More than meets the eye: Context effects in word identification

MICHAEL E. J. MASSON
University of Victoria, Victoria, British Columbia, Canada

and

RON BOROWSKY
University of Saskatchewan, Saskatoon, Saskatchewan, Canada

The influence of semantic context on word identification was examined using masked target displays. Related prime words enhanced a signal detection measure of sensitivity in making lexical decisions and in determining whether a probe word matched the target word. When line drawings were used as primes, a similar benefit was obtained with the probe task. Although these results suggest that contextual information affects perceptual encoding, this conclusion is questioned on the grounds that sensitivity in these tasks may be determined by independent contributions of perceptual and contextual information. The plausibility of this view is supported by a simulation of the experiments using a connectionist model in which perceptual and semantic information make independent contributions to word identification. The model also predicts results with two other analytic methods that have been used to argue for priming effects on perceptual encoding.

The identification of objects or events in the environment, including written and spoken language, benefits from the availability of conceptually relevant information provided by the context in which an object or event occurs. For example, the identification of visually presented individual letters is enhanced when they are placed in the context of a familiar word (see, e.g., McClelland & Rumelhart, 1981; Reicher, 1969; Wheeler, 1970), visually presented words are identified more quickly (see, e.g., Borowsky & Besner, 1991, 1993; Meyer, Schvaneveldt, & Ruddy, 1975; Neely, 1977) and accurately (see, e.g., Paul, Kellas, & Juola, 1992; Rhodes, Parkin, & Tremewan, 1993; Tulving & Gold, 1963) when presented in a relevant context, speech perception is improved by the availability of relevant linguistic context (Miller & Isard, 1963) or by visible aspects of speech (Erber, 1969; Sumby & Pollack, 1954), and identification of objects and faces is enhanced by relevant contextual information (Reinitz, Wright, & Loftus, 1989; Rhodes & Tremewan, 1993).

A fundamental issue that any account of contextual influences on perception must address is whether stimulus information produced by data-driven processes interacts with conceptually driven information, such as semantic context. *Interaction* here means that information about the context in which a stimulus occurs influences the formation of a perceptual representation of the stimulus, a process we refer to as "perceptual encoding." The working definition of *perceptual encoding* that we adopt here is the acquisition of perceptual stimulus information. In the case of word identification, for example, perceptual encoding creates a representation of a letter string's perceptual features. This characterization of perceptual encoding is compatible with a variety of word identification models, such as the accumulation of perceptual evidence in favor of a logogen that represents a particular word (Morton, 1969), the activation of feature- and letter-level units in the interactive activation model of word identification (McClelland, 1991; McClelland & Rumelhart, 1981), the formation of a pattern of activation among orthographic units in distributed memory models (see, e.g., Masson, 1995; Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989), or the acquisition of stimulus feature information in the fuzzy logical model of perception (FLMP; see, e.g., Massaro & Cohen, 1991).

The proposition that stimulus and contextual information interact, or are nonindependent, is inherent in connectionist models that propose an interactive activation architecture in which processing units that evaluate phonological features (in speech perception) or letters (in word identification) are influenced by higher level units

that represent words (e.g., McClelland, 1991; McClelland & Elman, 1986; McClelland & Rumelhart, 1981). In these models, the evaluation of stimulus information is influenced by contextual knowledge. The general view that context and stimulus information interact has also appeared in models that are not in the connectionist class. For example, Reinitz et al. (1989) proposed a multiplicative model in which context serves to increase by a constant factor the rate at which visual information is acquired from a stimulus.

In contrast to models that assume an interaction between contextual and stimulus information, independence models are founded on the proposition that stimulus information is acquired and evaluated independently of contextual information. A well-developed model of this class is the FLMP (Massaro, 1979, 1989; Massaro & Cohen, 1991; Oden & Massaro, 1978). Although in independence models it is assumed that stimulus and contextual sources of information are evaluated independently, it is also proposed that after evaluation, the two sources are integrated to determine a response. Thus, both sources of information contribute to performance, but they do so without affecting each other. The notion of modularity developed by Fodor (1983; see also Pylyshyn, 1980) is compatible with the fundamental assumption of independence models regarding the separate evaluation of stimulus and contextual information. In Fodor's account, input systems consist of independent modules that are specialized for perceptual encoding of a particular class of input (e.g., language, color, shape, or faces). These input systems, or modules, are assumed to be "informationally encapsulated" so that perceptual encoding is not affected by processing within other input modules or by semantic processing.

The many demonstrations of contextual influence on pattern recognition clearly indicate that, at some level, stimulus and contextual information must be combined. The question is, at what level? In the case, for example, of priming effects in word identification, the benefit of presenting a related context word prior to speeded identification of a target word might be taken as evidence in support of the view that contextual information increased the efficiency of perceptual processing. Alternatively, however, the benefit can be ascribed either to decision processes occurring after perceptual processing of the target word has taken place (see, e.g., Balota & Chumbley, 1984; Seidenberg, Waters, Sanders, & Langer, 1984; Stanovich & West, 1979) or to priming effects taking place within a lexical module (see, e.g., Forster, 1979, 1981; Seidenberg et al., 1984). More specifically, on the assumption that words are represented in a lexical module, mutual priming among related lexical entries within that module could produce context effects. Thus, context effects arising from these sources would not threaten independence models.

A stronger challenge to independence models is the interaction between context and target stimulus quality on response latency in word identification tasks. Response latency on lexical decision and naming tasks is reduced when a target word follows a semantically related context word than when it follows an unrelated context word, and this effect is amplified when the target word is visually degraded (see, e.g., Becker & Killion, 1977; Besner & Smith, 1992; Borowsky & Besner, 1991, 1993; Meyer et al., 1975). According to Sternberg's (1969) additive factors logic, if we assume that stimulus quality has a selective effect on the duration of processing at the earliest stage of perception, this interaction indicates that both context and stimulus quality affect the same stage of early perceptual processing, thereby compromising the view that stimulus information and contextual information are independently processed.

Although the context × stimulus quality interaction suggests that stimulus information and contextual information both affect a common stage of processing, this common stage need not be perceptual. An interactive effect on response latency does not preclude later stages of processing (i.e., any stage up to and including response execution) from being the common locus for the interactive effects of context and stimulus quality. For example, in the FLMP, stimulus information and contextual information are acquired independently, then integrated. The interaction between these two factors might occur during a multiplicative integration of the two sources of information, with contextual information having a larger impact on determining a response when stimulus information is impoverished (see, e.g., Massaro, Weldon, & Kitzis, 1991).

One means of overcoming this limitation on response latency measures is to establish empirically constraints regarding where the interactive effects probably occur. For example, Borowsky and Besner (1993) have shown that, within a single experiment, the semantic context effect has both a relatively early and a relatively late locus of influence in the lexical decision task. They demonstrated that semantic context interacts with both stimulus quality and word frequency, whereas the effects of stimulus quality and word frequency are additive with respect to response latency. By Sternberg's (1969) additive factors logic, given that stimulus quality must influence reading processes prior to the effect of word frequency, and given that these two variables must have separate loci of influence to accommodate their additivity, it follows that semantic context must have both a relatively early and a relatively late influence to accommodate the interactions of semantic context with stimulus quality and word frequency. This pattern of interaction and additivity led Borowsky and Besner (1993) to conclude that semantic context has its effect at the level of orthographic processing (where it interacts with stimulus quality) and at the level of semantic processing (where it interacts with word frequency via frequency-sensitive connections between orthographic and semantic stages of processing). Thus, in Borowsky and Besner's (1993) account of the context × stimulus quality interaction, semantic context affects orthographic processing by selectively lowering the activation threshold or raising the

baseline activation level of primed items in the system that represents orthography. On the other hand, stimulus quality affects the rate at which orthographic information is encoded, so that degraded visual information results in a lower rate of orthographic or "perceptual" encoding. It should be noted that although Borowsky and Besner (1993) chose this particular account of the context × stimulus quality interaction to create a parsimonious account of the joint effects of context and stimulus quality, additive factors logic does not permit one to discriminate whether one or both of these factors affect the rate of perceptual encoding, the thresholds for (or baseline levels of) activation, or some combination of the two.

SIGNAL DETECTION SENSITIVITY AS A MEASURE OF PERCEPTUAL ENCODING

An alternative strategy for testing between nonindependence and independence models is to consider a measure that is assumed to provide a direct assessment of stimulus information acquisition. Several researchers have considered signal detection theory (Tanner & Swets, 1954), which provides a means of separately measuring an observer's sensitivity to stimuli (ability to discriminate between targets and nontargets) and response bias (preference for responding "target" or "nontarget") as an appealing candidate for this role. For example, Farah (1989) argued that sensitivity effects reflect changes in perceptual processing, whereas changes in response bias are due to attentional and contextual processes. By this reasoning, if a factor such as context influences perceptual processing, an effect on sensitivity as measured by signal detection theory should be found.

A number of attempts to find an effect of contextual information on sensitivity have failed, suggesting that context effects are confined to influences on response bias (see, e.g., Antos, 1979; Hale & Johnston, 1983; Johnston & Hale, 1984; Masson, 1988). More recently, however, Rhodes and her colleagues (Rhodes et al., 1993; Rhodes & Tremewan, 1993; see also Paul et al., 1992) have obtained a reliable effect of semantic context on sensitivity in word and face identification tasks. For example, Rhodes et al. presented masked word and nonword targets that were preceded by a context word that was semantically related or unrelated to the target. The task was to determine whether or not the target was a word. They found that sensitivity in this masked lexical decision task was greater when semantically related context words were used as primes. Rhodes et al. took this result as evidence that context affects perceptual encoding, which would be a clear violation of independence and the modularity hypothesis.

Doubts About Sensitivity

Two questions can be raised with respect to the Rhodes et al. (1993) conclusion. First, although they obtained an effect of semantic priming on signal detection measures of sensitivity, both the primes and the targets were printed words. This configuration of primes and targets leaves open the possibility that priming effects may have occurred within a single processing module, such as a lexical module (see, e.g., Forster, 1979), in which links between related entries are responsible for producing the priming effect. In Fodor's (1983) terms, the priming effect obtained by Rhodes et al. might not be exogenous (i.e., might not arise outside the lexical module). If the source of a priming effect is not exogenous to the lexical module, the modularity assumption is not violated. This shortcoming was addressed in part by the Rhodes and Tremewan (1993) finding of an effect of priming on sensitivity measures of face identification when primes were printed names. Although this face identification result suggests that priming violated the assumption of modularity by operating across modules, perhaps via a conceptual level of representation, it does not resolve the question of whether exogenous effects can be found with word identification. Moreover, even the face priming result is subject to the other question arising from the word priming effect reported by Rhodes et al.

The second, more fundamental question, is whether an observed effect of context on signal detection measures of sensitivity can properly be taken as indicating an effect on perceptual processing. The problem is that in a task such as lexical decision, performance as measured by sensitivity may be determined both by perceptual and by conceptual processes that operate independently. For example, when a subject is presented with a well-known word, a positive lexical decision response can be supported both by the familiarity of the word's orthographic pattern and by the semantic information recruited by that pattern. If this is the case, a factor that affects sensitivity may operate through either or both of these processes. Thus, unless we can be certain that task performance is determined exclusively by perceptual processes, sensitivity effects cannot unambiguously be attributed to changes in perceptual processing.

Rhodes and her colleagues (Rhodes et al., 1993; Rhodes & Tremewan, 1993) were aware of this problem and marshaled arguments in support of the view that their sensitivity effects did indicate differences in perceptual processing. They offered as converging evidence the finding (reviewed above) of an interaction between context and target stimulus quality in latency measures of word identification. As discussed above, however, response latency measures leave open the possibility that stimulus and contextual information have a combined influence on a later stage of processing than perceptual encoding.

A further point made by Rhodes et al. (1993) to support the view that their sensitivity effect reflected a change in perceptual processing was that their lexical decision task was perceptual in nature because targets were masked and because the nonwords were orthographically similar to the target words. That is, in the related context condition, a nonword target (e.g., *voad*) was similar to a word (*road*) that was semantically related to the context word (*street*).

It is plausible, however, that even when targets are masked, sensitivity effects could arise from the inde-

pendent contributions of contextual information and stimulus information that combine, without influencing each other, to determine a subject's response. Massaro and Cohen (1991) demonstrated that the FLMP predicts improved forced-choice accuracy in letter identification when target letters appear in a word context and are followed by a mask. Moreover, Norris used a modular criterion bias model (see, e.g., Morton, 1969; Norris, 1986) to simulate the Rhodes et al. (1993, Experiment 1) effect of context on sensitivity and bias in lexical decisions.

In the Norris (1995) model, each word is represented as a logogen, and the effect of a related context word is to reduce the recognition threshold for logogens related to the context word. A positive lexical decision is made if at least one logogen is activated beyond its threshold value. In the simulations reported by Norris (1995), greater sensitivity was obtained when the target was related to the context word because reduced threshold due to priming acted in concert with stimulus information to increase the likelihood that activation of the target's logogen (or the logogen of some other word related to the context word) would exceed threshold. These simulation results indicate that the criterion bias model is capable of accounting for the sensitivity effect reported by Rhodes et al. (1993) without conceding that perceptual processes were affected by contextual information. To the extent that the criterion bias model is a plausible account of the lexical decision task, its successful simulation of the Rhodes et al. results is an existence proof that an effect of context on sensitivity is not sufficient evidence that context affects perception.

Appropriateness of Signal Detection Measures

In addition to simulating context effects on sensitivity with the criterion bias model, Norris (1995) made a general claim regarding the application of signal detection measures to the lexical decision task. He argued that changes in sensitivity reflect changes in perceptual encoding only if the assumptions of signal detection theory are met. In particular, a change in sensitivity brought about by a variable such as manipulation of context implies that signal and noise distributions have been moved further apart in a system in which decisions are made in just the way assumed by signal detection theory. Namely, a single criterion is placed on a single evaluative dimension.

Norris (1995) argued, however, that in models of word identification, decision processes are more complex than assumed by signal detection theory. For example, in the criterion bias model, individual word logogens have separately modifiable response criteria. When a context word is presented, the criteria of logogens representing words related to the context word will be reduced. In his simulation of lexical decision under degraded viewing conditions, Norris showed that this selective lowering of logogen criteria could produce an increase in sensitivity as measured by signal detection theory without any change in perceptual encoding. Standard signal detection theory has no means of capturing the selective reduction in response criteria and therefore misrepresents a selective

criterion shift as a change in sensitivity. Because sensitivity effects as measured using signal detection theory can be misleading when applied to word identification tasks, Norris argued against such application.

Although we agree that sensitivity measures alone cannot be taken as pure indices of perceptual encoding, we maintain that these measures are still useful in assessing performance accuracy, however that accuracy arises. There are at least two ways in which these measures of accuracy are useful in the examination of context effects on word identification. First, an effect of context on identification accuracy or sensitivity is important evidence in support of nonindependence models: If context affects perceptual encoding, then an effect of context should be seen when accuracy is measured. The demonstration by Norris (1995) shows that the converse is not necessarily true, indicating that an effect of context on accuracy is not sufficient evidence for the conclusion that context affects perceptual encoding.

Second, the finding that context affects word identification accuracy poses a challenge to independence models of word identification in general. The success of the criterion bias model developed by Norris (1995) in simulating the sensitivity effects reported by Rhodes et al. (1993) gives that class of model an advantage over competing independence models that have not been shown to produce such effects. It is important, therefore, to determine whether other models, particularly independence models, readily generate context effects on sensitivity.

A Distributed Memory Model of Context Effects

A view of context effects that is somewhat different from the criterion bias model is embodied in connectionist models that use distributed representations of word knowledge (see, e.g., Masson, 1991b, 1995; Plaut, 1995; Plaut & Shallice, 1993). In these models, word knowledge is represented as patterns of activation across an entire collection of processing units rather than as a set of logogens each of which represents a single word. In these connectionist models, presentation of a priming stimulus moves the processing units of a network into a pattern of activation that corresponds to the prime. The pattern of activation induced by the prime across the semantic units in the network is assumed to be similar to the patterns associated with words semantically related to the prime. When the target is presented, the semantic units are already in a pattern of activation compatible with the target and will help the network move more quickly into a pattern that fully instantiates knowledge about the target word.

In the distributed memory model we have developed (Borowsky & Masson, 1996; Masson, 1991b, 1995; Masson & Borowsky, 1995), separate sets of processing units are responsible for representing knowledge about a word's orthographic, phonological, and semantic features. Orthographic units are affected only by sensory input, but phonological and semantic units are affected by all units in the network. These architectural constraints are shown in Figure 1.[1] Thus, stimulus information and contextual

information make independent contributions to task performance. That is, the semantic units have no influence on the development of a pattern of activation across the units that represent the orthographic pattern of a stimulus. Only sensory input determines the orthographic pattern.

The question addressed in the work reported here was whether this distributed memory model would be capable of simulating a context effect on word identification accuracy such as that reported by Rhodes et al. (1993). Besides applying the model to results from the Rhodes et al. study, we extended our test in three ways. First, we collected additional data in a masked lexical decision task designed to produce more robust context effects than those obtained by Rhodes et al. Given that an independence model such as the distributed memory model might have difficulty simulating sensitivity effects, empirical demonstrations of large effects of this kind should pose a substantial challenge to the model.

Second, we empirically tested the hypothesis that sensitivity effects would be found in a word identification task that requires subjects to make a judgment about the identity of a masked target. This task requirement is not a necessary component of the masked lexical decision task, in which subjects do not have to decide exactly which word was shown, but only whether the stimulus was a word. In the identification task, a probe word followed the masked target and subjects decided whether the probe matched the target. It was not clear whether a context effect on sensitivity would be found in this task. Previous context effects on sensitivity measures have been restricted to judgments, such as lexical decision or fame (Rhodes et al., 1993; Rhodes & Tremewan, 1993), that do not require specification of the target's identity, but instead can be based on a sense of familiarity. In a further variant of this task, we used line drawings of objects, rather than words, as primes to produce an exogenous basis for semantic context.

Finally, we used results from the model's simulation of the probe task to examine the relation between accuracy and target exposure duration. Two analytic techniques using this relation have been proposed as methods of testing for an effect of priming on perceptual encoding (Reinitz & Alexander, 1996; Reinitz et al., 1989). Were the distributed memory model—based as it is on independent contributions of stimulus information and contextual information to task performance—able to fit the functions that these techniques identify as characteristic of priming effects on perceptual encoding, the interpretation of those techniques would be called into question.

EXPERIMENT 1

In Experiment 1, we attempted to replicate the effect of semantic context on sensitivity in the masked lexical decision task that was obtained by Rhodes et al. (1993). Unlike the nonword items used by Rhodes et al., the nonwords in Experiment 1 did not resemble any of the words used in the experiment. In contrast, Rhodes et al. used a nonword that differed by only one letter from a word related to the prime (e.g., *street*–VOAD) in the related-prime/nonword-target condition. We expected that by using nonwords that did not resemble related words, the false alarm probability would be the same in the related and the unrelated context conditions. Consequently, any increase in hit probability due to related context would not be accompanied by a corresponding increase in false alarms. This situation should lead to a more robust context effect than that obtained by Rhodes et al. A large context effect on sensitivity should provide a strong challenge to models that assume independence between stimulus information and contextual information. A second difference between our method and that used by Rhodes et al. was that we added a pre-mask to the post-masked displays to prevent accuracy from reaching ceiling.

Method

Subjects. Sixteen University of Victoria students participated in the experiment for extra credit in an introductory psychology course. All subjects had normal or corrected-to-normal vision and considered English to be their first language. Four additional subjects were tested, but their performance was not significantly greater than chance, so those data were not included in the analyses.

Materials and Design. The critical stimuli consisted of 120 word triplets and 120 nonwords. One word in each triplet served as the prime and the other two words were semantic associates of the prime (e.g., *king*; ROYAL, CROWN). Two associates were selected because we planned to use these materials in Experiment 2 and the design of that experiment required pairs of associates. The associates in each triplet were equated for length (either four or five letters). The nonwords were four- and five-letter pronounceable letter strings and were derived from words not used as critical items. In addition to the critical stimuli, we used five word pairs and five word–nonword pairs as practice stimuli.

The design of Experiment 1 was a factorial combination of context (related and unrelated) and target (word and nonword). The 120 word triplets were broken into four blocks of 30 triplets. For each subject, one block was assigned to each of the four conditions in the design. The assignment of blocks to conditions was counterbalanced across subjects so that each block served equally often
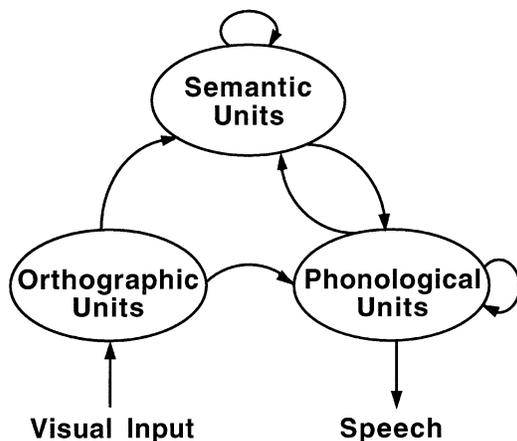

Figure 1. The architecture of the distributed memory model. Processing units are organized into three modules. Arrows indicate which modules affect activation in other modules and in which modules processing units affect one another.

in each condition. One of the two associates in each triplet was used as the target for half of the subjects and the other associate in each triplet was used for the other half of the subjects. For triplets assigned to the unrelated context condition, associates were reassigned to unrelated prime words. On trials that presented a nonword target, a nonword was randomly selected without replacement from the pool of 120 nonwords with the constraint that it be the same length as the associates that belonged to the triplet represented by the prime on that trial.

Procedure. An IBM-compatible computer with a color monitor was used to control stimulus displays, timing of events, and response recording. Subjects were tested individually in a quiet room. The stimuli appeared on the monitor and subjects responded by pressing one of two keys ("z" or "/" on the computer keyboard) to indicate their lexical decision response. The key under the subject's dominant hand was used to signal "word" and the key under the nondominant hand was used to signal "nonword." Subjects were instructed that on each trial they would see a word (the prime) followed by a letter string (the target) inserted between successive displays of a mask that consisted of a row of 8s that was the same length as the target. The task was to decide whether the letter string spelled a word. Subjects were informed that in some cases, when the letter string was a word it would be related to the word presented at the beginning of the trial. They were told that they should pay attention to that word because it might help them in making their decision.

Each critical trial began with a fixation cross in the center of the monitor; subjects initiated each trial by pressing the space bar on the computer keyboard. The fixation cross was erased and 250 msec later the prime word appeared in lowercase letters for 200 msec. The prime was erased and the monitor was blank for 600 msec. The pre-mask then appeared for 50 msec and was replaced by a 33-msec presentation of the target letter string in uppercase letters. Finally, the post-mask appeared for 50 msec, and then the monitor was blank while the subject responded.

A series of 10 practice trials was presented first, using the same procedure as that for the critical trials, except that the target dura-

tion was 1,000 msec on the first trial, 100 msec on the second trial, 50 msec on the third trial, and then 33 msec on the seven remaining trials. This procedure was adopted for the practice trials to help the subject understand the sequence of events before receiving critical trials with the very brief target duration. The practice trials were followed by a randomly ordered presentation of the 120 critical trials. Each subject received an independently determined random ordering of trials. The entire procedure lasted about 20 min.

Results

A hit was defined as a "word" response made when the target letter string was a word. A false alarm occurred when the "word" response was given on a trial involving a nonword target. The mean proportions of hits and false alarms for the related and unrelated conditions are shown in Figure 2. A signal detection analysis was used to compute separate measures of sensitivity and bias. The sensitivity measure A' was used because it does not rely on the strict assumptions of normality and equal variance of signal and noise distributions that are associated with d' (see, e.g., Grier, 1971). The A' measure has a maximum value of 1.0; a value of .5 indicates chance performance. Response bias was measured using B_D'', which ranges from -1 to 1, with negative values indicating a liberal response bias and positive values indicating a conservative bias (Donaldson, 1992). We chose to use B_D'' rather than B'' (which was used by Rhodes et al., 1993) because B'' has been shown not to be independent of A' (McNichol, 1972; Snodgrass & Corwin, 1988), whereas B_D'' is independent of A' (Donaldson, 1992). The mean A' and B_D'' values are shown in Figure 2.

Separate analyses of variance (ANOVAs) with context (related and unrelated) as the factor were applied to each
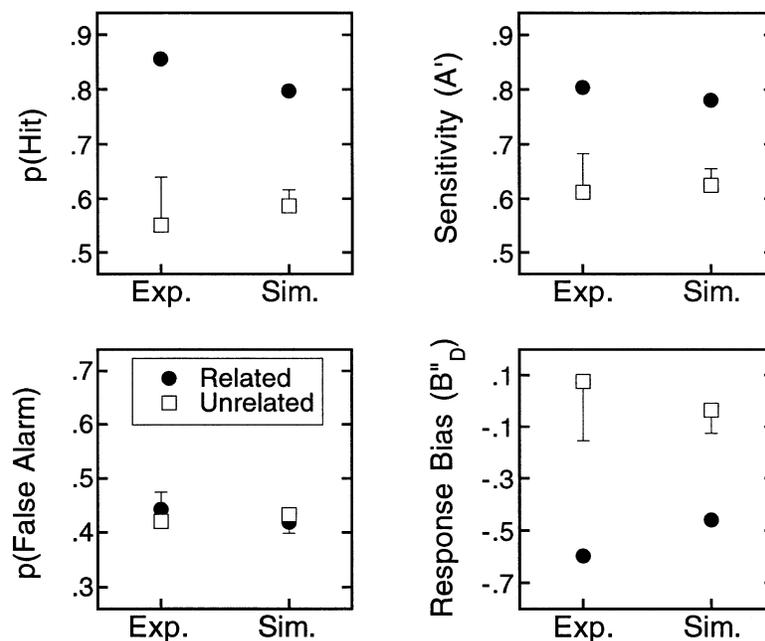

Figure 2. Mean hit and false alarm probabilities, sensitivity, and bias as a function of context for experimental (Exp.) and simulated (Sim.) results in Experiment 1. Error bars represent the 95% confidence interval for the difference between means.

of the four measures shown in Figure 2. In all analyses reported in this article, a Type I error rate of .05 was used. The mean proportion of hits was reliably greater in the related condition than in the unrelated condition [$F(1,15) = 54.84$, $MS_e = 0.014$, $\eta^2 = 0.78$], but there was no effect on false alarm proportion ($F < 1$). The primary measure of interest was sensitivity (A'). As suggested by the effect of context on hits but not on false alarms, there was a significant effect of context on sensitivity [$F(1,15) = 32.02$, $MS_e = 0.009$, $\eta^2 = 0.68$]. This effect shows that subjects were more sensitive in discriminating words from nonwords if the words were preceded by a related context. An analysis of response bias, as measured by B_D'', showed that there was a stronger bias in the related context condition [$F(1,15) = 38.93$, $MS_e = 0.093$, $\eta^2 = 0.72$], indicating that subjects were more likely to respond "word" in that condition.

Simulation results. The results of Experiment 1 and the results from the 250-msec stimulus onset asynchrony (SOA) condition in Experiment 1 of Rhodes et al. (1993) were simulated using the distributed memory model depicted in Figure 1 (Borowsky & Masson, 1996; Masson, 1991b, 1995). Details of the model are provided in the Appendix, but the basic principles of the model are as follows. A word is defined as a pattern of activation across the entire collection of units, rather than an independent unit or logogen, as in the Norris (1995) model. The model's knowledge about an entire vocabulary is represented in the connection weights that link the processing units. Word identification is simulated by first presenting the model with the orthographic pattern of a word. This presentation is done by randomly sampling individual orthographic units and setting a sampled unit to the state dictated by the target letter string's orthographic pattern. Each sampling of a unit is referred to as an "updating cycle." As the orthographic units are being updated, word identification goes forward by also randomly sampling and updating the activation states of phonological and semantic units. The state of a phonological or semantic unit is determined by activation sent to that unit from all other units in the network.

When presented with the orthographic pattern of a known word, the network, over the course of updating the processing units, usually will settle into a pattern of activation that corresponds to that word, achieving identification of the word. That is, the network is capable of instantiating a word's phonological pattern and meaning, given its orthographic representation. Semantic context effects emerge because semantically related words have similar patterns of activation in the semantic units (Masson, 1991b, 1995). When presented with a prime word, the network begins to settle into the pattern of activation associated with that item. When the prime is replaced by a target, the network picks up where it left off processing the prime. The orthographic units change their pattern because of changed visual input, and the phonological and semantic units follow suit. In the case of a semantically related prime and target, the semantic units will be in a favorable pattern of activation when the tar-

get takes over from the prime because the two items have similar patterns in those units. Therefore, it will take fewer updating cycles for the units to settle into the appropriate pattern (see Masson, 1991b, 1995).

To simulate the masked lexical decision task, the network was allowed to instantiate only partially a target's pattern before making a response. Unlike earlier simulations using this model, then, our interest was in the accuracy of the model's performance, rather than its latency. Simulation of lexical decisions made under time or informational constraints is based on the assumption that subjects respond to a global sense of familiarity induced by processing a letter string (Balota & Chumbley, 1984; Besner, 1983; Besner & Johnston, 1989). We assume that familiarity corresponds to a measure of how well the current pattern of activation across the processing units fits the constraints of the learned connection weights (Borowsky & Masson, 1996; Masson & Borowsky, 1995). As the network settles into a pattern of activation that represents a learned word, a better and better fit to the constraints of the connection weights is achieved. The fit to the connection weights can be computed by a measure known as *energy*, as shown in the Appendix. When presented an orthographic pattern that does not correspond to a known word (i.e., a nonword), the goodness of fit between the pattern of activation in the network and the connection weights is not as good as it is for words, so the energy value is not as large as it is for words. Thus, the energy value is diagnostic with respect to the lexical status of the target item.

Simulation of the masked lexical decision results of Experiment 1 involved 50 runs, each corresponding to an independent "subject." On each trial, the model was presented a prime word, a mask, and finally the target letter string. After the target had been processed for a brief time, energy was computed across the entire set of units in the network. If the computed energy surpassed a specified criterion, the letter string was classified as a word; otherwise it was classified as a nonword.

The mean proportions of hits and false alarms produced by the simulation in each of the context conditions (related and unrelated) are shown in Figure 2. Sensitivity (A') and bias (B_D'') scores were computed as well, and the means of these values are also shown in Figure 2. As in Experiment 1, separate ANOVAs were computed for hit and false alarm probabilities, sensitivity, and bias. The probability of a hit was reliably greater in the related condition [$F(1,49) = 199.96$, $MS_e = 0.006$, $\eta^2 = 0.80$], although there was no effect of context on false alarm probability ($F < 1$). There was a reliable effect of context on sensitivity [$F(1,49) = 107.18$, $MS_e = 0.006$, $\eta^2 = 0.69$], indicating that the model was more accurate in discriminating words from nonwords when the target was preceded by a related context. The mean bias value was reliably lower in the related context condition [$F(1,49) = 90.05$, $MS_e = 0.050$, $\eta^2 = 0.65$], reflecting a greater likelihood of a positive response in that condition.

The results of the simulation provide a good fit to the empirical data of the lexical decision task used in Ex-

periment 1. Moreover, the distributed memory model was also used successfully to simulate the lexical decision results reported by Rhodes et al. (1993, Experiment 1), in which nonwords resembled target words. The similarity between nonword and word targets that characterized the materials used by Rhodes et al. was simulated by using orthographic patterns for nonwords that were highly similar to those for words. The results of that simulation and the data from the Rhodes et al. study are shown in Figure 3. As in the empirical data reported by Rhodes et al., the simulation produced a greater false alarm probability in the related than in the unrelated context condition. The simulated effect of context on sensitivity was smaller than in the simulation of our Experiment 1 results, but still reliable [$F(1,49) = 17.99$, $MS_e = 0.002$, $\eta^2 = 0.27$].

The effect of context on lexical decision sensitivity in each of these lexical decision experiments was simulated by the distributed memory model even though that model does not make any provision for perceptual processes to be affected by higher level information. Instead, the effect of a related prime is to provide an additional source of relevant information to increase the sense of familiarity (energy) generated by the target item. The distributed memory model simulates effects of semantic context on sensitivity in the lexical decision task because it includes semantic units in its computation of energy. Like the criterion bias model used by Norris (1995) to challenge Rhodes et al.'s (1993) conclusions

regarding the effect of context on sensitivity, the distributed memory model also is capable of producing sensitivity effects without assuming a top-down influence of context on perceptual encoding.

Discussion

In finding a significant effect of context on sensitivity in the masked lexical decision task, we replicated the results of Rhodes et al. (1993) and Paul et al. (1992). Recall that an effect of context on sensitivity is taken by many to imply an effect of context on stimulus information acquisition, and therefore the effect is indicative of a violation of the modularity hypothesis and, more generally, the assumption that stimulus information and contextual information are processed independently. Rather than considering such results as evidence against the independence assumption, we agree with Norris (1995) that it could be argued that more than just stimulus information affects sensitivity in the masked lexical decision task. Rather, sensitivity can be influenced by joint contributions from stimulus information and contextual information. A similar point was made by Massaro (1987; Massaro & Cohen, 1991; Massaro & Oden, 1995) in his evaluation of context effects on speech and letter perception.

The important contribution of Experiment 1, however, was twofold. First, by using nonwords that bore no resemblance to word targets, we were able to produce a much larger effect of context on sensitivity than that obtained by Rhodes et al. (1993). Second, we were able to simulate
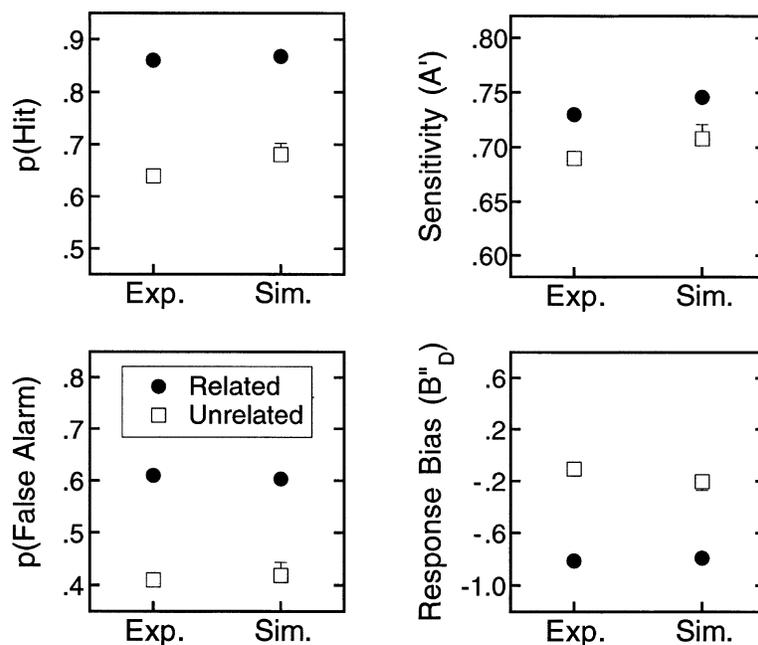
Figure 3. Mean hit and false alarm probabilities, sensitivity, and bias as a function of context for experimental (Exp.) and simulated (Sim.) results from Rhodes et al. (1993, Experiment 1, 250-msec SOA condition). Error bars represent the 95% confidence interval for the difference between means. These error bars are not shown for the experimental results because the information required to generate them was not provided in the Rhodes et al. article.

both this new result and the result reported by Rhodes et al. (Experiment 1) using the distributed memory model. The success of the simulation indicates that even though the architecture of the model dictates that stimulus information and contextual information make independent contributions to word identification, the model is able to generate an effect of semantic priming on sensitivity.

The method of simulating lexical decisions in the distributed memory model, in which energy is taken as a measure of a target's familiarity, can be compared to that of other connectionist models. The Seidenberg and McClelland (1989) model performed lexical decision by evaluating the pattern of activation in orthographic units (and possibly phonological units as well). The influence of semantics on these units was not implemented in the Seidenberg and McClelland model, so it could not produce any effect of semantic context. In their attractor network model of word identification, Hinton and Shallice (1991) used proximity of a generated pattern of activation in semantic units to semantic patterns of known words as a basis for lexical decisions. This model could, in principle, simulate an effect of semantic context on sensitivity in lexical decision because semantic units are directly involved in determining the lexical decision response.

Plaut and Shallice (1993) adopted a similar approach to that of Hinton and Shallice (1991), except that in their model, lexical decision was based on the proximity of the pattern of activation in graphemic units to the graphemic patterns of known words. The formation of a graphemic pattern of activation in the Plaut and Shallice model is determined in part by feedback from semantics, making it likely that semantic priming would influence sensitivity in the lexical decision task. Finally, in their connectionist model of word reading, Plaut et al. (1996) proposed that lexical decisions would be determined in part by evaluating activation in semantic units, although this process was not implemented. By including semantic units in the determination of lexical decisions, this model has the potential to produce context effects on sensitivity.

Thus, there is some variability in how connectionist models create the potential for semantic priming to affect sensitivity in the lexical decision task. In some cases (the distributed memory model used here; Hinton & Shallice, 1991; Plaut et al., 1996), semantic units directly influence decisions, making feedback from semantics to orthography (a violation of informational encapsulation) unnecessary with respect to simulation of context effects on sensitivity. In the Plaut and Shallice (1993) model, on the other hand, only graphemic units are used to determine lexical decisions, so feedback from semantics to graphemic units is required if the model is to generate sensitivity effects.

The ability of some models based on distributed representations to simulate the effects of priming on sensitivity, without assuming a top-down influence of context on perceptual encoding, supports the argument that sensitivity effects cannot be taken as clear evidence that context has affected perceptual encoding (Norris, 1995).

The success of the distributed memory model in simulating priming effects on sensitivity equals that of the criterion bias model described by Norris and adds to the plausibility of the model.

EXPERIMENT 2

A contrasting aspect of the distributed memory model and the criterion bias model (Norris, 1995) is the nature of knowledge representation. The criterion bias model is based on a localist representation of words, in which each word is represented by a single logogen. In this system, word identification tasks, such as lexical decision, are accomplished by evaluating the level of activation in each logogen. If one or more logogens are activated beyond some threshold, a positive decision is made. In this class of model, then, the output of the word recognition system is not simply a signal versus noise discrimination, but instead consists of some indication of which word is the input. In the masked lexical decision task, however, it seems plausible that subjects may have made accurate positive responses on many trials without being aware of the target word's identity, using instead a sense of familiarity, as suggested by the distributed memory model. Consistent with this suggestion, Bub and Arguin (1995) reported that a patient with pure alexia, who showed the characteristic pattern of laborious letter-by-letter reading when asked to identify words, achieved above-chance lexical decision accuracy when the exposure duration of targets was well below that normally required for the patient to report the word's identity. Therefore, the masked lexical decision task may have been particularly well suited for models—such as the distributed memory model—that rely on a single evaluative dimension for discriminating between words and nonwords.

In Experiment 2, we used a different word identification task, one that emphasized identification of the target word. The task we selected was based on a matching-to-sample procedure. In this task, a masked target word was followed by a clearly presented probe word and the subject decided whether the probe matched the target. We expected that performance on this task would depend much more on identification of the specific target item than is the case with the masked lexical decision task. If this identification task minimizes the reliance on a global evaluative dimension such as familiarity, it might be more difficult for the distributed memory model to simulate a priming effect on sensitivity, should such an effect be found.

It was not clear whether a semantic priming effect on sensitivity would be found using this task. We know of no reports of such an effect using a task that requires subjects to judge target identity (as opposed to categorizing targets into a broad class, such as word/nonword or famous/nonfamous). In anticipation of a small or null effect of context, a larger sample of subjects was used in Experiment 2 than in Experiment 1 to reach an adequate level of statistical power.

Method

Subjects. Forty-eight students were drawn from the same population as in Experiment 1. All had normal or corrected-to-normal vision and considered English to be their first language. Two additional subjects were tested. Their data were not included in the analyses because their performance was not significantly above chance.

Materials and Design. The critical stimuli were the 120 word triplets used in Experiment 1, each consisting of a prime and two related associates. The design of the experiment was a factorial combination of context (related and unrelated) and probe (match and nonmatch). The triplets were broken into four blocks of 30 for counterbalanced assignment of blocks to each of the four conditions in the design. Across subjects, each block of triplets appeared equally often in each condition. For half of the subjects tested in each counterbalanced assignment, one of the associates served as the target; for the other half of the subjects, the other associate was the target. For the triplets assigned to the unrelated context condition, prime words were reassigned to different triplets so that each pair of associates was assigned an unrelated prime. For triplets assigned to the matched-probe condition, the same associate was used as the target and the probe. For triplets assigned to the nonmatched-probe condition, one associate served as the target and the other was the probe. This procedure ensured that the target and probe always bore the same relationship to the prime. For example, the cue *king* and the target ROYAL would be followed by either a matching probe (ROYAL) or a foil related to the prime (CROWN).

In addition to the critical stimuli, a set of 10 word triplets was used for practice trials. Half of the practice triplets consisted of a prime and related target and the other half consisted of an unrelated prime and target pair. Furthermore, half of the practice items consisted of matching target and probe words, whereas the other half consisted of nonmatching target and probe words.

For all items, the target and probe were of the same length, so in the case of nonmatching probes, word length was not a cue that could assist discrimination between the two items. Moreover, se-

mantic association with the prime was not a cue in the case of nonmatching probes because both target and probe were related to the prime or both were unrelated to it.

Procedure. The procedure of Experiment 2 was identical to that of Experiment 1 except that on each trial the post-mask was followed by a blank interval of 50 msec and then the probe word appeared in uppercase letters for 500 msec. As in Experiment 1, the prime was in lowercase letters and the target was in uppercase letters. Although the target and probe words were both presented in uppercase letters, it is unlikely that subjects would base their responses exclusively on a physical match between the two items because the post-mask would disrupt the sensory representation of the target. A higher level representation of the target, presumably including aspects of its meaning and its identity on at least some trials, would be involved in the comparison with the probe item.

Subjects were instructed to decide whether or not the probe was the same as the target. They responded by pressing one of two keyboard keys to indicate their decision. Subjects were also told that on half the trials the target word would be related to the word presented at the beginning of the trial, so they should pay attention to that word. The procedure lasted approximately 20 min.

Results

A hit was defined as making a "same" response when the probe matched the target, whereas a false alarm was defined as making a "same" response when the probe was different from the target. The mean hit and false alarm probabilities, A' scores, and B''_D scores are shown in Figure 4. These data were analyzed in the same way as in Experiment 1. The proportion of hits was significantly greater in the related context condition [$F(1,47) = 9.26$, $MS_e = 0.006$, $\eta^2 = 0.16$], but there was no effect of context on the proportion of false alarms ($F < 1$). This pattern of results was consistent with the finding that A'
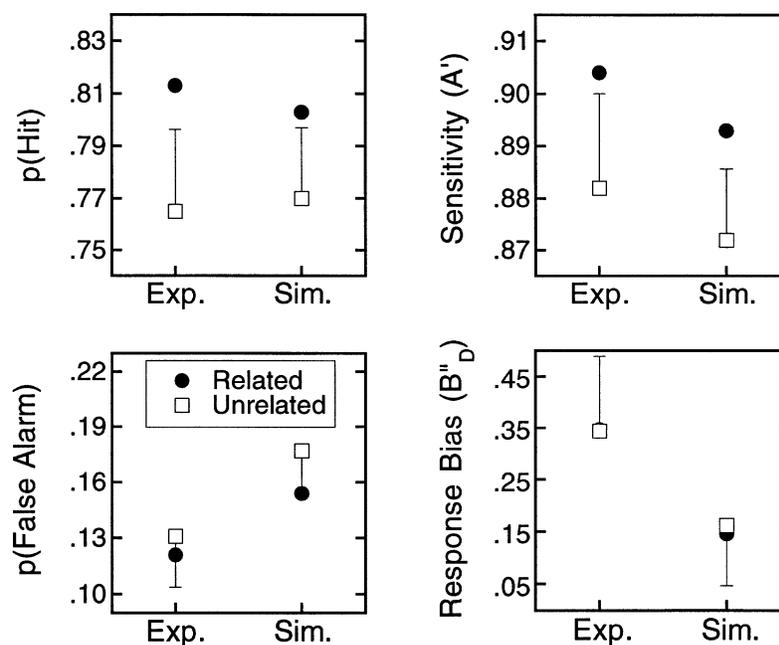

Figure 4. Mean hit and false alarm probabilities, sensitivity, and bias as a function of context for experimental (Exp.) and simulated (Sim.) results in Experiment 2. Error bars represent the 95% confidence interval for the difference between means.

scores were significantly higher in the related context condition [$F(1,47) = 6.46$, $MS_e = 0.002$, $\eta^2 = 0.12$]. There was no reliable effect of context on bias (B''_D, $F < 1$).

Simulation results. We attempted to simulate the observed effect of priming on sensitivity in the probe task using the distributed memory model characterized in Figure 1. We initially tried to use energy in the network as a means of determining whether the probe matched the target by computing energy as the probe was processed. A higher level of energy, or familiarity, would indicate that the probe matched the masked target that preceded it. Energy had been used as a measure of familiarity to simulate the lexical decision results of Experiment 1 and of Rhodes et al. (1993, Experiment 1). With the probe task, however, the model was not able to discriminate adequately between cases in which there was or was not a match; the network's energy was dominated by processing the probe, and the small effect of the difference between a matching and a nonmatching target's brief presentation on the growth of the energy function was too weak to be reliably detected. Thus, the distributed memory model could not use energy as a basis for reliably discriminating between matching and nonmatching probes.

Rather than using energy as a basis for simulating responses on this task, then, we assumed that subjects would obtain partial information from the target display and compare memory for this information with a full encoding of the clearly presented probe word. To do this, simulation runs were conducted just as in the simulation of Experiment 1, except that the network continued with updating cycles for the duration of the post-mask and during the blank period before presentation of the probe word. Just prior to the appearance of the probe word, the pattern of activation in the network was copied, with a small perturbation, into a memory buffer. The probe word was then fully instantiated in the network and its pattern was compared with the pattern in the memory buffer. If the pattern in the network, representing the probe, was sufficiently similar to the contents of the memory buffer, the simulation classified the probe as matching the target. Otherwise the probe was classified as different from the target.

Comparison of these two patterns of activation is different from the computation of energy that was used to simulate lexical decision. Although both approaches use information from all modules in the network, the computation of energy is determined by the goodness of fit between the connection weights and the state of the processing units. In contrast, comparison of patterns of activation does not consider the connection weights at all, but uses only the pattern of activation states in the processing units themselves. This comparison permits the model to make fine-grained discriminations between stimuli, particularly under conditions that might produce very similar energy values for different stimuli.

Two different criteria were used in making a decision about the probe, one for related-prime trials and one for unrelated-prime trials. A higher criterion was required for the related condition because the number of matching units on related-prime trials for nonmatching probes was, on average, greater than on unrelated-prime trials for matching probes. Using a single criterion would have rendered the model incapable of capturing the observed pattern of results. We view the use of two different criteria as psychologically plausible for the following reason. Subjects knew on each trial whether the probe was semantically related to the prime because both items could be clearly read. Under these circumstances, we suspect that subjects experienced a noticeably greater sense of similarity between probe and target in the case of related primes. Furthermore, it would be feasible for subjects to adjust their criterion for assessing this similarity because they knew whether the prime was related or unrelated.

In the simulation, hits were defined as correctly classifying a probe that matched the target, and false alarms were defined as incorrect responses to nonmatching probes. The mean hit and false alarm probabilities and the resulting mean sensitivity and bias scores from the simulation are shown in Figure 4. Separate ANOVAs were used to test each of these measures for an effect of context. The model produced a reliably greater probability of hits in the related condition [$F(1,49) = 6.16$, $MS_e = 0.005$, $\eta^2 = 0.11$]. The effect of context on false alarm probability approached significance [$F(1,49) = 3.26$, $MS_e = 0.003$, $p < .10$, $\eta^2 = 0.06$]. Sensitivity was reliably higher in the related context condition [$F(1,49) = 9.52$, $MS_e = 0.001$, $\eta^2 = 0.16$]. Finally, response bias did not significantly vary with context ($F < 1$).

The distributed memory model's simulation of the probe-matching task captured the primary features of the results of Experiment 2, particularly the significant effect of context on sensitivity. The context effect found in Experiment 2 can be simulated, then, in a model that does not include any mechanism whereby contextual information can directly influence perceptual processes. The effect of a related prime in this simulation was to move the semantic units into a pattern that was compatible with the upcoming target stimulus. During processing of the target, semantic units in the network were moved even further into a pattern consistent with the target, providing an additional source of information when the remembered target pattern of activation was compared with the probe item.

We note that unlike the results of Experiment 2, the results of the simulation showed a trend for an effect of context on false alarm probability, with a higher false alarm rate in the unrelated condition. This false alarm rate could have been reduced by increasing the criterion for a positive response in the unrelated condition, but the smallest increase possible would have produced too large a change. This problem arises because the model's response criterion was defined as the proportion (or number) of matching units when the buffered representation of the target and the full representation of the probe were compared. This basis for responding places a limit on how small an effect the smallest change in criterion can have. In particular, if we were to increase the criterion for declaring a match in the unrelated context

condition by one processing unit (the smallest change possible), the simulation would have produced hit and false alarm probabilities and a response bias that would be too low. We do not see this fault as a significant problem for the model since it could be overcome in a number of ways that would produce more finely grained criterion settings, such as by increasing the number of units in the network or by declaring a match with some probability when the number of matching units is exactly equal to the criterion. We were not seriously concerned about producing highly accurate quantitative fits to the data, so we did not pursue this issue.

We also attempted to simulate the results of Experiment 2 with a version of the criterion bias model described by Norris (1995) because that model has not been applied to the probe task that we used. Because the criterion bias model was designed to produce output that provides some indication of a target word's identity, we expected that it would successfully simulate the probe task results.

The details of the implementation of the criterion bias model are presented in the Appendix, but a general account of the simulation is as follows.[2] Words were represented as individual logogens, each with its own level of activation. In the model, a logogen's level of activation is determined by perceptual input (e.g., visual presentation of the word that corresponds to the logogen) and by activation that spreads from other, semantically related logogens. The probe task was simulated by determining whether any of the logogens were activated beyond a specified criterion by presentation of the target word. Priming had the effect of increasing the activation of logogens representing words related to the prime, and this effect, in turn, increased the probability that on related trials the target word's logogen would be included in the set of logogens activated beyond criterion. A positive response was made if an activated logogen matched the probe. A negative response was made if none of the activated logogens matched the probe. If no logogens were activated beyond criterion, the model guessed by selecting a positive or a negative response with equal probability. On matching trials, a related prime increased the likelihood of activating the target word's logogen and therefore increased the hit rate. On nonmatching trials, activating at least the target's logogen meant that the model did not have to guess, so a related prime reduced the probability of a false alarm.

The results of the simulation, in the form of hit and false alarm rates and A' and B_D'' measures, are shown in Table 1. These results are based on 100,000 trials in each of the four conditions of the design (related vs. unrelated prime crossed with matching vs. nonmatching probe), and therefore the simulation's estimates of hit and false alarm rates are very stable; following Norris (1995), no statistical tests of differences between conditions were conducted. The net effect of increasing the probability of activating the target word's logogen in the related-prime condition was to increase sensitivity in that condition and to establish a more liberal response bias. Al-

Table 1
Simulation Results for the Probe-Matching Identification Task Using the Criterion Bias Model

	Prime	
Measure	Related	Unrelated
Hit rate	.88	.80
False alarm rate	.19	.22
A'	.91	.87
B_D''	−.26	−.05

though the sensitivity effect produced by the criterion bias model fits the results observed in Experiment 2, the bias effect yielded by the model does not. The criterion bias model's bias effect, however, could have been eliminated by decreasing the probability of a positive response for guesses in the related-prime condition (this modification would correspond to the use of two different matching criteria with the distributed memory model). Because precise quantitative fits were not our goal, we did not pursue this issue.

Discussion

As expected, although a context effect on sensitivity was obtained with the probe-matching task, the effect was much smaller ($\eta^2 = 0.12$) than when a masked lexical decision task was used ($\eta^2 = 0.68$). Nevertheless, the distributed memory model and the criterion bias model both successfully simulated that sensitivity effect, even though the models make no provision for semantic information to influence perceptual processes. Once again, these models provide existence proofs that a priming effect on sensitivity does not necessarily imply that priming influences perceptual encoding. The distributed memory model's simulation of the probe-matching task demonstrates that the model's integration of contextual information and stimulus information is adequate to produce a sensitivity effect even when the task requires identification of a particular target, rather than a general assessment of target familiarity.

EXPERIMENT 3

It has been argued that context effects involving word primes and targets could simply reflect associations between representations within a word perception module (e.g., Fodor, 1983; Forster, 1979, 1981; Seidenberg et al., 1984). The criterion bias model used by Norris (1995) fits this assumption as well. In that model, no provision is made for logogens to be affected by higher level information. All that is required is that logogens representing semantically related words be capable of increasing one another's level of activation when any one of them is activated by visual input. Rhodes et al. (1993) acknowledged that the effect of word context on lexical decision is not necessarily an exogenous effect in that it could be construed as an effect occurring exclusively within a word perception module.

One way to test the possibility that the context effect observed in Experiments 1 and 2 was due to within-module associations is to use different modalities for contexts and targets. Rhodes and Tremewan (1993) adopted this approach in their study of face identification by using printed names as primes. They obtained a context effect on sensitivity in a task that required subjects to discriminate between famous and nonfamous faces. That context effect is not easily explained by appeal to within-module associations because identification of the prime and the target would occur in separate input modules. Priming across modalities would require that input modules be affected not only by perceptual input, but also by each other or by a top-down influence from a common semantic code, thereby violating the assumption of informational encapsulation.

A proposal of the latter kind, involving the influence of semantic information on an input module, is embodied in Morton's (1969) logogen model. In that model, it is assumed that there is a single, modality-independent, input logogen system. Input logogens are affected both by perceptual input and by a cognitive system that represents semantic codes. The top-down influence of the cognitive system provides for effects of context on word identification (e.g., semantic priming). More recent versions of the logogen model incorporate modality-specific input logogen systems (visual, auditory, and pictorial) to accommodate modality-specific effects in repetition priming (see, e.g., Clarke & Morton, 1983; Jackson & Morton, 1984; Morton, 1979). In these versions, however, a cognitive system consisting of semantic codes is assumed to have a top-down influence on the activation of input logogens. A model of this form would predict cross-modal priming effects on measures of sensitivity.

In Experiment 3, then, we conducted a test of cross-modal semantic priming analogous to that used by Rhodes and Tremewan (1993), but using the probe-matching task from Experiment 2. Line drawings were used as primes and words as targets. If the context effect in Experiment 2 was due only to within-module associations, no effect should be found in Experiment 3 on the assumption that modality-specific input modules are informationally encapsulated. The criterion bias model, as implemented by Norris (1995), would not predict a context effect under these conditions unless a system with semantic codes having a top-down influence on the input logogens were included.

In the distributed memory model, we assumed that pictorial input would act in much the same way as orthographic input to semantic units. No top-down influence of semantic units on either orthographic or pictorial input units is assumed. As before, however, we assumed that information about the target from all of the model's processing units would be integrated to produce a response on the probe-matching task. Therefore, we expected that even with line drawings as primes, an effect of priming on sensitivity would be observed and that the distributed memory model would simulate that result.

Method

Subjects. Seventy-two subjects were drawn from the same population as in Experiments 1 and 2. All subjects had normal or corrected-to-normal vision and considered English to be their first language. Nine additional subjects were tested, but their performance was not significantly better than chance, so their data were not included in the analyses. A larger sample of subjects was used in Experiment 3 than in Experiment 2 to increase the power to detect an effect of context.

Materials and Design. The critical stimuli were 100 triplets, each consisting of a line drawing taken from the Snodgrass and Vanderwart (1980) set and two words judged to be semantically related to the drawing (e.g., a drawing of a fork was grouped with the words *knife* and *spoon*). For many of the triplets, it was not possible to find two semantically related words of the same length, so in these cases the words varied in length. An additional 20 sets of items, each consisting of a line drawing and one or two words, were used on practice trials. Half of these items consisted of semantically related drawings and words, and half consisted of unrelated items; half consisted of matching target and probe words and half consisted of nonmatching target and probe words.

The design of Experiment 3 was the same as that of Experiment 2: a factorial combination of context (related and unrelated) and probe (match and nonmatch). The critical items were treated as four blocks of 25, and the assignment of blocks to conditions was counterbalanced across subjects so that each triplet was used equally often in each condition. One of the two words in a triplet served as the target for half of the subjects and the other word was the target for the other half. On related trials, the line drawing was followed by one of its related words; on unrelated trials it was followed by a word taken from a different triplet in the same block. For items assigned to the match condition, the probe word was the same as the target; for items in the nonmatch condition, the probe was the other word from the target's triplet. As in Experiment 2, then, the target and probe always had the same relationship to the prime.

Procedure. A Macintosh II microcomputer with a monochrome monitor was used to display stimuli and collect responses. The monitor refresh rate of this equipment dictated increments of 15 msec in the timing of displays. Subjects responded by pressing one of two keys mounted on a box that were labeled "yes" and "no." The sequence of events on each trial, similar to that in Experiment 2, was as follows: (1) a fixation cross in the center of the monitor, (2) initiation of the trial by the subject pressing one of the response keys, (3) a 255-msec blank interval, (4) a line drawing for 195 msec, (5) a 600-msec blank interval, (6) a pre-mask for 45 msec, (7) the target word in uppercase letters for 45 msec, (8) a post-mask for 45 msec, (9) a 45-msec blank interval, and (10) a probe word in uppercase letters, which remained on the screen until the subject responded. Subjects were instructed to try to identify the briefly presented target word. They were told to press the "yes" key if the probe word matched the target and to press the "no" key if it did not. They were also told that on half the trials the target word would be related to the drawing, so they should pay attention to the drawing.

Because we were unable to obtain word pairs of the same length, the target display included a set of ampersands that flanked the target, creating a display 15 characters wide (e.g., &&&&&HAMMER&&&) on each trial. The pre- and post-masks were also 15 characters in length. This procedure was used to prevent subjects from using word length information as a cue for discriminating the target and probe.

The experiment began with a set of 20 practice trials. Target duration was initially 225 msec but was decreased by 15 msec for each subsequent practice trial and remained at the critical duration (45 msec) for the remaining eight practice trials and for the subsequent critical trials. The practice trials were intended to help sub-
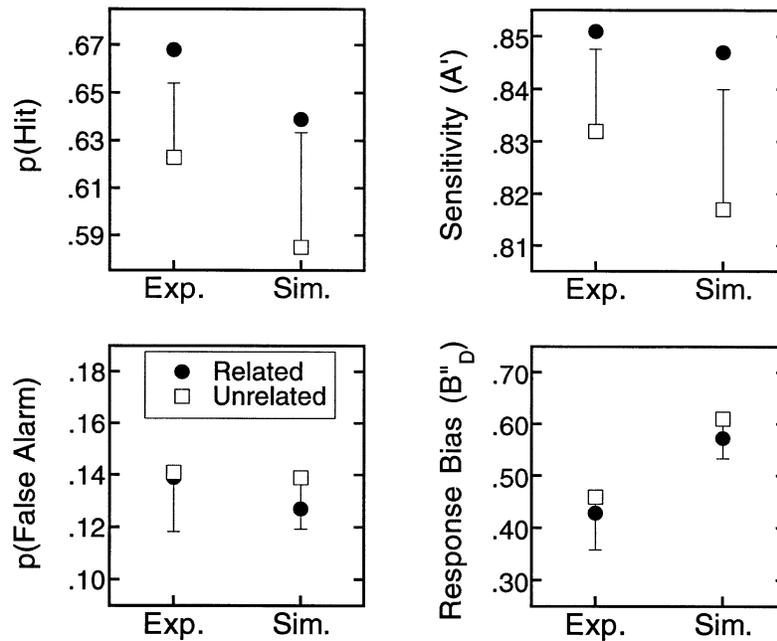
Figure 5. Mean hit and false alarm probabilities, sensitivity, and bias as a function of context for experimental (Exp.) and simulated (Sim.) results in Experiment 3. Error bars represent the 95% confidence interval for the difference between means.

jects understand the sequence of events before receiving trials with the very brief critical target duration. The 100 critical trials were then presented in random order. The procedure took about 20 min to complete.

Results

Hits and false alarms were defined as in Experiment 2. The mean hit and false alarm probabilities, A' scores, and B''_D scores are presented in Figure 5. Separate ANOVAs were used to test for context effects on each of these measures. These analyses revealed a pattern identical to that found in Experiment 2. Hits occurred with higher probability in the related context condition [$F(1,71) = 8.14$, $MS_e = 0.009$, $\eta^2 = 0.10$], but there was no effect of context on the probability of making a false alarm ($F < 1$). There was a reliable effect of context on sensitivity (A') in favor of the related condition [$F(1,71) = 5.71$, $MS_e = 0.002$, $\eta^2 = 0.08$]. Finally, context did not significantly affect response bias (B''_D, $F < 1$).

Simulation results. Simulation of the results of Experiment 3 using the distributed memory model required a change in the model's architecture so that it could handle processing of line drawings. We assumed that drawings or objects would be represented by a set of processing units (called "object units") that function much as orthographic units do, except that they do not directly influence phonological units. Instead, object units directly influence only semantic units. This arrangement is depicted in Figure 6. According to this architecture, a task such as naming an object requires instantiating, at least partially, the object's semantic representation. This constraint is consistent with models in the object-naming lit-

erature (e.g., Smith & Magee, 1980; Theios & Amrhein, 1989; Vitkovitch & Humphreys, 1991).

Note that this architecture is entirely modular with respect to both word perception (the orthographic units) and object perception (the object units). The only input into the orthographic and object units is the external stimulus. There are no top-down connections or connections between orthographic and object modules. Therefore, perceptual encoding processes are entirely independent of semantic context in this model.
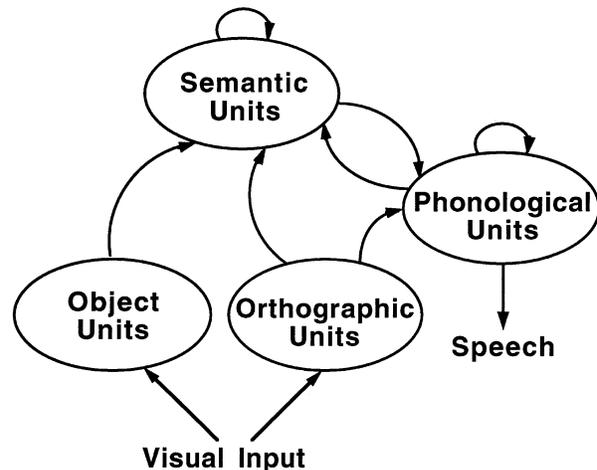

Figure 6. A modified version of the distributed memory model architecture for simulation of word identification with line drawings as primes (Experiment 3).

In the simulation of Experiment 3, the model worked in the same way as in the simulation of Experiment 2 except that primes were represented as line drawings and their input was channeled through the object units rather than the orthographic units. The mean hit and false alarm probabilities and the mean sensitivity and response bias scores are shown in Figure 5. Each of these measures was analyzed using an ANOVA. There was a reliably greater probability of a hit and reliably higher sensitivity in the related context condition [$F(1,49) = 4.88$, $MS_e = 0.014$, $\eta^2 = 0.09$, and $F(1,49) = 6.93$, $MS_e = .003$, $\eta^2 = 0.13$, respectively]. There was no effect of context on false alarm probability [$F(1,49) = 1.56$] or on response bias ($F < 1$).

No attempt was made to simulate the results of Experiment 3 using the criterion bias model (Norris, 1995). That model would be unable to simulate a cross-modal priming effect unless it is assumed that semantic codes, activated by pictorial input, have a top-down influence on the input logogens that respond to printed words. If that assumption is made, then simulation results like those shown in Table 1 would be obtained.

Discussion

A significant effect of semantic context on signal detection sensitivity was found in the probe-matching task with drawings of objects as primes. On the view held by Farah (1989), Rhodes et al. (1993; Rhodes & Tremewan, 1993), and others, this result contradicts models that assume that contextual information and stimulus information are processed independently. The results of Experiment 3 also satisfy the two conditions for violation of the modularity assumption put forward by Fodor (1983)—namely, an exogenous effect obtained on a perceptual metric. A similar demonstration of apparent cross-module priming on sensitivity was interpreted by Rhodes and Tremewan (1993) as a demonstration of a violation of modularity in a face recognition task. Rhodes and Tremewan found a significant effect of name primes on accuracy (measured by sensitivity in a signal detection analysis) in the classification of pictures of faces as famous or nonfamous (e.g., priming a subject with the name "Simon" increased the accuracy of a judgment about a picture of Garfunkel). They took that result as evidence that the face perception system appears to be nonmodular. From this perspective, the results of Experiment 3 constitute evidence for the claim that the word perception system also is nonmodular.

The criterion bias model proposed by Norris (1995) would be able to account for the priming effect found in Experiment 3 if it were to incorporate a top-down influence of semantic codes on the model's input logogens. This modification, however, would mean that the criterion bias model's perceptual input systems are not modular. Thus, the criterion bias model does not stand as an existence proof that cross-modal priming can be explained without recourse to violations of modularity.

In contrast, the distributed memory model successfully simulated the pattern of data from Experiment 3,

particularly the effect of context on sensitivity. The model succeeded in producing the sensitivity effect even though there was no means by which semantic context could affect perceptual processing or the units responsible for constructing an input code. We take this simulation result as a clear existence proof that, for the probe-matching task, an effect of semantic context on sensitivity does not necessarily imply that the effect involves perceptual encoding or a top-down influence of semantic context. Instead, the context effect can arise from the integration of multiple sources of information, including semantic, orthographic, and phonological evidence, but without any of these sources changing the amount or quality of stimulus information. Given that a plausible independence model is capable of simulating the sensitivity results obtained here, we conclude that these results cannot be taken as sufficient evidence for a violation of the general assumption of independent processing of stimulus and contextual information, or of the more specific assumption of modularity in word perception.

FURTHER APPLICATIONS OF THE DISTRIBUTED MEMORY MODEL

In addition to the sensitivity effects considered here, alternative analytic models to signal detection theory have been used to address the question of whether context affects perceptual encoding processes. Reinitz and his colleagues (Reinitz & Alexander, 1996; Reinitz et al., 1989) proposed two methods for testing the validity of a multiplicative model of priming effects. By the multiplicative model, a variable such as context affects the rate of information acquisition, with relevant context increasing that rate by a constant factor. In the version of the multiplicative model put forward by Reinitz, it is claimed that the rate of *stimulus* information acquisition is increased by a constant factor. Both methods of testing the multiplicative model involve measuring task performance across a range of target exposure durations and plotting performance as a function of these durations. For example, in testing the effect of semantic context, target identification accuracy is plotted separately for related and unrelated contexts as a function of exposure duration.

In the method developed by Reinitz et al. (1989), if the effect of related context is to increase, by a factor of *c*, the rate of stimulus information acquisition, then the amount of information obtained in the related context condition with *d* msec of exposure to the target should be equal to the amount obtained in the unrelated condition with *cd* msec of exposure. Therefore, a plot of performance as a function of exposure duration in the two context conditions should reveal horizontally diverging functions. More specifically, the divergence should be such that the target duration needed for performance to reach a given level will be greater by a factor of *c* in the unrelated condition than in the related condition.

Reinitz et al. (1989) proposed that a straightforward test of this prediction can be obtained by plotting per-

formance as a function of log target duration. Doing so will transform equal ratios in the linear scale into equal distances on the log scale. Thus, the performance functions plotted against log duration are predicted to be horizontally parallel, separated by a constant horizontal distance of log(c). Laterally shifting one of the functions by that distance should allow it to be superimposed on the other function. If the two functions can be superimposed in this way, then one has support for the conclusion that related and unrelated-prime conditions differ in the rate of stimulus information acquisition.

Reinitz et al. (1989) obtained horizontally parallel functions on a log duration scale in their data taken from a picture identification task in which target pictures were preceded by related or unrelated context words. Although this result suggests that information acquisition occurred at a higher rate for the related-prime condition, it is not clear that an effect of this kind involves perceptual information acquisition. We note that the relevant accuracy functions are not necessarily determined by exclusively perceptual processes. That is, accuracy may be codetermined by the integration of stimulus and contextual information as proposed in the distributed memory model. Therefore, we suspect that the general method proposed by Reinitz et al. method has the same limitation as the signal detection analysis approach.

To test this suspicion, we attempted to produce multiplicative effects of target duration and context on identification accuracy using the distributed memory model to simulate the probe-matching task. This simulation involved a replication of the simulation reported in Experiment 2 using a range of target exposure durations. The only change from that earlier simulation, besides varying exposure duration, involved varying the model's criteria (recall that there were separate criteria for related- and unrelated-prime trials) for making a "same" response; the criteria became more strict, with longer durations. If the distributed memory model, which has no connections in place to permit context effect to affect perceptual encoding, is capable of producing horizontally parallel accuracy functions for related and unrelated context conditions, then it would be inappropriate to conclude that such a pattern is diagnostic of an effect on perceptual encoding. That result would indicate that the Reinitz et al. technique faces the same constraint as the signal detection analysis approach.

The simulation results, expressed as accuracy (A') functions for related and unrelated context conditions, are shown in Figure 7. The top panel of the figure indicates an advantage of related context across the range of exposure durations. In the middle panel, a $\log_{10}$ transformation has been applied to exposure duration and the functions appear to be horizontally parallel. Shifting the function for the related condition to the right, as shown in the bottom panel of Figure 7, clearly indicates that the transformed functions satisfy the condition of being horizontally parallel. Even though the model precludes any influence of context on perceptual processing, it naturally produces horizontally parallel accuracy functions
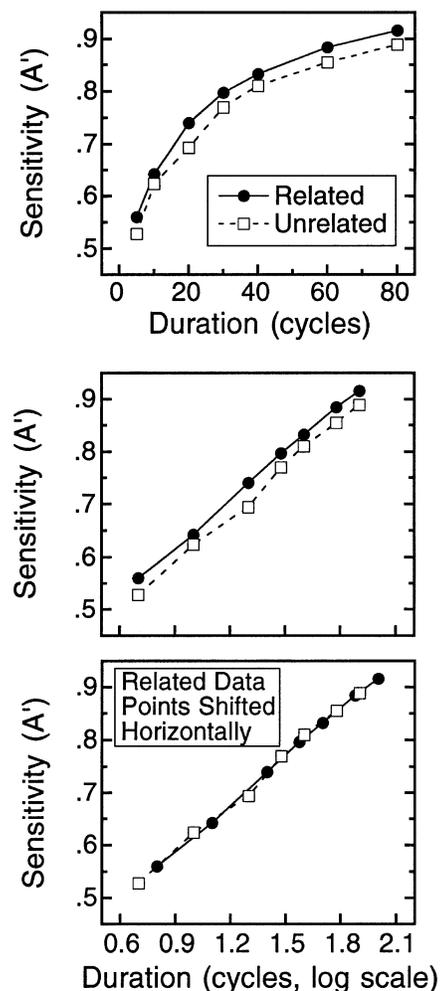

Figure 7. Simulation of the word identification task as a function of target exposure duration. The top panel shows exposure duration on a linear scale; the middle and bottom panels show exposure duration on a $\log_{10}$ scale. In the bottom panel, the function for the related condition has been shifted to the right to show that the functions for the related and unrelated conditions can be superimposed almost perfectly.

across a broad range of target durations. Thus, finding horizontally parallel functions, like finding an effect of context on sensitivity, does not by itself necessarily imply that perceptual encoding has been affected.

In fairness to Reinitz et al. (1989), we note that in addition to demonstrating horizontally parallel functions for the effect of priming on object identification, they obtained converging evidence to support their conclusion that priming specifically affected visual information acquisition. They conducted three experiments involving a memory test in which it was shown that priming the presentation of an object or scene with its category name (e.g., a picture of a dog primed with the category name DOG) led to improved recognition memory for that item when tested with a set of distractors from the same category (i.e., pictures of other types of dogs). These re-

sults, however, are not completely convincing with respect to the effect of context on the *acquisition* of visual information. Because a delayed memory test was used rather than an immediate discrimination test, it is possible that the nature of the prime may have affected rehearsal of the target item (e.g., appropriately primed targets received more effective rehearsal), which in turn would affect performance on the subsequent memory test in the absence of any influence on the original perceptual encoding of the item. A more powerful demonstration of the effect of context on perceptual encoding would be based on an immediate discrimination test, as proposed in the General Discussion.

A second method of testing the multiplicative model was proposed by Reinitz and Alexander (1996), building on the observation that performance, when measured as proportion correct, is an exponential function of target exposure duration (see, e.g., Loftus, Busey, & Senders, 1993; Loftus, Duncan, & Gehrig, 1992). Specifically, proportion correct, p, is equal to

$$1 - e^{-(d-L)/c}$$

when $d \geq L$ and is 0 when $d < L$. In this equation, d is target duration, L is the duration at which performance lifts off from 0, and c is the exponential growth constant. Transformation of p by the formula $-\ln(1.0 - p)$ results in a dependent measure, P, that is a linear function of exposure duration, d, when $d \geq L$,

$$P = (1/c)d - (L/c),$$

and is 0 when $d < L$. This linear function has slope equal to $1/c$, a y-intercept of $-L/c$, and an x-intercept of L (the duration at which performance begins to rise above 0). The slope of the linear function relating performance to duration is an indication of the rate of information acquisition. By the multiplicative model, the slope of the performance function should be steeper for primed than for unprimed targets.

Reinitz and Alexander (1996) tested this prediction of the multiplicative model in a masked word identification paradigm in which a target word was briefly presented

and the task was to identify the word. Rather than using context words as primes, Reinitz and Alexander examined repetition priming. Primed target words were presented in an earlier encoding task, prior to appearing in the masked word identification task. Performance on primed and unprimed targets was assessed across a range of exposure durations, and performance was transformed into P values, according to the equation above. Reinitz and Alexander found that these transformed performance functions were well fit by linear functions that had very similar x-intercepts and substantially different slopes. The finding of similar, nonzero x-intercepts indicated that visual encoding begins following some minimum amount of visual exposure (Busey & Loftus, 1994; Loftus & Ruthruff, 1994), independent of priming condition. Moreover, the different slopes were taken by Reinitz and Alexander as evidence that visual information acquisition occurred at a more rapid rate for primed than for unprimed targets. To support their conclusion that repetition priming had its effect on perceptual processing, Reinitz and Alexander cited various bodies of literature that suggest that priming effects in the masked word identification task are perceptually based.

We agree that it is advisable to seek converging evidence when using the results of this technique to support the conclusion that priming affects perceptual encoding processes. This approach is necessary because the finding of different slopes for primed and unprimed items is not a unique prediction of the multiplicative model. To show this, we applied the slope comparison technique to the results of the distributed memory model simulation that were used to test for horizontally parallel accuracy functions above. Recall that in that simulation, the effect of semantic priming on a probe-matching word identification task was simulated. Although those results do not involve simulation of repetition priming in the masked word identification task, they can be used to inform the more general issue of whether an independence model, like the multiplicative model, predicts linear functions with different slopes for primed and unprimed targets. To obtain performance values that would have the same
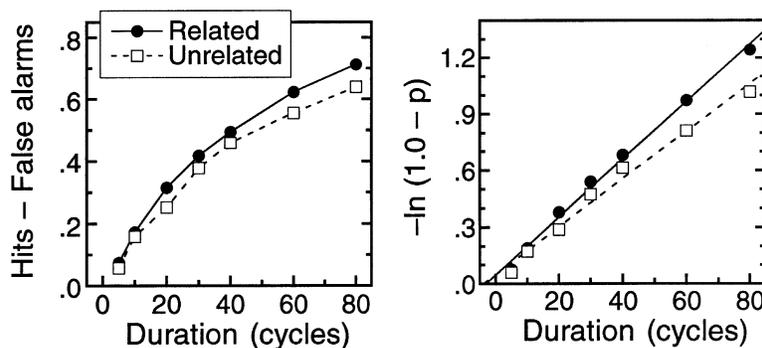
Figure 8. Simulation of the word identification task as a function of target exposure duration. Performance is shown in the left panel as hits minus false alarms. In the right panel, performance has been transformed using the formula $-\ln(1-p)$, where p is equal to the difference between hit and false alarm rates. Simulated data in the right panel are shown as open and filled symbols; lines indicate the best fitting linear functions.

property as proportion correct (i.e., range from 0 to 1.0), we used hits minus false alarms as a measure of accuracy, rather than A' scores. We did not use A' scores because normal performance on that measure ranges from 0.5 to 1.0. Moreover, A' is a nonlinear transformation of proportion correct (hit and false alarm rates) and may not conform to the exponential function described by Loftus and colleagues (Loftus et al., 1993; Loftus et al., 1992) that relates accuracy to target duration. The mean performance scores, expressed as hit rate minus false alarm rate, from the distributed memory model as a function of target exposure duration are shown in the left panel of Figure 8. These scores are simply an alternative expression of the model's performance, shown in the top panel of Figure 7 as A' scores. In the right panel of Figure 8, the model's performance has been transformed to P values; best fitting linear functions are shown as solid lines.

The linear functions fit the model's data very well, producing r^2 values of .995 and .986 for the related- and unrelated-prime conditions, respectively. These linear fits are comparable to those reported by Reinitz and Alexander (1996) for primed and unprimed word identification (.993 and .997, respectively). The x-intercept values, L, estimated by the linear functions derived from the model's data, were -3.00 and -3.83 cycles for the related and unrelated conditions, respectively. These estimates of L are about as similar to one another as were those obtained by Reinitz and Alexander (34.33 and 33.22 msec for the primed and unprimed conditions, respectively). The model's estimates of L, however, are very close to 0. The L parameter would be expected to be equal to 0 for data produced by the distributed memory model because that model begins to acquire relevant information on the first updating cycle. The model does not simulate the preliminary visual processes, prompted by stimulus onset, that lead up to the initiation of information acquisition (Busey & Loftus, 1994; Loftus & Ruthruff, 1994).

Finally, and most important, the slope of the linear function in the related condition was larger than the slope in the unrelated condition (.0154 vs. .0127). Although these slope values are comparable to those reported by Reinitz and Alexander (1996) for the repetition priming effect (.0160 and .0113 for primed and unprimed words, respectively), we do not place much emphasis on this similarity. The slopes obtained from the model are based on duration measured in updating cycles, whereas those found by Reinitz and Alexander are based on duration measured in milliseconds. We have assumed a 2:1 ratio of cycles to milliseconds, which suggests that the model's slopes are about two times steeper than those found by Reinitz and Alexander when expressed on a common scale. This difference in slopes is not troublesome because the two studies used different variants of the masked word identification task; the model simulated the probe-matching version, whereas Reinitz and Alexander used free report. We would expect that asymptotic performance is reached after a shorter

duration in the probe-matching task than in the free-report task (see Masson & MacLeod, 1996, for related evidence regarding visual contrast instead of duration).

These simulation results demonstrate that criteria set by two different quantitative tests of the multiplicative model of priming effects can be met by the distributed memory model. By the version of the multiplicative model advocated by Reinitz (Reinitz & Alexander, 1996; Reinitz et al., 1989), it is assumed that priming increases the rate of stimulus information acquisition. The significance of the distributed memory model's success in meeting these criteria is that it produces the appropriate relations between performance and exposure duration for related and unrelated contextual priming conditions without making any changes to the rate at which stimulus information accrues. The patterns predicted by the multiplicative model were a natural consequence of the distributed memory model's architecture.

We are led, therefore, to the same conclusion we reached in our application of the model to the interpretation of sensitivity effects based on signal detection theory. That is, the distributed memory model's successful simulation of the accuracy-duration functions predicted by the multiplicative model compromises the proposition that those functions can be taken as diagnostic of an effect of priming on perceptual encoding. At least in the case of word identification, those functions are readily produced by a model in which perceptual encoding is unaffected by priming. To construct a strong case for an effect of priming on perceptual encoding, one would need to obtain converging evidence to supplement the results involving accuracy-duration functions, as Reinitz and his colleagues have done (Reinitz & Alexander, 1996; Reinitz et al., 1989).

GENERAL DISCUSSION

The three experiments reported here replicate and extend results that have been taken as support for the conclusion that contextual information influences perceptual encoding. In each experiment, we found that a semantically related prime led to enhanced sensitivity on word identification tasks. In Experiment 1, we replicated the Rhodes et al. (1993) results with a masked lexical decision task. In Experiments 2 and 3, we extended that priming effect to a probe-matching task with both words and line drawings as primes. The additional significance of the priming effect with line drawing primes is that the priming effect is unlikely to have been due to associations between words within a lexical or word perception module (see Seidenberg et al., 1984). Rather, priming in this case appears to have been due to the influence of conceptual information, derived from the perception of a line drawing, on word processing. This kind of influence satisfies the criterion of an exogenous source of priming (Fodor, 1983).

The sensitivity effects reported here, taken together with the results obtained by Rhodes et al. (1993), could be used to support nonindependence models of context

effects. According to these models, context alters perceptual encoding, leading to more efficient or more accurate encoding of a stimulus. Improved encoding is reflected in the higher sensitivity associated with the related-prime condition.

Evaluating Evidence for Nonindependence Models

Although it has been proposed that sensitivity effects can be taken as evidence for changes in perceptual encoding (see, e.g., Farah, 1989; Fodor, 1983; Pitt, 1995), Rhodes et al. (1993) were cautious in their conclusion regarding the effect of context on sensitivity in their lexical decision experiments. Nevertheless, they argued that their priming effect could be interpreted as evidence for changes in perceptual encoding on the grounds that (1) conditions of their experiment maximized perceptual demands of the required discrimination; (2) semantic priming has been shown to interact with stimulus quality, suggesting a perceptual basis for the effect; and (3) Connine and Clifton (1987) demonstrated a dissociation between the effects of word context and monetary payoffs in a phoneme discrimination task, implying that word context had a perceptual effect rather than an effect on response bias. Rhodes et al. took this constellation of effects as converging evidence for the conclusion that semantic priming affects perceptual encoding.

Each of these sources of evidence put forward by Rhodes et al. (1993) is subject to an alternative interpretation. First, as discussed in the introduction, the finding of an interaction between priming and stimulus quality when word identification latency is measured (see, e.g., Becker & Killion, 1977; Borowsky & Besner, 1991, 1993) does not necessarily imply that priming affects early perceptual encoding. Interactive effects on a measure of response latency indicate a common influence that may occur at any stage of processing prior to the final response, not just the perceptual encoding stage.

Second, the Connine and Clifton (1987) finding of a dissociation between the effects of word context and monetary payoffs has been interpreted by Massaro and colleagues (Massaro & Cowan, 1993; Massaro & Oden, 1995) as evidence for two forms of bias, rather than as an effect on perceptual encoding and an effect on decision bias. In this view, monetary payoffs affect "decision bias," whereas word context affects "belief bias," which refers to a bias in the interpretation of a stimulus. As an example of the powerful nature of belief bias, Massaro and Oden pointed to the finding that viewing a speaker's face influences the interpretation of a spoken syllable. For example, hearing the syllable /ba/ while watching a speaker enunciate the syllable /ga/ leads an observer to experience yet a different syllable, such as /da/ (see, e.g., Massaro & Cohen, 1990; McGurk & MacDonald, 1976). This effect is accurately simulated by the FLMP, in which perceptual encoding is unaffected by context (the visual information), but these two independent sources of information are combined to determine an interpretation of the stimulus (Massaro, 1987; Massaro & Cohen, 1990); that is, watching the speaker's mouth movements biased

the interpretation of the auditory input without affecting the input itself. Thus, the dissociation revealed by Connine and Clifton may alternatively be interpreted as evidence for two different types of bias, rather than as evidence for a contextual influence on perceptual encoding.

Finally, the argument that the masked lexical decision task used by Rhodes et al. (1993) was perceptually demanding and therefore reflects an effect of context on perceptual encoding has been questioned on two fronts. First, Norris (1995) demonstrated that a criterion bias model in which priming has no effect on perceptual encoding could reproduce the Rhodes et al. effect of context on sensitivity. Second, we presented here a simulation of the Rhodes et al. results, as well as simulations of our three experiments, based on a connectionist model that also assumes no influence of contextual information on perceptual encoding. These simulations clearly show that plausible independence models can generate effects of context on sensitivity in word identification tasks. The alternative interpretations of sensitivity effects derived from these two models suggest that, contrary to the view held by Farah (1989) and others (e.g., Fodor, 1983; Pitt, 1995), sensitivity effects cannot readily be interpreted as an indicator of changed perceptual encoding.

Implications of Independence Models

Norris (1995) argued that priming effects on sensitivity in word identification tasks cannot be taken as evidence for enhanced perceptual encoding because task performance does not conform to the assumptions of signal detection theory, on which sensitivity measures are founded. In particular, Norris (1995) pointed out that in word identification tasks, sensitivity effects can arise from changes in bias associated with specific words in a lexicon, with no alteration whatsoever in the acquisition of perceptual information.

The validity of the argument put forward by Norris (1995) does not rest entirely on his claim about violation of assumptions underlying signal detection measures of sensitivity; it is also based on the demonstration that an independence model (the criterion bias model) is capable of simulating effects of context on sensitivity. This point is important because rather than using signal detection analysis to measure accuracy in the experiments reported by Rhodes et al. (1993) or in the experiments reported here, a measure of accuracy that is free of those assumptions could have been applied and would not be subject to the criticism raised by Norris (1995). An example of such a measure is the γ coefficient (Goodman & Kruskal, 1954). Nelson (1984, 1986) advocated a probabilistic interpretation of discrimination accuracy based on the gamma coefficient that does not depend on the assumptions underlying signal detection theory. Had the gamma coefficient been used to measure identification accuracy in the Rhodes et al. (1993) experiments or in the present experiments, there still would be doubt regarding the interpretation of any observed effect of context on accuracy. That doubt would stem from the fact that plausible independence models are capable of pro-

ducing context effects on accuracy without allowing context to affect perceptual encoding processes.

The critical aspect of the criterion bias and distributed memory models that enable those models to account for the effect of context on accuracy is the assumption the models make regarding which sources of information determine performance accuracy in word identification paradigms. In particular, if accuracy on a word identification task is influenced by conceptual information, sensitivity measures may be affected by semantic priming, even under an independence model. The simulation results found with the criterion bias and the distributed memory models clearly demonstrate this point.

The common ground for these two models is the fact that both simulate semantic priming effects by combining contextual information with information derived from the target stimulus in determining the model's response. In the criterion bias model, a prime reduces the response threshold for related logogens, which is equivalent to increasing the degree of evidence supporting each of those logogens. Thus, stimulus information and contextual information are combined to make it more likely that a primed logogen will achieve threshold activation.

In the distributed memory model, stimulus information is encoded as a pattern of activation in the orthographic units, whereas contextual information affects the pattern of activation in the semantic units. These two patterns are combined, along with the phonological pattern of activation, to determine a familiarity value for the stimulus (i.e., the energy measure) or the similarity between a target and a probe. Including information from the semantic units in these evaluations provides an additional source of information about the target because processing the target stimulus also affects the semantic units. In the case of a related prime, the target's influence on semantic units meshes with the pattern generated by the related prime to produce a pattern of activation in the semantic units that is closer to the target's meaning than is the case when an unrelated prime is used. This proximity enhances both the evaluated familiarity of the target and the similarity between the representation of the briefly presented target and a matching probe.

In either model, removal of the contextual contribution to stimulus evaluation would eliminate semantic priming effects. For the criterion bias model, if the effect of a prime on logogen criteria were removed, primed targets would no longer be more likely to reach threshold activation. Therefore, no priming effect would be obtained. Similarly, in the distributed memory model, if performance were determined only by the pattern of activation in the orthographic units, a semantically related prime would have no means of affecting performance because its relevant influence would be restricted to semantic and phonological units, and those units have no impact on the orthographic units. Again, no priming effect would be obtained. For both models, then, it is the combination of contextually induced information (criterion reduction in the criterion bias model and the pattern of activation in the semantic units of the distributed memory model) with information derived from perceptual processing of the target that produces the effect of priming on sensitivity.

Despite this similarity, the two models are distinguished by a number of important architectural features. Of particular importance, however, is the difference between the models regarding how context affects stimulus identification. This difference was highlighted by the results of Experiment 3, in which line drawings were used as primes. The only way the criterion bias model could produce a priming effect on sensitivity in word identification under those conditions would be to assume the existence of a top-down influence of semantic codes, activated by picture input, on logogens in the visual word input lexicon. A modification of this form would violate the modular nature of the criterion bias model. On the other hand, unlike the criterion bias model, no top-down influence of semantic units on orthographic units was required for priming to affect sensitivity in any of the simulations involving the distributed memory model. Thus, the distributed memory model stands as a clear example of an architecture in which acquisition of stimulus information remains fully independent of contextual information.

Plausibility of the Distributed Memory Model

The argument we have presented regarding the interpretation of context effects on signal detection measures of sensitivity depends in part on the plausibility of the independence models that we used to simulate those effects. Although we have no stake in the criterion bias model used by Norris (1995), we offer two points regarding the plausibility of the distributed memory model. First, the distributed memory model has been successfully applied to a number of word identification tasks. In addition to simulating accuracy and bias results of the masked lexical decision and probe-matching tasks used in the experiments reported here, the model also accounts for semantic priming and semantic ambiguity effects on response latency in word-naming and lexical decision tasks (Borowsky & Masson, 1996; Joordens, Masson, & Besner, 1995; Masson, 1989, 1991b, 1995).

Second, the distributed memory model's simulation of the probe-matching task naturally produced performance-duration functions predicted by a multiplicative model of priming effects. When accuracy was measured using A', accuracy functions for related and unrelated priming conditions were horizontally parallel when plotted against log target duration. Using a logarithmic transformation of hit rate minus false alarm rate as the performance measure produced performance functions that were linear when plotted against target duration, with a steeper slope for the related-prime condition. These performance-duration functions have been obtained in other paradigms involving priming effects on object identification (Reinitz et al., 1989) and word identification (Reinitz & Alexander, 1996). The successful application of the distributed memory model to this rather wide range of paradigms and analytic techniques provides evidence for its plausibility.

Multiplicative Model Revisited

Before leaving the issue of performance-duration functions, we wish to note an important parallel between our application of the distributed memory model to results based on signal detection theory and its application to results analyzed using a multiplicative model of priming effects. Although Reinitz and his colleagues (Reinitz & Alexander, 1996; Reinitz et al., 1989) interpreted certain patterns of performance-duration functions as evidence that priming increased by a constant factor the rate at which stimulus information was acquired from a stimulus, our simulation of those patterns with an independence model suggests that priming may have had no affect at all on stimulus information acquisition. Rather, the increase in information acquisition could have been due the integration of independently acquired stimulus and contextual information.

Just as with signal detection analyses of performance, the distributed memory model's simulation of results that were predicted by the multiplicative model do not impugn the validity of the model underlying the analytic technique. Instead, the simulations offer a different interpretation of the nature of the information that accrues at a higher rate when a related prime is used. As with the relation between signal detection theory and sensitivity effects, the generic multiplicative model is neutral with respect to the question of whether stimulus information acquisition is affected when slopes of performance-duration functions are increased by priming. Our general cautionary point, reified by the simulation results from the distributed memory model, is that if task performance is influenced by contextual information, sensitivity effects and slope effects cannot necessarily be ascribed to changes in perceptual encoding.

Tests of Perceptual Encoding Effects

The analysis we have provided indicates that tasks such as word identification are not good candidates for demonstrating clear effects of context on perceptual encoding. The problem is that accuracy in these tasks is likely to be affected directly by the availability of relevant semantic information, without any influence necessarily being exerted on perceptual encoding. A solution to this problem is to develop tasks in which contextual information cannot make a direct contribution to performance accuracy.

We know of four studies that have taken this approach. First, Krueger and Shapiro (1979) presented subjects lists of words or nonwords using rapid serial visual presentation of the lists. The task was to determine whether a mutilated as opposed to a normal letter A was present in the list. Krueger and Shapiro reasoned that if word context affects perceptual encoding of letters, accuracy at detecting the presence of mutilation should be higher for word lists than for nonword lists. Although they found a response bias against reporting mutilation in word lists, indicating that words and nonwords were processed differently, accuracy at detecting mutilation as

measured by signal detection sensitivity was no higher for word lists than for nonword lists. The lack of an effect of word context on mutilation detection accuracy is consistent with independence models. According to those models, context does not have a direct effect on perceptual encoding. Moreover, accuracy in the mutilation detection task depends entirely on stimulus information, so there is no basis for a context effect on accuracy in this task.

In a second study, Masson (1991a) reported an object discrimination task in which briefly presented objects were preceded by a word prime that named the target object or some other irrelevant object. The target object was followed by a probe object that was identical to it or a variant of it. For example, the context word *telephone* was followed by a drawing of a push-button telephone, and then a probe drawing of either an identical telephone or a rotary dial telephone was presented. The task was to determine whether the probe drawing was identical in form to the target drawing. As in the Krueger and Shapiro (1979) study, contextual information can make no direct contribution to accuracy in this task. By independence models, then, accuracy in this object discrimination task should not be affected by context. It was found that priming the target drawings with the appropriate object name had no effect on accuracy.

Third, Reinitz et al. (1989; Experiments 2–4) primed object and scene identification and later required subjects to discriminate between previously presented target items and visually similar members from the same categories as the targets. Performance on this memory test was enhanced if the earlier presentation of a target item was primed with the item's appropriate category name. This finding suggests, in contrast to the results obtained by Masson (1991a), that priming object or scene identification can improve perceptual encoding. As noted, however, this conclusion is compromised by the possibility that related and unrelated primes may have differentially affected subjects' rehearsal of the target items, thereby producing a priming effect on the subsequent memory test without necessarily influencing the initial acquisition of visual information.

Finally, Reingold and Jolicoeur (1993) tested the effects of word context on letter identification by embedding critical letter fragments in letter or nonletter contexts. Subjects more accurately identified the color of a critical letter fragment when it appeared in a letter context than when it appeared in a nonletter context. This result was taken as evidence for a top-down influence of context on perceptual processing. Although this demonstration did not involve semantic context of the form examined in the priming studies reviewed here, it constitutes an important example of a task in which it is unlikely that performance accuracy will be directly influenced by semantic information. Contextual influences on accuracy in tasks such as this would seem to be a compelling form of evidence in favor of the view that semantic context has a top-down influence on perceptual encoding. In our

view, pursuing the influence of context on this class of tasks holds substantial promise for settling the question of whether perceptual encoding can be altered by relevant contextual information.

REFERENCES

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

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 & Performance*, **10**, 340-357.

Becker, C. A., & Killion, T. H. (1977). Interaction of visual and cognitive effects in word recognition. *Journal of Experimental Psychology: Human Perception & Performance*, **3**, 389-401.

Besner, D. (1983). Basic decoding components in reading: Two dissociable feature extraction processes. *Canadian Journal of Psychology*, **37**, 429-438.

Besner, D., & Johnston, J. C. (1989). Reading and the mental lexicon: On the uptake of visual information. In W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 291-316). Cambridge, MA: MIT Press.

Besner, D., & Smith, M. C. (1992). Models of visual word recognition: When obscuring the stimulus yields a clearer view. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 468-482.

Borowsky, R., & Besner, D. (1991). Visual word recognition across orthographies: On the interaction between context and degradation. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **17**, 272-276.

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

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

Bub, D. N., & Arguin, M. (1995). Visual word activation in pure alexia. *Brain & Language*, **49**, 77-103.

Busey, T. A., & Loftus, G. R. (1994). Sensory and cognitive components of visual information acquisition. *Psychological Review*, **101**, 446-469.

Clarke, R., & Morton, J. (1983). Cross modality facilitation in tachistoscopic word recognition. *Quarterly Journal of Experimental Psychology*, **35A**, 79-96.

Connine, C. M., & Clifton, C. (1987). Interactive use of lexical information in speech perception. *Journal of Experimental Psychology: Human Perception & Performance*, **13**, 291-299.

Donaldson, W. (1992). Measuring recognition memory. *Journal of Experimental Psychology: General*, **121**, 275-277.

Erber, N. (1969). Interaction of audition and vision in the recognition of oral speech stimuli. *Journal of Speech & Hearing Research*, **12**, 423-425.

Farah, M. (1989). Semantic and perceptual priming: How similar are the underlying mechanisms? *Journal of Experimental Psychology: Human Perception & Performance*, **15**, 188-194.

Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Forster, K. I. (1979). Levels of processing and the structure of the language processor. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 27-85). Cambridge, MA: MIT Press.

Forster, K. I. (1981). Priming and the effects of sentence and lexical contexts on naming time: Evidence for autonomous lexical processing. *Quarterly Journal of Experimental Psychology*, **33A**, 465-495.

Goodman, L. A., & Kruskal, W. H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association*, **49**, 732-764.

Grier, J. B. (1971). Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin*, **75**, 424-429.

Hale, B. L., & Johnston, J. C. (1983, November). *How sentence context affects word identification: A signal detection theory analysis*. Paper presented at the annual meeting of the Psychonomic Society, San Diego.

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

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

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

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

Johnston, J. C., & Hale, B. L. (1984). The influence of prior context on word identification: Bias and sensitivity effects. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X* (pp. 243-255). Hillsdale, NJ: Erlbaum.

Joordens, S., Masson, M. E. J., & Besner, D. (1995, November). *Connectionist models and additive effects: Are distinct stages of processing necessary?* Paper presented at the annual meeting of the Psychonomic Society, Los Angeles.

Krueger, L. E., & Shapiro, R. G. (1979). Letter detection with rapid serial visual presentation: Evidence against word superiority at feature extraction. *Journal of Experimental Psychology: Human Perception & Performance*, **5**, 657-673.

Loftus, G. R., Busey, T. A., & Senders, J. W. (1993). Providing a sensory basis for models of visual information acquisition. *Perception & Psychophysics*, **54**, 535-554.

Loftus, G. R., Duncan, J., & Gehrig, P. (1992). On the time course of perceptual information that results from a brief visual presentation. *Journal of Experimental Psychology: Human Perception & Performance*, **18**, 530-549.

Loftus, G. R., & Ruthruff, E. (1994). A theory of visual information acquisition and visual memory with special application to intensity-duration trade-offs. *Journal of Experimental Psychology: Human Perception & Performance*, **20**, 33-49.

Massaro, D. W. (1979). Letter information and orthographic context in word perception. *Journal of Experimental Psychology: Human Perception & Performance*, **5**, 595-609.

Massaro, D. W. (1987). *Speech by ear and eye: A paradigm for psychological inquiry*. Hillsdale, NJ: Erlbaum.

Massaro, D. W. (1989). Testing between the TRACE model and the fuzzy logical model of speech perception. *Cognitive Psychology*, **21**, 398-421.

Massaro, D. W., & Cohen, M. M. (1990). Perception of synthesized audible and visible speech. *Psychological Science*, **1**, 55-63.

Massaro, D. W., & Cohen, M. M. (1991). Integration versus interactive activation: The joint influence of stimulus and context in perception. *Cognitive Psychology*, **23**, 558-614.

Massaro, D. W., & Cowan, N. (1993). Information processing models: Microscopes of the mind. *Annual Review of Psychology*, **44**, 383-425.

Massaro, D. W., & Oden, G. C. (1995). Independence of lexical context and phonological information in speech perception. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 1053-1064.

Massaro, D. W., Weldon, M. S., & Kitzis, S. N. (1991). Integration of orthographic and semantic information in memory retrieval. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **17**, 277-287.

Masson, M. E. J. (1988). The integration of sentence context and perceptual analysis in word identification. *Memory & Cognition*, **16**, 489-496.

Masson, M. E. J. (1989). Lexical ambiguity resolution in a constraint satisfaction network. In G. M. Olson & E. E. Smith (Eds.), *Proceedings of the Eleventh Annual Conference of the Cognitive Science Society* (pp. 757-764). Hillsdale, NJ: Erlbaum.

Masson, M. E. J. (1991a). Constraints on the interaction between context and stimulus information. In K. J. Hammond & D. G. Gentner (Eds.), *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society* (pp. 540-545). Hillsdale, NJ: Erlbaum.

MASSON, M. E. J. (1991b). A distributed memory model of context effects in word identification. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 233-263). Hillsdale, NJ: Erlbaum.

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

MASSON, M. E. J., & BOROWSKY, R. (1995). Unsettling questions about semantic ambiguity in connectionist models: Comment on Joordens and Besner (1994). *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 509-514.

MASSON, M. E. J., & MACLEOD, C. M. (1996). Contributions of processing fluency to repetition effects in masked word identification. *Canadian Journal of Experimental Psychology*, **50**, 9-21.

MCCLELLAND, J. L. (1991). Stochastic interactive processes and the effect of context on perception. *Cognitive Psychology*, **23**, 1-44.

MCCLELLAND, J. L., & ELMAN, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, **18**, 1-86.

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

MCGURK, H., & MACDONALD, J. (1976). Hearing lips and seeing voices. *Nature*, **264**, 746-748.

MCNICHOL, D. (1972). *A primer of signal detection theory.* London: Allen & Unwin.

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.

MILLER, G. A., & ISARD, S. (1963). Some perceptual consequences of linguistic rules. *Journal of Verbal Learning & Verbal Behavior*, **2**, 217-228.

MORTON, J. (1969). Interaction of information in word recognition. *Psychology Review*, **76**, 165-178.

MORTON, J. (1979). Facilitation in word recognition: Experiments causing change in the logogen model. In P. A. Kolers, M. E. Wrolstad, & H. Bouma (Eds.), *Processing of visible language* (Vol. 1, pp. 259-268). New York: Plenum.

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.

NELSON, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, **95**, 109-133.

NELSON, T. O. (1986). ROC curves and measures of discrimination accuracy: A reply to Swets. *Psychological Bulletin*, **100**, 128-132.

NORRIS, D. (1986). Word recognition: Context effects without priming. *Cognition*, **22**, 93-136.

NORRIS, D. (1995). Signal detection theory and modularity: On being sensitive to the power of bias models of semantic priming. *Journal of Experimental Psychology: Human Perception & Performance*, **21**, 935-939.

ODEN, G. C., & MASSARO, D. W. (1978). Integration of featural information in speech perception. *Psychological Review*, **85**, 172-191.

PAUL, S. T., KELLAS, G., & JUOLA, J. F. (1992). Priming effects on lexical access and decision processes: A signal-detection analysis. *Psychological Research*, **54**, 202-211.

PITT, M. A. (1995). The locus of the lexical shift in phoneme identification. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 1037-1052.

PLAUT, D. C. (1995). Semantic and associative priming in a distributed attractor network. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 17, pp. 37-42). Hillsdale, NJ: Erlbaum.

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.

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

PYLYSHYN, Z. (1980). Computation and cognition: Issues in the foundations of cognitive science. *Behavioral & Brain Sciences*, **3**, 111-132.

REICHER, G. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, **81**, 275-280.

REINGOLD, E. M., & JOLICOEUR, P. (1993). Perceptual versus postperceptual mediation of visual context effects: Evidence from the letter-superiority effect. *Perception & Psychophysics*, **53**, 166-178.

REINITZ, M. T., & ALEXANDER, R. (1996). Mechanisms of facilitation in primed perceptual identification. *Memory & Cognition*, **24**, 129-135.

REINITZ, M. T., WRIGHT, E., & LOFTUS, G. R. (1989). Effects of semantic priming on visual encoding of pictures. *Journal of Experimental Psychology: General*, **118**, 280-297.

RHODES, G., PARKIN, A. J., & TREMEWAN, T. (1993). Semantic priming and sensitivity in lexical decision. *Journal of Experimental Psychology: Human Perception & Performance*, **19**, 154-165.

RHODES, G., & TREMEWAN, T. (1993). The Simon then Garfunkel effect: Semantic priming, sensitivity, and the modularity of face recognition. *Cognitive Psychology*, **25**, 147-187.

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., SANDERS, M., & LANGER, P. (1984). Pre- and postlexical loci of contextual effects on word recognition. *Memory & Cognition*, **12**, 315-328.

SMITH, M. C., & MAGEE, L. E. (1980). Tracing the time course of picture-word processing. *Journal of Experimental Psychology: General*, **109**, 373-392.

SNODGRASS, J. G., & CORWIN, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, **117**, 34-50.

SNODGRASS, J. G., & VANDERWART, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning & Memory*, **6**, 174-215.

STANOVICH, K. E., & WEST, R. F. (1979). Mechanisms of sentence context effects in reading: Automatic activation and conscious attention. *Memory & Cognition*, **7**, 77-85.

STERNBERG, S. (1969). The discovery of processing stages: Extensions of Donders' method. In W. G. Koster (Ed.), *Attention and performance II* (pp. 276-315). Amsterdam: North-Holland.

SUMBY, W. H., & POLLACK, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, **26**, 212-215.

TANNER, W. P., & SWETS, J. A. (1954). A decision-making theory of visual detection. *Psychological Review*, **61**, 401-409.

THEIOS, J., & AMRHEIN, P. C. (1989). Theoretical analysis of the cognitive processing of lexical and pictorial stimuli: Reading, naming, and visual and conceptual comparisons. *Psychological Review*, **96**, 5-24.

TULVING, E., & GOLD, C. (1963). Stimulus information and contextual information as determinants of tachistoscopic recognition of words. *Journal of Experimental Psychology*, **66**, 319-327.

VITKOVITCH, M., & HUMPHREYS, G. W. (1991). Perseverant responding in speeded naming of pictures: It's in the links. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **17**, 664-680.

WHEELER, D. D. (1970). Processes in word recognition. *Cognitive Psychology*, **1**, 59-85.

NOTES

1. This feature of the model's architecture differs from that presented in Figure 2 of Masson (1995). Although the depiction of the model in that earlier figure shows connections from semantics and phonology to orthography, indicating the potential for a top-down influence on orthography, those connections were never implemented in the model.

2. We thank Dennis Norris for his suggestions regarding how the criterion bias model might be used to simulate the probe-matching task.

APPENDIX
Details of Simulations Using the Distributed Memory Model and the Criterion Bias Model

Distributed Memory Model

The distributed memory model is a Hopfield (1982) network that consists of three sets of processing units, one set representing orthography, another representing phonology, and another representing semantics. A *word* is defined as a pattern of activation across the entire collection of units. Each unit is binary valued, taking on either 1 or -1 as its activation state.

Learning and word identification. A word is learned by modifying connection weights between these units according to a learning rule derived from Hebb (1949). The connection weight between each pair of units is increased if the two units are in the same state when the word's pattern is instantiated across the network and decreased if they are in opposite states. These changes in connection weights are defined as

$$\Delta w_{ij} = n_i n_j,$$

where w_{ij} represents the connection weight between units i and j, and n_i and n_j represent the activation states (1 or -1) of units i and j when the learned pattern is instantiated. Across the entire set of patterns that the network learns, this rule allows the connection weights to capture the correlations between processing units (i.e., the tendency for two units to be in the same or different states across the full set of items).

Word identification is characterized as the instantiation of a word's pattern of activation across the network. This was accomplished by setting the network to a random pattern of activation, and then randomly sampling units without replacement for updating. Each sampling event is called an "updating cycle." Orthographic units were sampled at a higher rate than phonological units, which, in turn, were sampled at a higher rate than semantic units (see Borowsky & Masson, 1996, and Masson, 1995, for the rationale for this differential sampling rate). Sampling of units continued until a criterion was reached, as determined by specific task requirements. The process of updating was different for orthographic units than for the other units. Each time an orthographic unit was sampled for updating, it was set to the state dictated by the orthographic pattern of the target word. For all other units, activation coming into a unit was computed as

$$a_i = \sum_{i \neq j} w_{ij} n_j,$$

where a_i represents the amount of activation directed to unit i. A unit's activation state was determined by a threshold function: if $a_i > 0$ then $n_i = 1$, else $n_i = -1$.

This updating procedure moves the network through the space of possible patterns of activation and into a basin of attraction that, ideally, corresponds to the target word's learned pattern of activation. The procedure is a form of gradient descent that minimizes an energy function, E, defined as

$$E = -\sum_{i < j} w_{ij} n_i n_j.$$

The local minima of this function are basins of attraction that correspond to learned patterns of activation in the multidimensional space that represents all patterns of activation. Energy decreases monotonically as a function of the number of updating cycles and stabilizes when a local minimum is achieved (i.e., a word's pattern of activation has been instantiated). The form of the energy function is such that larger negative values occur as pairs of units take on values that are consistent with the connection weights between them (e.g., units with a strong positive connection weight are in the same state). Thus, energy can be taken as a measure of the network's progress toward a known state.

Simulation of context effects. The simulations reported here were based on the same architecture and procedures as those described by Masson (1995), with exceptions noted at appropriate points. The network consisted of 130 orthographic units, 40 phonological units, and 80 semantic units. Each unit was connected with every other unit in the network. Because of the way orthographic units were updated, they were functionally modular. The states of orthographic units were entirely determined by visual input, whereas the phonological and meaning units were influenced by every unit in the network.

During a word identification trial, the probability of sampling a unit from the orthographic module for updating was .7. If an orthographic unit was not selected, then a unit from the phonological module was sampled for updating with a probability of .7. Finally, if a phonological unit was not selected, then a unit from the semantic module was sampled. These updating probabilities are in keeping with the relative probabilities used in previous simulations (Borowsky & Masson, 1996).

For each experiment, performance of a single subject was simulated by training the network on a small vocabulary of five word triplets. The words in each triplet were semantically related, where relatedness was defined as having patterns of activation across semantic units that were identical in 52 of the 80 units. This degree of similarity is close to that used by Masson (1995), and was selected to produce simulated semantic context effects that were similar in magnitude to those observed in Experiments 1–3. Except for the constraint on semantically similar patterns, the patterns of activation for each word were randomly determined. Unrelated patterns would be identical, on average, in half of the units. The network was trained by setting all connection weights to zero, and then presenting each word's pattern once for training with the Hebbian learning rule.

Word identification trials were simulated by setting all units to a random starting value of 1 or -1, and then updating the network with learning turned off. One word from each of the five triplets was used as a target and was tested 40 times in each context condition, resulting in 200 observations per condition for each simulation run. Fifty separate runs, each using an independently generated set of words, was conducted to simulate testing of 50 subjects. The exposure durations of stimuli were simulated by translating millisecond values into updating cycles. In the translation, two updating cycles corresponded to 1 msec. This timing assumption contrasts with the one used by Masson (1995), in which one updating cycle was associated with 2 msec. The reason for the change is that in the current simulations, the orthographic units update over time, rather than having the entire orthographic pattern fully instantiated instantaneously, as was done in the simulations reported by Masson (1995). This change meant that the system required many more updating cycles for the processing units to reach stable patterns of activation. About 800–1,000 cycles were required for the present simulations to fully instantiate orthographic and phonological patterns. In the simulations of naming latency reported by Masson (1995), settling of the phonological units (the prerequisite for a word naming response) required only about 200 cycles. Thus, the current simulations take about four times as many updating cycles to settle into known states. The translation between updating cycles and milliseconds is, therefore, consistent with the general approach taken by Masson (1995).

Lexical decision. Simulation of trials in the masked lexical decision task used in Experiment 1 involved presenting the model with the orthographic pattern for the prime word and allowing 1,600 update cycles, corresponding to the 800-msec stimulus onset asynchrony between the prime and the pre-mask. Input to the orthographic units was then randomized to represent presentation of a mask and updating continued for 100 cycles (50 msec). Finally, the target item's orthographic pattern was presented to the orthographic units and 66 update cycles were run, corresponding to the 33-msec target duration. In the case of nonword targets, an orthographic pattern derived from an unrelated word learned by the system was used. The nonword patterns differed from their base words in 26 of the 130 orthographic units. The high degree of similarity to learned words is consistent with the fact that the nonwords in the experiment were pronounceable and orthographically legal, but was determined mainly by an effort to make the model's accuracy in discriminating between words and nonwords similar to that observed in Experiment 1. After the target had been processed, energy was computed. If the energy of the network was less (i.e., a larger negative number) than the criterion of −14,100, the target was classified as a word; otherwise it was classified as a nonword. The criterion was set so that the model's hit rate would be similar to that observed in Experiment 1.

Word identification with word context. Trials in the probe-matching task of Experiment 2 were simulated in the same way as lexical decision trials except that after processing the target, the model was presented with a random orthographic pattern representing the post-mask and continued to update for 200 cycles (100 msec). At that point, the pattern of activation across the entire network was read into a buffer with a small perturbation; each processing unit's activation state was correctly read into the buffer with probability .97; otherwise the incorrect state was read in. The perturbation prevented the model from achieving performance accuracy that was too high. The probe word's pattern of activation was then loaded into the network and compared, unit by unit, with the contents of the buffer. In the related context condition, the degree of match between the probe's pattern and the buffer generally was higher than in the unrelated context condition, even when the probe and target did not match. Therefore, two different criteria, one for each context condition, were used. In the related condition, a "same" response was given if the probe's representation and the buffer agreed in at least 146 of the 250 processing units; otherwise a "different" response was given. In the unrelated condition, the criterion was 133. These criteria were set so that the model would produce hit rates similar to those observed in Experiment 2.

Word identification with line drawing context. The simulation of the results of Experiment 3 was the same as that of Experiment 2, with two important exceptions. First, the context items in Experiment 3 were line drawings, not words. Although five word triplets were learned by the network, as in the other simulations, the orthographic patterns of the words that served as context items were specially treated. During processing of a context item, the relevant orthographic pattern was presented to the network, but the orthographic units were allowed to influence only the semantic units, not the phonological units. This procedure produced exactly the same effect as constructing an additional set of 130 units to represent the features of a line drawing, and connecting those units only to the semantic units in the network (as indicated in Figure 6). Second, while updating the network during presentation of the target word, when an orthographic unit was sampled, it was set to the appropriate state with probability .78; otherwise it was set to the incorrect state. This reduction in encoding accuracy was intended to represent the effect of lateral masking used in Experiment 3, which we assume was the cause of the lower accuracy in that experiment, relative to Experiment 2, despite the longer target exposure duration used in Experiment 3. Other changes reflected differences between Experiments 2 and 3 with respect to target and mask durations and different criteria for responding. The context duration was 1,590 cycles, the pre-mask and target durations were 90 cycles, and the post-mask duration was 180 cycles. The criterion for making a "same" response in the related condition was 148 matching units, and the criterion in the unrelated condition was 135 units.

Criterion Bias Model

A version of the criterion bias model described by Norris (1995) was implemented by creating a 500-word lexicon, with each word's orthographic pattern defined by a set of 20 orthographic features, randomly selected from a pool of 30 features. To simulate the brief presentation of a masked target, the feature pattern of the target word was degraded by randomly eliminating some of its features. The degraded pattern of the target word was taken as visual input, and the activation level of each word in the lexicon was determined by counting the number of features in common between that word and the visual input. If the number of features a lexical item shared with the visual input met the criterion of at least 13 out of 20, that item was deemed to have been activated. The decision about whether the target word matched the probe was based on the word or set of words that was activated by presentation of the target. If one of the activated words matched the probe, a positive response was made; if not, a negative response was made. If none of the words in the lexicon were activated by the target display, a positive response was made with probability .5. To simulate the effect of the prime, a randomly selected set of 20 words had their activation criterion reduced by one feature (as in the simulation reported by Norris, 1995). On related-prime trials, the target and probe words were included in the set of 20 primed words. A simulation of 100,000 trials in each condition defined by prime (related vs. unrelated) and probe (matching vs. non-matching) was conducted. On each trial, the model made either a positive or a negative response to the probe.