Computationally Rational Saccadic Control: An Explanation of Spillover Effects Based on Sampling from Noisy Perception and Memory

Michael Shvartsman
Department of Psychology
University of Michigan
mshvarts@umich.edu

Richard L. Lewis
Department of Psychology
University of Michigan
rickl@umich.edu

Satinder Singh
Computer Science & Eng.
University of Michigan
baveja@umich.edu

Abstract

Eye-movements in reading exhibit frequency spillover effects: fixation durations on a word are affected by the frequency of the previous word. We explore the idea that this effect may be an emergent property of a computationally rational eye-movement strategy that is navigating a tradeoff between processing immediate perceptual input, and continued processing of past input based on memory. We present an adaptive eye-movement control model with a minimal capacity for such processing, based on a composition of thresholded sequential samplers that integrate information from noisy perception and noisy memory. The model is applied to the *List Lexical Decision Task* and shown to yield frequency spillover—a robust property of human eye-movements in this task, even with parafoveal masking. We show that spillover in the model emerges in approximately optimal control policies that sometimes process memory rather than perception. We compare this model with one that is able to give priority to perception over memory, and show that the perception-priority policies in such a model do not perform as well in a range of plausible noise settings. We explain how the frequency spillover arises from a counter-intuitive but fundamental property of sequenced thresholded samplers.

1 Introduction and overview

Our interest is in understanding how eye-movements are controlled in service of linguistic tasks involving reading—more specifically, how saccadic decisions are conditioned on the moment-by-moment state of incremental perceptual and cognitive processing. The phenomena we are concerned with here are *spillover effects*, where fixation durations on a word are affected by linguistic properties of the prior word or words. The specific idea we explore is that spillover effects may be emergent properties of a computationally rational control strategy that is navigating a tradeoff between processing immediate perceptual input, and continued processing of past input based on a memory of recent stimuli.

The paper is organized as follows. We first review evidence that eye-movement control in reading is strategically adaptive, and describe our theoretical approach. We then review evidence from gaze-contingent eye-tracking paradigms—some existing and some new—that suggests that frequency spillover is not driven exclusively by parafoveal preview of upcoming words. We take this as evidence that frequency spillover may be driven in part by processing of words that continues after the eyes have moved away. We then extend an existing adaptive control model of eye-movements with a minimal capacity for such continued processing, by allowing it to process a memory of past input. The model is based on a simple composition of thresholded sequential samplers that integrate information from noisy perception and noisy memory. Threshold parameters define the control policy and their values determine how processing resources are allocated to perception and memory. We provide a computational rationality analysis of the model's policy space: First, we show that frequency spillover emerges in top-performing policies, where performance is evaluated on the same task and payoff given to human participants. Second, we show that a model capable of spillover does no worse than an otherwise identical model that can eliminate spillover by always attending to perception when it can, and that the spillover-capable policies in such a model do no worse than spillover-incapable ones across the speed-accuracy tradeoff curve, and in fact do better in some portions of the noise parameter space. Finally, we trace the origin of the effect to a counter-intuitive but fundamental property of the dynamics of sequenced thresholded samplers.

2 Adaptive control of eye-movements: Evidence and theoretical approach

A growing body of evidence suggests that eye-movements in reading are strategic adaptations that manifest at the level of individual fixations. For example, Rayner and Fischer (1996) showed that when participants are searching for a particular word in a text rather than reading for full comprehension, saccade durations are shortened and the magnitude of frequency effects is reduced. Wotschack (2009) showed that readers assigned the task of proofreading read more slowly and performed more second-pass reading with fewer skips than in a control reading-for-comprehension task.

People also adapt reading behavior to within-task manipulations of difficulty and payoff. Wotschack (2009) showed that people change their reading behavior in response to manipulations of the difficulty of comprehension questions. Lewis et al. (2013) showed that people adapt their eye movements in response to changes in quantitative task payoffs. Payoffs emphasizing speed at the expense of accuracy result in shorter fixation durations and lower accuracies.

We seek to develop a model that can explain such variation in eye-movement behavior as a rational adaptation to the task (including utility) and the internal oculomotor and cognitive architecture (Lewis et al., 2013). Such a model would permit a *computational rationality* analysis (Lewis et al., to appear) because the problem of rational behavior is defined in part by the bounded mechanisms of the posited computational architecture.

We constrain our architectural assumptions by building on existing theories of oculomotor architecture, such as E-Z Reader (Reichle et al., 2009). But we enrich these architectures with explicit assumptions about the policy space of saccadic control, and with assumptions about the processing of noisy perception and memory. This enriched architecture is then embedded in a minimal cognitive system that is capable of performing a complete experimental task. The complete model affords computational rationality analyses because it can be used to derive the implications of saccadic control policies for task performance.

3 The nature of spillover effects

Our aim in this section is to establish a link between spillover and the continued processing of past input based on memory. Consider a pair of words in sequence: word$_{n-1}$ and word$_n$. There are three natural explanations for how the frequency of word$_{n-1}$ could affect the duration of fixations on word$_n$. (1) During fixation of word$_n$, perceptual information from word$_{n-1}$ is available in the parafovea and continues to be processed.

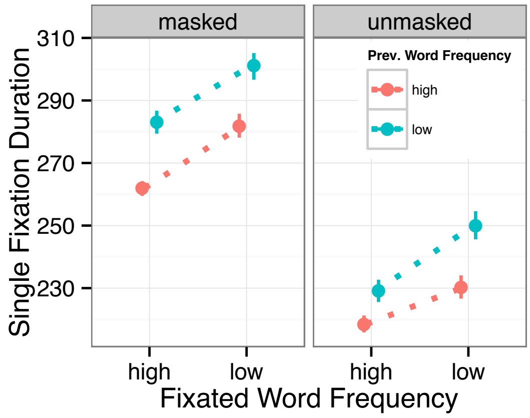

Figure 1: *Frequency spillover in the List Lexical Decision Task.* Single fixation durations (fixations when the word was fixated only once) on words as a function of the fixated and previous word's frequency. Frequencies are binned by a median split; error bars are bootstrapped standard errors.

We call this the *parafoveal review* explanation. (2) During fixation on word$_{n-1}$, perceptual information from word$_n$ is available in the parafovea; the frequency of word$_{n-1}$ affects the degree to which this information is processed, and this in turns affects the subsequent fixation duration on word$_n$. We call this the *parafoveal preview* explanation. (3) During fixation of word$_n$, processing of word$_{n-1}$ continues based on some memory of the perception of word$_{n-1}$, and this processing is affected by the frequency of word$_{n-1}$. We call this the *memory* explanation.

It is unlikely that spillover is driven by parafoveal review because the effective visual field in reading does not extend to the left of the current word (Rayner et al., 1980).

The standard paradigm for investigating the relationship between spillover effects and parafoveal preview is some form of parafoveal masking (Rayner, 1975): a nonveridical preview of word$_n$ is shown until the eye crosses an invisible boundary just before word$_n$, at which point word$_n$ is shown. When participants are not informed of the manipulation or do not notice it, they do not exhibit frequency spillover (Henderson and Ferreira, 1990; Kennison and Clifton, 1995; White et al., 2005). However, when participants are aware of preview being unavailable or not veridical, the spillover frequency effect remains (White et al., 2005; Schroyens et al., 1999). These results suggest that parafoveal preview (or review) cannot be the only explanation of spillover and therefore the

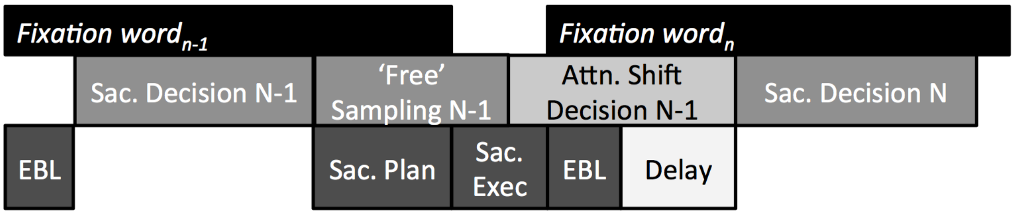

Figure 2: *Example dynamics of a decision to saccade from word$_{n-1}$ to word$_n$.* The memory-driven attention shift decision can delay the start of perceptual sampling on the next word, potentially creating spillover. A detailed description of the dynamics depicted in this figure is in §4.

memory explanation warrants consideration. We now summarize unpublished data consistent with these findings in a simple linguistic task that we also use to test the new model reported below.

Spillover in the List Lexical Decision Task (LLDT). We use the List Lexical Decision Task (LLDT) (Lewis et al., 2013), an extension of a task introduced by Meyer and Schvaneveldt (1971). In the LLDT participants must determine whether a list of six strings contains all words, or contains a single nonword. All strings are four characters in length and separated by six character spaces. The task was designed to require sequential eye-movements and contact with the mental lexicon (but not higher-level linguistic processing), to minimize parafoveal processing (via the wide spacing), and to yield a high proportion of single-fixation durations (via short strings).

Two versions of the task were performed by separate participant groups. In the masked condition, we used a gaze-contingent moving window paradigm wherein all strings but the fixated string were replaced with hashmarks (####). In the unmasked condition, all six strings remained visible.

Figure 1 shows the effects of word frequency on single fixation durations. The main result of current interest is that frequency spillover is evident in both conditions, despite the wide spacing in the unmasked condition, and the complete denial of parafoveal preview in the masked condition.

The work reviewed above and our new data are consistent with an account of spillover in which both parafoveal preview (if available) and memory-based processing are operative. Our concern here is with the latter: understanding how a noisy memory of recently seen stimuli might be incorporated into an adaptive oculomotor architecture, and exploring whether rational exploitation of that memory might lead to spillover.

4 A model of saccadic control with noisy memory for recent perception

Our new model extends the one presented in Lewis et al. (2013) to include a noisy memory that buffers perceptual input. We develop it in the context of the LLDT, but its essential elements are not tied to this task. It is most easily understood by first considering the dynamics of a single decision to saccade from one word to the next, as presented in Figure 2. After describing these dynamics we summarize the model's key assumptions and associated mathematical specification.

The dynamics of a decision to saccade from word$_{n-1}$ to word$_n$. The eye first fixates word$_{n-1}$. Some time passes before information from the retina becomes available for perceptual processing (the eye-brain lag, EBL in Figure 2). A sequence of noisy perceptual samples then arrive and are integrated via an incremental and noisy Bayesian update of a probability distribution over lexical hypotheses in a manner described below. The perceptual samples are also buffered by storing them in a memory that contains samples from only one word. When the probability of one of the hypotheses reaches the *saccade threshold*, saccade planning is initiated. Perceptual sampling (marked as *free sampling* in Figure 2 because its length is not under adaptive control) continues in parallel with saccade planning until the fixation ends, and then for another EBL amount longer (these are samples received at the retina during the fixation and only now arriving at the lexical processor).

The model then switches to sampling from its memory, continuing to update the distribution over lexical hypotheses until one of the hypotheses reaches an *attention shift threshold*. If this threshold had already been reached during the earlier perceptual sampling stages, attention shifts instantly. Otherwise attention remains on word$_{n-1}$ even if the eye has saccaded to word$_n$, and the eye-brain lag on word$_n$ is completed. Perceptual samples from word$_n$ will not be processed until attention is shifted away from the memory-based processing of word$_{n-1}$. Thus the memory processing on word$_{n-1}$ may delay processing of perceptual samples from word$_n$; perceptual samples arriving during this time are buffered in the memory. In this way the posterior update is a limited computational resource and its relative allocation to perception or memory is determined by the saccade and attention shift thresholds. To the extent that the time to reach the attention shift threshold is sensitive to the frequency of word$_{n-1}$, the model may exhibit a spillover frequency effect.

Lexical processing as rise-to-threshold decisionmaking. The decisions to plan a saccade, shift attention, and make a motor response are realized as Multi-hypothesis Sequential Probability Ratio Tests (Baum and Veeravalli, 1994; Dragalin et al., 2000). At each timestep, the model performs a Bayes update based on a noisy sample drawn from perception or memory, with the posterior at each timestep becoming the prior for the next timestep. Our choice of word representation follows Norris (2006) in representing a letter as a unit-basis vector encoding and a word as a concatenation of such vectors.

To generate a perceptual sample, mean-zero Gaussian *perception noise* with standard deviation (SD) σ_p is added to each component of the word representation vector. Each perceptual sample is also stored in a memory buffer, and memory samples are generated by uniformly drawing a stored sample from memory (with replacement), and adding an additional mean-zero Gaussian *memory noise* with SD σ_m to each position. Before each Bayesian update, whether using a sample from perception or memory, mean-zero Gaussian *update noise* with SD σ_u is added to each component of the word representation vector. Thus a Bayes update from a perceptual sample includes two noise terms, while a Bayes update from a memory sample includes three noise terms. All noises are drawn independently. The three SD's, σ_p, σ_m and σ_u, are free parameters in the model, and we explore their implications below.

The model uses the update specified in the appendix in Lewis et al. (2013) except for the noise generation specified above and the consequent change in the likelihood computation. The lexical hypotheses are updated as follows:

$$Pr_{new}(S^k|s^k,\mathcal{T}) = \frac{Pr(s^k|S^k,\mathcal{T})Pr_{old}(S^k,\mathcal{T})}{\sum_S Pr(s^k|S^k,\mathcal{T})Pr_{old}(S^k,\mathcal{T})} \tag{1}$$

where s^k is a sample generated as above from the letterstring (word or nonword) in the current position k, S^k is the hypothesis that the string at position k is S, and $\mathcal{T}$ is a multinomial distribution reflecting the current belief of (a) whether this is an all-words trial and (b) otherwise, where the nonword is located. The eye movement planning and attention shift decisions are conditioned on the distribution of probabilities $Pr(S^k)$ for all strings in the current position. When the maximum of these probabilities crosses a saccade planning threshold θ_s, saccade planning begins. When the maximum crosses the attention shift threshold θ_a, attention shifts to the next word[1]. Each sample takes 10ms, a fixed discretization parameter.

The likelihood of drawing perceptual or memory sample s for a string S is computed from the unit-basis word representation as follows:

$$Pr(s|S) = \prod_i f(s_i;\mu_i,\sigma) \tag{2}$$

where i indexes the unit-basis vector representation of sample s and some true letterstring S (and so μ_i is either 0 or 1), σ is the sampling noise (dependent on whether the samples are memory or perceptual samples as specified below), and $f(x;\mu,\sigma)$ is the probability density function of the normal distribution with mean μ and standard deviation σ.

We simplify the likelihood computation for memory samples by treating the perception and memory samples as independent. For present purposes this assumption may be treated as a bound on the architecture. The σ in Equation 2 is $\sqrt{(\sigma_p^2+\sigma_u^2)}$ for perceptual samples and $\sqrt{(\sigma_p^2+\sigma_m^2+\sigma_u^2)}$ for memory samples. At each sample the string-level probabilities in each position are aggregated to the multinomial trial-level decision variable $\mathcal{T}$ as described above. Given $\mathcal{T}$ the model computes the probability of a word trial $Pr(\mathcal{W})$ or nonword trial $Pr(\mathcal{N}) = 1 - Pr(\mathcal{W})$. When either of these probabilities exceeds the motor response threshold θ_r, motor response planning commences.

[1]Because there is a fixed set of memory samples available, the attention shift decision is not guaranteed to converge, unlike the saccade threshold. It nearly always converges, but we use a 30-sample deadline to prevent infinite sequences.

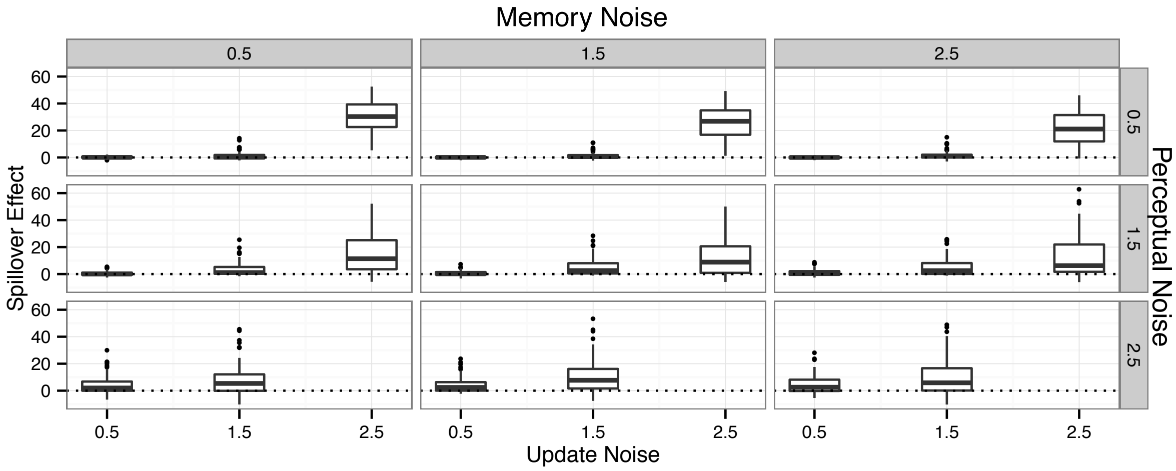

Figure 3: *Spillover effects generated by the top 5% of policies across different settings of memory, perception, and update noise*. On each distinct machine defined by a combination of noise settings, policies (settings of $\theta_s, \theta_m, \theta_r$) were evaluated by the same task payoff given to human participants in the experiment described in §3. Boxplots show spillover effects of the top-performing 5% of policies. Spillover effects are the difference in mean single fixation durations on word$_n$ when word$_{n-1}$ is low frequency and when word$_{n-1}$ is high frequency (low/high determined by median split). The highest noise settings in the bottom row are not shown because performance was near-chance even for the best policies.

The prior probability of an all-words trial is 0.5, so the prior probability of a word in each position k is $1 - \frac{0.5}{6}$. Therefore, we set the prior probabilities of words in each position to corpus frequency counts (Kučera and Francis, 1967), normalized to sum to this value, $1 - \frac{0.5}{6}$. Nonword probabilities are uniformly distributed over the remainder, $\frac{0.5}{6}$.

Oculomotor and Manual Architecture. The remainder of the architectural parameters are stage durations that are simulated as gamma deviates with means based on previous work or independently estimated from data. The key parameters for present purposes are the 50ms mean eye-brain lag and 125ms saccade planning time, following Reichle et al. (2009), and the 40ms mean saccade execution time, based on estimates from our own human participants. The standard deviation of each distribution is 0.3 times the mean. We transform the means and standard deviations into scale and shape parameters for a Gamma distribution and then draw duration values from these Gammas independently for every word and trial.

5 A computational rationality analysis

We explore whether spillover effects might be a signature of computationally rational behavior in two ways. First, we evaluate a space of policies (parameterized by $\theta_s, \theta_m, \theta_r$) against the task payoff given to our human participants, and show that top-performing policies yield frequency spillover consistent with human data, and poor-performing policies do not. Second, we extend the model's policy space to allow it to prioritize perception over memory samples when both are available (eliminating spillover in those policies), and show that the spillover portions of the policy space perform better than non-spillover ones under any imposed speed-accuracy tradeoff in plausible noise settings, and never perform worse.

In computational rationality analyses, we distinguish between policy parameters, fixed architecture parameters, and free architecture parameters. Policy parameters are determined by selecting those policies that maximize a given task payoff, given the hypothesized architectural bounds. Fixed architecture parameters are based on previous empirical or theoretical work. Free architecture parameters can be fit to data or explored to show the range of predictions with which the model is compatible. We focus here on the latter, showing not only that the model is compatible with human data, but that it is incompatible with results significantly different from the human data.

Our first evaluation of the model asks the question of whether we see spillover effects emerging in approximately optimal policies under our assumptions about mechanism and task. We evaluated our model in the LLDT, under the balanced payoff presented in Lewis et al. (2013), the same

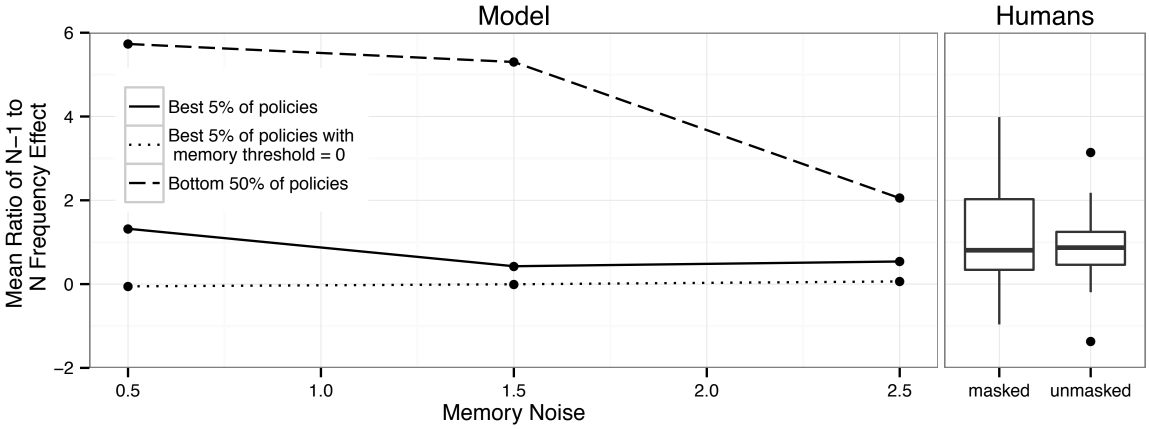

Figure 4: *Normalized spillover effect in model (vs. memory noise) and human participants.* We define normalized spillover as the ratio of the spillover (word$_{n-1}$) frequency effect size to the foveal (word$_n$) frequency effect size; this normalizes against scale differences between high and low noise architectures. *Left*: Mean normalized spillover effect at different memory noises for best performing 5% of policies with and without memory sampling, and worst 50% performing policies. *Right*: Mean human spillover effect sizes in masked and unmasked versions of LLDT.

payoff given to our participants in the unpublished masking experiment described above. We explored a discretized policy space as follows: we let θ_s range between 0.199 and 0.999 in steps of 0.05; θ_m between 0.19999 and 0.99999 in steps of 0.05, and also include $\theta_m = 0$ which prevents memory sampling; and θ_r between 0.599 and 0.999 in steps of 0.1. We explored all 1530 permutations.

Figure 3 shows the distribution of spillover effect sizes in the top 5% of policies (evaluated by task payoff, not fit to human data), for a range of noise parameter settings (at higher noise settings, even the best policies are close to chance performance). The top 5% of policies average 7.78 points per trial across the noise and policy range, and the bottom 50% average 1.32 points. The figure shows that top-performing policies show little to no spillover when update noise is low, positive but small spillover effects when update noise is moderate, and sizable positive spillover effects when update noise is relatively high. These results are consistent with spillover as a rational adaptation to belief update noise.

Figure 4 (left panel) shows normalized spillover effects (the ratio of the word$_{n-1}$ frequency effect to the word$_n$ frequency effect) for the best policies, the bottom 50% of policies, and the best policies constrained with a memory threshold of zero ($\theta_m = 0$). When $\theta_m = 0$, the spillover effect is zero as expected. The top performing policies in the unconstrained space generate nonzero spillover effects that are consistent with the human data, but the poor performing policies do not (Figure 4, right panel). We know that the top performing policies exploit memory because they do yield nonzero spillover effects, and the values of θ_m are nonzero for these policies.

Our second evaluation asks whether a model that is constrained to always give priority to processing perceptual samples over memory samples will perform better than the present model, which has the flexibility to give priority to memory over perception. To explore this, we added a single binary policy parameter, the *perceptual priority* bit. If this bit is set, then the model has the choice between memory sampling from word$_{n-1}$ and perceptual sampling from word$_n$, it always chooses the latter. Such an option is not available in the previous model—there is no setting of the saccade and memory thresholds that will always use memory samples when only they are available, but also never choose to use memory samples when perceptual samples can be used. With the perceptual priority bit set, the model is capable of exploiting the least noisy samples available to it, but is incapable of exhibiting spillover effects.

Figure 5 shows speed-accuracy tradeoffs for the model, with the perceptual-priority bit not set (spillover-capable) and set (spillover-incapable), in three representative noise settings. Individual points are policies and the lines mark the best accuracy available at a particular reaction time for the two classes of policies; i.e. these lines represent the best speed-accuracy tradeoff possible for

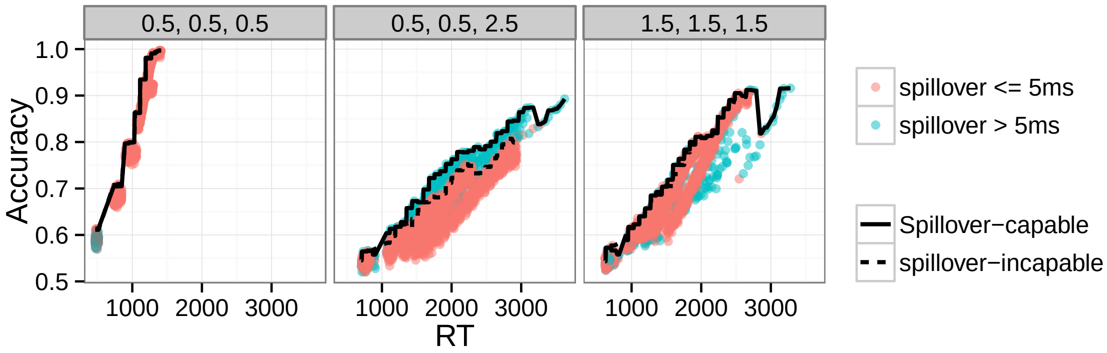

Figure 5: *Speed-accuracy tradeoff curves for some representative noise settings.* Each individual point corresponds to one policy (i.e. setting of the three decision thresholds). Plotted are mean trial RT and accuracy (computed from 5000 simulated trials), color-coded by whether the policies yielded spillover frequency effects. Lines mark the best speed-accuracy tradeoff available to spillover-capable and incapable policies. Each plot is labeled at the top with the noise setting *(perceptual, memory, update)*.

both spillover-capable and -incapable policies. In the left plot of the figure, noise is low enough overall such that responses are very fast and spillover-capable policies do no worse and no better than spillover-incapable policies. In the middle plot, update noise is higher, and the optimal speed-accuracy tradeoff is better for the model that can yield spillover, consistent with the exploitation of memory sampling to mitigate update noise. In the right plot, perception and memory noise are high enough that it is not useful to sample from memory at the expense of perception. All the noise settings we explored (see Figure 3 for the range) yield one of these three patterns, or the uninteresting case of near-chance performance. In no setting does the spillover-capable model perform worse than the spillover-incapable one. The noise settings cover a range from implausibly-high accuracy to chance performance, and so we conclude that spillover-capable policies dominate, in that they do no worse, and occasionally do better, than those constrained to give priority to perception over memory.

6 Why spillover arises from sequenced thresholded samplers

We have demonstrated through simulations that the model yields frequency spillover through a composed sequence of perception and memory sampling. We have not yet addressed the question of how or why this happens. Indeed, it is initially somewhat puzzling that an effect of priors (set by lexical frequency) would persist after the initial perceptual sampling threshold θ_p is passed, because this fixed threshold must be exceeded no matter the starting prior.

The crucial insight is that it is not always the case that the true word hypothesis reaches the threshold first; i.e., the decision to initiate saccade planning may be based on (partial) recognition of a different word than the true word. In such cases, at the start of memory sampling, the hypothesis for the true word is farther from the memory threshold θ_m than if the true word had been (partially) recognized. Incorrect decisions are more likely for low frequency words, so in expectation the memory-driven attention shift mechanism will start farther from its threshold for low-frequency words, and therefore take longer to reach threshold, delaying the following word more.

We constructed a minimal two-sampler example to clearly illustrate this phenomenon. The leftmost panel of Figure 6 illustrates the dynamics of such a trial. In this panel, the threshold is crossed for the incorrect hypothesis (green line) in the first sampler, triggering the start of the second sampler. The second sampler recovers from the mistake, allowing the correct (red) hypothesis to cross the threshold, but at the cost of additional time. The middle panel shows that incorrect (and thus eligible for recovery) trials are more frequent for low priors. The rightmost panel shows that the finishing time of the second sampler is proportional to the prior probability of the correct hypothesis for the first sampler. It is also inversely proportional to accuracy (middle plot), consistent with inaccurate trials driving the relationship between the first sampler prior and second sampler finishing times.

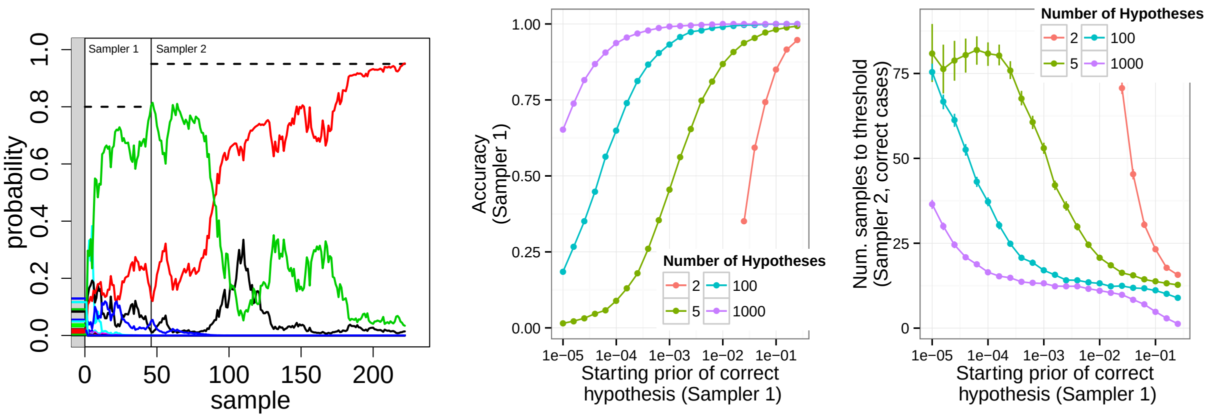

Figure 6: *A simple example illustrating how the prior for a thresholded sampler affects its final posterior, and therefore the prior for a subsequent coupled sampler, despite the fixed threshold. Left*: An example 'recovery' trial for 500 hypotheses (words). *Middle*: Accuracy for the first sampler as a function of the prior of the true hypothesis. *Right*: Second sampler finishing times as a function of to the true-hypothesis prior in the first sampler.

7 Discussion and Conclusion

We briefly highlight the key properties of the model that yield our result and how they may generalize beyond our particular implementation.

Post-perceptual processing. Although we adopted a second MSPRT sampler, spillover may arise from other processes with access to the posterior of the perceptual sampling, such that it can recover from perceptually misidentified words. In the present model we investigated the possibility that post-perceptual memory-based processing could be partially motivated by mitigating noise in the update process itself. But it is almost certainly the case that post-perceptual processing is required in the course of reading for independent reasons, and such processing could also yield spillover frequency effects in a way that the memory sampling process does. (A challenge for such an alternate process is that spillover effects persist in the LLDT in the absence of required higher level syntactic or semantic processing).

A tradeoff between processing perception and memory. The serial queuing model is a simple realization (inspired by EZ-Reader (Reichle et al., 1998)) of a limited resource that can be allocated to perceptual and memory processing, but an alternative parallel attention machine might recover the results, as long as it suffers from the same tradeoff that processing the previous word from memory will slow down processing of the fixated word.

Direct oculomotor control. In the present model saccade planning is triggered directly by the perceptual evidence accumulation process, and as such is not obviously compatible with autonomous saccade generation models like SWIFT (Engbert et al., 2005). It may be possible to layer SWIFT's time-delayed foveal inhibition over a sequential sampling process, but we note that spillover effects were part of the empirical motivation for such delayed control.

The present model and results open several avenues for future work. These include the interactions of memory-based or post-perceptual processing with models of saccade planning that include saccade targeting, re-targeting, and cancellation, as well as buttonpress behavior (e.g. in the self-paced moving window paradigm). The role that parafoveal preview plays in spillover effects can also be explored, including how the model (and thus human participants) might navigate the tradeoff between using parafoveal preview information (noisy due to eccentricity) and using memory of past input in the service of a reading task. Finally, it is possible to explore the spillover explanation in an architecture capable of higher-level sentence processing in service of different reading task goals.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant BCS-1152819 to RL and SB. We thank MindModeling@Home for invaluable computational resources.

References

C.W. Baum and V.V. Veeravalli. 1994. A sequential procedure for multihypothesis testing. *IEEE Transactions on Information Theory*, 40(6):1994–2007.

Vladimir P Dragalin, Alexander G Tartakovsky, Venugopal V Veeravalli, and Senior Member. 2000. Multihypothesis Sequential Probability Ratio Tests Part II : Accurate Asymptotic Expansions for the Expected Sample Size. *IEEE Transactions on Information Theory*, 46(4):1366–1383.

Ralf Engbert, Antje Nuthmann, Eike M. Richter, and Reinhold Kliegl. 2005. SWIFT: a dynamical model of saccade generation during reading. *Psychological review*, 112(4):777–813, October.

John M. Henderson and Fernanda Ferreira. 1990. Effects of foveal processing difficulty on the perceptual span in reading: Implications for attention and eye movement control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(3):417–429.

Sheila M. Kennison and Charles Clifton. 1995. Determinants of parafoveal preview benefit in high and low working memory capacity readers: implications for eye movement control. *Journal of experimental psychology. Learning, memory, and cognition*, 21(1):68–81, January.

Henry Kučera and W. Nelson Francis. 1967. *Computational analysis of present-day American English*. Brown University Press, Providence, RI.

Richard L. Lewis, Michael Shvartsman, and Satinder Singh. 2013. The Adaptive Nature of Eye Movements in Linguistic Tasks: How Payoff and Architecture Shape Speed-Accuracy Trade-Offs. *Topics in cognitive science*, pages 1–30, June.

Richard L. Lewis, Andrew Howes, and Satinder Singh. to appear. Computational rationality: Linking mechanism and behavior through utility maximization. *Topics in Cognitive Science*.

David E. Meyer and Roger Schvaneveldt. 1971. Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, 90:22–34.

Dennis Norris. 2006. The Bayesian reader: explaining word recognition as an optimal Bayesian decision process. *Psychological review*, 113(2):327–357, April.

Keith Rayner and Martin H. Fischer. 1996. Mindless reading revisited: eye movements during reading and scanning are different. *Perception & psychophysics*, 58(5):734–47, July.

Keith Rayner, Arnold D. Well, and Alexander Pollatsek. 1980. Asymmetry of the effective visual field in reading. *Perception & psychophysics*, 27(6):537–44, June.

Keith Rayner. 1975. The perceptual span and peripheral cues in reading. *Cognitive Psychology*, 7(1):65–81, January.

E D Reichle, a Pollatsek, D L Fisher, and K Rayner. 1998. Toward a model of eye movement control in reading. *Psychological review*, 105(1):125–57, January.

Erik D. Reichle, Tessa Warren, and Kerry McConnell. 2009. Using E-Z Reader to model the effects of higher level language processing on eye movements during reading. *Psychonomic bulletin & review*, 16(1):1–21, February.

Walter Schroyens, Françoise Vitu, Marc Brysbaert, and Géry D'Ydewalle. 1999. Eye movement control during reading: foveal load and parafoveal processing. *The Quarterly journal of experimental psychology*, 52(4):1021–46, November.

Sarah J. White, Keith Rayner, and Simon P. Liversedge. 2005. Eye movements and the modulation of parafoveal processing by foveal processing difficulty: A reexamination. *Psychonomic bulletin & review*, 12(5):891–6, October.

Christiane Wotschack. 2009. *Eye Movements in Reading Strategies*. Ph.D. thesis.