

## A diffusion model account of criterion shifts in the lexical decision task

Eric-Jan Wagenmakers<sup>a,\*</sup>, Roger Ratcliff<sup>b</sup>, Pablo Gomez<sup>c</sup>, Gail McKoon<sup>b</sup>

<sup>a</sup> Department of Psychology, University of Amsterdam, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands

<sup>b</sup> The Ohio State University, 291 Psychology Building, 1835 Neil Avenue, Columbus, Ohio 43210, USA

<sup>c</sup> Department of Psychology, DePaul University, 2219 North Kenmore Avenue, Chicago, IL 60614-3504, USA

Received 4 June 2006; revision received 18 April 2007

Available online 26 June 2007

### Abstract

Performance in the lexical decision task is highly dependent on decision criteria. These criteria can be influenced by speed versus accuracy instructions and word/nonword proportions. Experiment 1 showed that error responses speed up relative to correct responses under instructions to respond quickly. Experiment 2 showed that responses to less probable stimuli are slower and less accurate than responses to more probable stimuli. The data from both experiments support the diffusion model for lexical decision [Ratcliff, R., Gomez, P., & McKoon, G. (2004a). A diffusion model account of the lexical decision task. *Psychological Review*, 111, 159–182]. At the same time, the data provide evidence against the popular deadline model for lexical decision. The deadline model assumes that “nonword” responses are given only after the “word” response has timed out—consequently, the deadline model cannot account for the data from experimental conditions in which “nonword” responses are systematically faster than “word” responses.

© 2007 Elsevier Inc. All rights reserved.

**Keywords:** Diffusion model; Deadline model; Lexical decision; Response criteria; MROM; DRC

Over the last 20 years, the study of visual word recognition has made extensive use of the lexical decision task. This task requires participants to classify letter strings either as words or as nonwords (e.g., JOM), usually under the instruction to do so “as fast as possible without making errors”. The lexical decision task is often used to measure the ease with which words are activated or retrieved from lexical memory. However, the lexical decision task is not a pure measure of the ease with which lexical information becomes available, as a wide variety of

decisional and strategic factors have been shown to exert a powerful effect on task performance (e.g., Balota & Chumbley, 1984). For instance, classification performance is affected by the instruction to respond either accurately or fast (e.g., Grainger & Jacobs, 1996, p. 519). Also, the time it takes a participant to respond “word” to a stimulus letter string depends heavily on list composition, that is, on lexical characteristics of the experimental stimuli other than the presented letter string (e.g., Brown & Steyvers, 2005; Glanzer & Ehrenreich, 1979; Grainger & Jacobs, 1996; Ratcliff, Gomez, & McKoon, 2004a; Ratcliff, Van Zandt, & McKoon, 1999, Experiment 2; Wagenmakers et al., 2004); for instance, “word” responses are generally slower and less accurate when the accompanying nonword stimuli are very similar

\* Corresponding author.

E-mail address: [EJ.Wagenmakers@gmail.com](mailto:EJ.Wagenmakers@gmail.com) (E.-J. Wagenmakers).

The diffusion model offers an account of the lexical decision task that is fundamentally different from the one provided by the deadline model. In the diffusion model, “nonword” responses are mediated by the same decision mechanism that leads to “word” responses. Response criteria remain fixed during stimulus processing, and it is possible to obtain independent estimates of the ease of lexical processing and the setting of decision

We will show that the data from Experiments 1 and 2 are consistent with the diffusion model, but are inconsistent with the deadline model. When the deadline model is constrained to provide reasonable estimates for response accuracy, it incorrectly predicts that “word” responses are faster than “nonword” responses, regardless of the experimental manipulation of response criteria, and regardless of whether the presented letter string is a word or a nonword. Furthermore, the deadline model is unable to account for the ubiquitous right-skew of the RT distributions.

The mean rate of approach to a boundary is the drift rate  $v$  (“vee”), and the variation of sample paths around

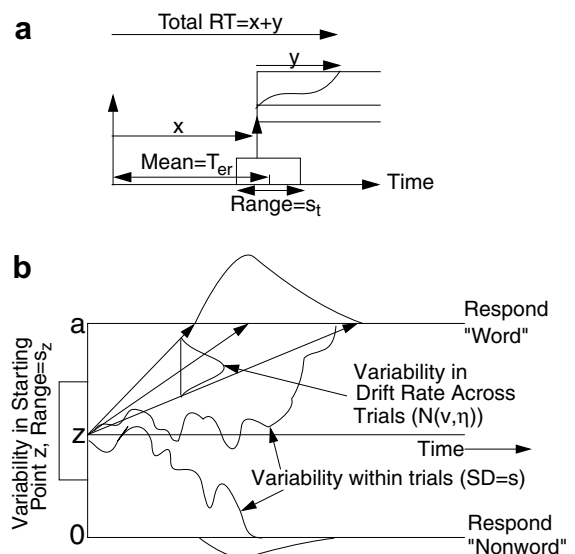


Fig. 1. The diffusion model and its parameters. See text for details.

this mean, called “within-trial” variability, is described by the diffusion coefficient  $s^2$ . This variability allows processes with the same drift rate to reach the same boundary at different times; it also allows processes to reach the wrong boundary by mistake, yielding error responses (the two undulating lines in Fig. 1, panel B). A criterion is placed on the distribution of drift rate values such that word stimuli and nonword stimuli generally have positive and negative drift rates, respectively (for a discussion of the drift rate criterion see Ratcliff, 1978, 1985; Ratcliff et al., 1999). Drift rate is also a function of the quality of the information extracted from the stimulus: when stimuli are relatively difficult to classify, as is the case for low frequency words or nonwords that are orthographically very similar to words, the absolute value of drift rate is relatively low. When stimuli are relatively easy to classify, as is the case for high frequency words or nonwords that are orthographically dissimilar to words, the absolute value of drift rate is relatively high.

Speed–accuracy tradeoffs occur when the boundaries are moved farther apart to produce slower and more accurate responses or closer together to produce faster and less accurate responses. Besides within-trial variability in drift rate, the model assumes across-trial variability in drift rate and across-trial variability in starting point. Drift rates vary across trials to reflect variability across nominally equivalent items (e.g., high frequency words).

In the diffusion model, components of processing other than the decision process, (e.g., encoding and response execution) are summarized into one parameter,  $T_{er}$ , which represents the mean duration of these non-decision processes. Like the other parameters just mentioned, the duration of the non-decision processes is assumed to vary across trials. In Fig. 1, panel A, the total RT generated by the model is a sum of the non-decision time  $x$  and the decision time  $y$ .

To summarize, the parameters of the diffusion model are: mean drift rate  $v$ ; within-trial variability in drift rate,  $s^2$  ( $s$  is a scaling parameter which is set to 0.1 in all fits); across-trial variability in drift rate, which is assumed to have a normal distribution with standard deviation  $\eta$ ; boundary separation  $a$ ; mean starting point  $z$ ; across-trial variability in  $z$ , which is assumed to be uniformly distributed with range  $s_z$ ; the mean time for non-decision RT components  $T_{er}$ ; and the across-trial variability in  $T_{er}$ , which is assumed to be uniformly distributed with range  $s_r$  (for more details see e.g., Ratcliff, 1978, 2002; Ratcliff et al., 2004a; Ratcliff, Thapar, Gomez, & McKoon, 2004b; Ratcliff & Tuerlinckx, 2002).

### Predictions of the diffusion model

The diffusion model makes several qualitative predictions, that is, predictions that hold regardless of the particular values for the parameters (cf. Ratcliff, 2002). For

instance, it follows from the geometry of the model that it can only predict right-skewed RT distributions, for both “word” responses and “nonword” responses, and for both correct responses and for error responses. The right-skew occurs because differences in high values of average drift on a trial produce small changes in RT while the same size differences in low values of drift rate produce large changes in RTs (cf. Ratcliff & Rouder, 1998, p. 348). The diagonal lines in Fig. 1, panel B that begin at the starting point and terminate at the top boundary illustrate this principle. The same size vertical difference between pairs of these lines leads to small differences for the shortest RTs and larger differences for the longer RTs.

A second prediction is that both correct and incorrect responses speed up when stimuli are easier to classify (i.e., when the absolute value of drift rate increases). That is, the diffusion model predicts that errors to relatively easy stimuli such as high frequency words are faster than errors to relatively difficult stimuli such as low frequency words. To appreciate the generality of this prediction, note that when the starting point  $z$  is equidistant from the two response boundaries and when there is no across-trial variability in drift rate and starting point, the model predicts that RT distributions for correct and error responses are identical (e.g., Laming, 1973, p. 192, footnote 7). In practical applications, the diffusion model comes with across-trial variability in both drift rate and starting point, and consequently the model no longer predicts that the RT distributions for correct and error responses are exactly identical. Nevertheless, the model still predicts that when drift rate increases (i.e., task difficulty decreases), both correct responses and error responses will speed up.

As mentioned above, the across-trial variability in drift rate (Ratcliff, 1978) and starting point allow the diffusion model to account for errors that are either faster or slower than correct responses. When across-trial variability in drift rate is sufficiently large, error responses are slower than correct responses (for details see Ratcliff et al., 1999, 2004a; Ratcliff & Rouder, 1998, Fig. 2). In contrast, sufficiently large across-trial variability in starting point causes error responses to be faster than correct responses: processes starting near the error boundary hit it with shorter RTs and greater probability than processes starting near the correct boundary, and their weighted sum gives faster errors than correct responses (Laming, 1968).

In a particular experimental situation, whether error responses are faster or slower than correct responses, or whether their relative speed varies across conditions of the experiment, depends on the relative amounts of across-trial variability in drift rate and starting point (and how large they are relative to the magnitude of the separation between boundaries and the magnitudes of the drift rates). Consider the situation in which

participants lower their response thresholds in order to follow instructions to respond faster. The diffusion model predicts that both correct responses and error responses will speed up, but that the increase in speed is larger for error responses. The explanation is that under speed stress, boundary separation is relatively small, and this increases the impact of variability in starting point and decreases the impact of variability in drift rate.

The above prediction is consistent with results from lexical decision experiments reported in Ratcliff et al. (2004a), who showed that fast participants (i.e., subjects with relatively low response thresholds) have fast errors and slow participants (i.e., subjects with relatively high response thresholds) have slow errors. Experiment 1 tests this prediction more directly using a within-subjects design in which response thresholds are manipulated through instructions and feedback.

### The deadline model for lexical decision

Two widely cited computational models of lexical decision performance are the multiple read-out model (MROM; Grainger & Jacobs, 1996) and the dual route cascaded model (DRC; Coltheart et al., 2001). MROM and DRC use the same mechanism for lexical decision; in particular, both models assume that “nonword” responses are given when a *temporal deadline* is exceeded (cf. Swenson, 1972; Yellott, 1971). The temporal deadline mechanism marks a major conceptual divide between MROM and DRC on the one hand and the diffusion model on the other.

In deadline models for lexical decision such as MROM and DRC, activation from sub-lexical units such as features (i.e., parts of letters) is transmitted to letters, and subsequently to whole word representations. The flow of activation between features, letters, and word units is governed by a local connectionist model widely known as the *interactive activation model* (McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982).

In the deadline model, it is assumed that during processing of a presented letter string, the system has access to and keeps track of three quantities: (1) the level of activation corresponding to each word representation; (2) the overall level of activation in the mental lexicon, summed over all word representations; and (3) the time since stimulus onset. Consequently, the deadline model posits three response criteria, all of which fluctuate from trial-to-trial according to a normal distribution. First, a “word” response is given when any of the representations for individual words reaches a threshold level of activation (cf. Morton, 1969). This threshold is termed the M-criterion and is the same for all word representations. The setting of the M-criterion is not under the system’s control. Second, a “word” response is given when

the total amount of activation in the lexicon reaches a threshold, and this threshold is termed the  $\Sigma$ -criterion. In contrast to the M-criterion, the  $\Sigma$ -criterion is under the system’s control. For instance, under instructions that stress speed over accuracy, the system lowers the  $\Sigma$ -criterion, allowing fast “word” responses to be made at the cost of increased erroneous classifications of nonword stimuli. Third, a “nonword” response is given when neither the M-criterion nor the  $\Sigma$ -criterion has been reached before a certain time criterion  $T$ . That is, a “nonword” response is based on the absence of evidence for a “word” response—after a certain amount of processing has been completed without detecting a word, the system times out and decides that the stimulus letter string is a nonword. Like the  $\Sigma$ -criterion, the  $T$ -criterion is not fixed and so can be lowered when response speed is stressed (cf. Swenson, 1972, p. 30), allowing fast “nonword” responses at the cost of more errors on word stimuli.

The main problem for the deadline model as formulated above is that it predicts that when participants correctly respond “nonword” at the same preset temporal deadline, all types of nonwords should have the same correct RTs. However, experiments have shown that nonwords that are orthographically very similar to words are correctly responded to slower than nonwords that are not very similar to words (e.g., Coltheart, Davelaar, Jonasson, & Besner, 1977). For example, correct responses to nonword letter strings such as DRAPA are slower than correct responses to nonword letter strings such as PDRAA. In order for the deadline model to account for this effect of nonword-to-word similarity, an adjustment is required. The remedy is to assume that the  $T$ -criterion is amendable to change during processing of the stimulus (Coltheart et al., 1977). Specifically, when early in processing considerable support is detected for the stimulus being a word, as would be the case for word-like nonwords, the system reacts by lowering the  $\Sigma$ -criterion and increasing the  $T$ -criterion, effectively facilitating “word” responses and inhibiting “nonword” responses. Thus, the  $T$ -criterion is increased for word-like nonwords, slowing down the correct “nonword” response.

At first consideration, the deadline model appears to make several qualitative predictions, for example, that “nonword” responses to high frequency words are slower than “nonword” responses to low frequency words (e.g., Ratcliff et al., 2004a, p. 178). However, because the deadline model is complex (in the same way as the diffusion model), intuitions are often incorrect and therefore we studied the behavior of the deadline model by Monte Carlo simulations. An additional advantage of this procedure is that it allows the deadline model to be tested against the complete set of observed phenomena simultaneously: the relative speeds of correct and error responses, the shapes of RT distributions,

response accuracy, and the effects of experimental manipulations.

In sum, the deadline model provides an account of performance in the lexical decision task that is different from the one provided by the diffusion model. In the deadline model, “nonword” responses originate when the system times out on the “word” response. Response criteria can change during stimulus processing, and lexical information may interact with response criteria. In the diffusion model, no fundamental distinction between “word” and “nonword” responses exists. Response criteria remain fixed during stimulus processing, and it is possible to obtain independent estimates of the rate of extraction of lexical information and the placement of response criteria. It is of course possible that the deadline model’s more flexible and complex account of lexical processing is warranted by the data. To address this issue, we will fit both the diffusion model and the deadline model to the data from two experiments.

### Experiment 1: Speed versus accuracy instructions

When the diffusion model was first applied to lexical decision (Ratcliff et al., 2004a), the focus was on experimental manipulations that affect the quality of lexical information. These experimental manipulations included word frequency and the use of pseudowords such as DRAPA versus random letter strings such as KDFEU. In contrast, Experiments 1 and 2 reported here focus on experimental manipulations that are thought to affect response criteria but not the quality of lexical information.

In Experiment 1, subjects were instructed either to respond as quickly as possible or to respond as accurately as possible. Speed versus accuracy instructions alternated across blocks of trials. In addition, word frequency was varied from high to low to very low (mean frequency values of 323.37, 4.41, and 0.38 per million; Kucera & Francis, 1967).

The goal for the diffusion model was to account for all aspects of the data for each condition of the experiment, and this includes both response probability and the RT distributions for correct and error responses. In the diffusion model, the effects of varying levels of stimulus difficulty are explained by changes only in drift rate, not any other components of the model. Ratcliff et al. (2004a) demonstrated that, with this assumption, the empirical effects of word frequency and words versus nonwords are well-described by the model. We expected to replicate this finding in this experiment. Hence, among the high, low, and very low frequency word conditions and the nonword condition, only drift rate was free to vary. As mentioned earlier, the diffusion model can account for the speed–accuracy tradeoff with changes solely in response criteria. Therefore, between the speed and accu-

racy conditions, only the distance between the nonword boundary and the starting point and the distance between the starting point and the word boundary were free to vary. All other parameters of the model were held constant across all the experimental conditions.

With the boundaries of the decision process closer together with speed instructions, responses should be faster and less accurate. Also, with the boundaries closer together, the amount of variability in the starting point is a larger proportion of the distance between the boundaries, and thus it has a larger effect compared to across-trial variability in drift rate than it would with the boundaries farther apart. Consequently, the diffusion model predicts that by moving from accuracy to speed instructions, error responses should speed up relative to correct responses.

In the deadline model, speed instructions decrease both the  $\Sigma$ -criterion and the T-criterion. Word frequency affects the resting level of activation, such that high frequency words start out closer to the M-criterion (e.g., Grainger & Jacobs, 1996, p. 541). The simulations presented below show how these assumptions affect the predictions of the deadline model.

### Method

#### Participants

Fifteen undergraduate students of Northwestern University participated for course credit in this single-session experiment. All participants were native speakers of English.

#### Stimulus materials

The stimuli were taken from Ratcliff et al. (2004a). The word stimuli consisted of 814 high frequency words with frequencies ranging from 78 to 10,600 occurrences per million (mean = 323.37,  $SD$  = 641.42, Kucera & Francis, 1967), 858 low frequency words with frequencies of 4 and 5 occurrences per million (mean = 4.41,  $SD$  = .49), and 741 very low frequency words with frequencies of 1 or 0 occurrences per million (mean = 0.38,  $SD$  = .59). All the very low frequency words occurred in the Webster’s Ninth Collegiate Dictionary (1990), and they were screened by three Northwestern undergraduate students—any words any one of the three students did not know were eliminated. For each word, a nonword was created by randomly replacing all vowels by other vowels (except for “u” after “q”), resulting in a total of 2413 nonwords.

#### Design

The experiment consisted of a sequence of 22 blocks. Speed versus accuracy instruction was manipulated between blocks such that blocks in which speeded performance was stressed alternated with blocks in which accurate performance was stressed. The first two blocks

(the first block stressing accuracy, the second block stressing speed) were for practice. Word frequency was manipulated within the 20 experimental blocks such that each block contained an equal proportion of high frequency, low frequency, and very low frequency items, and an equal proportion of nonword items derived from high frequency, low frequency, and very low frequency words. Each block contained an equal number of word and nonword stimuli. The two practice blocks each contained 15 stimuli, and the 20 experimental blocks each contained 96 stimuli. Thus, participants were tested on a total of 1950 letter strings. The stimuli were randomly selected without replacement, and no participant was presented with both a word and the nonword derived from it. The stimuli were presented in random order.

### Procedure

Stimuli were presented on a PC screen with responses collected from the keyboard. Stimulus presentation and response recording were controlled by a real-time computer system.

Participants received verbal instructions explaining the lexical decision task, the speed–accuracy requirements, and the alternation of speed and accuracy blocks. Participants were instructed to press the ‘/’ key with their right index finger when they believed the presented letter string to be an English word and to press the ‘z’ key with their left index finger when they did not believe the presented letter string to be an English word. In the accuracy blocks, each of which was preceded by the message “Try to respond accurately”, the feedback message “ERROR” was presented for 800 ms after every erroneous response. In the speed blocks, each of which was preceded by the message “Try to respond fast”, the feedback message “TOO SLOW” was presented for 800 ms after every trial for which the response latency exceeded 750 ms. In both the accuracy blocks and the speed blocks, anticipatory responding was discouraged by the 800 ms presentation of the feedback message “TOO FAST” after every trial for which the response latency was shorter than 200 ms. The response–stimulus interval was 150 ms.

The experimenter supervised performance during the first two practice blocks. After each experimental block, the participant had a self-paced break.

### Results

Table 1 lists the main empirical results. Approximately 2% of all trials were excluded from the analyses, either because the participant pressed an invalid key (i.e., any other key than ‘/’ or ‘z’) or because the response was diagnosed as an outlier (i.e., responses shorter than 300 ms or longer than 2500 ms). The results for words and nonwords were analyzed separately, as were the results for correct RT, error RT, and response

Table 1

Percentage errors, median correct RT (in ms), and median error RT for high frequency (HF) words, low frequency (LF) words, very low frequency (VLF) words, and nonwords (NW) as a function of speed versus accuracy instructions, as observed in the data from Experiment 1 and as obtained from the diffusion model

Stimulus	Observed data		Diffusion model	
	Accuracy	Speed	Accuracy	Speed
HF				
% Errors	1.9 (.004)	7.4 (.011)	0.5	6.5
Correct RT	564 (16)	471 (8)	569	489
Error RT	563 (40)	441 (10)	612	468
LF				
% Errors	6.0 (.010)	17.2 (.020)	5.0	16.4
Correct RT	636 (18)	510 (7)	639	508
Error RT	653 (48)	480 (11)	734	491
VLF				
% Errors	14.7 (.014)	29.1 (.023)	13.6	26.0
Correct RT	674 (18)	525 (8)	695	518
Error RT	760 (51)	498 (13)	789	504
NW				
% Errors	4.3 (.008)	11.4 (.015)	5.5	14.3
Correct RT	655 (19)	508 (9)	642	501
Error RT	718 (41)	488 (10)	748	504

Standard errors are given in parentheses.

accuracy. For word stimuli, the analysis of variance (ANOVA) included speed versus accuracy instruction and word frequency as within-subjects variables in a repeated measures design. In order to make the statistical analysis consistent with the fits of the diffusion model, which are based on quantiles, our statistical analyses are based on the .5 quantile of the RT distribution (i.e., median RT).

For word and nonword stimuli, instructions to focus on speed decreased accuracy and decreased RT, for correct and error responses. Also, as word frequency decreased, both correct and error responses slowed and accuracy decreased. Instructions to respond quickly decreased RT for error responses more than they decreased RT for correct responses. Further, error RTs for easy stimuli were shorter than error RTs for difficult stimuli. The next four paragraphs provide more detailed statistical analyses.

For word stimuli, instructions to respond fast reduced median correct RT,  $F(1, 14) = 95.21$ ,  $MSE = 3567$ ,  $p < .001$ , and decreased accuracy,  $F(1, 14) = 85.30$ ,  $MSE = 0.003$ ,  $p < .001$ . The effect of word frequency was significant both for median correct RT,  $F(2, 28) = 224.91$ ,  $MSE = 231$ ,  $p < .001$ , and for accuracy,  $F(2, 28) = 116.86$ ,  $MSE = 0.002$ ,  $p < .001$ . The linear contrast confirmed that an increase in word frequency is accompanied by a decrease in median RT,  $F(1, 14) = 383.46$ ,  $MSE = 260$ ,  $p < .001$ , and an increase in accuracy,  $F(1, 14) = 138.00$ ,  $MSE = 0.005$ ,  $p < .001$ .

Instructions to respond fast likewise caused a decrease in median RT for incorrect responses to word stimuli (i.e., the median error RTs),  $F(1,10) = 24.87$ ,  $MSE = 20,541$ ,  $p < .01$ . Word frequency also affected median error RTs,  $F(2,20) = 13.41$ ,  $MSE = 7849$ ,  $p < .001$ , and the linear contrast confirmed that an increase in word frequency is associated with a decrease in median error RT,  $F(1,10) = 21.58$ ,  $MSE = 9640$ ,  $p < .01$ .

Table 1 shows that speed versus accuracy instructions affected the relative speed of errors. Compared to median RT for correct responses, median RT for error responses was relatively long when subjects were told to respond accurately, and it was relatively short when subjects were told to respond quickly. However, four subjects did not make any errors for high frequency words, and because these participants were excluded from the ANOVA, the overall analysis including high frequency, low frequency, and very low frequency words just failed to reach significance at the .05 level,  $F(1,10) = 4.76$ ,  $MSE = 9111$ ,  $p = .054$ . When high frequency words were left out of the analysis, the effect of speed instructions on the relative speed of errors did reach the .05 level,  $F(1,14) = 6.90$ ,  $MSE = 13,845$ ,  $p < .05$ , as was the case when the analysis was performed over all four stimulus types (i.e., high frequency words, low frequency words, very low frequency words, and nonwords),  $F(1,10) = 8.17$ ,  $MSE = 11,523$ ,  $p < .05$ .

For nonword stimuli, the ANOVA included speed versus accuracy instruction as a within-subjects variable in a repeated measures design. Instructions stressing speed lowered accuracy,  $F(1,14) = 24.04$ ,  $MSE = 0.002$ ,  $p < .001$ , and caused a decrease in both median correct RT and median error RT [ $F(1,14) = 126.49$ ,  $MSE = 1290$ ,  $p < .001$  and  $F(1,14) = 36.17$ ,  $MSE = 10,974$ ,  $p < .001$ , respectively]. Further, speed versus accuracy instructions affected the relative speed of errors,  $F(1,14) = 6.84$ ,  $MSE = 7474$ ,  $p < .05$ , as median error RT for nonwords was relatively slow (compared to median correct RT for nonwords) when accuracy was stressed, and relatively fast when speed was stressed (cf. Table 1).

#### Diffusion model analysis

The results from Experiment 1 are qualitatively in good agreement with the predictions of the diffusion model discussed earlier. Specifically, the data show that the relative speed of error RTs increased when speed was stressed. In order to test whether the diffusion model also provides a good *quantitative* account of the present data, the model was fit to response proportions and to RT distributions for correct responses and error responses, for each experimental condition. Specifically, the model was fit to the .1, .3, .5, .7, and .9 quantiles of the group RT distributions. The group RT distributions

were obtained by averaging the quantiles from the individual participants' RT distributions. The diffusion model can be fit using several methods (cf. Ratcliff & Tuerlinckx, 2002)—here we used the  $\chi^2$  method because it provides the best balance between robustness and the ability to recover parameter values (for details see Ratcliff & Tuerlinckx, 2002).

Table 1 shows a comparison between the data in the 16 conditions and the results from the model for error rates and the .5 quantiles (i.e., the medians), confirming that the model can account for the observed results. Note that the diffusion model appears to be in closer correspondence to the data for correct responses than for error responses. This is due to the fact that error latencies are based on relatively few observations, and consequently receive relatively little weight in the  $\chi^2$  fitting procedure. Examination of the standard errors (see the numbers in parentheses in Table 1) confirms that the latencies for error responses are estimated less precisely than the latencies for correct responses.

Table 2 gives the best-fitting parameter values of the diffusion model for each condition. Recall that the only parameter free to vary among the word and nonword conditions was drift rate  $v$ : drift rate was larger for higher frequency words than for lower frequency words and drift rate was negative for nonwords. The only parameters free to vary between the speed conditions and the accuracy conditions were boundary separation  $a$  and starting point  $z$ . The effects of the  $a$  and  $z$  values interact with the effects of across-trial variabilities in starting point and drift rate. With the boundaries relatively close together (that is, with speed instructions), the amount of variability in the starting point is proportionally large relative to boundary separation and so error responses tend to be faster than correct responses. With the boundaries farther apart, variability in the starting point is proportionally smaller and so the effect

Table 2

Best-fitting parameter values for the diffusion model for the data from Experiment 1

Parameters	Accuracy instruction	Speed instruction
$a$	0.159	0.084
$z$	0.079	0.040
$T_{er}$	0.408	-
$s_z$	0.048	-
$s_t$	0.144	-
$v(\text{HF})$	0.452	-
$v(\text{LF})$	0.275	-
$v(\text{VLF})$	0.180	-
$v(\text{NW})$	-0.264	-
$\eta$	0.116	-

Note. HF, high frequency word; LF, low frequency word; VLF, very low frequency word; and NW, nonword. A hyphen indicates that the parameter value was constrained to be identical to the one in the adjacent column. See text for details.

of across-trial variability in drift rates dominates, with the consequence that error responses tend to be slower than correct responses.

Fig. 2, top-left panel, plots the RT distributions for correct responses under accuracy instructions. For each stimulus type (i.e., high frequency words, low frequency words, very low frequency words, and nonwords), the  $x$ -axis shows the probability of a correct response. The five vertical  $X$ s correspond to the observed RT quantiles. The separation between the  $X$ s at the top (i.e., the .7 and .9 quantiles) is much larger than the separation between the dots at the bottom (i.e., the .1 and .3 quan-

tiles), showing the general finding that RT distributions are skewed to the right. The grey dots scattered around the  $X$ s indicate the variability in the data, as obtained by a bootstrap procedure (Efron & Tibshirani, 1993) in which sampling with replacement was performed simultaneously on the level of participants and on the level of trials (for details see Ratcliff et al., 2004a). The five vertical  $+$ s indicate the quantile RT predictions of the diffusion model. The dispersion along the  $y$ -axis and the  $x$ -axis corresponds to variability in the RT quantiles and variability in the response probabilities, respectively (see also Ratcliff, Thapar, & McKoon, 2006, p. 357).

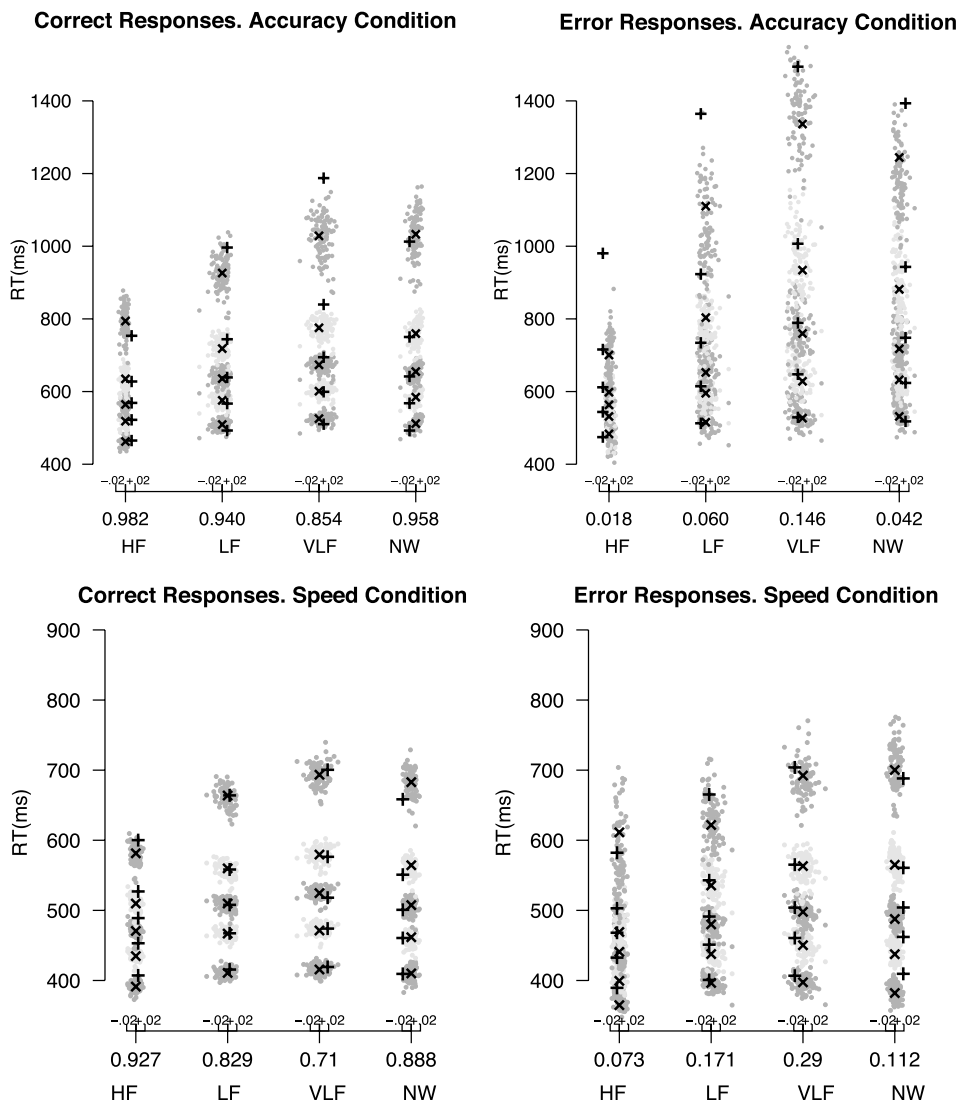


Fig. 2. Empirical ( $X$ s) and predicted ( $+$ s) .1, .3, .5, .7, and .9 quantiles for RT distributions in Experiment 1. The grey dots show variability from bootstrap simulations from the data. HF, high frequency word; LF, low frequency word; VLF, very low frequency word; and NW, nonword. The  $x$ -axis shows response accuracy. Top-left panel: correct responses in the accuracy condition; top-right panel: incorrect responses in the accuracy condition; bottom-left panel: correct responses in the speed condition; bottom-right panel: incorrect responses in the speed condition.

Fig. 2, top-right panel, plots the RT distributions for incorrect responses under accuracy instructions. Fig. 2, bottom-left and bottom-right panels, shows the RT distributions for correct and error responses, respectively, under speed instructions. The relatively good agreement between the data and the model further demonstrates that the models qualitative predictions are supported by a detailed and accurate quantitative account. The model's predictions generally fall within the range of variability of the data indicated by the grey dots. The fit is particularly good when instructions stress speed (bottom panels). When instructions stress accuracy, the top panels of Fig. 2 show that the model fits the higher quantiles worse than it does the lower quantiles. However, these high quantiles are also the most variable, especially for error responses.

By comparing the distributions of the correct responses in the top-left panel of Fig. 2 to the distributions of error responses in the top-right panel, it is immediately apparent that errors are slower than correct responses when instructions stress accurate responding. By comparing the bottom two panels it is also obvious that the differences between correct and error RT distributions are attenuated when instructions stress speed, and, in fact, that error RTs are slightly faster than correct RTs. Fig. 2 also shows that the distributions for "nonword" responses, whether correct or in error, are skewed to the right, much like the distributions of responses to low frequency and very low frequency words (cf. Ratcliff, 2002).

#### *Deadline model analysis*

To study the behavior of the deadline model, we repeatedly presented the model with a subset of the words and nonwords from Experiment 1. For each stimulus presentation, we monitored the number of cycles before a decision threshold was reached, and the identity of that threshold (i.e., the M-criterion and the  $\Sigma$ -criterion for a "word" response, and the T-criterion for a "nonword" response). The details of the simulation are presented below.

Note that, as mentioned earlier, the interactive activation model (McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982) prescribes the changes in activation of sub-lexical and lexical units from the time since stimulus onset. The three thresholds of the deadline model effectively monitor the output of the interactive activation model and allow the system to stop processing information and give a response. This means that the set of parameters for the deadline model comprises both the parameters of the interactive activation model and the parameters of the deadline decision structure.

#### *Materials and procedure*

All letter strings presented to the model were four letter words and nonwords that were used in Experiment 1

and that did not include the letters X, Q, or Z. The constraints of string length (i.e., only four-letter strings) and letter identity (i.e., no X, Q, or Z) were imposed by the version of the interactive activation model that was used.<sup>1</sup> Consequently, the set of letter strings presented to the deadline model consisted of (a) 114 high frequency words with frequencies higher than 100 per million (mean = 594,  $SD = 1090$ ); (b) 97 low frequency words with frequencies between 4 and 5 per million (mean = 4.5,  $SD = .5$ ); and (c) 98 nonwords. In order to obtain a reliable indication of the model's predictions and reduce the impact of noise in the simulations, 10,000 items were randomly selected with replacement from each of the three stimulus categories. The 30,000 items were then presented to the model one by one.

#### *Parameters*

The parameters of the interactive activation model were set at their default values (Grainger & Jacobs, 1996; McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982). Feature-to-letter excitation was set to .005, and feature-to-letter inhibition was equal to .15. Letter-to-word excitation was .07, and letter-to-word inhibition equaled .04. Finally, word-to-letter feedback excitation was .3, and word-to-word lateral inhibition was .21. The MROM implementation of the interactive activation model does not have lateral inhibition between letter units.

The parameter values for the deadline decision criteria are shown in Table 3. These parameter values are as close as possible to the ones reported by Grainger and Jacobs (1996), while maintaining a reasonable fit to the observed levels of accuracy. We experimented with a number of other criterion settings, but this only worsened the overall performance of the model. To accommodate the effect of instructions stressing speed instead of accuracy, the  $\Sigma$ -criterion was lowered from 1.5 to 0.7, and the T-criterion was lowered from 22 cycles to 18 cycles. In the deadline model, these criteria are the only two that are allowed to vary with speed-accuracy instructions. Note that the three decision criteria M,  $\Sigma$ , and T all vary normally around a mean value. This ensures that when the deadline model is presented with the same letter string, it does not always output the exact same response after exactly the same amount of processing (cf. Jacobs & Grainger, 1992).

Parameters  $\Delta\Sigma$  and  $\Delta T$  quantify the size of the within-trial adjustment for the  $\Sigma$ -criterion and the T-criterion, respectively. Consider the summed amount of activation across all the word units after seven cycles, and denote this quantity  $\sigma(7)$ . When  $\sigma(7)$  exceeds a

<sup>1</sup> We used the interactive activation model as programmed by Walter van Heuven. This program is publicly available at <http://www.psychology.nottingham.ac.uk/staff/vwh/jiam/>.

Table 3

Parameter values for the multiple read-out model as applied to Experiments 1 and 2

	M-criterion		$\Sigma$ -criterion				T-criterion			
	Mean	<i>SD</i>	Mean	<i>SD</i>	$\Delta\Sigma$	$L_\Sigma$	Mean	<i>SD</i>	$\Delta T$	$L_T$
Experiment 1										
Accuracy	.65	.04	<b>1.5</b>	.55	.15	.38	<b>22</b>	2	2	.22
Speed	.65	.04	<b>0.7</b>	.55	.15	.38	<b>18</b>	2	2	.22
Experiment 2										
75% Words	.65	.04	<b>0.7</b>	.55	.15	.38	<b>22</b>	2	2	.22
75% NW	.65	.04	<b>1.5</b>	.55	.15	.38	<b>18</b>	2	2	.22

*Note.* NW stands for “nonwords”, *SD* indicates the standard deviation of a normal distribution, and  $\Delta\Sigma$  and  $\Delta T$  quantify the size of the within-trial adjustment for the  $\Sigma$ -criterion and the T-criterion, respectively. The within-trial adjustments only occur when the threshold levels of global activation ( $L_\Sigma$  for the  $\Sigma$ -criterion and  $L_T$  for the T-criterion) are exceeded. The bold numbers indicate values that change as a function of the experimental conditions. See text for details.

prespecified threshold level (i.e.,  $L_\Sigma$ ), the mean value of the  $\Sigma$ -criterion is lowered by a value of  $\Delta\Sigma$ , in anticipation of the stimulus being a word. When  $\sigma(7)$  exceeds a possibly different prespecified threshold level (i.e.,  $L_T$ ), the mean value of the T-criterion is increased by a value of  $\Delta T$ , in anticipation of the stimulus not being a nonword. This means that when global activation is relatively high early in processing, the model anticipates that the stimulus will be a word and adjusts its decision criteria accordingly (cf. Grainger & Jacobs, 1996). In sum, the decision structure of the deadline model is characterized by ten parameters: the mean and standard deviation of the three criteria (i.e., M,  $\Sigma$ , and T), the within-trial adjustments  $\Delta\Sigma$  and  $\Delta T$ , and the corresponding threshold levels for the summed activation after seven cycles (i.e.,  $L_\Sigma$  and  $L_T$ ).

#### Simulation results

The simulation results for Experiment 1 are shown in Table 4 and Fig. 3. The model predicts that “word”

responses are faster than “nonword” responses, for all experimental conditions and regardless of whether the response was correct or in error. Specifically, for both high frequency and low frequency word stimuli the deadline model generates correct RTs (i.e., “word” responses) that are shorter than error RTs (i.e., “nonword” responses). This difference between correct RTs and error RTs is attenuated—but not reversed—when the emphasis is on speed rather than on accuracy. For nonword stimuli, the deadline model generates correct RTs (i.e., “nonword” responses) that are consistently longer than error RTs (i.e., “word” responses), both when instructions stress accuracy and when instructions stress speed. This reveals a deficiency of the model, as the data from Experiment 1 indicate that under instructions to respond accurately, correct “nonword” responses are in fact 63 ms faster than incorrect “word” responses. The reason that the deadline model incorrectly predicts correct “nonword” responses to be slower than incorrect “word” responses is due to the mecha-

Table 4

Percentage errors, median correct RT, and median error RT (mean RT is in brackets) for high frequency (HF) words, low frequency (LF) words, and nonwords (NW) as a function of speed instructions (Experiment 1) and list composition (Experiment 2), as obtained from the multiple read-out model

Stimulus	Experiment 1		Experiment 2	
	Accuracy	Speed	75% Words	75% Nonwords
HF				
% Errors	1.6	9.4	1.0	12.4
Correct RT	16 [16.7]	16 [15.0]	16 [15.4]	16 [16.4]
Error RT	23 [23.0]	18 [18.3]	23 [22.9]	18 [18.2]
LF				
% Errors	19.8	30.0	18.4	33.1
Correct RT	17 [17.5]	17 [16.1]	17 [16.5]	17 [17.1]
Error RT	23 [22.8]	18 [18.5]	23 [22.8]	18 [18.5]
NW				
% Errors	4.5	6.7	9.4	2.1
Correct RT	23 [23.2]	19 [19.2]	23 [23.2]	19 [19.2]
Error RT	19 [19.5]	14 [13.7]	16 [15.8]	17 [17.4]

All RTs are in cycles.

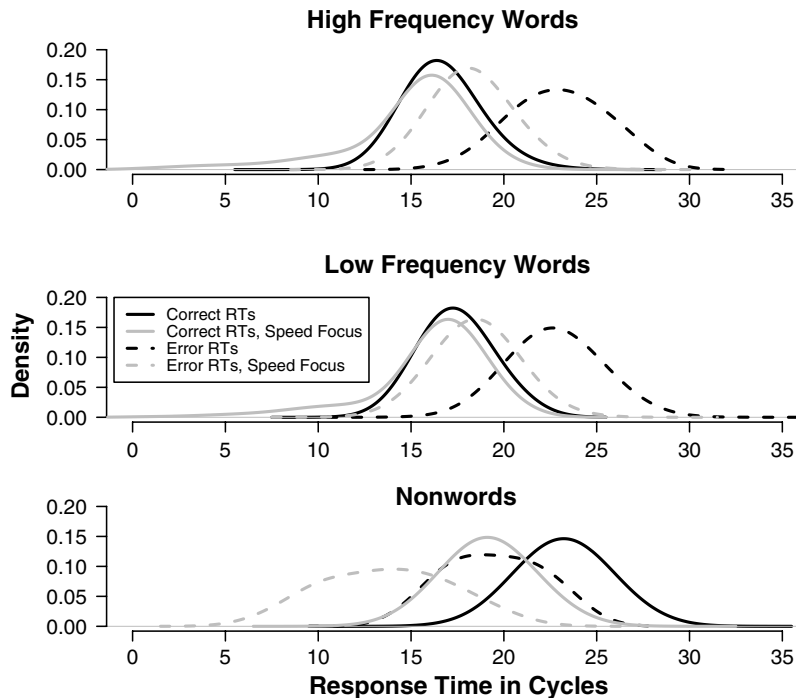


Fig. 3. MROM simulation results for Experiment 1. The distributions are density estimates based on discrete cycle times.

nism responsible for the “nonword” response. In the deadline model, the “nonword” response is a default response, that is, a response that is only made when the system has failed to detect sufficient evidence to support a “word” response. Such a temporal deadline mechanism makes it very difficult for the deadline model to generate “nonword” responses that are faster than “word” responses.

Fig. 3 also shows that the RT distributions are not skewed to the right. Most distributions are approximately normal, and some are even skewed to the left. This is in strong contradiction to the data from Experiment 1; Fig. 2 shows that all 16 RT distributions are clearly right-skewed. This can be seen most easily from Fig. 2 by comparing the distance between the lower RT quantiles to the distance between the upper quantiles.

### Discussion

As expected, speed versus accuracy instructions had a large impact on lexical decision performance. When instructions stressed response speed instead of response accuracy, error rate greatly increased, and median RT greatly decreased (cf. Table 1). Fig. 2 shows that the RT distributions also changed in shape: the characteristic right-skew is much less pronounced under instructions that stress speed than it is under instructions that stress accuracy (note the different scale on the y-axis between the top two panels and the bottom two panels).

The diffusion model accounted for all of the effects of speed versus accuracy instructions with just two parameters, boundary separation  $a$  and starting point  $z$ , free to vary between the speed condition and the accuracy condition. Of particular importance here is the fact that the decrease in boundary separation due to speed instructions reduced error RTs to a greater extent than it reduced correct RTs—a detailed and non-intuitive prediction that was borne out by the data.

The present lexical decision experiment used a within-subjects manipulation of speed versus accuracy instructions. In the diffusion model, this manipulation affects response conservativeness (i.e., boundary separation). Boundary separation indirectly affects the balance between trial-to-trial variability in starting point and trial-to-trial variability in drift rate: when boundary separation is relatively small, the impact of trial-to-trial variability in starting point is relatively strong, and this prominent role of starting point variability results in relatively fast errors. Swensson (1972) and Ratcliff and Rouder (1998, Experiment 1) also used this particular experimental design to study the RT difference between correct responses and error responses. Swensson used an intricate pay-off system to influence the tradeoff between speed and accuracy, whereas Ratcliff and Rouder used instructions that stressed either accurate or speedy performance. Both studies used a task that is different from lexical decision (i.e., judging the brightness of pixel arrays in the Ratcliff and Rouder study; judging the

slant of a rectangle in the Swensson study). The observed patterns of results with respect to the impact of instructions or pay-offs was consistent with the pattern observed in Experiment 1 here: errors were faster than correct responses in the speed conditions, and errors were slower than correct responses in the accuracy conditions.

Recently, several aging studies have also included a speed–accuracy manipulation (i.e., Ratcliff, Thapar, & McKoon, 2001; Ratcliff et al., 2004b; Thapar, Ratcliff, & McKoon, 2003). These studies used tasks in which error RTs were always longer than correct RTs; apart from the errors being slower than correct responses, the pattern of results was similar to the one reported here.

In addition to a manipulation of speed versus accuracy instructions, Experiment 1 also featured a manipulation of word frequency. The diffusion model accounted for the effects of word frequency with just one parameter, mean drift rate  $v$ , free to vary. The model predicts that the effect of speed instructions is particularly pronounced for words with low drift rates, that is, words that have a relatively low frequency of occurrence. Again, this prediction was confirmed by the data.

In contrast to the diffusion model, the deadline model could not account for the data from Experiment 1. In particular, the model consistently predicts that “word” responses should be faster than “nonword” responses. Under instructions that stress accurate responding, this prediction is correct for word stimuli, but it is incorrect for nonword stimuli. It appears to be problematic for any kind of temporal deadline model to handle situations in which responses that come about by exceeding the temporal deadline (i.e., “nonword” responses) are faster than responses that originate from processes that operate prior to the temporal deadline (i.e., “word” responses).

In addition, Fig. 3 shows that the deadline model does not produce RT distributions that are right-skewed. In contrast, the diffusion model predicts that all RT distributions, and particularly those distributions for stimuli that are difficult to classify, should be skewed to the right. The data from Experiment 1 support the diffusion model, but not the deadline model: both the distribution of correct responses to nonword stimuli and the distributions of error responses to word stimuli are markedly non-normal, having pronounced long tails.

## Experiment 2: Word/nonword proportion effects

From a diffusion model perspective, Experiment 1 influenced the distance between the starting point and the decision boundaries via speed versus accuracy instructions. In Experiment 2, this distance was influenced by manipulating the proportions of word stimuli

versus nonword stimuli from 75% words to 75% nonwords. With a high proportion of words, the starting point moves toward the word boundary and away from the nonword boundary, and vice versa with a high proportion of nonwords.<sup>2</sup> Consequently, with 75% words, correct responses to words and error responses to nonwords are speeded relative to 75% nonwords. With 75% nonwords, correct “nonword” responses and error responses for words are speeded relative to 75% words.

As mentioned earlier, the diffusion model also predicts that the RT distributions for the relatively fast decisions (i.e., correct responses to likely stimuli and incorrect responses to unlikely stimuli) will show much less spread than the RT distributions for the relatively slow decisions (i.e., correct responses to unlikely stimuli and incorrect responses to likely stimuli). In addition, when the starting point  $z$  is moved toward the boundary of the more likely response, the diffusion model predicts that the RT distributions will shift, including the leading edge of the distribution (Ratcliff & Smith, 2004, Experiment 3); this occurs because a change in starting point reduces the distance to the favored response boundary, reducing the time needed to reach that boundary for both fast and slow processes. The prediction of a shift in the leading edge is important, as the leading edge is usually relatively robust against experimental manipulations.

As in Experiment 1, the other model under consideration is the deadline model. The effect of word/nonword proportion can be accommodated in the deadline model by adjusting the  $\Sigma$ -criterion and the T-criterion. Specifically, with 75% words the model anticipates that the stimulus will be a word, and does so by lowering the  $\Sigma$ -criterion and heightening the T-criterion. Conversely, with 75% nonwords the model anticipates that the stimulus will be a nonword; this results in heightening the  $\Sigma$ -criterion and lowering the T-criterion.

Based on the simulations from Experiment 1, it was expected that the deadline model would not be able to account for the fact that “nonword” responses can be faster than “word” responses, and that the deadline model would again fail to capture the right-skew that is characteristic of RT distributions.

<sup>2</sup> Ratcliff (1985) and Ratcliff et al. (1999) found that changes in the drift criterion are sometimes capable of producing effects similar to those caused by changes in the starting point of the diffusion process. The drift criterion is the point on the drift dimension chosen so that drift rate equals zero, analogous to the zero point of strength in a signal detection analysis. Note, however, that the two parameters can be differentiated, as a change in starting point affects the leading edge more than the drift rate criterion does. For the data presented here, the fits show that the drift criterion is almost constant in fits where it is allowed to vary so we do not consider it further (for similar results, see Ratcliff & Smith, 2004, Experiment 3).

## Method

Experiment 2 was methodologically similar to Experiment 1. Blocks of trials with 75% words alternated with blocks with 75% nonwords, and each block included high, low, and very low frequency words.

## Participants

Nineteen undergraduate students of Northwestern University participated for course credit in this single-session experiment. All participants were native speakers of English.

## Stimulus materials

The same pools of words and nonwords were used as in Experiment 1.

## Design

The experiment consisted of a sequence of 21 blocks. List composition was manipulated between blocks such that blocks containing 75% words alternated with blocks containing 75% nonwords. The first block was for practice and contained 15 words and 15 nonwords. The first experimental block contained 75% words. As in Experiment 1, word frequency was manipulated within the 20 experimental blocks such that each block contained an equal proportion of high frequency, low frequency, and very low frequency items, and an equal proportion of nonword items derived from a high frequency, low frequency, or very low frequency word. The practice block contained 30 stimuli, and each of the 20 experimental blocks contained 96 stimuli.

## Procedure

The 75% word blocks were preceded by the message “Mainly Words in this list”, and the 75% nonword blocks were preceded by the message “Mainly Nonwords in this list”. The feedback message “ERROR” was presented for 800 ms after every erroneous response. Because the instructions did not stress speed, anticipatory responding was not discouraged via the feedback message “TOO FAST” as it was in Experiment 1. The experimenter supervised performance during the practice block.

## Results

The main empirical results from Experiment 2 (i.e., the effects of list composition and word frequency) are listed in Table 5. Approximately 1.5% of all trials were excluded from the analysis, either because the participant made an invalid key press, or because the response was diagnosed as an outlier (i.e., responses faster than 300 ms or slower than 2500 ms). The results for words and nonwords were analyzed separately, as were the results for correct RT, error RT, and accuracy. For

Table 5

Percentage errors, median correct RT (in ms), and median error RT for high frequency (HF) words, low frequency (LF) words, very low frequency (VLF) words, and nonwords (NW) as a function of list composition, as observed in the data from Experiment 2 and as obtained from the diffusion model

Stimulus	Observed data		Diffusion mode	
	75% W	75% NW	75% W	75% NW
<b>HF</b>				
% Errors	1.0 (.002)	5.5 (.012)	0.2	5.2
Correct RT	492 (16)	564 (11)	510	573
Error RT	554 (61)	469 (18)	623	487
<b>LF</b>				
% Errors	3.7 (.007)	19.0 (.029)	3.2	18.8
Correct RT	555 (19)	632 (14)	550	638
Error RT	743 (37)	538 (32)	739	526
<b>VLF</b>				
% Errors	10.8 (.010)	29.0 (.028)	9.2	31.4
Correct RT	595 (21)	673 (17)	574	674
Error RT	739 (29)	536 (29)	778	548
<b>NW</b>				
% Errors	13.7 (.017)	2.7 (.003)	16.0	4.0
Correct RT	669 (17)	548 (21)	671	532
Error RT	569 (34)	745 (35)	555	690

Standard errors are in parentheses.

word stimuli, the ANOVA included list composition and word frequency as within-subjects variables in a repeated measures design.

“Word” responses, whether correct or in error, were faster in the 75% word condition than in the 75% nonword condition. “Nonword” responses showed the opposite pattern: whether correct or in error, “nonword” responses were faster in the 75% nonword condition than in the 75% word condition. As in Experiment 1, both correct and error responses slowed as word frequency decreased, and accuracy decreased. The next three paragraphs provide more detailed statistical analyses.

Accuracy was higher and correct responses to word stimuli were faster in the 75% word condition as compared to the 75% nonword condition,  $F(1,18) = 80.70$ ,  $MSE = 2022$ ,  $p < .001$  and  $F(1,18) = 44.92$ ,  $MSE = 0.01$ ,  $p < .001$ , respectively. Word frequency affected both median correct RT,  $F(2,36) = 133.00$ ,  $MSE = 816$ ,  $p < .001$ , and accuracy,  $F(2,36) = 91.82$ ,  $MSE = 0.003$ ,  $p < .001$ . The linear contrasts confirmed that an increase in word frequency was accompanied by a decrease in median correct RT,  $F(1,18) = 157.00$ ,  $MSE = 1356$ ,  $p < .001$ , and an increase in accuracy,  $F(1,18) = 132.61$ ,  $MSE = 0.004$ ,  $p < .001$ .

Errors for word stimuli were faster in the 75% nonword condition than in the 75% word condition,  $F(1,11) = 19.95$ ,  $MSE = 25,630$ ,  $p < .01$ . Word frequency affected median error RT,  $F(2,22) = 3.16$ ,  $MSE = 1711$ ,  $p = .031$ , one-tailed. The linear contrast

showed that median error RT increased as word frequency decreased,  $F(1, 11) = 4.61$ ,  $MSE = 15,832$ ,  $p = .028$ , one-tailed. For words, list composition affected the relative speed of errors,  $F(1, 11) = 32.40$ ,  $MSE = 36,287$ ,  $p < .001$ , as median error RT was relatively slow in the 75% word condition, and relatively fast in the 75% nonword condition.

For nonword stimuli, the ANOVA included list composition as a within-subjects variable in a repeated measures design. Compared to the 75% nonword condition, nonword stimuli presented in the 75% word condition were responded to less accurately,  $F(1, 18) = 45.92$ ,  $MSE = 0.002$ ,  $p < .001$ , more slowly when correct,  $F(1, 18) = 141.13$ ,  $MSE = 987$ ,  $p < .001$ , but faster when in error,  $F(1, 18) = 28.25$ ,  $MSE = 10,340$ ,  $p < .001$ . List composition affected the difference between correct RT and error RT,  $F(1, 18) = 61.93$ ,  $MSE = 13,478$ ,  $p < .001$ , such that errors were faster than correct responses in the 75% word condition, but slower than correct responses in the 75% nonword condition.

In sum, the results of Experiment 2 are consistent with the qualitative predictions from the diffusion model. “Word” responses, whether correct or in error, were faster in the 75% word condition than in the 75% nonword condition. Similarly, “nonword” responses, whether correct or in error, were faster in the 75% nonword condition than in the 75% word condition.

#### *Diffusion model analysis*

The previous analyses demonstrated that results are consistent with the diffusion model on a qualitative level. To show that the diffusion model can also quantitatively account for the observed results, the model was fit to the data from Experiment 2 in the same fashion as for Experiment 1 (i.e., group RT distributions were constructed separately for correct responses and errors using quantile averaging, and the  $\chi^2$  method was used to minimize the discrepancy between the model and the data). Table 5 allows a quick comparison between the most important aspects of the data and the fits of the model, whereas Fig. 4 gives a complete summary of the fits of the diffusion model to the data over all five RT quantiles.

Table 5 shows a comparison between the data in the 16 conditions and the results from the model for error rates and the .5 quantiles (i.e., the medians), and this confirms that the model can account for the observed results. As in Experiment 1, the diffusion model appears to be in closer correspondence to the data for correct responses than for error responses, due to the fact that the error latencies are much more variable than the latencies for correct responses.

Fig. 4, top-left panel, plots the RT distributions for correct responses in the 75% word condition. As in Fig. 2, for each stimulus type (i.e., high frequency words, low frequency words, very low frequency words, and

nonwords), the x-axis shows the probability of a correct response. The five vertical Xs correspond to the observed RT quantiles. The grey dots scattered around the Xs indicate variability in the data. The five vertical +s correspond to the predicted RT quantiles. Fig. 4, top-right panel, plots the RT distributions for incorrect responses in the 75% word condition. For the 75% nonword condition, Fig. 4, bottom-left panel, plots the RT distributions for correct responses, and Fig. 4, bottom-right panel, shows the RT distributions for incorrect responses.

Fig. 4 demonstrates that the diffusion model is able to account for response probability and for the shape of RT distributions for correct and error responses. The diffusion model's predictions are very close to the observed data, and almost never fall outside of the variability in the data that is indicated by the gray dots. By comparing the left panels to the right panels, it is evident that the RT distribution of correct responses for a particular type of stimulus differs from its error RT counterpart. This difference is particularly pronounced in the leading edge of the distribution, that is, in the .1 quantile.

Also note that within each of the four panels, the leading edge of the RT distribution associated with nonword stimuli differs from that associated with word stimuli. That is, the leading edge of the RT distribution for nonword stimuli is either shorter than those for word stimuli (i.e., top-right and bottom-left panels), or it is longer (i.e., top-left and bottom-right panels). According to the diffusion model, in the 75% word condition the starting point is close to the word boundary. Consequently, the earliest “word” responses will be relatively fast, and the earliest “nonword” responses will be relatively slow. Fig. 4 shows that this pattern holds irrespective of whether the response was correct or in error. In the 75% nonword condition, the starting point is close to the nonword boundary, and the opposite pattern of results is obtained. That is, the earliest “word” responses will be relatively slow, and the earliest “nonword” responses will be relatively fast, regardless of whether the response was correct or incorrect (cf. Fig. 4, bottom two panels).

Table 6 provides the values for the parameter estimates. The diffusion model captures the large performance differences, 15% in response accuracy and 100–200 ms in the median RTs, with only two parameters free to vary. First, the starting point  $z$  was closer to the boundary for the more frequently presented stimulus type. Second, the boundary separation  $a$  increased by 10% when the stimuli consisted of 75% words. Also, as in Experiment 1, varying only the mean drift rate parameter,  $v$ , accounted for all the effects of word frequency.

#### *Deadline model analysis*

The simulations of the deadline model parallel those for Experiment 1. Word and nonword stimuli used in the experiment were repeatedly presented to the deadline

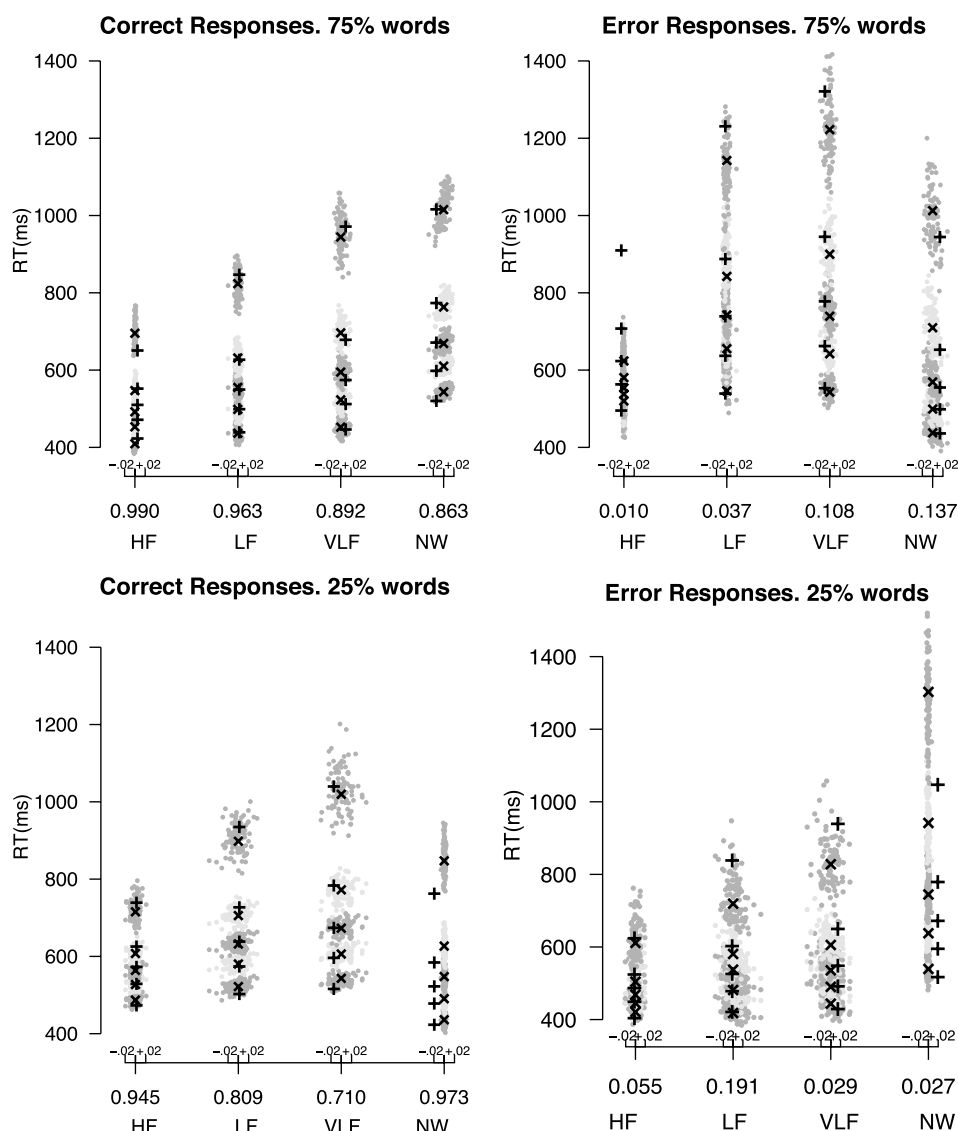


Fig. 4. Empirical (*Xs*) and predicted (*+s*) .1, .3, .5, .7, and .9 quantiles for RT distributions in Experiment 2. The grey dots show variability from bootstrap simulations from the data. HF, high frequency word; LF, low frequency word; VLF, very low frequency word; and NW, nonword. The *x*-axis shows response accuracy. Top-left panel: correct responses in the 75% word condition; top-right panel: incorrect responses in the 75% word condition; bottom-left panel: correct responses in the 75% nonword condition; bottom-right panel: incorrect responses in the 75% nonword condition.

model, and its performance was monitored in terms of RT (i.e., number of cycles until a decision threshold is reached) and response choice (i.e., “word” or “non-word”, depending on which decision threshold was reached first).

#### Materials and procedure

The materials and procedure were identical to ones used in the deadline model simulations that followed Experiment 1.

#### Parameters

The parameter values for the interactive activation model were left unchanged at their default values. The parameter values for the decision criteria are shown in Table 3. In Experiment 1, speed versus accuracy instructions affected the  $\Sigma$ -criterion and the T-criterion in the same fashion, that is, both criteria were lowered under instructions to respond quickly. In the simulation of Experiment 2, the experimental manipulation of response criteria affects the  $\Sigma$ -criterion and the T-criterion in

Table 6  
Best-fitting parameter values for the diffusion model for the data from Experiment 2

Parameters	75% Words	75% Nonwords
$a$	0.130	0.118
$z$	0.085	0.039
$T_{er}$	0.422	-
$s_z$	0.041	-
$s_t$	0.151	-
$v(\text{HF})$	0.476	-
$v(\text{LF})$	0.260	-
$v(\text{VLF})$	0.169	-
$v(\text{NW})$	-0.252	-
$\eta$	0.101	-

*Note.* HF, high frequency word; LF, low frequency word; VLF, very low frequency word; and NW, nonword. A hyphen indicates that the parameter value was constrained to be identical to the one in the adjacent column. See text for details.

opposite directions. That is, the  $\Sigma$ -criterion is lower in the 75% word condition than in the 75% nonword condition, whereas the T-criterion is higher in the 75% word condition than it is in the 75% nonword condition. Again, the parameter values were as close as possible to the ones reported by Grainger and Jacobs (1996), while providing a reasonable fit to the observed levels of accuracy.

#### Simulation results

The simulation results for Experiment 2 are shown in Table 4 and Fig. 5. For both high frequency and low fre-

quency words, the deadline model generates incorrect “nonword” responses that are slower than correct “word” responses. This is inconsistent with the data: in the 75% nonword condition, incorrect “nonword” responses are about 100 ms *faster* than correct “word” responses. For nonword stimuli, the deadline model predicts that incorrect “word” responses should be faster than correct “nonword” responses. This prediction holds true for the 75% word condition, but it fails for the 75% nonword condition, in which errors are about 200 ms *slower* than correct responses. Thus, as in the previous simulation, the deadline model has a major problem accounting for “nonword” responses that are faster than “word” responses.

In addition, the deadline model does not generate RT distributions that have the correct shape. An examination of Fig. 4 shows that the data from all 16 conditions are clearly skewed to the right, having a pronounced right tail (i.e., in Fig. 4, the difference between the lower RT quantiles is much smaller than the difference between the upper RT quantiles). In contrast, Fig. 5 shows that the model predicts RT distributions to be normal, or even skewed to the left.

#### Discussion

The present results show that participants adjust their criterion settings to reflect the statistical regularities in the lists of stimuli: responses to the less probable stim-

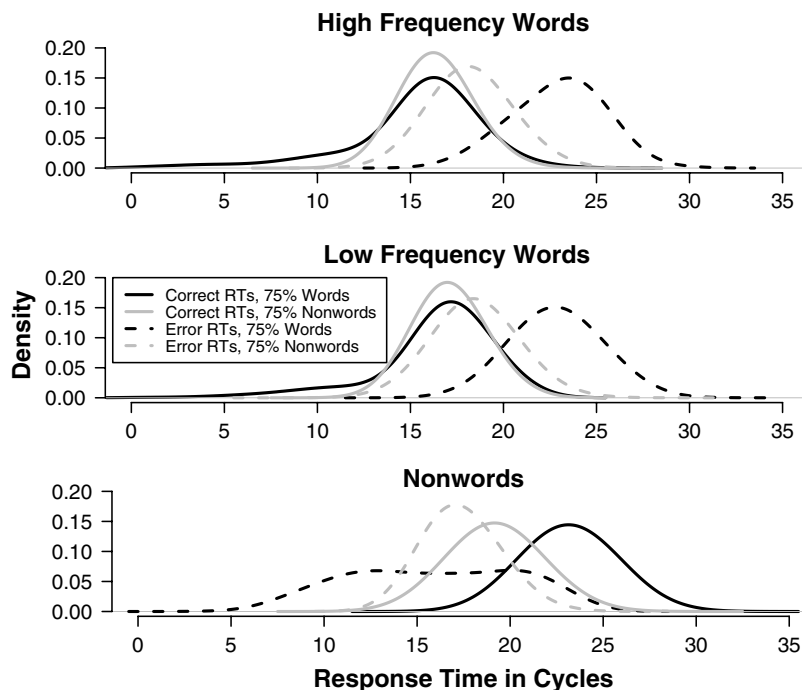


Fig. 5. MROM simulation results for Experiment 2. The distributions are density estimates based on discrete cycle times.

uli are slower and less accurate than responses to the more probable stimuli. That is, with 75% words in a list, median correct RT for word stimuli is shorter than median error RT for word stimuli. With 75% nonwords in a list, median correct RT for word stimuli is longer than median error RT for word stimuli. This pattern of results reverses for nonword stimuli. The above effects on RT are evident for the entire RT distribution, including the leading edge. Response accuracy is also affected by list composition: for word stimuli, error rate is lowest for lists with 75% words, whereas for nonword stimuli, error rate is lowest for lists with 75% nonwords.

Quantitative fits demonstrate that the diffusion model can parsimoniously account for the effects of list composition in lexical decision. By altering only those parameters that are associated with criterion settings, the diffusion model captures the effects of list composition on error rate and on the shape of the RT distributions for both correct and error responses. In particular, the diffusion model correctly predicts a shift in the leading edge of the RT distributions as a function of the experimental manipulation. By varying only the drift rate parameter, the diffusion model also captures the effects of word frequency: low frequency words are responded to slower than high frequency words, and this effect is particularly pronounced in the tails of the RT distributions (i.e., the .9 quantile).

As in Experiment 1, the deadline model failed to account for the data, even qualitatively. For levels of response accuracy in the range of the empirical data, the model does not produce “nonword” responses that are faster than “word” responses. In the 75% nonword condition, “nonword” responses are generally much faster than “word” responses, regardless of whether the response is correct or in error. Also, as in Experiment 1, the observed RT distributions are markedly non-normal. Fig. 4 demonstrates that all RT distributions are skewed to the right, having long tails. The deadline model produces normal or left-skewed distributions.

## General discussion

In Experiments 1 and 2, we manipulated response criteria settings in lexical decision through speed–accuracy instructions and through the proportions of words versus nonwords in the test lists. The question was whether the diffusion and deadline models could account for the experimental results.

The diffusion model makes several strong predictions about the effects of manipulations that lead to changes in criterion settings. Moving from accuracy to speed instructions, error RTs should decrease relative to correct RTs. When the proportions of words versus nonwords are varied, responses to the less probable stimuli should be slower and less accurate than responses to the more

probable stimuli. The RT distributions for the less probable stimuli are predicted to both shift and spread relative to the RT distributions for the more probable stimuli.

Experiment 1 confirmed these predictions for the effects of speed versus accuracy instructions and Experiment 2 confirmed them for the effects of proportion of words versus nonwords. In both cases, the data were modeled with only the starting point of the diffusion process and the separation between the decision criteria varying across conditions. These two parameters accounted for changes in accuracy, the relative speeds of correct and error responses, and the shapes of the RT distributions, including the relative positions of the leading edges of the RT distributions.

The deadline model also has two parameters to account for the effects of the speed–accuracy and proportion manipulations, namely the  $\Sigma$ -criterion and the T-criterion. However, simulations revealed that when the deadline model was constrained to maintain a reasonable level of classification accuracy, the model was unable to generate fast “nonword” responses. We believe that the deadline model’s inability to generate fast “nonword” responses is a result of a fundamental feature of the model, namely that a “nonword” response is engaged only after the system fails to find sufficient evidence for a “word” response. This assertion implies that minor modifications of the standard deadline model will not allow it to generate fast “nonword” responses.

### *Additional explorations of the deadline model*

In order to explore the generality of our claim that the deadline model cannot generate fast “nonword” responses while maintaining a reasonable level of response accuracy, we conducted several additional simulation of the MROM. In most of our simulations, we used the MROM parameters for Experiment 2 as the point of departure (i.e., the bottom two rows of Table 3)—recall that in the 75% nonword condition of this experiment, fast “nonword” responses were reliably observed for all stimulus categories.

In one set of simulations, we allowed the M-criterion to vary from one condition to the next. Recall that the M-criterion represents the activation threshold for individual word representations (cf. Morton, 1969); The standard MROM model assumes that the mean of the M-criterion is fixed and outside of the system’s control. Recently, however, Perea, Carreiras, and Grainger (2004, p. 1096) suggested that the M-criterion may be “strategically variable”. It could be that the addition of a strategically variable M-criterion allows the MROM model to generate fast “nonword” responses. In the 75% nonword condition of Experiment 2, the M-criterion was heightened to facilitate “nonword” responses.

In another set of simulations, we let the system monitor the global level of lexical activation at different times

after stimulus onset. Recall that in the standard version of the MROM model, the system computes the summed amount of activation across all the word units after seven cycles (i.e.,  $\sigma(7)$ ); this summed amount of activation may lead to within-trial changes in the  $\Sigma$ -criterion and in the T-criterion. However, the number of cycles after which the system computes the global amount of lexical activation (i.e., the cycle check number C) is arbitrary—there is no *a priori* reason why the system should not monitor global lexical activation earlier or later than after seven cycles. Therefore, it is important to confirm that MROM cannot generate fast “nonword” responses when the level of global activation is assessed after different numbers of cycles.

Our simulations showed that neither of the above MROM adjustments let the model produce fast “nonword” responses. Specifically, in the 75% nonword condition of Experiment 2, the most important effect of an increase in the M-criterion was to bias the system to respond “nonword”, and this resulted primarily in more errors to word stimuli. Nevertheless, the MROM continued to predict that incorrect “nonword” responses to word stimuli were slower than correct “word” responses, contrary to what the data show. Changes in cycle check time did not remedy this situation. These simulations provide further support for our general claim that the deadline model cannot generate fast “nonword” responses while keeping response accuracy at an acceptable level.

The deficiencies of the deadline model are striking, but they do not imply that the deadline model’s underlying representational assumptions are incorrect. The model might well provide a better account of the data if the temporal deadline mechanism was replaced by a decision mechanism similar to that of the diffusion model. This is an important point, as one of the greatest attractions of the MROM deadline model is that it allows the researcher to make predictions about the relative speeds of processing for specific words. The diffusion model, in contrast, is not a model of lexical representation. In the diffusion model, no mention is made of how words are represented or organized in memory (cf. De Moor, Verguts, & Brysbaert, 2005; Joordens, Piercey, & Azarbehi, 2003; Ratcliff et al., 2004a). Rather, the diffusion model describes the decision components of the lexical decision task. The possibility of combining models of lexical representation with the decisional machinery of the diffusion model was briefly discussed in Ratcliff et al. (2004a), and here we discuss this issue in more detail.

#### *Neural networks and the diffusion model*

Neural networks are among the most popular methods to represent the mental lexicon. A neural network can represent words by single units (i.e., a local connec-

tionist model) or by patterns of activation across the network units (i.e., a distributed connectionist model). Regardless of the specific type of neural network used, the single most problematical issue that these networks face in the modeling of lexical decision is that nonwords, by definition, have no lexical representation. This raises the question as to how exactly evidence accumulates to support a “nonword” response. In addition, performance for word stimuli depends greatly on the extent to which the nonwords are similar to words (e.g., Wagenmakers et al., 2004). That is, a word such as TANGO is responded to faster and more accurately when it is presented in a list with easy nonwords such as MRLOP than when it is presented in a list with difficult nonwords such as DRAPA.

One method to address this phenomenon is to assume a temporal deadline mechanism for the “nonword” response; however, the present experiments and modeling cast doubt on the validity of such a deadline mechanism. A second method is to assume that the system uses an estimate of the amount of lexical activation that can be expected in case the stimulus is a nonword (e.g., Joordens et al., 2003; Plaut, 1997). For instance, if the system is forced to respond “word” or “nonword” after some fixed time  $t$  following stimulus onset, then the lexical activation at time  $t$  could be compared to what is expected in case the stimulus is a word and to what is expected in case the stimulus is a nonword. The complication is that in the standard lexical decision task, the participant accumulates information until he or she feels confident enough to respond—how can a neural network model simulate this standard lexical decision paradigm?

One answer is to assume that the system has continuous access to estimates for the amount of lexical activation in case the stimulus is a word or a nonword (e.g., Joordens et al., 2003). Consider a stimulus that generates a lexical activation value of 0.5 after time  $t$ . The evidential impact of this value critically depends on  $t$ , that is, on the amount of processing that the stimulus has already undergone. If the value of 0.5 is reached almost immediately after stimulus onset, this may provide substantial evidence for the hypothesis that the stimulus is a word. If the value of 0.5 is reached only after the stimulus has been thoroughly analyzed, this may actually provide strong evidence *against* the hypothesis that the stimulus is a word. Thus, it is not enough that the system knows two distributions that reflects the system’s expectation about the amount of lexical activation for words and nonwords. Because the total amount of lexical activation is time-dependent, the system needs to know about how these two distributions increase and diverge over time and use this information to continuously adjust its computations. This places a heavy burden on the system’s computational resources.

### Bayesian approaches and the diffusion model

Bayesian approaches for modeling lexical decision (Norris, 2006; Wagenmakers et al., 2004) are similar to the diffusion model in the sense that both approaches depend on the sequential accumulation of noisy information that acts as a measure of relative evidence. That is, information that increases the likelihood that the stimulus is a word will simultaneously decrease the likelihood that the stimulus is a nonword. Moreover, Bayesian approaches are generally optimal (e.g., Geisler, 2003), and so is the diffusion model (i.e., the diffusion model minimizes RT for a given level of accuracy; Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006; Wald & Wolfowitz, 1948).

In the REM-LD model (Wagenmakers et al., 2004), the stimulus is represented as a vector of features. These stimulus features are matched to the features from word representations in lexical memory. As processing time increases, more and more stimulus features become available to the comparison process. The outcome of the noisy comparison process is a number of matches and mismatches. Based on the probability of a feature match given that the stimulus is a word, and the probability of a feature match given that the stimulus is a nonword, the system is able to calculate the overall odds that the stimulus is a word. Note that nonwords are not represented in memory, and processing time does not need to be taken into account explicitly. The model was applied to data from a signal-to-respond paradigm, but the extension of the REM-LD model to a free-response paradigm is relatively straightforward. The resulting model would capture many of the same qualitative trends as captured by the diffusion model.

In the Bayesian Reader (Norris, 2006), the stimulus is represented in a multi-dimensional perceptual space. As in REM-LD, the Bayesian Reader calculates the odds that a stimulus is a word versus a nonword. In order to account for data from the lexical decision task, the Bayesian Reader needs to have some knowledge about nonwords. It is assumed that the system knows that nonwords are relatively similar to existing words (i.e., nonwords generally differ from words in one letter only), and so each nonword is located relatively close to at least one word in multi-dimensional perceptual space. These nonword representations are “virtual”, in that they are merely postulated to quantify the system’s expectancy for word-to-nonword similarity.

The Bayesian Reader was developed to demonstrate how a rational analysis of task performance can parsimoniously account for a set of robust phenomena in visual word recognition. The model has not yet been fit to data with a high degree of accuracy. Given its conceptual similarity to the diffusion model (i.e., optimal decision making based on sequential sampling of noisy information), we expect that the Bayesian Reader would

be able to capture many of the qualitative patterns of results obtained in the current study.

To conclude, the ability of the diffusion model to explain how and why behavior changes with manipulations of speed–accuracy instructions and stimulus proportions adds to the growing body of support for the model, further attesting to its descriptive and explanatory power. The model is severely constrained: few parameters are free to vary in fits of the model to data, but the model still explains large differences in performance across experimental conditions. Other models for lexical decision have not been developed to a level of detail that would allow quantitative comparisons to the diffusion model. The model’s good performance with the data presented in this article demonstrates that the model can help to disentangle the effects of lexical processing from other effects such as those of decision thresholds.

### Acknowledgments

Preparation of this article was supported by NIMH Grant R37-MH44640, NIA Grant R01-AG17083, NIMH Grant K05-MH01891, and a Veni grant from the Netherlands Organisation for Scientific Research (NWO). We thank Steve Lupker and Marius Usher for comments on an earlier draft of this paper.

### References

- Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 340–357.
- Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced choice tasks. *Psychological Review*, 113, 700–765.
- Brown, S. D., & Steyvers, M. (2005). The dynamics of experimentally induced criterion shifts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 587–599.
- Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance: Vol. VI* (pp. 535–555). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, 108, 204–256.
- De Moor, W., Verguts, T., & Brysbaert, M. (2005). Testing the multiple in the multiple read-out model of visual word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 1502–1508.

- Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.
- Geisler, W. S. (2003). Ideal observer analysis. In L. Chalupa & J. Werner (Eds.), *The visual neurosciences* (pp. 825–837). Boston: MIT press.
- Glanzer, M., & Ehrenreich, S. L. (1979). Structure and search of the internal lexicon. *Journal of Verbal Learning and Verbal Behavior*, 18, 381–398.
- Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review*, 103, 518–565.
- Jacobs, A. M., & Grainger, J. (1992). Testing a semistochastic variant of the interactive activation model in different word recognition experiments. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 1174–1188.
- Joordens, S., Piercey, C. D., & Azarbeh, R. (2003). From word recognition to lexical decision: A random walk along the road of harmony. In F. Detje, D. Dörner, & H. Schaub (Eds.), *Proceedings of the fifth international conference on cognitive modeling* (pp. 141–146).
- Kucera, H., & Francis, W. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- Laming, D. R. J. (1968). *Information theory of choice-reaction times*. London: Academic Press.
- Laming, D. R. J. (1973). *Mathematical psychology*. New York: Academic Press.
- McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part I. An account of basic findings. *Psychological Review*, 88, 375–407.
- Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, 76, 165–178.
- Norris, D. (2006). The Bayesian reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Review*, 113, 327–357.
- 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.
- Plaut, D. C. (1997). Structure and function in the lexical system: Insights from distributed models of word reading and lexical decision. *Language and Cognitive Processes*, 12, 765–805.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85, 59–108.
- Ratcliff, R. (1981). A theory of order relations in perceptual matching. *Psychological Review*, 88, 552–572.
- Ratcliff, R. (1985). Theoretical interpretations of the speed and accuracy of positive and negative responses. *Psychological Review*, 92, 212–225.
- Ratcliff, R. (2002). A diffusion model account of response time and accuracy in a brightness discrimination task: Fitting real data and failing to fit fake but plausible data. *Psychonomic Bulletin & Review*, 9, 278–291.
- Ratcliff, R., Gomez, P., & McKoon, G. (2004a). A diffusion model account of the lexical decision task. *Psychological Review*, 111, 159–182.
- Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, 9, 347–356.
- Ratcliff, R., & Rouder, J. N. (2000). A diffusion model account of masking in two-choice letter identification. *Journal of Experimental Psychology: Human Perception and Performance*, 26, 127–140.
- Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, 111, 333–367.
- Ratcliff, R., Thapar, A., Gomez, P., & McKoon, G. (2004b). A diffusion model analysis of the effects of aging on recognition memory. *Journal of Memory and Language*, 50, 408–424.
- Ratcliff, R., Thapar, A., & McKoon, G. (2001). The effects of aging on reaction time in a signal detection task. *Psychology and Aging*, 16, 323–341.
- Ratcliff, R., Thapar, A., & McKoon, G. (2006). Aging, practice, and perceptual tasks: A diffusion model analysis. *Psychology and Aging*, 21, 353–371.
- Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review*, 9, 438–481.
- Ratcliff, R., Van Zandt, T., & McKoon, R. (1999). Connectionist and diffusion models of reaction time. *Psychological Review*, 102, 261–300.
- Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: Part II. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review*, 89, 60–94.
- Swenson, R. G. (1972). The elusive tradeoff: Speed versus accuracy in visual discrimination tasks. *Perception & Psychophysics*, 12, 16–32.
- Thapar, A., Ratcliff, R., & McKoon, G. (2003). A diffusion model analysis of the effects of aging on letter discrimination. *Psychology & Aging*, 18, 415–429.
- Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory & Cognition*, 32, 1206–1220.
- Wagenmakers, E.-J., Steyvers, M., Raaijmakers, J. G. W., Shiffrin, R. M., van Rijn, H., & Zeelenberg, R. (2004). A Bayesian model for lexical decision. *Cognitive Psychology*, 48, 332–367.
- Wald, A., & Wolfowitz, J. (1948). Optimum character of the sequential probability ratio test. *Annals of Mathematical Statistics*, 19, 326–339.
- Webster, M. (1990). *Merriam-Webster's ninth collegiate dictionary*. NY: Merriam-Webster.
- Yellott, J. I. Jr., (1971). Correction for guessing and the speed–accuracy tradeoff in choice reaction time. *Journal of Mathematical Psychology*, 8, 159–199.