Blocking by word frequency and neighborhood density in visual word recognition: A task-specific response criteria account

MANUEL PEREA
Universitat de València, València, Spain

MANUEL CARREIRAS
Universidad de La Laguna, Tenerife, Spain

and

JONATHAN GRAINGER
CNRS and University of Provence, Aix-en-Provence, France

Effects of blocking words by frequency class (high vs. low) and neighborhood density (high vs. low) were examined in two experiments using progressive demasking and lexical decision tasks. The aim was to examine the predictions of a task-specific response criteria account of list-blocking effects. Distinct patterns of blocking effects were obtained in the two tasks. In the progressive demasking task, a pure-list disadvantage was obtained to low frequency–high density words, whereas high frequency–low density produced a trend toward a pure-list advantage. In lexical decision, high-frequency words showed a pure-list advantage that was strongest in high-density words, whereas low frequency–low density words produced a trend toward a pure-list disadvantage. A simulation study implementing task-specific response criteria within the framework of the multiple read-out model provided an accurate description of the blocking effects obtained in the experiments. It is argued that adjustments of task-specific response criteria determine changes in list-blocking effects across different tasks.

In the typical setup in a visual word recognition experiment, the participant is presented with a letter string and has to make an overt response (e.g., deciding whether a letter string is a word or not). The most important dependent variable is the time elapsed between the onset of the target stimulus and the beginning of the participant's response (i.e., the reaction time, RT). Although the many earlier experiments over the past three decades have provided valuable information on the underlying processes involved in accessing the mental lexicon, speeded RTs in any laboratory word identification task may be sensitive to a number of extrastimulus influences (e.g., Grainger, Spinelli, & Ferrand, 2000; Jared, 1997; Kello & Plaut, 2000; Lupker, Brown, & Colombo, 1997; Ratcliff, Gómez, & McKoon, 2004; Stone & Van Orden, 1993). To optimize their responses in a RT task (i.e., minimize the RTs while keeping a low rate of errors), participants may use different strategies depending on a number of factors, one of them being the characteristics of the experimental list. For instance, the criterion settings adopted by the participants may be quite different when the experimental list is composed of only familiar, "easy" stimuli than when the list is composed of "easy" and "difficult" stimuli: This would result in a blocking effect (i.e., a difference between the mean RT in the mixed list and the mean RT in the pure list). Pure versus mixed stimulus presentation is a methodological variable that can exercise considerable control over the setting of response criteria in studies of word processing (see Carr, Posner, Pollatsek, & Snyder, 1979). Furthermore, contrasting pure versus mixed lists is an ideal "strategy" manipulation (see Stone & Van Orden, 1993), because a given trial is identical in the mixed versus pure conditions (i.e., only the composition of the overall stimulus set differs). Understanding how participants can exercise strategic control over the reading process is undoubtedly important for developing more general models of visual word recognition and reading.

At a more general level of theorizing, a number of researchers (e.g., Kiger & Glass, 1981; Lupker et al., 1997; Strayer & Kramer, 1994a, 1994b; Taylor & Lupker, 2001; Treisman & Williams, 1984) have suggested that participants can somehow judge the average difficulty of performing a given task in a given context. For instance, Kiger and Glass found that, in a sentence verification task, RTs

were faster for a given item when placed in a context of easy items compared with a context of difficult items. Kiger and Glass pointed out that a list composed of difficult and easy trials may raise RTs as a result of a criterion shift. Participants who receive a list with a high proportion of difficult items will perceive their task as being harder than participants receiving lists with a high proportion of easy items. Participants with the difficult list might take more time, participants with the easy list, less time, by strategically adjusting the setting of response criteria necessary to perform the task. In pure lists, participants can choose an optimal strategy for each trial and condition. In mixed lists, participants cannot anticipate the nature of the following trial and thus must choose a common strategy for the experimental condition. Although the strategy may be optimal for the mixed list as a whole, it is likely to be suboptimal on a trial-by-trial basis.

The Frequency-Blocking Manipulation

In studies of visual word recognition, the variable most often associated with ease of performance is the frequency of occurrence of words in a representative corpus of written language: High-frequency words are identified faster and/or more accurately than low-frequency words. Further, in the lexical decision task (word/nonword discrimination), high-frequency words are responded to faster when the word list is composed of other high-frequency words (pure list) than when the word list is composed of both high- and low-frequency words (mixed list). In other words, the word frequency effect is modulated by the frequency properties of the other words in the stimulus list. The *frequency-blocking* effect was first observed by Glanzer and Ehrenreich (1979) using the lexical decision task (see also Dorfman & Glanzer, 1988; Gordon, 1983; Ratcliff et al., 2004; Stone & Van Orden, 1993). Recently, Lupker et al. (1997) have reported a similar effect in the naming task (but see Forster, 1981, for a failure to observe this pattern).

The frequency-blocking phenomenon in all its complexity poses problems for most models of visual word recognition (for review, see Dorfman & Glanzer, 1988; Grainger & Jacobs, 1996; Stone & Van Orden, 1993). Recently, a model has been proposed that appears to capture most aspects of frequency-blocking effects obtained with the lexical decision task. This is the multiple read-out model (MROM; Grainger & Jacobs, 1996). The MROM is an extension of the interactive activation model (McClelland & Rumelhart, 1981), which incorporates the design principle of multiple read-out: A response in a given experimental task is generated (read-out) when at least one of the codes that is appropriate for responding in that task reaches a critical activation level. In this model, a criterion set on activity in whole-word representations is used to trigger a positive word recognition response. This is the so-called M criterion. But the most distinctive feature of the model is the inclusion of two additional response criteria for the lexical decision task. One is the Σ decision criterion, which reflects general activity across

all word-detector units (for similar ideas, see Balota & Chumbley, 1984; Coltheart, Rastle, Perry, Ziegler, & Landon, 2001; Monsell, Doyle, & Haggard, 1989; Paap & Johansen, 1994). That is, participants may make a positive lexical decision response on the basis of global lexical activity before even identifying the letter string as a real word. For negative lexical decision responses, Grainger and Jacobs (1996) hypothesized that participants set a time limit (the T criterion) that is adjusted on each trial as a function of global lexical activation (see Coltheart, Davelaar, Jonasson, & Besner, 1977). When the M or the Σ criteria are not reached before the T criterion, participants make a "nonword" decision. (It is worth mentioning that these three response criteria have also been adopted in the dual-route cascaded [DRC] model; Coltheart et al., 2001; we address this model in the General Discussion.)

Grainger and Jacobs (1996) proposed that strategic modifications of the Σ and T criteria (as a function of list characteristics) were responsible for frequency-blocking effects obtained in the lexical decision task. As the proportion of high-frequency words increases, so does the proportion of words that generate critical levels of summed lexical activity in early stages of processing (simply because, other things being equal, high-frequency words tend to generate higher values of summed lexical activation than do low-frequency words). Thus, faster responses can be generated in the model through two criteria adjustments: (1) To reduce positive RTs, the Σ criterion for a positive response is lowered; (2) to reduce negative RTs, the T decision criterion is maintained as low as possible. By combining these two adjustments, the gain in positive RT does not necessarily provoke an increase in false positive errors, because the nonwords will have less time, on average, to reach the Σ criterion. Moreover, lower values of the T criterion in pure high-frequency lists lead to an increase in false negative errors, which is typically observed in experimentation (e.g., Dorfman & Glanzer, 1988; Glanzer & Ehrenreich, 1979; Gordon, 1983; Stone & Van Orden, 1993).

The Density-Blocking Manipulation

An interesting prediction of the MROM concerns what we will refer to as the density-blocking manipulation, where *density* refers to the number of orthographic neighbors of a given word target (also referred to as N; i.e., any words that can be created by changing one letter of the stimulus item, preserving letter positions; e.g., the word *spoon* has an N index of 4: *swoon*, *spook*, *spool*, and *spoor*; see Coltheart et al., 1977). The density-blocking effect (Grainger & Jacobs, 1996) indicates that the facilitative effects of N should increase as a function of the proportion of high-N words in the experimental list (with word frequency and type of nonword held constant). According to the MROM (Grainger & Jacobs, 1996), blocking the word stimuli by neighborhood density encourages participants to adjust the Σ response criterion. With only high-density words in the list—for a given set of non-

words with average levels of neighborhood density—the Σ criterion can be lowered, thus producing a global decrease in RTs to word stimuli. Recently, we (Carreiras, Perea, & Grainger, 1997) carried out a series of experiments including a density-blocking manipulation intended to exaggerate the facilitative effects of number of orthographic neighbors in the lexical decision task. Within the framework of the MROM, it was predicted that blocking word stimuli by neighborhood density (high or low) while having both high- and low-density nonwords in each block should allow increased use of the Σ criterion in the blocked high-density condition. This should enhance the facilitative effects of neighborhood density on correct responses to word stimuli in the lexical decision task, while increasing the number of false positive errors to high-density nonword stimuli, as was actually observed.

Description of the Experiments

In the present study, we test a general account of list-composition effects on visual word recognition. In the framework of the MROM (Grainger & Jacobs, 1996), list-composition influences task-specific response criteria. According to our approach, modifications in task-specific response criteria will account for list-composition effects, so different patterns should be obtained in different tasks. (We present a simulation study using the general framework of the MROM in the General Discussion.) To that end, in the present experiments we manipulated word frequency (high vs. low), neighborhood density (high vs. low), and the composition of the lists (pure or mixed). Prior research on the frequency-blocking phenomenon has often used between-subjects rather than within-subjects designs (Glanzer & Ehrenreich, 1979; Gordon, 1983; Ratcliff et al., 2004; Stone & Van Orden, 1993; but see Dorfman & Glanzer, 1988; Forster, 1981). However, using different participants for the pure and the mixed lists reduces statistical power (given the variability across participants). It seems reasonable to assume that participants change their criterion settings during the experiment, once they have seen several stimuli of the same type (see Balota & Chumbley, 1984; Forster, 1981; Gordon, 1983; Grainger & Jacobs, 1996; Lima & Huntsman, 1997; Lupker et al., 1997; McNamara, 1992; Perea & Carreiras, 2003; Plaut, 1997; Strayer & Kramer, 1994a, 1994b; Taylor & Lupker, 2001). Some of these adjustments are likely to occur on a trial-by-trial basis (see Lima & Huntsman, 1997; Perea & Carreiras, 2003; Taylor & Lupker, 2001). In the present series of experiments, a within-subjects design was used to test the presence of blocking effects, thereby increasing power considerably relative to a between-subjects design. (Of course, this option has the potential drawback that the number of items per condition is less than in the case of a between-subjects manipulation, but the final balance is likely to be favorable to the within-subjects manipulation.) The procedure employed in the experiments is very much similar to that used by Forster (1981, Experiment 3).

In addition, we used a multitask approach. The same word stimuli were used in two different paradigms: speeded identification (the progressive demasking task), and word–nonword classification (the lexical decision task). Experiment 1 explored the frequency- and density-blocking effect with the progressive demasking task (Carreiras et al., 1997; Grainger & Jacobs, 1996; Grainger & Segui, 1990), whereas Experiment 2 explored the frequency- and density-blocking effect with the lexical decision task.

EXPERIMENT 1
Progressive Demasking Task

In the progressive demasking task, the presentation of a target word is alternated with that of a mask. On each alternation cycle, the target presentation time increases slowly while that of the mask decreases (i.e., the signal-to-noise ratio increases). The participant's task is to press a button as soon as the target word is identified. Compared with other tasks (e.g., lexical decision), progressive demasking reduces the rate of presenting the sensory information to the participant (i.e., the average RT, as measured from the onset of the masking process, usually falls between 1 and 2 sec), thus effectively slowing the word identification process. Previous work using progressive demasking and related techniques has demonstrated the sensitivity of the paradigm to the influence of orthographic neighborhood on word recognition (neighborhood density: Carreiras et al., 1997; Grainger & Jacobs, 1996; Grainger & Segui, 1990; neighborhood size: Carreiras et al., 1997; Snodgrass & Mintzer, 1993; van Heuven, Dijkstra, & Grainger, 1998) as well as various other aspects of lexical processing (see Dijkstra, Grainger, & van Heuven, 1999; Schreuder & Baayen, 1997).

Method

Participants. Forty psychology students from the Universidad de La Laguna took part in this experiment to fulfill a course requirement. All of them had normal or corrected-to-normal vision and were native speakers of Spanish.

Design and Materials. A set of 96 disyllabic Spanish words were selected from the Spanish word pool (Alameda & Cuetos, 1995) as a function of word frequency (low vs. high) and neighborhood density (high vs. low; i.e., words with many neighbors vs. words with few neighbors, Coltheart et al., 1977). All words were four or five letters long (48 of four letters and 48 of five letters). A word was considered high frequency when its written frequency was higher than 42 per 1 million words, whereas a word was considered low frequency when its printed frequency was less than 8 per 1 million words. High-density words had at least eight orthographic neighbors, and low-density words had fewer than four orthographic neighbors. In order to control for effects of neighborhood frequency (see Grainger, O'Regan, Jacobs, & Segui, 1989) within each frequency class, none of the high-frequency words had any higher frequency neighbors, whereas all the low-frequency words had at least one higher frequency neighbor. The characteristics of the words used in the experiment are presented in Table 1. The materials are presented in the Appendix.

Two groups of 20 participants were tested. For each participant, there was one list for each of the following conditions: (1) high frequency–low density; (2) high frequency–low density; (3) low frequency–high density; (4) low frequency–low density. In addition to the four pure lists, there were four mixed lists composed of an equal number of words of the four categories. Assignment of words to conditions was arranged so that each word occurred both in a

Table 1
Characteristics of the Words in Experiments 1–3

Length	M-WF	Range-WF	M-N	Range-N	M-Length	Range-Length
HF-LD	159.5	44–396	1.3	0–3	4.5	4–5
HF-HD	169.5	57–433	11.2	8–19	4.5	4–5
LF-LD	3.2	1–7	1.7	1–3	4.5	4–5
LF-HD	3.5	1–7	10.9	8–16	4.5	4–5

Note—M-WF, mean frequency of words based on a count of 1 million Spanish words; M-N, average number of neighbors; M-Length, average number of letters; HF, high frequency; LD, low density; HD, high density; LF, low frequency.

pure list and in a mixed list, but not for the same participant. Thus, if a word occurred in a pure list for Group 1, it occurred in a mixed list for Group 2, and vice versa. Each experimental list contained 12 words. In order to establish a context for the words in each experimental block, four warm-up trials with the same characteristics as those in the block were also included. In this way, the participants would be familiar with the average difficulty of stimuli in each block. These warm-up trials were not considered in the analysis of the data. The words were presented in blocks of 12 items throughout (not including the warm-up trials). The eight experimental lists (four pure and four mixed) were presented in a fixed semi-random order to all participants in one group (the other group receiving the reverse sequence), but within each list, a different random ordering of the 12 items occurred for each participant.[1]

Procedure. Word stimuli were presented in alternation with a pattern mask. Each presentation cycle was composed of a given stimulus word followed immediately by a pattern mask of seven hash marks (#######). On each successive cycle, the presentation of the stimulus was increased by 14 msec and the presentation of the mask decreased by 14 msec. The total duration of each cycle remained constant at 300 msec. Each trial consisted of a succession of cycles wherein stimulus presentation increased and mask presentation decreased. On the first cycle of each trial, stimuli were presented for 14 msec and the mask for 286 msec. On the second cycle, stimuli were presented for 28 msec and the mask for 272 msec, and so on. There was no interval between cycles. This succession of cycles continued until the participant pressed a response key on the computer keyboard to indicate that he/she had recognized the stimulus word. Response latencies were measured from the beginning of the first cycle until the participant's response. Participants were instructed to focus their attention on the center of the visual display and to press the response key with the forefinger of their preferred hand as soon as they had recognized the word. They were instructed to type in the identified word using the keyboard of the computer. Pressing the return key initiated the following trial. Participants were asked to carefully check that they had correctly typed the word they thought they had been presented before initiating the following trial. Each participant received a total of 160 trials across eight blocks (20 words per block, of which the first 4 were filler trials that were not included in the data analysis). The whole session lasted approximately 15 min.

Results

Incorrect responses or identification times greater than 3,500 msec (1.09%) were excluded from the latency analysis. Mean RTs on words were submitted to an analysis of variance (ANOVA), with word frequency (high vs. low), neighborhood density (high vs. low), type of list (pure vs. mixed), and group (List 1 vs. List 2) as factors. Group was included in the analysis to extract the variance due to the lists (Pollatsek & Well, 1995). Group was the only non–repeated measures factor in the analysis by participants (F_1). In the analysis by items (F_2), word frequency, neighborhood density, and group were non–repeated measures factors, whereas type of list was a repeated measures factor. The mean identification time and the error rate on the stimulus words in each experimental condition are displayed in Table 2.

The ANOVA on the latency data showed that the main effect of word frequency was significant [$F_1(1,38) = 196.36$, $MS_e = 33,854$, $p < .001$; $F_2(1,88) = 117.649$, $MS_e = 34,801$, $p < .001$]. The inhibitory effect of neighborhood density was significant in the analysis by participants [$F_1(1,38) = 11.17$, $MS_e = 19,961$, $p < .002$; $F_2(1,88) = 3.01$, $MS_e = 34,891$, $p < .09$]. The main effect of type of list was not significant (both $Fs < 1$). The interaction between word frequency and neighborhood density was significant in the analysis by participants [$F_1(1,38) = 5.66$, $MS_e = 11,043$, $p < .025$; $F_2(1,88) = 1.21$]: There was an inhibitory effect of neighborhood density for high-frequency words [$F_1(1,38) = 32.09$, $MS_e = 8,125$, $p < .001$; $F_2(1,88) = 4.02$, $MS_e = 34,891$, $p < .05$], but not for low-frequency words [$F_1(1,38) = 1.08$, $F_2 < 1$].

The interaction between word frequency and type of list was significant in the analysis by participants [$F_1(1,38) = 7.28$, $MS_e = 15,051$, $p < .02$; $F_2(1,88) = 3.94$, $MS_e = 12,632$, $p = .050$]: Low-frequency words were responded to faster when embedded in the mixed list than when embedded in the pure list [$F_1(1,38) = 6.43$, $MS_e = 14,994$, $p < .02$; $F_2(1,88) = 3.24$, $MS_e = 12,632$, $p = .075$]. In contrast, there were no effects of type of list for high-frequency words [$F_1(1,38) = 1.80$; $F_2(1,88) = 1.01$]. Thus, the word frequency effect was larger in the pure lists than in the mixed lists (325 vs. 251 msec, respectively). The interaction between neighborhood density and type of list was also significant [$F_1(1,38) = 5.42$, $MS_e = 16,788$, $p < .03$; $F_2(1,88) = 5.79$, $MS_e = 12,632$, $p < .02$], which reflected a significant effect of list for high-density words [$F_1(1,38) = 5.28$, $MS_e = 15,889$, $p < .03$; $F_2(1,88) = 4.40$, $MS_e = 12,632$, $p < .04$] but not for low-density words [$F_1(1,38) = 1.28$; $F_2(1,88) = 1.70$]. That interaction also reflected that an inhibitory effect of neighborhood density occurred in the pure list [$F_1(1,38) = 11.31$, $MS_e = 26,470$, $p < .003$; $F_2(1,88) = 11.83$, $MS_e = 14,934$, $p < .002$] but not in the mixed list [$F_1(1,38) = 1.41$; $F_2 < 1$]. The interaction between word frequency, neighborhood density, and type of list was not significant (both $Fs < 1$).

The ANOVA on the error data showed that the main effect of word frequency was significant [$F_1(1,38) = 17.42$, $MS_e = 27.7$, $p < .001$; $F_2(1,88) = 9.23$, $MS_e = 29.9$, $p < .004$]. The effect of neighborhood density was not significant (both $Fs < 1$). The interaction between word frequency and neighborhood density was not significant [$F_1(1,52) = 3.07$, $MS_e = 19.1$, $p = .0880$; $F_2(1,88) = 1.12$]. None of the interactions with type of list were significant.

Discussion

For high-frequency words, we found an inhibitory effect of neighborhood density (see also Carreiras et al., 1997; van Heuven et al., 1998). For low-frequency words,

Table 2
Mean Identification Time (in Milliseconds) and
Error Rates (PEs) for the Words in Experiment 1
(Progressive Demasking Task)

| | Type of List | | | | | |
| | Pure | | Mixed | | Mixed − Pure | |
Words	M	PE	M	PE	M	PE
HF-LD	1,565	1.1	1,612	1.6	47	0.5**
HF-HD	1,668	1.1	1,671	2.9	3	1.8
LF-LD	1,906	3.5	1,903	5.8	−3	2.3
LF-HD	1,977	3.5	1,882	3.7	−95	0.2*

Note—*$p < .05$ by participants and by items; **$p < .10$ by participants. HF, high frequency; LD, low density; HD, high density; LF, low frequency.

there was a robust 71-msec inhibitory effect of neighborhood density in the pure list. However, this inhibitory effect vanished in the mixed list, in which there was a nonsignificant 21-msec facilitative effect of neighborhood density. It is worth noting that a post hoc regression analysis on the item means for the low-frequency words revealed that there was an inhibitory cumulative effect of the number of higher frequency neighbors in the pure lists ($r = -.30$, $p < .05$) when the influence of the log of word frequency and the number of lower frequency neighbors were controlled for (see also Perea & Pollatsek, 1998; Pollatsek, Perea, & Binder, 1999, for evidence in a reading task). However, we failed to find a cumulative effect of the number of higher frequency neighbors in the mixed lists ($r = -.10$, n.s.), again when the influence of the log of word frequency and the number of lower frequency neighbors were controlled for. This particular result constitutes evidence against any type of guessing account of performance in the progressive demasking task according to which inhibitory effects of higher frequency orthographic neighbors arise from a bias to guess that such words are present on the basis of partial bottom-up information. If this were the case, one would expect the influence of higher frequency neighbors on the recognition of low-frequency words to be *stronger*—not weaker—in (mixed) lists containing high-frequency words.

We discuss the implications of these results for the task-specific response criteria account in the General Discussion, since it is the pattern of blocking effects across different tasks that puts this account to test.

EXPERIMENT 2
Lexical Decision Task

Method

Participants. Forty psychology students from the Universidad de La Laguna took part in this experiment to fulfill a course requirement. All of them either had normal or corrected-to-normal vision and were native speakers of Spanish. None of them had participated in the previous experiment.

Materials and Design. The design and the materials were the same as in Experiment 1, except that nonword trials were also presented. Ninety-six orthographically legal nonwords were constructed

for the purposes of the lexical decision task. These nonwords had been created by changing an interior letter from disyllabic Spanish words with five or six orthographic neighbors. Each experimental list contained 12 words and 12 nonwords. The 96 experimental nonwords were assigned at random to eight experimental lists (12 nonwords each). The lists of nonwords were counterbalanced in the same way as the lists of words. In order to establish a context for the words in each experimental block, eight warm-up trials (four words and four nonwords) with the same characteristics as those in the block were also included. In this way, the participants would be familiar with the average difficulty of stimuli in each block. These warm-up trials were not considered in the data analysis.

Procedure. Participants were tested individually in a quiet room. Presentation of stimuli and recording of latencies were controlled by a PC computer. Before the experiment, there were 10 practice trials. The stimuli were presented in blocks of 32 trials throughout, each block constituting a complete list. There were eight such blocks. The eight experimental blocks were presented in a fixed semi-random order to all participants in one group (the other group receiving the reverse sequence), but within each block, a different random ordering of items occurred for each participant, except for the constant positioning of the eight preliminary warm-up items in each list. On each trial, a ready signal (*) was presented for 500 msec on the center of the screen. Next, the letter string was presented centered until the participant's response. Participants were instructed to press one of two buttons on the keyboard ("l" for yes and "a" for no, which had the labels "yes" and "no" on them) to indicate whether the letter string was a legitimate Spanish word or not. This decision had to be made as quickly and as accurately as possible. After an intertrial interval of 1,500 msec, the next trial was presented. Each participant received a total of 256 trials across eight blocks (32 stimuli per block, 16 words and 16 nonwords, of which the first 8 were filler trials that were not included in the data analysis). The whole session lasted approximately 15 min.

Results

Incorrect responses or RTs less than 300 msec or greater than 1,500 msec (0.85%) were excluded from the latency analysis. The mean lexical decision time and the error rate on the stimulus words and nonwords in each experimental condition are displayed in Table 3.

Analysis of words. Not surprisingly, high-frequency words were responded to more rapidly than low-frequency words [$F_1(1,38) = 208.45$, $MS_e = 5,340$, $p < .001$; $F_2(1,88) = 135.42$, $MS_e = 4,920$, $p < .001$]. In addition, high-density words were responded to more rapidly than low-density words, as revealed in the analysis by participants [$F_1(1,38) = 26.60$, $MS_e = 1,086$, $p < .001$; $F_2(1,88) = 3.47$, $MS_e = 4,920$, $p < .07$]. However, the neighborhood density effect must be interpreted in the light of the significant interaction between word frequency and neighborhood density [$F_1(1,38) = 43.16$, $MS_e = 1,878$, $p < .001$; $F_2(1,88) = 9.51$, $MS_e = 4,920$, $p < .03$]. The effect of neighborhood density was facilitative for low-frequency words [$F_1(1,38) = 46.58$, $MS_e = 2219$, $p < .001$; $F_2(1,88) = 12.23$, $MS_e = 4,920$, $p < .001$] but inhibitory for high-frequency words [$F_1(1,38) = 8.83$, $MS_e = 745$, $p < .006$; $F_2(1,88) < 1$].

Type of list interacted significantly with word frequency [$F_1(1,38) = 7.18$, $MS_e = 2,124$, $p < .02$; $F_2(1,88) = 8.67$, $MS_e = 936$, $p < .005$]. High-frequency words were responded to faster when embedded in the pure list than

Table 3
Mean Lexical Decision Time (in Milliseconds) and Error Rates (PEs) for the Words and Nonwords in Experiment 2 (Lexical Decision Task)

	Pure M	Pure PE	Mixed M	Mixed PE	Mixed − Pure M	Mixed − Pure PE
Words						
HF-LD	612	1.3	623	1.3	11	(0.0)*†
HF-HD	617	1.5	644	1.7	27	(0.2)***
LF-LD	780	11.7	754	14.0	−26	(2.3)†
LF-HD	713	3.5	721	5.4	9	(1.9)

	M	PE
Nonwords		
Pure HF-LD	746	5.6
Pure HF-HD	748	4.2
Pure LF-LD	800	4.8
Pure LF-HD	775	3.7
Mixed	776	5.7

Note—***$p < .001$ by participants and by items; *†$p < .10$ by participants and $p < .05$ by items; †$p < .10$ by participants. The nonword RTs refer to nonwords in particular lists; they are not derived from words with particular characteristics. HF, high frequency; LD, low density; HD, high density; LF, low frequency.

when embedded in the mixed list [$F_1(1,38) = 14.18$, $MS_e = 1,016, p < .001; F_2(1,88) = 10.36, MS_e = 936, p < .003$]. In contrast, there were no effects of type of list for low-frequency words [$F_1(1,38) = 1.02, F_2(1,88) < 1$]. That interaction also reflected that the word frequency effect was smaller in the mixed list (104 msec) than in the pure list (132 msec).

The interaction between neighborhood density and type of list was also significant [$F_1(1,38) = 7.33, MS_e = 1,666, p < .02; F_2(1,88) = 9.51, MS_e = 4,920, p < .003$]: High-density words were responded to faster in the pure lists than in the mixed lists [$F_1(1,38) = 9.74, MS_e = 1,261, p < .004; F_2(1,88) = 7.67, MS_e = 936, p < .008$], whereas the effect of list composition on low-density words was not significant (both $Fs < 1$). Finally, the three-way interaction between word frequency, neighborhood density, and type of list was not significant [$F_1(1,38) = 1.42; F_2(1,88) < 1$].

As in the latency analysis, an analysis of the error data showed significant main effects of word frequency [$F_1(1,38) = 53.37, MS_e = 78.6, p < .001; F_2(1,88) = 31.72, MS_e = 79.3, p < .001$] and neighborhood density [$F_1(1,38) = 27.67, MS_e = 46.5, p < .001; F_2(1,88) = 9.73, MS_e = 79.3, p < .003$]. The interaction between these two factors was also significant [$F_1(1,38) = 30.06, MS_e = 49.7, p < .001; F_2(1,88) = 11.31, MS_e = 79.3, p < .002$], which reflected facilitative effects of neighborhood density for low-frequency words [$F_1(1,38) = 32.02, MS_e = 86.8, p < .001; F_2(1,88) = 21.01, MS_e = 79.3, p < .001$] but not for high-frequency words (both $Fs < 1$). Neither word frequency nor neighborhood density interacted significantly with type of list.

Analysis of nonwords. The ANOVA on the latency data showed that the nonwords embedded in pure high-frequency lists were responded to faster than the non-

words embedded in pure low-frequency lists [$F_1(1,38) = 18.15, MS_e = 3,646, p < .001; F_2(1,91) = 7.29, MS_e = 4,984, p < .01$] or in mixed lists [$F_1(1,38) = 9.37, MS_e = 2,435, p < .005; F_2(1,91) = 4.43, MS_e = 4,984, p < .04$]. No significant differences were found between nonwords embedded in pure low-frequency lists and words embedded in mixed lists (both $ps > .10$). No differences were found with respect to the type of neighborhood either. The analysis of errors did not yield any reliable effects.

Discussion

First, we found the usual interaction between neighborhood density and word frequency; facilitative effects of neighborhood density were found for low-frequency words (see Andrews, 1989, 1992; Forster & Shen, 1996; Perea & Rosa, 2002; Perea, Rosa, & Gómez, 2003; Pollatsek et al., 1999; Sears, Hino, & Lupker, 1995; Siakaluk, Sears, & Lupker, 2002), whereas there was some inhibitory effect of neighborhood density for high-frequency words (although only in the analysis by participants; see Perea & Rosa, 2002; Perea et al., 2003, for a similar pattern of results). As usual, we found an overall interaction between word frequency and type of list: A frequency-blocking effect was found for high-frequency words but not for low-frequency words (e.g., see Dorfman & Glanzer, 1988; Glanzer & Ehrenreich, 1979; Gordon, 1983; Stone & Van Orden, 1993).[2]

As in Experiment 1, the conditions that yielded the strongest positive blocking effect (i.e., high frequency–high density and high frequency–low density) also yielded the fastest RTs in the pure list, whereas the condition that yielded the strongest negative blocking effect also yielded the slowest RTs in the pure list (high frequency–low density). Furthermore, the fact that the pattern of blocking effects changed completely from Experiment 1 to Experiment 2 (the conditions that yielded the strongest blocking effects in Experiment 1 yielded the smallest effects in Experiment 2) suggests that task-specific response criteria are at play. In the General Discussion, we present one possible description of the type of criteria that are being used to perform the progressive demasking and lexical decision tasks, based on the MROM (Grainger & Jacobs, 1996).

The pattern obtained to the nonword stimuli shows basically that mean nonword RT in a given list is strongly related to the mean RT to word stimuli in that list. The Pearson correlation between mean RT to word stimuli in the four pure lists and the mixed list and the corresponding mean RTs to nonword stimuli is .99 ($n = 5$). Within the framework of the MROM, the faster the average RTs to word stimuli, the lower the T criterion can be set without increasing the false negative error rate. Thus, the variable temporal deadline model of nonword responding that is implemented in the MROM (see also Coltheart et al., 1977; Coltheart et al., 2001) predicts that average nonword RTs will be correlated with average word RTs in a given list, since the setting of the T criterion is determined by speed of correct positive response to words.

GENERAL DISCUSSION

The present study was designed to examine a task-specific response criteria account (Grainger & Jacobs, 1996) of list-composition effects. In this framework, we consider that list-composition effects must always be interpreted in light of the particular task being performed. Indeed, distinct patterns of frequency- and density-blocking effects arose in each of the two tasks that were tested. Specifically, in the progressive demasking task, a pure-list disadvantage was obtained to low frequency–high density words, whereas high frequency–low density produced a trend toward a pure-list advantage. In lexical decision, high-frequency words showed a pure-list advantage that was strongest in high-density words, whereas low frequency–low density words produced a trend toward a pure-list disadvantage. As we will show below, this diversity of blocking effects can be explained in terms of the different response criteria, or different combinations of response criteria, that are used to trigger a speeded response in the different paradigms. Thus, blocking effects are hypothesized to reflect context-dependent adjustments of these response criteria in the framework of the MROM.

A Task-Specific Response Criteria Account of List-Composition Effects

The account described here is based on the principle that trial-to-trial adjustments in response criteria are the basis of all list-composition effects. We assume that participants try to optimize performance on each trial by applying on-line, *within-trial* adjustments of response criteria. These within-trial shifts in response criteria are maintained or not for the following trial as a function of the estimated accuracy of responding on that trial (which can be modified by explicit feedback). Thus, two factors determine the final setting of a given response criterion on a given trial: (1) the initial value of the criterion (influenced by the preceding trials); and (2) the estimated signal strength during target processing on that trial (above average values will pull the criterion down, and lower than average values will push it up). Following such within-trial modifications of response criteria, the modified criterion setting is maintained for the following trial if the participant believes that her/his response was accurate. Trials on which errors occur will modify this process of trial-by-trial adjustment. Given the low overall error rate and the absence of speed–accuracy trade-offs, this possibility will not be examined in the present study, but will provide the focus for future work manipulating speed and accuracy of performance.

It should be noted that our approach differs slightly from that described by Grainger and Jacobs (1996). In order to provide a general account of list-context effects, we assume that *all* response criteria are subject to the same mechanisms of strategic adjustment (this was not the case for the M criterion in the MROM). The presence of blocking effects in the progressive demasking task—in which only the M criterion is posited to be operational

(see Grainger & Jacobs, 1996)—strongly suggests that identification times were sensitive to influences beyond those involved in isolating a unique word representation. Thus, it seems reasonable to assume that, under time pressure, participants could initiate their response before the system has settled on a stable pattern of activity (i.e., before the activation level of the most activated word unit has reached asymptote). List context could influence the degree to which a response is initiated before such asymptotic levels have been reached. Within the framework of the MROM, this amounts to introducing a strategically variable M criterion.

According to the present account, in pure lists, within-trial adjustments of response criteria accumulate across trials. Since each adjustment is driven by the estimated signal strength (relevant for a given response criterion) on a given trial, it is the average values of these estimations that drive blocking effects. In mixed lists, criterion settings fluctuate up and down from trial to trial, preventing any systematic movement toward a higher or lower criterion setting. The major obstacle for such a general account of list-blocking effects is to operationalize the measure of signal strength for a given set of stimuli in a given task. Here we build on the approach developed by Grainger and Jacobs (1996) in the MROM: Signal strength for a given trial is operationalized as activation generated in early phases of processing (set at seven cycles in the model). This framework allows us to estimate the size of blocking effects in the progressive demasking and lexical decision tasks used in Experiments 1 and 2 of the present study.

In previous simulation work with the MROM (Grainger & Jacobs, 1996), a measure of signal strength during stimulus processing was used to modify the setting of the Σ criterion during each trial. This measure was the summed activity of all positively activated word units after seven cycles of processing, which was appropriate for a response criterion set on global word activation. For the M criterion in the MROM, the equivalent measure of signal strength is the activation of the most activated word unit at a given point in processing (we will again apply seven cycles to comply with our earlier simulation work). These values can be used to modify the M criterion setting on a given trial, high values causing a slight lowering of the criterion and low values causing a slight increase in the criterion setting. The average activation values at seven cycles of the most activated word unit, and the summed activation of all activated word units, are given in Table 4 for each stimulus category tested in the present study. For the progressive demasking, a task that requires unique word identification, we assume that only the M criterion is operational. Blocking effects are determined by the average values of unit activation at seven cycles for each stimulus category. For the lexical decision task of Experiment 2 we assume, following Grainger and Jacobs, that both the M and Σ criteria will be used in generating responses to word stimuli. The average values of m and σ for each stimulus category relative to the grand average of m and σ for word stimuli (see Table 4) will

Table 4
Activation of Most Activated Word Unit (*m*), and Summed
Lexical Activation (σ) After Seven Cycles of Processing
in the MROM for the Four Categories of Words
Tested in the Present Study

	m(7)	σ(7)
HF-LD	.19+	.24−
HF-HD	.16	.36+
LF-LD	.16	.26−
LF-HD	.13−	.37+
Average	.16	.31

Note—Plus and minus signs refer to whether the condition means are greater than or smaller than the grand average. HF, high frequency; LD, low density; HD, high density; LF, low frequency.

determine whether the respective response criteria will be lowered or raised in the pure lists relative to the mixed lists. The M criterion setting is the same as for Experiment 1, and this is combined with the Σ criterion setting to predict blocking effects in Experiment 2. According to these average values, when only the M criterion is used for read-out, the model should generate a positive blocking effect for the high frequency–low density words and a negative blocking effect for the low frequency–high density words. On the other hand, when both the M and the Σ criteria are used for response read-out, the model should generate a positive blocking effect for the high frequency–high density words and a negative blocking effect for the low frequency–low density words. This is tested in the simulation study below.

Simulation Study

The simulation study was based on an implementation of our general theory of list blocking effects expressed within the framework of the extended MROM model. The aim of this study was to demonstrate that (1) such an implementation is possible and (2) that the model does generate a pattern of responses that are compatible with the observed pattern of blocking effects, and are in line with the general predictions of our theory. It should be noted that our aim was not to provide the best fits to the empirical data by trial-and-error parameter setting. No such parametric adjustments were used in our simulations. The same parameter settings as in our previous simulation work (Carreiras et al., 1997; Grainger & Jacobs, 1996) were adopted for the present simulations. The parameters related to mechanisms that have been added to the original MROM (flexible M criterion and trial-by-trial criterion adjustments) were set a priori on the basis of the average activation values shown in Table 4, and were not adjusted in order to provide better fits to the data. The only parametric adjustment that was necessary in the present simulation study was the size of trial-by-trial adjustments of the σ criterion (an initial value of 0.1 was found to be too large). As stated earlier, in order to keep the simulation as straightforward and as transparent as possible, we did not attempt to simulate criterion adjustments following an error or criterion adjustments following a nonword trial.

Progressive demasking task. As in our previous simulation work, only the M criterion was used in the simulation of performance in the progressive demasking task. A time-out of 50 cycles was employed, and trials on which this time-out was reached were treated as errors. Contrary to our prior simulation work, in the new version of the MROM, the M criterion flexibly adjusts as a function of the value of the activation of the most activated word unit at seven cycles. This within-trials adjustment of the M criterion is directly analogous to the within-trials adjustments of the Σ criterion implemented in the original MROM (Grainger & Jacobs, 1996). In both the pure-list and the mixed-list simulations, the default starting value for M was set at 0.70, and the values changed from trial to trial using the following equations: if $m(7) \geq 0.16$ then $M = M - 0.1$; if $m(7) < 0.16$ then $M = M + 0.1$. Following Grainger and Jacobs, the response criteria varied around their mean value (determined using the above equations) following a truncated normal distribution. A fixed pseudorandom order of presentation was used for all simulations. The results of the simulations averaged over 20 simulated participants are given in Table 5.

A comparison of Table 2 (empirical data) and Table 5 (predicted data) shows that the model does a reasonable job of capturing the relative size and direction of blocking effects in the progressive demasking task. It successfully captures the fact that the strongest blocking effects emerge in the high frequency–low density and low frequency–high density conditions, and these two conditions show effects that go in opposite directions. The simulation results are perfectly in line with what was predicted on the basis of the average activation values of the most activated word unit after seven cycles of processing [$m(7)$], as shown in Table 4.

Lexical decision task. Variable Σ and T criteria were added for the lexical decision simulation. The M criterion was adjusted in the same way as for the progressive demasking simulation, except that the average value of M was decreased by 0.2 (average M = 0.68 for lexical decision compared with 0.70 for progressive demasking). The Σ and T criteria were adjusted on each trial on the basis of the summed activation of all word units activated above zero at seven cycles [$σ(7)$]. In both the pure-list and mixed-list simulations, the Σ and T criteria were adjusted in the following way: if $σ(7) \geq 0.31$ then $Σ = Σ - 0.03$; if $σ(7) < 0.31$ then $Σ = Σ + 0.03$; if $σ(7) \geq 0.18$ then $T = T + 1$; $σ(7) < 0.18$ then $T = T - 1$. The starting values of the M, Σ, and T criteria for each simulated pure block were 0.68, 0.92, and 21 cycles, respectively.

The results of these simulations averaged over 20 simulated participants are given in Table 6. A comparison of Table 3 (empirical data) and Table 6 (predicted data) shows that the model captures the direction of blocking effects in the lexical decision task. The inclusion of the Σ criterion has, as predicted by our theory, completely modified the pattern of blocking effects compared with those obtained with the progressive demasking task. As

Table 5
Mean Time (in Milliseconds) to Reach the Flexible M Criterion and Percent Errors (PEs) in the MROM Simulation of the Progressive Demasking Task

	Type of List					
	Pure		Mixed		Mixed − Pure	
	M	PE	M	PE	M	PE
HF-LN	17.38	0.0	20.27	1.7	2.89	1.7
HF-HN	21.52	0.0	21.22	0.7	−0.30	0.7
LF-LN	22.43	6.7	22.49	14.2	0.06	7.5
LF-HN	26.24	2.3	23.19	0.7	−3.05	−1.6

Note—HF, high frequency; LD, low density; LN, low number of neighbors.

Table 6
Time (in Milliseconds) to Reach Either the M or the Σ Criterion and Percent Errors (PEs) in the MROM Simulation of the Lexical Decision Task

	Type of List					
	Pure		Mixed		Mixed − Pure	
	M	PE	M	PE	M	PE
HF-LN	16.68	0.8	17.75	0.8	1.07	0.0
HF-HN	17.17	1.9	18.83	1.4	1.66	−0.5
LF-LN	20.90	14.2	19.31	13.9	−1.59	−0.3
LF-HN	18.18	8.1	19.07	8.4	0.89	0.3

Note—HF, high frequency; LD, low density; LN, low number of neighbors.

in the empirical data, it is now the high frequency–high density and low frequency–low density conditions that produce the strongest blocking effects, and these go in opposite directions. Again, this is what was predicted on the basis of the average values of summed lexical activity $[\sigma(7)]$ combined with the activation values of the most activated word unit $[m(7)]$, as shown in Table 4.

The most important aspect of the present demonstration is that two distinct measures of lexical activation (unit activation and summed activation) produce a very different pattern of means for the four categories of stimuli tested in the present study. At a very general level, therefore, our approach accounts for the fact that distinct patterns of blocking effects were obtained in Experiments 1 and 2. The values generated from the MROM go further than this general prediction and provide quite an accurate description of the blocking effects obtained in these experiments.

In sum, a task-specific response criteria account such as that proposed in the framework of the MROM not only provides a good description of the list-blocking effects observed in the present study, but also has the potential to account for variations in RTs and error rates as a function of factors such as frequency and neighborhood density in different experimental tasks.

List Composition and Strategic Control of Processing

The pattern of results obtained in the present series of experiments has shown that the effects of word frequency and neighborhood density differ as a function of the type of list. As a result, it is important to examine what type of list (pure or mixed) gives the most useful information on the process of visual word recognition. Forster and Shen (1996) pointed out that the effects obtained in pure lists (e.g., a word list composed of high-frequency words) are strategic in nature because it is only under these conditions that the type of item that is going to be presented on the next trial can be anticipated. Similarly, Andrews (1997) concluded her review of neighborhood effects by saying that manipulations such as blocking are difficult to interpret because they allow participants to adjust their decision criteria or the information they use to make a response for each stimulus type. Andrews (1997) indicated that the most relevant data are those obtained when stimuli vary across the full range of possible stimulus characteristics (i.e., mixed lists), so that participants cannot modify their strategy across trials in anticipation of particular stimuli.

In contrast, one might argue that using pure blocks would be the best way to verify the presence of an effect. The reason is that if we assume that participants are always trying to optimize their behavior (i.e., minimize their RTs while keeping a low error rate), then performance in pure blocks would be the one that is best adjusted to the characteristics of a given type of stimulus. Bear in mind that performance in mixed blocks may be optimal for the list as a whole, but not for each type of stimulus in particular. Consistent with this view, it has been argued that "the use of pure blocks should be encouraged whenever it makes sense" (Lupker et al., 1997, p. 586). (As noted by Lupker et al., pure blocks can create a situation in which criteria can inappropriately affect the results; e.g., an associative priming experiment comparing a block of pure related trials and a block of pure unrelated trials.)

We would argue that the effects obtained with blocked lists of stimuli and the effects obtained with mixed lists are both interesting in their own right. One simply needs a quantitative, computational model that specifies exactly how list composition will influence performance. This model should specify which mechanisms are involved in task- and context-independent processing of printed words, and which mechanisms are specific to a given task and a given context (cf. the concept of *functional overlap* discussed by Grainger & Jacobs, 1996). Our approach builds on one such model, the MROM, which we believe to be a promising start to such an enterprise.

Alternative Accounts

Recently, Lupker and colleagues (Lupker et al., 1997; Taylor & Lupker, 2001; see also Jared, 1997) described an alternative approach to explaining list-composition effects. This account, referred to as the time criterion account, assumes the presence of small trial-to-trial adjustments in the setting of a time criterion used to determine response initiation. These adjustments reflect a gradual shift of the time criterion's value over the course of a series of (similar) trials toward the optimal value for

a set of stimuli: Participants lower a given response criterion in blocked lists of words that are easy to identify (i.e., high-frequency words) and raise the criterion in blocked lists of difficult (i.e., low-frequency) words. (This account is neutral as to what factors make some stimuli "easy" or "difficult.") Thus, grouping the easy words together will produce a blocking advantage, whereas putting the difficult words together will produce a blocking disadvantage. In other words, this account predicts that there would be a "homogenization" effect in the mixed relative to the pure lists. There is empirical evidence that supports this view in naming (see Taylor & Lupker, 2001).

According to the time criterion approach, it is the same response criterion that is involved in list-composition effects, independently of task. Any changes in list-composition effects across tasks will be due to differences in overall ease of processing the different categories of stimuli in the different tasks. That is, if a variable time criterion is the only mechanism responsible for frequency- and density-blocking effects, the "slowest" word conditions in the pure list should show a blocking disadvantage (i.e., faster RTs in the mixed list than in the pure list), independently of word frequency (or neighborhood density), whereas the "fastest" word conditions should show a blocking advantage. More specifically, the time criterion account predicts that there should be a general homogenization effect across stimulus categories when shifting from the pure block condition to mixed lists independently of the task being performed. Although there was some evidence for homogenization, this process falls well short of accounting for the complete and rather complex pattern of blocking effects obtained in the present study. The fact that each experiment generated a very specific pattern of list-composition effects (as a function of word frequency and neighborhood density) suggests that factors other than homogenization are playing a role in list-blocking effects.

A computational implementation of the time-criterion account within a model of visual word recognition and naming would be necessary to examine in detail the advantages/limitations of such an account. Bear in mind that one obvious limitation of the time-criterion account, as it stands, is that it remains neutral as to what factors are responsible for the pattern of effects (word frequency, neighborhood density, etc.); instead, it is posited to be a general principle that predicts which conditions will show a blocking advantage or a blocking disadvantage on the basis of mean RTs in a pure list. But what we should note is that the pattern of blocking effects is not always consistent with a single criterion account. For instance, in the lexical decision task, there is clear evidence that there is a blocking advantage for high-frequency words (including our data in Experiment 2), whereas the evidence that the presence of an overall blocking disadvantage for low-frequency words—such as that predicted by the time-criterion account—is not supported by most empirical data (e.g., Dorfman & Glanzer, 1988; Gordon, 1983; Glanzer & Ehrenreich, 1979; Stone & Van Orden,

1993). In any event, it is worth noting that Taylor and Lupker (2001) acknowledged that the use of a time criterion is not the only strategic adjustment possible in cognitive tasks, and that it was likely that participants would employ other strategies depending on the characteristics of the experimental tasks.

In the framework of a connectionist network, Kello and Plaut (2000, 2003) suggested that these context-dependent changes can be ascribed to the rate of processing (or input gain) within the network. In this model, which has been proposed for the naming task, pure blocks of relatively easy stimuli encourage higher levels of input gain, since responses can be both fast and accurate under these conditions. In contrast, pure blocks of relatively difficult stimuli require moderate levels of input gain in order to keep errors to a minimum. The specific predictions of this model are quite similar to those of the time-criterion account (with some exceptions; see Kello & Plaut, 2000).

The diffusion model (Ratcliff, 1978) is probably the most successful model in accounting for performance in binary decision tasks (see Ratcliff & Rouder, 1998; Ratcliff, Van Zandt, & McKoon, 1999). The diffusion model assumes a gradual accumulation of noisy information over time toward a positive or a negative boundary (i.e., "nonword" responses are not guided by a temporal deadline, but by a bottom boundary responsible for "nonword" decisions). A recent implementation of the diffusion model has been suggested to successfully account for the lexical decision task (Ratcliff et al., 2004; see also Stone & Van Orden, 1993, for a qualitative account of the model). The model can cope with the presence of frequency-blocking effects (Ratcliff et al., 2004) in terms of variations in the drift rate (a parameter responsible for "quality" of information) and in terms of variations in the distance between the boundaries and the starting point (a parameter responsible for bias). This, in fact, is very similar to our own account, in which both stimulus quality evaluated during target processing and the initial setting of response criteria are at the basis of frequency- and density-blocking effects. One possible limitation of the diffusion model for researchers in visual word recognition is that it is a model of the decision process in binary tasks, so other word identification tasks (e.g., speeded identification, naming) are beyond the scope of the model.

Finally, unlike the previous models, the DRC model (Coltheart et al., 2001) can cope with the two experimental tasks in the present study, especially lexical decision. However, Coltheart et al. did not discuss how the model would deal with blocking effects in the lexical decision task. Nonetheless, it seems reasonable to assume that the DRC model may account for the present effects in the lexical decision task by strategic variations of some of the parameters. Keep in mind that "word" and "nonword" responses are driven by similar mechanisms in the MROM and in the DRC model (i.e., unique word identification and summed lexical activation for "word" responses, and a temporal deadline for "nonword" responses).

Blocking Effects and Sequential Dependencies

Given an account of blocking effects in terms of criterion adjustments, an important question is whether the response criteria are set for the entire list or on a trial-by-trial basis. Treisman and Williams (1984; see also Strayer & Kramer, 1994a, 1994b) suggested that list-context effects reflect the operation of a system with two components: a long-term macroscopic criterion setting process that determines a reference criterion and a short-term microscopic process of fine adjustments. Once the system has specified a value for the reference criterion, this should remain substantially constant over short sequences of trials. Nonetheless, from trial to trial, fine adjustments may be necessary to ensure that the criterion in use is optimally placed. The resultant response criteria used on a given trial are determined by a combination of these long- and short-term processes. Our task-specific response criteria account of blocking effects also appeals to trial-by-trial fine adjustments of response criteria. As discussed above, it is an on-line measure of ease of processing on each trial, combined with a check for errors once the response is generated, that governs the criterion setting procedure. Global list-context effects are assumed to be the cumulative result of such adjustments. However, participants must obviously begin the experiment with a given criterion value, which is likely to be determined by factors such as task instructions and any prior experience with the task. Our account is therefore essentially an adaptation of Treisman and Williams's (1984) criterion setting theory to the specific tasks used in the present study.

In line with this general approach, recent evidence suggests that criterion placement in word recognition experiments may be adjusted on an item-by-item basis. Lima and Huntsman (1997) found sequential dependencies in the lexical decision task when the trials are not related to each other: Both word and nonword responses were significantly slower when the previous trial involved a nonword than when it involved a word. Similarly, Lupker et al. (1997; Taylor & Lupker, 2001) found that both high- and low-frequency words are named faster when a high-frequency word has been presented in the previous trial than when another low-frequency word has been presented in the previous trial. Finally, Perea and Carreiras (2003) found first-order sequential item-frequency effects for low-frequency words and nonwords in the lexical decision task, although no such effects due to the frequency of the previous trial were found for high-frequency words.

Taken together, these results appear to imply a dynamic model in which participants continually adjust their criterion settings on a momentary, trial-by-trial basis (Treisman & Williams, 1984). These momentary adjustments could also be modulated by more global adjustments made on the basis of an appreciation of the list composition as a whole. Thus it is possible that a part of the blocking effects observed in the present experiments may reflect processes other than the accumulation of first-order sequential effects (e.g., a word list composed of high-frequency words could induce participants to use more extreme response criteria for making "word" responses than those induced in a mixed list with high- and low-frequency words). Clearly, the absence of first-order effects in certain conditions may simply imply that measurement sensitivity is not great enough to capture such fine adjustments. Work examining the influence of increasingly larger sequences of preceding stimuli will be critical for testing criterion adjustment accounts of list-blocking phenomena.

Summary

To summarize, the results of the present study indicate that list-composition effects are a complex phenomenon: Distinct patterns of blocking effects were obtained in two different tasks (progressive demasking and lexical decision) as a function of word frequency and neighborhood density. The modulation of task-specific response criteria provides a coherent account of the pattern of blocking effects observed in the present series of experiments. This account builds on the modeling approach developed by Grainger and Jacobs (1996) for visual word recognition, while being inspired by more general approaches to the sequential adjustment of response criteria (Treisman & Williams, 1984). Undoubtedly, more research is needed to analyze what the participant does when making a response in a laboratory word identification task, as well as to explore how the criterion settings can be strategically modified.

REFERENCES

ALAMEDA, J. R., & CUETOS, F. (1995). *Diccionario de frecuencia de las unidades lingüísticas del castellano* [Dictionary of word frequency in Spanish]. Oviedo: Servicio de publicaciones de la Universidad de Oviedo.

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

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

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

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.

CARR, T. H., POSNER, M. I., POLLATSEK, A., & SNYDER, C. R. R. (1979). Orthography and familiarity effects in word processing. *Journal of Experimental Psychology: General*, **108**, 389-414.

CARREIRAS, M., PEREA, M., & GRAINGER, J. (1997). Effects of orthographic neighborhood in visual word recognition: Cross-task comparisons. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **23**, 857-871.

COLTHEART, M., DAVELAAR, E., JONASSON, J. F., & BESNER, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535-555). Hillsdale, NJ: Erlbaum.

COLTHEART, M., RASTLE, K., PERRY, C., ZIEGLER, J., & LANGDON, R. (2001). DRC: A Dual Route Cascaded model of visual word recognition and reading aloud. *Psychological Review*, **108**, 204-256.

DIJKSTRA, A. F. J., GRAINGER, J., & VAN HEUVEN, W. J. B. (1999). Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory & Language*, **41**, 496-518.

Dorfman, D., & Glanzer, M. (1988). List composition effects in lexical decision and recognition memory. *Journal of Memory & Language*, **27**, 633-648.

Forster, K. I. (1981). Frequency blocking and lexical access: One mental lexicon or two? *Journal of Verbal Learning & Verbal Behavior*, **20**, 190-203.

Forster, K. I., & Shen, D. (1996). No enemies in the neighborhood: Absence of inhibitory neighborhood effects in lexical decision and semantic categorization. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **22**, 696-713.

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

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

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

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

Grainger, J., & Segui, J. (1990). Neighborhood frequency effects in visual word recognition: A comparison of lexical decision and masked identification latencies. *Perception & Psychophysics*, **47**, 191-198.

Grainger, J., Spinelli, E., & Ferrand, L. (2000). Effects of baseword frequency and orthographic neighborhood size in pseudohomophone naming. *Journal of Memory & Language*, **42**, 88-102.

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

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

Kello, C. T., & Plaut, D. C. (2003). Strategic control over rate of processing in word reading: A computational investigation *Journal of Memory & Language*, **48**, 207-232.

Kiger, J. I., & Glass, A. L. (1981). Context effects in sentence verification. *Journal of Experimental Psychology: Human Perception & Performance*, **7**, 688-700.

Lima, S. D., & Huntsman, L. A. (1997). Sequential dependencies in the lexical decision task. *Psychological Research*, **60**, 264-269.

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

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, **88**, 375-405.

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

Monsell, S., Doyle, M. C., & Haggard, P. N. (1989). Effects of frequency in visual word recognition tasks: Where are they? *Journal of Experimental Psychology: General*, **118**, 43-71.

Paap, K. R., & Johansen, L. S. (1994). The case of the vanishing frequency effect: A retest of the verification model. *Journal of Experimental Psychology: Human Perception & Performance*, **20**, 1129-1157.

Perea, M., & Carreiras, M. (2003). Sequential effects in the lexical decision task: The role of the item-frequency of the previous trial. *Quarterly Journal of Experimental Psychology*, **56A**, 385-401.

Perea, M., & Pollatsek, A. (1998). The effects of neighborhood frequency in reading and lexical decision. *Journal of Experimental Psychology: Human Perception & Performance*, **24**, 767-777.

Perea, M., & Rosa, E. (2002). Does "whole-word shape" play a role in visual word recognition? *Perception & Psychophysics*, **64**, 785-794.

Perea, M., Rosa, E., & Gómez, C. (2003). Influence of neighborhood size and exposure duration on visual-word recognition: Evidence with the yes/no and the go/no-go lexical decision task. *Perception & Psychophysics*, **65**, 273-286.

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

Pollatsek, A., Perea, M., & Binder, K. (1999). The effects of neighborhood size in reading and lexical decision. *Journal of Experimental Psychology: Human Perception & Performance*, **25**, 1142-1158.

Pollatsek, A., & Well, A. (1995). On the use of counterbalanced designs in cognitive research: A suggestion for a better and more powerful analysis. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 785-794.

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

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

Ratcliff, R., & Rouder, G. (1998). Modeling response times for decisions between two choices. *Psychological Science*, **9**, 347-356.

Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychological Review*, **106**, 261-300.

Schreuder, R., & Baayen, R. H. (1997). How complex simple words can be. *Journal of Memory & Language*, **37**, 118–139.

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

Siakaluk, P. D., Sears, C. R., & Lupker, S. J. (2002). Orthographic neighborhood effects in lexical decision: The effects of nonword orthographic neighborhood size. *Journal of Experimental Psychology: Human Perception & Performance*, **28**, 661-681.

Snodgrass, J. G., & Mintzer, M. (1993). Neighborhood effects in visual word recognition: Facilitatory or inhibitory? *Memory & Cognition*, **21**, 247-266.

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

Strayer, D. L., & Kramer, A. F. (1994a). Strategies and automaticity: I. Basic findings and conceptual framework. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 318-341.

Strayer, D. L., & Kramer, A. F. (1994b). Strategies and automaticity: II. Dynamic aspects of strategy adjustment. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 342-365.

Taylor, T., & Lupker, S. J. (2001). Sequential effects in naming: A time criterion account. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 117-138.

Treisman, M., & Williams, T. C. (1984). A theory of criteria setting with an application to sequential dependencies. *Psychological Review*, **91**, 68-111.

van Heuven, W. J. B., Dijkstra, T., & Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory & Language*, **39**, 458-483.

NOTES

1. The order of experimental lists was low frequency–low density, mixed, mixed, low frequency–low density, mixed, mixed, high frequency–low density, high frequency–high density, and vice versa; the order was the same in Experiments 1 and 2.

2. Of course, as an anonymous reviewer pointed out, one can never have pure lists of high- or low-frequency words in the lexical decision task, because the lists always include nonwords.

APPENDIX
Word Targets Used in the Experiments

High Frequency–Low Density Words

agua (water), amor (love), ayer (yesterday), azul (blue), edad (age), isla (island), jefe (boss), lujo (luxury), ocho (eight), país (country), real (real), zona (zone), campo (field), carne (meat), común (common), dulce (sweet), fácil (easy), hotel (hotel), miedo (fear), negro (black), orden (order), papel (paper), radio (radio), siglo (century)

High Frequency–High Density Words

aire (air), base (base), boca (mouth), doña (lady), duda (doubt), lado (side), mesa (table), rato (period), risa (laugh), rojo (red), ropa (clothing), tema (theme), calle (street), carta (letter), corte (cut), costa (coast), gesto (gesture), grado (degree), marco (frame), medio (middle), moral (moral), patio (courtyard), plano (flat), razón (reason)

Low Frequency–Low Density Words

afín (similar), añil (indigo), asno (donkey), bidé (bidet), nazi (Nazi), ogro (ogre), olmo (elm), orla (border), oval (oval), peón (laborer), rail (rail), yoga (yoga), belga (Belgian), fósil (fossil), garbo (glamour), hiena (hyena), laico (lay), legua (league), mugre (filth), potro (colt), rapaz (predatory), rosal (rose bush), sauce (willow), sepia (squid)

Low Frequency–High Density Words

faja (girdle), foso (pit), gasa (gauze), mazo (mallet), moho (mold), nata (cream), saga (saga), soja (soy), topo (mole), vara (stick), vena (vein), batín (dressing gown), dólar (dollar), hacha (ax), lapa (limpet), menta (mint), morro (thick lip), patán (rustic), porra (stick), socio (associate), suela (sole), suero (serum), tallo (stem), torta (cake)

Note—Approximate English translations given in parentheses.