An incremental information-theoretic buffer supports sentence processing

Francis Mollica, Steven T. Piantadosi
{mollicaf | spiantado}@gmail.com
Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY 14627 USA

Abstract

People have the capability to process text three times faster than they would naturally read it, yet many current theories of sentence processing rely on natural reading times as a proxy for processing difficulty. How can people read material so quickly in spite of information processing limitations suggested by sentence processing theories? One possibility is that surprisal effects on reading time, the hallmark of processing difficulty under sentence processing theories, might arise from perceptual processing, implying no relation between surprisal and sentence processing difficulty. In this paper, we conducted a novel self-paced rapid serial visual presentation (RSVP) experiment, which controlled perceptual processes to probe for sentence processing related surprisal effects. We further tested how readers might compensate for information processing limits during RSVP. We find support for sentence processing related surprisal effects, the pattern of which is consistent with a First-In, First-Out (FIFO) buffer model.

Keywords: language processing; linguistic memory; RSVP

Introduction

One of the remarkable feats of language processing is the quick pace at which various forms of information are seamlessly integrated (e.g., Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Researchers have noticed that language can be processed much faster than how we naturally engage with it (e.g., Potter, 1984). For example, by presenting text faster than the human eye would naturally fixate, people are able to read a story $3-4$ times faster than natural. These observations have led to several apps training people how to read faster—although the efficacy of these apps has been contested (see Rayner, Schotter, Masson, Potter, & Treiman, 2016). Even though researchers have noted that information processing in reading is faster than people's natural reading pace, psycholinguistics has relied upon the assumption that natural reading times are a reflection of processing difficulty. Using reading time as an index of processing difficulty has provided the data behind several theories in sentence processing (e.g., Grodner & Gibson, 2005; Levy, 2008). Recently, this assumption has been formalized in terms of information theory: words that carry more information tend to increase reading times relative to words that contain less information, suggesting a fixed processing rate that is measured in bits of information per unit time (e.g., Hale, 2001; Levy, 2008). The amount of information conveyed by a word, colloquially referred to as surprisal, is equal to the negative log probability of the word given its context.

Given the importance of reading times as a measure of processing difficulty, it is surprising that people can read several times faster than they naturally would and it poses a very important question: Are the effects of surprisal observed in reading times reflective of linguistic information processing limitations, or do they arise from some alternate perceptual process? We conducted a novel rapid serial visual presentation (RSVP) experiment in which we controlled perceptual processing to test if surprisal effects reflect language processing. Our experiment was also designed to probe how readers might be compensating for linguistic information processing limits when faced with rapidly presented text. Specifically, we consider two hypotheses that would reconcile information processing limits with rapid presentation of text. First, language users might suspend incremental information processing to store information in a buffer, until it can be processed. Given the quick response times in RSVP experiments, this would suggest fast rates of information processing, giving us real reason to check our assumption that reading time is a valid measure of sentence processing difficulty. Second, language users might compensate for RSVP by utilizing an incremental First-In, First-Out (FIFO) memory buffer, where information is immediately copied into a buffer and processed out of the buffer at a fixed rate. Many researchers using the RSVP paradigm have proposed the use of a buffer (e.g., Mitchell, 1984); however, buffer models have never been formalized in terms of information processing. In the remainder of the introduction, we motivate the perceptual and linguistic sources of surprisal effects and our buffer models.

Perceptual Information Maintenance

When the reader manages visual presentation time, it is unclear if we see surprisal effects because readers maintain the visual input while they are syntactically processing the word or if we see surprisal effects for perceptual processing reasons. For example, both Bayesian models of word identification (Norris, 2006) and rational models of eye movements in reading (Bicknell & Levy, 2012) posit that readers maintain information until they reach a level of confidence in a word's identity, resulting in log effects of word predictability—i.e., numerically the same value as surprisal. This hypothesis is corroborated by work on individual differences in eye-tracked reading that demonstrate a relationship between word-identification ability and reading speed (Kuperman & Van Dyke, 2011). In our experiment, we control for a perceptual explanation of surprisal effects by using RSVP, which prohibits readers from influencing presentation time. Importantly, the same processes underlying natural reading are still thought to be at play in RSVP (Potter, 1984). Therefore, if surprisal effects were a result of readers maintaining the visual input until they were confident in the word's identity or had finished syntactically processing the word, we would expect that RSVP would eliminate all surprisal effects.

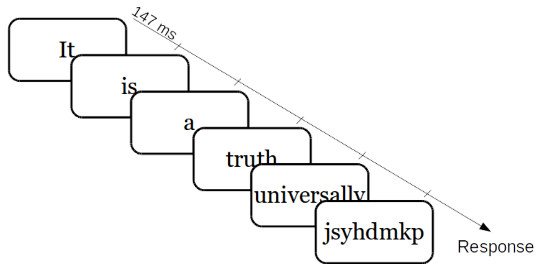

Figure 1: In 5-RSVP, each word is presented for 147 ms and immediately masked by the next word. The last word in each 5 word chunk is masked with random lowercase letters remaining until the participant presses a button to continue.

Linguistic Information Processing Limits

One of the foundational questions in sentence processing research is how do people construct the correct syntactic parse for a sentence. As mentioned above, the primary method for teasing apart sentence processing theories is making predictions about processing difficulty for the constituent words of a sentence. Information theory has provided one way to formalize the amount of processing required for each word in a sentence. Hale (2001) first noticed that word reading times are proportional to the total probability of all syntactic structures that are no longer possible after observing a word, which happens to be the word's surprisal. Levy (2008) showed that surprisal can be given an alternative interpretation in terms of parallel-resource allocation. As each new word is processed, the syntactic parses under consideration are re-ranked and resources are allocated accordingly. The processing difficulty of that word is reflected in the amount of re-ranking required as measured by Kullback-Leibler divergence, which simplifies to the word's surprisal. Either way, a word's surprisal is a measure of the amount of processing, and surprisal effects in reading suggest processing difficulty caused by information processing limits. In our experiment, we have controlled for a perceptual origin of surprisal effects so that if we still see surprisal effects, we can be more confident in interpreting them as stemming from language processing.

If surprisal effects reflect information processing limits, the contradiction between readers being capable of reading rapidly presented text but possessing hard information processing limits still remains. Our novel five word self-paced RSVP task presents text in five word RSVP chunks for each button press (see Figure 1), allowing us to consider two hypotheses that would reconcile these observations. First, readers might focus on buffering linguistic information during rapid presentation and delay processing until the perceptual barrage has ended. Under this account, we would expect a **small** uniform profile of surprisal weights for each of the five words presented together (see middle panel of Figure 2). In RSVP tasks, readers only require a one second break between sentences to perform as well as control groups in recall tasks (Potter, 1984), which suggests an information processing rate roughly 100 bits/second[1]. This rate is too fast to predict surprisal effects at naturalistic reading paces, which makes this an important hypothesis to rule out for expectation based sentence processing theories.

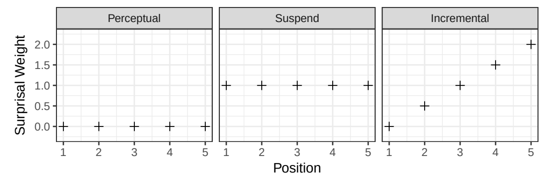

Figure 2: Predicted surprisal weights (beta values for surprisal) in the 5-RSVP task as a function of word position for our hypotheses.

Second, readers might incrementally process linguistic information, but perception and language processing might not be as tightly coupled as previously thought. Perception might quickly place information into a buffer, where language processing then removes that information at a slower rate, but still very quickly. This allows perception to move ahead while language processing is ongoing. In this FIFO buffer model, the expected profile of surprisal weights depends on the rate of information processing. Roughly, the model yields three different profiles of surprisal weights (see Figure 3). At slow rates of information processing, multiple words will be seen before the processing of the initial word has ended. As a result, the majority of processing time for all of the words is reflected in reading time post visual presentation, which gives rise to a **large** uniform profile of surprisal weights. At rates of information processing closer to the average amount of information conveyed in a word, sometimes a word is completely processed before the next word appears and sometimes a word's processing briefly carries over to when the next word appears. As a result, early words contribute less to post-presentation reading time than later words, and the profile of surprisal weights increases with a word's presentation position. At fast rates of information processing, each word is completely processed before the next word appears. Therefore, there should be no influence of surprisal on post-presentation reading time, corresponding to a uniform zero profile of surprisal weights. Our a-priori prediction for the FIFO buffer assumed a rate of processing near the average amount of information (see right panel of Figure 2).

It is important to note that the FIFO buffer model shows the traditional surprisal effect with the rate of information processing parameter. The profile of surprisal weights in the FIFO buffer model do not necessarily reflect processing difficulty as in the typical surprisal effect; rather, the surprisal weights reflect each word's contribution to the information remaining in the buffer post-presentation.

Experiment

Our plan of attack is to see if surprisal effects are present in a masked word-by-word self-paced reading (mSP) task and a novel 5-word chunk self-paced RSVP task (5-RSVP), which holds perceptual information constant but allows the reader unlimited time to process their input. As a result of the per-

[1]We assumed each word conveys on average 10 bits, which is true of our stimuli, and an average sentence length of 10 words.

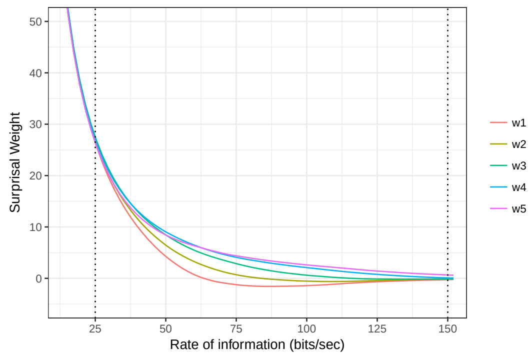

Figure 3: Predicted surprisal weights from an incremental FIFO buffer model for each word position (w1-w5) as a function of information processing rate. At slow rates, there is a large uniform surprisal effect. At very fast rates, there is no effect of surprisal. In the middle, surprisal weights increase as a function of their position in a chunk.

ceptual controls, we can interpret an effect of surprisal as reflecting linguistic information processing limits. Further, if there is an effect of surprisal, the profile of surprisal weights for each word in the 5-RSVP condition can distinguish between two hypotheses reconciling readers' ability to read rapidly presented text in spite of incremental processing limits (see Figure 2). Specifically, a uniform profile of surprisal weights would suggest a buffer model in which readers suspend incremental linguistic processing until after presentation ends. Whereas, a profile of surprisal weights that increases as a function of a word's position in the five word chunk would suggest a a FIFO buffer model (formalized below). In addition, participants completed a self-paced reading task (SP) to replicate the standard surprisal effect with our materials. Participants also read an additional story at their natural pace to provide a baseline accuracy and reading rate, which ensured our presentation rate was faster than our participants natural reading rate.

Participants. Sixty-four participants were recruited from the University of Rochester community. Participants were compensated for their participation, which lasted approximately 45 minutes. All participants were screened for normal or corrected-to-normal vision, and English as their native language.

Materials and Design[2]. Participants read four excerpts from articles published in 2008 issues of *The New Yorker*. To equate the length of the excerpts while preserving the naturalness of the reading material, articles were truncated to end with the paragraph containing the 1000^{th} word of the article. Word frequency and bigram probabilities with Laplace smoothing ($\alpha = 0.1$) were estimated using ZS (Smith, n.d.) to access Google N-grams (Michel et al., 2011). For each story, ten yes-no questions were constructed (5-yes, 5-no) to serve as a comprehension check. Four lists were constructed such that each list contained a story in every presentation condition (Baseline, SP, mSP and 5-RSVP) and across lists each story occurred with every condition. To ensure precise presentation timing, stimuli were presented using a monitor with a 144 Hz refresh rate. Responses and reading times were collected using the keyboard. Stimulus presentation and timing and response collection was controlled by Psychopy (Peirce, 2007).

Procedure. Sixteen participants were randomly assigned to each list. Participants were told that they would be reading and answering questions about four short stories. All participants began the experiment with the baseline reading condition. In the baseline condition, the full body of text for one story was displayed on the screen in 17.5 point Free Sans Bold font. Participants could scroll through the text using the up and down arrow keys. Participants were instructed to press the space bar when they had finished reading the text, which started the comprehension check.

Following the baseline reading task, participants were introduced to the mSP condition with three practice trials. On the first trial, each word was presented in the center of the screen for 245 ms and then immediately masked with a random string of 14 lower-case letters. Participants were instructed to press the space-bar to reveal the next word. For the next two practice trials, the presentation duration of each word was decreased to 196 ms and 175 ms, respectively. Before each trial and story, a fixation cross appeared for 147 ms to orient the participant to the center of the screen. After the practice trials, the order of the remaining conditions was randomized for each participant.

In the SP condition, text was presented similar to the practice trials with one exception: each word was not masked and remained visible to the participant until they pressed the space-bar to advance. In the mSP condition, text was presented similar to the practice trials, except the duration for each word was set at 147 ms. In the 5-RSVP condition, each button press triggered the presentation of the next five words in the text. Each word was flashed on the screen for 147 ms. The first four words were masked by the next word's appearance. The fifth word was masked by a 14 character lowercase letter string.

After each story, the comprehension check for that story began immediately. For each of the ten questions, participants registered their answer by pressing the Y (yes) or N (no) key on the keyboard, which prompted the next question. The order of the questions was randomized for each participant. After the comprehension check, participants were instructed to find their experimenter to advance to the next story. Participants were strongly encouraged to take breaks in between stories and often did. In line with our experience and intuitions, our participants reported that reading in these presentation conditions is taxing[3].

[2] All code and data are available at: `github.com/mollicaf/speedread`

[3] We set our presentation rate at 147 ms/word even though previous tasks have gone as fast as 84 ms/word because we were concerned about enervating participants after one story.

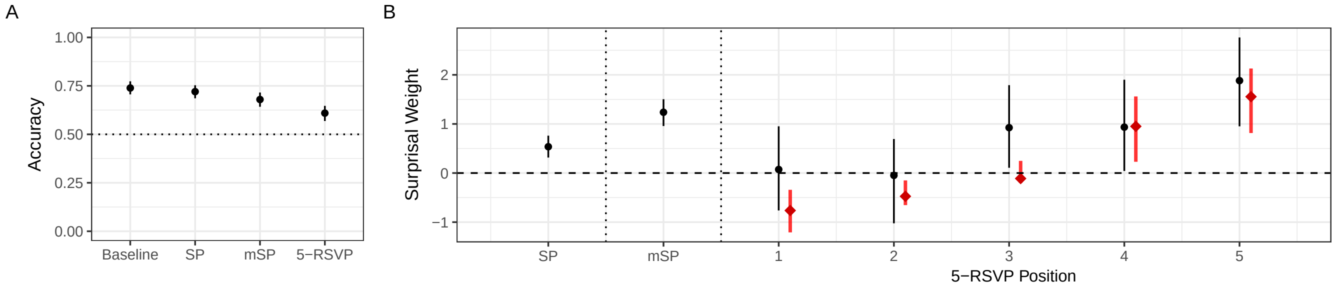

Figure 4: **A.** Average participant accuracy across presentation conditions. **B.** Surprisal weights from a linear mixed-effect regression of reading time in the SP, mSP and 5-RSVP conditions (as a function of each word's order in the 5-RSVP condition). Red diamonds represent the FIFO buffer model predictions with a best fit rate of 123 bits/sec. All intervals reflect 95% bootstrapped confidence intervals.

Results

To check that our presentation rate of 147 ms/word was less than participant's natural reading rate, we calculated each participant's average reading time per word in the baseline condition. The mean and median natural reading rate was 300 ms/word (SEM=11). Three participants had average natural reading rates faster than 147 ms/word, we retained their data in the remainder of the analysis[4]. Accuracy data was analysed using a mixed effect logistic regression with random intercepts for participants, item, story and list. Across all conditions, participant accuracy was on average greater than chance[5] (see Figure 4A). The baseline accuracy was 74% suggesting there was no ceiling effect. In the mSP and 5-RSVP conditions, accuracy was significantly lower than baseline; however, the effect size is small (mSP $OR = 0.70$, 5-RSVP $OR = 0.48$). To account for spurious button presses and abnormally long reading times, we removed the top and bottom 2.5% of the reading time data for each participant, resulting in a 5% data loss overall.

We analysed the reading times for each condition separately. With the SP condition, we aimed to replicate the effect of surprisal reported in the literature when presentation is central and perceptual information management is still in the reader's hands. Reading times were analysed using a mixed effect linear regression with surprisal, log frequency, zero-centered length, and position as fixed effects and random intercepts for participant and story. Replicating the prior literature, there was a significant increase in reading time as surprisal increased ($\beta = 0.54, t = 4.60$, $p < 0.05$).

With the mSP condition, we aimed to test if there would be an effect of surprisal when perceptual information management is no longer in control of the reader. Reading times were analysed using a mixed effect linear regression with the same effect structure as in the SP condition. Even though perceptual control was removed from the reader, there was a significant increase in reading time as surprisal increased ($\beta = 1.24$, $t = 8.60$, $p < 0.05$), suggesting that there are sentence processing related surprisal effects without perceptual control.

Having observed a surprisal effect in mSP, we analyzed the 5-RSVP condition to tease apart whether readers compensate for information processing limits during rapid presentation by buffering information and suspending information processing entirely—resulting in a uniform non-zero profile of surprisal weights, or by incrementally buffering and processing linguistic information—resulting in an increasing profile of surprisal weights across subsequent words. To test these hypotheses, we fit two mixed effect linear models: one with a single beta weight for surprisal and one with a unique beta weight for each word's surprisal. Both models contained random intercepts for participant and story and fixed effects for zero-centered length, position and a frequency term for each of the five words presented in a chunk. We find a significant surprisal effect in both models (single beta model: $\beta = 0.73$, $t = 3.93$, $p < 0.05$), which suggests that the observed surprisal effects are reflective of language processing. As can be seen in the profile of the individual beta weights for surprisal in Figure 4B, the profile is more consistent with a non-uniform profile of weights than the uniform profile ($\chi^2(4) = 11.7$, $p < 0.05$).

First-In, First-Out (FIFO) Buffer Model

To further explore the idea of an incremental linguistic information buffer, we formalize a FIFO buffer model with one free parameter, i.e., the rate of information processing, and fit the model to our data to provide us an estimate of the rate parameter. Our model operates in five 147 ms windows matching our presentation duration for each of the five words. In each window, the new word's information is added to the buffer, and the amount of information that could be processed in that window according to our rate parameter is removed from the buffer. As a result, if a word could not be completely processed in one window, its processing carries over to the next window. If all the information in the buffer is processed, the buffer is empty when the next word appears. The information left in the buffer after the fifth word's window is the information whose processing should contribute to our measured reading time. We use the rate of information processing to predict the expected reading time for that five word chunk. We repeat this process for every chunk in our stories. To arrive at predictions for the surprisal weights, we analyse

[4]Analyses excluding their data do not significantly alter the results we present here.

[5]We did not have an exclusionary cut-off for participant accuracy because we designed our questions to be difficult and inspire engagement with the text.

our expected reading times using linear regression with separate surprisal weights for each of the five words. We do not include an intercept value as we expect the entire reading time to be a function of each chunk's surprisal values. Predicted surprisal weights for a variety of processing rate parameters are shown in Figure 3. Our a-priori prediction was that surprisal weights should increase across word positions, as early words will tend to be fully processed so that later words have a larger effect on post-presentation reading time. As can be seen in Figure 3, our a-priori prediction holds true for a large range of plausible rate parameters.

We fit our buffer model to the observed pattern of surprisal weights using gradient descent to minimize squared error. We find that the best fit rate of information processing is 121 bits/second. The best model fit profile of surprisal weights is illustrated in Figure 4B as the red diamonds.

Discussion

To address the apparent contradiction between the speed at which readers can comprehend text and the information processing limits proposed by sentence processing theories, we suggested that surprisal effects observed in reading tasks might actually be a result of perceptual processing (e.g., Norris, 2006) rather than linguistic information processing. If this were the case, we would expect to see no surprisal effects in the mSP and 5-RSVP conditions, where perceptual control was stripped from the reader. Instead, we found effects of surprisal, suggesting that surprisal effects in RSVP reading have a language processing origin.

Having established the existence of an information processing bottleneck in RSVP, we proposed two possible alternatives[6] that reconcile information processing limits and rapid text comprehension. First, readers may compensate for RSVP by postponing information processing until presentation has ended to ensure all the input is copied into a buffer. Under this account, we would expect a small uniform profile of surprisal weights in our 5-RSVP condition. Second, readers may compensate for RSVP by incrementally buffering and processing linguistic information, allowing perceptual processing and information processing to occur on quick but separate timescales. Our a-priori prediction for the FIFO buffer, in line with the best fit prediction from our model, was that early words should contribute less to the post-presentation reading time than later words, resulting in an increasing profile of surprisal weights in our 5-RSVP condition. As can be seen in Figure 4B, the data are more consistent with the predictions of a FIFO incremental buffer model, suggesting that surprisal effects are present in RSVP and readers compensate for the rapid presentation of text by incrementally buffering and processing linguistic information.

While our experiment was designed to control for a perceptual origin of surprisal effects, it also rules out two decision/motor origins. First, surprisal effects would be expected under Hick (1952)'s law if for each word, readers were choosing the identity of the word from multiple alternatives. Presumably, such decision processes would be disrupted by our 5-SP-RSVP condition. Second, the programming of saccadic eye movements is related to the log predictability of the target (Carpenter & Williams, 1995). RSVP and central presentation would not require eye movements nor provide sufficient time to plan them. It is unclear if our experiment rules out the optimal preparation model of Smith and Levy (2008), which predicts surprisal effects. To summarise their argument, people respond faster to expected events and slower to unexpected events, suggesting that they are preparing for future events according to their predictability. If this finding holds true for all levels of linguistic processing—i.e., under a scale free assumption, the effect of predictability must be on a log scale so as to satisfy the multiplicative nature of joint probabilities with the additive nature of reaction times. As a result, if a word is optimally prepared for, its processing time should be a multiple of its surprisal. Unfortunately, the model does not specify if there is a time cost to preparation or minimum required preparation time. If there is a cost to optimal preparation, our experiment might not have provided sufficient time to prepare. Similarly, the predictions of Smith and Levy (2013)'s highly incremental processing account of surprisal are unclear for our experiment.

Our results have several implications for sentence processing research. First, even with a linguistic information buffer the best fitting rate of information processing is large—i.e., 121 bits/sec, which corresponds to a word processing rate of 12 words/sec. This rate is consistent with the upper limit on comprehension in RSVP tasks (Potter, 1984). Interestingly, even under the suspended processing buffer model the rate of information would be large[7]. So if readers are processing information at such a fast rate, why do surprisal effects show up at all in natural reading tasks? One possibility is that even though readers can process information at these rates, they might prefer to maintain the information they are processing before proceeding. Whether this processing is perceptual—i.e., waiting for some level of certainty in the percept of the word, linguistic—e.g., parallel resource allocation in syntactic parsing, or optimal preparation is still an open question, with the important implication that surprisal effects in natural reading times might not be a measure of syntactic processing difficulty. Another important possibility is that rate of information processing is not consistent across different tasks. In our opinion, RSVP is a demanding task and as such the rate of information processing might differ from natural rates of processing. In this case, surprisal effects in natural reading could reflect a real processing bottleneck, inspiring the new question, why don't we adjust our processing rates to alleviate the bottleneck?

[6]Although, it is clear that there are several alternative proposals to be explored in the future including parallel processing and retrieval accounts (e.g., Lewis & Vasishth, 2005).

[7]Average reading time 516 ms (SEM=2.25) and average bits per five word chunk 50 bits (SEM=0.5) give an information processing rate around 100 bits/sec.

Second, our buffer model suggests that there is a decoupling of perceptual and linguistic information processing[8], which is potentially relevant for two sentence processing phenomenon: spill-over effects and right context effects. In self-paced reading and eye-tracking data (e.g., Smith & Levy, 2013; Shvartsman, Lewis, & Singh, 2014), researchers sometimes discover spill-over effects—i.e., properties of previously fixated/presented words are reflected in the reading time of the current fixated/presented word. Usually spill-over effects are realized as increased reading times following a word thought to be difficult to process (or conveying a large amount of information). Spill-over effects could be explained as perceptual processes continuing to advance through the sensory input while being sensitive to the information processing slightly lagging behind in a buffer. For example, if a buffer had a fixed capacity, perceptual processing might stall on words further in the input than is currently being processed until the buffer has room for more information.

Right context effects occur when information further in the sensory input influences previously perceived information. This plays out differently depending on modality. Readers maintain perceptual uncertainty about word identities (Levy, Bicknell, Slattery, & Rayner, 2009) and their regressive eye movements can be linked to future input increasing uncertainty about past input (Bicknell & Levy, 2010). Eavesdroppers, on the other hand, do not have the luxury of playback. In speech processing, listeners maintain uncertainty about words (Bicknell, Tanenhaus, & Jaeger, 2015) and have to hope that the future context will disambiguate the signal for them. The current proposal is that the processing of a segment of speech operates beyond the duration of the speech segment (Dahan, 2010). Arguably, the maintenance of unprocessed information implicates a linguistic buffer. Future research should look into using buffer models to account for these phenomenon.

Our results, in line with previous RSVP studies, show that readers can process text faster than they would process text naturally. Both the surprisal effect in our mSP condition and the information processing rate parameter in our model of the 5-RSVP condition are the first pieces of evidence for surprisal effects in RSVP. Our data are consistent with a FIFO buffer model suggesting that when readers are quickly bombarded with information, they store linguistic information in a buffer and immediately begin processing that information serially at a fixed rate. Our buffer model suggests a looser temporal coupling between perceptual processing and linguistic processing than had previously been theorized. Our initial analyses using a FIFO buffer model prompt further research on the nature of the buffer and how the buffer may be implicated in other sentence processing phenomena.

References

Bicknell, K., & Levy, R. (2010). A rational model of eye movement control in reading. In J. Hajivc, S. Carberry, S. Clark, & J. Nivre (Eds.), *Proceedings of the 48th annual meeting of the association for computational linguistics (acl)* (pp. 1168–1178). Uppsala, Sweden: Association for Computational Linguistics.

Bicknell, K., & Levy, R. (2012). Word predictability and frequency effects in a rational model of reading. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th annual conference of the cognitive science society.* Austin, TX: Cognitive Science Society.

Bicknell, K., Tanenhaus, M., & Jaeger, T. (2015). Listeners can maintain and rationally update uncertainty about prior words. *Manuscript submitted for publication.[KB].*

Carpenter, R., & Williams, M. (1995). Neural computation of log likelihood in control of saccadic eye movements. *Nature*, *377*(6544), 59.

Dahan, D. (2010). The time course of interpretation in speech comprehension. *Current Directions in Psychological Science*, *19*(2), 121–126.

Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentenial complexity. *Cognitive science*, *29*(2), 261–290.

Hale, J. (2001). A probabilistic earley parser as a psycholinguistic model. In *Proceedings of the second meeting of the north american chapter of the association for computational linguistics on language technologies* (pp. 1–8).

Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology*, *4*(1), 11–26.

Kuperman, V., & Van Dyke, J. A. (2011). Effects of individual differences in verbal skills on eye-movement patterns during sentence reading. *Journal of memory and language*, *65*(1), 42–73.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 11261177.

Levy, R., Bicknell, K., Slattery, T., & Rayner, K. (2009). Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proceedings of the National Academy of Sciences*, *106*(50), 21086–21090.

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive science*, *29*(3), 375–419.

Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., & Brockman, W. (2011). The google books team, joseph p. *Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig, Jon Orwant, Steven Pinker, Martin A. Nowak, and Erez Lieberman Aiden*, 176–182.

Mitchell, D. C. (1984). An evaluation of subject-paced reading tasks and other methods for investigating immediate processes in reading. *New methods in reading comprehension research*, 69–89.

Norris, D. (2006). The bayesian reader: explaining word recognition as an optimal bayesian decision process. *Psychological review*, *113*(2), 327.

Peirce, J. W. (2007). Psychopypsychophysics software in python. *Journal of neuroscience methods*, *162*(1), 8–13.

Potter, M. C. (1984). Rapid serial visual presentation (rsvp): A method for studying language processing. *New methods in reading comprehension research*, *118*, 91–118.

Rayner, K., Schotter, E. R., Masson, M. E., Potter, M. C., & Treiman, R. (2016). So much to read, so little time how do we read, and can speed reading help? *Psychological Science in the Public Interest*, *17*(1), 4–34.

Shvartsman, M., Lewis, R. L., & Singh, S. (2014). Computationally rational saccadic control: An explanation of spillover effects based on sampling from noisy perception and memory. *ACL 2014*, 1.

Smith, N. J. (n.d.). Zs: A file format for efficiently distributing, using, and archiving record-oriented data sets of any size.

Smith, N. J., & Levy, R. (2008). Optimal processing times in reading: a formal model and empirical investigation. In *Proceedings of the cognitive science society* (Vol. 30).

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, *128*(3), 302319.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632.

[8]The optimal preparation model (Smith & Levy, 2008) also suggests this decoupling, although the predictions are less clear.