# Is perceptual acuity asymmetric in isolated word recognition? Evidence from an ideal-observer reverse-engineering approach

**Nathaniel J. Smith and Wen-Hsuan Chan**
**njsmith@cogsci.ucsd.edu, whchan@ucsd.edu**
University of California at San Diego, Department of Cognitive Science
9500 Gilman Drive #515, La Jolla, CA 92093-0515 USA

**Roger Levy**
**rlevy@ling.ucsd.edu**
University of California at San Diego, Department of Linguistics
9500 Gilman Drive #108, La Jolla, CA 92093-0108 USA

**Abstract**

An asymmetrical optimal viewing position (OVP) effect in isolated word recognition has been well documented, such that recognition speed and accuracy are highest when the point of fixation within the word is slightly to the left of center. However, there remains disagreement as to the source of the asymmetry in the OVP effect. One leading explanation is that perceptual acuity in isolated word recognition is asymmetric, falling off more rapidly to the left than to the right. An alternative explanation is that of lexical constraint: perceptual acuity may be symmetric, but that the distributional statistics of the lexicon are such that the letters near the beginning of a word are on average of greater value in discriminating word identity than the letters near the end. On both these accounts, a left-of-center fixation point optimizes the efficient accrual of perceptual input from the word string, but for different reasons. These accounts have been difficult to tease apart experimentally due to the ubiquitous potential influence of lexical constraint. Here we take a novel approach, constructing an ideal-observer model of isolated word recognition which takes into account word frequency information and thus intrinsically accounts for the role of lexical constraint. Within this model, the shape of the perceptual acuity curve is governed by free parameters that can be estimated from purely behavioral response data from word recognition experiments. Fitting our model to the experimental data of Stevens & Grainger (2003), we find that the asymmetric version, in which perceptual acuity can differ to the left and to the right, fits human behavioral responses significantly better than symmetric versions in which the perceptual acuity curve is constrained to be the same to the left and to the right. Furthermore, in both parametric and nonparametric versions of the asymmetric model, perceptual acuity falls off more rapidly to the left than to the right. These results support the position that the perceptual acuity curve in isolated word recognition is indeed asymmetric.

**Keywords:** Psychology, Cognitive Science, Perception, Language Understanding, Decision Making, Bayesian modeling

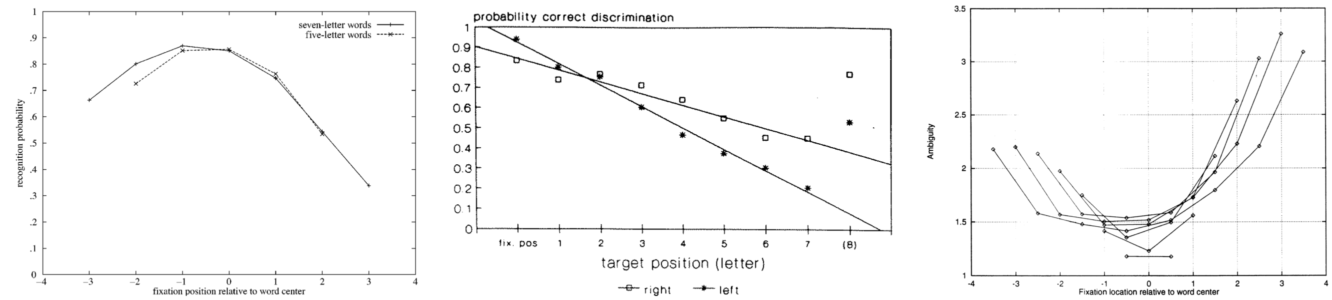
## Introduction

Literate native speakers are exquisitely adapted to the visual and linguistic processing of written text in their language. The naturalistic task underlying most of this adaptation is reading (Rayner, 1998). In the study of eye movements in reading, one of the most striking examples of this adaptation in the last several decades has been discovery of *asymmetry of the perceptual span*: in languages which are written from left to right, readers are more sensitive to material to the right of the center of fixation on the page than they are to material on the left. In languages which are written from right to left, however, this sensitivity is reversed (Rayner, Well & Pollatsek, 1980). Since visual acuity per se is not itself asymmetric (as evidenced by experimental work on perception of non-linguistic visual inputs), the most intuitive interpretation of this finding is that, in ordinary progressive reading, because readers of languages written left to right have already seen what lies to the left of their eyes, they differentially attend to what lies to the right (and vice versa for languages written right to left).

However, the discovery by O'Regan, Lévy-Schoen, A., Pynte, J. & Brugaillére (1984) of the optimal viewing position (OVP) in *isolated* word recognition makes the picture more complex. The OVP in isolated word recognition can be succinctly described as follows: word recognition is fastest and most accurate when the initial fixation point of the eyes is slightly to the left of the center of the word (Figure 1a). The discovery of the OVP launched considerable discussion as to its nature and implications, since it cannot be obviously accounted for by an asymmetry in perceptual acuity that would be adaptive for the task.

At present, there are two leading explanations that have been proposed for the OVP in isolated word recognition. One is that the asymmetry of perceptual acuity to the left and to the right within reading may affect *all* processing of visual linguistic input. If acuity drops off more rapidly to the left of the fixation point than to the right, then the best strategy to recognize a word would be to fixate at a left-of-center location, maximizing average acuity across the word as a whole. Evidence supporting this position has been adduced by Nazir, O'Regan & Jacobs (1991) and Nazir, Heller & Sussman (1992), who demonstrated left-right acuity differences in tasks involving the detection of a target letter at a variable position within a masking letter string (e.g., *kkkkkykk*). They found that the drop-off in performance was a monotonic function of visual eccentricity, and the left visual field showed steeper drop-off than the right visual field (Figure 1b).

The other leading explanation was put forth by Clark & O'Regan (1999), who argued that a better way to understand the contributions of these different mechanisms may lie in the distributional statistics of the written lexicon itself. They investigated the contributions of orthographic constraints, con-




(a) OVP effect (Stevens & Grainger, 2003) (b) Asymmetric perceptual acuity in letter-in-mask recognition (Nazir et al., 2003) (c) Asymmetry of lexical constraint (Clark & O'Regan, 1999)

Figure 1: The OVP effect in isolated word recognition and possible explanations for it.

structing a simple measure of residual lexical ambiguity that would hold assuming only that two letters near the fixation point and at the end of the word are known. For a range of word lengths in both French and English, this ambiguity measure is minimized just to the left of the word's center, capturing the OVP effect through lexical constraint without resorting to an asymmetric perceptual acuity curve (Figure 1c).

It is difficult to adjudicate between these two possible explanations through purely experimental means, because well-established effects on word recognition such as those of word frequency and neighborhood density make it is fairly clear that lexical constraint plays a ubiquitous role in the process but varies slightly for every word in the lexicon, making it difficult to design a word-recognition experiment that controls for lexical constraint while testing perceptual acuity. Conversely, our understanding of precisely how the perceptual acuity curve affects word recognition remains limited, making it difficult to hold it constant and test only lexical constraint as a possible source of OVP effects. In this paper, we take an alternative approach, constructing an ideal-observer model (Marr, 1982; Anderson, 1990) of isolated word recognition in which lexical constraints are assumed to be available. The perceptual acuity curve in this model is determined by a set of parameters which are free and can be fit to behavioral data using well-established techniques of statistical inference. Our ideal-observer model thus allows us to "reverse-engineer" the perceptual acuity curve active in isolated word recognition on the basis of a lexicon (with word frequencies) and a behavioral data set, and assess whether and how the reverse-engineered acuity curve may be asymmetric.

## Data

The dataset used in our study is from Experiment Two of Stevens & Grainger (2003). In this experiment, words were presented for 50ms each at various positions relative to the center of the fixation. After presentation of each word, the participant was asked to type the presented word back into the computer.. The dataset contains 75 five-letter words and 105 seven-letter words and the human performance (correctness of the response) for each word from seventy subjects, for a total of 12,600 observations balanced across word identity and fixation position. We obtained a word frequency database from the Agence France Presse (AFP) French Corpus, which contains French journal articles from 1993-1996. Since the human dataset has both five-letter and seven-letter French words, we extracted frequency counts for all five-letter and seven-letter words in the corpus. The resulting lexicon contains 14,379 five-letter French words and 21,569 seven-letter French words.

## Model

### The Bayesian Reader

We use an ideal-observer model of isolated word recognition, the Bayesian Reader (Norris, 2006, 2009). Our version of the Bayesian Reader introduces several key assumptions regarding the nature of word recognition which determine the form of the probabilistic model:

- Word recognition is a Bayesian hypothesis test in which prior expectations regarding what word is likely to be presented are combined with perceptual evidence **d** to determine posterior beliefs about what word $w$ is being seen:[1]

$$P(w|\mathbf{d}) = \frac{P(\mathbf{d}|w)P(w)}{P(\mathbf{d})};$$

- The prior probability of each word $P(w)$ is proportional to its corpus frequency of occurrence;
- Perceptual evidence consists of a sequence of independent identically distributed (i.i.d.) input samples $d^{(1)}, \ldots, d^{(N)}$ drawn from a NOISE DISTRIBUTION $P(d|w)$, where samples accrue at a constant rate over time;
- If we denote the letters of a word $w$ of length $L$ as $w_1, \ldots, w_L$, then an input sample $d$ can be decomposed into $L$ samples $d_1, \ldots, d_L$, with $d_i$ conditionally independent of $d_j$ given $w_i$ and $w_j$ for all $i \neq j$, so that

[1] Tasks in which the decision to be made is something other than the identity of the word—e.g., a lexical decision about whether the input string is a word in the participant's language—can be formulated as Bayesian hypothesis tests among the possible choices (Norris, 2006; del Prado Martìn, 2008).

$$P(d|w) = \prod_{i=1}^{L} P(d_i|w_i).$$

We will refer to the term $P(d_i|w_i)$ as the noise distribution for letter $w_i$.

Thus far, this version of the Bayesian Reader is simpler and more general than that introduced by Norris (2006), who assumed (a) a specific representation of each sample $d$ as a point in a high-dimensional space; (b) a multivariate Gaussian form for the noise distribution $P(d|w)$; and (c) a specific estimate of the noise variance used by the ideal observer in computing the posterior distribution over words given perceptual evidence.

## Adaptation to modeling visual acuity curves

In order to adapt the Bayesian Reader to the task of estimating visual acuity curves, however, we need to introduce a dependence of the noise distribution for each letter on its physical positioning. In particular, we assume that the noise distribution for each letter is dependent on its eccentricity as measured in number of characters from the point of fixation, with negative values corresponding to left-of-fixation and positive values to right-of-fixation; and on its proximities $v^L$ and $v^R$ to the left and right edges of the word, also measured in characters. If we define these physical positioning characteristics of a letter $w_i$ as $k_i \equiv \langle e_i, v_i^L, v_i^R \rangle$, then its position-contingent noise distribution can be denoted as $P(d_i|w_i, k_i)$.

For the empirical modeling studies presented here, we make the additional assumption that the value of

$$E_{d_i|w_i^*,k_i}\left[P(d_i|w_i,k_i)\right], \tag{1}$$

where $E_{d_i|w_i^*,k_i}$ denotes expectation under the conditional distribution $P(d_i|w_i^*, k_i)$ for the true letter $w_i^*$ being presented, depends only on $k$ and on whether $w_i = w_i^*$. This assumption can be interpreted as stating that every letter is equally confusable with all letters other than itself; the quantity in (1) for $w_i \neq w_i^*$ can be interpreted as the level of confusability of a letter as a function of its physical positioning. This assumption is not necessary within the overall framework, and indeed could be relaxed in order to incorporate letter confusability matrices (Engel, Dougherty & Jones, 1973; Geyer, 1977) into the model and even to learn them directly from behavioral word-recognition data. In the present studies, however, this assumption greatly simplifies and facilitates both the statistical learning problem and its computational implementation.

## Learning visual acuity from word identification data

Recall that in their word-identification study, Stevens & Grainger (2003) presented experimental participants with five- and seven-letter words one at a time for a brief, fixed interval too short to permit refixation, with fixation position varying across trials. The behavioral response $r$ in each trial was the participant's guess as to which word they saw. Our goal is to use these behavioral responses to learn the dependence of the visual acuity of a letter—as quantified by confusability in (1)—on its physical positioning.

We model the naming task using the assumptions outlined in the previous two sections. In general, the experimental participant must choose their response $r$ through some possibly stochastic decision process based on their posterior beliefs $P(w|\mathbf{d}, k)$ about what word they saw. We further assume that the participant makes their choice of response through probability matching, so that the probability of any response $r$ given the word $w^*$ actually being presented is given by its expected posterior probability:

$$P(r|w^*, k) = E_{\mathbf{d}|k,w^*}\left[P(r|\mathbf{d}, k)\right] \tag{2}$$

where $E_{\mathbf{d}|k,w^*}$ represents the expectation marginalizing over possible perceptual input samples given the true word and its physical positioning. For notational simplicity, we omit the subscript on the expectation whenever it is clear from context.

Equation (2) can be rewritten using Bayes' rule as

$$P(r|w^*, k) = P(r)E\left[\frac{P(\mathbf{d}|r,k)}{P(\mathbf{d}|k)}\right]. \tag{3}$$

Equation (3) is the expectation of a ratio of random variables $E[Y/X]$, an expression which cannot in general be manipulated exactly. Using the method of propagation of error, however, a second-order approximation for the expectation of a ratio can be found (Rice, 1995):

$$E\left[\frac{Y}{X}\right] \approx \frac{E[Y]}{E[X]} + \frac{1}{E[X]^2}\left(\mathrm{Var}[X]\frac{E[Y]}{E[X]} - \mathrm{Cov}[X,Y]\right).$$

We now turn our attention to the rightmost part of this expression, the covariance between the numerator and the denominator—in Equation (3), these terms are $P(\mathbf{d}|r,k)$ and $P(\mathbf{d}|k)$ respectively. Insofar as any individual word plays only a small part in the calculation of the marginal probability $P(\mathbf{d}|k)$, we would expect the covariance of this marginal probability with $P(\mathbf{d}|r,k)$ to be small (with the important caveat that because words tend to look more like each other than like non-words, there will generally be some positive covariance, and its magnitude may depend on the orthographically typicality of $w^*$ and $r$). Dropping the covariance from the above approximation allows us to approximate our posterior probability as

$$P(r|w^*, k) \approx P(r)\frac{E\left[P(\mathbf{d}|r,k)\right]}{E\left[P(\mathbf{d}|k)\right] + \frac{\mathrm{Var}[P(\mathbf{d}|k)]}{E[P(\mathbf{d}|k)]^3}}.$$

Ignoring the denominator (which is constant with respect to $r$) and decomposing the perceptual input $\mathbf{d}$ into its component independent samples at each time $j$ and letter position $i$, we obtain

$$P(r|w^*,k) \propto P(r) \prod_{i,j} E_{d_i^{(j)}|k_i,w_i^*}\left[P(d_i^{(j)}|r_i,k_i)\right].$$

We take advantage of the identical distribution of the $N$ samples to obtain the approximate unnormalized probability

$$P(r|w^*,k) \propto P(r) \prod_i \left(E_{d_i|k_i,w_i^*}\left[P(d_i|r_i,k_i)\right]\right)^N.$$

We are now ready to take advantage of our assumption from the previous section that each letter is equally confusable with all letters other than itself—that is, the value of each of the above terms $\left(E_{d_i|k_i,w_i^*}\left[P(d_i|r_i,k_i)\right]\right)^N$ depends only on $k_i$ and on whether $w_i = w_i^*$. For each $k_i$, let us denote the value taken when $w_i = w_i^*$ as $p_i$, the value taken when $w_i \neq w_i^*$ as $q_i$, and the ratio $\frac{p_i}{q_i}$ as $l_i$. Substituting these terms in and dividing the entire expression by $q_1 \dots q_L$ gives us our final approximate expression for the probability of the participant's response:

$$P(r|w^*,k) \propto P(r) \prod_{i:w_i=w_i^*} l_i \quad (4)$$

where $l_i$ is dependent on the physical positioning of the letter in question. On the original assumption of the Bayesian Reader that the number of input samples $N$ accumulates at a constant rate over time, the value $\log l_i$ can be interpreted as the average rate at which perceptual information accrues at position $i$.

### Model parameterization and estimation

Within the context of our model, the goal of inferring a visual acuity curve from behavioral word-recognition data entails estimating these input-accrual rate parameters $\log l_i$. In the studies presented here, we assume as stated before that the $\log l_i$ for each letter is a function of three properties of its physical position: its eccentricity from the fixation point, its proximity from the left edge of the word, and its proximity from the right edge of the word. All measurements are made in characters. We consider two functional forms for the eccentricity parameters: a PARAMETRIC form in which the values of $\log l_i$ is assumed to follow a Gaussian curve centered at the fixation point with maximum value $\alpha$ and standard deviation $\sigma$; and a NONPARAMETRIC form in which each eccentricity has its own arbitrary parameter value. For each form, we consider a SYMMETRIC version in which the eccentricity parameters $\log l_i$ are determined by the absolute eccentricity, and an ASYMMETRIC version in which the parameters for negative and positive eccentricity values of the same magnitude can be different (in the parametric Gaussian case, the asymmetric model allows different standard deviations $\sigma_L$ and $\sigma_R$ to the left and the right of the fixation point). Additionally, all models include one left-edge and one right-edge "bonus" parameter, $b_L$ and $b_R$, added to the eccentricity parameters to determine the input-accrual rate parameters $\log l_i$ for the first and last letters of a word respectively. That is, if the fixation position is on the $f$-th character of the word and the function $e(\cdot)$ maps eccentricities to values in $\log l_i$ space, we have

$$\begin{aligned} \log l_1 &= e(1-f) + b_L && \text{(first letter)} \\ \log l_i &= e(i-f) && \text{(middle letters; } i \notin \{1,L\}) \\ \log l_L &= e(L-f) + b_R && \text{(last letter)} \end{aligned}$$

In all cases, we fit the parameters of our models using maximum likelihood estimation. Fortunately, the gradient of our model is readily calculable and allows estimation using standard gradient-descent techniques.

## Results

For each of our nonparametric ($-$PAR) and parametric (+PAR) models, we fit a symmetric (+SYM) and an asymmetric ($-$SIM) variety to the 12,600-observation dataset of Stevens & Grainger (2003, Experiment 2) estimation. This dataset contains presentations of both five-letter and seven-letter words, with every letter of each word serving as the fixation point for an equal number of trials. For each model, we used a single set of eccentricity and edge-bonus parameters parameters to cover all trials, giving us four parameters in the symmetric parametric case (one maximum acuity parameter, one standard deviation, and two edge bonuses), five in the asymmetric parametric case (two standard deviations instead of one), nine in the symmetric nonparametric case (seven eccentricity parameters and two edge bonuses), and fifteen in the asymmetric nonparametric case (thirteen eccentricity parameters instead of seven). These models are nested in the classical statistical sense as follows:

$$[-\text{PAR},-\text{SYM}] \prec \{[-\text{PAR},+\text{SYM}],[+\text{PAR},-\text{SYM}]\} \prec [+\text{PAR},+\text{SYM}]$$

Since we use maximum likelihood estimation with far more observations than parameters, we can use likelihood-ratio tests for pairwise comparisons of all models except between [$-$PAR,+SYM] and [+PAR,$-$SYM]. These tests indicate that asymmetric models explain participant response behavior far better than symmetric models in both parametric ($\chi^2(1) = 250, p \ll 0.001$) and non-parametric cases ($\chi^2(6) = 345.7800, p \ll 0.001$). Among asymmetric model variants, the nonparametric model explains participant response behavior significantly better than the Gaussian model ($\chi^2(10) = 564.7, p \ll 0.001$).

For the asymmetric nonparametric model, we estimated standard deviations for our parameter estimates using 100 bootstrap replicates. Figure 2 graphs the value of $\log l$ as a function of eccentricity, together with edge-bonus parameter

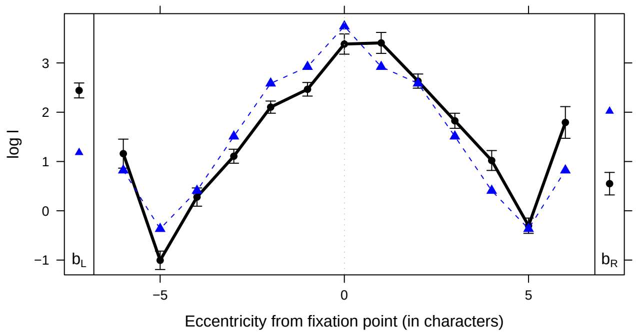



Figure 2: Symmetric (dashed blue lines) and asymmetric (solid black lines) non-parametric model parameter estimates. Error bars on the asymmetric model parameter estimates are standard deviations estimated from bootstrap replicates. $b_L$ and $b_R$ are "edge bonus" parameters added to the appropriate eccentricity parameters to obtain $\log l$ for the leftmost and rightmost letters in a word.

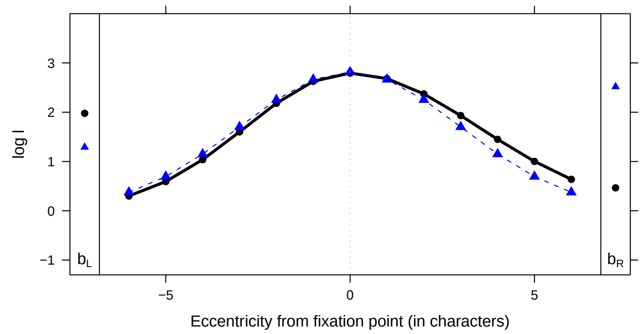



Figure 3: Symmetric (dashed blue lines) and asymmetric (solid black lines) parametric model results, assuming that eccentricity is piecewise Gaussian centered around the fixation point. $b_L$ and $b_R$ are "edge bonus" parameters added to the appropriate eccentricity parameters to obtain $\log l$ for the leftmost and rightmost letters in a word.

estimates (and error bars to indicate bootstrapped standard deviations for the asymmetric variant), for the two varieties of the nonparametric model. Figure 3 graphs these quantities for the two varieties of the parametric model. In both figures, eccentricity falls off to the left more rapidly in the asymmetric model than in the symmetric model, and to the right more rapidly in the symmetric model than in the asymmetric model.[2] This consistent pattern suggests that the data of Stevens & Grainger (2003) provide evidence for an asymmetry in the perceptual acuity curve in visual recognition of isolated words even when lexical constraint is taken explicitly into account.

## Conclusion

The results of our modeling studies provide additional evidence for the idea that an asymmetric visual acuity curve contributes to the OVP effect documented in many studies on isolated word recognition (O'Regan et al., 1984; Vitu, O'Regan & Mittau, 1990; O'Regan & Jacobs, 1992; Stevens & Grainger, 2003). Even while explicitly accounting for the role of lexical constraint, we consistently found that the models which best accounted for the distribution of response accuracies found in Experiment 2 of Stevens & Grainger (2003) had an asymmetric perceptual acuity curve in which acuity dropped off more slowly as a function of visual eccentricity to the right than to the left. Although Stevens & Grainger (2003) also presented results of another experiment using the letter-within-mask identification task which called into question the generalizability of the findings of Nazir et al. (1991, 1992) regarding asymmetry of the perceptual acuity curve, the fact that an asymmetric perceptual acuity curve is required to provide the best account for their word-recognition data by an ideal observer with the knowledge of lexical statistics of the language calls into question the strong position staked out by Clark & O'Regan (1999) that the left-of-center position of the OVP may derive purely from lexical constraint. In our view, the most plausible theoretical position reconciling our modeling findings with empirical findings from the reading literature on the asymmetric perceptual span (Rayner et al., 1980) and with those of Stevens & Grainger (2003) on character-within-mask recognition is that isolated word recognition does indeed involve an asymmetry of perceptual acuity, but that it is a parasitic byproduct of cognitive adaptation to the naturalistic task of reading. The character-within-mask recognition task may be sufficiently unlike naturalistic reading that it does not trigger this asymmetric perceptual span. It is worth noting that the experiments of Nazir et al. (1991) and Nazir et al. (1992) always involved fixation on the leftmost or rightmost character of the word, and used a letter (instead of the hash mark # used by Stevens & Grainger) as the mask character. It is possible that the lower overall visual acuity due to fixation at the word's right or left edge together with the letter mask may have induced the language and perceptual systems to categorize input in this experiment more like that obtained in natural reading, inducing asymmetric perceptual acuity.

Although we believe our modeling approach represents an important step forward in resolving these issues, it is important to emphasize that several simplifying assumptions we introduced which led from the general ideal-observer formulation to the specific, highly tractable model of Equation (4) have the potential to significantly affect our modeling results and deserve more careful consideration in the future. The first

[2]In the nonparametric model, the most extreme eccentricities have oddly-behaving parameter estimates that indicate possible problems with model specification, perhaps because the edge bonus parameters are so often implicated in model predictions for these extreme eccentricities. We expect that fitting the nonparametric model to behavioral data involving presentation of words of a larger variety of lengths would be likely to reduce or eliminate this problem.

of these simplifying assumptions was equal confusability of letter pairs. Of course, it is well known that some letter pairs are more confusable than others (e.g., *o* is much more confusable with *e* than with *l* for native English speakers; Geyer, 1977). It is possible that the assumption of equal confusability could interact with the lexical statistics of French and the items chosen by Stevens & Grainger (2003) to create a confound in the explanation of our modeling results—for example, such a confound might arise if the letters near the end of their items were more highly confusable than the letters near the beginning of their items. This possibility can be explored in future work by relaxing the assumption of equal confusability and using established letter confusion matrices to scale our model parameters, or even to allow our model to learn confusability parameters directly from word-recognition behavioral data.

The second simplifying assumption deserving discussion is that participants's behavioral responses arose from probability matching. In other cognitive domains, this assumption seems to have reasonable theoretical and empirical support (Vulkan, 2000; Mozer, Pashler & Homaei, 2008; Vul & Pashler, 2008; Vul, Goodman, Griffiths & Tenenbaum, 2009). That being said, it is entirely possible that participants' responses may reflect maximization or some other similar decision process, and that most inter-trial response variability derives from the variation inherent in noisy perception. The consequences of this possibility may be explored in future work within our framework by explicit simulation of inter-trial noise instead of marginalizing over perceptual input as we have done here.

The final simplifying assumption is that of minimal covariance between the probability of noisy perceptual samples given the word under consideration and the marginal probability of those perceptual samples. As discussed earlier, this assumption is clearly wrong insofar as words in any given language tend to look more like each other than like non-words; but the sheer number of words in the lexicon, combined with the considerable variability that does exist among wordforms, implies that this covariance should in general be rather small. It is also not obvious to us how this simplifying assumption might introduce a confound to our specific result of an asymmetric perceptual acuity curve in isolated word recognition. Nevertheless, there are two ways that this simplifying assumption could be relaxed in future work. First, explicit simulation of inter-trial noise would permit us to quantify the discrepancy between the simplified results we report here and the results which would obtain under the model more generally. Alternatively, we might try to quantify the covariance between $P(\mathbf{d}|w,k)$ and $P(\mathbf{d}|k)$ through explicit simulation, and then use these covariance estimates to adjust our expectation-based model directly. This latter alternative might have the added benefit of giving us more direct insight into the full range of top-down effects that are present in word recognition. Intuitively, this covariance should be larger for more "prototypically word-like" words, which should *decrease* the expected posterior belief for such words relative to less word-like words, a sort of second-order neighborhood-density effect. Further elucidation of all these issues awaits future research.

## Acknowledgments

We are grateful to Michaël Stevens and Jonathan Grainger for graciously sharing their experimental data with us; and to Rebecca Colavin for her expertise in French. This research was partially supported by NIH Training Grant T32-DC000041 to the Center for Research in Language at UCSD to NJS, by a Taiwan Graduate Fellowship of government sponsorship for overseas study to WHC, and by NSF grant 0953870 to RL.

## References

Anderson, J. R. (1990). *The Adaptive Character of Human Thought*. Lawrence Erlbaum.

Clark, J. J. & O'Regan, J. K. (1999). Word ambiguity and the optimal viewing position in reading. *Vision Research*, *39*, 843–857.

del Prado Martín, F. M. (2008). A fully analytical model of the visual lexical decision task. In *Proceedings of the Cognitive Science Society*.

Engel, G. R., Dougherty, W. C., & Jones, G. B. (1973). Correlation and letter recognition. *Canadian Journal of Psychology*, *27*(3), 317–326.

Geyer, L. H. (1977). Recognition and confusion of the lowercase alphabet. *Perception & Psychophysics*, *22*, 487–490.

Marr, D. (1982). *Vision*. San Francisco: Freeman.

Mozer, M. C., Pashler, H., & Homaei, H. (2008). Optimal predictions in everyday cognition: The wisdom of individuals or crowds? *Cognitive Science*, *32*(7), 1133–1147.

Nazir, T. A., Heller, D., & Sussman, C. (1992). Letter visibility and word recognition: The optimal viewing position in printed words. *Perception & Psychophysics*, *52*(3), 315–328.

Nazir, T. A., O'Regan, J. K., & Jacobs, A. M. (1991). On words and their letters. *Bulletin of the Psychonomics Society*, *29*(2), 171–174.

Norris, D. (2006). The Bayesian Reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Review*, *113*(2), 327–357.

Norris, D. (2009). Putting it all together: A unified account of word recognition and reaction-time distributions. *Psychological Review*, *116*(1), 207–219.

O'Regan, J. K. & Jacobs, A. M. (1992). Optimal viewing position effect in word recognition: A challenge to current theory. *Journal of Experimental Psychology: Human Perception & Performance*, *18*(1), 185–197.

O'Regan, J. K., Lévy-Schoen, A., Pynte, J., & Brugaillére, B. (1984). Convenient fixation location within isolated words of different length and structure. *Journal of Experimental Psychology: Human Perception & Performance*, *10*, 250–257.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422.

Rayner, K., Well, A. D., & Pollatsek, A. (1980). Asymmetry of the effective visual field in reading. *Perception & Psychophysics*, *27*(537–544).

Rice, J. A. (1995). *Mathematical Statistics and Data Analysis* (2 ed.). Duxbury Press.

Stevens, M. & Grainger, J. (2003). Letter visibility and the viewing position effect in visual word recognition. *Perception & Psychophysics*, *65*(1), 133–151.

Vitu, F., O'Regan, J. K., & Mittau, M. (1990). Optimal landing position in reading isolated words and continuous text. *pp*, *47*(6), 583–600.

Vul, E., Goodman, N. D., Griffiths, T. L., & Tenenbaum, J. B. (2009). One and done? Optimal decisions from very few samples. In *Proceedings of the Cognitive Science Society Conference*.

Vul, E. & Pashler, H. (2008). Measuring the crowd within: Probabilistic representations within individuals. *Psychological Science*, *19*(7), 645–647.

Vulkan, N. (2000). An economist's perspective on probability matching. *Journal of Economic Surveys*, *14*(1), 101–118.