2014

A new perspective on visual word processing efficiency

Joseph W Houpt, *Wright State University - Main Campus*
James T Townsend, *Indiana University - Bloomington*
Christopher Donkin, *University of New South Wales*

A New Perspective on Visual Word Processing Efficiency

Joseph W. Houpt[a,1,*], James T. Townsend[a], Christopher Donkin[b]

[a]*Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN 47405, USA*
[b]*School of Psychology, The University of New South Wales, Sydney, NSW 2052, Australia*

Abstract

As a fundamental part of our daily lives, visual word processing has received much attention in the psychological literature. Despite the well established advantage of perceiving letters in a word or in a pseudoword over letters alone or in random sequences using accuracy, a comparable effect using response times has been elusive. Some researchers continue to question whether the advantage due to word context is perceptual. We use the capacity coefficient, a well established, response time based measure of efficiency to provide evidence of word processing as a particularly efficient perceptual process to complement those results from the accuracy domain.

Keywords: word perception, word superiority effect, capacity coefficient

As a fundamental part of our daily lives, visual word processing has received much attention in the psychological literature. However, the interest in visual word perception extends beyond its value in communication. The written word is a complex stimulus with which most adults have a large amount of experience. Unlike faces, there is no reason to believe we have any innate ability to perceive words. Thus, word perception may represent the limit of perceptual learning in the absence of innate ability.

Due to the relative ease with which most adults read, it is reasonable to assume that word perception is an efficient process. This is further supported by the intuition that with more experience with a process we become more efficient and we are quite experienced with the written word. Often, the efficiency is measured using single letter perception as a base line. When word context offers an advantage in the accuracy or processing time of perceiving a letter, this supports the claim that word perception is efficient.

From the early days of experimental psychology, researchers have been interested in the value of a word context for perceiving letters. In one study, letters were displayed sequentially to participants at faster and faster rates until they could no longer correctly identify the letters. They found that participants maintained accuracy with shorter durations when the letters were presented as part of a word compared with random letter sequences (Cattell, 1886).

One problem with studies of this nature is that they do not control for the fact that forcing a string to be a word constrains the number of possible letters in the string. Hence, it is not clear from those early results whether the advantage is a perceptual advantage or a decisional advantage. For example, if the last letter of a four letter word is "h", then the second to last is most likely an "s", "t" or "c." Thus, there is redundant information about the identity of the second to last letter: both the perceptual information about the shape of that letter and the decisional information about the letter conditioned on the last letter being an "h." If random letter strings are used, there is no longer the same constraint on the likely identity of the second to last letter: "x" is just as likely as "s" so the only information is the perceptual information about the second to last letter.

In the late 1960's an alternative task was designed to eliminate the decisional advantage of word context so as to examine the perceptual effects. In this task a letter or word was tachistoscopically displayed to a participant. Participants then chose from two possible choices, one of which was correct. In the letter condition, the choices were letters. In the word condition, both choices were words that differed in only a single letter. This design is depicted in Figure 1. Since both alternatives were words, the word context was no longer informative as to the identity of the letter. Participants were still more accurate at perceiving letters in the word condition than the letter condition (Reicher, 1969). Furthermore, they found that participants are also more accurate when identifying letters in words than random letter sequences. This is known as the word superiority effect.

In a follow-up paper, Wheeler (1970) falsified a number of alternative explanations for the word superiority effect. One possible explanation that Wheeler tested was

*Corresponding author
Email addresses: joseph.houpt@wright.edu (Joseph W. Houpt), jtownsen@indiana.edu (James T. Townsend), c.donkin@unsw.edu.au (Christopher Donkin)

[1]This research took place while the first author was a graduate student at Indiana University. He is now at Wright State University, Dayton OH 45435, USA. This work was supported by NIH-NIMH MH 057717-07 and AFOSR FA9550-07-1-0078 awarded to JTT.

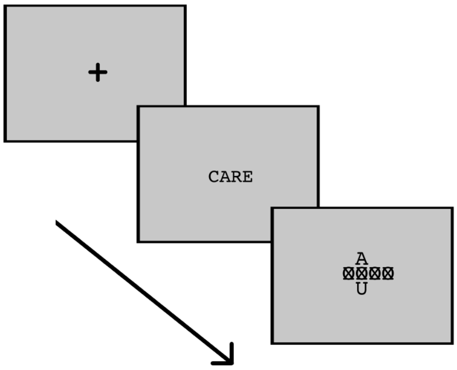

Figure 1: An example trial in the Reicher (1969) task. The participant initiated stimulus onset and the presentation time was calibrated such that participants had 90% accuracy.

that the fact that response choices in the Reicher task were letters, the choices may interfere more with letter stimuli than word stimuli. That interference could lead to worse performance when the stimuli were letters, and hence result in a word superiority effect. Although Wheeler found evidence for interference from the response choices, when the responses were delayed long enough such that there was no longer an effect of increased delay there was still a word superiority effect. In the task described below, we use a different method to eliminate this alternative: We use word responses to word stimuli and letter response to letter stimuli. A second possible explanation Wheeler tested, which also foreshadows our experimental design, is that people may focus their attention on only the positions within a word that disambiguate that word with its orthographic neighbors. For example, the word "wren" can be morphed into "when" by changing the second letter, but cannot be changed into another word by changing the last letter. Like Wheeler, we use words that can be morphed into another word by a single letter change in any position within the word.

An efficiency gain of context over letters alone is not unique to words. If a sequence of letters conformed to the pronunciation rules of English (pseudoword), then participants were again more accurate than letters alone (e.g., McClelland & Johnston, 1977; Carr et al., 1978). This is the pseudoword superiority effect. Researchers have even found a superiority effect for familiar acronyms and initialisms such as DVD (Laszlo & Federmeier, 2007), however unfamiliar sequences of letters tend to be as bad or perhaps worse than letters alone (e.g., Reicher, 1969).

Despite the robustness of the word and pseudoword superiority effects, a comparable effect using response times (and controlling for decisional information due to context) has been elusive. In many studies, response times were not recorded or at least not reported (e.g., Estes & Brunn, 1987; Allport, 2009; Ferraro & Chastain, 1997).[2] Wheeler (1970), for example, found that response times to words were slower than response times to letters, regardless of whether responses were correct or incorrect. The absence of a response time word superiority effect may be in part explained by the possibility that people will read an entire word even if the task does not require it. Indeed, the concept that words are always fully read has been put forth as further evidence that word perception is special (LaBerge & Samuels, 1974). One of the goals of this paper is to demonstrate a response time based word superiority effect, and possibly a pseudoword superiority effect as well.

The word superiority effect has another limitation, it has only been found with stimulus masking. When the stimulus screen is followed by a blank screen, letters can be identified with the same accuracy whether the letters were alone or in a word context (Johnston & McClelland, 1973; Massaro & Klitzke, 1979).

Even in the accuracy domain, some researchers continue to question whether there is a *perceptual* advantage due to word context. For example, Pelli et al. (2003) demonstrated evidence for a model of word perception in which letters are perceived independently and with separate detection decisions on each letter. Their evidence comes from comparing the efficiency of word perception as the number of letters in the word increases. Depictions of longer words have more information about their identity, since the more letters that are known, the fewer possibilities there are for the others. Hence, if a person is able to take advantage of this global information, they should need less per letter information as the number of letters increases. However, a model of word perception based on independent, separate decisions on the letters predicts that as the word length increases, the reader will still need the same amount of information per letter to maintain accuracy. In fact participants did need roughly the same amount of per letter information as the number of letters increased, supporting the latter model.

Pelli et al. (2003) were not the first to propose an independent parallel processing model for word perception. Massaro (1973) and Estes (1975), for example, proposed models in which letters are independently recognized during an initial stage, then word level information is used in a second stage. The second stage of processing accounts for the word superiority effect without appealing to dependence among the perception of the letters in the early stage and without any word to letter level feedback.

In the next section we describe the capacity coefficient, a response time based measure of efficiency. We propose that this measure, along with a task that controls for both the available information and possibly mandatory word

[2]Krueger (1970) found that participants were faster at searching for target letters in words than letters; however, the search task in which participants are focused on a particular letter differs significantly from the Reicher-Wheeler discrimination tasks.

reading, provides evidence of word processing as a particularly efficient process to complement and extend those results from the accuracy domain.

0.1. The Capacity Coefficient

The capacity coefficient, $C(t)$, is a response time based measure of the effect of increased load on processing efficiency (Townsend & Nozawa, 1995; Townsend & Wenger, 2004; Houpt & Townsend, 2012). Specifically, $C(t)$ is a measure of the change in processing rates as the task requires attention to more targets, or possibly more dimensions of a single target. The basic idea of the measure is to compare response times when performing a task with all parts of the stimulus present to the times that would be predicted if each part is processed in parallel, with no difference in speed whether they are alone or with other parts. In terms of word perception, the baseline model for comparison assumes that letters are identified equally as fast when alone or in a word context and, when the letters are in words, they are perceived in parallel. We will refer to this baseline model as the standard parallel model.

The capacity function for an exhaustive task is defined using the cumulative reverse hazard function, $K(t) = \ln F(t); F(t) = \mathrm{P}\{\mathrm{RT} \leq \mathrm{t}\}$, and is similar to the cumulative hazard function used in survival analysis (cf. Chechile, 2011). If K_{c1} is the cumulative reverse hazard for the first character response times, K_{c2} is the cumulative reverse hazard for the second character, etc., and K_S is the cumulative reverse hazard for the string condition, the capacity coefficient is given by,

$$C(t) = \frac{\left[\sum_{i=1}^{4} K_{c_i}\right]}{K_S}. \tag{1}$$

More details on the motivation for this particular form and its connection to the baseline model are given in Appendix A.

Interpretation of the capacity coefficient is based on the participant's performance relative to the standard parallel model baseline. If a person performs better than the standard parallel model, $C(t) > 1$, their performance is referred to as super-capacity. This may happen if there is facilitation of perception between characters. Performance worse than the standard parallel model, $C(t) < 1$, is limited capacity. Inhibition between characters or serial processing of each character individually would lead to limited capacity. When performance is about the same as the standard parallel model, $C(t) \approx 1$, then we refer to it as unlimited capacity.

Houpt & Townsend (2012) developed a null-hypothesis-significance test for workload capacity analysis. If the null hypothesis that the capacity coefficient is equal to one (unlimited capacity) is true then the test statistic will have a standard normal distribution. Conclusions about the capacity coefficient for each individual can be made using a z-test and group level hypothesis can be tested by appropriately combining individuals' statistics. Despite the fact that the capacity coefficient and thus the Houpt-Townsend statistic are nonparametric, the statistic is quite powerful. Furthermore, because the measure is not based on particular distribution of the underlying processes, the conclusions are quite general. Further details of the capacity coefficient are included in Appendix A.

1. Experiment 1

1.1. Method

To properly compare perceptual efficiency across words, pseudowords, nonwords, upside-down nonwords and unfamiliar characters, our task must eliminate the extra information available given a word context. Furthermore, the possibility that words are exhaustively processed automatically may lead to a disadvantage for words on response time measures. To address these issues, we adapted a task from Blaha (2010) which forces exhaustive processing of the characters in a string using an approach similar to Baron & Thurston (1973). This experiment consists of two components. First, we measure the participants' response times to correctly identifying the target string. To ensure exhaustive processing, i.e., that participants base their identification on the entire string and not any subset, we include a distractor of a string with a single character different in each position in the string. For example if the target is "care" then "bare," "cure," "cave" and "card" are used as distractors (see Table 1). Second, the participants distinguish between letters in isolation. Whereas in the exhaustive case the participant needed to distinguish between "bare" and "care," we now only require them to distinguish between "b" and "c." The response times on these tasks are used for computing the predicted performance of the standard parallel model.

1.1.1. Participants

Participants were recruited from the Indiana University population. Eight females and two males participated in this study, all of whom were native English speakers and reported that they did not read or speak Japanese. Their ages ranged from 19-34. All participants reported having normal or corrected to normal vision, no difficulty reading English, and no prior diagnoses of a reading disorder.

1.1.2. Stimuli

Table 1 gives the complete list the stimuli used for both the single character and exhaustive trials for each type.[3] There were five types of stimuli used: words, pronounceable nonwords (pseudowords), unpronounceable nonwords, upside-down unpronounceable nonwords, and strings of Katakana characters. All strings used were four characters

[3]The use of only one target stimulus per version facilitated the modeling, but in future it will be important to test these effects with a variety of target strings.

	Target	Distractors				Single Character							
Word	care	bare	cure	cave	card	c	b	a	u	r	v	e	d
Pseudoword	lerb	nerb	larb	lemb	lerf	l	n	e	a	r	m	b	f
Non-Word	rlkf	vlkf	rtkf	rlhf	rljk	r	v	l	t	k	h	f	k
Upside-down	ɟʞlɹ	ɟʞlʌ	ɟʞʇɹ	ɟɥlɹ	ʞſlɹ	ɹ	ʌ	l	ʇ	ʞ	ɥ	ɟ	ʞ
Katakana	サイクオ	ヘイクオ	サナクオ	サイフオ	サイクノ	サ	ヘ	イ	ナ	ク	フ	オ	ノ

Table 1: Full set of stimuli used for capacity analysis.

long. Word frequency counts (based on Kucera & Francis, 1967) are listed in Appendix B. Pseudowords were taken from the ARC Nonword Database (Rastle et al., 2002). The neighborhood size and summed frequency of the neighbors for each of the pseudowords are also included in Appendix B.

Strings and characters were presented in black Courier font on a gray background. Each character subtended roughly 0.33° degrees of visual angle horizontally and between 0.3° and 0.45° vertically. Strings subtended 1.5° horizontally.

1.1.3. Procedure

All experimental conditions were run using DMDX version 2.9.06 developed at Monash University and at the University of Arizona by K.I.Forster and J.C.Forster. Stimuli were presented on a 17" Dell Trinitron CRT monitor running in 1024x720 mode. Participants used a two-button mouse for their responses.

Participants were paid $8 per session, and received a $20 bonus upon completion of all 10 sessions. Each session lasted between 45 and 60 minutes and was dedicated to one of the five types of stimuli (e.g., word, pseudoword, . . .), so there were two sessions of each type. At the beginning of each session, we read the participant the general instructions for the task while those instructions were presented on the screen. The instructions encouraged participants to respond as quickly as possible while maintaining a high level of accuracy. Each session was divided into five blocks, one block of string stimuli and a block for each of the corresponding single character stimuli.

Each block began with a screen depicting the button corresponding to each of the categories. An example instruction screen is shown in Figure 3. Participants had 40 practice trials, 20 of each category. Next, participants were given 240 trials divided evenly between the two categories, the first 40 of which were not used in the analysis. The trial structure is show in Figure 2. Each trial began with a 300 ms presentation of a fixation cross. After a random delay (300-600 ms), the stimulus was presented for 80 ms. Participants had a maximum of 2500 ms to respond. If the participant responded correctly, the next trial started after a 400 ms delay. If the participant responded incorrectly, a tone was played during the 400 ms delay. The session order was counterbalanced among the participants so that participants completed the different types on different days and in different orders.

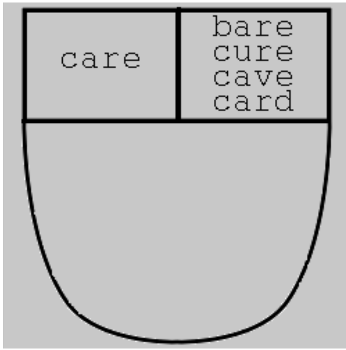

Figure 3: Example instruction screen indicating that the participant should click left if they see care and right if they see bare, cure, cave, or card.

1.1.4. Analysis

All data were analyzed using R statistical software (R Development Core Team, 2011). We computed a repeated measures ANOVA of the correct target response times in each condition using the ez package (Lawrence, 2012) and capacity analyses were completed using the sft package (Houpt et al., 2013).

A repeated measures ANOVA on the string response times (top left of Figure 4) indicated a crossover interaction between version and target/distractor ($F(4, 36) = 20.5, p < 0.05, \eta^2_G = 0.044$) and a significant effect of version on response time ($F(4, 36) = 22.6, p < 0.05, \eta^2_G = 0.49$) but not a main effect of target/distractor ($F(1, 9) = 0.685, p = 0.43$). Post-hoc analysis on target response times was done with repeated measures ANOVA on each pair of versions of the task. Using Bonferroni correction ($\alpha = 0.05/20 = 0.0025$), the following comparisons were significant: Word versus Upside-Down ($F(1, 9) = 50.85, p < 0.0025, \eta^2_G = 0.529$); Word versus Katakana ($F(1, 9) = 57.56, p < 0.0025, \eta^2_G = 0.697$); Pseudoword versus Upside-Down ($F(1, 9) = 34.8, p < 0.0025, \eta^2_G = 0.438$); Pseudoword versus Katakana ($F(1, 9) = 53.9, p < 0.0025, \eta^2_G = 0.643$); Random versus Katakana ($F(1, 9) = 22.1, p < 0.0025, \eta^2_G = 0.398$).

The ANOVA on the string condition accuracy (bottom left side of Figure 4) indicated that there was an interaction between version and target/distractor ($F(4, 36) = 3.69, p < 0.05, \eta^2_G = 0.079$) and main effects of both version ($F(4, 36) = 3.64, p < 0.05, \eta^2_G = 0.11$) and target/distractor ($F(1, 9) = 17.6, p < 0.05, \eta^2_G = 0.081$). Both the interac-

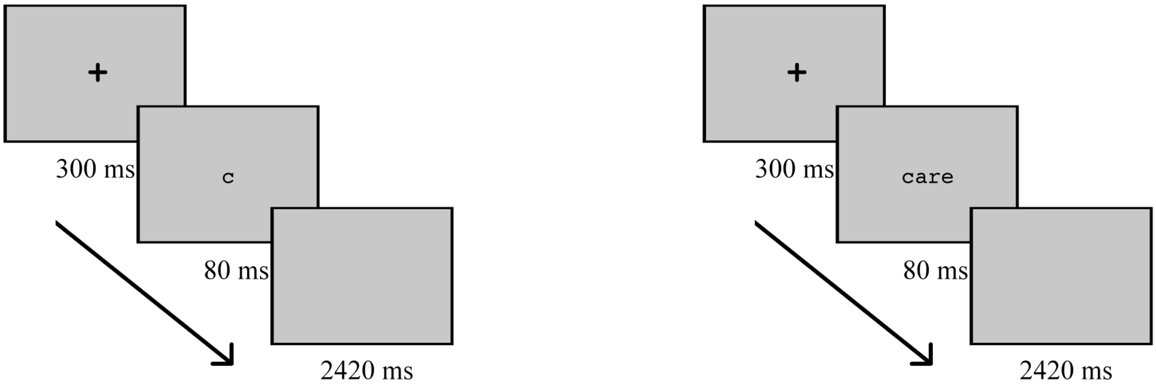

Figure 2: Trial structure for Experiment 1. Trials began with a fixation cross, followed by a blank screen. After a brief, random delay the probe appeared for 80 milliseconds. The probe was followed by a blank screen. Instructions indicating the probe and distractors were given at the beginning of each block.

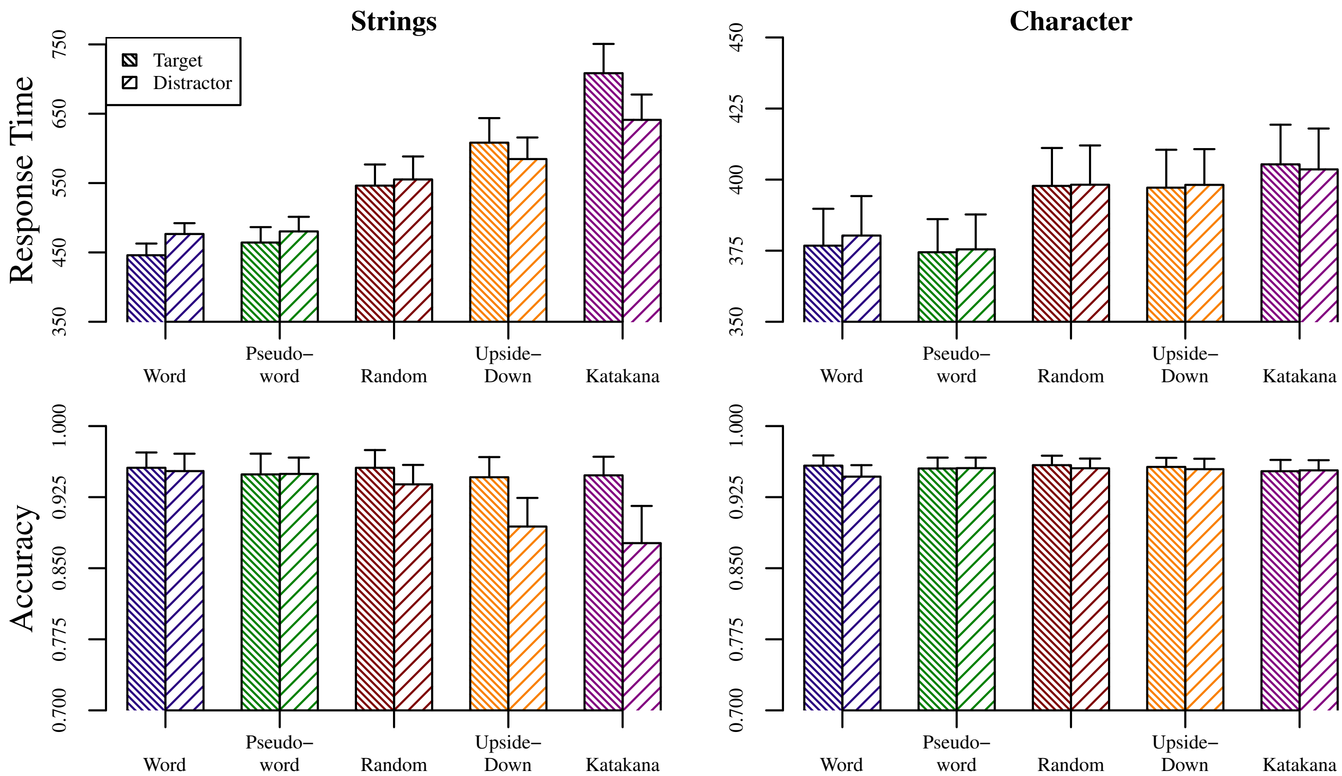

Figure 4: Response times and accuracy from Experiment 1. Error bars indicate the standard error of the mean across trials and participants. The top two graphs show mean response times; the bottom two show accuracy. The left graphs are data from the string tasks; the right graphs are from the corresponding single character tasks. To highlight variation across task version, the character response times are shown on a smaller scale than the string response times. Both accuracy plots are on the same scale.

tion ($W = 0.072, p < 0.05$) and the main effect of version ($W = 0.033, p < 0.05$) failed Mauchly's test of sphericity and only the interaction effect was significant after a Greenhouse-Geisser correction ($GGe = 0.518, p < 0.05$), not version ($GGe = 0.376, p = 0.065$). The effects may be driven entirely by the accuracy on the distractors because there is no significant effect of version when the analysis is limited to the hit rate ($F(4, 36) = 0.411, p = 0.31$).

We found a similar pattern with the single character conditions (right side of Figure 4). There was a significant effect of version on response time ($F(4, 36) = 4.64, p < .05, \eta^2_G = 0.089$), but the main effect of target/distractor ($F(1, 9) = 0.424, p = 0.53$) and the interaction ($F(4, 36) = 0.335, p = 0.85$) were not significant. Post-hoc analysis on target response times was done using repeated measures ANOVA on each pair of versions. Using Bonferroni correction ($\alpha = 0.05/20 = 0.0025$, the only significant differences in response times were between the letters in the pseudoword and upside-down versions ($F(1, 9) = 20.27, p < 0.0025, \eta^2_G = 0.098$) and pseudoword and Katakana versions ($F(1, 9) = 20.0, p < 0.0025, \eta^2_G = 0.092$). The other test results were as follows: Word versus pseudoword ($F(1, 9) = 0.104, p = 0.754$); Word versus Random ($F(1, 9) = 3.29, p = 0.103$); Word versus Upside-Down ($F(1, 9) = 7.55, p = 0.023$); Word versus Katakana ($F(1, 9) = 8.40, p = 0.018$); Pseudoword versus Random ($F(1, 9) = 7.07, p = 0.026$); Random versus Upside-Down ($F(1, 9) = 0.0045, p = 0.948$); Random versus Katakana ($F(1, 9) = 0.592, p = 0.461$). There were not significant effects on accuracy of version ($F(4, 36) = 0.433, p = 0.784$), target/distractor ($F(1, 9) = 4.55, p = 0.062$) and there was no significant interaction ($F(4, 36) = 1.28, p = 0.295$).

Individual capacity coefficients are shown in Figure 5. z-scores for individual and group data, using the statistic in Houpt & Townsend (2012) are shown in Table 2. Each z-score indicates a test of the null-hypothesis that a participant performs equally to a standard parallel model. Significance values are based on a two-sided test. Nearly all participants are significantly different from standard parallel, usually better in the word and pseudoword versions and worse in the Random, Upside-Down and Katakana versions.

Using repeated measures ANOVA, we found a significant effect of version on capacity ($F(4, 36) = 22.64, p < 0.05, \eta^2_G = 0.58$). For post-hoc analyses, we used the z-scores resulting from the mean difference between subjects' capacity z-scores in each pair of version of the task. Word capacity was significantly higher than pseudoword capacity ($z = 7.27, p < 0.0025$), random letter capacity ($z = 22.9, p < 0.0025$), upside-down capacity ($z = 36.7, p < 0.0025$), and Katakana capacity ($z = 45.9, p < 0.0025$). Pseudoword capacity was significantly higher than random letter capacity ($z = 15.6, p < 0.0025$), upside-down capacity ($z = 29.4, p < 0.0025$), and Katakana capacity ($z = 38.6, p < 0.0025$). Random letter capacity was higher than upside-down capacity ($z = 13.8, p < 0.0025$), and Katakana capacity ($z = 22.9, p < 0.0025$). Upside-down capacity was significantly higher than Katakana capacity ($z = 9.19, p < 0.0025$).

1.2. Discussion

Participants responded faster to words and pseudowords than to upside-down nonwords and Katakana strings, following a word and pseudoword superiority effect respectively. However, the comparisons between response times to words and response times to nonwords and pseudowords were not significant, and thus do not indicate superiority effects.

One possible explanation of the basic string response time results is that the individual characters were more difficult to process when they were unfamiliar or upside-down. Even the trend toward faster performance on words compared to nonwords could be due to differences in the speed with which the particular letters are processed: Words tend to contain more common letters and include vowels, compared to unpronounceable random letter sequences and more common letters are perceived faster than less common letters (Appelman & Mayzner, 1981).

Herein lies the advantage of the capacity coefficient. By design, the measure accounts for the processing time of each character in measuring the performance of the string. Despite accounting for faster processing with letters than unfamiliar or upside-down characters, the capacity results still indicate word and pseudoword superiority over Katakana and upside-down strings. Furthermore, unlike the raw response time data, the capacity coefficient indicates word and pseudoword superiority over random letter sequences.

Figure 5 and Table 2 show that there are also superiority effects for words and pseudowords over individual letters, i.e., assuming parallel processing of characters, participants were slower when the characters were presented in isolation rather than in a string. In contrast, capacity for upside-down and Katakana was limited.

Finding word and pseudoword superiority effects with response times, by using workload capacity analysis, is notable because the superiority effects have only been reported in accuracy in the past. Furthermore, the accuracy superiority effects are dependent on post stimulus masking. We have demonstrated a clear superiority of words and pseudowords over single characters, random letter strings, upside-down strings and unfamiliar characters without any masking.

Results from Experiment 1 demonstrate that the capacity coefficient can be used to find a more robust word and pseudoword superiority effects than the traditional Reicher-Wheeler paradigm. With Experiment 2, we verify that the response time superiority effects will hold up in this design when there is post-stimulus masking, as in the original paradigm. Additionally, in Experiment 1, the participants were only shown the instruction screen once, at the beginning of a block. Thus, differences in performance may be due to differences in ability to remember

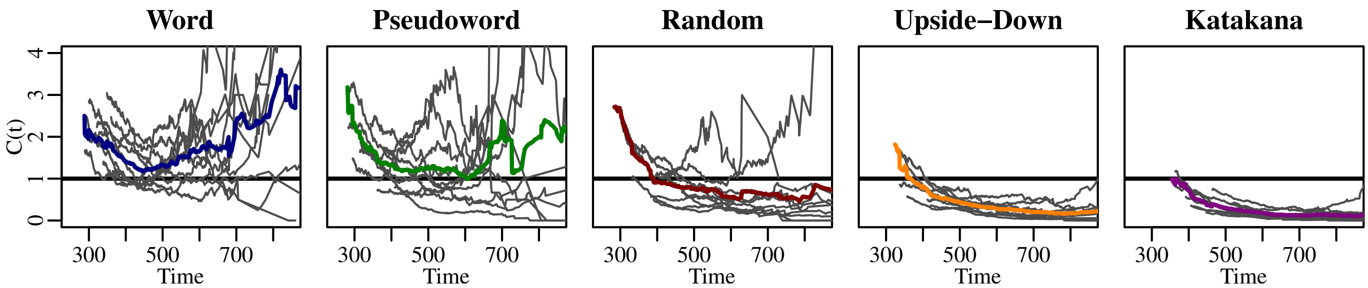

Figure 5: Capacity coefficients for Words, Pseudowords, Random letters, Upside-down random letters and Katakana in Experiment 1. Grey lines indicate individual participants' capacity coefficients and the thick line indicates the average function across participants. The capacity coefficients for each participant are only plotted in regions where reasonable estimates are possible based on individual response time distributions.

	Word	Pseudoword	Random	Upside-Down	Katakana
1	9.97***	3.92***	7.19***	−2.62**	−4.43***
2	11.92***	4.44***	−0.73	−5.95***	−10.02***
3	8.19***	−6.29***	−6.88***	−10.88***	−12.34***
4	0.13	−3.38***	−7.34***	−6.60***	−10.58***
5	0.79	10.70***	−2.36*	−6.27***	−6.86***
6	7.34***	5.19***	10.61***	−2.58**	−11.99***
7	9.34***	3.25**	−2.27*	−2.49*	−5.78***
8	7.17***	7.84***	4.68***	2.86**	−1.79
9	5.71***	13.34***	−8.43***	−9.52***	−7.37***
10	3.88***	2.45*	−2.46*	−7.44***	−9.40***
Group	20.38***	13.11***	−2.52*	−16.28***	−25.47***

Table 2: Workload capacity statistics for each participant in each version of the task in Experiment 1. Under the null hypothesis the limit distribution of the statistic has a standard normal distribution. Significance levels of z-tests are indicated by: $*** : p < 0.001, ** : p < 0.01, * : p < 0.05$.

the target–response mapping across string type. In Experiment 2, we display the instruction screen on every trial. A final potential issue with Experiment 1 is the use of lower case letters. Words with lower case letters can vary more in their global shape than those with only upper case letters (e.g., “BARD” and “CARE” versus “bard” and “care”). This can bias a participant to use global shape information in distinguishing between letter strings. In Experiment 2, we use the same letter strings, but in upper case.

2. Experiment 2

2.1. Method

2.1.1. Participants

As in Experiment 1, participants were recruited from the Indiana University population. Ten females and 2 males participated in this study, all of whom were native English speakers and reported that they did not read or speak Japanese. Their ages ranged from 19-34. All participants reported having normal or corrected to normal vision, no difficulty reading English, and no prior diagnoses of a reading disorder. None of the participants from Experiment 1 participated in Experiment 2.

2.1.2. Stimuli

The stimuli were essentially the same as those used in the Word, Pseudoword and Random versions in Experiment 1, except with capital letters. Because the main effects of interest are the Word, Pseudoword, and Random versions, we did not run the Upside-down and Katakana versions in Experiment 2.

2.1.3. Procedure

Unlike Experiment 1, the stimuli were immediately followed by a mask made of Xs and Os overlayed in each position that a letter was shown (following Reicher, 1969). To allow participants to maintain high accuracy despite the mask, we increased the stimulus presentation time to 100 ms, which we chose based on pilot data. The trial structure is shown in Figure 6.

2.2. Results

Response times in the string condition (top left of Figure 7), showed was a significant effect of version ($F(2, 22) = 12.6, p < 0.05, \eta_G^2 = 0.25$) and a significant interaction between version and target/distractor ($F(2, 22) = 6.36, p < 0.05, \eta_G^2 = 0.0046$), both of which failed Mauchley's Test for Sphericity (Version: $W = 0.194, p < 0.05$; Interaction: $W = 0.532, p < 0.05$) but both remained significant after a Greenhouse-Geisser correction (Version: $GGe = 0.554, p < 0.05$; Interaction $GGe = 0.532, p < 0.05$). There was not a significant main effect of target/distractor ($F(1, 11) = 0.177, p = 0.68$).

There were also significant effects on accuracy in the string condition (bottom left of Figure 7). Both main effects were significant (Version: $F(2, 22) = 14.42, p < 0.05, \eta_G^2 = 0.41$; Target/Distractor: $F(1, 11) = 12.6, p < 0.05, \eta_G^2 = 0.072$) as was the interaction ($F(2, 22) = 5.33, p < 0.05, \eta_G^2 = 0.033$). Again, both version and the interaction failed test for sphericity (Version: $W = 0.132, p < 0.05$; Interaction: $W = 0.531, p < 0.05$) but remained significant after correction (Version: $GGe = 0.536, p < 0.05$; Interaction: : $GGe = 0.536, p < 0.05$).

After Bonferroni correction ($\alpha = 0.05/3 = 0.0167$), all but one of the pairwise comparisons on the target data were significant, the comparison of response times in the Word and Pseudoword versions ($F(1, 11) = 6.49, p = 0.027$). Accuracy comparisons: Word versus Pseudoword ($F(1, 11) = 20.0, p < 0.0167, \eta_G^2 = 0.374$); Word versus Random ($F(1, 11) = 20.12, p < 0.0167, \eta_G^2 = 0.475$); Pseudoword versus Random ($F(1, 11) = 13.23, p < 0.0167, \eta_G^2 = 0.293$). Response time comparisons: Word versus Random ($F(1, 11) = 19.5, p < 0.0167, \eta_G^2 = 0.302$), Pseudoword versus Random ($F(1, 11) = 13.0, p < 0.0167, \eta_G^2 = 0.202$).

In the single character condition (right side of Figure 7), there were no significant response effects of target/distractor ($F(1, 11) = 0.413, p = 0.53$), version ($F(2, 22) = 1.59, p = 0.23$) nor any significant interaction ($F(2, 22) = 1.15, p = 0.33$). There was a significant effect of version on response time ($F(2, 22) = 3.48, p < 0.05$) but neither target/distractor ($F(1, 11) = 0.187, p = 0.67$) nor the interaction were significant ($F(2, 22) = 0.0731, p = 0.93$).

Individual capacity coefficients are shown in Figure 8 and z-scores for individual and group data are shown in Table 3. Nearly all participants are significantly better than the standard parallel model in the word and pseudoword versions. In the random letter condition, half of the participants did not have high enough accuracy to apply the capacity coefficient. The accuracy results for these participants indicate limited capacity because the pseudoword string condition had particularly low accuracy at the group level while the letter level accuracy was not significantly different from the other letter conditions. In fact, all participants except 10 and 11 had worse performance on the random letters strings than would be predicted by independent identification of each letter.[4] However, of those participants that had high enough accuracy, four had significantly super-capacity performance at the $\alpha = 0.05$ level.[5]

[4]This was measured using an accuracy analog to the capacity coefficient: To be correct on the string, one must correctly identify each character. Independent (and unlimited capacity) processing would lead to $\mathrm{P}\{\mathrm{Correct}_\mathrm{S}\} = \mathrm{P}\{\mathrm{Correct}_{\mathrm{c}_1}\} \times \mathrm{P}\{\mathrm{Correct}_{\mathrm{c}_2}\} \times \mathrm{P}\{\mathrm{Correct}_{\mathrm{c}_3}\} \times \mathrm{P}\{\mathrm{Correct}_{\mathrm{c}_4}\}$ where S is the string and c_i is the ith character.

[5]An alternative, parametric approach for measuring capacity is given in Eidels et al. (2010) that accounts for both response time and accuracy differences. We attempted to fit their model for analyzing these data but there were too few condition across which we could constrain parameters, leading to unreliable parameter estimates. Townsend & Altieri (2012) provide a generalized capacity coefficient accounting for both accuracy and response time, although we chose not to include it here because it currently lacks a methods for statistical hypothesis testing.

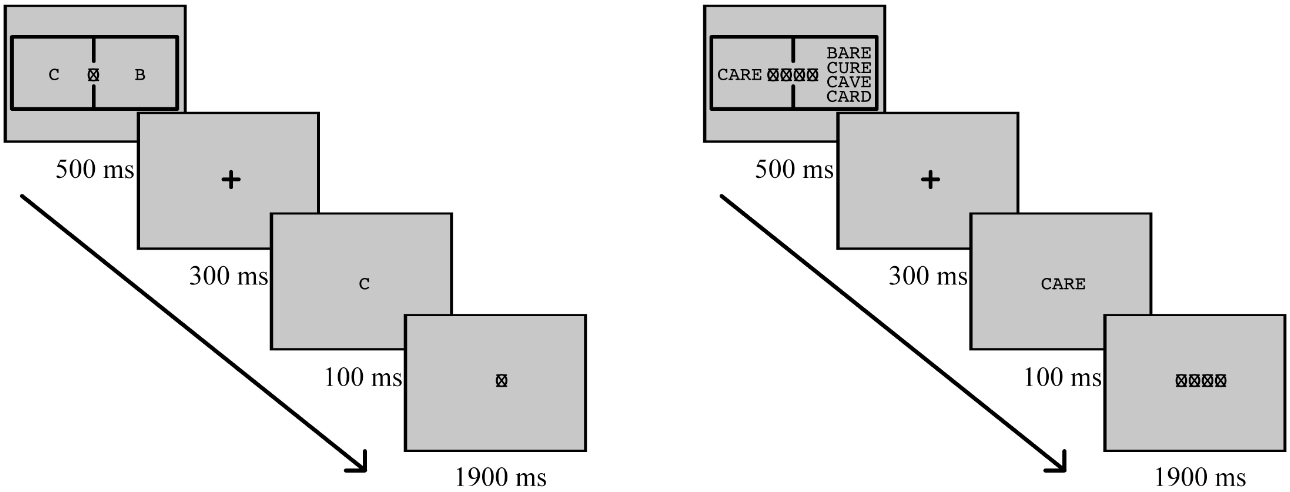

Figure 6: Trial structure for Experiment 2. Trials began with a fixation cross, followed by a blank screen. After a brief, random delay the probe appeared for 100 milliseconds. A mask was presented immediately following the probe. Instructions indicating the target and distractors were given before each trial.

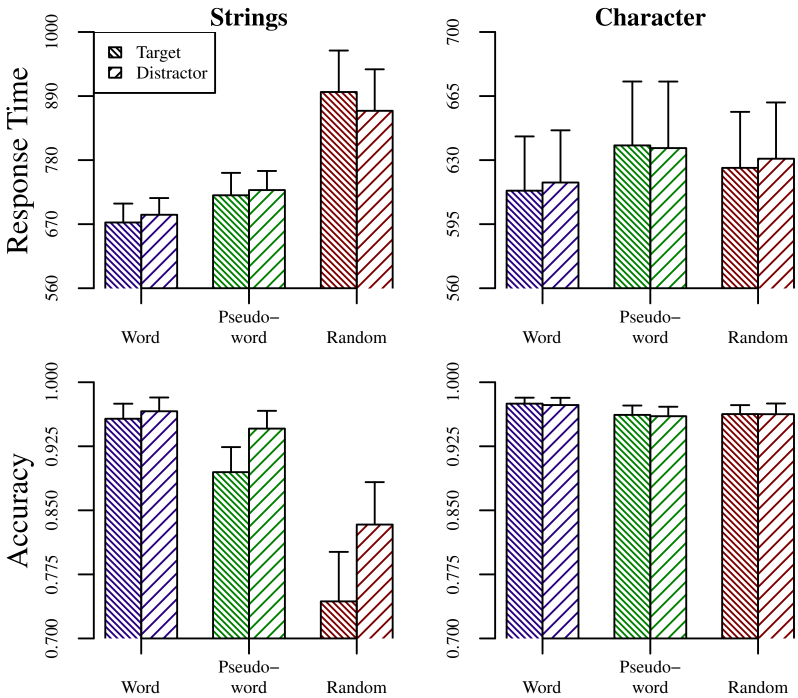

Figure 7: Response times and accuracy from Experiment 2. Error bars indicate the standard error of the mean. The top two graphs show mean response times; the bottom two show accuracy. The left graphs are data from the string tasks; the right graphs are from the corresponding single character tasks. To highlight variation across task version, the character response times are shown on a smaller scale than the string response times. Both accuracy plots are on the same scale.

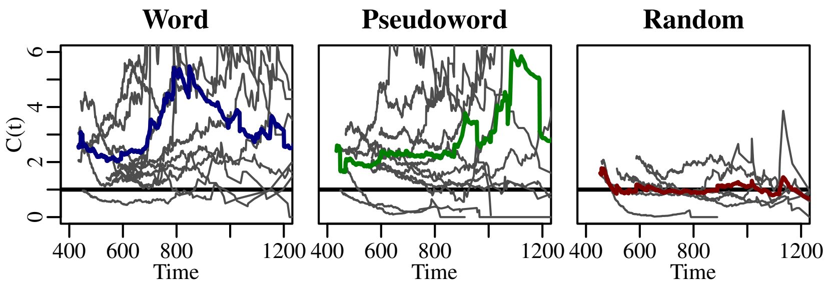

Figure 8: Capacity coefficients for Words, Pseudowords, and Random letters in Experiment 2. Grey lines indicate individual participants' capacity coefficients and the thick line indicates the average function across participants. The capacity coefficients for each participant are only plotted in regions where reasonable estimates are possible based on individual response time distributions.

Due to the missing capacity values, we performed a series of paired t-tests, in lieu of an ANOVA. With Bonferroni correction ($\alpha = .05/3 = .0167$), word capacity was significantly higher than nonword capacity ($t(5) = 5.92, p < 0.0167$) and pseudoword capacity was higher than nonword capacity ($t(5) = 5.95, p < 0.0167$), but word and pseudoword capacity were not significantly different ($t(10) = 0.773, p = 0.458$.

2.3. Discussion

In Experiment 2, all of the single characters were letters, so the lack of any significant effect of version on letter response time and accuracy is not surprising. The random letter version differs from other the two in that all of the characters are consonants, which may be processed slower or less accurate than vowels (Appelman & Mayzner, 1981), but there was no evidence of that difference here. Instead, the capacity differences among the versions are due to the differences in response times in the string conditions. Words and pseudowords were processed faster than random letters and had higher capacity values, consistent with Experiment 1 and the word and pseudoword effects. Also in keeping with Experiment 1, words and pseudowords were super-capacity, indicating superior performance of the letters in those contexts over letter in isolation. Thus, even using masking and upper case letters and minimizing the reliance on memory, there is still a clear indication of the standard superiority effects.

One unexpected result was that the random letter sequences were also super-capacity for many participants, despite being significantly lower capacity than the word and pseudoword version. This may be due to the extensive practice participants had with the target string. Even consonant sequences can show superiority effects if they are highly familiar (Laszlo & Federmeier, 2007). Alternatively, the generally lower accuracy in random version may explain the super-capacity, as the traditional capacity coefficient assumes high accuracy (this is why a half of the participants have NA listed in the Random column of Table 3: their accuracy was too low). Hence, participants may have weighed the relative importance of speed and

	Word	Pseudoword	Random
1	6.01^{***}	4.44^{***}	NA
2	9.64^{***}	2.60^{**}	-0.07
3	7.75^{***}	12.6^{***}	NA
4	7.74^{***}	9.49^{***}	5.44^{***}
5	4.31^{***}	7.03^{***}	NA
6	13.5^{***}	9.06^{***}	0.79
7	5.23^{***}	5.72^{***}	NA
8	9.66^{***}	11.0^{***}	1.96^{*}
9	17.7^{***}	15.3^{***}	5.76^{***}
10	14.5^{***}	11.3^{***}	3.95^{***}
11	10.8^{***}	NA	NA
12	-1.31	-2.66^{**}	NA
Group	30.5^{***}	25.9^{***}	7.28^{***}

Table 3: Workload capacity statistics for each participant in each version of the task in Experiment 2. Capacity coefficients for participants with lower than 80% accuracy on any of the single character conditions or the string condition in a particular version were not calculated. Under the null hypothesis the limit distribution of the statistic has a standard normal distribution. Significance levels of z-tests are indicated by: $*** : p < 0.001, ** : p < 0.01, * : p < 0.05$.

accuracy differently in each version, despite receiving the same instructions for each.

3. General Discussion

We demonstrated clear word and pseudoword superiority effects in response times using the capacity coefficient. This includes a superiority of letter perception in word and pseudoword contexts over letters alone and over letters embedded in random consonant sequences. Furthermore, unlike the standard accuracy based effect, these superiority effects are not dependent on the presence of a mask. By using response times, we can also draw conclusions about the structure of the perceptual processes that were not possible to determine based on the accuracy effect. Foremost, we have clearly rejected an unlimited capacity, independent, parallel processing (standard parallel) model of word and pseudoword perception; the models presented in Massaro (1973) and Pelli et al. (2003) are not consistent with the findings the results reported herein.

To further explore the implications of the capacity coefficient results, we examine each of the multiple plausible explanations for those results. When the Houpt & Townsend (2012) z-test is significant, then at least one of the assumptions of the standard parallel model must have been violated. Note that each of these violations have been considered previously for explanations of the accuracy based superiority effects.

One assumption that may have been violated is that of independence. If there is any type of facilitation between the letter processes, each letter would be processed faster within a word or pseudoword context which would explain the capacity coefficient values above one. There could be many explanations of this facilitation. For example, word processing mechanisms may in fact take advantage of the considerable amount of co-occurrence between letters in English. As is often observed, there are only a fraction of possible four letter combinations used for words and it would be surprising if we did not take some advantage of this reduction in uncertainty. This correlation between letters is an important part of how connectionist models explain the word superiority effect (McClelland & Rumelhart, 1981; Plaut et al., 1996; Coltheart et al., 2001). When the characters are less familiar, such as upside-down letters or Katakana characters, then their confusability may lead to inhibition among the perceptual processes and thus limited capacity.

A related component of many visual word processing models is the phonological pathway (e.g., Coltheart et al., 2001). If a phoneme is activated as a possible interpretation of some letter combination, then it may in turn send positive feedback to those letters, speeding up their processing. Hence, a phonological component of visual word processing could also lead to capacity coefficient values above one. Both the correlation between letters and the lack of a regular pronunciation of the nonwords imply that these predictions are consistent with lack of evidence against the standard parallel model of nonword processing. The phonological explanation is also supported by the evidence of a pseudoword superiority effect.

Another assumption of the standard parallel model is that the letters are processed in parallel, with a separate detection of each letter. An alternative architecture that does predict capacity coefficient values above one is the coactive architecture (Townsend & Nozawa, 1995; Colonius & Townsend, 1997; Townsend & Wenger, 2004; Houpt & Townsend, 2012) which pools information from multiple parallel sources for single decision. By pooling activation from each of the letters when processing a word, the word is processed much faster than if each letter is processed separately. A coactive architecture in this sense can be thought of as an extreme version of a facilitatory parallel model, in which all activation in each of the letters is shared (Eidels et al., 2011). Many connectionist models of visual word perception assume a type of coactive architecture. In these models the activation accumulated in favor of a letter is immediately passed on to the word level. In this framework the type of parallel model assumed in the standard parallel would not pass on any activation until the letter process is complete. Similarly, a holistic model of word perception (e.g., Drewnowski & Healy, 1977) has a coactive form: information pooled for a single identification. There is some middle ground between these two extremes. One example is that of squelching suggested by Pelli et al. (2003). In this case, the activation from the letter process would only be passed on once it is above a certain threshold.

The particularly low capacity for the upside-down and Katakana versions could be due to serial processing of the individual characters. With unfamiliar characters, participants may be forced to check each position in the string, one at a time. All else being equal, serial processing is much less efficient than parallel processing, so it leads to limited capacity. It is important to note that the word and pseudoword results are not necessarily inconsistent with serial processing, but for a serial model to predict capacity-values above one it would need to include large amounts of facilitation and/or require faster processing of individual characters as the number of characters increases (cf. Whitney, 2001).

A coactive architecture could also lead to violations of the assumption of unlimited capacity, so that seemingly more resources are available to each component when more components are present. Capacity values above one imply that the participant dedicated more than four times the resources in the word task compared to the letter task: Each individual letter process in the word has at lease the resources available that were available when that letter was presented in isolation. In this sense the advantage is similar to chunking; when groups of letters are recognized as a single unit, the resources that would have been divided across two individual letter units can be dedicated to a single chunked unit. Participants probably do not have truly unlimited resources to dedicate to the task, there is no doubt an upper limit on the number of letters a person can perceive at once, but having enough resources available to act super-capacity with four letters is not so unreasonable.

In addition to the group level findings, there intriguing individual differences indicted in these data, particularly in word and pseudoword processing capacity. This finding mirrors results reported in accuracy based studies (e.g., Reicher, 1969) and it will be an interesting extension of this work to compare the capacity measure to established measures of individual differences in reading. In fact, research is currently underway using the capacity coefficient to study dyslexia (Sussman et al., 2011).

Another important finding in this paper is that the word superiority effect, as measured by the capacity coefficient, is *not* eliminated in the absence of a post-stimulus mask. This raises the question as to why the accuracy based word superiority effect is less robust. One possibility, raised in the introduction, is that words may be fully processed, even if the task only requires a decision on a part. Thus, the accuracy advantages of a word context might be mitigated by the fact that more is processed in a word context than in a nonword context. This is a special case of the more general issue that response time is more sensitive to certain aspects of perception, such as distinguishing exhaustive and self-terminating strategies and distinguishing coactive and parallel processing, than accuracy (cf., Townsend & Ashby, 1983; Townsend & Nozawa, 1995). In future research, it will be important to determine if capacity coefficient measure of word superiority is robust against other manipulations that may disrupt the accuracy based effect, such as attentional allocation and fixation location (e.g., Johnston & McClelland, 1974; Purcell et al., 1978) or the size of the word Purcell et al. (1978).

We can also examine these results in the context of other configural superiority effects measured by the capacity coefficient. For example, Eidels et al. (2008) demonstrated super-capacity performance when participants could distinguish targets based on global topological properties of the stimulus. In contrast, they found limited or unlimited capacity when the stimuli were made of the same parts as the super-capacity task, but the parts were organized in such a way that the targets were not distinguishable based on their topology. If the same perceptual mechanisms underly the super-capacity in the Eidels et al. (2008) and the current study, this would suggest that the super-capacity performance is driven by global shape of the word, including both the outline as well as the shapes defined by neighboring letters. Without additional assumptions, the global shape explanation would imply super-capacity performance even in the nonwords. It may be that through many years of experience we are specially attuned to the differences between shapes generated by words but not so well attuned for nonword sequences. The shape as the lone explanation of the superiority effect may be a bit of a stretch, but global shape may still play a role in word perception, particularly if there is some sort of unitized representation of the words that is used for recognition (cf., Healy, 1994).

Whether or not learning specific global shapes contributes to word superiority, it is likely perceptual learning is an important part of many configural superiority effects. Blaha (2010) examined the effects of perceptual learning on the capacity coefficient. Using stimuli that Goldstone (2000) had demonstrated could lead to perceptual unitization, Blaha measured the capacity coefficient for targets over the course of multiple days of learning. When the stimuli were novel, participants were extremely limited capacity. Over the course of about a week of training (relatively few trials compared to the number of times we see common words), most participants reached high levels of super-capacity. The parts used in those stimuli were randomly generated "squiggly" lines, for which, like letters, there is no reason to believe people have any innate ability to form unitized representations. Given that Blaha used the same task structure (with squiggly lines in place of letters) and found similar levels of super-capacity at the end of training, we believe that perceptual learning plays an important part in the capacity coefficient word superiority effect. In future work, we hope to explore this connection by using the capacity coefficient to measure word superiority at different stages of the development of reading ability.

Finally, we reiterate the importance of going beyond the simple ANOVA analysis of these data. Merely finding an ordering of the means in the string conditions says nothing about the relative processing efficiencies. For example, faster word processing than nonword processing could be due to the letters in "care" being relatively faster to process than the letters "rlkf". Workload capacity analysis, however, takes the processing of the components into account in estimating efficiency.

3.1. Summary

We have demonstrated response time based evidence for visual word perception as a particularly efficient process using the capacity coefficient. This includes evidence that words are more efficiently perceived than predicted by the individual letter reading times, and evidence from comparing word perception efficiency to nonword stimuli. Based on the workload capacity analysis, there is also evidence for a pseudoword superiority effect in the response time domain although not as strong as for word superiority. The evidence we present negates models of word processing that assume parallel, independent processing of letters with separate decision thresholds on each channel. This deeper level of understanding of visual word perception required a shift from statistics based on comparing means toward a more theoretically rich, modeling-based approach.

Allport, A. (2009). Recognition units in reading: backward masking experiments. *Spatial Vision*, *22*, 473–491.

Appelman, I. B., & Mayzner, M. S. (1981). The letter-frequency effect and the generality of familiarity effects on perception. *Perception & Psychophysics*, *30*, 436–446.

Baron, J., & Thurston, I. (1973). An analysis of the word-superiority effect. *Cognitive Psychology*, *4*, 207–228.

Blaha, L. M. (2010). *A Dynamic Hebbian-Style Model of Configural Learning*. Ph.D. thesis Indiana University Bloomington, Indiana.

Carr, T. H., Davidson, B. J., & Hawkins, H. L. (1978). Perceptual flexibility in word recognition: Strategies affect orthographic computation but not lexical access. *Journal of Experimental Psychology: Human perception and performance*, *4*, 674–690.

Cattell, J. M. (1886). The time it takes to see and name objects. *Mind*, *11*, 63–65.

Chechile, R. A. (2011). Properties of reverse hazard functions. *Journal of Mathematical Psychology*, *55*, 203–222.

Colonius, H., & Townsend, J. T. (1997). Activation-state representation of models for the redundant-signals-effect. In A. J. Marley (Ed.), *Choice, Decision, and Measurment: Volume in Honor of R. Duncan Luce* (pp. 245–254). Mahweh, N.J.: Erlbaum Associates.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*, 204–256.

Drewnowski, A., & Healy, A. F. (1977). Detection errors on the and and: Evidence for reading units larger than the word. *Memory and Cognition*, *5*, 636–647.

Eidels, A., Donkin, C., Brown, S. D., & Heathcote, A. (2010). Converging measures of workload capacity. *Psychonomic Bulletin & Review*, *17*, 763–771.

Eidels, A., Houpt, J. W., Pei, L., Altieri, N., & Townsend, J. T. (2011). Nice guys finish fast, bad guys finish last: Facilitatory vs. inhibitory interaction in parallel systems. *Journal of Mathematical Psychology*, *55*, 176–190.

Eidels, A., Townsend, J. T., & Pomerantz, J. R. (2008). Where similarity beats redundancy: The importance of context, higher order similarity, and response assignment. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 1441–1463.

Estes, W. K. (1975). The locus of inferential and perceptual processes in letter identification. *Journal of Experimental Psychology: General*, *104*, 122–145.

Estes, W. K., & Brunn, J. L. (1987). Discriminability and bias in the word-superiority effect. *Perception & Psychophysics*, *42*, 411–422.

Ferraro, F. R., & Chastain, G. (1997). An analysis of Reicher-task effects. *The Journal of General Psychology*, *124*, 411–442.

Goldstone, R. L. (2000). Unitization during category learning. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 86–112.

Healy, A. F. (1994). Letter detection: A window into unitization and other cognitive processes in reading text. *Psychonomic Bulletinn & Review*, *3*, 333–344.

Houpt, J. W., Blaha, L. M., McIntire, J. P., Havig, P. R., & Townsend, J. T. (2013). Systems Factorial Technology with R. *Behavior Research Methods*, . doi:10.3758/s13248-013-0377-3. Advance online publication.

Houpt, J. W., & Townsend, J. T. (2012). Statistical measures for workload capacity analysis. *Journal of Mathematical Psychology*, *56*, 341–355.

Johnston, J. C., & McClelland, J. L. (1973). Visual factors in word perception. *Perception & Psychophysics*, *14*, 365–370.

Johnston, J. C., & McClelland, J. L. (1974). Perception of letters in words: Seek not and ye shall find. *Science*, *184*, 1192–1194.

Krueger, L. E. (1970). Search time in a redundant visual display. *Journal of Experimental Psychology*, *83*, 391–399.

Kucera, H., & Francis, W. (1967). *Computational Analysis of Present-day American Engish*. Providence: Brown University Press.

LaBerge, D., & Samuels, S. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, *6*, 293–323.

Laszlo, S., & Federmeier, K. D. (2007). The acronym superiority effect. *Psychonomic Bulletin & Review*, *14*, 158–1163.

Lawrence, M. A. (2012). *ez: Easy analysis and visualization of factorial experiments.*. URL: `http://CRAN.R-project.org/package=ez` r package version 4.1-1.

Massaro, D. W. (1973). Perception of letters, words, and nonwords. *Journal of Experimental Psychology*, *100*, 349–353.

Massaro, D. W., & Klitzke, D. (1979). The role of lateral masking and orthographic structure in letter and word recognition. *Acta Psychologica*, *43*, 413–426.

McClelland, J. L., & Johnston, J. C. (1977). The role of familiar units in perception of words and nonwords. *Perception & Psychophysics*, *22*, 249–261.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter percpetion: Part 1. an account of basic findings. *Psychological Review*, *88*, 375–407.

Pelli, D. G., Farell, B., & Moore, D. C. (2003). The remarkable inefficiency of word recognition. *Nature*, *423*, 752–756.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, *103*, 56–115.

Purcell, D. G., Stanovich, K. E., & Spector, A. (1978). Visual angle and the word superiority effect. *Memory & Cognition*, *6*, 3–8.

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna, Austria. URL: `http://www.R-project.org` ISBN 3-900051-07-0.

Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC nonword database. *Quarterly Journal of Experimental Psychology*, *55A*, 1339–1362.

Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, *81*, 274–280.

Sussman, B. L., Houpt, J. W., Townsend, J. T., & Newman, S. D. (2011). EEG correlates of visual word processing efficiency in dyslexia. Poster presented at: Society for Neuroscience Annual Meeting.

Townsend, J. T., & Altieri, N. (2012). An accuracy-response time capacity assessment function that measures performance against standard parallel predictions. *Psychological Review*, *119*, 500–516.

Townsend, J. T., & Ashby, F. G. (1983). *The Stochastic Modeling of Elementary Psychological Processes*. Cambridge: Cambridge University Press.

Townsend, J. T., & Nozawa, G. (1995). Spatio-temporal properties of elementary perception: An investigation of parallel, serial and coactive theories. *Journal of Mathematical Psychology*, *39*, 321–360.

Townsend, J. T., & Wenger, M. J. (2004). A theory of interactive parallel processing: New capacity measures and predictions for a response time inequality series. *Psychological Review*, *111*, 1003–1035.

Wheeler, D. D. (1970). Processes in word recognition. *Cognitive Psychology*, *1*, 59–85.

Whitney, C. (2001). How the brain encodes the order of letters in a printed word: The seriol model and selective literature review. *Psychonomic Bulletin & Review*, *8*, 221–243.

Appendix A. Derivation of Standard Parallel Capacity

The mathematical formulation of this construct can be derived as follows. Suppose, as in our tasks, the participant can only respond when they have identified all of the letters (c_i) in the word (S). Then the probability that she has responded to the word is the probability that she has already identified each of the letters,

$$\mathrm{P}\{RT_S \leq t\} = \mathrm{P}\{RT_{c_1} \leq t, RT_{c_2} \leq t, RT_{c_3} \leq t, RT_{c_4} \leq t\}. \tag{A.1}$$

If we assume that the letters are perceived independently and in parallel, then Equation A.1 can be simplified to,

$$\mathrm{P}\{RT_S \leq t\} = \mathrm{P}\{RT_{c_1} \leq t\}\mathrm{P}\{RT_{c_2} \leq t\}\mathrm{P}\{RT_{c_3} \leq t\}\mathrm{P}\{RT_{c_4} \leq t\}. \tag{A.2}$$

The capacity function for an exhaustive task is defined using the cumulative reverse hazard function, $K(t) = \ln F(t); F(t) = \mathrm{P}\{\mathrm{RT} \leq \mathrm{t}\}$, and is similar to the cumulative hazard function used in survival analysis (cf. Chechile, 2011). If K_{c1} is the cumulative reverse hazard for the first character response times, K_{c2} is the cumulative reverse hazard for the second character, etc., and K_S is the cumulative reverse hazard for the string condition, the capacity coefficient is given by,

$$C(t) = \frac{\left[\sum_{i=1}^{4} K_{c_i}\right]}{K_S}. \tag{A.3}$$

By taking the logarithm of both sides of Equation A.2, we see that the baseline model predicts capacity equal to 1,

$$\begin{aligned}
\log\left[\mathrm{P}\{RT_S \leq t\}\right] &= \log\left[\mathrm{P}\{RT_{c_1} \leq t\}\mathrm{P}\{RT_{c_2} \leq t\}\mathrm{P}\{RT_{c_3} \leq t\}\mathrm{P}\{RT_{c_4} \leq t\}\right] \\
\log\left[\mathrm{P}\{RT_S \leq t\}\right] &= \sum_{i=1}^{4} \log\left[\mathrm{P}\{RT_{c_i} \leq t\}\right] \\
K_S &= \sum_{i=1}^{4} K_i \\
C(t) &= 1
\end{aligned}$$

To measure a participant's performance against the baseline model, performance must be measured when each of the single characters are presented in isolation and when all characters are used together. Response times from each of the single character conditions are used to estimate the cumulative reverse hazard for each term in the sum in the numerator of Equation A.3. The times to respond to all of the characters together are used to estimate the cumulative reverse hazard function in the denominator.

Following Houpt & Townsend (2012), we use the Nelson-Aalen type estimator for the cumulative reverse hazard function. We use $G(t)$ for the number of responses that have occurred in a given condition up to and including time t and T_j to indicate the jth response time in the ordered list of all of the correct response times for that condition. Using that notation, the estimate is,

$$\hat{K}(t) = -\sum_{T_j \leq t} \frac{1}{G(T_j)}.$$

Appendix B. Word and Pseudoword Details

Word	Kucera & Francis Frequency	Pseudoword	Neighborhood Size	Summed Frequency of Neighbors
CARE	162	LERB	2	12
BARE	29	NERB	2	12
CURE	28	LARB	5	27
CAVE	9	LEMB	2	26
CARD	26	LERF	2	15