How lexical decision is affected by recent experience: Symmetric versus asymmetric frequency-blocking effects

SACHIKO KINOSHITA
Macquarie University, Sydney, New South Wales, Australia

and

MICHAEL C. MOZER
University of Colorado, Boulder, Colorado

In a lexical decision task (LDT) in which list composition is manipulated, a typical finding to date has been a slowdown for easy items (e.g., high-frequency words) but little speedup for hard items (e.g., low-frequency words) when they are mixed together. This asymmetric frequency-blocking effect contrasts with the symmetric pattern (both a speedup for hard items and a slowdown for easy items when they are mixed together) observed with the naming task. In the present study, we investigated the mechanism responsible for the asymmetric blocking effect in the LDT within a model of blocking effect proposed by Mozer, Kinoshita, and Davis (2003), termed the *adaptation-to-the-statistics-of-the-environment* (ASE) model. Experiments 1A and 1B showed that when the same high- and low-frequency words were used, consistent with the existing literature, an asymmetric blocking effect was found in the LDT and a symmetric blocking effect was found in the naming task. Within the ASE model, a symmetric versus asymmetric blocking effect can be explained in terms of different asymptotic rates in subjective estimates of error probability. Experiments 2 and 3 tested and confirmed a prediction of the model based on this assumption that a speedup of hard items would be observed in an LDT with hard items whose subjective error probability asymptotes near zero (low-frequency words with high familiarity ratings that subjects could be certain were words). Implications of the model for task differences in reaction times are discussed.

Reaction time (RT) tasks are probably the most popular tools used by experimental psychologists to study cognitive processes. Put simply, the assumption underlying the use of the task is that the longer the RT, the more complex the underlying processes required to produce the response. There is a growing appreciation, however, of the fact that RT to a stimulus is affected by factors other than those directly related to the stimulus itself. One such factor is the difficulty of the preceding stimuli, particularly those from the most recent trials. Using the naming task, Lupker and colleagues (e.g., Lupker, Brown, & Colombo, 1997; Taylor & Lupker, 2001) have shown that the speed with which preceding items are responded to has a major impact on the RTs to the succeeding stimuli. In particular, when items differing in response latency (e.g., high-frequency vs. low-frequency words or words vs. nonwords) are presented together in a mixed block, the easy items slow down and the hard items speed up by about the same amount, relative to when these items are presented in separate, pure blocks. Lupker et al. (1997) termed this pattern of results the *blocking effect*.

Lupker, Kinoshita, Coltheart, and Taylor (2003) found blocking effects when subjects named stimuli other than words—specifically, pictures and sums (where the subjects were required to say aloud the solution to addition problems such as 8 + 9 = ?). They even observed cross-stimulus blocking effects when these stimuli were mixed with word stimuli. Also, Rastle, Kinoshita, Lupker, and Coltheart (2003) found cross-task effects: The difficulty of the stimuli in a naming task affected responses in a lexical decision task (LDT), and vice versa, when these tasks were interleaved. These findings suggest that blocking effects are a very common and robust phenomenon, as was noted by Kiger and Glass (1981) over 20 years ago. These authors commented that these effects will "continue to be rediscovered in many circumstances" (p. 697). The cross-task effects have led Rastle et al. (2003) to suggest that the locus of the effect must be at a stage common to all the

Note—This article was accepted by the previous editorial team, when Colin M. MacLeod was Editor.

tasks—probably, at the stage of response initiation. The aim of the present study was to further our understanding of the mechanism underlying blocking effects.

The Blocking Effect

Using the naming task, Lupker and colleagues (Lupker et al., 1997; Taylor & Lupker, 2001) presented verbal stimuli (words and nonwords) in a pure block of easy items, a pure block of hard items, and a mixed block in which the easy and hard items were presented randomly. Across a series of experiments, the following patterns were observed consistently.

1. Subjects were faster and less error prone for easy items than for hard items.

2. When easy and hard items were intermixed, there was a slowdown for easy items and a speedup for hard items. However, the convergence of easy and hard items was not complete.

3. Speed–accuracy trade-offs were observed: A drop in error rate accompanied easy item slowdown; a rise in error rate accompanied hard item speedup (although in the naming task, the error rate is generally low overall and, because the errors must be perceived by the experimenter, effects on error rate may be difficult to observe).

4. The magnitudes of easy item slowdown and the hard item speedup were similar. A significant speedup for hard items in the absence of a slowdown for easy items was never observed.

5. A *sequence effect* was observed—that is, the RT in trial n was shorter if the RT in trial $n-1$ (and trial $n-2$ to a lesser extent) was short; trial n was slowed down if the RT in trial $n-1$ was long.

These five patterns have also been observed in naming tasks involving other types of stimuli (e.g., pictures or sums; Lupker et al., 2003; Rastle et al., 2003). Thus, we regard them as signatures of a blocking effect that any model of the blocking effect must be able to explain.

Existing Accounts of the Blocking Effect

As has been noted by Kiger and Glass (1981), observation of the blocking effect is not new, and various accounts have been proposed. Also as was noted by Kiger and Glass, most of these accounts have been domain specific, and this continues to be the case. For example, in the word recognition literature, the dominant interpretation of the blocking effect has been one framed in terms of the dual-route model of reading aloud, which assumes a change in relative speeds of lexical versus nonlexical routes for generating phonology from print (e.g., Rastle & Coltheart, 1999; but see Chateau & Lupker, 2001, for evidence against their account). Another word recognition model used to explain list composition effects is Grainger and Jacobs's (1996) multiple readout model (MROM). Like many other word recognition models, MROM assumes that a criterion set on activity in a word detector unit (the M-criterion) is used to trigger a positive word recognition response. In addition, in an LDT, the Σ criterion, which reflects general activity across all word detector units (global lexical activity), could be used to

make a positive response before a specific word detector reaches the criterion. Grainger and Jacobs explained the frequency-blocking effect observed in the LDT by a change in the value of the Σ criterion. More recently, this model has been extended to other tasks, including naming (Perea, Carreiras, & Grainger, 2004); to simulate blocking effects in different tasks, Perea et al. reported that changes to the value of task-specific parameters were required. In picture naming, Meyer, Roelofs, and Levelt (2003) observed blocking effects with pictures with mono- and disyllabic names. Meyer et al. explained the blocking effect by assuming that different criteria are used for initiating articulation in pure and mixed blocks: In pure blocks, a whole-word–based criterion is used, whereas in a mixed block, a criterion based on only the first syllable is used, so that the difference between mono- and disyllabic words is eliminated (see Lupker et al., 2003, for arguments against this account).

In all of the accounts described above, to explain blocking effects, the value of the parameter that is changed between blocks to fit the observed data is task specific, and its scope is limited to the specific model under discussion. Although we do not deny the possibility that subjects may strategically change the weighting on different type of information used to make decisions within a given task, we agree with Kiger and Glass (1981) that the cross-task blocking effects (as observed, e.g., by Kiger & Glass, 1981, and by Rastle et al., 2003) call for a unitary mechanism of adaptation, as a function of list composition, that is general across tasks.

A candidate mechanism with an intuitive appeal is the trial-by-trial adjustment-of-evidence criterion. For example, the random-walk/diffusion model (Ratcliff, 1978) employs such a mechanism. In brief, the diffusion model is a continuous stochastic decision process that integrates evidence toward two competing responses—for example, *word* and *nonword* responses. It assumes that evidence accumulates over time following a random walk with a certain mean rate of evidence accumulation (drift rate) and that a response is output when the accumulated evidence reaches one of the evidence criteria (absorption boundary). The parameter values governing the mean drift rate and the positive and negative absorption boundaries are estimated by best fits to the distribution of RTs and errors. To explain blocking effects within this framework, the evidence criterion is set lower in the pure easy block, higher in the pure hard block, and at an intermediate value in the mixed block, leading to a slowdown for easy items and a speedup for hard items in the mixed block.

Although the adjustment-of-evidence criterion has an intuitive appeal, we are skeptical that this is the mechanism responsible for the blocking effects that have been observed in the naming task. The main theoretical reason is that a simple averaging-of-evidence criterion cannot produce the signature pattern of the blocking effect— namely, the fact that the magnitude of speedup for hard items in a mixed block is similar to the slowdown for easy items (Pattern 4, described earlier). To appreciate this problem, consider Figure 1, which schematically depicts
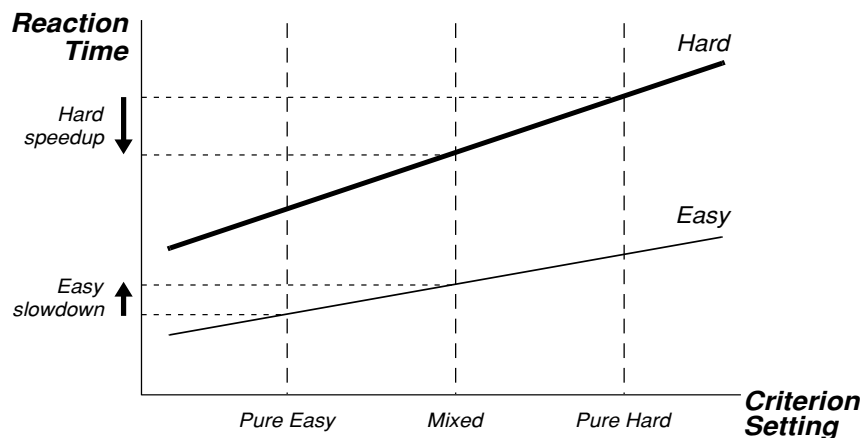
Figure 1. Schematic depiction of reaction time as a function of criterion setting for any model of blocking effects based on the adjustment-of-evidence criterion. The thin line is for easy items; the thick line is for hard items. The dashed vertical lines indicate the criterion setting for the three block types, with a mixed block having a criterion intermediate between that of pure easy and pure hard blocks.

the relationship between the criterion setting (*x*-axis) and the RT (*y*-axis) for easy and hard items.

As the criterion is raised, RTs for both hard and easy items increase. If the criteria indicated are selected for pure easy and pure hard blocks and a mixed block achieves an intermediate criterion, one will obtain a speedup for hard items and a slowdown for easy items, consistent with the blocking effect. However, for the curves depicted in the figure, the speedup will be *greater* than the slowdown for easy items, inconsistent with the observed pattern. In the figure, this effect arises because the slope of the function relating criterion setting to RT is steeper for hard than for easy items. This is not an arbitrary assumption; it is incorporated into most RT models, because RT effects scale multiplicatively. Within the diffusion model, this property is reflected in the assumption that the drift rate (the mean rate of evidence accumulation) for hard items is less than the drift rate for easy items; hence, raising (lowering) the absorption boundary by the same amount would have a greater impact on the RTs for hard items than on those for easy items. We believe that this property poses a fundamental obstacle to an explanation of blocking effects in terms of the adjustment-of-evidence criterion, and our own attempts to find an algorithm for locating the evidence criterion for the mixed block, relative to the pure blocks, met with failure.

Thus, although there have been previous studies based on the diffusion model that have claimed to have successfully accounted for blocking effects in terms of the adjustment-of-evidence criterion in a memory search task (Strayer & Kramer, 1994) and an LDT (Ratcliff, Gomez, & McKoon, 2004), we have some reservations about this claim. First, inconsistent with the evidence criterion adjustment account, the parameter whose value changed as a function of composition of block was not the absorption boundary but the drift rate parameter in Ratcliff et al.'s (2004) study. In contrast, Strayer and Kramer did report

different absorption boundary parameter values in the pure easy, pure hard, and mixed blocks. However, it is important to note that in a mixed block, there must be a single criterion setting for the easy and hard items, because subjects do not know in advance of a trial in a mixed block whether the target will be easy or hard. The fits obtained by Strayer and Kramer violated this assumption. Thus, on the basis of the available evidence, it is unclear whether the adjustment-of-evidence criterion within the framework of the diffusion model provides a satisfactory account of blocking effects.

Mozer, Kinoshita, and Davis's (2003) Adaptation-to-the-Statistics-of-the-Environment Model

Recognizing the need for an alternative mechanism, Mozer, Kinoshita, and Davis (2003) proposed a model to account for the blocking effect data observed with the naming task reported by Lupker and colleagues (Lupker et al., 1997; Taylor & Lupker, 2001), summarized above. They used a decision model suggested earlier by Mozer, Colagrosso, and Huber (2002). The main tenet of this decision model is that instead of simply choosing a threshold in error rate or in RT, a response cost is computed that depends on both RT (i.e., the cost of waiting) and expected error rate (i.e., the cost of possibly making an incorrect response), and that a control mechanism initiates a response at the point in time at which a minimum in net cost is attained. The model assumes that the response initiation system obtains an estimate of its expected error rate as processing within a trial unfolds.

This estimate is obtained in the following way. At any point in processing, the model has accumulated evidence in favor of each of the various response alternatives, much as a random walk or accumulator model would. The model translates this evidence into a probability distribution over responses. Assuming that the most probable response is correct, the error estimate is simply the sum

of the probabilities of the other response alternatives. The solid curve in Figure 2A represents this expected error, monotonically decreasing over time. At any point in time, a response can be initiated, and the corresponding value of the curve represents the probability of an error. The dotted straight line in Figure 2A represents the elapsed time. A response initiation cost depends on both the expected error rate and the RT; it decreases as the error rate drops and increases as processing time builds up. Figure 2B shows a sample response initiation cost function. The cost function is constructed so that the minimum of the function is at the point at which the solid error curve crosses the dotted elapsed time curve in Figure 2A. We will refer to the error and elapsed time curves as *cost* curves, even though technically, it is their integrals that are subtracted from one another to obtain the cost function.

A key claim in accounting for blocking effects is that the error estimate is noisy, either because the most probable response is not correct or because information about accumulated evidence is not well communicated to response initiation systems. To overcome the unreliability of the error estimate, a sensible strategy in a stationary environment is to average the error estimate being computed for the current stimulus with the error estimate from recent trials. By *stationary environment*, we mean an environment in which item difficulty—and hence, shape of the error curve—does not vary much from one trial to the next. By *average*, we mean that the error curve used to estimate the cost is a combination of the current error curve and a *historical* error curve, which itself is an average of the error curves from recent trials, with the most recent trials weighted more heavily. (The weighting is exponential, which permits a simple updating function that redefines the historical curve to be a weighted average of the previous historical curve and the current error curve.)

According to this framework, blocking effects reflect the adaptation to the stimulus environment. In pure blocks, the historical error curve would be the same as the current stimulus and is depicted in Figure 2C by the leftmost (pure easy block) and rightmost (pure hard block) solid curves. A mixed block, on the other hand, comprises both easy items and hard items. The historical error curve is, therefore, an average of easy and hard error curves and is depicted in Figure 2C by the dotted curve in the middle. This average historical error curve is, in turn, combined with the error curve of the current stimulus, and it is this weighted average of current and historical error curves that is used to estimate the optimal point for initiating a response. In Figure 2C, the error curves in a mixed block are depicted by the dashed curves: the dashed curve on the left for the easy stimulus, and the dashed curve on the right for the hard stimulus.

Figure 3 depicts a *zoomed-in* section of where the elapsed time curve (dotted straight line) intersects the error curves (from left to right) for the easy stimulus in a pure block, the easy stimulus in a mixed block, the hard stimulus in a mixed block, and the hard stimulus in a pure block. The top panel depicts the condition in which zero asymptotes are assumed for all stimuli; that is, subjects are confident of generating a correct response, given sufficient time. (The bottom panel depicts a condition in which the subjective error estimate for hard items asymptotes at a nonzero value and will be discussed later.) The various aspects of the blocking phenomenon are explained as follows. It can be seen that the point of intersection is shifted to the right for an easy stimulus (in this example, from 448 msec in a pure block to 471 msec in a mixed block) and to the left for a hard stimulus (from 539 msec in a pure block to 521 msec in a mixed block), indicating a slowdown for the easy stimulus and a speedup for the hard stimulus in the mixed block, relative to the pure blocks. The difference in latency between an easy stimulus and a hard stimulus in a mixed block (in this example, 471 vs. 521 msec) is explained by the model as reflecting the contribution of the stimulus difficulty of the current trial as a factor in the computation of the error curve used to estimate the optimal RT. Because the historical error curve is updated trial by trial, with greater weight given to
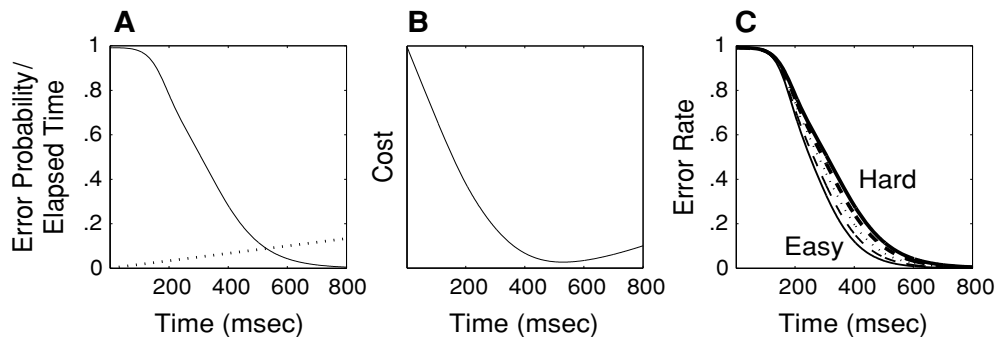

Figure 2. Mechanism of response initiation according to the adaptation-to-the-statistics-of-the-environment model. (A) Error probability estimate (solid) and reaction time cost (dotted) over time. (B) Combined error cost function over time. (C, from left to right) Current error curve for an easy stimulus, averaged error curve for an easy stimulus in a mixed block, historical error curve in a mixed block, averaged error curve for a hard stimulus in a mixed block, and current error curve for a hard stimulus.
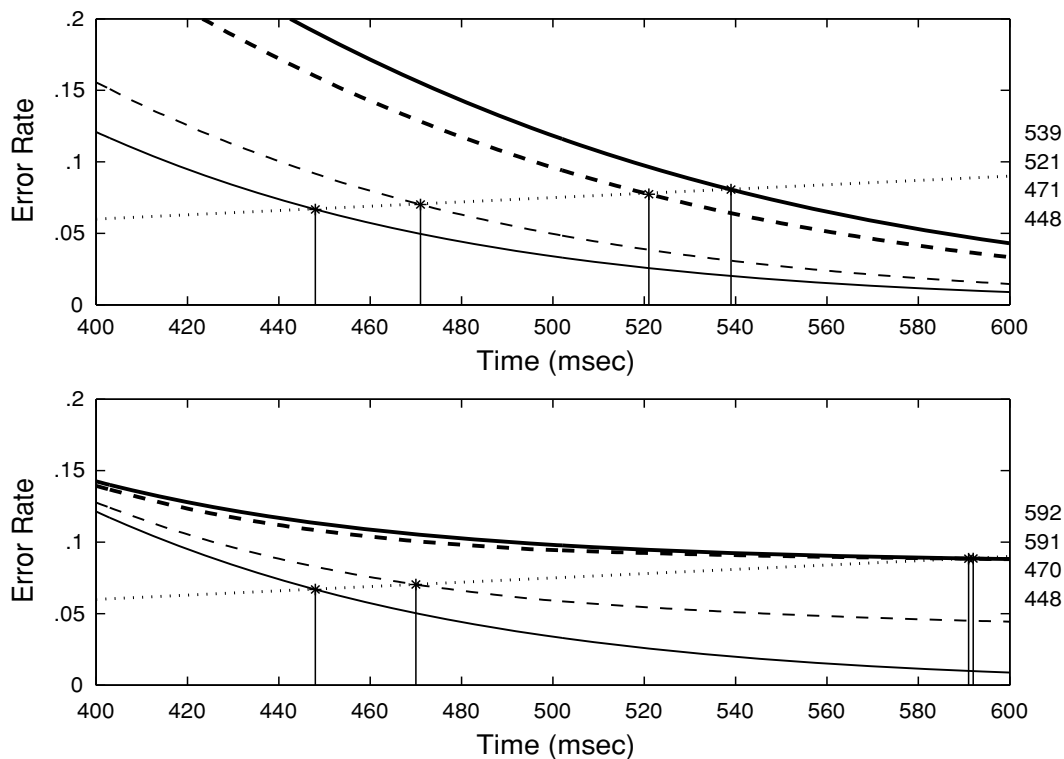
Figure 3. Close-up of the error curves for easy and hard stimuli in pure blocks (solid) and mixed block (dashed), along with the reaction time cost function (dotted). Top panel assumes zero asymptotes for the easy and hard stimuli; bottom panel assumes nonzero asymptote for the hard stimuli. See text for further explanation.

more recent trials, the sequence effect reported by Taylor and Lupker (2001) falls out naturally from the model. Finally, the model correctly predicts the trade-off between speed and accuracy in the blocking effect (the difference between easy and hard items in error rate is *magnified* in the mixed block, relative to the pure blocks), on the basis that responses are emitted at a later point in time for easy stimuli and at an earlier point in time for hard stimuli in the mixed block.

Although Mozer et al. (2003) developed their model to account for the blocking effects observed with the naming task, the proposed mechanism is a general one and is not tied to the temporal dynamics of a specific task. One finding that might be taken to argue against a general mechanism operating in all RT tasks, however, is the variation in the pattern of blocking effects observed. Specifically, whereas both a speedup for hard items and a slowdown for easy items of roughly equal magnitude are observed in the naming task, such is not the case in other, binary classification tasks. In particular, a different pattern has been reported in studies in which the effect of blocking frequency of word targets has been examined in an LDT (e.g., Glanzer & Ehrenreich, 1979; Gordon, 1983; Stone & Van Orden, 1993). In these studies, one group of subjects performed lexical decisions in a block containing only high-frequency words as word targets (pure high-frequency block) and a block containing only

low-frequency words as word targets (pure low-frequency block), and another group of subjects was presented with a block containing both high- and low-frequency words (mixed block). A typical finding observed across the studies (observed in Dorfman & Glanzer, 1988; Glanzer & Ehrenreich, 1979; Gordon, 1983; Stone & Van Orden, 1993, Experiment 1 vs. 3, legal nonword condition) is a pure-block advantage for high-frequency words (easy items) and a weak or nonexistent block type effect for low-frequency words (hard items)—that is, an asymmetric pattern of blocking effects.

The aim of Experiment 1 was to test whether this asymmetric pattern of the blocking effect (a limited speedup for hard items, accompanied by a slowdown for easy items) in the LDT is indeed the norm. To this end, we used the same word stimuli in an LDT (Experiment 1A) and a naming task (Experiment 1B). Experiment 1A also extended the frequency blocking manipulation used in the previous studies in two ways. One was the use of a within-subjects manipulation of block type. The between-subjects manipulation of block type used in the earlier studies is not ideal for establishing the pattern of speedup versus slowdown, due to a large variation in individual RTs. The use of a within-subjects manipulation of block type was assumed to allow a more sensitive assessment of relative speedup and slowdown. Second, the previous studies manipulated only the difficulty of *word* targets, by manipulating fre-

quency. It is well known that the difficulty of lexical discrimination depends on both the nonwords and the words used (see Wagenmakers et al., 2004, for an explicit account of this process). It may be argued, therefore, that the manipulation of frequency alone was not as strong a manipulation of *difficulty* as it could have been. The present Experiment 1A therefore manipulated the difficulty of words and nonwords simultaneously.

EXPERIMENT 1

In Experiment 1A, we investigated the pattern of blocking effects in an LDT. Difficulty of lexical discrimination was manipulated by varying the difficulty of both words and nonwords simultaneously. The difficulty of words was manipulated in terms of frequency: Easy words were of high frequency (e.g., *river*), and hard words were of low frequency (e.g., *scald*). The difficulty of nonwords was manipulated in terms of number of (orthographic) neighbors, using Coltheart, Davelaar, Jonasson, and Besner's (1977) N-metric (i.e., the number of words that can be generated by changing a letter in the same position). It has been found consistently (e.g., Andrews, 1989; Coltheart et al., 1977) that in the LDT, nonwords with many neighbors (high-N nonwords; e.g., *scart*) are responded to more slowly than nonwords with few neighbors (e.g., *crylc*; Andrews, 1989; Coltheart et al., 1977). Thus, the pure easy condition contained high-frequency words and zero-N nonwords; the pure hard condition contained low-frequency words and high-N nonwords. Subjects performed lexical decisions in three blocks: a pure easy block, a pure hard block, and a mixed block in which the easy and hard items were mixed randomly. It was expected that the difference in latencies between the easy and the hard items would be reduced in the mixed block, relative to the pure blocks. The critical question concerned the specific pattern of the blocking effect. In the naming task, the slowdown for easy items and the speedup for hard items in a mixed block are equal. In contrast, in previous lexical decision studies in which frequency blocking has been used, the speedup for hard items (low-frequency words) has been weak or absent.[1] Our central interest was, therefore, which of these symmetric versus asymmetric patterns of blocking would be observed in Experiment 1A.

In addition, Experiment 1B was conducted with the same high-frequency and low-frequency word stimuli as those in Experiment 1A being used in a naming task. On the basis of previous findings of a *symmetric* pattern for the blocking effect (i.e., equal speedup for slow items and slowdown for easy items in a mixed block) typically observed in the naming task, it was expected that this pattern would be found in Experiment 1B.

Experiment 1A
Lexical Decision Task

Method

Subjects. Twenty-seven volunteer Macquarie University students participated in Experiment 1A for course credit. All the subjects were native Australian English speakers.

Design. In Experiment 1A, the LDT was used. The experiment consisted of a 2 (block type: pure vs. mixed) $\times$ 2 (stimulus difficulty: easy vs. hard) factorial design, with both factors manipulated within subjects. The dependent variables were decision latency and error rate. Each subject performed the task in three blocks: pure easy, pure hard, and mixed.

Materials. The critical stimulus materials used in this experiment were 54 high-frequency words, 54 low-frequency words, 54 high-N nonwords, and 54 zero-N nonwords, all of which were five letters long. The high-frequency words had a frequency of at least 100 ($M = 234$) per million, according to Kučera and Francis (1967). Examples are *river*, *floor*, and *study*. The low-frequency words ranged between 1 and 10 occurrences per million, with a mean of 3.39. Examples are *scald*, *eject*, and *colon*. The high-N nonwords had 5 or more neighbors (range, 5–10; $M = 6.39$), all with existing bodies (mean number of body neighbors = 12.41)—for example, *scart*, *ludge*, and *grake*. The zero-N nonwords had no neighbors—for example, *crylc*, *sckaz*, and *sawch*. The word items were selected using the MRC Psycholinguistic Database (Coltheart, 1981; available at the Internet address www.psy.uwa.edu.au/MRCDataBase/uwa_mrc.htm), and the nonword items were selected using the ARC Nonword Database (Rastle, Harrington, & Coltheart, 2002; available at the Internet address www.maccs.mq.edu.au/~nwdb). The easy items are listed in Appendix A; the hard items are listed in Appendix B.

Each item type was divided into three sets containing 18 items each; Sets A, B, and C. Nine list versions were constructed for the purpose of counterbalancing assignment of sets to the three block types—namely, a pure block of easy items (containing 36 high-frequency words and 36 zero-N nonwords), a pure block of hard items (containing 36 low-frequency words and 36 high-N nonwords), and a mixed block (containing 18 of each of the four types of stimuli).

Prior to each test block, the subjects were given 4 practice items and 32 warm-up items that were representative of the block type. Warm-up and practice items were selected according to the same criteria as the test stimuli. These items were not included in the analysis.

Apparatus and Procedure. The subjects were tested in groups of up to 6 at a time, seated approximately 40 cm in front of an NEC V700 monitor, upon which the stimuli were presented. Each subject completed three blocks of trials: one mixed block and two pure blocks. The order of blocks was counterbalanced across subjects, so that an equal number of subjects were presented with each of the three blocks (pure easy, pure hard, and mixed) in each of the three possible positions (first, second, or third). Thus, full counterbalancing of order and version was realized with 27 subjects.

Stimulus presentation and data collection were achieved through the use of the DMDX display system developed by K. I. Forster and J. C. Forster at the University of Arizona (Forster & Forster, 2003), running on Deltacom Pentium III personal computers. The subjects were instructed at the outset of the experiment that they would be presented with a series of letter strings and that their task was to decide whether it was a word or a nonsense word as quickly and accurately as possible. They were instructed to press the right shift key on the keyboard if it was a word and the left shift key if it was a nonword.

Each trial started with the presentation of a target letter string, presented in lowercase letters in 12-point Courier font in the center of the screen. The letter string remained on the screen for a maximum of 2,000 msec or until the subject's response. Following a blank screen for 300 msec, the next trial started. The subjects were given no feedback on either latencies or error rates during the experiment.

Results

For this and the subsequent analyses, the preliminary treatment of data was as follows. Any trial on which a subject made an error was excluded from the latency analy-

sis; remaining latency data were winsorized to the three standard deviation boundary for that subject. In Experiment 1A, this affected 1.44% of the trials. Word targets and nonwords were analyzed separately.

RT and error rates were analyzed using a two-way ANOVA with difficulty (easy vs. hard) and block type (pure vs. mixed) as factors. In the by-subjects analysis, both factors were within-subjects factors; in the by-items analysis, block type was a within-item factor, and response type and difficulty were between-item factors. Effects were considered to be significant when both subject and item analyses were significant at the .05 level. Mean lexical decision latencies and error rates are presented in Table 1.

Word targets. For latency, the main effect of difficulty was highly significant [$F_1(1,26) = 109.61$, $MS_e = 5,344.31$; $F_2(1,106) = 185.29$, $MS_e = 7,558.79$]. The main effect of block type was nonsignificant [$F_1(1,26) = 1.49$, $MS_e = 3,038.06$; $F_2(1,106) = 3.37$, $MS_e = 4,960.83$, $p = .07$]. These two factors interacted [$F_1(1,26) = 11.42$, $MS_e = 1,530.51$; $F_2(1,106) = 5.14$, $MS_e = 4,690.83$]. Critically, simple effects analyses indicated that whereas the 38-msec slowdown for the easy words in the mixed block was significant [$F_1(1,26) = 13.01$, $MS_e = 1,530.51$; $F_2(1,53) = 19.05$, $MS_e = 2,192.47$], the 13-msec speedup for the hard words was not [$F(1,26) = 1.37$, $MS_e = 1,530.51$, $p = .25$; $F_2(1,53) = 0.06$, $MS_e = 7,729.19$, $p = .81$].

For error rate, the main effect of difficulty was highly significant [$F_1(1,26) = 56.70$, $MS_e = 147.32$; $F_2(1,106) = 53.17$, $MS_e = 2.55$]. The main effect of block type was not significant [$F_1(1,26) = 0.99$, $MS_e = 56.75$; $F_2(1,106) = 1.79$, $MS_e = 0.51$]. The interaction between these two factors was nonsignificant by subjects [$F_1(1,26) = 1.50$, $MS_e = 83.96$] but was significant by items [$F_2(1,106) = 4.02$, $MS_e = 0.51$]. The pattern of interaction was opposite to that observed with the latency data, in that the difference between the easy and the hard items was magnified in the mixed block relative to the pure blocks (19.7% vs.15.4%). Simple effects analyses showed that the 3.6% decrease in errors for the easy items in the mixed block did not reach significance by subjects [$F_1(1,26) = 3.05$, $MS_e = 57.58$, $p = .09$] but was significant by items [$F_2(1,53) = 15.76$,

$MS_e = 0.18$]. The 0.7% increase in errors for the hard items in the mixed block was nonsignificant [$F_1(1,26) = 0.08$, $MS_e = 83.13$; $F_2(1,53) = 0.14$, $MS_e = 0.84$].

Nonword targets. For latency, the main effect of difficulty was highly significant [$F_1(1,26) = 306.77$, $MS_e = 3,022.52$; $F_2(1,106) = 370.13$, $MS_e = 4,577.93$]. The main effect of block type was significant [$F_1(1,26) = 4.55$, $MS_e = 3,351.91$; $F_2(1,106) = 3.89$, $MS_e = 5,573.99$, $p = .05$]. These two factors interacted [$F_1(1,26) = 16.01$, $MS_e = 1,872.10$; $F_2(1,106) = 11.62$, $MS_e = 5,573.99$]. Again, the simple effects analyses showed that whereas the (57-msec) slowdown for the easy items in the mixed block was significant [$F_1(1,26) = 23.49$, $MS_e = 1,872.10$; $F_2(1,53) = 19.75$, $MS_e = 4,086.42$], the (9-msec) speedup for the hard items in the mixed block was not [$F_1(1,26) = 0.66$, $MS_e = 1,872.10$, $p = .42$; $F_2(1,53) = 0.82$, $MS_e = 7,061.56$, $p = .37$].

For error rate, the main effect of difficulty was significant [$F_1(1,26) = 29.31$, $MS_e = 182.92$; $F_2(1,106) = 66.96$, $MS_e = 1.30$]. The main effect of block type was also significant [$F_1(1,26) = 4.52$, $MS_e = 64.83$; $F_2(1,106) = 5.99$, $MS_e = 0.79$]. These two factors interacted [$F_1(1,26) = 8.11$, $MS_e = 91.87$; $F_2(1,106) = 15.20$, $MS_e = 0.79$]. As with the word targets, the pattern of the interaction was opposite to that of the latency data in showing that the difference between the easy and the hard items was magnified in the mixed block, relative to the pure blocks (19.3% vs. 8.8%). Simple effects analyses showed that the 2.0% increase in errors for the easy items in the mixed block was nonsignificant [$F_1(1,26) = 0.79$, $MS_e = 65.99$; $F_2(1,53) = 3.76$, $MS_e = 11.79$]. The 8.5% increase in errors for the hard items in the mixed block, on the other hand, was significant [$F_1(1,26) = 10.86$, $MS_e = 90.72$; $F_2(1,53) = 11.71$, $MS_e = 72.18$].

Experiment 1B
Word-Naming Task

Method

Subjects. Twenty-four volunteer Macquarie University students participated in Experiment 1B for course credit. All the subjects were native Australian English speakers.

Design. In Experiment 1B, the word-naming task was used. The experiment consisted of a 2 (block type: pure vs. mixed) × 2 (stimulus difficulty: easy vs. hard) factorial design, with both factors manipulated within subjects. The dependent variables were naming latency and error rate. Each subject performed the task in three blocks: pure easy, pure hard, and mixed.

Materials. The critical stimulus materials used in this experiment were the 54 high-frequency words and 54 low-frequency words used in Experiment 1A. Each item type was divided into two sets containing 27 items each, Sets A and B. Two list versions were constructed for the purpose of counterbalancing assignment of sets to the two block types—namely, a pure block (containing 27 high-frequency words or 27 low-frequency words) and a mixed block (containing 27 high-frequency and 27 low-frequency words).

Prior to each test block, the subjects were given four practice items and two warm-up items that were representative of the block type. Warm-up and practice items were selected according to the same criteria as the test stimuli. These items were not included in the analysis.

Apparatus and Procedure. The subjects were tested individually, seated approximately 40 cm in front of an NEC Multisync 4FG

Table 1
Mean Lexical Decision Latencies (RTs, in Milliseconds)
and Percentages of Error (%E) in Experiment 1A

| | Block Type | | | | | |
| | Pure | | Mixed | | Difference | |
Stimuli	RT	%E	RT	%E	RT[a]	%E
Words						
High frequency	525	5.1	563	1.5	+38*	−3.6
Low frequency	698	20.5	685	21.2	−13	+0.7
Frequency effect	173	15.4	122	19.7		
Nonwords						
Zero N	556	5.5	613	3.5	+57*	−2.0
High N	774	14.3	765	22.8	−9	+8.5
N effect	218	8.8	152	19.3		

[a]A positive difference denotes a slowdown and a negative difference a speedup from a pure block to the mixed block.   *Significant speedup/slowdown at the .05 level.

monitor, upon which stimuli were presented. Each subject completed three blocks of trials: one mixed block and two pure blocks. The order of blocks was counterbalanced across subjects so that half of the subjects were presented with a pure block first and the other half with the mixed block first. Within each group, half of the subjects were presented with the pure hard block before the pure easy block, and the other half in reversed order. Thus, full counterbalancing of order and version was realized with 8 subjects.

Stimulus presentation and data collection were achieved through the use of the DMASTR display system, which was a precursor to the DMDX system (Forster & Forster, 2003). The subjects were instructed at the outset of the experiment that they would be presented with a series of words and that their task was to read aloud each word as quickly and accurately as possible. RTs were recorded using a voice key that triggered when a criterial level of amplitude was reached. Subject errors and possible measurement errors due to inappropriate triggering of the voice key (e.g., coughing) were recorded manually by the experimenter.

Each trial started with the presentation of a target word, presented in lowercase letters in 12-point Courier font in the center of the screen. The target word remained on the screen for a maximum of 2,000 msec or until the subject's response. Following a blank screen for 300 msec, the next trial started. The subjects were given no feedback on either latencies or error rates during the experiment.

Results

The preliminary treatment of data was identical to that in Experiment 1A, and the three standard deviation winsorization procedure affected 1.35% of the trials. Mean naming latencies and error rates are presented in Table 2.

Both naming latency and error rate were analyzed using a 2 (difficulty: easy vs. hard) × 2 (block type: pure vs. mixed) ANOVA. For latency, the main effect of difficulty was significant [$F_1(1,23) = 30.75$, $MS_e = 3,713.29$; $F_2(1,106) = 115.45$, $MS_e = 2,279.88$]. The main effect of block type was nonsignificant [$F_1(1,23) = 0.08$, $MS_e = 55.51$; $F_2(1,106) = 0.06$, $MS_e = 1,382.97$]. These two factors interacted [$F_1(1,23) = 19.83$, $MS_e = 897.33$; $F_2(1,106) = 30.65$, $MS_e = 1,382.97$]. Critically, simple effects analyses indicated that the 28-msec slowdown for the easy words in the mixed block was significant [$F_1(1,23) = 12.53$, $MS_e = 791.45$; $F_2(1,53) = 31.72$, $MS_e = 728.82$], as well as the 26-msec speedup for the hard words [$F(1,23) = 9.38$, $MS_e = 845.15$; $F_2(1,53) = 9.50$, $MS_e = 2,037.13$].

For error rate, the main effect of difficulty was significant [$F_1(1,23) = 38.49$, $MS_e = 10.02$; $F_2(1,106) = 9.39$, $MS_e = 92.65$]. The main effect of block type was significant [$F_1(1,23) = 4.72$, $MS_e = 5.93$; $F_2(1,106) =$

Table 2
Mean Naming Latencies (RTs, in Milliseconds) and Percentages of Error (%E) in Experiment 1B

| Word Stimuli | Block Type | | | | | |
| | Pure | | Mixed | | Difference | |
	RT	%E	RT	%E	RT[a]	%E
High frequency	445	0.5	473	0.5	+28*	0
Low frequency	541	5.6	515	3.4	−26*	+2.2
Frequency effect	96	5.1	42	2.9		

[a]A positive difference denotes a slowdown and a negative difference a speedup from a pure block to the mixed block.   *Significant speedup/slowdown at the .05 level.

4.16, $MS_e = 15.20$]. The interaction between these two factors was significant [$F_1(1,23) = 5.90$, $MS_e = 4.74$; $F_2(1,106) = 4.16$, $MS_e = 15.20$]. The pattern of interaction mirrored that observed with the latency data in that the difference between the easy and the hard items was reduced in the mixed block, relative to the pure blocks (5.1% vs. 2.9%). Simple effects analyses showed that the 2.2% decrease in errors for the hard items in the mixed block was significant [$F_1(1,23) = 5.90$, $MS_e = 9.47$; $F_2(1,53) = 4.77$, $MS_e = 26.50$]. The error rate for the easy items did not differ between the pure and the mixed blocks.

Discussion

Consistent with the results of previous studies, a blocking effect was observed in an LDT (Experiment 1A), with the latency difference between the easy items and the hard items reduced in the mixed block, relative to the pure blocks. (Also consistent with the results of previous studies, a trade-off was observed with error rate, with the difference between the easy and the hard items being *magnified* in the mixed block, relative to the pure blocks.) Critically, the specific form of blocking effect was such that although responses to the easy items slowed down in the mixed block, the speedup for the hard items in the mixed block was small and did not reach statistical significance. This asymmetric pattern is consistent with that observed in previous studies investigating the effects of frequency blocking in the LDT (e.g., Gordon, 1983; Stone & Van Orden, 1993). In contrast, in the naming task (Experiment 1B), the same high- and low-frequency words produced a symmetric pattern—that is, the equal speedup for hard items and slowdown for easy items in the mixed block, relative to the pure blocks, that has been reported consistently with the naming task (e.g., Lupker et al., 1997).

The fact that the pattern of the blocking effect is different for the naming task and the LDT may be taken to argue that the mechanism responsible for blocking effects is different in these two tasks. Indeed, such a view was suggested by Perea and Carreiras (2003), who reported that the pattern of sequence effects (the effect of trial $n-1$ on trial n) in an LDT was different from that observed by Taylor and Lupker (2001) in the naming task. Perea and Carreiras hinted that these different patterns may have their origins in the fact that in a naming task, "there is a gradual accumulation of information about the correct item (i.e., it is an incremental task)" (p. 394), whereas the LDT is "a two-alternative forced choice task in which words are to be distinguished from nonwords" (p. 395). However, the process of evidence accumulation for the purpose of making lexical decisions is generally assumed to be gradual (e.g., Ratcliff et al., 2004). In any case, Perea and Carreiras did not develop their argument beyond describing these task differences, and it is unclear to us how such a putative nonincremental nature of the LDT would cause the slowdown for easy items without the speedup for hard items.

As was mentioned earlier, within the adaptation-to-the-statistics-of-the-environment (ASE) model of blocking effects proposed by Mozer et al. (2003), there is no intrinsic

reason why the different task requirements of the LDT would produce a pattern of blocking effects different from that produced with the naming task. Accordingly, we explored the properties of the model to see whether it could produce the asymmetric pattern. We discovered that the model produces this pattern when the asymptote in the estimated error rate for the hard items is nonzero. We will elaborate on this property below.

As is depicted in Figure 2C and the top panel of Figure 3, when the blocking effects observed with the naming task were modeled, the estimated error rate was assumed to asymptote at zero. To put it another way, it was assumed that the subjects would be confident of generating a correct response given sufficient time. It seems reasonable to assume this in the naming task. Reading aloud is not a difficult task, and error rates are quite low. Importantly, the subjects are confident of being able to generate a correct reading of a letter string, given sufficient time. In contrast, this assumption may not always be met in the LDT. In particular, when the hard items are particularly difficult—as would be the case when an unfamiliar low-frequency word (e.g., *hovel* or *covet*) or a very wordlike nonword (e.g., *shink* or *ludge*) is used as a stimulus, as in Experiment 1A—the subjects may just not be sure whether it is a word that is unfamiliar to them or a made-up item, even if there was no requirement to make a decision quickly. In other words, the subjective error probability would not asymptote at zero, but at a level higher than zero.

The bottom panel of Figure 3 depicts the condition in which the hard items are associated with a nonzero asymptote. To explain the asymmetric pattern of blocking effects depicted here, it needs to be pointed out first that because the error curves are time-varying functions, the notion of averaging is ambiguous. It is possible to average the error rate of points with the same time value, which we refer to as *vertical averaging*; alternatively, it is possible to average the times of points with the same error rate value, which we refer to as *horizontal averaging*. Horizontal averaging cannot be performed in real time during ongoing processing, because it requires that the system "looks ahead" in time. However, vertical averaging is extremely sensitive to noise in regions where the error curve is positively accelerating, such as the point where the error is approaching its asymptotic minimum. Thus, the ASE model uses horizontal averaging for the historical error curves, which need not be performed in real time, and vertical averaging for combining the historical error curve and the current error curve, which is computed as a trial progresses.

With the notion of averaging as just described, the level of asymptotic error probability has an implication for the magnitude of blocking effect. To see that this is the case, consider the historical error curve. Because the error rate in horizontal averaging is undefined for values below the maximum error rate, the model assumes that the historical error curve asymptotes at the nonzero value of the item with the maximum error rate. Psychologically, what this means is that the control mechanism assumes that in a mixed block, on the basis of past history, the accuracy of response will not improve over time beyond the

hardest item experienced recently. As can be seen from the bottom panel of Figure 3, for a hard item, this will make the averaged error curves (averaged over the current and historical error curves) in the pure hard block and the mixed block closer, resulting in the elapsed time curve's intersecting the error curves at points much closer to each other than those in the zero asymptote case. This also fits the intuition that if the items are hard—in the sense that even without the time pressure, subjects cannot be sure whether the response is correct—there is no point in delaying the response, since extra waiting will not improve the accuracy of the response. The main consequence of a nonzero asymptote in a subjective estimate of error probability, therefore, is a more limited speedup of hard items in the mixed block. Thus, the model could produce the asymmetric pattern of blocking effects observed in Experiment 1 if such a nonzero asymptote of estimated error rate is assumed. A corollary of this assumption is that hard items should speed up in the mixed block if their estimated error probability asymptotes at zero.

EXPERIMENT 2

In Experiment 2, we tested the prediction above, derived from the model, by using *easier* hard items. Specifically, we used low-frequency words that were rated as highly familiar (e.g., *spoon* or *thief*), and nonwords (e.g., *flawk* or *droil*) that were less wordlike than the high-N nonwords used in Experiment 1, by virtue of the fact that they had fewer neighbors. We expected that the subjects would know that these familiar words were words for sure and that the less wordlike nonwords were not words and, hence, assumed that these words would be associated with zero asymptotes in estimated error rate.

We should note that the error traces assumed by the model are based on *subjective* estimates of error probability and are not directly observable. Nevertheless, objective error rates provide unbiased estimates of such subjective error probability, and we used the error rates in a nonspeeded paper-and-pencil LDT as an index of asymptotic subjective error probabilities. This task was similar to that used by Paap and colleagues (Paap, Johansen, Chun, & Vonnahme, 2000), in which subjects were presented with a list of letter strings on a sheet of paper and were asked to indicate for each letter string whether it was a word or a nonword. In addition to indicating their decision, we asked the subjects to rate their confidence in their decision. For the easier hard items used in Experiment 2, we expected the error rate to approach zero; as well, we expected the subjects' confidence ratings in their decision to be high, corresponding to a zero asymptote of estimated error rate. In contrast, for the low-frequency words used in Experiment 1, we expected a higher error rate and lower confidence ratings.

Method

Speeded LDT

Subjects. An additional 27 volunteer Macquarie University students participated in Experiment 2 for course credit. None of the subjects had participated in Experiment 1A or 1B.

Design. The design of Experiment 2 was identical to that in Experiment 1. The dependent variables were decision latency and error rate.

Materials. The easy items were identical to those used in Experiment 1. For the hard items, both words and nonwords, which were different from those used in Experiment 1, were selected from the same source. All the items were five letters long. The (hard) word targets had a frequency that was somewhat higher (range, 1–20; $M = 7.78$) than in Experiment 1 (range, 1–10; $M = 3.39$) and had a high familiarity rating (minimum = 490, $M = 541.07$ on a scale of 100–700). Examples are *spoon* and *thief*. N was lower for nonwords (range, 3–5; $M = 3.67$) than for the hard nonwords used in Experiment 1 (range, 5–10; $M = 6.39$). Examples are *droil* and *flawk*. These items are listed in Appendix B.

The procedure for counterbalancing the assignment of items to the pure versus mixed blocks, the order of conditions, and the number of practice and warm-up trials (selected according to the same criteria as those for the test items) were identical to those in Experiment 1A.

Apparatus and Procedure. The apparatus, instructions given to the subjects, timing parameters, and procedure were identical to those in Experiment 1A.

Paper-and-Pencil LDT

Subjects. Twenty-six subjects from the same population as that in Experiments 1A and 1B and 2 participated in this experiment.

Materials. The stimulus materials were the hard items used in Experiments 1A and 1B—namely, the two sets of 54 low-frequency words and 54 nonwords—and are listed in Appendix B.

Apparatus and Procedure. The 108 low-frequency words and 108 nonwords were mixed together and typed in three columns over two pages in alphabetical order, with the characters "$+ / - 3\ 2\ 1$" written next to each item. The subjects were told to circle for each item "$+$" if it was a word and "$-$" if it was a made-up word and then to rate their confidence in their response on a scale of 1 to 3, where $3 = $ *very confident*, $2 = $ *somewhat confident*, and $1 = $ *guessing*. As in Paap et al.'s (2000) study, the subjects were instructed not to work so fast that they might make careless mistakes but, on the other hand, not to agonize over their decision.

Results

Speeded LDT

The treatment of data prior to analysis was identical to that in Experiment 1. The winsorizing procedure affected 1.83% of the trials. Mean lexical decision latencies and error rates are presented in Table 3. Latency and error rate data were analyzed in the same way as in Experiment 1A.

Word targets. For latency, the main effect of difficulty was highly significant [$F_1(1,26) = 72.82$, $MS_e = 1,699.54$; $F_2(1,106) = 99.85$, $MS_e = 2,435.05$]. Difficulty interacted with block type [$F_1(1,26) = 17.92$, $MS_e = 1,557.95$; $F_2(1,106) = 28.38$, $MS_e = 1,498.69$]. Importantly, simple effects analysis showed that the 42-msec slowdown of the easy items in the mixed block was significant [$F_1(1,26) = 15.39$, $MS_e = 1,557.95$; $F_2(1,53) = 34.62$, $MS_e = 1,320.57$], as well as the 22-msec speedup of hard items [$F_1(1,26) = 4.28$, $MS_e = 1,557.95$; $F_2(1,53) = 9.38$, $MS_e = 1,676.81$].

For error rate, the main effect of difficulty was significant [$F_1(1,26) = 13.91$, $MS_e = 9.50$; $F_2(1,106) = 7.01$, $MS_e = 37.67$]. The interaction between difficulty and block type was significant by subjects [$F_1(1,26) = 7.36$, $MS_e = 8.16$] and marginally significant by items

Table 3
Mean Lexical Decision Latencies (RTs, in Milliseconds) and Percentages of Error (%E) in Experiment 2

| | Block Type | | | | | |
| | Pure | | Mixed | | Difference | |
Stimuli	RT	%E	RT	%E	RT[a]	%E
Words						
High frequency	496	2.5	538	1.5	+42*	−1.0
Low frequency	596	3.2	574	5.2	−22*	+2.0
Frequency effect	99	0.7	33	3.7		
Nonwords						
Zero N	549	3.1	571	1.9	+22	−1.2
Medium N	725	10.6	680	12.2	−45*	+1.6
N effect	174	7.5	100	10.3		

[a]A positive difference denotes a slowdown and a negative difference a speedup from a pure block to the mixed block. *Significant speedup/slowdown at the .05 level.

[$F_2(1,106) = 3.46$, $MS_e = 34.72$, $p = .065$]. Simple effects analysis indicated that the 1.03% decrease in errors for the easy words in the mixed block was nonsignificant [$F_1(1,26) = 1.82$, $MS_e = 7.81$, $p = .19$; $F_2(1,53) = 1.81$, $MS_e = 15.77$, $p = .18$]. The 1.96% increase in errors for the hard words in the mixed block was significant by subjects [$F_1(1,26) = 5.28$, $MS_e = 9.79$] but not by items [$F_2(1,53) = 1.92$, $MS_e = 53.67$].

Nonword targets. For latency, the main effect of difficulty was highly significant [$F_1(1,26) = 192.88$, $MS_e = 2,844.83$; $F_2(1,106) = 243.36$, $MS_e = 4,129.33$]. As with word decisions, this factor interacted with block type [$F_1(1,26) = 16.21$, $MS_e = 1,842.58$; $F_2(1,106) = 32.20$, $MS_e = 1,774.56$], indicating the presence of a blocking effect. The 22-msec slowdown of zero-N nonwords in the mixed block approached significance by subjects [$F_1(1,26) = 3.55$, $MS_e = 1,842.58$, $p = .07$] and was significant by items [$F_2(1,53) = 9.69$, $MS_e = 1,179.43$]. Importantly, the 45-msec speedup of medium-N nonwords was significant [$F_1(1,26) = 14.52$, $MS_e = 1,842.58$; $F_2(1,53) = 22.54$, $MS_e = 2,369.69$].

For error rate, the main effect of difficulty was significant [$F_1(1,26) = 16.02$, $MS_e = 133.67$; $F_2(1,106) = 32.64$, $MS_e = 131.01$]. None of the other main effects or interactions was significant [$F_1(1,26) < 1.32$, $p > .26$; $F_2(1,106) = 1.69$, $p > .20$, in all cases].

Paper-and-Pencil LDT

The error rate and confidence rating data are presented in Table 4, and the data for individual items are listed in Appendix B.

As can be seen in Table 4, there was a substantial difference in both error rates and confidence ratings between the sets of hard (low-frequency) words used in Experiments 1A and 2. The number of errors differed significantly between the two sets [$F_2(1,106) = 25.51$, $MS_e = 7.01$], as did confidence ratings [$F_2(1,106) = 35.34$, $MS_e = 33.81$]. The results showed that even when the decisions were not speeded, the subjects were not very accurate at judging the lexical status of the hard items used in Experiment 1A: They were unable to recognize 11.53% of the low-

Table 4
Mean Error Rates and Confidence Ratings for the
Hard Items in the Paper-and-Pencil Lexical Decision Task

| | Words | | | | Nonwords | | | |
	Example	%E	Confidence Rating[a]	%Guess[b]	Example	%E	Confidence Rating[a]	%Guess[b]
Items								
Experiment 1	Aloof Banal	11.53	2.24	8.69	Skell Ludge	17.30	−1.74	23.22
Experiment 2	Spoon Thief	1.65	2.89	0.34	Flawk Droil	10.88	−2.02	21.65

[a]Confidence ratings ranged from/to −3 to +3, with −3 indicating *very confident that it is a made-up word* and +3 indicating *very confident it is a word*.   [b]% Guess refers to the percentage of subjects who rated their confidence at +1 or −1.

frequency words, and a substantial percentage (8.69%) of responses were classified as *guessing*. In contrast, they recognized almost all of the low-frequency words used in Experiment 2 correctly: The error rate was a low 1.65%. Importantly, the proportion of *guessing* responses (0.34%) was substantially lower for the easier hard words used in Experiment 2 and was close to zero. We take this as evidence that the estimated error rate for the more difficult hard items asymptotes above zero, whereas that for the easier hard words asymptotes at zero or close to zero.

The data for nonwords also showed that the items used in Experiment 1A were more difficult than those used in Experiment 2. Specifically, the number of errors differed significantly between the two sets [$F_2(1,106) = 13.27$, $MS_e = 5.65$], as did confidence ratings [$F_2(1,106) = 11.04$, $MS_e = 0.19$]. However, the percentage of responses classified as *guessing* was quite high (21.65%) for the medium-*N* nonwords used in Experiment 2 and was not very different from that for the more difficult nonwords (23.22%). The percentage is certainly higher than what would be expected from the view that the estimated error rate asymptotes at zero. One explanation for this result is that unlike words, there are no definitive bases for classifying a nonword: If it is an orthographically legal letter string, it is always possible that it is an unknown word. Since discrimination was quite difficult in this paper-and-pencil LDT, because it contained only low-frequency words as word targets, the subjects may have been particularly conservative in their decisions for nonwords.

Discussion

In Experiment 2, we set out to test a novel prediction of the ASE model—namely, that a speedup of hard items would emerge in an LDT if the estimated subjective error probability for the hard items asymptotes at zero. To this end, we replaced the hard items used in Experiment 1A with the easier words (low-frequency words with high familiarity ratings) and easier nonwords (medium-*N* nonwords). The results confirmed this prediction and showed a significant speedup of hard items in the mixed block. To our knowledge, this is the first time a significant speedup of hard items has been found in the LDT. Also consistent with the prediction of the ASE model, these *easier* hard items produced lower error rates than did the hard items used in Experiment 1A.

With regard to the paper-and-pencil LDT, the main result was that the error rate was substantially higher and the confidence ratings substantially lower for the less familiar low-frequency words used in Experiment 1A than for the low-frequency words used in Experiment 2. The high error rate for words used in Experiment 1A (11.53% failed to be recognized as words) may be a cause of concern for those who lament the literacy level of university students. We note that Paap et al. (2000) also reported a low level of performance in their university student subjects. In their study, subjects failed to recognize 26.8% of low-frequency words in the range of 1–5 per million. (These items—e.g., *mirth*, *salve*, or *pert*—were referred to by Paap et al. as "clinkers.") We will make two comments about this low level of recognition. One is that our results are not atypical. Second, as was pointed out earlier, Mozer et al.'s (2003) model is concerned with the *subjective* estimate of error probability, and not with the *objective* error rate, but nonetheless, the two are related. In this regard, Paap et al.'s assertion that "many very low-frequency words may have incomplete and unstable representations" (p. 1709) is relevant. Paap et al. pointed out that, for example, only 47% of their subjects correctly identified *dined* as a word, yet these subjects would probably have no problem understanding the word in the context "The Republicans ate crow when they *dined* at the White House." (Similarly, we believe that our subjects would recognize one of our star clinkers, *covet* [which only 61.5% of our subjects recognized as a word] in the context of "Thou shalt not *covet* thy neighbor's wife.") Paap et al. suggested that the difficulty that their subjects had with these words is, therefore, likely to be a temporary accessibility problem, rather than a permanent absence of these words in the subjects' lexicon. Such intermittent accessibility is likely to reduce subjects' confidence in their judgment of the word's lexical status, unlike items that are permanently absent from the subject's lexicon (which would be confidently rejected as nonwords). Thus, the high objective error rate accompanied by a low confidence rating observed with the low-frequency words used in Experiment 1A observed in the present paper-and-pencil LDT is consistent with the unstable representation of these items in the subjects' lexicon. Furthermore, such unstable representation is likely to lead to a nonzero asymptote in subjective estimate of error probability when lexical decisions are made. Previous studies of frequency blocking

(e.g., Glanzer & Ehrenreich, 1979; Gordon, 1983; Stone & Van Orden, 1993) did use unfamiliar low-frequency words; thus, the lack of speedup of low-frequency words observed in these studies may also have its origin in the nonzero asymptote of estimated error probability.

EXPERIMENT 3

Experiment 2 was the first experiment to show a significant speedup for hard items in an LDT, in contrast to those in previous studies (e.g., Glanzer & Ehrenreich, 1979; Gordon, 1983; Stone & Van Orden, 1993). It may be argued, however, that numerically, the slowdown for high-frequency words (42 msec) was substantially larger than the speedup for low-frequency words (22 msec). Although this pattern was reversed in the nonword data (showing a greater speedup for hard items than the slowdown for easy items), nevertheless it was felt that the finding of a significant speedup for hard items was in need of replication. This, then, was the aim of Experiment 3. In addition, we wished to test whether it is possible to observe a significant speedup for low-frequency words without simultaneously varying the difficulty of nonword targets: After all, previous studies of frequency blocking held the nonwords constant between blocks. To this end, the relatively familiar low-frequency words used in Experiment 2 were used as the hard word targets, and the easy and hard blocks differed only in terms of the frequency of word targets. We used "easy," zero-N nonwords as foils in all the blocks, because using harder nonwords would make the discrimination between low-frequency words and nonwords difficult and could elevate the asymptotic subjective error probability estimate.

Method

Subjects. An additional 48 volunteer Macquarie University students participated in Experiment 3 for course credit. None of the subjects had participated in Experiment 1 or 2.

Design. The design of Experiment 3 was identical to that in Experiments 1A and 2, except that nonword type was held constant across all block types. Thus, for nonwords, the *difficulty* factor refers to the difficulty of the word targets in the block (i.e., the difficulty of lexical discrimination), not to the difficulty of the nonword targets. The dependent variables were decision latency and error rate.

Materials. Forty-eight high-frequency words, 48 low-frequency words, and 96 zero-N nonwords were used as stimuli in this experiment. The word targets were identical to those used in Experiment 2, except for 6 each of high- and low-frequency words not used in Experiment 3. All the nonwords were five-letter long, zero-N nonwords like the *easy* nonwords used in Experiments 1 and 2 and were selected from the same source, using the same criteria. These items are listed in Appendix B.

Unlike in Experiments 1A and 2, the mixed block contained twice as many items (24 high-frequency words, 24 low-frequency words, and 48 nonwords) as the pure blocks (24 words from either frequency band and 24 nonwords). The assignment of items to the pure and mixed blocks was counterbalanced across every pair of subjects, and the order of conditions was counterbalanced across every 4 subjects, so that full counterbalancing was achieved across every 8 subjects. The number of practice and warm-up trials (selected according to the same criteria as the test items) were identical to those in Experiments 1A and 2.

Apparatus and Procedure. The apparatus, instructions given to the subjects, timing parameters, and procedure were identical to those in Experiments 1 and 2.

Results and Discussion

The treatment of data prior to analysis was identical to that in Experiment 1. The winsorizing procedure affected 1.73% of the trials. Mean lexical decision latencies and error rates are presented in Table 5.

Word targets. For latency, the main effect of difficulty was highly significant [$F_1(1,47) = 73.40$, $MS_e = 1,683.02$; $F_2(1,94) = 50.62$, $MS_e = 2,466.29$]. The main effect of block type was nonsignificant [$F_1(1,47) = 0.22$, $MS_e = 2,386.37$, $p = .64$; $F_2(1,94) = 1.26$, $MS_e = 767.48$, $p = .27$]. Difficulty interacted with block type [$F_1(1,47) = 4.57$, $MS_e = 1,033.37$; $F_2(1,94) = 8.09$, $MS_e = 767.48$], indicating that a blocking effect occurred. Most important, simple effects analysis indicated that the 13-msec speedup for low-frequency words was significant [$F_1(1,47) = 4.08$, $MS_e = 1,033.37$; $F_2(1,47) = 8.24$, $MS_e = 732.21$]. The 7-msec slowdown for high-frequency words, on the other hand, was nonsignificant [$F_1(1,47) = 1.00$, $MS_e = 1,033.37$, $p = .32$; $F_2(1,47) = 1.42$, $MS_e = 802.75$, $p = .24$].

For errors, the main effect of difficulty was significant [$F_1(1,47) = 7.30$, $MS_e = 29.88$; $F_2(1,94) = 5.80$, $MS_e = 2.16$]. The main effect of block type was nonsignificant

Table 5
Mean Lexical Decision Latencies (RTs, in Milliseconds) and
Percentages of Error (%E) in Experiment 3

Stimuli	Pure		Mixed		Difference	
	RT	%E	RT	%E	RT[a]	%E
Word						
High frequency	550	3.9	557	1.9	+7	−2.0
Low frequency	611	4.6	598	5.5	−13*	+0.9
Frequency effect	61	0.7	41	3.6		
Nonword						
Mixed with high-frequency words	612	7.4	641	7.6	+29*	−0.2
Mixed with low-frequency words	674	7.5	655	8.1	−19	+0.6

[a]A positive difference denotes a slowdown and a negative difference a speedup from a pure block to the mixed block.   *Significant speedup/slowdown at the .05 level.

$[F_1(1,47) = 0.72, MS_e = 21.57, p = .40; F_2(1,94) = 1.02, MS_e = 0.86, p = .31]$. The interaction between difficulty and block type was significant $[F_1(1,47) = 6.13, MS_e = 16.10; F_2(1,94) = 6.59, MS_e = 0.86]$. As in the previous experiments, the pattern of interaction was opposite to that of latency, indicating that the frequency effect in error rate was *magnified* in the mixed block (3.6%), relative to the pure blocks (0.7%). Simple effects analysis indicated that the 2% reduction in error rate for high-frequency words in the mixed block was significant $[F_1(1,47) = 5.96, MS_e = 16.10; F_2(1,47) = 9.97, MS_e = 0.55]$, but the 0.9% increase for low-frequency words in the mixed block was not $[F_1(1,47) = 1.12, MS_e = 16.10, p = .30; F_2(1,47) = 0.89, MS_e = 1.17, p = .35]$.

Nonword targets. The main effect of difficulty was significant $[F_1(1,47) = 27.05, MS_e = 2,531.35; F_2(1,94) = 34.14, MS_e = 2,119.41]$. Difficulty interacted with block type $[F_1(1,47) = 9.38, MS_e = 3,024.24; F_2(1,94) = 4.31, MS_e = 2,119.41]$. Simple effects analysis indicated that the 29-msec slowdown from the pure high-frequency block to the mixed block was significant $[F_1(1,47) = 6.77, MS_e = 3,024.24; F_2(1,47) = 4.06, MS_e = 2,236.70]$. The 19-msec speedup from the pure low-frequency block to the mixed block was nonsignificant $[F_1(1,47) = 2.99, MS_e = 3,024.24, p = .09; F_2(1,47) = 0.80, MS_e = 2,002.13, p = .38]$.

For errors, none of the main or interaction effects was significant $[F_1(1,47) < .13, p > .71; F_2(1,94) < 1.78, p > .19]$. As can be seen from Table 4, the error rate remained constant at about 7%–8% across all the conditions.

The results of Experiment 3 showed that, replicating Experiment 2, there was a significant speedup for hard items (low-frequency words). This was not due to a greater bias toward responding *words* in the mixed block, relative to the pure hard block, because a similar speedup (albeit not statistically significant) was observed with the nonword targets from the pure hard block to the mixed block. Because Experiment 3 manipulated only the difficulty (frequency) of word targets, we conclude that the manipulation of both word and nonword difficulty is not necessary to observe significant speedup of low-frequency words in an LDT. Instead, the results of Experiment 3 suggest that the weak or absent speedup for low-frequency words in the previous experiments was due to nonzero asymptotes in subjective estimates of error probability for these items. The substantially lower error rates for the hard items in Experiment 3, relative to the hard items used in Experiment 1A (which did not show a significant blocking advantage in the LDT), are also consistent with this view.

Some readers may feel unconvinced by our claim that we have demonstrated speedup for hard items in the mixed block in Experiments 2 and 3 by using easier hard items, because the speedup for low-frequency words in Experiment 2 (22 msec) and Experiment 3 (13 msec), which was statistically significant, was numerically similar to the speedup for the low-frequency words in Experiment 1A (13 msec), which was statistically nonsignificant. However, this ignores the fact that the effect of difficulty (frequency) was much larger in Experiment 1. Expressed as a percentage of the frequency effect observed in the pure blocks, the speedup for low-frequency words was 7.5% in Experiment 1A, 22.2% in Experiment 2, and 21.3% in Experiment 3. Similarly expressed, for nonwords, the speedup for hard items was 4.1% in Experiment 1A, 25.9% in Experiment 2, and 30.6% in Experiment 3. We are therefore confident that the emergence of a speedup for hard items in the mixed block observed in the LDT when using *easier* hard items is real.

GENERAL DISCUSSION

The present study investigated the speedup and slowdown for easy and hard items (high- and low-frequency words) in the LDT based on the composition of items in a trial block. Previous studies in which word frequency has been used as the blocking variable have reported a slowdown for easy items but little speedup for hard items. This pattern was confirmed in Experiment 1A, using relatively difficult hard items (low-frequency words in the range of 1–10 occurrences per million and high-N nonwords) typical of the hard items used in previous studies. This asymmetric pattern is different from the symmetric pattern of blocking effects (equal speedup for hard items and slowdown for easy items) typically observed with the naming task and confirmed in Experiment 1B. The ASE model of blocking effects proposed by Mozer et al. (2003) can explain this variation in the pattern of blocking effects in terms of a difference in the subjective estimate of error probability. Specifically, the model predicts an essentially symmetric blocking effect on the assumption that subjective error probability asymptotes at zero and provides an explanation for an asymmetric pattern of blocking in terms of an above-zero asymptote. Experiments 2 and 3 tested the ASE model's prediction by using relatively easier hard items (low-frequency words in the range of 1–20 per million that had high familiarity ratings). Error rate and confidence-rating data for a paper-and-pencil lexical decision test were consistent with the assumption that these *easier* low-frequency words had a lower asymptotic rate of subjective error probability, close to zero, which was lower than that of the *harder* low-frequency words used in Experiment 1A. The RT data in the LDT supported the model's prediction, by showing a significant speedup for hard items. To our knowledge, this is the first time low-frequency words have shown a significant frequency-mixing advantage in an LDT.

In the present study, both asymmetric and symmetric patterns of blocking effects were observed in the LDT, depending on the nature of *hard* items: Responses for hard items that were both slower and (subjectively) more error prone than those for the easy items did not speed up in a mixed block, whereas responses for hard items that were primarily just slower showed a speedup that was equal to the slowdown for easy items. The distinction between responses for hard items that are primarily slow and those for hard items that are *both slow and error prone* has not been emphasized previously in the RT literature, but we believe that it is a useful distinction.

For one thing, the distinction seems to characterize the difference between the naming task and the binary decision tasks reasonably well. Typically, stimuli presented in a naming task (whether they are pictures, words, or nonwords) are easy to name, and the error rates to even the hard items are not high. Consequently, subjects are generally confident of being able to generate a correct response in this task. In contrast, in many binary decision tasks, the hard items are associated with substantially higher error rates, as well as with longer RTs, than are the easy items. Thus, the asymptote in subjective estimate of error probability for the hard items is likely to be above zero for these binary decision tasks. This characterization of naming task and binary decision tasks could explain why blocking effects have typically been symmetric in the former and asymmetric in the latter, as in the studies in which the LDT has been used to investigate the frequency-blocking effect (e.g., Gordon, 1983; Stone & Van Orden, 1993), as well as with other tasks, such as the semantic categorization task (e.g., Kiger & Glass, 1981).

Another potential implication of the distinction between the hard items for which responses are and are not error prone concerns the shape of the RT distribution. It has been reported that the RT distributions are more (positively) skewed in the LDT than in the naming task; moreover, this underlies the greater word frequency effect typically observed in the former than in the latter task (Andrews & Heathcote, 2001; Balota & Spieler, 1999). The theoretical implication of this finding has been much debated—in particular, whether the greater skew reflects a task-specific decision component (e.g., a *familiarity detection mechanism*) external to the word recognition process. An alternative possibility is that the greater skew is a consequence of a higher asymptotic level of subjective error probability estimate. It can be seen from the comparison of the top and bottom panels of Figure 3 that everything else being equal, hard items with an above-zero asymptote (bottom panel) produce longer RTs than do hard items with a zero asymptote (top panel). This is because the error curve for those items approaches the asymptote more slowly than does that for the items with a zero asymptote and, hence, the optimal RT will be shifted further to the right. Consistent with this possibility, Andrews and Heathcote (2001) found that the skew in the RT distribution for the same low-frequency words was just as large in a semantic categorization task as in the LDT; moreover, the semantic categorization task showed a greater skew for the high-frequency words than did the LDT. In all of these cases, the greater skew was associated with higher error rates. Thus, the ASE model might be able to provide an explanation for the task difference in the extent of the positive skew in terms of difference in asymptotic error estimates, as well as for the task difference in the pattern of blocking effects. Although the ASE model is currently silent with regard to the distribution of RTs (its scope currently being limited to mean RTs), by building a variance component to the error curves, the model could generate predictions regarding RT distributions. Such extensions to the model remain on our future agenda.

REFERENCES

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., & HEATHCOTE, A. (2001). Distinguishing common and task-specific processes in word identification: A matter of some moment? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 514-544.

BALOTA, D. A., & SPIELER, D. H. (1999). Word frequency, repetition, and lexicality effects in word recognition tasks: Beyond measures of central tendency. *Journal of Experimental Psychology: General*, **128**, 32-55.

CHATEAU, D., & LUPKER, S. J. (2001). Strategic effects in word naming: Examining the route-emphasis versus time-criterion accounts. *Journal of Experimental Psychology: Human Perception & Performance*, **29**, 139-151.

COLTHEART, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology*, **33A**, 497-508.

COLTHEART, M., DAVELAAR, E., JONASSON, J. T., & BESNER, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 539-555). Hillsdale, NJ: Erlbaum.

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

FORSTER, K. I., & FORSTER, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, **35**, 116-124.

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.

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

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

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.

LUPKER, S. J., KINOSHITA, S., COLTHEART, M., & TAYLOR, T. (2003). Mixing costs and mixing benefits in naming words, pictures, and sums. *Journal of Memory & Language*, **49**, 556-575.

MEYER, A. S., ROELOFS, A., & LEVELT, W. J. M. (2003). Word length effects in object naming: The role of a response criterion. *Journal of Memory & Language*, **48**, 131-147.

MOZER, M. C., COLAGROSSO, M. D., & HUBER, D. E. (2002). A rational analysis of cognitive control in a speeded discrimination task. In T. Dietterich, S. Becker, & Z. Ghahramani (Eds.), *Advances in neural information processing systems 14* (pp. 51-57). Cambridge, MA: MIT Press.

MOZER, M. C., KINOSHITA, S., & DAVIS, C. (2003, November). *Control and adaptation of the speed–accuracy tradeoff: A mathematical theory*. Paper presented at the 44th Annual Meeting of the Psychonomic Society, Vancouver.

PAAP, K. R., JOHANSEN, L. S., CHUN, E., & VONNAHME, P. (2000). Neighborhood frequency does affect performance in the Reicher task: Encoding or decision? *Journal of Experimental Psychology: Human Perception & Performance*, **26**, 1691-1720.

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., CARREIRAS, M., & GRAINGER, J. (2004). Blocking by word frequency and neighborhood density in visual word recognition: A task-specific response criteria account. *Memory & Cognition*, **32**, 1090-1102.

Rastle, K., & Coltheart, M. (1999). Serial and strategic effects in reading aloud. *Journal of Experimental Psychology: Human Perception & Performance*, **25**, 482-503.

Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC Nonword Database. *Quarterly Journal of Experimental Psychology*, **55A**, 1339-1362.

Rastle, K., Kinoshita, S., Lupker, S. J., & Coltheart, M. (2003). Cross-task strategic effects. *Memory & Cognition*, **31**, 867-876.

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

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

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

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

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

Wagenmakers, E.-J., Steyvers, M., Raaijmakers, J. G. W., Shiffrin, R. M., van Rijn, H., & Zeelenberg, R. (2004). A model for evidence accumulation in the lexical decision task. *Cognitive Psychology*, **48**, 332-367.

NOTE

1. Rastle et al. (2003, Experiment 1B) conducted a lexical decision experiment that was identical in design to that of Experiment 1. In that experiment, a blocking effect was observed, but neither the slowdown for easy items nor the speedup for hard items was specifically reported. Subsequent analysis showed that the slowdown for easy items (37 msec for words and 35 msec for nonwords) was statistically significant [$F_1(1,23) = 14.39$, $MS_e = 1,154.22$, $F_2(1,39) = 24.08$, $MS_e = 1,155.65$, and $F_1(1,23) = 10.27$, $MS_e = 1,438.26$, $F_2(1,39) = 17.82$, $MS_e = 1,434.66$, for words and nonwords, respectively]. In contrast, the speedup for hard items was numerically smaller (19 msec for both words and nonwords) and was nonsignificant by subjects [for words, $F_1(1,23) = 1.50$, $MS_e = 2,636.38$], although significant by items [$F_2(1,39) = 7.13$, $MS_e = 1,941.19$; for nonwords, $F_1(1,23) = 1.03$, $MS_e = 4,207.20$, and $F_2(1,39) = 2.26$, $MS_e = 2,080.91$].

APPENDIX A
Easy Stimuli (Used in Experiments 1A, 2, and 3)

High-Frequency Words

river, floor, great, start, blood, large, hotel, heart, leave, money, early, stand, brown, music, speak, class, death, paper, right, stage, story, cause, clear, below, level, heavy, house, night, later, table, force, earth, field, month, green, party, study, north, child, front, final, drive, bring, light, mouth, least, space, peace, short, close, black, horse, every, group

Zero-*N* Nonwords

crylc, zoate, swesk, cwelb, snaup, spolv, spyph, gnylp, dieph, dwagn, joobe, valgn, phurv, sawch, rurmb, crozz, quool, bieph, dwylm, klyke, glulf, preth, squyl, gnymp, gnooc, whylb, raugg, phyke, zourf, smulv, sckaz, kworg, yainn, puinn, splyv, knyzz, gnorg, mewve, twobb, ghlik, jorfe, felmb, fewve, whowt, cwoac, twysk, psarr, phylt, preib, reuck, blumn, sorle, plewc, dweip

APPENDIX B
Hard Stimuli Used in Experiments 1A and 2
and Their Paper-and-Pencil Lexical Decision Data

	Experiment 1A				Experiment 2		
Item	Mean Confidence Rating	%E	%Guess	Item	Mean Confidence Rating	%E	%Guess
			Word Targets				
scald	2.46	3.85	7.69	spoon	2.92	0	0
marsh	2.69	3.85	7.69	clock	3.00	0	0
drone	1.62	19.23	23.08	towel	3.00	0	0
chalk	2.77	3.85	0.00	steal	3.00	0	0
filth	2.50	7.69	11.54	marry	2.77	3.85	0
adore	2.96	0.00	0.00	salad	3.00	0	0
plush	2.23	7.69	11.54	float	3.00	0	0
lunge	1.77	15.38	19.23	broom	2.81	3.85	0
balmy	1.31	26.92	30.77	trash	2.77	3.85	0
canal	2.58	3.85	15.38	nurse	2.81	3.85	0
bland	2.42	3.85	7.69	blush	2.69	3.85	0
colon	1.96	15.38	7.69	thief	3.00	0	0
greed	3.00	0.00	0.00	gravy	2.96	0	0
hinge	2.35	7.69	0.00	spice	2.65	7.69	3.85
goose	3.00	0.00	0.00	pedal	2.92	0	0
flirt	2.54	7.69	0.00	clown	3.00	0	0
vinyl	2.92	0.00	3.85	nasty	3.00	0	0
ebony	2.27	11.54	11.54	arrow	3.00	0	0
maple	2.58	3.85	7.69	apple	3.00	0	0
vicar	1.77	23.08	7.69	toast	2.96	0	0
lurch	1.96	19.23	15.38	juice	2.77	3.85	0
chick	2.96	0.00	0.00	chalk	2.77	3.85	0
groan	2.65	3.85	0.00	shout	2.58	7.69	0
crate	2.19	11.54	15.38	brake	2.69	3.85	3.85
gouge	1.92	7.69	26.92	flame	3.00	0	0
fairy	2.96	0.00	0.00	frost	2.92	0	0
flare	2.77	3.85	0.00	steam	3.00	0	0
banal	−1.00	73.08	30.77	smash	3.00	0	0
poach	2.31	11.54	7.69	skate	2.96	0	0
hoist	2.58	3.85	7.69	dense	2.92	0	0
covet	0.81	38.46	34.62	jelly	3.00	0	0
droop	2.08	7.69	15.38	lemon	3.00	0	0
eject	3.00	0.00	0.00	linen	2.65	7.69	3.85
aloof	2.38	7.69	3.85	spear	2.65	3.85	3.85
scalp	2.54	7.69	7.69	slang	2.85	3.85	3.85
broth	1.65	23.08	15.38	skull	3.00	0	0
scarf	3.00	0.00	0.00	glove	2.77	3.85	0
groom	2.77	3.85	0.00	elbow	3.00	0	0
vowel	2.85	3.85	3.85	slope	2.77	3.85	0
cheat	3.00	0.00	0.00	stair	2.96	0	0
amaze	3.00	0.00	0.00	nerve	2.96	0	0
medal	2.96	0.00	0.00	onion	3.00	0	0
lucid	2.23	11.54	3.85	cheat	3.00	0	0
graft	1.73	23.08	19.23	ankle	3.00	0	0
drake	1.69	19.23	15.38	fancy	3.00	0	0
hovel	0.31	50.00	19.23	peach	2.96	0	0
carve	2.54	7.69	7.69	purse	3.00	0	0
easel	0.58	42.31	3.85	grape	2.77	3.85	0
barge	1.62	23.08	11.54	puppy	3.00	0	0
fling	2.35	11.54	7.69	hobby	3.00	0	0
brunt	1.54	23.08	15.38	apron	2.77	3.85	0
crave	3.00	0.00	0.00	tiger	3.00	0	0
flute	3.00	0.00	0.00	salty	3.00	0	0
pluck	1.73	19.23	7.69	tribe	2.27	11.54	0

APPENDIX B (Continued)

	Experiment 1A				Experiment 2		
Item	Mean Confidence Rating	%E	%Guess	Item	Mean Confidence Rating	%E	%Guess
				Nonword Targets			
scart	−1.62	19.23	11.54	cleam	−1.85	7.69	34.62
ludge	−0.77	42.31	30.77	clinn	−2.69	0.00	15.38
broon	−2.15	11.54	19.23	clume	−1.92	15.38	23.08
sweak	−1.85	11.54	26.92	frane	−1.65	19.23	26.92
clane	−1.54	19.23	26.92	frook	−2.46	3.85	15.38
crine	−1.54	23.08	23.08	groon	−2.00	11.54	26.92
swate	−1.77	19.23	30.77	guive	−2.12	3.85	23.08
scall	−1.15	30.77	34.62	krail	−2.38	7.69	11.54
farge	−1.48	23.08	23.08	lurse	−1.58	19.23	19.23
grake	−1.81	15.38	23.08	pleab	−2.38	3.85	19.23
vaint	−1.77	23.08	23.08	scalm	−1.73	15.38	42.31
vatch	−2.31	0.00	19.23	shoit	−2.42	0.00	15.38
slock	−2.00	15.38	19.23	stass	−2.27	3.85	23.08
jouth	−1.73	19.23	19.23	teigh	−2.15	7.69	7.69
snape	−0.62	38.46	23.08	thete	−2.54	3.85	15.38
slone	−1.23	26.92	26.92	thref	−2.69	0.00	0.00
hount	−1.31	30.77	30.77	trock	−2.08	11.54	19.23
swink	−1.46	19.23	26.92	zoach	−2.46	3.85	11.54
scalt	−1.96	11.54	23.08	barch	−2.31	3.85	11.54
putch	−1.54	23.08	15.38	blafe	−2.77	0.00	7.69
pight	−2.00	11.54	19.23	blime	−1.62	15.38	15.38
darge	−2.46	7.69	15.38	brait	−1.65	19.23	34.62
prawl	−0.73	38.46	30.77	droil	−1.81	15.38	34.62
clant	−1.27	23.08	30.77	drong	−1.50	23.08	26.92
fench	−0.85	34.62	15.38	flawk	−2.08	11.54	23.08
frack	−1.85	19.23	23.08	floth	−2.04	7.69	19.23
spope	−2.00	11.54	11.54	foach	−2.42	3.85	19.23
shink	−1.08	30.77	26.92	gurse	−1.88	11.54	15.38
flate	−1.42	23.08	26.92	hidge	−1.69	15.38	23.08
spart	−1.65	19.23	19.23	junge	−1.62	15.38	26.92
drace	−2.04	7.69	23.08	rutch	−2.04	11.54	15.38
stant	−1.58	23.08	42.31	seash	−2.15	7.69	19.23
natch	−1.42	23.08	34.62	slive	−1.77	19.23	19.23
footh	−2.42	3.85	15.38	steef	−1.77	19.23	26.92
speet	−2.15	7.69	26.92	stoom	−2.23	0.00	34.62
swake	−1.85	11.54	26.92	virth	−2.08	3.85	30.77
stonk	−2.04	7.69	38.46	blobe	−2.50	0.00	15.38
trank	−2.23	7.69	15.38	carse	−1.73	19.23	34.62
prown	−1.77	19.23	23.08	cronk	−1.62	23.08	23.08
haint	−2.23	0.00	34.62	curge	−2.27	11.54	19.23
crose	−2.31	7.69	23.08	drell	−2.15	7.69	19.23
spave	−1.92	15.38	19.23	flage	−1.38	19.23	30.77
tream	−1.38	23.08	23.08	furse	−1.92	15.38	19.23
tooch	−2.31	3.85	19.23	harch	−1.46	19.23	38.46
skack	−2.19	7.69	11.54	moutt	−2.35	3.85	19.23
pribe	−1.35	26.92	19.23	nolse	−2.27	7.69	11.54
frawl	−1.23	26.92	34.62	preek	−2.23	7.69	23.08
criss	−1.85	11.54	30.77	smint	−0.88	34.62	30.77
prace	−1.73	19.23	30.77	smoge	−1.77	15.38	26.92
dight	−2.00	11.54	11.54	snove	−1.88	15.38	26.92
skell	−2.31	3.85	23.08	sporm	−2.04	11.54	23.08
snain	−2.69	0.00	0.00	stath	−2.50	0.00	11.54
prope	−2.19	11.54	15.38	trawn	−1.00	30.77	26.92
skake	−2.00	11.54	15.38	wooth	−2.50	3.85	15.38