OBSERVATION

How to Say “No” to a Nonword: A Leaky Competing Accumulator Model of Lexical Decision

Stéphane Dufau, Jonathan Grainger, and Johannes C. Ziegler
Centre National de la Recherche Scientifique, Marseille, France, and Aix-Marseille University

We describe a leaky competing accumulator (LCA) model of the lexical decision task that can be used as a response/decision module for any computational model of word recognition. The LCA model uses evidence for a word, operationalized as some measure of lexical activity, as input to the YES decision node. Input to the NO decision node is simply a constant value minus evidence for a word. In this way, evidence for a nonword is a function of time from stimulus onset (as in standard deadline models) modulated by lexical activity via the competitive dynamics of the LCA. We propose a simple mechanism for determining the value of this constant online during the first trials of a lexical decision experiment, such that the model can rapidly optimize speed and accuracy in discriminating words from nonwords. Further optimization is achieved via trial-by-trial adjustments in response criteria as a function of task demands and list context. We show that the LCA model can simulate mean response times and response distributions for correct and incorrect YES and NO decisions for a number of benchmark experiments that have been shown to be fatal for deadline models of lexical decision. Finally, using lexical activity calculated by a computational model of word recognition as input to the LCA decision module, we provide the first item-level simulation of both word and nonword responses in a large-scale database.

Keywords: lexical decision, leaky competing accumulator, response criteria, item-level simulations, nonword responses

The lexical decision task, first used by Rubenstein, Garfield, and Millikan (1970), remains one of the most widely used behavioral measures of silent word reading. A recent smartphone application even enables lexical decision data to be collected from thousands of participants all over the world (Dufau et al., 2011). Participants in this task are requested to classify as rapidly and accurately as possible a given string of letters as being a word or not. Variation in response times and in accuracy to different word stimuli is taken to reflect variation in the ease with which these words are read. Performance in the lexical decision task has therefore played a primary role in developing and testing theories of visual word recognition. However, as with that for any other measure of word recognition, lexical decision performance is prone to measurement biases and response strategies, which complicate the interpretation of data obtained with this task (see Balota & Chumbley, 1984). One solution to this general problem, proposed by Grainger and Jacobs (1996), is to develop computational models that incorporate a key distinction between task-specific and task-independent core processes. In doing so, Grainger and Jacobs applied the notion of functional overlap to illustrate how different tasks used to investigate a given phenomenon can each contribute to uncovering the basic mechanisms underlying the phenomenon of interest, as long as the contribution of task-specific mechanisms is well understood.

Continuing this tradition, we argue that lexical decision involves some very important differences with respect to normal, everyday, word recognition and that understanding these differences is absolutely critical for making progress in this field. However, we also believe that when performing the lexical decision task, participants apply very general principles of human information processing expressed in terms of optimization of performance (Anderson, 1990; Oaksford & Chater, 1998). In our opinion, these general principles do not dispense with a more fine-grained analysis of the specific mechanisms involved in performing a particular task, and it is such fine-grained analyses that advance understanding, once the more general principles have been accepted (e.g., Perry, Ziegler, & Zorzi, 2010).

Lexical Decision as Evidence Accumulation

The present work follows the tradition of recent accounts of performance in the lexical decision task (Norris, 2006; Ratcliff, Gomez, & McKoon, 2004; Wagenmakers, Ratcliff, Gomez, & Milikan (1970), remains one of the most widely used behavioral measures of silent word reading. A recent smartphone application even enables lexical decision data to be collected from thousands of participants all over the world (Dufau et al., 2011). Participants in this task are requested to classify as rapidly and accurately as possible a given string of letters as being a word or not. Variation in response times and in accuracy to different word stimuli is taken to reflect variation in the ease with which these words are read. Performance in the lexical decision task has therefore played a primary role in developing and testing theories of visual word recognition. However, as with that for any other measure of word recognition, lexical decision performance is prone to measurement biases and response strategies, which complicate the interpretation of data obtained with this task (see Balota & Chumbley, 1984). One solution to this general problem, proposed by Grainger and Jacobs (1996), is to develop computational models that incorporate a key distinction between task-specific and task-independent core processes. In doing so, Grainger and Jacobs applied the notion of functional overlap to illustrate how different tasks used to investigate a given phenomenon can each contribute to uncovering the basic mechanisms underlying the phenomenon of interest, as long as the contribution of task-specific mechanisms is well understood.

Continuing this tradition, we argue that lexical decision involves some very important differences with respect to normal, everyday, word recognition and that understanding these differences is absolutely critical for making progress in this field. However, we also believe that when performing the lexical decision task, participants apply very general principles of human information processing expressed in terms of optimization of performance (Anderson, 1990; Oaksford & Chater, 1998). In our opinion, these general principles do not dispense with a more fine-grained analysis of the specific mechanisms involved in performing a particular task, and it is such fine-grained analyses that advance understanding, once the more general principles have been accepted (e.g., Perry, Ziegler, & Zorzi, 2010).

Lexical Decision as Evidence Accumulation

The present work follows the tradition of recent accounts of performance in the lexical decision task (Norris, 2006; Ratcliff, Gomez, & McKoon, 2004; Wagenmakers, Ratcliff, Gomez, &
McKoon, 2008) in applying the basic notion of noisy accumulation of evidence over time. Input to the decision process is therefore a measure of evidence in favor of a word response and evidence in favor of a nonword response. However, although it is fairly straightforward to specify what might constitute evidence for a word, defining what might be evidence for a nonword is not so straightforward. This is one specificity of the lexical decision task, as opposed to other binary decision tasks, such as deciding whether a stimulus is moving to the left or to the right, is a square or a circle, or is red or blue. In a sense, the problem here is that evidence for a nonword is just less of the same measure that is used to provide evidence for a word. This is unlike binary decision tasks, where qualitatively different kinds of evidence are associated with the two responses, and it is this particular characteristic of the lexical decision task that was the major motivation for deadline accounts of nonword decision making (Grainger & Jacobs, 1996).

Ratcliff et al. (2004) described a diffusion model account of the lexical decision task. They assigned a positive value to word stimuli, whose strength depended on word frequency, and a negative value to nonword stimuli, whose strength was determined by how much the nonword resembled real words. Of course, this approach begs the question of how a lexical processor could actually transform the evidence extracted from a string of letters into positive and negative values. One possibility would be that the system learns, via exposure to words and nonwords during practice, the amounts of activation generated by word and nonword stimuli and, as a function of that experience, sets a zero point on the activation (wordness) dimension separating positive values from negative values. This corresponds to the drift rate criterion in the diffusion model.

Norris’ (2006) model, the Bayesian Reader, performs lexical decision by calculating the posterior probability that the stimulus is a word and comparing this with the probability that it is a nonword. The posterior probability of a word response for a particular stimulus is calculated using the probabilities of each individual word, given the stimulus (something akin to a global lexical activation value). The posterior probability of a nonword response is calculated by assuming the existence of a single virtual nonword stimulus, whose distance from the input is greater than the distance of the nearest word by a value that varies as a function of the similarity of words and nonwords in an experiment. This implies that participants in a lexical decision experiment must evaluate the average frequency of the nonword stimuli in the experiment. They could do so by estimating the amount of lexical activation that nonwords compared with the word stimuli generate, in a way similar to the setting of the drift rate criterion in the diffusion model.

The Present Approach

In the present work we apply an alternative framework for understanding speeded binary decision making, the leaky competing accumulator (LCA) model of Usher and McClelland (2001). The specific architecture of the LCA model provides an ideal framework for conceptualizing the basic mechanisms involved in making a lexical decision. In particular, the LCA framework offers a straightforward implementation of what could be conceptualized as a dynamic deadline account of lexical decision and, therefore, an alternative to standard deadline models, such as the multiple read-out model (MROM; Grainger & Jacobs, 1996). One key aspect of the MROM is that the temporal deadline is adjusted as a function of lexical activity generated by the stimulus, such that the deadline is adjusted upward with greater values of lexical activity and downward with lower values of lexical activity (see also Coltheart, Davelaar, Jonasson, & Besner, 1977). In this way, the temporal deadline in the MROM is essentially a mechanism for combining evidence for a word with elapsed time (time from stimulus onset) in order to generate a NO response. A NO response is generated if not enough evidence for a YES response has accumulated before the deadline has been reached. However, Ratcliff et al. (2004) and, more recently, Wagenmakers et al. (2008) have outlined a number of deficiencies with this particular mechanism. The deficiencies that were pinpointed in these studies are summarized in Table 1. All concern either correct response times (RTs) to nonword stimuli or error RTs to word stimuli. Thus, these deficiencies are all related to performance driven by the temporal deadline mechanism. Here we describe a solution for generating NO responses in the LCA model that is conceptually similar to the temporal deadline mechanism. We examine whether this dynamic deadline model overcomes the deficiencies listed by Ratcliff et al. (2004) and Wagenmakers et al. (2008) with respect to standard deadline models such as the MROM (Grainger & Jacobs, 1996).

A primary goal in the present work is to understand the mechanism used by participants in a lexical decision task to discriminate words from nonwords. After all, that is the definition of the lexical decision task. As described above, deadline models provide a straightforward solution here. For models like the diffusion model and the Bayesian Reader, the solution is not so straightforward (e.g., the virtual nonword mechanism of the Bayesian Reader). Here we offer an easily implementable solution within the framework of LCA models.

Another important goal in the present study is to investigate the role of adjustments in response criteria within the LCA framework as a means of capturing effects of list context in the lexical decision task. This second goal is a straightforward extension of the criterion adjustment approach developed by Grainger and Jacobs (1996) in their account of list-context effects on lexical decision (see also Perea, Carreiras, & Grainger, 2004). The approach adopted here is not to optimize data fitting via parameter tuning routines but to understand changes in model performance (mimicking changes in human performance) by controlled adjustments of a handful of parameters. Put differently, the philosophy of our approach is not to seek which parameters need to be adjusted in order to accurately fit data but rather to test whether theoretically motivated adjustments of parameters allow the model to account for the data.2

Finally, it is important to note that our aim is not to compare an LCA model of lexical decision with other eminent approaches.

1 The lexical decision task is not the only task in this category of binary decisions, one other prominent example being the standard old–new decision task used to study recognition memory.

2 We are not claiming that our approach is superior in any way, just that the focus is different.
such as the diffusion model (Ratcliff et al., 2004). This particular comparison has already been made in the work of Ratcliff and Smith (2004). We also acknowledge the obvious similarities across these different approaches and the fact that they can be shown to be mathematically equivalent under certain ranges of parameter settings (e.g., Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006). However, we believe that important differences emerge at a conceptual level and that in this respect the LCA framework offers a more straightforward implementation of the concepts we wish to express. Most important, the LCA model can be seen as a module for lexical decision making that can be appended to any computational model of visual word recognition that outputs a measure of lexical activation for a given string of letters as input (see Ziegler, Grainger, & Brysbaert, 2010, for a review of such models). In this way, lexical input values can be fed into the LCA response/decision module on a trial-by-trial basis, thus providing the opportunity for performing item-level analyses of simulation results.

An LCA Account of Lexical Decision

Our LCA model of lexical decision is shown in Figure 1. Applied to the specific case of lexical decision, bottom-up input to the YES response node is evidence for a word that is derived from the stimulus. In this respect we follow standard practice. What is original in our approach is that the input to the NO response node is simply a constant value minus the evidence for a word. This implies that in the absence of any positive evidence for a nonword (the solution adopted here), the bottom-up input to the NO response node is equal to the constant total input value minus the evidence for a word extracted from the stimulus. The use of a constant total input also differentiates the current LCA model from previous proposals (Davis, 2010; Grainger, Dufau, & Ziegler, 2009), in which input to the NO decision node was only a function of elapsed time. The second source of lexical influences in our LCA model arises via the mutually inhibitory connections that exist between the two response nodes, such that a rise in activity in one automatically causes a reduction in activity in the other and vice versa. Both of these mechanisms are fundamental to the LCA approach in general. Finally, the other ingredient of LCA models is the “leak” or decay of information that is associated with the noisy accumulation of information over time.

Implementation

We implemented a linear version of LCA in which accumulators are updated according to the following equations:

\[ y_{\text{YES}}(i + 1) = y_{\text{YES}}(i) + \Delta_{\text{YES}}(i) \]
\[ y_{\text{NO}}(i + 1) = y_{\text{NO}}(i) + \Delta_{\text{NO}}(i) \]

where \( y_{\text{YES}} \) and \( y_{\text{NO}} \) are variables representing the activities of the YES and NO accumulators (here, the calculation of the \( i \)th + 1 cycle), and deltas are variation terms of the corresponding cycle. Deltas are calculated with the following equations:

\[ \Delta_{\text{YES}}(i) = c_{\text{YES}}(i) \cdot [-w \cdot y_{\text{NO}}(i) - k \cdot y_{\text{YES}}(i) + I_{\text{YES}} + G_{\text{YES}}(i)] \]
\[ \Delta_{\text{NO}}(i) = c_{\text{NO}}(i) \cdot [-w \cdot y_{\text{YES}}(i) - k \cdot y_{\text{NO}}(i) + I_{\text{NO}} + G_{\text{NO}}(i)] \]
where $k$ and $w$ are the leakage and inhibition factors, $I_{YES}$ and $I_{NO}$ are the input to the $YES$ and $NO$ accumulators, $G_{YES}$ and $G_{NO}$ are Gaussian noise terms associated with activity in the $YES$ and $NO$ accumulators, and $c_{YES}$ and $c_{NO}$ are bounding factors ensuring that $y_{YES}$ and $y_{NO}$ values stay within $[0;1]$.\(^4\)

\[ I_{YES} + I_{NO} \text{ is a constant.} \tag{5} \]

The model therefore makes decisions on the basis of the accumulation of noisy, leaking, and competing information over time (see Figure 2, Panel A). Whenever the accumulation of activation in either the $YES$ or the $NO$ decision node reaches the criterion value specified for that trial by the setting of response thresholds, a lexical decision response is made and processing is terminated. The response corresponds to the decision node whose threshold was reached first, and the response time corresponds to the time from stimulus onset to the time at which the threshold was reached.

In our simulations, we also maintained the principle of adding noise to the decision criteria that was an essential ingredient of MROM (Figure 2, Panel B; Grainger & Jacobs, 1996; Jacobs & Grainger, 1992). One key aspect of the model is the implementation of trial-by-trial adjustments of response criteria that occur within a simulation. For example, in simulating a lexical decision experiment where the focus is on response speed rather than

\[^4\] Following the equations of interactive activation (McClelland & Rumelhart, 1981), activation is bounded between 0 and 1, but removing this constraint does not affect the performance of our LCA model.
accuracy, the YES and NO threshold values decrease by a constant value on each trial as long as a correct response is given by the model. When an error is made, thresholds are reset to their initial values. In experiments where task instructions stress accuracy rather than speed, a smaller constant value for threshold adjustments is used (see Table 2).

Parameter Settings

Following the parameter setting approach adopted by Grainger and Jacobs (1996), the core parameters of the LCA model were hand-tuned to a target set of data (Experiments 1 and 2 from Ratcliff et al., 2004). These parameter values are given in Table 3. Once the core parameters of our LCA model were fixed, the only parameters that were allowed to vary in order to accommodate specific patterns of results were (a) the values of bottom-up input to the model (to capture changes in stimulus-driven factors) and (b) the values of response criteria, which were modified on a trial-by-trial basis (to capture influences of task and list context).

Bottom-Up Input

The different word frequency conditions were simulated by a gradient of lexical input values ranging from .11 (very low-frequency words) to .12 (high-frequency words; see Table 2). Lexical input differed for the two classes of nonword stimuli tested in the simulations. Orthographically regular, pronounceable pseudowords were assigned a lexical input value of .0875, whereas nonwords formed of random strings of letters were assigned a lexical input value of .0825. Input to the NO decision node was simply a constant value (.2) minus the lexical input. For example, a lexical input of .11 would generate a NO input of .09 (word stimulus), and one of .09 would generate a NO input of .11 (nonword stimulus).

Response Criteria Adjustments

As noted above, one key feature of the present theoretical work is the implementation of trial-by-trial adjustments in response criteria. The initial settings for YES and NO response criteria were drawn independently from a normal distribution of mean .5 and standard deviation .01. The initial value of the mean and the value of the standard deviation did not change across simulations. At the end of every trial, the mean values of both response criteria were adjusted as a function of response accuracy, such that if a correct response was made these values were decreased by a small constant. Whenever an error occurred, the mean values of response thresholds were reset to their initial values. The threshold adjustment constant varied as a function of task instructions, type of nonwords, and percentage of word versus nonword stimuli (see Table 2).

Simulation Studies

Our LCA model was first evaluated against the results of Experiments 1 and 2 of Ratcliff et al. (2004) in Simulation Study 1. This seemed the most appropriate set of benchmark phenomena for an initial evaluation, given that (a) these were the data used to falsify the MROM and (b) these data were used as the unique benchmarks in a recent evaluation of the Bayesian Reader model (Norris, 2009). The model was also evaluated against the experiments of Wagenmakers et al. (2008) in Simulation Study 2 and the English Lexicon Project (ELP) lexical decision data (Balota et al., 2007) in Simulation Study 3. The results of Wagenmakers et al. provide another important benchmark for models of the lexical

### Table 2

<table>
<thead>
<tr>
<th>Simulation</th>
<th>Stimuli/conditions</th>
<th>Lexical input values</th>
<th>Threshold values YES / NO (adjustment)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Words and pseudowords</td>
<td>HF = .12 LF = .11 VLF = .11 NW = .11</td>
<td>YES = .0875 NO = .0875 YES/NO (.003)/.5 (.003)</td>
</tr>
<tr>
<td>2</td>
<td>Words and random strings</td>
<td>HF = .12 LF = .11 VLF = .11 NW = .11</td>
<td>YES = .0875 NO = .0875 YES/NO (.005)/.5 (.005)</td>
</tr>
<tr>
<td>3</td>
<td>Words and pseudowords (accuracy stressed)</td>
<td>HF = .12 LF = .11 VLF = .11 NW = .11</td>
<td>YES = .0875 NO = .0875 YES/NO (.005)/.5 (.005)</td>
</tr>
<tr>
<td>4</td>
<td>Words and pseudowords (speed stressed)</td>
<td>HF = .12 LF = .11 VLF = .11 NW = .11</td>
<td>YES = .0875 NO = .0875 YES/NO (.005)/.5 (.005)</td>
</tr>
<tr>
<td>5</td>
<td>Words and pseudowords (75% pseudowords)</td>
<td>HF = .12 LF = .11 VLF = .11 NW = .11</td>
<td>YES = .0875 NO = .0875 YES/NO (.005)/.5 (.005)</td>
</tr>
</tbody>
</table>

Note. HF = high-frequency; LF = low-frequency; VLF = very low-frequency; NW = nonwords; BIAM = bimodal interactive activation model (Diependale et al., 2010).
decision task, and the ELP simulation provided an initial test of the possible integration of the LCA decision mechanism within the larger framework of models of word recognition and lexical decision (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Grainger & Jacobs, 1996).

In the experiments of Ratcliff et al. (2004), high-frequency (HF), low-frequency (LF), and very low-frequency (VLF) words were intermixed with regular pronounceable nonwords (pseudowords) in Experiment 1 and random letter strings in Experiment 2. Wagenmakers et al. (2008) tested HF, LF, and VLF words intermixed with pseudowords while manipulating speed versus accuracy instructions in their Experiment 1. In these experiments, correct responses to words were faster and accuracy was higher when nonwords were random strings than when they were pseudowords, and responses were faster but accuracy was lower when speed was stressed in the instructions to participants. Wagenmakers et al. (2008) reported the results of a second experiment in which they manipulated the proportion of word versus nonword stimuli, with one condition having 75% words and another condition having 75% nonwords. Overall, their results showed that RTs were faster and accuracy was higher for the stimulus category (word vs. nonword) that occurred the most frequently in a given condition.

The simulations reported here were structured like a behavioral experiment. Each simulation study consisted of 50 consecutive runs, equivalent to 50 virtual participants. Each run was composed of 750 words and 750 nonwords randomly ordered, except for the simulations in which word/nonword proportion was manipulated, where there were 1,125 words and 375 nonwords in the 75% word simulation and 375 words and 1,125 nonwords in the 75% nonword simulation. Number of cycles to reach either the YES or the NO decision threshold was recorded on each trial, as well as the winning response (YES or NO).

### Simulation Study 1

Simulation Study 1 used the conditions tested in Experiments 1 and 2 of Ratcliff et al. (2004), in which HF, LF, and VLF words were either intermixed with orthographically regular and pronounceable nonwords (pseudowords, Experiment 1) or with random letter strings (random strings, Experiment 2). The effects of this manipulation of type of nonword were simulated by a change in the rate of the trial-by-trial adjustments of response criteria (see Table 2). The rationale behind this particular simulation is discussed below. Figure 3 shows the scatterplots of simulated and experimental mean RTs for the different conditions tested in these experiments. The 16 mean RTs are derived from the combination of stimulus type (HF, LF, and VLF words, and nonwords), correctness of response (correct and error), and type of nonword (pseudowords and random strings). The results of these simulations clearly show that our LCA model provides a good fit to the experimental data. It should be noted that the LCA model does not exhibit the same problem as does the Bayesian Reader model, which systematically underestimated error RTs in these experiments (see Norris, 2009, Tables 1 and 2).

The simulations reported here were structured like a behavioral experiment. Each simulation study consisted of 50 consecutive runs, equivalent to 50 virtual participants. Each run was composed of 750 words and 750 nonwords randomly ordered, except for the simulations in which word/nonword proportion was manipulated, where there were 1,125 words and 375 nonwords in the 75% word simulation and 375 words and 1,125 nonwords in the 75% nonword simulation. Number of cycles to reach either the YES or the NO decision threshold was recorded on each trial, as well as the winning response (YES or NO).

**Figure 3.** Performance of the LCA model in simulating the response time (RT) condition means from Ratcliff et al. (2004, Experiments 1 and 2). $R^2$ is the percentage of explained variance. Correct responses are in dark gray, errors are in light gray. PW = pseudowords; HF = high frequency; LF = low frequency; VLF = very low frequency; RS = random strings; LCA = leaky competing accumulator.
lexical decision, as implemented in our LCA model, can overcome the fundamental limitations of standard deadline models. We return to this point in greater detail in the General Discussion.

**Simulation Study 2**

Simulation Study 2 tested the LCA against the data from Wagenmakers et al. (2008) in which HF, LF, and VLF words were intermixed with orthographically regular, pronounceable pseudowords. The same parameter set as in Simulation Study 1 was maintained in these simulations. In Experiment 1 of Wagenmakers et al. (2008), participants were given either instructions that stressed accuracy of responding or instructions that stressed speed of responding. The effects of this manipulation were simulated by changes in the rate of trial-by-trial adjustments in response criteria. In the speed condition, adjustment of both response criteria was greater than in the accuracy condition (see Table 2). In Experiment 2, Wagenmakers et al. manipulated the percentage of words versus pseudowords in the experiment (75% words in one condition and 25% in the other condition). Once again, these manipulations were simulated in LCA by trial-by-trial adjustments of response criteria. In the condition with a high proportion of words, adjustment of the YES decision node was greater than adjustment of the NO decision node, and in the condition with a high proportion of nonwords, the opposite adjustment was implemented (see Table 2).

Figure 4 presents the results of simulations run on the LCA model. The core parameters of the LCA model were the same as those used in Simulation Study 1. Quite remarkably, by simply modifying the rate of adjustment of response criteria, the LCA model captured practically the same amount of variance in mean RT (82%) as in Simulation Study 1 that involved not only a different experimental manipulation (type of nonword), but also different stimuli and participants. Once again, the LCA model captures key patterns in the data that the MROM could not capture. First, in the speed instructions condition tested by Wagenmakers et al. (2008), the LCA model correctly simulates the faster RTs obtained to error responses than to correct responses (diamond shapes in Figure 4), and one can also see a clear frequency effect in both the correct and the error RTs that is well captured by the LCA model. Figure 5 (left panel) shows the distribution of simulated correct and error RTs to word targets in the speed instructions condition, and it therefore demonstrates that RTs can be faster on errors than on correct responses without RTs in the leading edge of the distribution being too fast. Second, correct RTs to nonword stimuli can be faster than error RTs to nonword stimuli, as was the case in the accuracy instructions condition of Wagenmakers et al. and as captured by LCA (circle shapes in Figure 4). This is further demonstrated in Figure 5 (right panel), showing the distribution of simulated RTs for correct and error responses to nonwords in the accuracy instructions condition of Wagenmakers et al. Here we clearly see a shift in the complete RT distribution, with correct RTs being faster than error RTs for nonword stimuli. Furthermore, Wagenmakers et al. demonstrated that this pattern of correct and error RTs to nonword stimuli is exaggerated in the 75% nonword condition they tested, but the opposite pattern is seen in the 75% word condition. Figure 4 shows that LCA perfectly captures this pattern. Finally, Figure 6 shows the mean accuracy per condition

![Figure 4](image-url)
in Simulation Studies 1 and 2, in order to demonstrate that the LCA model’s success in simulating mean RT is not achieved at the cost of accuracy.

Summarizing the results of Simulation Studies 1 and 2, we conclude that the LCA model provides an excellent fit to the major data patterns presented by Ratcliff et al. (2004) and Wagenmakers et al. (2008). Most important, with the same set of core (nonmodifiable) parameters, our model generated excellent fits to the data of Wagenmakers et al. (2008). Via theoretically motivated, trial-by-trial adjustments in the values of response criteria.

Simulation Study 3

Simulation Study 3 tested the LCA model against the data from the English Lexicon Project (ELP; Balota et al., 2007) in which words were intermixed with orthographically regular, pronounceable pseudowords. From the 40,000+ words in the ELP database, we extracted the 5-letter and 6-letter words (842 words) that had been tested in Ratcliff et al. (2004). We also randomly selected 842 5-letter and 6-letter nonwords from the ELP database. These words and nonwords were then submitted as input to the bimodal version of the interactive activation model (Diependaele, Ziegler, & Grainger, 2010). For each stimulus, we took the activation level of the most activated word unit at the 15th cycle (i.e., before asymptote). These activation values were then linearly transformed to fit the range of input values used in Simulation Studies 1 and 2, and these values were used as input to the YES response node in LCA. In this simulation, the same core parameter values as in Simulation Studies 1 and 2 were maintained (see Tables 2 and 3). The value of the threshold adjustment parameter was set to a small value for both thresholds, given the large number of trials in these simulations.

Figure 5. Left panel: Distributions (quantile means) of simulated correct and error response times (RTs) in the LCA model for high-frequency (HF), low-frequency (LF), and very low-frequency (VLF) words under the speed instruction conditions of Wagenmakers et al. (2008). Right panel: Distributions (probability density) of simulated correct and error RTs to nonwords under the accuracy instruction conditions of Wagenmakers et al. (2008). LCA = leaky competing accumulator.

Figure 6. Mean accuracy in the six experiments and simulations of the present study. In black, pseudowords; in dark gray, very low-frequency words; in light gray, low-frequency words; in white, high-frequency words. PW = pseudowords (Ratcliff et al., 2004, Experiment 2); RS = random strings (Ratcliff et al., 2004, Experiment 1); Accuracy = accuracy stressed (Wagenmakers et al., 2008, Experiment 1); Speed = speed stressed (Wagenmakers et al., 2008, Experiment 2); 75W = 75% words (Wagenmakers et al., 2008, Experiment 3); 75NW = 75% nonwords (Wagenmakers et al., 2008, Experiment 4).
bases is probably no more than 40% (see Rey, Courrieu, Schmidt-Weigand, & Jacobs, 2009). Furthermore, this result is quite encouraging, given that no parameter tuning was performed in order to increase the amount of variance that could be explained. Figure 7 also shows that LCA generates RT distributions that change as a function of word frequency and lexical status (word vs. nonword), much in the same way as the empirical distributions. To our knowledge, this is the first simulation of large-scale lexical decision data that includes a simulation of nonword responses as well as word responses. It is also the first simulation of large-scale lexical decision data involving trial-by-trial adjustments in response criteria (see Perea et al., 2004, for similar simulations of a standard lexical decision experiment). The relative success of this simulation opens up a whole new perspective for analyzing sequential dependencies in such large-scale data and examining the consequences of adjustments in response criteria on the evolution of correct and error RTs as well as error rates in such data.

**General Discussion**

The lexical decision task is probably the most popular task used to study silent single-word reading. Yet, we still have no clear understanding of how participants in a lexical decision experiment perform this task. In particular, it is not at all clear how participants respond negatively when the stimulus is not a word. One simple solution to this problem, the temporal deadline mechanism (implemented in the MROM; Grainger & Jacobs, 1996), was rejected on the basis of recent analyses of the patterns of RTs obtained to correct and error responses in the lexical decision task (Ratcliff et al., 2004; Wagenmakers et al., 2008; see Table 1). In the present work, we tested a dynamic deadline account of lexical decision implemented within the framework of an LCA (Usher & McClelland, 2001). This model has two decision nodes, one for a YES decision and one for a NO decision. Input to the YES decision node is a function of lexical activity generated by the stimulus, whereas input to the NO decision node is equal to a constant value minus...
the input to the **YES** decision node. The activation dynamics of the two decision nodes follow the standard equations of LCAs (Usher & McClelland, 2001), with decay over time (leak) and lateral inhibition between nodes (competition).

Simulation Studies 1 and 2 provided a test of LCA on two independent sets of experimental data. Simulation Study 1 was used to hand-tune the core parameters of LCA. Having fixed these core parameter values, LCA was able to capture the effects of type of nonword (pseudowords vs. random strings) by adjusting the lexical input generated by these two types of stimuli and by adjusting the rate of adjustment of response criteria. LCA captured 83% of the variance across the 16 mean RTs reported by Ratcliff et al. (2004, Experiments 1 and 2).

The same core parameters whose values were adjusted to fit the empirical data in Simulation Study 1 were used in the second simulation study. The only parameter that changed between these two simulation studies was the value of the trial-by-trial adjustment of response criteria. Despite the absence of any systematic parameter tuning, LCA accounted for 82% of the variance across the 32 mean RTs reported in Wagenmakers et al. (2008). More precisely, the model provided an accurate account of variations in mean RT and error rate to word and nonword stimuli as a function of changes in speed versus accuracy instructions and the percentage of words versus nonwords in the experiment. Overall, the results of Simulation Studies 1 and 2 showed that the LCA model easily passed the tests that had been fatal for the MKRM (see Table 1) and that proved problematic for Norris’ (2009) Bayesian Reader model (i.e., incorrect predictions concerning error RTs). We analyze this particular aspect of the simulation results in more detail in the following section.

Simulation Study 3 tested LCA against a subset of words and nonwords extracted from the English Lexicon Project database of item-level lexical decision latencies. Input to the **YES** response was word activation level as calculated by the bimodal interactive activation model (Diependaele et al., 2010), which provides a measure of word activity for word and nonword stimuli of varying length. The core parameter values used in Simulation Studies 1 and 2 remained unchanged. The model provided a good fit to the item-level data, suggesting that it can be usefully appended to word recognition models in order to account for performance in the lexical decision task.

In the following sections, we first analyze precisely how the LCA model overcomes the deficiencies of its predecessor: the multiple read-out model (MRM) of lexical decision (Grainger & Jacobs, 1996). Then we discuss the main contributions of the present approach in increasing our understanding of how participants actually perform the lexical decision task. In particular, we describe the LCA solution to the hard question in modeling lexical decision, that is, what constitutes evidence for a nonword? Finally, we discuss a key novel aspect to the current approach, that is, the implementation of trial-by-trial adjustments in response criteria.

### The Four Deficiencies of MROM

In Table 1 we listed the four deficiencies of the MROM (Grainger & Jacobs, 1996), as identified in the work of Ratcliff et al. (2004) and Wagenmakers et al. (2008). Here, we analyze the success of the LCA model in passing these tests.

Considering first of all Points 1 and 2, both of which concern RTs in response to word stimuli. MROM was found to be incapable of producing fast error RTs to word targets without generating excessively fast RTs in the first decile of the RT distribution (Point 1), and error RTs to word targets in MROM did not show a word frequency effect that is seen in the empirical data (Point 2). The LCA simulation of the speed instruction condition tested by Wagenmakers et al. (2008) revealed faster error RTs than correct RTs to word targets (see Figure 4) without distorting the RT distribution (see Figure 5, left panel). Also, the results of Simulation Studies 1 and 2 both revealed a word frequency effect in the error RTs to word stimuli (see Figures 3 and 4).

Next, let us consider Points 3 and 4, both of which concern RTs in response to nonword stimuli. MROM was found to generate RT distributions for correct responses to nonword stimuli that did not have the same rightward skew as the distributions for correct responses to words (Point 3), and MROM could not generate fast correct RTs to nonwords (and never faster than error RTs to nonwords) without generating too many errors (Point 4). Figure 5 (right panel) reveals that the RT distributions for nonword stimuli in the LCA model show the distinctive rightward skew. The simulation results reported in Figure 3 clearly demonstrate that the LCA model can generate RTs for correct responses to nonword targets that are faster than the error RTs to the same targets. Finally, accuracy in the LCA model compares favorably with accuracy in the experiments, as shown in Figure 6.

### Input to the **NO** Response

Having abandoned a deadline account of generating **NO** responses in lexical decision, as implemented in the MROM (Grainger & Jacobs, 1996), we note that the hard question for alternative accounts of lexical decision, such as the diffusion model (Ratcliff et al., 2004), the Bayesian Reader (Norris, 2009), or the present LCA model, is to define what constitutes evidence for a nonword. Defining evidence for a word (i.e., a **YES** response) is relatively straightforward. In the first two simulation studies of the present work, we followed Ratcliff et al. in using word frequency to determine these input values. In the third simulation study, we used a measure of lexical activity derived from a computational model of visual word recognition (Diependaele et al., 2010). The main contribution of the present work concerns our proposal for how to calculate evidence for a nonword.

In the LCA model, input to the **NO** decision node (i.e., evidence for a nonword) is equal to a constant value minus the input to the **YES** decision node (i.e., a measure of lexical activity generated by the stimulus). The model therefore has to know only two things in order to calculate input to the **NO** decision node: the amount of lexical activity generated by the stimulus and the total input value that optimizes speed and accuracy of responses to words and nonwords.

So, a key remaining question here concerns how the total input value could be computed online during the first trials of a lexical decision experiment? One simple solution is to monitor the lexical activity generated by word and nonword stimuli on each trial (these correspond to the lexical input values given in Table 3, and to compute the sum of these two values. This sum can be used to provide an initial estimate of the total input value after having seen only one word and one nonword in a lexical decision experiment.
This initial value can then be adjusted in order to optimize performance, much in the same manner as response criteria are adjusted on a trial-by-trial basis in our model. Increasing the total input value will generate faster RTs to nonwords but will also generate more errors to word stimuli, as input to the NO response will eventually be greater than input to the YES response. Error-driven adjustments of this value can therefore be used to rapidly home in on an optimal setting. A similar mechanism could be applied to the diffusion model (Ratcliff et al., 2004), where evidence for a word and evidence for a nonword would be averaged, rather than summed, in order to provide an initial estimate of the drift rate criterion.

It is important to note here that our proposed mechanism for computing the optimal value of the total constant input in LCA is analogous to the way in which response criteria are adjusted in our model. This implies that the modifiable parameters in our LCA model all involve either stimulus-driven changes in lexical input values or trial-by-trial adjustments of the mean values of the YES and NO decision thresholds and the total constant input in order to optimize performance.

Online Adjustments of Response Criteria

One of the key features of the LCA model, in line with the general approach to modeling adopted in the MROM (Grainger & Jacobs, 1996), is that effects of list context and task instructions are simulated via adjustments of response criteria. In the present study, we demonstrated that such adjustments in the thresholds of the YES and NO decision nodes provided an accurate account of how different types of nonword affect responding to word stimuli (an effect of list context), how responses to word and nonword stimuli are affected by the proportion of words versus nonwords in the experiment (another effect of list context), and how responses to word and nonword stimuli are affected by stressing speed over accuracy or vice versa (an effect of task instructions). The general principle that was applied in this endeavor is that participants in a lexical decision experiment are trying to optimize their performance in line with the instructions they receive and do so by monitoring their performance on a trial-by-trial basis. This critical information provided by such monitoring is whether or not a mistake was made, and it is this information about accuracy of responding that was used to adjust response criteria in the present work.

In implementing such trial-by-trial adjustments of response criteria, we follow in the steps of a number of prior attempts to describe such adjustments in the light of empirical data concerning post-error responding (e.g., Laming, 1968; Rabbit, 1966; Rabbit & Rodgers, 1977). In particular, conflict monitoring theory (Botvinick, Braver, Barch, Carter, & Cohen, 2001) accounts for error-related changes in performance by postulating the occurrence of conflict in association with errors. Conflict is used to change the activation level of output units, which is of course functionally equivalent to modifying response thresholds, as in the present approach. This allowed Botvinick et al. to account for the standard pattern of slower RTs and higher levels of accuracy following an error (but see Nottebohm et al., 2009, for different patterns of post-error performance).

We implemented a relatively simple version of this general “monitor and adjust” principle, which despite its simplicity provided an accurate account of the behavioral data. The mean values of the thresholds of YES and NO decision nodes were lowered on every correct trial by a small constant value that was determined by task instructions and list context, and the rate of change could be different for the two decision nodes. Every time there was an error, the values of both thresholds were reset to their initial values.

To simulate the influence of type of nonword on responses to both nonword and word targets, we increased the rate of adjustment of response thresholds in the easy nonword condition (i.e., random strings), and this increase was slightly greater for the NO threshold than the YES threshold. The NO threshold can be lowered to a certain extent without an increase in false negative errors, as long as the YES threshold is lowered at the same time. The YES threshold can be lowered to a certain extent without an increase in false positive errors because random string nonword stimuli have high input activation to the NO response node. Similar adjustments of the rate of change of response thresholds were used to simulate the effects of the proportion of word versus nonword stimuli in an experiment. With more words than nonwords, the YES response threshold was lowered by a greater amount on each correct trial, and with more nonwords than words it was the NO response threshold that was adjusted by a greater amount. The lowering of response thresholds as a function of the proportion of words and nonwords enables optimization of responding for a majority of trials in the experiment. Finally, with speed instructions, the trial-by-trial threshold adjustment was increased by the same amount for both decision nodes, thus simulating participants’ attempts to respond faster to both word and nonword stimuli. This generated more false negative and false positive errors, and the resulting error RTs were faster due to the fact that they are mostly generated with the lowest settings of response criteria. Because errors immediately generate a resetting of response criteria to their initial values, correct responses in these conditions are therefore slower on average than error responses.

Thus, overall our simple implementation of trial-by-trial adjustments in response criteria provided an accurate simulation of the effects of list context and task instructions as tested in the studies of Ratcliff et al. (2004) and Wagenmakers et al. (2008). Although this study is clearly only a first approximation with respect to how participants in a lexical decision experiment adjust to such contextual factors, the relative success of this approach encourages further tests of more sophisticated adjustments in response criteria with the aim to capture trial-by-trial changes in human performance.

Conclusions

In 1996, Grainger and Jacobs presented an influential account of the lexical decision task (the multiple read-out model; MROM), in which time from stimulus onset was used to trigger NO responses via a temporal deadline. This account has been rejected on the basis of results showing fast NO responses in certain conditions in lexical decision experiments (Ratcliff et al., 2004; Wagenmakers et al., 2008). Here we implemented and tested a dynamic deadline model of the lexical decision task using the framework of leaky competing accumulators (Usher & McClelland, 2001). The LCA model was shown to overcome the deficiencies associated with the MROM, and overall it provided an accurate account of variations
in RTs for correct and error responses to word and nonword stimuli as a function of the type of nonword, the proportion of words versus nonwords in the experiment, and whether or not task instructions stressed speed or accuracy of responding. Finally, by using lexical activity calculated by a computational model of visual word recognition as lexical input to the LCA decision module, we were able to provide the first item-level simulation of both word and nonword responses in a large-scale database.

References


Received February 18, 2011
Revision received December 1, 2011
Accepted December 5, 2011
AUTHOR PLEASE ANSWER ALL QUERIES

AQ1: Author. Table 3 was cited before Table 2, so they were switched.

AQ2: Author. Figure 7. The caption could make a clearer distinction between the panels.