, whereas the errors of younger and poor readers tend to be semantically or syntactically related. Overall, younger and poor readers seem to show larger contextual effects than older and good readers.

The most commonly cited account of the developmental differences in word priming effects is Stanovich's (1980) *interactive compensatory* model (see also Perfetti & Lesgold, 1977; Perfetti & Roth, 1981). According to this model, older and good readers have fast and automatic word decoding skills and, thus, rely less on expectancy-based processes to facilitate or inhibit word recognition. By contrast, because their word decoding is slower and less automatic, younger and poor readers rely more heavily on expectancy-based processes (Raduege & Schwantes, 1987) and, thus, are expected to exhibit a greater degree of inhibition.

With regard to effects of target frequency, a further relevant property of the interactive compensatory model is that higher level operations (e.g., expectancy-based processes) affect lower level processes (e.g., word recognition) only when the latter are slow or strategic. It is well known that, for adults, word recognition is faster for high- compared with low-frequency words (see Monsell, 1991, for review). For children, however, the decoding of most high-frequency words is not automatized to adult levels until the middle elementary school years (see, e.g., Golinkoff & Rosinki, 1976; Guttentag & Haith, 1979; Perfetti & Hogaboam, 1975). Children should, therefore, use priming context to an equal degree for recognizing high- and low-frequency words because they decode all words slowly. In this way, the interactive compensatory model predicts a three-way Age (Ability) × Priming Context × Target Frequency interaction. Specifically, older and high-ability readers should show greater priming effects for low- compared with high-frequency target words, whereas younger and low-ability readers should show equivalent priming for these items. Moreover, the model predicts that older and high-ability readers should show both facilitation and inhibition for low-frequency targets but only facilitation for high-frequency targets, whereas the younger and low-ability readers should show both facilitation and inhibition for both low- and high-frequency targets.

The existing literature on the relative contribution of facilitation and inhibition to priming context effects for younger and low-ability readers is inconclusive. Most studies on developmental differences in priming effects have used entire sentences as context manipulations and have found that older children and adults ex-

[2] Note, however, that when Nation and Snowling (1998a) compared the performance of children with good versus poor comprehension who were matched on decoding skills, they found that good comprehenders benefitted more from context than poor comprehenders when contextual facilitation was normalized relative to baseline performance on isolated targets.

hibit both less facilitation and less inhibition compared with younger children (Schwantes, Boesl, & Ritz, 1980; Stanovich, Nathan, West, & Vala-Rossi, 1985; West & Stanovich, 1978; West, Stanovich, Feeman, & Cunningham, 1983). However, these sentence priming paradigms differ in important ways from single-word priming paradigms. Sentence priming is influenced by syntactic and discourse-level factors, such that the developmental decrease in inhibition from sentence contexts may be due to greater efficiency of sentence integration processes with increased reading experience (Simpson & Lorsbach, 1987). By contrast, single-word priming is assumed to reflect lexical processing more purely and may give rise to a different developmental pattern of facilitation and inhibition.

To our knowledge, only two studies have examined developmental differences in facilitation and inhibition using the single-word priming context paradigm with a neutral prime baseline (i.e., a string of *X*s). In a naming study involving 2nd-grade, 4th-grade, 6th-grade, and college students, Simpson and Lorsbach (1983) found that younger students exhibited more facilitation and less inhibition compared with older students for stimulus lists with a high proportion of related trials (75%).[3] The same result was obtained in a second naming study that contrasted good versus poor readers among 4th-grade and 6th-grade students (Simpson & Lorsbach, 1987). These results suggest that, contrary to the predictions of the interactive compensatory model, facilitation at the lexical level decreases with development, whereas inhibition increases with development. This conclusion is also supported by research indicating that the ability to use inhibitory processes in picture naming and Stroop tasks is still developing during the early elementary school years (Guttentag & Haith, 1979; Schadler & Thissen, 1981).

As far as we know, no empirical studies of developmental or individual differences have investigated whether the magnitude of the interaction between a single-word priming context and target frequency is greater for older and high-ability readers than for younger and low-ability readers, and no developmental studies have investigated the time course of facilitation and inhibition using nonword primes as a neutral baseline condition. Our empirical and computational modeling work did exactly this.

Most studies of individual differences in semantic priming effects have focused on the impact of overall reading skill, as indexed by standardized tests of naming accuracy and reading comprehension. Few studies have examined the relative contribution of specific aspects of reading skill. Thus, we do not know whether the differences in semantic priming result from variations in higher level reading skills, such as vocabulary knowledge and inferential processes, or in lower level reading skills, such as perceptual encoding ability. Although vocabulary knowledge is a strong determinant of reading ability (Stahl, Hare, Sinatra, & Gregory, 1991), perceptual efficiency, as measured by match-to-sample tasks, is also related to reading proficiency (Vernon, 1987). Moreover, deficits in rapid perceptual processing are strongly associated with abnormal reading acquisition (Booth, Hunt, Perfetti, & MacWhinney, 1998; Eden et al., 1996; Lovegrove, Martin, & Slaghuis, 1986) and abnormal language development more generally (e.g., Tallal, Miller, & Fitch, 1993; Tallal & Piercy, 1975; see Farmer & Klein, 1995, for a review). In fact, perceptual efficiency may be particularly relevant in the early stages of reading acquisition. Detterman and Daniel (1991) found that the correlation of perceptual efficiency measures with the Wechsler IQ score was .26 for high-IQ individuals but .60 for low-IQ individuals. Given the typically strong relationship found between IQ and reading skill, the latter high correlation suggests that lower level perceptual abilities play an important role in the development of reading skill. Our empirical studies considered directly whether individual differences in perceptual ability, as measured by a match-to-sample task, has an important impact on the use of priming context for recognizing high- versus low-frequency target words.

Distributed Network Models of Semantic Priming

A spreading-activation framework can account for the interaction of priming context with target frequency, as well as the relative time course of facilitation and inhibition, but not without being elaborated to include discrete processing stages and separate, expectancy-based processes. Our central theoretical claim is that distributed network models can also account for these empirical findings, as well as the novel ones reported below, without these added complexities, and hence provide a more parsimonious explanation than do spreading-activation theories. The approach has the added benefit of being supported by computational simulations that make fully explicit the underlying mechanism that actually gives rise to the appropriate effects.

Our account takes as its starting point a preliminary distributed network simulation developed by Plaut (1995). Although the simulation was not applied to modeling specific empirical data, it exhibited a number of effects that are relevant in the current context. The network was trained on an abstract version of the task of mapping from written words to their meanings. The written form of each word was represented by a particular pattern of activity over a set of orthographic units, and its meaning was represented by another pattern over a set of semantic units. Categorical relatedness among words was encoded by the degree of feature overlap in their semantic representations, whereas associative relatedness was encoded by the frequency with which one word followed another during training (also see Moss et al., 1994). Words also differed in their frequency of presentation during training and in their degree of category dominance (i.e., how similar they were to a prototype pattern for their category), and target words were presented both at normal and reduced contrast (i.e., less binary orthographic input). When tested for priming effects following training, the network exhibited greater associative priming for low-frequency targets than for high-frequency targets, as has been found with experiment participants (C. A. Becker, 1979; Borowsky & Besner, 1993).[4]

[3] Note that lists with a lower relatedness proportion (25%) produced only the increase in facilitation but no reliable decrease in inhibition.

[4] In addition to target frequency, stimulus quality also interacted with priming context (i.e., greater priming for degraded compared with intact stimuli), but stimulus quality did not interact with target frequency. Thus, the Plaut (1995) network exhibited the pattern of results found empirically by Borowsky and Besner (1993) and was taken by them to imply that frequency and context effects must be located at distinct processing stages. Note, however, that frequency and context effects are not restricted to specific stages or levels of representation within the network; rather, these factors influence weight changes throughout the network over the course of learning.

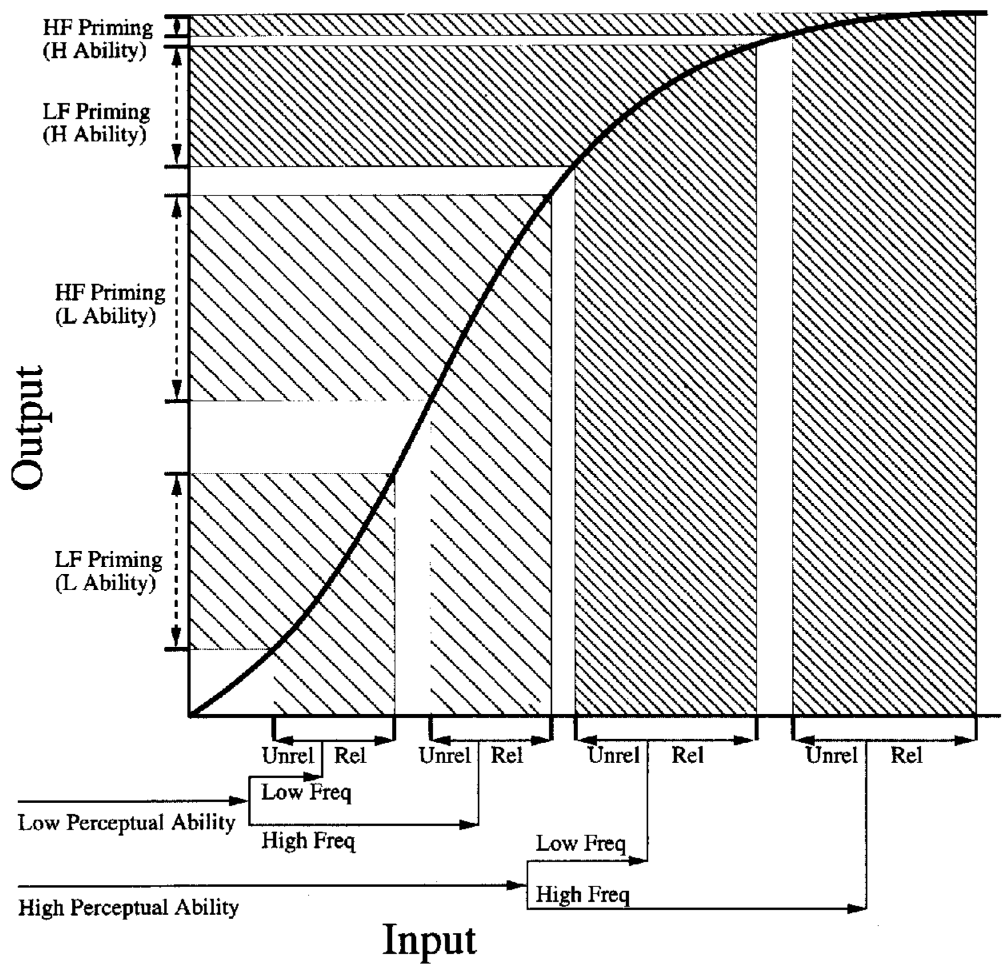

Figure 1. A depiction of how nonlinearities in the sigmoid activation function for semantic units in a distributed attractor network can give rise to greater priming (i.e., the difference in performance following related [Rel] versus unrelated [Unrel] primes) of low- versus high-frequency target words (LF and HF, respectively) for participants with high perceptual ability (H Ability; narrow-lined regions) but approximately equal priming for low- and high-frequency words for participants with low perceptual ability (L Ability; wide-lined regions). The combination of arrows at the bottom depicts the separate contributions of perceptual ability, target frequency (Freq), and priming context, which are summed together to form the input to a given semantic unit (indicated by the small vertical lines on the x-axis), to which the sigmoid function is applied to determine the activation of the unit. Note that relative magnitudes of these contributions are assumed to be greater for high- compared with low-ability participants, greater for high- compared with low-frequency targets, and positive for related primes but negative for unrelated primes (reflecting both facilitation and inhibition, respectively). Moreover, the magnitudes of the contributions of target frequency and priming context are assumed to be greater for high-ability participants because they can process both primes and targets more effectively than low-ability participants. The bottom portion of the sigmoid function is omitted for clarity.

The basis of the Frequency × Context interaction in the network's performance can be understood in terms of the nonlinear effects of the S-shaped sigmoid activation function that relates the input of each unit to its activation (see Figure 1). The use of a sigmoid or logistic function for units is standard in connectionist modeling and can be understood as optimal for tasks involving binary output patterns (Rumelhart, Durbin, Golden, & Chauvin, 1995).

In processing a target word, the RT of the network is taken to be the point at which the activations of all semantic units approach asymptote (either 0 or 1) and stop changing. This, in turn, depends on the magnitude of the input to the unit—stronger input (positive or negative) drives a unit to asymptote more quickly than weaker input.[5] One factor that influences the strength of the input to the semantic units for a given target word is the frequency with which

[5] Strictly speaking, the magnitude of the input to a semantic unit controls not only the time to reach asymptote but also the level of activation reached. We assume that the time required for interactions with other units to push the unit to an extreme activation value (1 or 0) depends primarily on the initial activation that would be produced by the strength of "bottom–up" input, as depicted in Figure 1.

the word was presented during training. In general, high-frequency words generate stronger input to the semantic units than do low-frequency words because, by being trained more often, they have a greater impact on the weights learned by the network. All else being equal, this stronger input causes the network to settle faster, and thus respond more quickly, to high- compared with low-frequency targets. Another relevant factor is associative relatedness. The network was trained to process the target when starting from the representation of an associated prime more often than when starting from the representation of any particular nonassociated prime, resulting in stronger input to semantic units, and thus faster responses, in a related versus unrelated priming context.

The magnitude of the effect of priming context is, however, modulated by target frequency (see the narrow-lined regions in Figure 1). Specifically, high-frequency words provide sufficient input to the semantic units to boost their activation near asymptote, leaving little room for priming context to have an additional effect on the units' output. By contrast, the input for low-frequency words remains closer to the linear region of the activation function, where further differences due to priming context are reflected more directly in the output of units. Thus, frequency and context interact in the activation of semantic units and, hence, in the settling time of the network, because of the "diminishing returns" of one factor when another factor is sufficiently strong on its own.[6]

With regard to developmental and individual differences, the current formulation of a distributed network account makes a similar prediction as the interactive compensatory model (Stanovich, 1980)—namely, that the Frequency × Context interaction should hold only for older and high-ability readers; younger and low-ability readers should show equal priming effects for high- and low-frequency words. This follows under the assumption that the overall strength of input is weaker for younger and low-ability readers compared with older and high-ability readers, such that both low- and high-frequency words fall within the linear range of the activation function (see the wide-lined regions in Figure 1).[7] In this case, the relative effects of related versus unrelated primes are roughly equivalent for both low- and high-frequency words.

There are fundamental differences, however, in how the distributed network model and the interactive compensatory model view the changes in performance between children and adults. An important implication of our single-mechanism account is that children and adults should differ only quantitatively. For example, because children have less reading experience than adults, they benefit from fewer learning episodes with any given word (i.e., their effective word frequencies are lower) and, hence, they are less effective at processing a prime. Consequently, a longer SOA in children may result in the same degree of activation in the semantic system as a shorter SOA in adults. Thus, on a distributed network account, priming effects in children at a long SOA would be expected to be similar to priming effects in adults at a short SOA. By contrast, a dual-mechanism model like the interactive compensatory model holds that children are qualitatively different from adults because priming effects in children result more from inhibitory expectancy-based processes, whereas priming effects in adults result more from spreading activation processes.

A distributed network approach may also be able to account for the occurrence and time course of facilitation and inhibition in priming, without invoking separate expectancy-based processes. As mentioned earlier, associative priming seems to result from facilitation at both short and long SOAs, whereas categorical priming seems to result from facilitation at short SOAs but from both facilitation and inhibition at long SOAs. In the Plaut (1995) model, associative priming is due to the increased frequency with which targets are preceded by associated versus nonassociated primes during training. Plaut (1995) showed that associative priming effects increase as the duration of the prime is lengthened, because the resulting pattern more closely approximates the representation of the prime that is associated with that of the target. By contrast, categorical priming effects, which are due to semantic feature overlap among category members, peak at a relatively short prime duration and then decrease with additional processing of the prime. This decrease is caused by the semantic units, including those that differ between the prime and target, being driven toward their asymptotic values. These differences take time to be reversed when the target is presented, thereby diminishing the advantage of starting with some overlapping features due to a categorically related prime.

In other words, at longer SOAs, the network exhibits a greater degree of *hysteresis* in moving from one stable state to another, even when those states share many active units. It is important to understand the basis for this effect because, as we argue later, it may explain the shift from facilitation dominance at short SOAs to inhibition dominance at long SOAs (C. A. Becker, 1980). Hysteresis in the network arises from the operation of *attractors* over semantic representations. During the course of training, the network learned to make particular patterns of semantic activity stable, such that unit interactions caused any activity pattern to be cleaned up into the nearest stable, attractor pattern (corresponding to the meaning of a word). This process can be conceptualized in terms of movement within a high-dimensional *state space* containing a dimension for each semantic unit (see Figure 2). Within this space, different semantic patterns correspond to different points, and the settling process consists of movement of the point for the initial pattern downhill into a bowl-like *basin of attraction* to the attractor point at the bottom of the basin (depicted in the figure by a set of concentric circles). When a target word follows a prime, the network must alter its activity pattern to move up and out of the attractor basin for the prime in order to settle to the bottom of the basin for the target. To the extent that the processing of the prime is particularly strong or prolonged (e.g., at a long SOA), the network requires more time to move out of the prime's basin of

[6] Directly analogous explanations based on the nonlinearity of the sigmoid function have been proposed by Cohen, Dunbar, and McClelland (1990) to account for the balance of facilitation and inhibition in Stroop tasks and by Plaut, McClelland, Seidenberg, and Patterson (1996) to account for the interaction in naming latencies of frequency and spelling–sound consistency (e.g., Seidenberg, Waters, Barnes, & Tanenhaus, 1984; Taraban & McClelland, 1987; Waters & Seidenberg, 1985), and for the three-way Frequency × Consistency × Imageability interaction (Strain, Patterson, & Seidenberg, 1995).

[7] The assumption that good and poor readers differ only in how strongly semantics is activated by orthography is clearly a simplification. It is likely that good and poor readers differ in many ways, including the quality of their semantic representations, but this is not the focus of our empirical or modeling work.

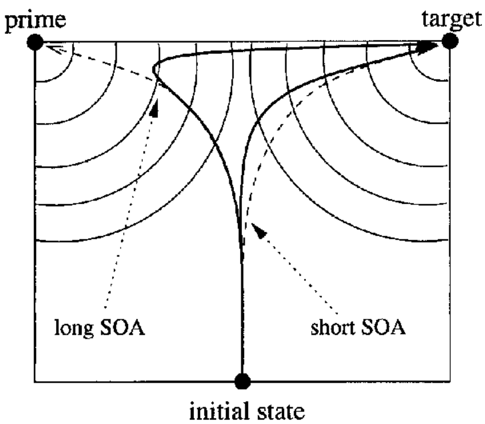

Figure 2. A depiction of the effect of stimulus-onset asynchrony (SOA) on categorical priming in terms of movement within a high-dimensional state space; for clarity, only two dimensions are depicted. The point corresponding to the network's current activity pattern moves over time as the network settles to a response. The trajectories of this point produced by processing of the prime or target in isolation (dashed arrows, partly occluded) are similar initially but eventually diverge. Thus, at a short SOA, processing of a related prime moves the current pattern toward the target, producing facilitation relative to the effects of a neutral prime (which can be thought of as placing the system near the middle of the space). By contrast, at a long SOA, the pattern moves sufficiently into the attractor basin of an unrelated prime that it has moved away from the target, producing inhibition relative to a neutral prime (and, perhaps, also facilitation for a related prime).

attraction, resulting in a slower RT to the target compared with when a prime produces only weak activation (e.g., at a short SOA or if the prime is a nonword). Note that associative priming, by contrast, is not subject to this hysteresis because the network learned to make associated prime–target transitions effectively because of their elevated frequency during training.

Although the pattern of results exhibited by the Plaut (1995) network is broadly consistent with the empirical findings, the relative magnitudes of facilitation and inhibition were not established explicitly by comparing RTs to target words following neutral primes (e.g., nonwords or *X*s) to those following related and unrelated primes. In fact, as far as we know, the computational model of semantic priming presented below is the first to be tested directly in this manner.

In summary, distributed network models suggest an account for individual and developmental differences in the interaction between priming context and target frequency. These models are fundamentally different from dual-mechanism models, which postulate a strategic, expectancy-based mechanism separate from an automatic, spreading-activation mechanism. Indeed, the defining characteristics of multiple-mechanism accounts are that they invoke processing mechanisms that are based on distinct sets of computational principles and could easily be stipulated independently of each other (i.e., a system could have spreading activation without retrospective matching or vice versa). By contrast, a distributed network constitutes a single mechanism because it instantiates a single, integrated set of computational principles. These principles may interact in complex ways to give rise to distinct patterns of performance in different contexts. However, even when it is useful to describe the operation of a network in terms of multiple influences or processes (e.g., sources of facilitation and inhibition), these processes are inextricably bound together and cannot be stipulated independently.

In the current work, we examined the degree to which the interaction between single-word priming context and target frequency depends on individual and developmental differences at long and short SOAs, and we present a distributed network model that accounts for the findings and makes additional empirical predictions. In light of the evidence summarized earlier that reading and language skills are strongly influenced by age and perceptual encoding ability, we examined whether these factors modulate the typical interaction found in adults between priming context and target frequency (C. A. Becker, 1979; Borowsky & Besner, 1993). Note that this investigation involved evaluating the joint impact of five factors: priming context, target frequency, SOA, age, and perceptual ability. Because it was infeasible to fully cross these factors in a single experiment, we conducted three empirical studies that each involved a subset of these factors. The first experiment examined differences in perceptual ability among college students at a long SOA (800 ms). The second experiment investigated ability and age differences among 3rd- and 6th-grade children at the same long SOA. The third experiment examined ability differences among college students at a short SOA (200 ms). We predicted that older and high-ability readers would exhibit an interaction between priming context and target frequency, whereas younger and low-ability readers would not. Furthermore, we expected different patterns of facilitation and inhibition for the adults versus the children. On the basis of previous results (C. A. Becker, 1980; Heyer et al., 1985; L. C. Smith et al., 1987), we expected adults to show facilitation at the short SOA but both facilitation and inhibition at the long SOA. By contrast, according to the distributed network account, children at the long SOA should show primarily facilitation (see Simpson & Lorsbach, 1983, 1987) whereas, according to the interactive compensatory account, they should show inhibition. Previous results have also established that the relative degree of facilitation and inhibition in priming is influenced by whether stimuli are related categorically or associatively and by the nature of the neutral priming condition. Although we did not vary these latter factors experimentally, we consider their impact on the results in the General Discussion.

Following the empirical experiments, a computational simulation of a distributed network model is presented. A network was trained to map orthographic representations of words onto semantic representations, including both associative relatedness (increased transition probabilities) and semantic relatedness (increased semantic feature overlap). Individual differences in perceptual ability were implemented by manipulating the strength of the orthographic input, and developmental differences were modeled by examining the performance of the network at different points in training. Word frequency was reflected in the frequency with which words were presented during training. Finally, SOA corresponded directly to the timing of prime and target presentation during testing. Simulation results support our claim that distributed network models offer a viable alternative to dual-mechanism models of semantic priming.

Experiment 1

Our primary purpose in Experiment 1, in which we used a long SOA (800 ms), was to test the prediction that, among college students, only those with high perceptual ability should show greater priming for low- compared with high-frequency targets. Given the long SOA, we also expected this interaction to result from both facilitation and inhibition.

Method

Participants

Ninety-four college students (*M* age = 20.9, *SD* = 5.7) at the University of Maryland participated to fulfill a psychology course requirement. All students had English as a first language and reported that their vision was normal or corrected to normal.

Apparatus

Participants viewed all stimuli for the priming task on a VGA monitor controlled by an IBM 286 computer with Micro Experimental Laboratories (MEL) software (Schneider, 1990). The participants controlled stimulus presentation and recorded their responses with a computer keyboard. MEL computed RTs by measuring the time lapse between the onset of the target word and the participant's response. MEL also recorded the error rates.

Materials and Design

The critical stimuli for the priming task, listed in Appendix A, were 120 prime–target pairs in each of three conditions: unrelated word prime and word target (e.g., EIGHT–BELOW), related word prime and word target (e.g., ABOVE–BELOW), and nonword prime and word target (e.g., KARBS–BELOW). Each type of prime–target condition had an equal probability of being presented to each participant (i.e., 40 trials). The nonword target pairs were 40 different word primes (e.g., HAPPY–GORPH) and 40 different nonword primes (e.g., ZENOX–AJUPE) paired with 80 different nonword targets.[8] These five conditions totaled 200 test pairs. Note that the word and nonword stimuli were not matched orthographically. For example, the nonwords have lower summed positional bigram and trigram frequencies (based on the Kučera & Francis, 1967, corpus) than the words: for bigrams (*M*s = 62.6 vs. 82.0, respectively), $F(1, 518) = 37.86$, $MSE = 1{,}106$, $p < .001$; for trigrams (*M*s = 6.3 vs. 11.9, respectively), $F(1, 518) = 74.50$, $MSE = 46.4$, $p < .001$. This difference is relevant to the design of the stimuli used in the computational simulation.

Each prime was presented in white lowercase letters on a black background and was followed by the target words presented in lowercase letters.[9] Targets were presented at an 800 ms SOA with a 200 ms interstimulus interval (ISI). The three conditions for the critical prime–target pairs were counterbalanced between participants. Specifically, a related prime, a nonword prime, and an unrelated prime preceded the same target word equally often across three different experimental lists. Because three counterbalancing lists were used, the same stimulus item was never seen by a single participant on more than one occasion. Within each list, the order of item presentation was randomized for each participant. There were also 30 practice trials that consisted of 6 pairs of the 5 prime–target conditions. The practice trials were excluded from all statistical analyses.

The strength of association between the related prime and target word ($M = .47$, $SD = .17$) was controlled by using established association norms (Nelson, McEvoy, & Schreiber, 1994), because controlling for this association enables a better understanding of priming effects on word recognition (C. A. Becker, 1980). All stimuli were also restricted to be five letters in length because demonstrating interactions of target frequency with priming context is more convincing if confounding factors such as word length are controlled for. In addition, target frequencies were chosen such that they were normally distributed after transforming frequency using the formula $40 + \log_{10}(f + 1)$, where f is the frequency of the target in Kučera and Francis (1967). This logarithmic transformation reduced the very large variability typical of word frequency counts (see Borowsky & Besner, 1993). Frequency was also dichotomized into high frequency ($M = 232.6$, $SD = 167.7$) and low frequency ($M = 30.7$, $SD = 20.5$) for ease of interpretation in the figures; however, all analyses of covariance (ANCOVAs) treated frequency as a continuous variable.

Nonword primes were used as neutral primes instead of repetitive stimuli like the word READY or a string of Xs because these latter stimuli may not engage the linguistic substrates involved in word recognition, and they may also lose their alerting qualities over repeated presentations (Antos, 1979; Jonides & Mack, 1984). RTs to target words following these repetitive primes may therefore be inflated, resulting in an underestimation of inhibition by unrelated primes and an overestimation of facilitation by related primes (see Neely, 1991, pp. 278–281, for a discussion of using nonwords as neutral primes). Note, however, that recently there has been controversy over whether nonword primes are, in fact, good neutral baselines (see, e.g., McKoon & Ratcliff, 1992; McNamara, 1994).

The Symbol Search Test of the Wechsler Intelligence Scale for Children (3rd ed.; WISC–III; Wechsler, 1991) was used as a measure of perceptual processing ability. The Symbol Search Test loads on the Processing Speed factor of the WISC–III. This paper-and-pencil task required each participant to correctly indicate as quickly and as accurately as possible whether either of two meaningless symbols to the left appeared in a row of five meaningless symbols to the right. This test consisted of 3 pages of 15 items each, and each page was timed from the moment in which the participants checked the first yes or no box to the moment in which they made their last response on that page. A speed–accuracy score was calculated for each participant by dividing his or her time needed to complete the test by his or her accuracy. All mean accuracy scores in Experiments 1–3 were more than 43 out of 45.

In addition, as a measure of vocabulary knowledge, the participants were given the revised Peabody Picture Vocabulary Test (PPVT–R). The

[8] The likelihood of a word target is .67 following word primes but only .5 following nonword primes. Consequently, participants could potentially derive and use information about the lexicality of the prime to bias lexical decisions to the target. The effect of this would presumably be to facilitate word responses following word primes (related or unrelated) relative to word responses following nonword primes, thereby underestimating inhibition relative to facilitation. This is not a problem in the current context because, as we report, participants exhibited clear inhibition despite this possible underestimation. Moreover, this concern is likely to apply only to adults tested at a long SOA (as in the current experiment) given that, at short SOAs, adults are generally insensitive to experimental manipulations that induce strategic effects (see Neely, 1991), and younger children are relatively insensitive to experimental factors, such as relatedness proportion and SOA, that influence priming effects in older children and adults (Simpson & Lorsbach, 1983). Thus, the effect should be negligible for children at a long SOA (Experiment 2) and for adults at a short SOA (Experiment 3).

[9] The target words were low-intensity in half of the trials and high-intensity in the other half. The low-intensity targets were dark gray letters presented on a black background, and the high-intensity targets were white letters presented on a black background. At each level of intensity, half of the targets were high-frequency and half were low-frequency. The target words in the related, unrelated, and nonword priming conditions were also counterbalanced between participants for low- and high-intensity. It turned out that the intensity manipulation produced no reliable effects in any of the current experiments, so all analyses were collapsed across target intensity.

PPVT–R was used because of the large vocabulary knowledge range anticipated. The PPVT–R has been normed for 2- to 33-year-olds (Dunn & Dunn, 1981). The PPVT–R is correlated ($rs > .60$) with other standardized vocabulary and verbal measures such as the WISC–III and the Wechsler Adult Intelligence Scale (WAIS; Wechsler, 1958).

The Symbol Search Test and PPVT–R were chosen specifically so as not to tap reading processes directly. We assumed that performance on the Symbol Search Test reflects the speed with which letters and words can be encoded by the orthographic system but that it measures it in a way that avoids confounds from other orthographic, lexical, or semantic factors. The PPVT–R was administered in order to control for higher level semantic and vocabulary knowledge.

Procedure

All participants were individually administered the Symbol Search Test, the PPVT–R, and the priming task in a room that was separated by two doors from any sound interference. The Symbol Search Test and the PPVT–R were always administered first. The testing period lasted approximately 60 min.

The priming task began with the experimenter reading instructions, which were presented on a computer monitor placed about 50 cm in front of the participant. At this distance, the 5-letter target words subtended about 1.5° of visual angle. The experimental session proceeded as follows. The participants were told that the first stimulus would be either a word or a nonword and that the second stimulus would also be a word or a nonword. The participants were then told to decide whether the second stimulus spells a word they know and to respond as accurately and quickly as possible by pressing the red key (*z*) on the keyboard with their left hand if the stimulus was not a word and by pressing the green key (*/*) with their right hand if the stimulus was a word. They were then told that they could control the rate at which each trial would be presented. Pressing the space bar would make the "Get Ready" indicator disappear and start the trial by causing a fixation cross (+) to be presented on the screen. The participants were asked to fixate on this cross and were told that after 2 s the first stimulus would appear for less than a second, and then the second stimulus would appear shortly thereafter. The participants were told that the second stimulus would remain on the screen until they responded. There was then a mandatory 2-s intertrial interval.

Results and Discussion

Participants were dichotomized into high- or low-perceptual-ability groups on the basis of their Symbol Search Test speed–accuracy score (see Table 1). The low-ability group scored significantly poorer than the high-ability group, $t(92) = 13.34, p < .05$. The vocabulary measure (PPVT–R) did not correlate significantly with the perceptual measure, suggesting that these instruments were measuring two distinct underlying abilities ($r = .09$). This independence was supported further by the finding of no significant differences in vocabulary knowledge between the high- and low-perceptual-ability groups ($|t| < 1$). Therefore, any differences in priming between the high- and low-perceptual-ability groups could not be due to vocabulary differences.

Table 1
Means and Standard Deviations for Age and for the Perceptual and Vocabulary Ability Measures for Experiment 1

Group	Age (in years)		Perceptual speed–accuracy		Vocabulary raw score	
	M	*SD*	*M*	*SD*	*M*	*SD*
High perceptual ability	20.1	4.3	2.2	0.2	159.2	9.7
Low perceptual ability	21.9	6.8	2.8	0.3	159.7	9.1

Note. For each group, $N = 47$. Perceptual speed–accuracy is mean time (in seconds) divided by number correct (out of 45) on the Symbol Search Test of the Wechsler Intelligence Scale for Children (3rd ed.). Vocabulary raw score is the mean raw score (out of 171) on the Peabody Picture Vocabulary Test—Revised.

In all subsequent ANCOVAs, perceptual ability (high vs. low) was a between-subjects factor, priming context (related vs. unrelated) was a within-item factor, and target frequency was a continuous between-items factor. Target frequency was treated as a continuous factor to more accurately reflect the underlying variable of frequency, because the range of this variable was limited by the selection of relatively high-frequency items with the expectation that a similar list of stimuli would be used for the developmental study in Experiment 2. Because of the limited frequency range, a dichotomous frequency manipulation would have been too weak to detect frequency differences. Because frequency was treated as a continuous factor, this meant that only item analyses could be computed, and therefore perceptual ability could not be treated as a continuous variable in the analyses. Note, however, that all figures treat target frequency as a dichotomous variable (high vs. low) for ease of interpretation. Planned *t* tests were then calculated to compare the related, unrelated, and nonword priming conditions in order to determine whether priming effects resulted from facilitation, inhibition, or both.

For the semantic priming task, an item with a correct mean RT of greater or less than 2.5 *SD*s from the overall mean in the related, unrelated, or nonword prime condition for either the high- or low-ability participants was eliminated (3 items, or 2.5%). Correct mean RTs or error rates for each word were then entered into a 2 (perceptual ability: high, low) × 2 (priming context: related, unrelated) ANCOVA with target frequency as a continuous regressor variable. The mean RTs are shown in Figure 3, and the numeric values of mean RTs and error rates for this and subsequent experiments are listed in Appendix B. The RT analysis yielded a main effect of perceptual ability, $F(1, 468) = 227.50$, $MSE = 811{,}834$, $p < .001$; priming context, $F(1, 468) = 12.02$, $MSE = 42{,}897$, $p < .001$; and target frequency, $F(1, 468) = 17.94$, $MSE = 64{,}031$, $p < .001$. However, these main effects were qualified by a significant three-way Perceptual Ability × Priming Context × Target Frequency interaction, $F(1, 468) = 4.47$, $MSE = 15{,}947$, $p < .05$. As predicted, for high-ability participants, priming context influenced the recognition of low-frequency targets (priming difference $d = 33$ ms) more than the recognition of high-frequency targets ($d = 7$ ms). By contrast, for low-ability participants, priming context influenced the recognition of high- and low-frequency targets to a similar degree (ds = 20 and 17 ms, respectively). As suggested in the introduction, the observed interaction among perceptual ability, priming context, and target frequency can be understood in terms of the asymptotic response of semantic units within a distributed network model (see Figure 1).

Planned *t* tests were calculated to compare the related, unrelated, and nonword prime conditions in order to determine whether the priming effects resulted from facilitation, inhibition, or both (see Figure 3). Facilitation is indicated by faster recognition of a target following a related prime compared with a nonword prime,

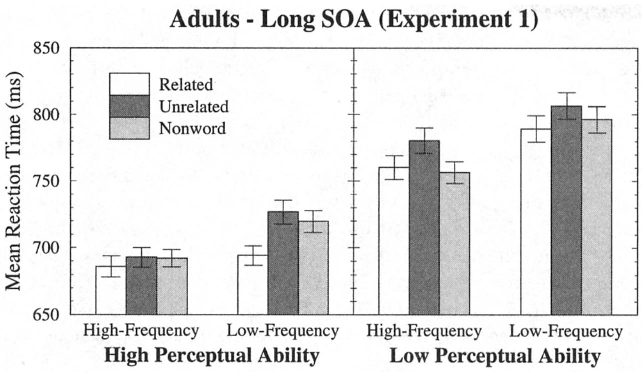

Figure 3. Item means for Experiment 1 of correct mean reaction times to high- and low-frequency target words following related, unrelated, and nonword primes (800-ms stimulus-onset asynchrony [SOA]), for high- and low-perceptual-ability college students. Error bars indicate 1 standard error.

whereas inhibition is indicated by slower recognition of a target following an unrelated prime compared with a nonword prime (see Neely, 1991). The results indicated that there were no significant differences among priming conditions for the high-ability participants reading high-frequency words or for the low-ability participants reading low-frequency words ($ps > .05$). However, the related prime condition was faster than both the unrelated and nonword prime conditions for the high-ability participants reading low-frequency words, and the unrelated prime condition was slower than the nonword prime condition for the low-ability participants reading high-frequency words ($ps < .05$). These effects suggest that priming resulted from facilitation as well as inhibition in this group of participants. The role of inhibition was further supported by the finding that RTs to target words were numerically slower after unrelated primes compared with nonword primes for all participants regardless of their perceptual ability and for all targets regardless of their frequency. These findings support models of word recognition that predict that priming effects should result from facilitation and inhibition at long SOAs. As discussed in the introduction, both the dual-mechanism and distributed network models predict this result, but they do so in very different ways (see Neely, 1977; Plaut, 1995; Stanovich, 1980).

The error analysis yielded a significant main effect of priming context, $F(1, 468) = 8.46$, $MSE = 82.8$, $p < .001$, but this effect was qualified by an interaction between priming context and target frequency, $F(1, 468) = 6.02$, $MSE = 59.0$, $p < .05$. The difference between the related and unrelated prime conditions was larger for low-frequency targets (1.0% vs. 2.5%, respectively) than for high-frequency targets (1.2% vs. 1.4%, respectively). This effect was produced primarily by facilitation through related primes and not by inhibition through unrelated primes, because the error rates for the unrelated prime conditions were less than 0.4% different from the error rates for the nonword prime conditions. In contrast to the RT results, the interaction between priming context and target frequency was not modulated by differences in perceptual ability, probably because the mean accuracy rates were above 96% for all ability levels.

For reasons given later, in Experiment 2 (with children) we used only 72 of the items from the current experiment. To enable direct comparison, we calculated an additional ANCOVA, parallel to the one reported earlier, using only these items. This RT analysis yielded essentially the same results as the 120-item analysis (see Figure 4). There was a main effect of perceptual ability, $F(1, 276) = 141.13$, $MSE = 428{,}306$, $p < .001$, and priming context, $F(1, 276) = 33.79$, $MSE = 102{,}553$, $p < .001$, and a trend for target frequency, $F(1, 276) = 3.16$, $MSE = 9{,}593$, $p = .07$. These main effects were qualified by a significant three-way Perceptual Ability $\times$ Priming Context $\times$ Target Frequency interaction, $F(1, 276) = 3.41$, $MSE = 17{,}303$, $p < .05$. In addition, the patterns of facilitation and inhibition differences were similar for the analyses based on the 72 items and the full 120 items.

In summary, when adults performed lexical decisions to target words following primes at a long SOA (800 ms), only those with high perceptual ability—as assessed by a match-to-sample pretest—exhibited the standard finding of greater priming for low-frequency targets compared with high-frequency targets. Participants with low perceptual ability exhibited equivalent levels of priming for low- and high-frequency targets. Moreover, by comparison with a nonword prime baseline, these effects were due to a combination of facilitation from related primes and inhibition from unrelated primes.

Experiment 2

The main purpose of Experiment 2 was to determine whether the priming differences observed among college students varying in perceptual ability in Experiment 1 could be replicated in a population of 3rd- and 6th-grade children. We expected that grade-school students with high perceptual ability would use priming context more for recognizing low-frequency words than for recognizing high-frequency words, whereas those with low perceptual ability would not exhibit an interaction between priming context and target frequency.

A second goal of Experiment 2 was to investigate whether priming context effects in grade-school students result from facilitation, inhibition, or both. Based on the Simpson and Lorsbach (1983) study discussed in the introduction, we expected that the priming effects exhibited by grade-school students at a long (800

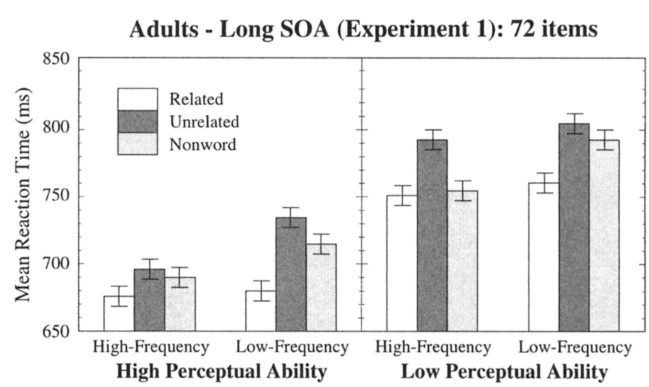

Figure 4. Mean reaction times in Experiment 1 for the 72 items that were also presented to children in Experiment 2, for both high- and low-frequency target words following related, unrelated, and nonword primes (800-ms stimulus-onset asynchrony [SOA]), for high- and low-perceptual-ability college students. Error bars indicate 1 standard error.

ms) SOA would reflect mostly facilitation rather than inhibition. Specifically, we predicted that priming effects in both high- and low-ability grade-school students would be due to faster RTs to target words following related primes compared with nonword primes, but that there would be no difference between RTs to target words following unrelated primes compared with nonword primes. As discussed in the introduction, our prediction that children should exhibit smaller inhibitory effects than adults is in direct contrast to the prediction of the dual-mechanism, interactive compensatory model (Stanovich, 1980).

Method

Participants

Forty-four 3rd-graders (*M* age = 8.9, *SD* = 0.5) and forty-six 6th-graders (*M* age = 11.8, *SD* = 0.5) from 2 private schools in the metropolitan Washington, DC, area participated in this study. All participants spoke English as a first language and reported that their vision was corrected to normal.

Apparatus

The experimental apparatus was the same as that used in Experiment 1.

Materials and Design

Because of limits on the amount of time children could miss classroom activities, the priming task used in Experiment 2 was shortened for use with the grade-school students based on college student performance in Experiment 1. Specifically, items were eliminated if their mean RT for the adults in Experiment 1 fell greater than 2.0 *SD*s above the correct mean RT for the related prime (7 items), unrelated prime (9 items), or nonword prime (9 items) conditions. Items were also eliminated if the correct mean RT to a nonword target fell greater or less than 2.5 *SD*s above the correct mean RT for the word prime (13 items) or nonword prime (19 items) conditions. Items with slow RTs were eliminated because they might have been too difficult for the grade-school students. Items were also eliminated if their correct mean RT in the related prime condition was more than 10 ms greater than their correct mean RT in the unrelated prime condition (see Appendix A). Because we were interested in the effect of a reliable priming context on the recognition of words varying in target frequency, removing items that did not yield a reliable priming effect was justified (Borowsky & Besner, 1993).

The resulting critical stimuli for the priming task in Experiment 2 were 72 prime–target pairs in each of 3 conditions: unrelated prime and word target pairs, related prime and word target pairs, and nonword prime and word target pairs. There were also 24 word prime and nonword target pairs and 24 nonword prime and nonword target pairs. These five conditions totaled 120 test pairs. All other aspects of the priming task were the same as in Experiment 1.

The strength of association between the related prime and target word in Experiment 2 ($M = .48$, $SD = .18$) was very similar to that in Experiment 1 ($M = .47$, $SD = .17$). Frequency was dichotomized into high frequency ($M = 242.2$, $SD = 173.0$) and low frequency ($M = 37.9$, $SD = 25.6$). The high- and low-frequency mean values in Experiment 2 were less than 10 per million greater than the same values in Experiment 1 (Kučera & Francis, 1967). The similarity of these values makes it possible to compare results across the experiments, even though a different set of items was used.

The grade-school students were also administered the PPVT–R and the Symbol Search Test (see Experiment 1).

Procedure

The testing procedure was the same as in Experiment 1. All participants were tested within their school building during regularly scheduled school hours.

Results and Discussion

Participants were divided into a 3rd-grade group or 6th-grade group and, independently of age, into a high- or low-perceptual-ability group on the basis of their Symbol Search Test speed–accuracy score (see Table 2). The low-ability group scored significantly poorer on the perceptual ability measure than the high-ability group, $t(88) = 11.99$, $p < .001$, and the 3rd graders scored significantly poorer on the perceptual ability measure than the 6th graders, $t(88) = 9.17$, $p < .001$. The latter correlation of age and perceptual ability precluded treating these measures as independent factors in a crossed design; instead, the effect of each was evaluated in a separate analysis. By contrast, the PPVT–R was only marginally correlated with the perceptual measure partialed for age ($r = -.19$, $p = .07$), and the high- and low-perceptual-ability groups did not differ reliably after effects of age were partialed out of the PPVT–R scores ($t < 1$). These results suggest that the Symbol Search Test and the PPVT–R measure distinct underlying abilities and that ability differences in priming should be interpreted as resulting from perceptual ability and not from vocabulary knowledge.

Table 2
Means and Standard Deviations for Age and for the Perceptual and Vocabulary Ability Measures for Experiment 2

		Age (in years)		Perceptual speed–accuracy		Vocabulary raw score	
Group	*N*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Sixth graders	46	11.8	0.5	3.2	0.5	133.7	12.1
Third graders	44	8.9	0.5	4.4	0.8	110.0	12.2
High perceptual ability	45	11.3	1.3	3.1	0.4	129.5	15.1
Low perceptual ability	45	9.4	1.2	4.5	0.7	114.8	15.7

Note. Perceptual speed–accuracy is mean time (in seconds) divided by number correct (out of 45) on the Symbol Search Test of the Wechsler Intelligence Scale for Children (3rd ed.). Vocabulary raw score is the mean raw score (out of 171) on the Peabody Picture Vocabulary Test—Revised.

Effects of Perceptual Ability

For the semantic priming task, all items with a correct mean RT of greater than 2.5 *SD*s from the overall mean in the related, unrelated, or nonword prime condition for either the high- or low-ability participants were eliminated (1 item, or 1.3%). Correct mean RTs or error rates for each word were entered into a 2 (perceptual ability: high, low) × 2 (priming context: related, unrelated) ANCOVA, with target frequency as a continuous regressor variable (see Experiment 1 for description of the ANCOVA design and inferential analysis strategy). Figure 5 shows the mean RT for each condition. The RT analysis yielded significant main effects of perceptual ability, $F(1, 284) = 148.09$, $MSE = 1{,}422{,}020$, $p < .001$; priming context, $F(1, 284) = 12.00$, $MSE = 115{,}210$, $p < .001$; and target frequency, $F(1, 284) = 17.80$, $MSE = 170{,}879$, $p < .001$. However, these main effects were qualified by a significant interaction among priming context, target frequency, and perceptual ability, $F(1, 284) = 3.54$, $MSE = 24{,}414$, $p < .05$. As predicted, only high-ability participants showed larger priming effects for low-frequency words ($d = 52$ ms) than for high-frequency words ($d = 17$ ms). Low-ability participants showed similar priming effects for both high- and low-frequency words ($ds = 47$ and 46 ms, respectively). This replicates the Perceptual Ability × Priming Context × Target Frequency interaction found for the college students in Experiment 1.

The related, unrelated, and nonword prime conditions were then compared with planned *t* tests in order to determine whether the priming context effects resulted from facilitation, inhibition, or both (see Figure 5). There were no significant priming context differences for the high-ability participants reading high-frequency words ($p > .05$). However, the related prime condition was faster than the unrelated and nonword prime conditions for the high-ability participants reading low-frequency words and for the low-ability participants reading both low- and high-frequency words ($ps < .05$). By contrast, there were no significant differences between the unrelated and nonword priming contexts in any of the conditions. These results suggest that priming effects resulted from facilitation to related target words and not from inhibition to unrelated target words. Consistent with this, Simpson and Lorsbach (1983, 1987) found no evidence for inhibition in 2nd- through 6th-grade students at a low relatedness proportion (25%; cf. 20% in the current experiment). As mentioned above, the interactive compensatory model predicts that younger and low-ability readers should use inhibitory expectancy-based processes to a greater degree than older and high-ability readers (Stanovich, 1980). This prediction is disconfirmed by our findings and those of Simpson and Lorsbach. The findings are more compatible with a distributed network account that predicts weaker inhibition earlier compared with later in training.

The error analysis yielded significant main effects of perceptual ability, $F(1, 284) = 21.35$, $MSE = 522.4$, $p < .001$; priming context, $F(1, 284) = 4.58$, $MSE = 112.1$, $p < .05$; and target frequency, $F(1, 284) = 6.75$, $MSE = 165.3$, $p < .05$. High-ability participants had lower error rates than low-ability participants (1.8% vs. 4.5%), a related priming context resulted in lower error rates than an unrelated priming context (2.5% vs. 3.8%), and high-frequency targets had lower error rates than low-frequency targets (2.1% vs. 4.3%). Whereas the college students in Experiment 1 exhibited a significant interaction between priming context and target frequency for error rates, the grade-school students in the current experiment did not. This difference suggests that the Frequency × Context interaction for error rates becomes stronger with age and supports the prediction that the interaction between priming context and target frequency should be stronger for older and high-ability readers than for younger and low-ability readers (see Figure 1).

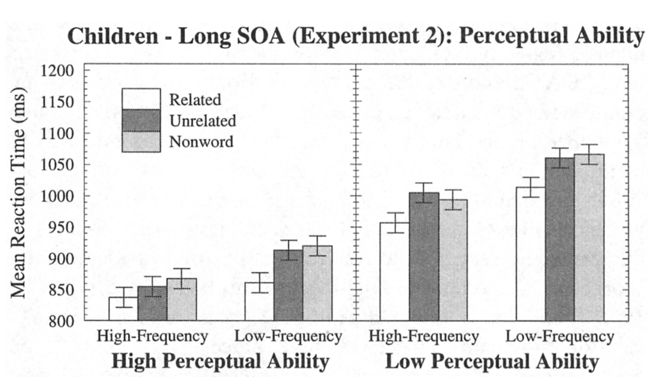

Figure 5. Item means for Experiment 2 of correct mean reaction times to high- and low-frequency target words following related, unrelated, and nonword primes (800-ms stimulus-onset asynchrony [SOA]), for high- and low-perceptual-ability grade school students. Error bars indicate 1 standard error.

Effects of Age

So far we have only considered differences in priming related to perceptual ability. It was not possible to analyze the effects of perceptual ability and age jointly without reducing the number of participants in each cell of the design to unacceptable levels. Therefore, in order to examine age differences, correct mean RTs or error rates for each word were entered into a 2 (age: 3rd-grade, 6th-grade) × 2 (priming context: related, unrelated) ANCOVA, with target frequency as a continuous regressor variable. Figure 6 shows the mean RT for each condition. The RT analysis revealed significant main effects of age, $F(1, 284) = 187.59$, $MSE = 1{,}805{,}640$, $p < .001$; priming context, $F(1, 284) = 12.08$, $MSE = 116{,}267$, $p < .001$; and target frequency, $F(1, 284) = 17.59$, $MSE = 169{,}347$, $p < .001$. The Age × Priming Context × Target Frequency interaction indicated a weak trend toward significance, $F(1, 284) = 1.86$, $MSE = 17{,}931$, $p = .17$. As predicted, older participants tended to show larger priming effects for low-frequency words ($d = 43$ ms) than for high-frequency words ($d = 16$ ms). Younger participants tended to show more similar priming effects for both high- and low-frequency words ($ds = 47$ and 57 ms, respectively).

Although there was not a significant interaction between age and priming context, the trends were in the predicted direction. The 6th-grade children tended to show less priming than the 3rd-grade children ($ds = 29$ and 52 ms, respectively). This finding is consistent with the literature showing that younger children exhibit larger priming effects than older children (e.g., Schwantes, 1985, 1991; Stanovich, 1980; West & Stanovich, 1978).

The error analysis revealed significant main effects of age, $F(1, 284) = 26.01$, $MSE = 677.5$, $p < .001$; priming context, $F(1, 284)$

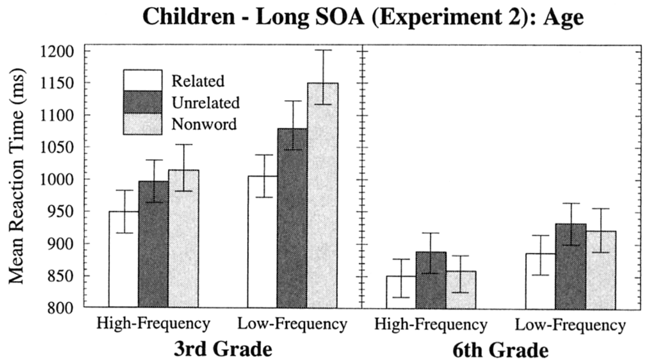

Figure 6. Item means for Experiment 2 of correct mean reaction times to high- and low-frequency target words following related, unrelated, and nonword primes (800-ms stimulus-onset asynchrony [SOA]), for 3rd- and 6th-grade students. Error bars indicate 1 standard error.

= 4.87, MSE = 126.9, $p < .05$; and target frequency, $F(1, 284)$ = 6.10, MSE = 158.9, $p < .05$. Sixth-grade participants had lower error rates than 3rd-grade participants (1.7% vs. 4.8%, respectively; see the analysis of perceptual ability above for mean differences in priming context and target frequency effects).

Comparison With Adults

A direct comparison of the performance of the adults in Experiment 1 with the children in Experiment 2 is complicated by the substantial difference in perceptual ability between the two groups (see Tables 1 and 2). However, as an approximation to controlling for this factor, it is possible to compare the adults with low perceptual ability (mean score of 2.8 on the Symbol Search Test) with the children with high perceptual ability (mean score of 3.1; recall that a lower score indicates better performance). We carried out a 2 × 2 analysis of variance (ANOVA) of RTs to the 72 target words used in both experiments, comparing age (adults vs. children) and context (unrelated vs. nonword primes), collapsing across target frequency. There was, of course, a main effect of age, $F(1, 276)$ = 118.2, MSE = 768,995, $p < .001$, but, more importantly, also a reliable Age × Context interaction, $F(1, 276)$ = 3.88, MSE = 22,031, $p < .05$, such that the difference between the unrelated and nonword prime conditions was less for the children (d = −10 ms) than for the adults (d = 17 ms). This means that children show less inhibition than adults when the two groups have comparable perceptual ability. Note that the source of this difference is unclear, given that adults and children presumably differ on a wide variety of other factors that were not controlled in the current work.

This latter point is relevant for interpreting another comparison between adults and children that, on the surface, appears to be inconsistent with our theoretical account. The account predicts that, all else being equal, groups with lower perceptual ability should show a weaker Frequency × Context interaction. However, even though children overall have lower perceptual ability than adults, they show a trend toward a Frequency × Context interaction of comparable magnitude to that shown by adults in Experiment 1 (in the 72-item analysis). Specifically, the children showed more priming for low-frequency targets (d = 50 ms) than for high-frequency targets (d = 32 ms); the adults showed more priming for the low-frequency targets (d = 49 ms) than for the high-frequency targets (d = 30 ms). We merely note that this comparison is difficult to interpret because, as just pointed out, "all else" is not equal among children and adults.

In summary, like the adults in Experiment 1, when tested at a long SOA, the children in Experiment 2 exhibited greater priming for low- compared with high-frequency words only when they were of high perceptual ability. The low-ability children showed equal levels of priming regardless of target frequency. Moreover, by comparison with a nonword priming baseline, the children showed evidence of weaker inhibition than the adults, contrary to the predictions of the interactive compensatory model but consistent with those of distributed network models. To our knowledge, no other developmental studies have used the single-word priming paradigm to determine whether the interaction between priming context and target frequency varies as a function of age; the few that have investigated this issue (e.g., Stanovich et al., 1981) have used a sentential priming paradigm. In addition, Experiment 2 revealed a nonsignificant trend for an interaction among age, priming context, and target frequency. Age may not have explained a significant amount of variance in the interaction between priming context and target frequency because of individual differences in perceptual ability within the 3rd and 6th graders (see Table 2). An important implication of this finding is that developmental researchers should be careful when using age as an independent variable, at least in studies of word recognition. Because adults as well as children vary to a large degree in many different ways, one must consider individual differences in performance when interpreting developmental effects (see Epelboim, Booth, & Steinman, 1994, 1996).

Experiment 3

Considerable empirical work has focused on differences in the patterns of priming that result at short versus long SOAs. Experiment 1 established that perceptual ability in college students modulates the interaction between priming context and target frequency at a long (800 ms) SOA. Our first goal in Experiment 3 was to determine whether the same pattern of modulation holds also at a shorter (200 ms) SOA.

Our second goal in Experiment 3 was to determine whether the relative degree of facilitation and inhibition differs at short versus long SOAs. Inhibitory effects that hold only at long SOAs are often ascribed to strategic, expectancy-based processes. Although a number of previous studies have shown a predominance of facilitation at short SOAs (e.g., C. A. Becker, 1980; Heyer et al., 1985; L. C. Smith et al., 1987), none of them used a nonword prime baseline. Our expectation, based on these earlier findings, is that priming effects should result primarily from facilitation at a short SOA. According to dual-mechanism models (Neely, 1977, 1991), inhibition is absent at short SOAs because there is insufficient time to deploy expectancy-based processes. By contrast, on our distributed network account, there is insufficient time at a short SOA for the prime to settle very deeply into its attractor basin; hence, there is little if any hysteresis in moving on to the representation of the target. Indeed, this was how we explained the weakened inhibition at a long SOA shown by the children in Experiment 2 compared with the adults at the same SOA in

Experiment 1. In essence, we predict that adults at a short SOA should show similar priming effects to children at a long SOA.

Method

Participants

Fifty-three college students (M age = 19.1, SD = 2.1) at the University of Maryland participated to fulfill a psychology course requirement. All participants spoke English as a first language and reported that their vision was corrected to normal.

Apparatus

The experimental apparatus was the same as Experiment 1.

Materials and Design

The materials and experimental design were the same as those used in Experiment 1, except that the SOA was 200 ms with a 100-ms ISI. This ISI was chosen as a compromise between maintaining the absolute value of ISI (decreased by 50%) and maintaining the relative proportion of SOA occupied by ISI (increased by 50%).

Procedure

The procedure used was the same as in Experiment 1.

Results and Discussion

Participants were dichotomized into a high- or low-perceptual-ability group on the basis of their Symbol Search Test speed–accuracy score (see Table 3). The low-perceptual-ability group scored significantly worse on the perceptual-ability measure than the high-perceptual-ability group, $t(53) = 9.62$, $p < .001$. The vocabulary measure (PPVT–R) did not correlate significantly with the perceptual measure ($r = .17$), suggesting that these instruments were measuring two distinct underlying abilities. This independence was supported further by the finding of no significant vocabulary ability differences between high- and low-perceptual-ability groups ($|t| < 1$). These results indicate that any priming differences between the perceptual ability groups cannot be due to vocabulary differences.

For the semantic priming task, an item was eliminated from the analyses if it had a mean RT of greater or less than 2.5 *SD*s from the overall mean in the related, unrelated, or nonword prime condition for either the high- or low-perceptual-ability participants (2 items, or 1.6%). Correct mean RTs or error rates for each remaining word were entered into a 2 (perceptual ability: high, low) × 2 (priming context: related, unrelated) ANCOVA, with target frequency as a continuous regressor variable (see Experiment 1 for a description of the ANCOVA design and inferential analysis strategy). The error analysis yielded no significant main effects or interactions, so this analysis is not reported here. The mean RT for each condition is shown in Figure 7. The RT analysis yielded significant main effects of perceptual ability, $F(1, 472) = 142.24$, $MSE = 763{,}699$, $p < .001$; priming context, $F(1, 472) = 17.88$, $MSE = 96{,}005$, $p < .001$; and target frequency, $F(1, 472) = 25.19$, $MSE = 135{,}262$, $p < .001$. However, these main effects were qualified by a significant three-way Perceptual Ability × Priming Context × Target Frequency interaction, $F(1, 472) = 4.35$, $MSE = 23{,}356$, $p < .05$. As predicted, only high-ability participants showed larger priming effects for low-frequency targets (d = 52 ms) than for high-frequency words (d = 17 ms). Low-ability participants showed similar priming effects for high- and low-frequency targets (ds = 27 and 22 ms, respectively). In this way, Experiment 3, with a 200-ms SOA, replicated Experiments 1 and 2 with adults and children at long SOAs (800 ms).

An additional ANCOVA, parallel to the one reported above, was calculated with only 72 items, because only these items were used in Experiment 2 with the children (see Figure 8). The RT analysis yielded essentially the same results as the 120-item analysis. There was a main effect of perceptual ability, $F(1, 272) = 88.51$, $MSE = 481{,}932$, $p < .001$; priming context, $F(1, 272) = 16.83$, $MSE = 91{,}632$, $p < .001$; and target frequency, $F(1, 272) = 11.02$, $MSE = 60{,}005$, $p < .01$. These main effects were qualified by a significant three-way Perceptual Ability × Priming Context × Target Frequency interaction, $F(1, 272) = 3.98$, $MSE = 21{,}686$, $p < .05$. In addition, the patterns of facilitation and inhibition differences reported below were similar for the analyses based on the 72 items and the full 120 items.

Planned t tests were then calculated to compare the related, unrelated, and nonword priming conditions in order to determine whether the priming effects resulted from facilitation, inhibition, or both (see Figure 7). The results indicated that the nonword priming condition was slower than the related and unrelated priming conditions for the high- and low-ability participants reading high-frequency words (ps < .05). The related priming condition was faster than both the unrelated and nonword priming condition for the high-ability participants reading low-frequency words, and the related prime condition was faster than the nonword prime condition for the low-ability participants reading low-frequency

Table 3
Means and Standard Deviations for Age and for the Perceptual and Vocabulary Ability Measures for Experiment 3

		Age (in years)		Perceptual speed–accuracy		Vocabulary raw score	
Group	*N*	*M*	*SD*	*M*	*SD*	*M*	*SD*
High perceptual ability	26	18.9	2.1	2.1	0.1	154.1	13.8
Low perceptual ability	27	19.3	2.1	2.6	0.2	156.0	11.0

Note. Perceptual speed–accuracy is mean time (in seconds) divided by number correct (out of 45) on the Symbol Search Test of the Wechsler Intelligence Scale for Children (3rd ed.). Vocabulary raw score is the mean raw score (out of 171) on the Peabody Picture Vocabulary Test—Revised.

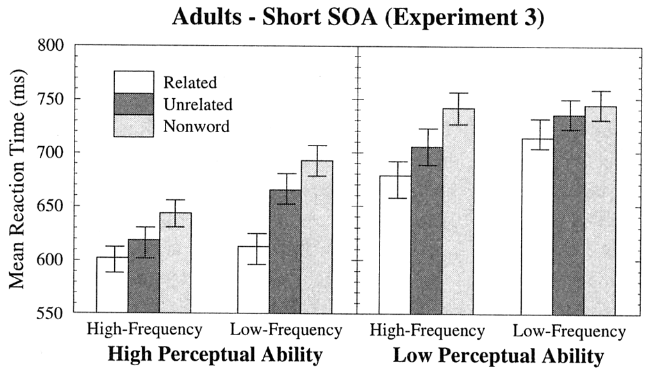

Figure 7. Item means for Experiment 3 of correct mean reaction times to high- and low-frequency target words following related, unrelated, and nonword primes (200-ms stimulus-onset asynchrony [SOA]), for high- and low-perceptual-ability college students. Error bars indicate 1 standard error.

words (*ps* < .05). In addition, across conditions, performance following unrelated primes was generally faster than performance following nonword primes. Thus, the recognition of target words was facilitated by related primes but not inhibited by unrelated primes. These results, taken together with those of Experiment 1, support earlier empirical findings (C. A. Becker, 1980; Heyer et al., 1985; Neely, 1977, 1991; L. C. Smith et al., 1987) that only facilitation operates at short SOAs but both facilitation and inhibition operate at long SOAs.

Computational Model

The central question that motivated Experiments 1–3 is this: What underlies age and ability differences in the use of priming context for influencing visual word recognition? Our experiments revealed that individuals with high perceptual ability showed larger single-word priming effects for low-frequency targets than for high-frequency targets, but that individuals with low perceptual ability showed equal priming effects for high- and low-frequency targets. These findings were very consistent and robust, holding for both children and adults and, for the latter, at both long and short SOAs. We also found that the priming effects in adults resulted from only facilitation at a short SOA but from both facilitation and inhibition at a long SOA. By contrast, children showed little if any inhibition at the same SOA at which the adults showed strong inhibition (see also Simpson & Lorsbach, 1983).

We have argued that a distributed network model of semantic priming can account both for the three-way Priming Context × Target Frequency × Perceptual Ability interaction and for the patterns of facilitation and inhibition across age and SOA. If so, such a model would provide a more parsimonious account than dual-mechanism models (e.g., C. A. Becker, 1980; Neely & Keefe, 1989; Stanovich, 1980), which must invoke distinct automatic, spreading-activation processes and strategic, expectancy-based processes, and yet still have difficulty explaining the lack of inhibition for children at long SOAs. However, to this point, our claim is based solely on verbal characterizations of rather complex properties of distributed networks, only some of which have been demonstrated in existing simulations (e.g., the Frequency × Context interaction shown by Plaut, 1995). To substantiate our account, and to demonstrate that a distributed network model can, in fact, exhibit the behavior we are ascribing to it, we developed a computational simulation of semantic priming in lexical decision and applied it to the empirical findings from Experiments 1–3. The approach taken is closely related to the one used in the Plaut (1995) simulation.

Method

Stimuli

The actual stimuli used in the experiments were not used in the simulation because of the complexity of their orthographic structure and because it was not feasible to derive realistic semantic representations for them. Rather, the network was trained on an abstract version of the task of mapping orthography to semantics. Although this task was simplified relative to the realistic mapping, it retained what we claim to be its most important property: that similarity in orthographic form is unpredictive of similarity in meaning.

Orthographic representations consisted of three-letter sequences constructed from 10 consonants (B, D, K, L, M, N, P, R, S, T) and 5 vowels (A, E, I, O, U). Letters were described in terms of six possible binary "features" such that each letter was assigned two of the six features. No attempt was made to assign similar codes to visually similar letters; codes were assigned to letters in alphabetic order. Words were restricted to consonant–vowel–consonant (CVC) strings; of the 500 possible CVC strings, 128 were chosen randomly to constitute the orthographic inputs on which the network was trained. Nonwords, by contrast, were restricted to VCV strings, with 128 chosen randomly out of the 250 possible strings. Nonwords were given orthographic representations that differed systematically from those of words because, in the empirical studies, the word and nonword stimuli were not matched orthographically (see Experiment 1 *Method* section) and included many orthographically unwordlike nonwords (see Appendix A). Note, however, that because vowel and consonant letters share features, there is some overlap between the orthographic representations for words and nonwords—just less than the overlap among the words themselves. Specifically, the average number of shared features among words was 2.22 (SD = 1.16), whereas the average number of shared features between words and nonwords was 1.87 (SD = 0.4), $t(16, 382) = 22.1$, $p < .001$. In fact, as a measure of relative word–nonword similarity, the ratio of these values, 1.19, is quite similar to the ratio of the

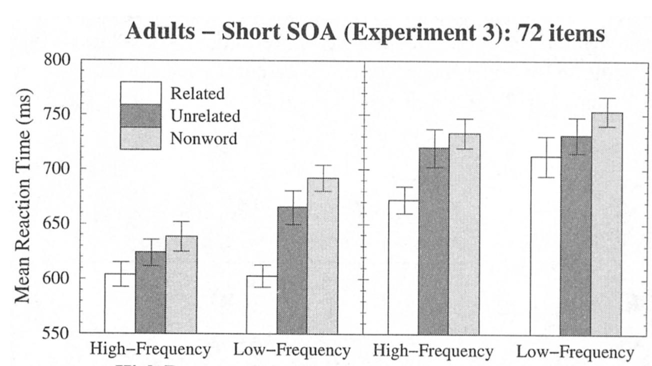

Figure 8. Mean reaction times in Experiment 3 for the 72 items that were also presented to children in Experiment 2, for both high- and low-frequency target words following related, unrelated, and nonword primes (200-ms stimulus-onset asynchrony [SOA]), for high- and low-perceptual-ability college students. Error bars indicate 1 standard error.

summed positional bigram frequencies of the words and nonwords used in the empirical studies, 82.0/62.6 = 1.31.

The semantic representations of words were the same as those used by Plaut (1995). They were generated to cluster into artificial semantic "categories."[10] Eight different random binary patterns were generated over 100 semantic features, in which each feature had a probability $p_a = .1$ of being active. These patterns served as the prototypes for eight separate semantic categories. Sixteen exemplars were generated from each prototype pattern by randomly altering some of its features (Chauvin, 1988). Eight of these were high-dominance exemplars in which relatively few features of the prototype were changed (each feature had a probability of .2 of being resampled with $p_a = .1$). The remaining eight were low-dominance exemplars in which more features were altered (resampling probability of .4). In addition, all pairs of patterns were constrained to differ by at least four features. The effect of this manipulation is to make all exemplars within a category cluster around the prototype, with high-dominance exemplars more similar to the category prototype than low-dominance exemplars, and for all semantic patterns to have an average of 10 active features (range = 4–18) out of a total of 100. Although the effect of target category dominance was not explored in the current work, the Plaut (1995) simulation exhibited greater semantic priming of high- compared with low-dominance targets, in keeping with empirical findings (Lorch, Balota, & Stamm, 1986; Schwanenflugel & Rey, 1986).

Semantic patterns were assigned to orthographic patterns randomly to ensure, as is true of monomorphemic words in English, that there was no systematic relationship between orthography and semantics. Words were considered semantically related if their semantics were generated from the same prototype. Half of the words in each category were designated as high-frequency and, as described below, were presented twice as often during training as the remaining, low-frequency words.

Network Architecture

The architecture of the network is shown in Figure 9. Eighteen orthographic units (three banks of six features) encoded the three-letter input. These units were fully connected to 100 hidden units which, in turn, were fully connected to 100 semantic units. The semantic units were fully connected to each other as well as to the hidden units. In addition, each hidden and semantic unit had a bias connection from a unit whose activity was always 1.0; the weight on this connection determines the unit's activation in the absence of input and is learned in the same way as the other weights in the network. Including biases, the network had a total of 34,024 connections. The weights on connections were initialized to random values sampled uniformly between ±0.25.

The states of units in the network change smoothly and continuously in time in response to influences from other units. For the purposes of simulation on a digital computer, it is convenient to approximate continuous units with finite difference equations in which time is discretized into ticks of some duration τ. Thus, the activation of unit j at time t is given by

$$a_j^{[t]} = \tau\sigma\left(\sum_i w_{ij} a_i^{[t-\tau]}\right) + (1-\tau)a_j^{[t-\tau]}, \qquad (1)$$

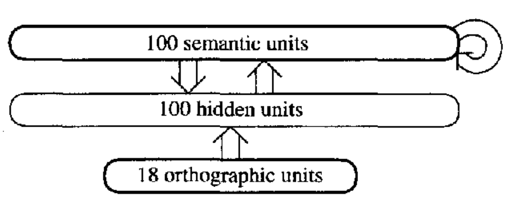

Figure 9. The architecture of the network. Ovals represent groups of units, and arrows represent full connectivity between these groups.

where w_{ij} is the weight from unit i to unit j and $\sigma(x) = (1 + \exp(-x))^{-1}$ is the standard sigmoid function (the top portion of which is depicted in Figure 1). According to this equation, a unit's activation at each point in time is a weighted average of its current activation and the one dictated by other units, where τ is the weighting proportion. As τ approaches zero, the discrete system more closely approximates the true underlying continuous system, but the demands on computational resources increase. In the current simulation, a relatively large value of τ was used during training, when minimizing computation time is critical, whereas a much smaller value of τ was used during testing, when a more precise measure of settling time was required. Note, however, that this manipulation of τ did not significantly change the settling behavior of the network nor the final pattern it produced for any given input.

Training Procedure

The network was trained in the following way. On most training trials, units started with the activations they had at the end of processing the previous word. However, for the very first word, and with a probability of .01 throughout training, the activations of units were initialized to 0.2. This was done to insure that the network was capable of processing words correctly in the absence of context from a preceding word, as was the case for the presentation of primes in the testing procedure. A word was presented to the network by providing each orthographic unit with external input that was positive if the corresponding orthographic feature was present in the word's representation and negative if it was absent. The strength of this external input—controlled by a parameter termed *input strength*—was intended to reflect the relative effectiveness of lower level perceptual processes not implemented in the current simulation. Specifically, the input strength parameter specifies the fraction of the distance from the neutral activation of 0.2 to the relevant extreme of the sigmoid function (1.0 for present features and 0.0 for absent features) that would be produced by the external input in the absence of other input to a unit.[11] This external input remained constant during the processing of the word. Although the input strength parameter was held constant at 0.8 during training, it was manipulated during the testing of the network to model the performance of individuals with high versus low perceptual ability, as described below.

Given the external input resulting from the presentation of a word, all of the units in the network (including the orthographic units) updated their states according to Equation 1 for every time tick of duration $\tau = 0.2$ over a total of 4.0 units of time. (Note that the absolute time scale of the network is determined by the time constants of the underlying differential equations, which were assumed to be equal to 1.0.) The performance error of the network was measured by the cross-entropy, C (Hinton, 1989), between the activations of the semantic units, a_j, and the assigned semantic pattern for the presented word, s_j, throughout the last unit of time.

$$C = \tau \sum_{3<t\le 4} \sum_j s_j \log(a_j^{[t]}) + (1-s_j)\log(1-a_j^{[t]}). \qquad (2)$$

A continuous version of back-propagation through time (Pearlmutter, 1989) was then used to calculate the partial derivative of this measure with respect to each weight in the network. The weights were updated immediately after each word presentation p according to

[10] We characterize the basis for similarity (i.e., feature overlap) among semantic representations in terms of categories for ease of exposition. Note, however, that the general concept of semantic relatedness extends beyond simple taxonomic category membership (see, e.g., Moss, Ostrin, Tyler, & Marslen-Wilson, 1995).

[11] For example, an input strength of 0.8—the value used during training—specifies an external input of 0.575 for present features and −3.18 for absent features, because $0.8 \times (1.0 - 0.2) = 0.64 = \sigma(0.575)$ and $0.2 - 0.8 \times (0.2 - 0.0) = 0.04 = \sigma(-3.18)$.

$$\Delta w_{ij}(p) = \epsilon \frac{\partial C}{\partial w_{ij}} + \alpha \Delta w_{ij}(p-1), \tag{3}$$

using a learning rate $\epsilon = 0.005$ and momentum $\alpha = 0.8$. Although the network received error only during the last of four units of time, the back-propagated error exerted a pressure on the network to settle to the correct semantic pattern as quickly as possible.

The selection of the next word for training was influenced both by relatedness and by frequency. Given the high co-occurrence of categorical and associative relatedness in the experimental stimuli (see Appendix A) and in natural language (see, e.g., Postman & Keppel, 1970), word pairs that were categorically related were also made associatively related by increasing their transition probabilities (Plaut, 1995). Specifically, with a probability of 1/7, the next word to be presented was selected randomly from among the other words in the same semantic category as the current word; on the remaining 6/7 of trials, the next word was selected randomly from the entire set of words. The value of 1/7 was chosen so that the next word was twice as likely to come from the same category as from another category. The approximation of using complete co-occurrence of categorical and associative relatedness in the model was considered adequate in the current context because our empirical work did not attempt to dissociate these factors.

Words designated as high frequency were twice as likely to be selected for training as low-frequency words. In this way, the relative influence of frequency and semantic–associative relatedness were equated in the network. Frequency was not varied continuously as in the empirical studies because the full crossing of category and typicality left only eight items per condition over which frequency could vary, and it seemed unlikely that a smooth frequency distribution over so few items would produce different effects than a dichotomous one.

The network was trained for a total of 200,000 word presentations, at which point it was completely accurate in settling into the semantic representation of each word regardless of the preceding context. However, as described below, we also examined the performance of the network at earlier points in training as an approximation to the reading experience of the 3rd- and 6th-grade children in Experiment 2.

Testing Procedure

During testing, stimuli were presented to the network in prime–target pairs. First, the network was initialized to activations of 0.2. Then the prime was presented as external input to the orthographic units and processed for some specified duration. The prime was then replaced by a "blank" input (all zeros), and processing continued until some specified SOA had elapsed. Following this, the target was presented and the network continued processing until the semantic activation stopped changing—specifically, until the activation of each semantic unit differed from the sigmoid of its summed input from other units (see Equation 1) by no more than 0.05. At this point, the network was considered to have responded and the time elapsed since the presentation of the target was taken as its RT. To compute these RTs precisely, the network was tested using a much finer temporal discretization ($\tau = 0.01$) than was used during training ($\tau = 0.2$). As mentioned earlier, however, this manipulation had a negligible effect on the final activations produced by the network during testing.

The network was tested under 12 conditions by fully crossing three factors. The factor Age specified the amount of training experienced by the network—either 60,000 presentations (3rd-grade children), 80,000 presentations (6th-grade children), or 200,000 word presentations (adults). The values for the children were chosen so that the relative difference in overall RT between the child and adult conditions was approximately the same for the network as for the humans. The second factor, Perceptual Ability, was instantiated by setting the input strength parameter higher in the high-ability conditions (0.9) than in the low-ability conditions (0.82). These values were chosen to approximate the relative difference in performance between the high- and low-ability individuals in the empirical studies.[12] The third factor, SOA, reflected the timing of stimuli—prime–target pairs were presented either at an SOA of 1.0 unit of time with an ISI of 0.5 (short SOA) or at an SOA of 4.0 with an ISI of 1.0 (long SOA). Note that the SOA and ISI values are directly proportional to those used in the empirical studies (short SOA of 200 ms with 100-ms ISI; long SOA of 800 ms with 200-ms ISI).

For each level of Age, Perceptual Ability, and SOA, the RTs for each word and nonword as target were measured when preceded by every other word and nonword as prime.[13] Lexical decisions were based on a measure of the familiarity of the resulting semantic pattern (Atkinson & Juola, 1973; Balota & Chumbley, 1984). The specific measure of familiarity that was used is termed *semantic stress* and reflects the degree to which semantic activations are binary (also see Plaut, 1997). More formally, the stress S_j of unit j is a measure of the information content (entropy) of its activation a_j, corresponding to the degree to which it differs from the "neutral" output of 0.5 (the value generated by the sigmoid given zero input):

$$S_j = a_j \log_2(a_j) + (1 - a_j)\log_2(1 - a_j) - \log_2(0.5). \tag{4}$$

The stress of a unit is 0 when its activation equals 0.5 and approaches 1 as its activation approaches either 0 or 1. Because the semantic patterns generated by words come to approximate their binary target patterns over the course of training, the average semantic stress for words approaches 1. By contrast, nonwords are novel stimuli that typically do not drive semantic units as strongly as words do, resulting in lower semantic stress values (see Figure 10 in the *Results and Discussion* section below). We assume that individuals can adopt a decision criterion that optimally distinguishes words from nonwords on the basis of the distribution of semantic stress values. Plaut (1997) showed that, under this assumption, semantic stress provided a reliable basis for lexical decision in a feedforward network that was trained to map from orthography to semantics for the 2,998 words in the Plaut et al. (1996) training corpus. For the current network, presentations of word targets that generated stress values below the decision criterion were considered errors and were not included in the RT analyses.

For each word target in each network condition, three item means were computed: (a) the mean RT of correct responses to the target word when preceded by each of the 15 nonidentical primes that were both associatively and categorically related to it (*related* condition); (b) the mean RT of correct responses to the target word when preceded by each of the 112 primes that were neither associatively nor categorically related to it (*unrelated* condition); and (c) the mean RT of correct responses to the target word when preceded by each of the 128 nonword primes (*nonword* condition).

In summary, the network was tested in a fully crossed, five-factor design involving age (3rd-grade vs. 6th-grade vs. adult), perceptual ability (high vs. low), SOA (short vs. long), target frequency (high vs. low), and priming context (related vs. unrelated vs. nonword), for a total of 72 cells. Target

[12] A more direct match to the empirical situation would be to vary perceptual ability (i.e., input strength) over the course of learning rather than only at testing. We chose not to do this in order to be able to account for all of the experimental conditions using a single network. If input strength were manipulated during training, a somewhat lower value would have been needed to match the low-perceptual conditions in the empirical data, but the overall pattern would not be expected to change substantially. The reason for this is that weaker input strength would yield more error in performance and, hence, larger weight changes. This effect would, however, diminish over time as the network gained competence on the task. Thus, the general difference between the high- versus low-perceptual conditions would be reduced in magnitude but not eliminated.

[13] Note that the network was reinitialized before each prime presentation, so there is no possibility of cross-trial contaminating effects caused by target or prime repetition, as there would be with experiment participants.

frequency is a between-items factor, whereas all others are within-item factors.

Results and Discussion

Figure 10 shows the distribution of semantic stress values produced by word and nonword targets in the child conditions (3rd- and 6th-grade combined) and in the adult conditions, averaged over all other factors. When compared with the child conditions, the adult condition produces higher stress values, particularly for word targets, and much less overlap between the word and nonword distributions. Given the strong effect of age (i.e., amount of training) on the discriminability of words and nonwords, separate decision criteria were applied in performing lexical decisions in the adult and child conditions. These criteria, shown as vertical lines in the figure, were chosen to roughly approximate the proportion of hits to false alarms exhibited by the corresponding individuals in the empirical studies. Specifically, for the adult conditions in the model, making a "yes" response when semantic stress equals or exceeds a criterion of 0.91 yielded 99.8% hits and 0.32% false alarms; the corresponding values were 98% and 6.8% for the adults in Experiments 1 and 3. For the child conditions (collapsing 3rd- and 6th-grade), a criterion of 0.87 yielded 97.6% hits and 10.0% false alarms (cf. 97% and 11% for the children in Experiment 2). Thus, overall, the network was somewhat more accurate at lexical decision than were the participants, particularly in the adult conditions.

For all correct responses to word targets (hits), RTs outside ±2.5 *SD*s within each cell of the design (i.e., each combination of age, perceptual ability, SOA, target frequency, and priming context) were withheld from the latency analysis. Overall, this removed 2.9% of the observations, with a maximum over cells of 5.4%. In general, there was slightly greater trimming for cells involving the weaker level of each factor, although an ANOVA revealed no statistically reliable effects of any of the factors or their interactions on the proportion of items removed.

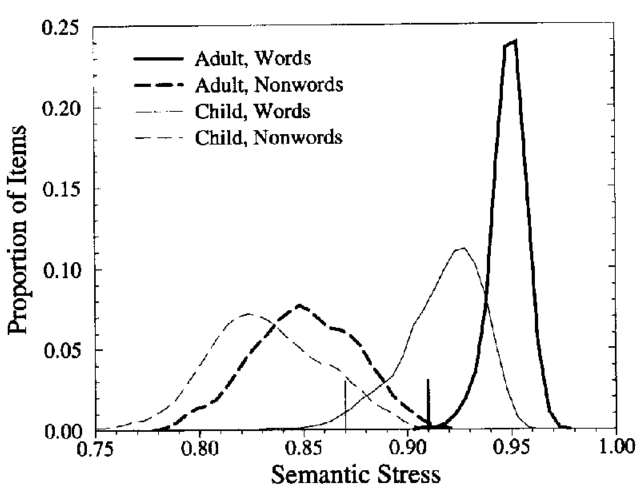

Figure 10. Distribution of semantic stress values produced by the network for word targets (solid lines) and nonword targets (dashed lines) in the child conditions (light lines; 3rd- and 6th-grade collapsed) and in the adult condition (dark lines). The small vertical lines indicate the decision criteria used for lexical decision for the child conditions (left, light line) and the adult conditions (right, dark line).

To facilitate comparison of the network's performance with that of the participants', the network's RTs were converted to milliseconds by computing a mean RT for each condition and then linearly regressing these means against the participant means from the corresponding conditions from Experiments 1–3. A separate regression was carried out for each of the experiments, because different populations and testing conditions were used. Frequency was treated as a dichotomous variable because it was instantiated as such in the simulation. Planned comparisons among the related, unrelated, and nonword conditions were used to determine whether the priming effects were due to facilitation, inhibition, or both. As in the experiments, all of the analyses are over items (N = 128) and, therefore, all F tests have degrees of freedom of (1, 126).

The network can be made to trade speed for accuracy by manipulating the criteria for determining settling times and lexical decisions (see the General Discussion). However, given the very low error rates in the current empirical studies, particularly for the adults, we fixed these criteria in the network to values that produced comparable high levels of accuracy. Consequently, many testing conditions produced no errors, and there were relatively few reliable effects in the analysis of errors produced by the network. Those effects that did hold were all in the same direction as those found in the RT analyses reported below. In fact, across conditions, there was a high correlation (r = .74, p < .001) between mean RT and error rate. Moreover, the only reliable interaction in the error data from the empirical studies was between target frequency and priming context for adults at the long SOA (Experiment 1), and this was in the same direction as the effect in RTs. Given these considerations, only RT analyses for the network are presented below. Means of both RTs and errors are presented in Appendix B.

Simulation of Experiment 1: Adults, Long SOA

Analogous to Experiment 1, correct mean RTs of the network for each word in the adult condition (200,000 word presentations) tested at the long SOA (SOA = 4.0, ISI = 1.0) were entered into a three-factor ANOVA, with perceptual ability (high vs. low) and priming context (related vs. unrelated) as within-item factors and target frequency (high vs. low) as a between-items factor. Figure 11 shows the means for each condition both for the participants in Experiment 1 and for the network. The analysis revealed clear main effects of perceptual ability, F = 76.06, MSE = 4,481, p < .001; target frequency, F = 47.33, MSE = 7,284, p < .001; and priming context, F = 75.62, MSE = 1,081, p < .001; and an Ability × Frequency interaction, F = 9.59, MSE = 4,482, p < .005. However, these effects were qualified by a reliable three-way Ability × Frequency × Context interaction, F = 10.26, MSE = 781.2, p < .001. These findings agree with those of Experiment 1, except for the two-way interaction. In the network, the frequency effect was smaller in the high-ability condition (41 ms) than in the low-ability condition (85 ms). For the participants, this difference (31 vs. 40 ms, respectively) was in the same direction numerically but was not reliable.

As can be seen in Figure 11, there was a trend in the low-ability condition in the opposite direction to the interaction for the high-ability condition. Specifically, in the high-ability condition, the

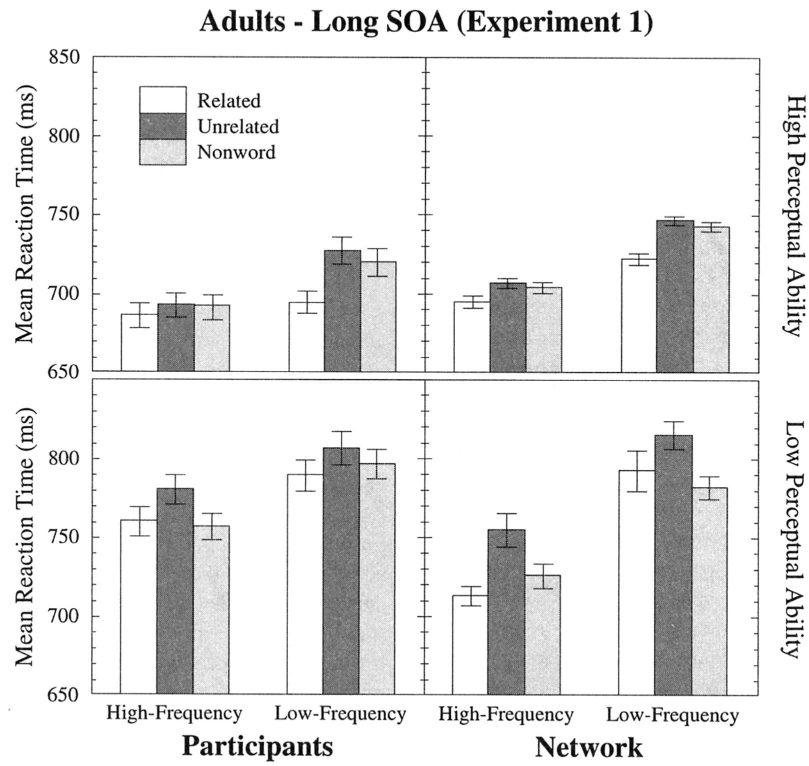

Figure 11. Mean reaction times (RTs) for high- and low-perceptual-ability adults at the long stimulus-onset asynchrony (SOA; from Experiment 1) and the mean RTs for the network from the corresponding conditions. Error bars are 1 standard error by items.

network exhibited less priming for high-frequency targets than for low-frequency targets (ds = 12 and 24 ms, respectively), which agrees with the empirical findings (ds = 7 and 33 ms). In the low-ability condition, the network's priming effect was larger for high- versus low-frequency targets (ds = 42 and 23 ms, respectively). The corresponding numeric difference for participants (ds = 20 and 17 ms) was in the same direction but was not reliable.

We have explained the standard finding of greater priming for low-frequency targets in the high-ability condition in terms of the nonlinearity of sigmoid activation function (see Figure 1). In fact, the same principles can explain a trend toward the reverse interaction—greater priming for high-frequency targets—in the low-ability condition. If the bottom–up contribution of perceptual ability to the input of semantic units is sufficiently weak, the low-frequency targets may start to fall within the lower tail of the sigmoid (for active units)—this is, in fact, reflected in Figure 1. As a result, the effects of priming context would be reduced relative to those for high-frequency targets, which still fall within the linear range of the function. However, given that this reverse Frequency × Context interaction was not reliable in either the empirical or simulation data, this account should be considered merely suggestive until the finding is verified by additional empirical and computational investigation.

As in the empirical analyses, planned comparisons with the nonword prime condition within each combination of target frequency and perceptual ability were calculated to determine whether context effects were due to facilitation (related vs. nonword conditions) or inhibition (unrelated vs. nonword conditions). In the high-ability conditions, context effects were due mostly to facilitation ($ps < .001$) with only marginal inhibition ($p = .013$ for low-frequency targets; $p = .069$ for high-frequency targets). In the low-ability conditions, context effects were due to both facilitation ($p < .012$) and inhibition ($p < .05$) for high-frequency targets, but only inhibition for low-frequency targets ($p < .001$). These findings are broadly consistent with those from Experiment 1, except for the findings of reliable facilitation for high-frequency targets in the high-ability condition and reliable inhibition for low-frequency targets in the low-ability condition. In both of these cases, however, the numeric differences in the empirical data agree with the effects in the network. Also, as was true of the empirical findings, mean RTs for unrelated primes were numerically slower than for nonword primes regardless of target frequency and perceptual ability.

Overall, there was a good qualitative match between the pattern of RTs produced by the network in the adult, long-SOA conditions and the pattern of results produced in Experiment 1 by adults tested at the long (800 ms) SOA. Most important, the network exhibited the appropriate three-way Perceptual Ability × Target Frequency × Priming Context interaction, with greater priming for low-frequency targets only in the high-ability condition. In fact, the low-ability condition showed a trend toward greater priming for high-frequency targets. Although this reverse interaction held numerically in the participant data and can be understood within the general framework of distributed network models, it requires further verification. Finally, the network also produced the observed empirical pattern of a mixture of facilitation and inhibition at the long SOA, with primarily facilitation in the high-ability conditions and primarily inhibition in the low-ability conditions.

Simulation of Experiment 2: Children, Long SOA

By analogy with Experiment 2, two separate ANOVAs were carried out on the network's correct mean RTs for each word in the child, long-SOA conditions (60,000 and 80,000 word presentations; SOA = 4.0, ISI = 1.0). The first ANOVA analyzed the effects of perceptual ability by collapsing across age, and the second analyzed age by collapsing across perceptual ability.

Effects of perceptual ability. The analysis of perceptual-ability effects was analogous to the one just described for the adult, long-SOA conditions, except that the data were collapsed across the 3rd- and 6th-grade conditions. Figure 12 shows the means for each condition both for the participants in Experiment 2 and for the network. Like the participant data, the network's settling times at the long SOA were much slower in the child condition than in the adult condition, but the pattern of results was very similar. There were reliable main effects of perceptual ability, $F = 203.22$, $MSE = 4{,}886$, $p < .001$; target frequency, $F = 63.36$, $MSE = 12{,}395$, $p < .001$; and priming context, $F = 109.71$, $MSE = 1{,}138$, $p < .001$. Perceptual ability also interacted with target frequency, $F = 6.41$, $MSE = 4{,}886$, $p = .0126$, and with priming context, $F = 30.34$, $MSE = 645.2$, $p < .001$. These interactions were not reliable in the empirical data from Experiment 2, but both were in the same direction numerically as in the network's data. The network's two-way interactions were, however, qualified by a three-way Ability × Frequency × Context interaction, $F = 6.21$, $MSE = 2{,}698$, $p = .014$. As in the corresponding analysis of the empirical data from Experiment 2, low-frequency words produced greater priming than high-frequency words only in the high-ability condition (ds = 29 and 9 ms, respectively), not in the low-ability condition (ds = 44 and 43 ms, respectively).

The relative contributions of facilitation and inhibition to context effects were determined by planned comparisons of the mean RTs for the nonword prime conditions to the related and unrelated conditions. In the high-ability condition, there was only facilitation for high-frequency targets ($p < .001$) but both facilitation and inhibition for low-frequency targets (ps $< .001$). In the low-ability condition, there was both facilitation and inhibition for high-frequency targets (ps $< .05$) but only inhibition for low-frequency targets ($p < .001$). As Figure 12 makes clear, the main discrepancies between the network findings and those in Experiment 2 are due to the nonword priming conditions for the low-ability participants. Particularly for low-frequency targets, the network exhibited inhibition—faster responses following nonword primes as compared with unrelated primes—whereas the participants did not.

Effects of age. Figure 13 shows the means for each condition both for the 3rd- and 6th-grade participants in Experiment 2 and for the network. An ANOVA revealed a clear main effect of age, $F = 57.48$, $MSE = 2{,}941$, $p < .001$, but no reliable interactions of

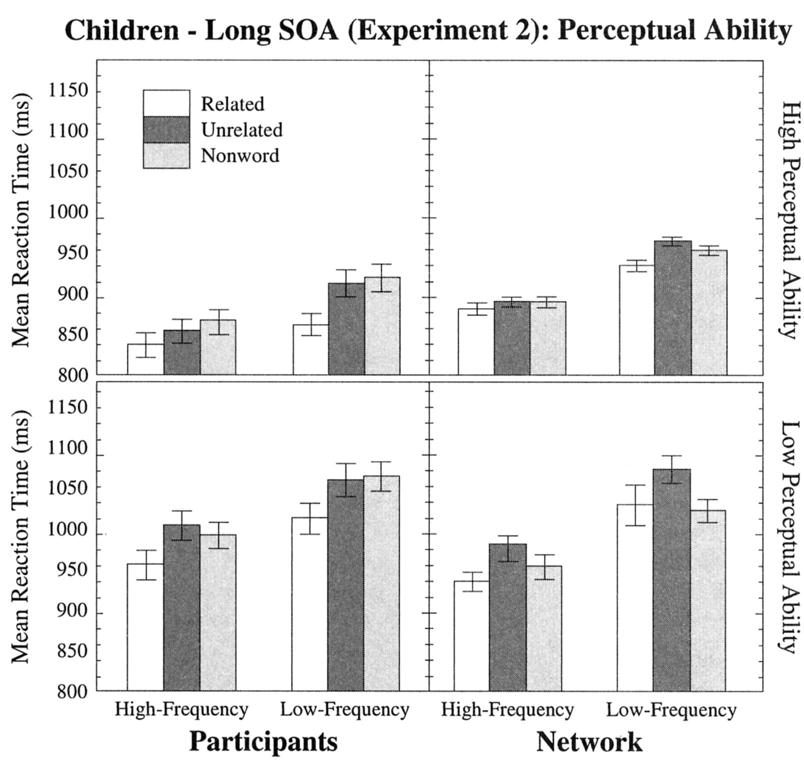

Figure 12. Mean reaction times (RTs) of high- and low-perceptual-ability children at the long stimulus-onset asynchrony (SOA; from Experiment 2) and the mean RTs for the network from the corresponding conditions. Error bars are 1 standard error by items.

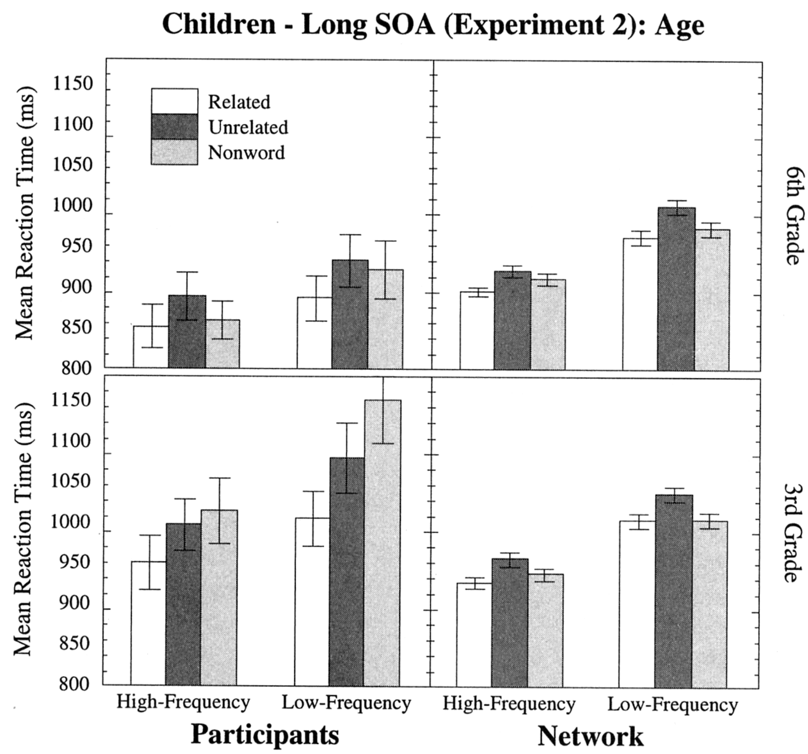

Figure 13. Mean reaction times (RTs) of 6th- and 3rd-grade children at the long stimulus-onset asynchrony (SOA; from Experiment 2) and the mean RTs for the network from the corresponding conditions. Error bars are 1 standard error by items.

other factors with age. In particular, the three-way Age × Target Frequency × Perceptual Ability interaction showed only a weak trend toward significance, $F = 1.54$, $MSE = 437.5$, $p = .22$. This is the same pattern of results as was found for the children in Experiment 2. Note that, apart from a general speedup, the network shows very little change in its pattern of performance as a function of age, whereas the 3rd-grade participants are much more variable than the 6th-grade participants. This discrepancy indicates that amount of reading experience alone does not provide a full account of the differences in reading behavior between 3rd- and 6th-graders.

Comparison with adult conditions. In the analysis of the empirical data from Experiment 2, we carried out a 2 × 2 ANOVA comparing adults with low perceptual ability (from Experiment 1) to children with high perceptual ability in terms of RTs following unrelated versus nonword primes and found that the adults showed greater inhibition than the children. The corresponding analysis of the network's RTs yields equivalent result: a reliable Age × Context interaction, $F = 40.27$, $MSE = 213.2$, $p < .001$, such that the difference between the unrelated and nonword prime conditions was less in the child condition ($d = 6$ ms) than in the adult condition ($d = 22$ ms).

In summary, as in the empirical studies, the network in the child, long-SOA conditions produced a pattern of priming rather similar to that of the adult condition at the long SOA. Most important, the standard finding of greater priming for low-frequency targets held only under conditions of high perceptual ability; high- and low-frequency targets produced nearly equal levels of priming under low perceptual ability. Also in agreement with the empirical findings, differences in age produced a general speed-up but no interactions with frequency or context. The model did, however, produce inhibition in the low-ability condition, particularly for low-frequency targets, that was absent from the empirical data. We consider the implications of this discrepancy in detail in the General Discussion.

Simulation of Experiment 3: Adults, Short SOA

Analogous to Experiment 3, another ANOVA was carried out on the correct mean RTs of the network for each word in the adult, short-SOA conditions (200,000 word presentations; SOA = 1.0, ISI = 0.5). Figure 14 shows the means for each condition both for the participants in Experiment 3 and for the network. Consistent with the empirical findings, the network analysis showed main effects of perceptual ability, $F = 33.05$, $MSE = 4{,}181$, $p < .001$; target frequency, $F = 54.10$, $MSE = 8{,}412$, $p < .001$; and priming context, $F = 121.37$, $MSE = 732.8$, $p < .001$. These main effects were qualified, however, by a three-way Ability × Frequency × Context interaction, $F = 7.52$, $MSE = 180.4$, $p < .01$. This interaction is due to the network producing greater priming for low- versus high-frequency targets in the high-ability condition ($ds = 28$ and 21 ms, respectively) but a trend toward the opposite pattern in the low-ability condition: less priming for low- versus high-frequency targets ($ds = 26$ and 31 ms, respectively). That priming for low-frequency targets was numerically smaller than for high-frequency targets in the low-ability condition agrees with

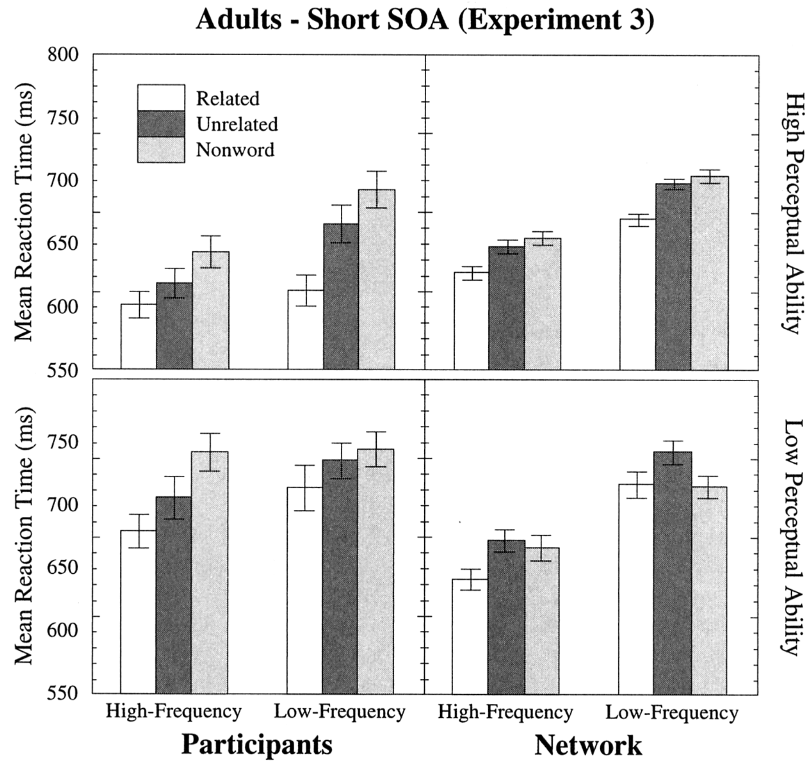

Figure 14. Mean reaction times (RTs) for high- and low-perceptual-ability adults at the short stimulus-onset asynchrony (SOA; from Experiment 3) and the mean RTs for the network from the corresponding conditions. Error bars are 1 standard error by items.

the empirical findings from Experiment 3 (ds = 22 and 27 ms) and echoes the trend toward a reverse Frequency × Context interaction discussed in the context of the simulation of adults at the long SOA (Experiment 1).

Planned comparisons of the related and unrelated prime conditions with the nonword prime condition within each combination of frequency and perceptual ability revealed that, in the high-ability conditions, context effects were due only to facilitation ($ps < .001$). Note that this contrasts with the corresponding conditions for adults at the long SOA, in which the simulation exhibited both facilitation and inhibition. These findings agree with the empirical results from Experiments 1 and 3 for the high-ability condition. In the low-ability conditions, context effects on high-frequency targets were also due primarily to facilitation ($p < .001$; $p = .070$ for inhibition). The effects on low-frequency targets, by contrast, were due entirely to inhibition ($p < .001$). Similar to the simulation results for Experiment 2 (children at the long SOA), the current simulation results differ from the empirical results for the low-ability condition, particularly for low-frequency targets, for which nonword primes produced faster RTs than unrelated primes.

To further clarify the pattern of facilitation and inhibition exhibited by the network, we carried out an additional analysis of the changes in these measures as a function of SOA (by collapsing the data from the simulations of Experiments 1 and 3). The magnitudes of facilitation and inhibition were normalized relative to the mean RT of the relevant priming conditions (i.e., facilitation was measured by the difference in mean RT between the nonword and related priming conditions, divided by their average; inhibition was measured by the difference between the unrelated and nonword priming conditions, divided by their average). This analysis revealed that facilitation was greater at the short compared with long SOA, $F = 10.50$, $MSE = 206.15$, $p < .005$, whereas inhibition was greater at the long compared with short SOA, $F = 28.71$, $MSE = 156.10$, $p < .001$. This general shift from facilitation at the short SOA to inhibition at the long SOA also holds numerically for the empirical data from Experiments 1 and 3 (see Figures 11 and 14) and is consistent with the findings of a number of previous studies (e.g., Heyer et al., 1985; L. C. Smith et al., 1987).

In summary, the network in the adult, short-SOA conditions produced a pattern of results similar to the empirical results from Experiment 3 for adults at the short SOA (200 ms) except that, for the network, nonword primes produced overly fast RTs to low-frequency targets in the low-perceptual-ability condition.

In addition to modeling the results from Experiments 1–3, the network also provides a basis for making predictions concerning both conditions that have yet to be investigated empirically and interactions that could not be tested with the empirical data because of the inability to control certain factors across participant groups. Specifically, the next subsection presents predictions of the performance of children at a short SOA, which was not tested empirically. The subsequent subsection presents predictions of how effects of perceptual ability, priming context, and target frequency interact with age and SOA. These interactions could not be tested reliably because perceptual ability and frequency of exposure to targets cannot be equated across age groups, and SOA was manipulated as a between-subjects factor. These predictions are important because they broaden the generality of the central empirical and theoretical claims of the current work beyond what we know from existing data.

Predictions

Children, short SOA. A three-factor ANOVA was carried out on the network's correct mean RTs for each word in the child, short SOA conditions (60,000 and 80,000 word presentations; SOA = 1.0, ISI = 0.5). Given the lack of any interactions with age in the previous analyses, data were collapsed across this factor. Network settling times were converted to RTs using the regression equation from Experiment 2. Figure 15 shows the means for each condition for the network; the numeric values can be found in Appendix B. As was found for the other combinations of age and SOA, there were main effects of perceptual ability, $F = 90.64$, $MSE = 4,604$, $p < .001$; target frequency, $F = 47.85$, $MSE = 56,952$, $p < .001$; and priming context, $F = 156.34$, $MSE = 404.5$, $p < .001$. There was also an Ability × Context interaction, $F = 26.61$, $MSE = 218.98$, $p < .001$, but these effects were qualified by a three-way Ability × Frequency × Context interaction, $F = 9.00$, $MSE = 218.99$, $p = .0033$. Note that, at least numerically, the network showed the same trend of a reverse Frequency × Context interaction in the low-ability condition as found in the other simulation conditions.

Planned comparisons of related and unrelated prime conditions with the nonword prime condition within each combination of frequency and perceptual ability indicated a pattern of facilitation and inhibition rather similar to that for the adult, short-SOA conditions (simulation of Experiment 3). Specifically, in the high-ability conditions, context effects were due only to facilitation ($ps < .001$), whereas in the low-ability conditions, context effects were due to facilitation for high-frequency targets ($ps < .005$) but only to inhibition for low-frequency targets ($p < .001$). However,

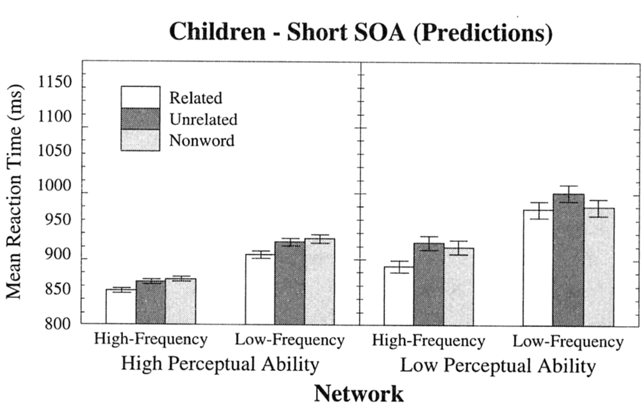

Figure 15. Mean reaction times for the network under the child condition at the short stimulus-onset asynchrony (SOA), as a function of perceptual ability, target frequency, and priming context. Error bars are standard errors by items.

given the discrepancy between the network's and participants' performance for low-frequency targets following nonword primes in the low-ability conditions for adults as the short SOA, the prediction of inhibition for children in this condition must be interpreted with caution.

Finally, an analysis of facilitation and inhibition effects across SOA, analogous to the one for adults reported above, was carried out for the child conditions. This analysis showed that facilitation was greater at the short compared with long SOA, $F = 11.73$, $MSE = 34.79$, $p < .001$, whereas inhibition was greater at the long compared with short SOA, $F = 117.15$, $MSE = 11.38$, $p < .001$. Thus, the network predicts that, as with adults, facilitation should diminish and inhibition should increase in moving from short to long SOA with children.

In summary, the network in the child condition at the short SOA produced a pattern of results that is generally consistent with the findings for the other combinations of age and SOA. Specifically, there was more priming for low-frequency targets compared with high-frequency targets under conditions of high perceptual ability, but equal amounts of priming for low- and high-frequency targets under conditions of low perceptual ability. Moreover, the high-ability conditions produced only facilitation, whereas the low-ability conditions produced a mixture of facilitation and inhibition.

Interactions with age and SOA. As a final analysis, we entered correct mean RTs of the network for each word into a five-factor ANOVA, with age (adult vs. child), SOA (long vs. short), perceptual ability (high vs. low), and priming context (related vs. unrelated) as within-item factors and target frequency (high vs. low) as a between-items factor. This fully crossed comparison was not possible for the empirical data because the children were not tested in a short SOA condition, they were presented with a fewer number of items, and they were not matched to the adults in terms of perceptual ability. The primary purpose of this analysis was to determine whether the three-way Perceptual Ability × Priming Context × Word Frequency interaction depended on SOA or Age. Given the large number of comparisons involved in this analysis (32), we considered an effect reliable only at $p < .001$. Not surprisingly, the three-way interaction was reliable, $F = 15.93$, $MSE = 975.87$, $p = .0001$, but its magnitude did not depend on age, SOA, or their interaction ($p > .019$). Thus, the most central finding of the current work, that greater priming for low- vs. high-frequency targets holds only for participants with high perceptual ability, is predicted to be independent of age and SOA.

In a separate study involving 130 participants and 116 target words (Booth & Plaut, 2000), we replicated the results of Experiments 1 and 3 with SOA manipulated as a within-subjects factor. This new study allowed us to overcome the limitation of the present study in comparing participants with different perceptual abilities at the long and short SOA. The study yielded significant main effects of perceptual ability, priming context, target frequency, and SOA ($ps < .001$). These main effects were qualified by the same pattern of three-way Perceptual Ability (high vs. low) × Priming Context (related vs. unrelated) × Target Frequency (high vs. low) interaction, $F(1, 928) = 4.95$, $MSE = 42,520$, $p < .05$, as found in the current studies. In addition, this three-way interaction was of similar magnitude at short and long SOAs—there was no four-way interaction.

Summary of Simulation Results

The current simulation demonstrated that a distributed network model of lexical processing can account for the findings of Experiments 1–3 that priming context interacts with target frequency for participants with high perceptual ability but not for those with low perceptual ability. As was true for the empirical studies, this finding in the network was remarkably general, holding across differences in both age and SOA.

There were, however, some discrepancies between the empirical data and the results of the computational model. In particular, both for children at the long SOA and for adults at the short SOA, the network produced inhibition for low-frequency targets under conditions of low perceptual ability, because of overly fast RTs following nonword primes under these conditions. Nonetheless, the network did exhibit the general pattern found in the empirical data: Context effects caused by facilitation were larger at the short SOA, but context effects caused by inhibition were larger at the long SOA.

General Discussion

In our empirical studies and distributed network modeling, we examined the influence of several theoretically important factors on the magnitude of semantic priming. Our empirical results support the extensive literature in naming and lexical decision, which shows that the degree to which a prime, such as NURSE, affects the recognition of a target, such as DOCTOR, depends on the frequency of the target: Low-frequency targets are influenced more by priming context than high-frequency targets (e.g., C. A. Becker, 1979; Borowsky & Besner, 1993; Stanovich & West, 1981). The most theoretically important finding of the empirical studies was that, across differences in both age and SOA, the interaction between priming context and target frequency depended on a reader's perceptual ability. Only those with high perceptual ability exhibited greater priming for low-frequency targets than for high-frequency targets. Participants with low perceptual ability showed equal priming for high- and low-frequency targets (see Figures 3, 5, and 7).

We then demonstrated that a distributed network model exhibited the same pattern of results when perceptual ability was instantiated in terms of the strength with which orthographic input was presented to the network (see Figures 11, 12, and 14). In the model, the orthographic input drives the semantic representations of high-frequency targets more strongly than those of low-frequency targets, because of differences in the frequency of training on these words (also see Borowsky & Besner, 1993). As a result, the semantic system settles into a stable pattern of activity faster for high- compared with low-frequency targets. Preceding a high-frequency target by a related prime produces little facilitation relative to an unrelated prime because frequency alone is sufficient to drive the activations of semantic units near the asymptote of the sigmoid function. By contrast, priming context has a much larger effect on low-frequency targets because their initial activations fall closer to the linear range of the sigmoid, where other factors can still have clear effects (see Figure 1). Therefore, the same priming context yields a larger priming effect for low-frequency targets than for high-frequency targets because of the "diminishing returns" of the asymptotic nature of the sigmoid activation function (see also Cohen et al., 1990; Plaut, 1995; Plaut et al., 1996). This Frequency × Context interaction is modulated by perceptual ability because the weaker orthographic activation in low-ability readers causes the input for both high- and low-frequency targets to fall within the linear range of the activation function for semantic units. As a result, the related and unrelated primes influence the recognition of high- and low-frequency targets to a similar degree.

Note that alternative models, such as the interactive compensatory model (Stanovich, 1980), can also account for the three-way Perceptual Ability × Priming Context × Target Frequency interaction, but they must invoke multiple mechanisms to do so. By contrast, the distributed network model provides a single-mechanism account of these effects. Moreover, the underlying computational principles embodied in the model are not specific to the domain of lexical processing, but apply in essentially unaltered form across the full range of cognitive processes (see McClelland, Rumelhart, & the PDP Research Group, 1986; McLeod, Plunkett, & Rolls, 1998; Quinlan, 1991).

With respect to facilitation and inhibition relative to a neutral priming context, dual-mechanism models assume that inhibition can influence word recognition only at long SOAs because it is a slow, strategic expectancy-based process, whereas facilitation influences word recognition regardless of SOA because spreading activation is fast and automatic (see Neely, 1977, 1991; Posner & Snyder, 1975). Consistent with this account, our empirical studies with adults showed both facilitation and inhibition at the long SOA, but only facilitation at short SOAs (see Figures 3 and 7).

Under conditions corresponding to adult performance (i.e., training for 200,000 word presentations), our distributed network model also exhibited inhibition dominance at a long SOA and facilitation dominance at a short SOA (see Figures 11 and 14). This finding is of fundamental importance because it is often assumed that inhibitory priming effects imply a contribution from separate expectancy-based processes. Our results indicate that the increased inhibition at long SOAs can arise from the same mechanism that produces only facilitation at short SOAs. On our account, the shift from facilitation to inhibition across SOAs reflects the degree of hysteresis in moving from the representation of the prime to that of the target, corresponding to the depth to which the system settles into the attractor basin for the prime when encountering the target. At a short SOA, the system has only enough time to move partially into the prime's basin. This corresponds to fairly weak semantic activity that nonetheless facilitates the processing of a semantically related target. By contrast, at a longer SOA, the network settles deeply into the attractor basin for the prime. On presentation of the target, the system must then move up and out of the prime's basin to derive the representation of the target. Although semantic similarity between prime and target may help this process to some extent, the semantic features for which the prime and target differ must nevertheless be reversed, and this process is prolonged as the prime's features (including those not shared with the target) are activated more strongly. Thus, in the adult conditions, our distributed network model provides an alternative and more parsimonious account than dual-mechanism models of the time course of facilitation and inhibition as a function of SOA.

Dual-mechanism models also predict that children should exhibit inhibition as well as facilitation because their word recognition processes are slow and not automatic (Stanovich, 1980). Our empirical results did not support this prediction—both high- and low-perceptual-ability children exhibited facilitation but no inhibition (see Figure 5). It must be acknowledged, however, that our distributed network model also provided a less than adequate account of these findings (see Figure 12). Specifically, the model showed clear inhibitory effects for the low-perceptual-ability conditions that were absent in the empirical data for the participants.

Two factors may have contributed to this discrepancy. The first is that the nonwords may not have engaged the network sufficiently strongly, particularly in conditions of low perceptual ability where the input strength was relatively weak. In terms of feature overlap, the relative similarity of nonwords and words in the simulation matched the relative summed positional bigram frequencies of the empirical stimuli, but individual features may not be the only relevant level of structure to consider. It is likely that, over the course of training, the network learned to take advantage of the reliable CVC structure of words to facilitate processing; this higher order orthographic knowledge would not have generalized to nonwords with VCV structure. As a result, compared with unrelated word primes, nonword primes may have produced far less hysteresis, and hence much faster RTs, for subsequent target words, leading to an exaggeration of inhibitory effects in the network. In essence, our nonwords may be subject, to some extent, to the same criticism leveled against neutral baselines like Xs and words like READY—such repeated stimuli may not have the same attentional effects or engage the same levels of linguistic processing as word or wordlike nonword primes (see Antos, 1979; Jonides & Mack, 1984; McKoon & Ratcliff, 1992; McNamara, 1994; and Neely, 1991, for discussion).

The second factor that may have exaggerated the inhibition shown by the model concerns the instantiation of semantic and associative relatedness in the model. The prime–target stimuli for our empirical studies were chosen from free-association norms (Nelson et al., 1994). Although some of these pairs were in the same semantic category, many were not. By contrast, in the model, all associatively related prime–target pairs were also categorically related—there were no purely associatively related pairs. Empirical studies show that inhibition effects are larger in categorical priming than in associative priming (Lupker, 1984), and Plaut (1995) observed the same tendency in a simulation very similar to the current model but which separated categorical and associative relatedness. Thus, the greater predominance of categorical relatedness among prime–target pairs for the model compared with the participants may have contributed to the overly strong inhibitory effects in the former.

Overall, whereas there are some aspects of the performance of our distributed network model that are inconsistent with the findings from our empirical studies, these differences can be understood in terms of approximations made in developing the model. We leave it to future research to determine whether a more realistic model can provide an even closer fit to the empirical data.

In the remainder of the General Discussion, we begin by articulating the specific empirical predictions made by our model. We then consider the implications of some of the important simplifications made in constructing the model. Following this, we discuss other empirical findings that may appear to challenge our account and describe empirical and computational extensions of our approach to address related phenomena.

Empirical Predictions of the Model

An important benefit of developing an explicit computational implementation that instantiates the core principles of a theory is that it can be used not only to predict qualitative patterns of performance under novel conditions, but also to make specific quantitative predictions of performance under such conditions. A case in point concerns the predictions of the network for the performance of children when tested at a short SOA. Applying the regression equation from Experiment 2 to the network's settling times when tested in the child condition at the short SOA yields specific predictions of mean RTs in milliseconds across 12 additional conditions (3 priming contexts $\times$ 2 target frequencies $\times$ 2 levels of perceptual ability; see Figure 15). Most important, the model predicts that, at a short SOA, children should exhibit the same interaction among perceptual ability, priming context, and target frequency shown by adults at both the long and short SOA and by children at the long SOA.

The second prediction generated by the network model was that the interaction among perceptual ability, target frequency, and priming context (when defined as related versus unrelated) was independent of age and SOA. When facilitation and inhibition effects were examined relative to a nonword prime baseline, the model predicted a general shift from facilitation at the short SOA to inhibition at the long SOA, for both adults and children. With respect to adults, these predictions are broadly consistent with the past literature, which shows no increase in facilitation but an increase in inhibition with longer SOAs (Heyer et al., 1985; L. C. Smith et al., 1987). Again, this has not been tested in children.

Finally, although more tentatively, the network produced a trend toward a reverse Frequency $\times$ Context interaction (i.e., greater priming for high- compared with low-frequency targets) when tested in the adult, low-perceptual-ability condition at the long SOA. The same pattern held numerically for these conditions at the short SOA. The corresponding empirical data showed the same direction of effects numerically although these were not statistically reliable. As discussed in the context of the simulation of Experiment 1, this reverse interaction, like the standard one in the high-ability condition, can be understood in terms of the nonlinear effects of the sigmoid activation function (see Figure 1). Given that a reverse Frequency $\times$ Context interaction held only weakly in the model, however, additional simulation work is needed to verify that this effect is indeed a robust prediction of the model.

Simplifications of the Model

In evaluating the current account of lexical processing, it is important to keep in mind that the implemented model necessarily incorporates a number of simplifications relative to the theoretical account on which it is based. The first of these concerns the means by which RTs and lexical decisions are determined. In the current implementation, the RT of the network to a stimulus is determined by the number of processing cycles required for semantic activation levels to stabilize below some specified criterion. At that point, a lexical decision is made on the basis of the degree to which semantic representations have been driven strongly toward binary activation levels, operationalized by a measure termed *stress* (see Equation 4 and Figure 10), such that a stress level above a particular criterion indicates a "yes" response (see also Plaut, 1997).

Stress provides a reliable basis for discriminating words from nonwords because of the lack of systematicity between orthography and semantics. Words generate high stress levels because they are trained to generate semantic representations consisting of binary features. Nonwords, by definition, were not presented during training. The network's behavior for these items is solely a function of generalization from its knowledge of orthographically similar words. Given that all of the network's knowledge is encoded in the same set of connection weights, processing a nonword partially engages the mappings for all of the trained words, in proportion to the orthographic similarity of each word to the nonword. For a systematic mapping, like that between orthography and phonology, the mappings for orthographically similar words generally agree with each other and thus conspire effectively to generate strong output activation for nonwords (see Plaut et al., 1996). By contrast, for an unsystematic mapping, like that between orthography and semantics, orthographically similar words map to unrelated sets of semantic features. Nonwords still engage a combination of the mappings for similar words, but now these mappings conflict with each other—semantic units that are activated by the mapping for one similar word are inhibited by the mappings for different similar words. As a result of this inconsistency, semantic activations are driven less strongly toward extreme values—yielding lower stress—when the network processes a nonword compared with when it processes a familiar word. Although the current model used only a relatively small number of words (128), Plaut (1997) showed that semantic stress can support accurate lexical decision for 2,998 words (Plaut et al., 1996).

A number of researchers (e.g., Borowsky & Masson, 1996; Joordens & Becker, 1997; Masson & Borowsky, 1998; Rueckl, 1995) have proposed a related measure—the negative of energy (Hopfield, 1982), sometimes termed *harmony* (Smolensky, 1986) or *goodness* (McClelland & Rumelhart, 1988)—as the basis on which individuals make lexical decisions. One drawback of this measure, $\Sigma_{i<j} a_i a_j w_{ij}$, is that it requires decision processes to have direct access to the weights, w_{ij}, among units in the lexical system. By contrast, computing the stress measure requires only a fairly simple combination of unit activations. It should be pointed out, however, that the two measures are closely related—as long as all output activations are on the correct side of "neutral" (i.e., 0.5 for standard [0, 1] units), then increasing stress by moving activations toward more binary values generally also decreases energy.

Regardless of whether stress or energy is used to make lexical decisions, however, there is still a problem with the procedure used in the current simulation, stemming from the use of separate criteria for determining when the network responds (stability) and how it responds (stress). A more satisfactory approach would be to define a response criterion on the basis of how stress values change over the course of settling in response to word and nonword inputs (see Joordens & Becker, 1997, for a related proposal in terms of "harmony"). We used the simpler procedure of defining RTs in terms of settling times partly because this approach has been used successfully to model response latency data from lexical tasks (e.g., S. Becker et al., 1997; Borowsky & Masson, 1996; Kawamoto, 1993; Masson, 1995; McRae et al., 1997; Plaut et al., 1996), but also because implementing the actual mechanism that generates "yes" and "no" responses was considered beyond the scope of the current work.

It should be noted, however, that decision processes can be modeled effectively using the same computational principles as used in the current work. For example, Usher and McClelland (in press) have demonstrated recently that competition among linear, stochastic, time-averaging units representing alternative responses gives rise to a number of basic properties of empirical findings in standard choice RT tasks. Their approach could be applied in the current context by adding to the lexical network a response layer consisting of two units (see Figure 16A): a "yes" unit whose input is the level of semantic stress for the current target, and a "no" unit whose input is the value of the decision criterion used in the current work to distinguish words from nonwords (which could be

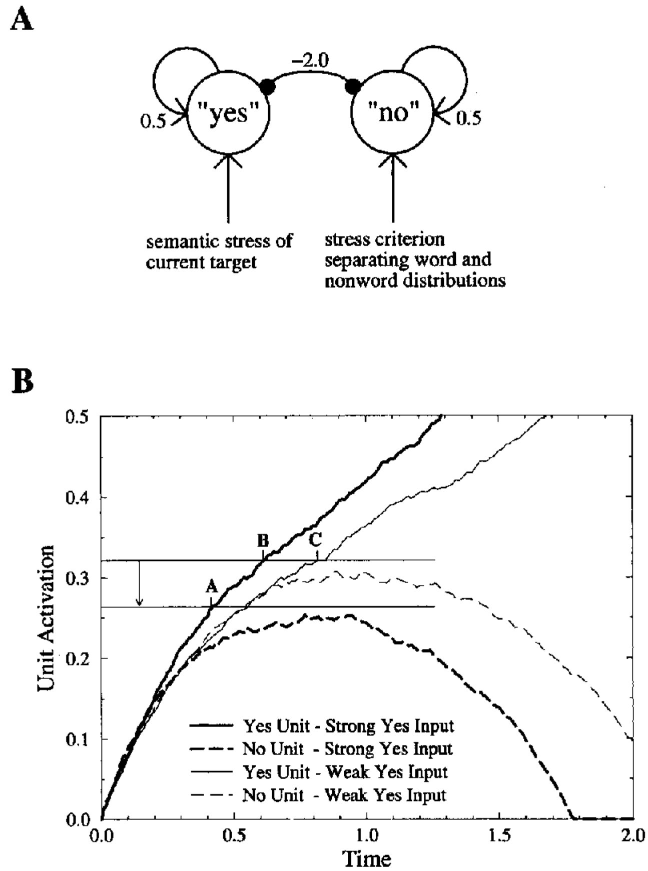

Figure 16. A: A depiction of competitive "yes" and "no" response units (Usher & McClelland, 1995, in press). B: Response unit activations when the input to the "yes" unit is relatively strong and much larger than the input to the "no" unit (i.e., 0.9 vs. 0.8, respectively) compared with when it is weaker and more similar to the "no" input (i.e., 0.85 vs. 0.8, respectively). These input values are intended to reflect the semantic stress for a target word ("yes" inputs) and for the stress criterion separating word and nonword responses ("no" inputs). In addition to these inputs, each unit has an excitatory weight of 0.5 from itself and an inhibitory weight of -2.0 from the other unit, and its activation is corrupted with noise (SD = 0.2) and integrated with time constant τ = 0.01. The horizontal lines indicate alternative response criteria. Note that, when strong and weak word inputs are mixed, responses to the former are faster (cf. B vs. C). When the strong word inputs are blocked, the response criterion can be shifted downward (indicated by the arrow) to yield even faster reaction times for these items (cf. A vs. B).

estimated from a running average of stress values across all stimuli). The "yes" and "no" units compete on the basis of the relative strength of their inputs, and the network responds when the activation of one of the units exceeds a threshold response criterion. An important property of this type of competitive response system is that the competition takes longer to resolve, and hence RTs are prolonged, when the inputs to the response units are of similar magnitude (i.e., when a word or nonword target produces a stress value close to the decision criterion separating the word and nonword distributions). Moreover, under experimental conditions in which the separation of the word and nonword distributions is increased, the competition generally resolves more quickly, and thus a more aggressive response criterion (i.e., lower activation threshold) can be used to speed overall responding while keeping error rates acceptably low (see Figure 16B). These characteristics play an important role in our accounts of stimulus blocking effects, as discussed below.

The second and perhaps more obvious simplification in the current implementation was that it did not include phonological representations and processes. This omission might seem particularly problematic in light of recent findings of very rapid phonological influences on lexical processing (e.g., Booth, Perfetti, & MacWhinney, 1999; Lukatela, 1994; Lukatela & Turvey, 1994; although see Jared & Seidenberg, 1991; Verstaen, Humphreys, Olson, & d'Ydewalle, 1995). In fact, other network models and empirical findings have illustrated the importance of examining the interaction among orthographic, phonological, and semantic representations when trying to account for behavioral data from naming and lexical decision tasks (see, e.g., Harm, 1998; Kawamoto, 1993; Plaut, 1997; Plaut et al., 1996; Stone, Vanhoy, & Van Orden, 1997; Strain et al., 1995; Van Orden & Goldinger, 1994; Van Orden, Pennington, & Stone, 1990). Thus, the current simulation, which involved only a mapping from orthography to semantics, cannot be expected to provide a full account of lexical processing in general, or even of lexical decision performance in particular.

Nonetheless, apart from considerations of issues that relate to phonology per se, such as pseudohomophone effects (McCann, Besner, & Davelaar, 1988), the central properties exhibited by the current implementation would also be expected to hold for a more general implementation that included phonology. The reason is that, in English, orthography and phonology bear a similar relationship to semantics—the similarity of monomorphemic words within each domain is essentially unrelated to their semantic similarity. As explained above, it is this lack of structure between the surface forms of words and their meanings that, on the current account, provides the most reliable basis for distinguishing words from nonwords. Thus, the rapid derivation of phonological information from orthography allows two unstructured mappings to contribute to performance instead of one, but it does not fundamentally alter the relative effectiveness of familiar versus novel surface forms (i.e., words vs. nonwords) to engage semantics. Of course, pseudohomophones (e.g., BRANE) are precisely those stimuli that violate the more general functional similarity of orthography and phonology, because they are orthographically unfamiliar but phonologically familiar. For exactly this reason, pseudohomophone effects in lexical decision are beyond the scope of the current implementation (but not outside the scope of the more general theoretical framework of distributed network models; see, e.g., Plaut, 1997; Seidenberg & McClelland, 1989).

Finally, the current model bases lexical decisions on stress calculated only over semantic representations. Although there is strong evidence that semantics plays an important role in lexical decision performance (see, e.g., Azuma & Van Orden, 1997; Balota, Ferraro, & Connor, 1991; Borowsky & Masson, 1996; Chumbley & Balota, 1984; Hino & Lupker, 1996; James, 1975; Millis & Button, 1989), it is also clear that readers can base lexical decisions, at least in part, on orthographic or phonological information, particularly when the nonword foils are relatively unwordlike (James, 1975; Waters & Seidenberg, 1985). Our nonword foils were, for the most part, orthographically legal, so reliance solely on orthography is unlikely. In a more comprehensive version of our account of lexical decision, we would assume that individuals can base their decisions on any available information in the lexical system and that they adopt a strategy that optimizes their performance given the composition of the stimuli (also see Seidenberg & McClelland, 1989). Moreover, given that orthographic information is available earlier than either phonological or semantic information, we would expect individuals to rely on orthographic information to whatever extent possible. Given our focus on semantics rather than orthography as a basis for lexical decision, we would not expect the current form of our model to account for the influence of orthographic factors, such as neighborhood density, on performance (see, e.g., Andrews, 1992; Sears, Hino, & Lupker, 1995).[14]

In summary, there are a number of ways in which the current implemented model falls short of a fully comprehensive account of lexical processing. Nonetheless, despite the simplifications incorporated into the model, we claim that the central computational principles underlying its performance can be extended to account for the full range of relevant phenomena.

Additional Empirical Issues

Having discussed limitations of our implemented model of semantic priming in lexical decision, we now turn to a consideration of related empirical phenomena that would seem to challenge our more general single-mechanism, distributed network account. These phenomena relate to blocking and strategy effects, priming across unrelated items, categorical versus associative priming, and backward associative priming. We devote considerable attention to blocking and strategy effects because these would seem to be the most problematic for our account.

[14] Note, however, that Sears, Hino, and Lupker (1999) recently carried out a detailed analysis of neighborhood effects in Seidenberg and McClelland's (1989) model and in the version of Plaut et al.'s (1996) model that included a semantic contribution to naming (Simulation 4), and they found that the models provided a good account of the facilitatory neighborhood size effect (Andrews, 1989, 1992) and the facilitatory neighborhood frequency effect (Sears et al., 1995; Sears, Lupker, & Hino, 1999). Furthermore, Andrews (1997) has argued that other, inhibitory neighborhood effects (e.g., Carreiras, Perea, & Grainger, 1997; Grainger & Jacobs, 1996; Grainger, O'Regan, Jacobs, & Segui, 1989) may arise from sophisticated guessing strategies in unusual stimulus environments or may be restricted to languages with very high spelling–sound consistency.

Blocking and Strategy Effects

It is often claimed that single-mechanism models cannot account for effects of experimental manipulations that induce apparent changes in the processing strategies adopted by individuals (see, e.g., Neely, 1991). These effects have been operationalized in a number of ways. The most common way is by manipulating properties of the experimental stimuli—for example, the similarity of nonword foils to words (Stone & Van Orden, 1993), the proportion of prime–target pairs that are related (Neely, 1977), or the proportion of pairs that are associatively versus categorically related (C. A. Becker, 1980). For example, increasing the proportion of related trials yields larger semantic priming effects at long but not short SOAs (Groot, 1984; Heyer, 1985; Neely, 1977; Seidenberg, Waters, Sanders, & Langer, 1984; Tweedy, Lapinski, & Schvaneveldt, 1977). Another type of manipulation is to compare the effects of blocking versus mixing experimental conditions for the same stimuli. For example, M. C. Smith et al. (1994) found that the magnitude of semantic priming at a short SOA depended on whether such trials were blocked or mixed with long SOA trials; adults exhibited priming at the short SOA in the blocked condition, but such priming effects were minimal in the mixed condition. Moreover, Booth and Plaut (2000) found that, when SOA was mixed, priming context interacted with target frequency only at the long SOA; there were no context effects at the short SOA. These findings suggest that semantic priming is at least partially subject to strategic effects and, therefore, is not entirely automatic.

It is clear, then, that the nature of processing a given target word depends not only on the immediately preceding (priming) context but also on more general aspects of the experimental situation. The critical question is what sort of processes must be postulated to account for these types of strategic effects and to what extent are they specific to the lexical system per se. Blocking and list context effects are typically explained by reference to changes in the operation of expectancy-based processes (see C. A. Becker, 1980; Neely, 1991; Neely & Keefe, 1989). The fact that such processes are assumed to be slow is used to account for why many strategic effects arise only at relatively long SOAs (but see M. C. Smith et al., 1994; Stolz & Besner, 1997; Stolz & Neely, 1995).

An alternative approach to explaining these types of effects, however, is to assume that individuals adjust the operation of decision processes, which are separate from the lexical system per se, as a function of the composition of the stimuli and testing conditions within the current block (see, e.g., Gordon, 1983; Grainger & Jacobs, 1996; Seidenberg, Waters, Sanders, & Langer, 1984; Stone & Van Orden, 1993). In particular, as suggested earlier, individuals might adjust the response criterion within a competitive response system (Usher & McClelland, 1995, in press; see Figure 16A).[15] Plaut (1997; also see Gordon, 1983; Seidenberg & McClelland, 1989) has argued that, if lexical decisions are based on semantic stress, this approach can provide an account of the frequency blocking effect (Glanzer & Ehrenreich, 1979), in which RTs to high- but not low-frequency targets are reduced under blocked compared with mixed presentation. The essence of the account is that, because high-frequency words tend to produce higher stress values than low-frequency words, a more conservative response criterion is needed to produce acceptable levels of accuracy when low-frequency words are among the stimuli (whether blocked or mixed). However, when high-frequency words are blocked, there is greater separation between the word and nonword distributions, so that a more aggressive response criterion can be adopted that, for the same error rate, produces faster responding (see Figure 16B). The same form of account can explain why the effects of target frequency are increased when nonword foils are more wordlike (Stone & Van Orden, 1993).

An analogous account may explain the SOA blocking effects (Booth & Plaut, 2000; M. C. Smith et al., 1994; Stolz & Besner, 1997).[16] In our model, RTs are faster and priming effects are weaker at the short SOA because, compared with the long SOA, the network does not settle as deeply into the prime's attractor basin, so there is less hysteresis in moving from the pattern produced by the prime to the representation of the target. Moreover, given that the network is settling to a binary semantic pattern, the fact that the network settles faster at the short SOA means that, at any point in time following the presentation of the target, the value of semantic stress is generally higher at the short versus long SOA. By contrast, the stress levels for nonword targets are relatively unaffected by manipulation of SOA because they are not settling to patterns that are as binary (see Figure 10). Consequently, there is greater separation between the distributions of stress values for words and nonwords when short-SOA trials are blocked compared with when they are mixed with long-SOA trials. Following the logic used in explaining the frequency blocking effect, a more conservative response criterion is required in the mixed versus blocked condition for short SOA trials. Under this more conservative criterion, the relative difference in stress for targets following related versus unrelated primes is reduced (compared to the blocked condition), thereby reducing and perhaps even eliminating priming effects at the short SOA in the mixed condition. By contrast, essentially the same response criterion is needed for long-SOA trials whether they are mixed or blocked, so there is no effect of this manipulation on the magnitude of semantic priming at long SOAs.

Now consider the findings that increasing the proportion of related trials produces stronger semantic priming (e.g., Groot, 1984; Heyer, 1985; Tweedy et al., 1977). Note that the basic semantic priming effect is that RTs to targets are faster following related compared with unrelated primes. Based on the argument just presented for short versus long SOAs, this means that, at a

[15] Similarly, with regard to the naming task, Lupker, Brown, and Colombo (1997) and Jared (1997) have provided evidence that shifts in a criterion for when to initiate articulation provides a better account of blocking effects in naming (e.g., Monsell, Patterson, Graham, Hughes, & Milroy, 1992; Paap & Noel, 1991) than accounts that rely on changes in the relative contribution of lexical and sublexical pathways. Note, however, that Lupker and colleagues proposed a time-based criterion, whereas the current proposal involves an activation-based criterion (see also Kello & Plaut, 2000).

[16] M. C. Smith et al. (1994) provided an explanation for the SOA blocking effect, which they and Stolz and Besner (1997) described as a "signal-detection" account. This account is, however, rather different than the current proposal, in that it postulates the adjustment of a criterion that controls whether lexical activation spreads from the orthographic system to the semantic system. By contrast, the relevant criterion on the current account is an activation threshold applied to "yes" and "no" units within a competitive response system (Usher & McClelland, 1995, in press), which is not considered part of the lexical system per se.

given point in processing the target, stress values are generally higher following related versus unrelated primes. Thus, increasing the proportion of related trials permits a more aggressive response criterion, which increases the relative difference in stress values for the related versus unrelated priming conditions, thereby leading to a larger priming effect. Relatedness proportion may not affect semantic priming at short SOAs because, compared with long SOAs, the basic priming effects are weaker at short SOAs (see Neely, 1991) and thus are less susceptible to modulation.

To be clear, the proposals we have outlined here are not fully adequate accounts of the relevant phenomena; rather, they are intended to sketch out an approach to explaining these effects that relies only on the adjustment of a response criterion, rather than on the existence of complicated expectancy-based processes. It seems unlikely, however, that changes in a response criterion provide an account of other types of strategic effects—particularly those involving changes in the instructions to individuals (Favreau & Segalowitz, 1983; Neely, 1977) and other depth-of-processing manipulations (e.g., Henik, Freidrich, & Kellogg, 1983; Kaye & Brown, 1985; M. C. Smith, Theodor, & Franklin, 1983; Stolz & Besner, 1996). Instructions clearly must induce strategic effects on lexical processing at some level; otherwise, how can it be that, when presented with the same stimulus, individuals perform lexical decision in one experiment and naming or letter search in others? Thus, the relevant question is not whether higher level strategies influence processing, but rather whether the lexical system itself must incorporate a strategic mechanism such as C. A. Becker's (1980) generation of expectancy sets or Neely and Keefe's (1989) retrospective semantic matching. Our position is that the mechanisms that underlie these types of strategy changes apply generally across all cognitive domains and are not specific to the lexical system (see also Kello & Plaut, 2000). In fact, on the current account, even the response criterion is not part of the lexical system but is part of a more general cognitive mechanism for making forced-choice decisions (Ratcliff, 1978; Usher & McClelland, in press). Thus, although we do not, at present, have a fully adequate theory of the nature and operation of these general-purpose mechanisms, we can nonetheless make progress in articulating the principles of operation of the lexical system quite apart from such a theory. In this way, our model of the lexical system per se remains a single-mechanism account.

Priming Across Unrelated Items

Another apparent challenge for distributed network models is to account for the finding that associative priming can span an intervening item, such as in the word sequence NURSE → CAT → DOCTOR (e.g., Joordens & Besner, 1992; McNamara, 1992; Meyer & Schvaneveldt, 1971). Although these priming effects are weak, they have challenged distributed network models because if the network settles completely to the meaning of the intervening word CAT, then the pattern of activity representing the meaning of NURSE is completely eliminated, leaving no opportunity for it to facilitate the processing of DOCTOR. However, the intervening word might be processed only partially, leaving residual semantic activation for NURSE to influence processing of DOCTOR (Masson, 1995). Indeed, Plaut (1995) showed that a distributed network model exhibited associative priming across an intervening item, particularly under conditions that encourage fast responding (also see Masson, 1995). Therefore, distributed network models seem able to account for the existence of priming across an intervening item without recourse to another mechanism.

Categorical Versus Associative Priming

A model of semantic priming should also account for the time course of facilitation and inhibition in categorical versus associative priming. At short SOAs, there is facilitation dominance for both categorical and associative priming, whereas at long SOAs there is facilitation dominance for associative priming, but inhibition dominance for categorical priming (Heyer et al., 1985; L. C. Smith et al., 1987). Plaut (1995) showed that a distributed network model exhibited greater associative priming with longer SOAs, and that same model also exhibited a decrease in categorical priming from short to long SOAs.

Although the Plaut (1995) model replicated the basic finding in the literature, conclusions could not be drawn about the relative magnitudes of facilitation and inhibition because the performance of the model was not evaluated relative to a neutral baseline condition. Although the current simulation did use a neutral (nonword) priming baseline, it involved a training environment with complete co-occurrence of categorical and associative relatedness and, thus, it cannot be used to evaluate the relative contributions of these factors.

However, a version of our simulation that at least partially separated categorical and associative relatedness might be able to account for the time course of these types of priming. Associative priming occurs in our model because, during training, the network learned to make a rapid transition from the representation of a prime to that of a target. This learned transition produces strong facilitation, which increases with SOA because the representation generated by the prime becomes increasingly accurate. Note, however, that actual words typically have only one or at most a few strong associates; thus, the majority of words that precede a given target word during training are unrelated to it. As a result, there is minimal inhibition from unrelated primes because the network has learned to ignore nonoverlapping previous patterns except in the few specific cases involving associative relatedness.

On the other hand, categorical priming occurs in a distributed network model because a related prime activates features that overlap with those of the target. Categorical facilitation tends to be weak because only some features overlap between the prime and target, whereas categorical inhibition tends to be strong because many features do not overlap. In fact, this may explain why some authors have not found priming for categorically but not associatively related prime–target pairs (Moss et al., 1995; Shelton & Martin, 1992) and why others have found pure categorical priming only when the prime–target pairs are very highly related (Lund, Burgess, & Atchley, 1995; McRae & Boisvert, 1998; Perea & Gotor, 1997) or, among children, only for those with good reading comprehension (Nation & Snowling, 1999). Categorical priming shows inhibition dominance at longer SOAs on this account because, with additional processing, semantic units that differ between the prime and target are driven to more extreme values. In order to correctly identify the target, all of these differences must be corrected, so the magnitude of inhibition is greater at longer SOAs.

Backward Associative Priming

A model of visual word recognition should be able to account not only for forward but also backward associative priming (Chwilla, Hagoort, & Brown, 1998; Kahan, Neely, & Forsythe, 1999; Koriat, 1981; Peterson & Simpson, 1989; Seidenberg, Waters, Sanders, & Langer, 1984; Thompson-Schill, Kurtz, & Gabrieli, 1998). In backward priming, the prime and target are related only through a backward association from target to prime (RACK → COAT) rather than a forward association from prime to target. For example, Kahan et al. (1999) found robust backward priming in lexical decision at both a short (150 ms) and long (500 ms) SOA whereas, in naming, backward priming was much weaker and occurred only at the short SOA. These results held both for associates that form compound words (e.g., COATRACK) and for asymmetrically associated noncompounds (CRY → ONION; note that CRY is often given as an associate of ONION, but not vice versa). One of the greatest triumphs of compound-cue theory (Ratcliff & McKoon, 1988) is that it can account for backward priming by assuming that individuals use a familiarity value of the prime–target combination in order to make lexical decisions to the target. Compound-cue theory has not been extended to account for the different pattern of backward priming in naming. Can distributed network models account for these paradigm differences in backward priming?

The current simulation cannot be used to test for the existence of backward priming because, as just pointed out, associative relatedness always co-occurred with categorical relatedness and, in our formulation, the latter involves a symmetric relationship. However, Plaut (1995) did not find backward associative priming in a network in which associated prime–target pairs were not categorically related. Thus, it seems unlikely that backward associative priming is an inherent consequence of the presence of forward associations in a distributed network model.

An alternative possibility is that backward priming arises not from an associative relationship but from feature overlap. In English, the type of object referred to by a compound is determined by its second component (e.g., a coatrack is a type of rack; Marchand, 1969). Thus, the representation generated by the prime RACK should overlap with that of COATRACK because of a categorical relationship. This representation, in turn, would facilitate the processing of COAT because of the typical semantic or functional relationship between compounds and their first components (e.g., a coatrack is a rack used for coats; Moss et al., 1995). For noncompound associates, the prime often designates a salient or distinctive property of the target (e.g., an important characteristic of onions is that they make people cry). If such information is included in the semantic representation of ONION, then pre-activating CRY should facilitate the processing of ONION. In this way, a small amount of feature overlap between prime and target may explain backward priming in lexical decision and in naming at short SOAs.[17] That the latter is weaker is consistent with the general observation that standard semantic priming is weaker in naming than in lexical decision (see Neely, 1991). As elaborated below, we ascribe this difference to the following: Because the spelling–sound consistency in English is relatively high, the contribution of semantics to naming is relatively weak for most words (other than low-frequency exception words; see Plaut et al., 1996).

Why, then, is backward priming in naming not just weak but absent at long SOAs? One possibility is that, at a long SOA, the system has had sufficient time to generate the full phonology of the prime before the target is presented. There is good evidence from cross-modal priming of picture naming (e.g., Levelt et al., 1991; Schriefers, 1992) that pre-activation of a competing phonological representation interferes with the generation of a naming response. This interference from the phonology of the prime may be sufficient to eliminate the weak facilitation from feature overlap with the target. By contrast, standard priming in naming at long SOAs survives in the face of this interference because of the much stronger categorical or associative relationship between prime and target than in the backward priming paradigm.

Extensions of the Approach

The most natural extension of the current work, both empirically and computationally, would be to address semantic priming effects in naming. Semantic priming is generally weaker in naming than in lexical decision, often being present only in a subset of the conditions that produce semantic priming in lexical decision (see Neely, 1991). One possible exception to this pattern is mediated priming (i.e., LION → STRIPES, presumably through TIGER), which has been observed in naming but not in lexical decision when tested under comparable conditions (Balota & Lorch, 1986; but see McKoon & Ratcliff, 1992; McNamara, 1992; McNamara & Altarriba, 1988; Shelton & Martin, 1992).

Semantic priming effects in naming could be addressed within a distributed network model in which orthographic, phonological, and semantic representations interact to settle simultaneously on the appropriate meaning and pronunciation for written words (Harm, 1998; Kawamoto, 1993; Plaut et al., 1996; Seidenberg & McClelland, 1989). The current approach to modeling semantic priming in lexical decision could be extended directly to the naming task by basing responses on phonological rather than semantic activation and by allowing residual activation from a previous stimulus to influence the processing of a current target word.

An important aspect of the English lexical system is that there is a high degree of systematicity between orthography and phonology but very little systematicity between either of these and semantics (at a monomorphemic level). A fundamental property of distributed network models is that they learn systematic mappings more quickly and strongly than unsystematic mappings (Plaut et al., 1996; Van Orden & Goldinger, 1994; Van Orden et al., 1990). As a result, whereas orthographic input activates both phonological and semantic representations simultaneously, the phonological representations settle far more quickly and are less sensitive to pre-existing activation (see, e.g., Harm, 1998; Kawamoto, 1993; Kawamoto & Zemblige, 1992). These properties provide a natural account of why semantic priming effects are weaker in naming (based on phonological activation) than in lexical decision (based

[17] Note that our explanation of backward priming in terms of feature overlap does not imply that the prime–target pairs (e.g., RACK–COAT, CRY–ONION) would be considered directly related in semantic similarity ratings. For both compound and noncompound pairs, the association that underlies the relatedness of the concepts in backward priming is not likely to be considered by individuals when generating similarity ratings.

on semantic activation). Consistent with this account, naming performance can exhibit similar semantic priming effects to that found in lexical decision if slowed to a comparable rate by experimental manipulations such as target masking (Flores d'Arcais, Schreuder, & Glazenborg, 1985) or lateralized presentation (Chiarello, Burgess, Richards, & Pollock, 1990).

A number of predictions concerning semantic priming in a distributed network model of naming arise directly from the greater systematicity of the orthography–phonology mapping compared with the orthography–semantics mapping. First, given that semantic priming effects are weaker in naming, effects of variables such as perceptual ability and target frequency on priming should also be weaker in naming than in lexical decision. Second, given that generating the semantic representation of the prime is relatively slow compared with generating its phonology, semantic priming effects in naming should be larger at long compared with short SOAs (although Cortese, Simpson, & Woolsey, 1997, did not find an interaction of context and SOA). More generally, effects of semantic priming should be larger when the contributions to phonology of other factors or combination of factors are weaker. For example, given that the derivation of phonology for low-frequency exception words is slowed relative to items higher in either word frequency or spelling–sound regularity or consistency (Andrews, 1982; Seidenberg, Waters, Barnes, & Tanenhaus, 1984; Taraban & McClelland, 1987; Waters & Seidenberg, 1985), there should greater semantic priming when naming these items—that is, there should be a three-way interaction of frequency, consistency, and priming context. This prediction mirrors the three-way Frequency × Consistency × Imageability in Naming interaction found by Strain et al. (1995). In both cases, the effect of semantics is greatest for items with the weakest spelling–sound mapping—low-frequency exception words. Indeed, Cortese et al. (1997) recently investigated semantic priming among low-frequency targets varying in regularity and imageability. They found a two-way Context × Regularity interaction, such that irregular targets exhibited greater semantic priming than regular words (although see Kawamoto, Goeltz, Agbayani, & Groel, 1998, for evidence that this effect may be restricted to onset durations rather than latencies). However, the interaction of context and imageability—greater priming for low-imageable words—was only marginal, and Cortese and colleagues did not report tests of the three-way interaction.

A model of reading acquisition with differential development of the orthographic–phonological mapping versus the orthographic–semantic mapping would have important, but untested, developmental implications (also see Share, 1995). We have suggested that good readers show larger semantic priming effects in lexical decision than in naming (Keefe & Neely, 1990; Lorch et al., 1986; Lupker, 1984) because their well-developed spelling–sound mapping allows them to pronounce words rapidly, thereby reducing the effects of semantics on naming. However, poor readers should exhibit comparable semantic priming effects in lexical decision and naming because their underdeveloped spelling–sound mapping allows semantic information to influence their slow naming processes. A differential development model also predicts that any factor that increases the ability of semantics to influence naming, such as reading low-frequency exception words at long SOAs, increases priming effects in naming for good readers. By contrast, these factors should minimally influence the magnitude of semantic priming effects in naming for poor readers because the orthographic system does not strongly drive the phonological system, so semantics can influence naming regardless of frequency, regularity, or SOA.

Notice that our argument here is very similar to the one formulated to explain the cross-linguistic semantic priming differences in naming. For example, Katz and Feldman (1983) and Frost, Katz, and Bentin (1987) compared semantic priming effects on naming in English with the effects in Serbo-Croatian, a more shallow orthography with greater spelling–sound consistency, and with the effects in unpointed Hebrew, a deeper, less consistent orthography. They found no semantic priming in Serbo-Croatian and greater priming in Hebrew than in English. These findings can be understood as natural consequences of the basic properties of distributed network models, given their sensitivity to the relative degree of systematicity of the orthography–phonology mapping (see Seidenberg, 1992, for a discussion). Specifically, across languages, greater systematicity should produce weaker semantic effects on naming.

Our empirical studies and computational simulation showed that the adult condition exhibited about half as much priming as the child condition. Our finding of age-related decreases in the magnitude of semantic priming is supported by other empirical studies (Schwantes, 1981; Simpson & Lorsbach, 1983; West & Stanovich, 1978). It appears that, in English, higher level (semantic) information influences word recognition less as children become skilled readers. By contrast, lower level (orthographic and phonological) information appears to influence the reading process more as children develop. Only two earlier studies of single-word priming have directly examined the relative influences of orthographic and phonological processes in children's visual word recognition (Goswami, 1990; Hansen & Bowey, 1992). However, these studies did not examine developmental differences. More recently, Booth et al. (1999) have shown that there is a strong positive relationship between the magnitude of orthographic–phonological priming and both naming accuracy and age. Moreover, older and high-ability children can activate this orthographic and phonological information more quickly than younger and low-ability children. Unfortunately, no investigation to date has examined developmental differences in orthographic, phonological, and semantic priming effects in a single group of children.

One possibility is that beginning readers compensate for their deficient knowledge of spelling–sound correspondences by bringing to bear semantic knowledge about the world (Nation & Snowling, 1998a, 1998b). However, as children learn the statistical regularities between phonology and orthography, they rely less on semantics and more on interactions between orthographic and phonological representations for rapid word recognition. These developmental differences have important implications for models of visual word recognition because reading acquisition does not consist simply of age-related increases in all component skills. Rather, some effects, such as semantic priming, appear to decrease with age in English, whereas other effects, such as orthographic and phonological priming, appear to increase with development.

Kang and Simpson (1996) have demonstrated a pattern of developmental effects in Korean, a shallow orthography like Serbo-Croatian, that are exactly the opposite of those found in English. Specifically, children learning to read Korean appear to exhibit a decrease in phonological priming and an increase in semantic priming with age. Korean-reading children may show greater phonological priming than English-reading children in the initial phases of learning to read because they benefit from the greater spelling–sound consistency of Korean. Over the course of development, however, mappings with semantics become stronger, so meaning has a larger influence on word recognition and this reduces the phonological priming effect. Unfortunately, to our knowledge, there is no well-controlled cross-linguistic study that directly examines developmental differences in phonological and semantic priming.

Conclusion

Semantic priming phenomena have played a critical role in constraining theories of lexical processing. It is almost universally accepted that a comprehensive account of these phenomena must incorporate multiple mechanisms. Single-mechanism accounts of semantic priming phenomena were abandoned by most researchers because they did not seem capable of accounting for strategic effects (Neely, 1991). A central goal of the current work is to reconsider this conclusion in light of more recent progress in understanding the computational properties of distributed network models and in applying them to complex empirical phenomena.

Our empirical work demonstrates that frequency effects on semantic priming depend on perceptual ability. Our computational work demonstrates that this pattern of data is a natural consequence of the nonlinear effects within a distributed network model that derives the meanings of written words. The model also exhibits the shift from facilitation dominance at short SOA to inhibition dominance at long SOA, without recourse to expectancy-based processes. Moreover, other phenomena thought to implicate strategic processes may instead be explained in terms of shifts in response criteria for decision processes outside the lexical system. It is certainly the case that considerable work remains to be done in order to extend the current approach into a full account of semantic priming phenomena in lexical processing. Even so, the relative success of a single-mechanism, distributed network model in accounting for data that have heretofore been taken to necessitate additional, expectancy-based processes suggests that such models may provide a viable alternative to multiple-mechanism accounts of lexical processing.

References

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Andrews, S. (1982). Phonological recoding: Is the regularity effect consistent? *Memory & Cognition, 10,* 565–575.

Andrews, S. (1989). Frequency and neighborhood effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 802–814.

Andrews, S. (1992). Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 234–254.

Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin & Review, 4,* 439–461.

Antos, S. J. (1979). Processing facilitation in a lexical decision task. *Journal of Experimental Psychology: Human Perception and Performance, 5,* 527–545.

Atkinson, R. C., & Juola, J. F. (1973). Factors influencing speed and accuracy of word recognition. In S. Kornblum (Ed.), *Attention and performance IV* (pp. 583–612). New York: Academic Press.

Azuma, T., & Van Orden, G. C. (1997). Why SAFE is better than FAST: The relatedness of a word's meanings affects lexical decision times. *Journal of Memory and Language, 36,* 484–504.

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.

Balota, D. A., Ferraro, F. R., & Connor, L. T. (1991). On the early influence of meanings in word recognition: A review of the literature. In P. Schwanenflugel (Ed.), *The psychology of word meanings* (pp. 187–222). Hillsdale, NJ: Erlbaum.

Balota, D. A., & Lorch, R. F. (1986). Depth of automatic spreading activation: Mediated priming effects in pronunciation but not in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 336–345.

Becker, C. A. (1979). Semantic context and word frequency effects in visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 5,* 252–259.

Becker, C. A. (1980). Semantic context effects in visual word recognition: An analysis of semantic strategies. *Memory & Cognition, 8,* 493–512.

Becker, S., Moscovitch, M., Behrmann, M., & Joordens, S. (1997). Long-term semantic priming: A computational account and empirical evidence. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 1059–1082.

Biemiller, A. (1970). The development and use of graphic and contextual information as children learn to read. *Reading Research Quarterly, 6,* 75–96.

Booth, J. R., Hunt, S. B., Perfetti, C. A., & MacWhinney, B. (1998, April). *The influence of rapid sequential perception on orthographic and phonological processing in reading disabled children and adults.* Paper presented at the meeting of the Society for the Scientific Study of Reading, San Diego, CA.

Booth, J. R., Perfetti, C. A., & MacWhinney, B. (1999). Quick, automatic, and general activation of orthographic and phonological representations in young readers. *Developmental Psychology, 35,* 3–19.

Booth, J. R., & Plaut, D. C. (2000). *Blocking effects in naming and lexical decision: Further support for a single-mechanism account of lexical processing.* Manuscript in preparation.

Borowsky, R., & Besner, D. (1993). Visual word recognition: A multistage activation model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 813–840.

Borowsky, R., & Masson, M. E. J. (1996). Semantic ambiguity effects in word identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 63–85.

Carreiras, M., Perea, M., & Grainger, J. (1997). Effects of orthographic neighborhood in visual word recognition: Cross-task comparisons. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 857–871.

Chauvin, Y. (1988). *Symbol acquisition in humans and neural (PDP) networks.* Unpublished doctoral dissertation, University of California, San Diego.

Chiarello, C., Burgess, C., Richards, L., & Pollock, A. (1990). Semantic and associative priming in the cerebral hemispheres: Some words do, some words don't . . . sometimes, some places. *Brain and Language, 38,* 75–104.

Chumbley, J. I., & Balota, D. A. (1984). A word's meaning affects the decision in lexical decision. *Memory & Cognition, 12,* 590–606.

Chwilla, D. J., Hagoort, P., & Brown, C. M. (1998). The mechanism underlying backward priming in a lexical decision task: Spreading activation versus semantic matching. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 51*(A), 531–568.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review, 97,* 332–361.

Cortese, M. J., Simpson, G. B., & Woolsey, S. (1997). Effects of association and imageability on phonological mapping. *Psychonomic Bulletin & Review, 4,* 226–231.

Cree, G. S., McRae, K., & McNorgan, C. (1999). An attractor model of lexical conceptual processing: Simulating semantic priming. *Cognitive Science, 23,* 371–414.

Detterman, D. K., & Daniel, M. H. (1991). Correlations of mental tests with each other and with cognitive variables are highest with low IQ groups. *Intelligence, 15,* 247–250.

Dosher, B. A., & Rosedale, G. (1989). Integrated retrieval cues as a mechanism for priming in retrieval from memory. *Journal of Experimental Psychology: General, 2,* 191–211.

Dunn, L. M., & Dunn, L. M. (1981). *Peabody Picture Vocabulary Test—Revised.* Circle Pines, MN: American Guidance Service.

Eden, G. F., VanMeter, J. W., Rumsey, J. M., Maisog, J. M., Woods, R. P., & Zeffiro, T. A. (1996). Abnormal processing of visual motion in dyslexia revealed by functional brain imaging. *Nature, 382,* 66–69.

Epelboim, J., Booth, J. R., & Steinman, R. M. (1994). Reading unspaced text: Implications for theories of reading eye movements. *Vision Research, 34,* 1735–1766.

Epelboim, J., Booth, J. R., & Steinman, R. M. (1996). Much ado about nothing: The place of space in text. *Vision Research, 36,* 465–470.

Farmer, M. E., & Klein, R. M. (1995). The evidence for a temporal processing deficit linked to dyslexia: A review. *Psychonomic Bulletin & Review, 2,* 460–493.

Favreau, M., & Segalowitz, N. S. (1983). Automatic and controlled processes in the first- and second-language reading of fluent bilinguals. *Memory & Cognition, 11,* 565–574.

Flores d'Arcais, G. B., Schreuder, R., & Glazenborg, G. (1985). Semantic activation during recognition of referential words. *Psychological Research, 47,* 39–49.

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.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91,* 1–67.

Glanzer, M., & Ehrenreich, S. L. (1979). Structure and search of the internal lexicon. *Journal of Verbal Learning and Verbal Behaviour, 18,* 381–398.

Goldsmith-Phillips, J. (1989). Word and context in reading development: A test of the interactive-compensatory hypothesis. *Journal of Educational Psychology, 81,* 299–305.

Golinkoff, R. M., & Rosinki, R. R. (1976). Decoding, semantic processing, and reading comprehension skill. *Child Development, 47,* 252–258.

Gordon, B. (1983). Lexical access and lexical decision: Mechanisms of frequency sensitivity. *Journal of Verbal Learning and Verbal Behaviour, 22,* 24–44.

Goswami, U. (1990). Phonological priming and orthographic analogies in reading. *Journal of Experimental Child Psychology, 49,* 323–340.

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

Grainger, J., O'Regan, J. K., Jacobs, A. M., & Segui, J. (1989). On the role of competing word units in visual word recognition: The neighbourhood frequency effect. *Perception and Psychophysics, 45,* 189–195.

Groot, A. M. B. de (1984). Primed lexical decision: Combined effects of the proportion of related prime–target pairs and the stimulus-onset asynchrony of prime and target. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 36*(A), 253–280.

Guttentag, R. E., & Haith, M. M. (1979). A developmental study of automatic word processing in a picture classification task. *Child Development, 50,* 196–206.

Hansen, J., & Bowey, J. A. (1992). Orthographic rimes as functional units of reading in fourth-grade children. *Australian Journal of Psychology, 44,* 37–44.

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

Henik, A., Freidrich, F. J., & Kellogg, W. A. (1983). The dependence of semantic relatedness effects upon prime processing. *Memory & Cognition, 11,* 366–373.

Heyer, K. den. (1985). On the nature of the proportion effect in semantic priming. *Acta Psychologica, 60,* 25–38.

Heyer, K. den, Briand, K., & Smith, L. (1985). Automatic and strategic factors in semantic priming: An examination of Becker's model. *Memory & Cognition, 13,* 228–232.

Hino, Y., & Lupker, S. J. (1996). Effects of polysemy in lexical decision and naming: An alternative to lexical access accounts. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 1331–1356.

Hinton, G. E. (1989). Connectionist learning procedures. *Artificial Intelligence, 40,* 185–234.

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

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Science, USA, 79,* 2554–2558.

Jackson, N. E., & Biemiller, A. J. (1985). Letter, word, and text reading times of precocious and average readers. *Child Development, 56,* 196–206.

James, C. T. (1975). The role of semantic information in lexical decisions. *Journal of Experimental Psychology: Human Perception and Performance, 104,* 130–136.

Jared, D. (1997). Evidence that strategy effects in word naming reflect changes in output timing rather than changes in processing route. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 1424–1438.

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

Jonides, J., & Mack, R. (1984). On the cost and benefit of cost and benefit. *Psychological Bulletin, 96,* 29–44.

Joordens, S., & Becker, S. (1997). The long and short of semantic priming effects in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 1083–1105.

Joordens, S., & Besner, D. (1992). Priming effects that span an intervening unrelated word: Implications for models of memory representation and retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 483–491.

Kahan, T. A., Neely, J. H., & Forsythe, W. J. (1999). Dissociated backward priming effects in lexical decision and pronunciation tasks. *Psychonomic Bulletin & Review, 6,* 105–110.

Kang, H., & Simpson, G. B. (1996). Development of semantic and phonological priming in a shallow orthography. *Developmental Psychology, 32,* 860–866.

Katz, W. F., & Feldman, L. B. (1983). Relation between pronunciation and recognition of printed words in deep and shallow orthographies. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 157–166.

Kawamoto, A. H. (1993). Nonlinear dynamics in the resolution of lexical ambiguity: A parallel distributed processing approach. *Journal of Memory and Language, 32,* 474–516.

Kawamoto, A. H., Goeltz, K., Agbayani, J. T., & Groel, K. (1998). Locus of semantic priming effects in speeded naming. *Psychonomic Bulletin & Review, 5,* 676–682.

Kawamoto, A. H., & Zemblige, J. H. (1992). Pronunciation of homographs. *Journal of Memory and Language, 31,* 349–374.

Kaye, D. B., & Brown, S. W. (1985). Levels and speed of processing effects on word analysis. *Memory & Cognition, 13,* 425–434.

Keefe, D. E., & Neely, J. H. (1990). Semantic priming in a pronunciation task: The role of prospective prime-generated expectancies. *Memory & Cognition, 18,* 289–298.

Kello, C. T., & Plaut, D. C. (2000). Strategic control in word reading: Evidence from speeded responding in the tempo naming task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 719–750.

Koriat, A. (1981). Semantic facilitation in lexical decision as a function of prime–target association. *Memory & Cognition, 9,* 53–65.

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

Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T., & Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological Review, 98,* 122–142.

Lorch, R. F., Balota, D., & Stamm, E. (1986). Locus of inhibition effects in the priming of lexical decisions: Pre- or post-lexical access? *Memory & Cognition, 14,* 95–103.

Lovegrove, W., Martin, F., & Slaghuis, W. (1986). A theoretical and experimental case for a visual deficit in specific reading disability. *Cognitive Neuropsychology, 3,* 225–267.

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

Lund, K., Burgess, C., & Atchley, R. A. (1995). Semantic and associative priming in high-dimensional semantic space. In Cognitive Science Society (Ed.), *Proceedings of the 17th Annual Conference of the Cognitive Science Society* (pp. 660–665). Hillsdale, NJ: Erlbaum.

Lupker, S. J. (1984). Semantic priming without association: A second look. *Journal of Verbal Learning and Verbal Behaviour, 23,* 709–733.

Lupker, S. J., Brown, P., & Colombo, L. (1997). Strategic control in a naming task: Changing routes or changing deadlines? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 570–590.

Marchand, H. (Ed.). (1969). *The categories and types of present-day English word-formation.* Munich, Germany: C. H. Becksche Verlagsbuchhandlung.

Masson, M. E. J. (1995). A distributed memory model of semantic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 3–23.

Masson, M. E. J., & Borowsky, R. (1998). More than meets the eye: Context effects in word identification. *Memory & Cognition, 26,* 1245–1269.

McCann, R. S., Besner, D., & Davelaar, E. (1988). Word recognition and identification: Do word frequency effects reflect lexical access? *Journal of Experimental Psychology: Human Perception and Performance, 14,* 693–706.

McClelland, J. L., & Rumelhart, D. E. (1988). *Explorations in parallel distributed processing: A handbook of models, programs, and exercises.* Cambridge, MA: MIT Press.

McClelland, J. L., Rumelhart, D. E., & the PDP Research Group. (Eds.). (1986). *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 2. Psychological and biological models.* Cambridge, MA: MIT Press.

McKoon, G., & Ratcliff, R. (1992). Spreading activation versus compound cue accounts of priming: Mediated priming revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 1155–1172.

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

McNamara, T. P. (1992). Theories of priming: I. Associative distance and lag. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 1173–1190.

McNamara, T. P. (1994). Theories of priming: II. Types of primes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 507–520.

McNamara, T. P., & Altarriba, J. (1988). Depth of spreading activation revisited: Semantic mediated priming occurs in lexical decisions. *Journal of Memory and Language, 27,* 545–559.

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

McRae, K., Sa, V. R. de, & 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. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90,* 227–234.

Meyer, D. E., Schvaneveldt, R. W., & Ruddy, M. G. (1975). Loci of contextual effects on visual word recognition. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and performance V* (pp. 98–118). New York: Academic Press.

Millis, M. L., & Button, S. B. (1989). The effect of polysemy on lexical decision time: Now you see it, now you don't. *Memory & Cognition, 17,* 141–147.

Monsell, S. (1991). The nature and locus of word frequency effects in reading. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 148–197). Hillsdale, NJ: Erlbaum.

Monsell, S., Patterson, K., Graham, A., Hughes, C. H., & Milroy, R. (1992). Lexical and sublexical translation of spelling to sound: Strategic anticipation of lexical status. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 452–467.

Moss, H. E., Hare, M. L., Day, P., & Tyler, L. K. (1994). A distributed memory model of the associative boost in semantic priming. *Connection Science, 6,* 413–427.

Moss, H. E., Ostrin, R. K., Tyler, L. K., & Marslen-Wilson, W. D. (1995). Accessing different types of lexical semantic information: Evidence from priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 863–883.

Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review, 89,* 609–626.

Nation, K., & Snowling, M. J. (1998a). Individual differences in contextual facilitation: Evidence from dyslexia and poor reading comprehension. *Child Development, 69,* 996–1011.

Nation, K., & Snowling, M. J. (1998b). Semantic processing and the development of word-recognition skills: Evidence from children with reading comprehension difficulties. *Journal of Memory and Language, 39,* 85–101.

Nation, K., & Snowling, M. J. (1999). Developmental differences in sensitivity to semantic relations among good and poor comprehenders: Evidence from semantic priming. *Cognition, 70,* B1–B13.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited capacity attention. *Journal of Experimental Psychology: General, 106,* 226–254.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading* (pp. 264–336). Hillsdale, NJ: Erlbaum.

Neely, J. H., & Keefe, D. E. (1989). Semantic context effects on visual word processing: A hybrid prospective/retrospective processing theory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 24, pp. 207–248). New York: Academic Press.

Neely, J. H., Keefe, D. E., & Ross, K. (1989). Semantic priming in a lexical decision task: Roles of prospective prime-generated expectancies and retrospective semantic matching. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 1003–1019.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. (1994). *The University of South Florida word association, rhyme, and word fragment norms.* Unpublished manuscript.

Paap, K. R., Newsome, S. L., McDonald, J. E., & Schvaneveldt, 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.

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

Perea, M., & Gotor, A. (1997). Associative and semantic priming effects occur at very short stimulus-onset asynchronies in lexical decision and naming. *Cognition, 62,* 223–230.

Perfetti, C. A., & Hogaboam, T. (1975). Relationship between single word decoding and reading comprehension skill. *Journal of Educational Psychology, 67,* 461–469.

Perfetti, C. A., & Lesgold, A. M. (1977). Discourse comprehension and sources of individual differences. In M. Just & P. Carpenter (Eds.), *Cognitive processes in comprehension* (pp. 141–183). Hillsdale, NJ: Erlbaum.

Perfetti, C. A., & Roth, S. (1981). Some of the interactive processes in reading and their role in reading skill. In A. Lesgold & C. Perfetti (Eds.), *Interactive processes in reading* (pp. 269–297). Hillsdale, NJ: Erlbaum.

Peterson, R. R., & Simpson, G. B. (1989). Effect of backward priming on word recognition in single-word and sentence contexts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 1020–1032.

Plaut, D. C. (1995). Semantic and associative priming in a distributed attractor network. In Cognitive Science Society (Ed.), *Proceedings of the 17th Annual Conference of the Cognitive Science Society* (pp. 37–42). Hillsdale, NJ: Erlbaum.

Plaut, D. C. (1997). Structure and function in the lexical system: Insights from distributed models of naming and lexical decision. *Language and Cognitive Processes, 12,* 767–808.

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

Posner, M. I., & Snyder, C. R. R. (1975). Facilitation and inhibition in the processing of signals. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and performance V* (pp. 669–682). New York: Academic Press.

Postman, L., & Keppel, G. (1970). *Norms of word associations.* New York: Academic Press.

Quinlan, P. (1991). *Connectionism and psychology: A psychological perspective on new connectionist research.* Chicago: University of Chicago Press.

Raduege, T. A., & Schwantes, F. M. (1987). Effects of rapid word recognition training on sentence context effects in children. *Journal of Reading Behavior, 19,* 395–414.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85,* 59–108.

Ratcliff, R., & McKoon, G. (1988). A retrieval theory of priming in memory. *Psychological Review, 88,* 385–408.

Rueckl, J. G. (1995). Ambiguity and connectionist networks: Still settling into a solution—Comment on Joordens and Besner (1994). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 501–508.

Rumelhart, D. E., Durbin, R., Golden, R., & Chauvin, Y. (1995). Backpropagation: The basic theory. In Y. Chauvin & D. E. Rumelhart (Eds.), *Back-propagation: Theory, architectures, and applications* (pp. 1–34). Hillsdale, NJ: Erlbaum.

Schadler, M., & Thissen, D. M. (1981). The development of automatic word recognition and reading skill. *Memory & Cognition, 9,* 132–141.

Schneider, W. (1990). *MEL user's guide: Computer techniques for real time experimentation.* Pittsburgh, PA: Psychology Software Tools.

Schriefers, H. (1992). Lexical access in the production of noun phrases. *Cognition, 45,* 33–54.

Schwanenflugel, P. J., & Rey, M. (1986). Interlingual semantic facilitation: Evidence for a common representational system in the bilingual lexicon. *Journal of Memory and Language, 25,* 605–618.

Schwantes, F. M. (1981). Locus of the context effect in children's word recognition. *Child Development, 52,* 895–903.

Schwantes, F. M. (1985). Expectancy, integration, and interactional processes: Age differences in the nature of words affected by sentence context. *Journal of Experimental Child Psychology, 39,* 212–229.

Schwantes, F. M. (1991). Children's use of semantic and syntactic information for word recognition and determination of sentence meaningfulness. *Journal of Reading Behavior, 23,* 335–350.

Schwantes, F. M., Boesl, S. L., & Ritz, E. G. (1980). Children's use of context in word recognition: A psycholinguistic guessing game. *Child Development, 51,* 730–736.

Sears, C. R., Hino, Y., & Lupker, S. J. (1995). Neighborhood size and neighborhood frequency effects in word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 876–900.

Sears, C. R., Hino, Y., & Lupker, S. J. (1999). Orthographic neighborhood effects in parallel distributed processing models. *Canadian Journal of Experimental Psychology, 53,* 220–230.

Sears, C. R., Lupker, S. J., & Hino, Y. (1999). Orthographic neighbourhood effects in perceptual identification and semantic categorization tasks: A test of the multiple read-out model. *Perception & Psychophysics, 61,* 1537–1554.

Seidenberg, M. S. (1992). Beyond orthographic depth: Equitable division of labor. In R. Frost & K. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 85–118). Amsterdam: Elsevier.

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

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 Behaviour, 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.

Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition, 55,* 151–218.

Sharkey, A. J., & Sharkey, N. E. (1992). Weak contextual constraints in text and word priming. *Journal of Memory and Language, 31,* 543–572.

Shelton, J. R., & Martin, R. C. (1992). How semantic is automatic semantic priming? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 1191–1210.

Simpson, G. B., & Lorsbach, T. C. (1983). The development of automatic and conscious components of contextual facilitation. *Child Development, 54,* 760–772.

Simpson, G. B., & Lorsbach, T. C. (1987). Automatic and conscious context effects in average and advanced readers. *Journal of Research in Reading, 10,* 102–112.

Smith, L. C., Briand, K., Klein, R. M., & Heyer, K. den (1987). On the generality of Becker's verification model. *Canadian Journal of Experimental Psychology, 41,* 379–386.

Smith, M. C., Besner, D., & Miyoshi, H. (1994). New limits to automaticity: Context modulates semantic processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 104–115.

Smith, M. C., Theodor, L., & Franklin, P. E. (1983). On the relationship between contextual facilitation and depth of processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 697–712.

Smolensky, P. (1986). Information processing in dynamical systems: Foundations of harmony theory. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 1. Foundations* (pp. 194–281). Cambridge, MA: MIT Press.

Stahl, S. A., Hare, V. C., Sinatra, R., & Gregory, J. F. (1991). Defining the role of prior knowledge and vocabulary in reading comprehension. *Journal of Reading Behavior, 23,* 487–508.

Stanovich, K. E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly, 16,* 32–71.

Stanovich, K. E., Nathan, R. G., West, R. F., & Vala-Rossi, M. (1985). Children's word recognition in context: Spreading activation, expectancy, and modularity. *Child Development, 56,* 1418–1428.

Stanovich, K. E., & West, R. F. (1981). The effect of sentence context on ongoing word recognition: Tests of a two-process theory. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 658–672.

Stanovich, K. E., West, R. F., & Feeman, D. (1981). A longitudinal study of sentence context effects in second-grade children: Tests of an interactive-compensatory model. *Journal of Experimental Child Psychology, 32,* 185–199.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologia, 30,* 276–315.

Stolz, J. A., & Besner, D. (1996). The role of set in visual word recognition: Activation and activation blocking as nonautomatic processes. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 1166–1177.

Stolz, J. A., & Besner, D. (1997). Visual word recognition: Effort after meaning but not (necessarily) meaning after effort. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 1314–1322.

Stolz, J. A., & Neely, J. H. (1995). When target degradation does and does not enhance semantic context effects in word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 596–611.

Stone, G. O., Vanhoy, M., & Van Orden, G. C. (1997). Perception is a two-way street: Feedforward and feedback phonology in visual word recognition. *Journal of Memory and Language, 36,* 337–359.

Stone, G. O., & Van Orden, G. C. (1993). Strategic control in word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 744–774.

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.

Tallal, P., Miller, S., & Fitch, R. H. (1993). Neurobiological basis of speech: A case for the preeminence of temporal processing. In P. Tallal, A. M. Galaburda, R. R. Llinas, & E. von Euler (Eds.), *Temporal information processing in the nervous system: Special reference to dyslexia and dysphasia* (pp. 27–47). New York: New York Academy of Sciences.

Tallal, P., & Piercy, M. (1975). Developmental aphasia: The perception of brief vowels and extended stop consonants. *Neuropsychologia, 13,* 69–74.

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

Thompson-Schill, S. L., Kurtz, K. J., & Gabrieli, J. D. E. (1998). Effects of semantic and associative relatedness on automatic priming. *Journal of Memory and Language, 38,* 440–458.

Tweedy, J. R., Lapinski, R. H., & Schvaneveldt, R. W. (1977). Semantic-context effects on word recognition: Influence of varying the proportion of items presented in an appropriate context. *Memory & Cognition, 5,* 84–99.

Usher, M., & McClelland, J. L. (1995). *On the time course of perceptual choice: A model based on principles of neural computation* (Tech. Rep. No. PDP.CNS.95.5). Pittsburgh, PA: Carnegie Mellon University, Department of Psychology.

Usher, M., & McClelland, J. L. (in press). On the time course of perceptual choice: The leaky competing accumulator model. *Psychological Review.*

Van Orden, G. C., & Goldinger, S. D. (1994). Interdependence of form and function in cognitive systems explains perception of printed words. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 1269–1291.

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.

Vernon, P. A. (1987). *Speed of information processing and intelligence.* Norwood, NJ: Ablex.

Verstaen, A., Humphreys, G. W., Olson, A., & d'Ydewalle, G. (1995). Are phonemic effects in backward masking evidence for automatic prelexical phonemic activation in visual word recognition? *Journal of Memory and Language, 34,* 335–356.

Waters, G. S., & Seidenberg, M. S. (1985). Spelling–sound effects in reading: Time course and decision criteria. *Memory & Cognition, 13,* 557–572.

Wechsler, D. (1958). *The measurement and appraisal of adult intelligence.* Baltimore: Williams & Wilkins.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children* (3rd ed.). San Antonio, TX: Harcourt, Brace.

West, R. F., & Stanovich, K. E. (1978). Automatic contextual facilitation in readers of three ages. *Child Development, 49,* 717–727.

West, R. F., Stanovich, K. E., Feeman, D., & Cunningham, A. (1983). The effect of sentence context on word recognition in second- and sixth-grade children. *Reading Research Quarterly, 19,* 6–15.

Wijnen, F. (1992). Incidental word and sound errors in young speakers. *Journal of Memory and Language, 31,* 734–755.

(Appendixes follow)

Appendix A

Nonword Primes, Unrelated Primes, Related Primes, and Target Words Used in Experiments 1–3

Nonword primes	Unrelated primes	Related primes	Target words	Nonword primes	Unrelated primes	Related primes	Target words
KARBS	EIGHT	ABOVE	BELOW	YAMBE	FINAL	ADULT	CHILD
DOFER	CRAWL	AGONY	PAINS	KARPE	SCREW	ALARM	CLOCK
RAIFY	SPLIT	ARGUE	FIGHT	STELT	BASIC	BEING	HUMAN[d]
TREGS	ALONE	BIRTH	DEATH[d]	FWIST	GUARD	BLADE	KNIFE[d]
VIGHT	VENUS	BLANK	EMPTY	SLELS	FAVOR	BLAZE	FIRES
STELI	TRUNK	BORED	TIRED[d]	DILCH	CRACK	BRIDE	GROOM
MEASH	LOWER	BRIEF	SHORT	EROWN	EARLY	BRING	TAKES
CHESA	TOWER	CANOE	BOATS	SHILT	NERVE	CHAIN	LINKS
SLOVE	MOIST	CHUCK	THROW[d]	JAMOR	GIANT	CIGAR	SMOKE
KOUGH	UNITE	CLEAN	DIRTY[d]	DIGRI	BEGIN	CLOSE	OPENS
RIFEY	SHOOT	COACH	TEAMS[d]	REVRI	ARROW[c]	CORAL	REEFS
SUMIC	CROWD	COURT	JUDGE	GLANE	TOPIC	CRANE	LIFTS[d]
PEESH	USUAL	CREEK	RIVER	DORCH	COUNT	CYCLE	BIKES
SHALS	CHEST	DEATH	LIVES	SLIND	STRAW	DITCH[b]	HOLES[d]
KESPO	PLAIN	DONOR	BLOOD[d]	SMONT	BLIND	ENTER	EXITS[d]
SHEDA	BENCH[c]	FAIRY	TALES	OMOSE	ALIKE[c]	FENCE[b]	POSTS
YESRA	CREAM	FLAME	FIRES[d]	OCKSO	TODAY	FLOOD	WATER
ROBAD	PATCH	FRESH	FRUIT[d]	TOLBS	HURRY	FUNNY	LAUGH[d]
WRUPP	CLIMB	GHOUL	GHOST	VANGE	SIGHT	GLOVE	HANDS
LAVUE	WIDTH	GRAIN	WHEAT	THEET	SCORE	GRASP	HOLDS
SKALT	EVENT	GRASS	GREEN	BISER	READY	HEAVY	LIGHT[d]
STEAF	PRIZE	HONEY	SWEET	DRAFU	SHAPE	HOUSE	HOMES
DUTSY	ALLOW	JOINT	KNEES	SALGS[a]	LEVEL	KNOCK	DOORS
TORMS	NEVER	LABOR	WORKS	EKAPS[a]	PARTY	LARGE	SMALL
TRUIF	ROUGH	LEMON	LIMES[d]	FIETH	ANGLE	LOOSE[b]	TIGHT
WEASH	SHINE	MAJOR	MINOR	SYAMP	BEAST	MAPLE	TREES
EOUSH	PITCH	MARCH	APRIL	SNOGS	STEAM	MINTS	CANDY[d]
POUGH	COLOR	MONTH	YEARS[d]	CARCK	CHECK	MOTEL	HOTEL
SLAKE	CAUSE	NORTH	SOUTH	GOWAN	CHEEK	NOVEL	BOOKS
JUDIT	SOLID	PAINT	BRUSH	KLOPS	CABIN	PASTE	GLUES
APULT	FAITH[c]	PAUSE	STOPS	KABES	EXTRA	PHONE	CALLS
RUESH[a]	DENSE[c]	PHONY	FAKES	GLAFS	REPLY	PIANO	PLAYS
BLOGE	HABIT	PILOT	PLANE	SPLAY	STALK	POKER[b]	CARDS[d]
SHOAT	LEAVE	PRINT	WRITE	PRAPE	PEARL	QUACK	DUCKS[d]
NAIRT	SHOCK	QUEEN	KINGS[d]	KULLS	CLEAR	RADIO	MUSIC
FOROL	CLOTH	RAZOR	SHARP	SCRIE[a]	WORSE	REACH	GRABS
NOACE	STIFF[c]	SCENT	SMELL	RENGE	SWIFT	SHAME	GUILT
RAPEL	VOICE	SHARE	GIVES	GWINS	GOING	SHEET	PAPER
THRON[a]	TENSE[c]	SHIFT[b]	GEARS	LANEG	PUPIL	SHIRT	PANTS
OPINT	NOTES	SHORE	BEACH	GANRY	BURST	SHOUT	YELLS
MILTI	DRINK	SKIRT	DRESS	HASLY	AHEAD	SLICE	PIECE[d]
HIFTA	CHARM	SMILE	HAPPY	GATCH[a]	PROUD	SNAKE	BITES
CHIRD	RAISE	SOCKS	SHOES	LUPPE	DREAM	SOUND	NOISE
WETCH	STORE	SPARE	TIRES	MOTTU	AVOID	SPEAK	TALKS
ROWEL	FLOOR	SPEND	MONEY	HESET	RAPID	SPOON	FORKS
LINDS	QUICK	STALL	HORSE	NACLE	ANGER	STARE	LOOKS[d]
VOBAE	MOTOR	STEEL[b]	METAL	TILOP	CURVE	STILL	MOVES
CANFY	STAMP	STONE	ROCKS	RAICH	BOUND	STORM	RAINS
BLACE	STAND	STUFF	THING	STANT	STATE	SUPER	GREAT
NUIST[a]	CRASH	SWEAR	CURSE	SNISP[a]	DRILL	SWEEP	BROOM
TOOFA	RIFLE	TABLE	CHAIR	ODEAS	ADMIT	TEACH	LEARN[d]
LEJLY	GUEST	THIEF	STEAL	FIRCH	METER	TIGER	LIONS
RUSOI	FROST	TOAST	BREAD	GAMIK	GLORY	TOOTH	DECAY
DREAB	VISIT	TOUCH	FEELS[d]	GELEA	NURSE	TRAIL	PATHS
ECHAT	SCALE	TRAIN	TRACK[d]	ROCAL	EQUAL	TRICK	TREAT
TAFAL	TOTAL	TRUCE	PEACE[d]	SNOLE	MODEL	TWIST	TURNS
VIGES[a]	APART[c]	UNCLE	AUNTS	JIETZ	SWAMP	WAGON	WHEEL[d]
CHUTH	PLATE	WAVES[b]	OCEAN	SMOUL	CHIEF	WHITE	BLACK[d]
DUKOL	CHINA	WINGS	BIRDS[d]	QUARF	CLOUD	WRIST	WATCH
LERRI	FOUND	WRONG	RIGHT	KLIGS	FRONT	YOUTH	YOUNG[d]

Note. All items that were outliers or yielded no reliable priming effect in Experiment 1 with college students were eliminated from the word list for Experiment 2 with the elementary students.
[a] Outlier in nonword condition. [b] Outlier in related condition. [c] Outlier in unrelated condition. [d] No reliable priming effect.

Appendix B

Mean Reaction Times (RTs; in Milliseconds) and Error Rates (ERs) for Participants From Experiments 1–3 and for the Network

	High-frequency targets						Low-frequency targets					
	Related primes		Unrelated primes		Nonword primes		Related primes		Unrelated primes		Nonword primes	
Condition	RT	ER	RT	ER	RT	ER	RT	ER	RT	ER	RT	ER
Adults												
Short SOA												
High perceptual ability												
Participants	601	1.0	618	2.1	643	2.8	613	3.5	665	2.3	692	4.5
Participants[a]	604	1.3	624	1.8	638	3.0	603	2.6	665	1.8	692	1.9
Network	627	0.0	647	0.0	654	0.0	669	0.0	697	0.0	703	0.0
Low perceptual ability												
Participants	678	2.0	705	3.3	742	3.8	713	2.1	735	2.6	744	3.9
Participants[a]	672	2.4	720	2.5	733	4.3	713	1.5	732	2.6	754	3.6
Network	641	0.0	672	0.0	667	0.0	717	0.1	742	0.0	715	0.4
Long SOA												
High perceptual ability												
Participants	686	1.0	693	1.0	692	1.3	694	1.0	727	2.0	720	1.2
Participants[a]	675	0.9	696	1.4	699	1.7	679	1.6	733	1.5	714	0.6
Network	695	0.0	707	0.0	704	0.0	722	0.1	746	0.0	743	0.0
Low perceptual ability												
Participants	760	1.3	780	1.7	756	2.4	789	1.0	806	3.0	796	2.5
Participants[a]	750	1.5	790	1.6	753	2.2	758	1.4	802	2.7	780	1.8
Network	713	0.0	755	0.3	726	0.0	793	1.0	815	1.7	782	0.6
Children												
Short SOA												
High perceptual ability[b] network	848	0.0	861	0.0	865	0.0	900	0.9	918	1.9	923	2.0
Low perceptual ability[b] network	883	0.2	917	0.5	911	0.4	966	6.4	990	7.5	969	5.4
Long SOA												
High perceptual ability[b]												
Participants	837	0.6	855	2.5	868	2.1	861	1.3	913	5.4	920	4.2
Network	882	0.0	890	0.0	890	0.0	934	0.3	963	1.3	952	1.3
Low perceptual ability[b]												
Participants	956	1.8	1,004	2.9	992	2.0	1,014	4.3	1,059	5.4	1,064	4.1
Network	936	0.5	980	1.5	954	0.8	1,031	7.1	1,074	10.2	1,023	7.3
Third graders[c]												
Participants	949	1.4	996	3.8	1,013	2.4	1,004	3.5	1,079	8.4	1,150	5.7
Network	926	0.4	955	1.1	936	0.4	1,002	6.4	1,035	10.0	1,003	7.6
Sixth graders[c]												
Participants	851	1.0	888	1.7	858	1.8	886	2.2	932	2.6	921	2.7
Network	894	0.1	919	0.4	910	0.4	960	1.0	998	1.5	971	0.9

Note. SOA = stimulus-onset asynchrony.
[a] Data are from only the 72 items also used with children in Experiment 2. [b] Data are collapsed across 3rd- and 6th-grade children. [c] Data are collapsed across children with high and low perceptual ability.

Received December 26, 1997
Revision received January 18, 2000
Accepted January 19, 2000 ■