TBD

Matthew J. C. Crump[1], Walter Lai[1], & Nicholaus Brosowsky[1]

[1] Brooklyn College of the City University of New York

Author Note

In this draft, listed authorship order simply indicates who is participating in the project

Correspondence concerning this article should be addressed to Matthew J. C. Crump, Brooklyn College of CUNY, 2900 Bedford Avenue, Brooklyn, NY, 11210. E-mail: mcrump@brooklyn.cuny.edu

Abstract

Enter abstract here. Each new line herein must be indented, like this line.

Keywords: keywords

Word count: X

TBD

Hicks Law refers to the noticeable slowing of responses in choice reaction experiments when the possible stimuli are less predictable.

Hicks in his experiment, “On the rate of gain of information”, tested his reaction speed on a choice reaction test. During each trial he had to determine which of n lights was turned on. He manipulated the set size n, and found that the greater the number of alternatives the slower his response. His data also suggested that reaction speed is linearly related to the predictability of responses; Trials in which certain responses were disproportionately more likely ellicited faster response times than those in which each response was equally likely.

Hick’s law at larger set sizes however does not have consistent experimental support.

Conrad (1962) in an experiment which required subjects to name 320 nonsense syllables within trials of variable number of distinct syllables (number of alternatives) found that the response time was linearly correlated with the logarithm of number of alternatives. However, Pierce and Karlin in 1957 in an experiment which required subjects to read pages of words as quickly as possible, reported no differences in response times between trials with set sizes that ranged from 4 to 256 words.

Proctor and Schneider in a review of studies testing Hick’s Law attributes the inconsistency in large set size experiments to differences in skill level or practice between subjects and the arbitrariness of Stimulus Response coding/mappings when creating more number of alternatives.

Our experiment sought to test Hick’s law on typing speeds.

In this study, we used typing data from 346 typists to analyze the effect of letter uncertainty on response times. In essence, every letter typed represents one choice reaction test trial from Hick’s experiment. In Hick’s experiment, the probability of each light turning on was manipulated to change the H value. This is analagous to the probability of certain letters appearing at a specific position in a word. Peter Norvig analyzed words from numerous texts and has determined the probability for each letter appearing at a specific position in 2 to 9 letter words.

For example, the first letter of a two letter long word is has a 9.44% chance to be an "a", 14.9% chance to be a "t", and a 25% chance to be an "i". The third letter of a five letter long word is has a 13.5% chance as an "e", 4.91% chance as a "t", and a 11.4 % chance as a "a". The H indices for each letter position - word length pair can be calculated. The H for the 1st Letter, 2-letter-long word (abbreviated 1:2) is 2.85. The H for the 3rd Letter, 5-letter-long word (abbreviated 3:5) is 3.94. The differences in H tell us that the first letter of a two letter long word has lower uncertainty ie. it is more predictable than the third letter of a five letter long word.

This combination of letter probabilities creates an H value for each unique pair of letter position and word length.

One potential problem that can hinder our experiment's ability to test the Hick's Law is the fact that our typists vary in terms of their skill level. This means that differences in RT between subjects will generate noise that may hide the H effect on RTs. An advantage of testing Hick's Law with this data set is that the stimulus response coding is non-arbitrary. The letters that typists see on the screen correspond exactly to the letters on their keyboards.

Methods

Data analysis and pre-processing

We used R (Version 3.4.2; R Core Team, 2017) and the R-packages *bindrcpp* (Version 0.2.2; Müller, 2018), *Crump* (Version 1.0; Crump, 2017), *data.table* (Version 1.10.4.3; Dowle & Srinivasan, 2017), *dplyr* (Version 0.7.4; Wickham, Francois, Henry, & Müller, 2017), *ggplot2* (Version 2.2.1; Wickham, 2009), *knitr* (Version 1.20; Xie, 2015), *papaja* (Version 0.1.0.9655; Aust & Barth, 2018), *Rcpp* (Eddelbuettel, 2017; Eddelbuettel & Balamuta, 2017; Version 0.12.16; Eddelbuettel & François, 2011), *RcppZiggurat* (Version 0.1.4; Eddelbuettel, 2017), and *Rfast* (Version 1.8.8; Papadakis et al., 2018) for all our analyses.

For each subject, we recorded timestamps for each keystroke using JavaScript. We applied the following pre-processing steps. We included IKSIs only for keystrokes involving a lower case letter, and only for correct keystrokes that were preceded by a correct keystroke. Outlier IKSIs were removed for each subject, on a cell-by-cell basis, using Van Selst & Jolicoeur's (1994) non-recursive moving criterion procedure, which eliminated approximately X% of IKSIs from further analysis.

Results

Typing Performance

For each subject, we calculated mean IKSIs as a function of letter position and word length. The letter position and word length factors were not factorially crossed. To determine whether there were differences among the means we submitted the means to a single factor repeated measures design with 45 levels (e.g., letter position|word length: 1|1, 1|2, 2|2, … 9|9). Figure 1 shows mean IKSIs collapsed over subjects, as a function of letter position and word length.

The omnibus test indicated differences among the means were significantly different from chance, $F(44,15180) = 276.74$, $MSE = 1{,}269.66$, $p < .001$. Visual inspection of figure shows several trends across the means consistent with first-letter slowing and mid-word slowing reported by Ostry (1983).

Our more important aim was to determine whether variation among these means can be explained by variation in letter uncertainty. For this reason we do not exhaustively discuss all of the possible 990 differences among these 45 conditions. Nevertheless, we did conduct all 990 comparisons using bonferroni corrected paired samples t-tests. The results are displayed Figure 2, which shows absolute mean differences between conditions color coded for significance (light grey is significant).

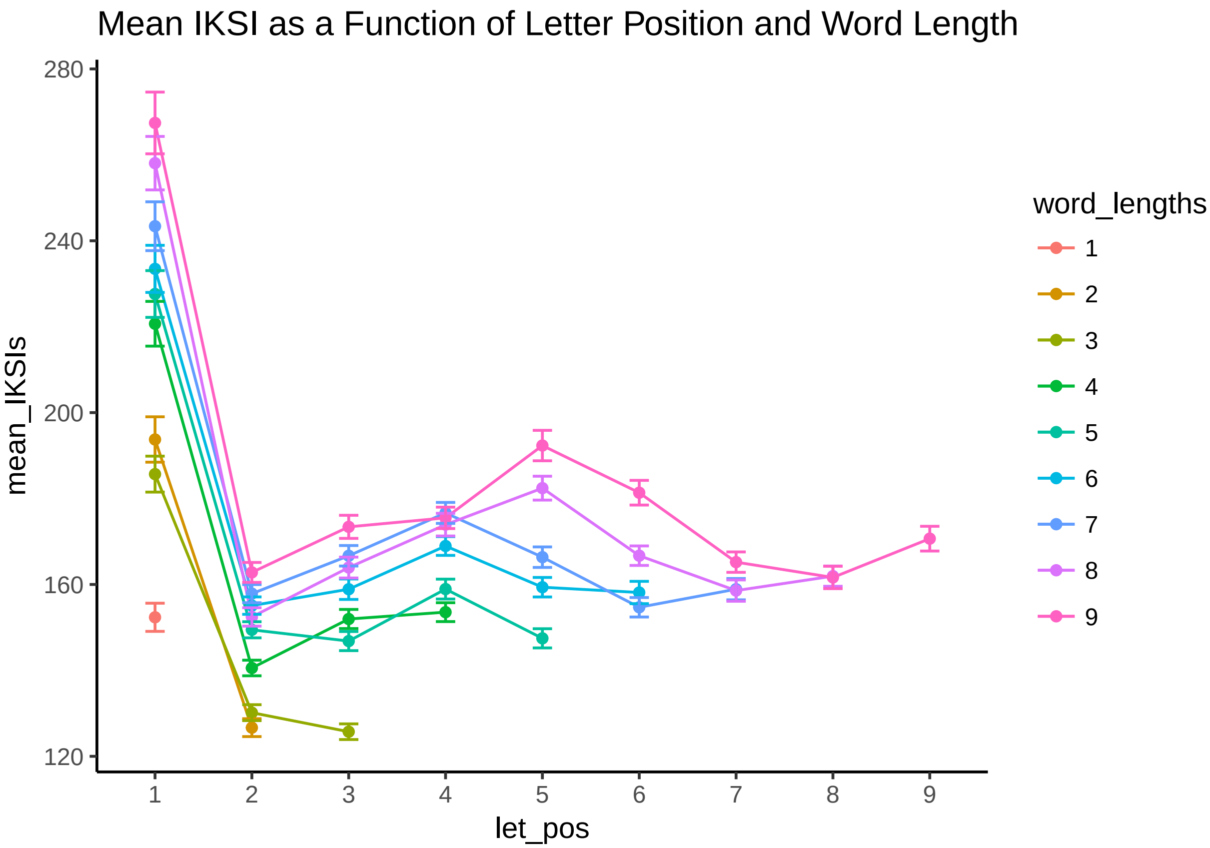

Figure 1

Letter Uncertainty

The primary question of interest was whether natural variation in letter uncertainty by position and word length explains variance in mean IKSI by position and word length. We estimated letter uncertainty by position and word length from google's ngram database, which provides frequency counts of letters and words occuring in Google's massive corpus (X million) of digitized books. Letter frequency counts for letters a to z, for each position in words from length one to nine, were obtained from Norvig ().

For each letter frequency distribution, we computed Shannon's H (entropy) to quantify letter uncertainty. Shannon's H is defined as:

$H = -\sum p \log_2 p$

where, p is the probability of occurence for each letter in a given distribution. H is the

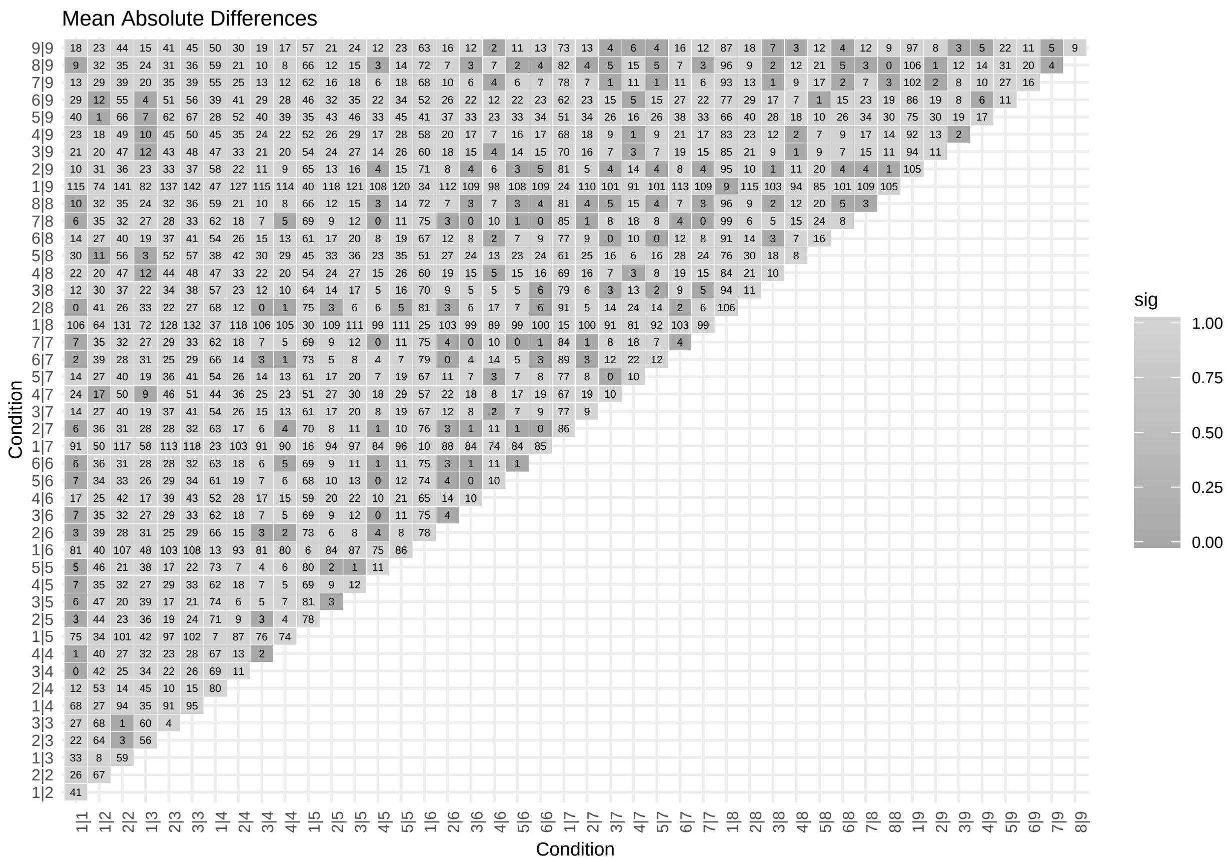

Figure 2

number of bits needed to represent the distribution. We considered only the set of the 26 lowercase letters from a to z. For this set, H can range from 0 to ~4.7. H approaches 4.7 as letter probabilities approach a uniform distribution, indicating all letters are equiprobable, $H = -\sum \frac{1}{26} \log_2 \frac{1}{26} = 4.7004$. H by definition less than 4.7 for all unequal letter probability distributions, where some letters occur with higher/lower probabilities than others.

We converted each letter frequency distribution to a probability distribution then calculated H for each distribution. Figure 3 displays estimates of letter uncertainty (H) as a function of letter position and word length. Visual inspection of the graph shows that variation in letter uncertainty maps closely onto variation in mean IKSI (Figure 1) as a function of position and word length. In particular, letter uncertainty and mean IKSI for position one as a function of word length appear highly similar. And for the remaining positions, letter uncertainty shows an inverted U- shape with greater letter uncertainty in the middle rather than the beginning and endings of words. This suggests that natural variation of letter uncertainty across position and word in English may account for aspects of the first-letter and mid-word slowing phenomena in typing.

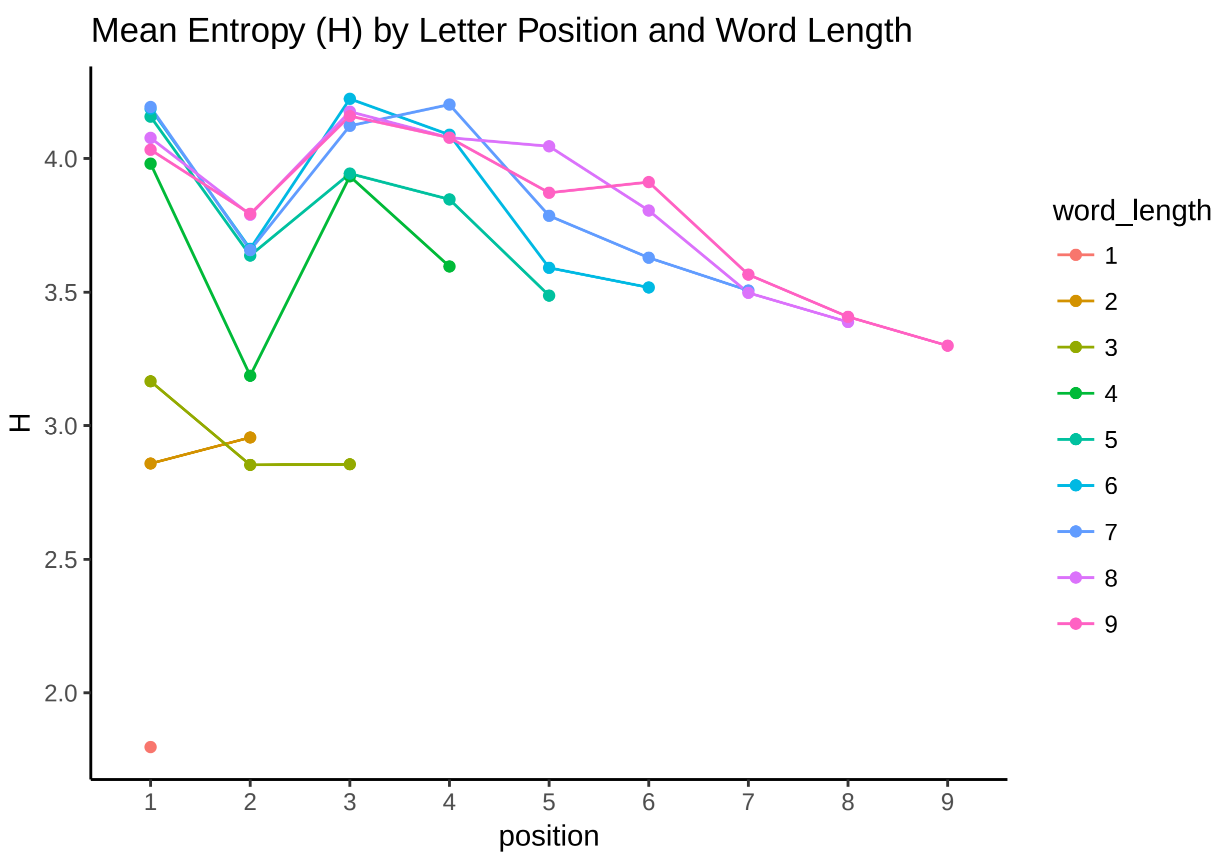

Figure 3

Letter Uncertainty and Mean IKSI

If the Hick-Hyman law applied to continuous typing we would expect a neat linear relationship between mean IKSIs and letter uncertainty. Figure 4 shows a plot of mean IKSIs taken from all positions and word lengths against letter uncertainty. The scatterplot shows a general trend for mean IKSI to increase as a function of letter uncertainty.

A linear regression on the mean IKSIs collapsed over subjects as the dependent variable, and letter uncertainty as the independent variable showed a significant positive correlation, r = 0.46, r^2 = 0.22 , p = 0.0013

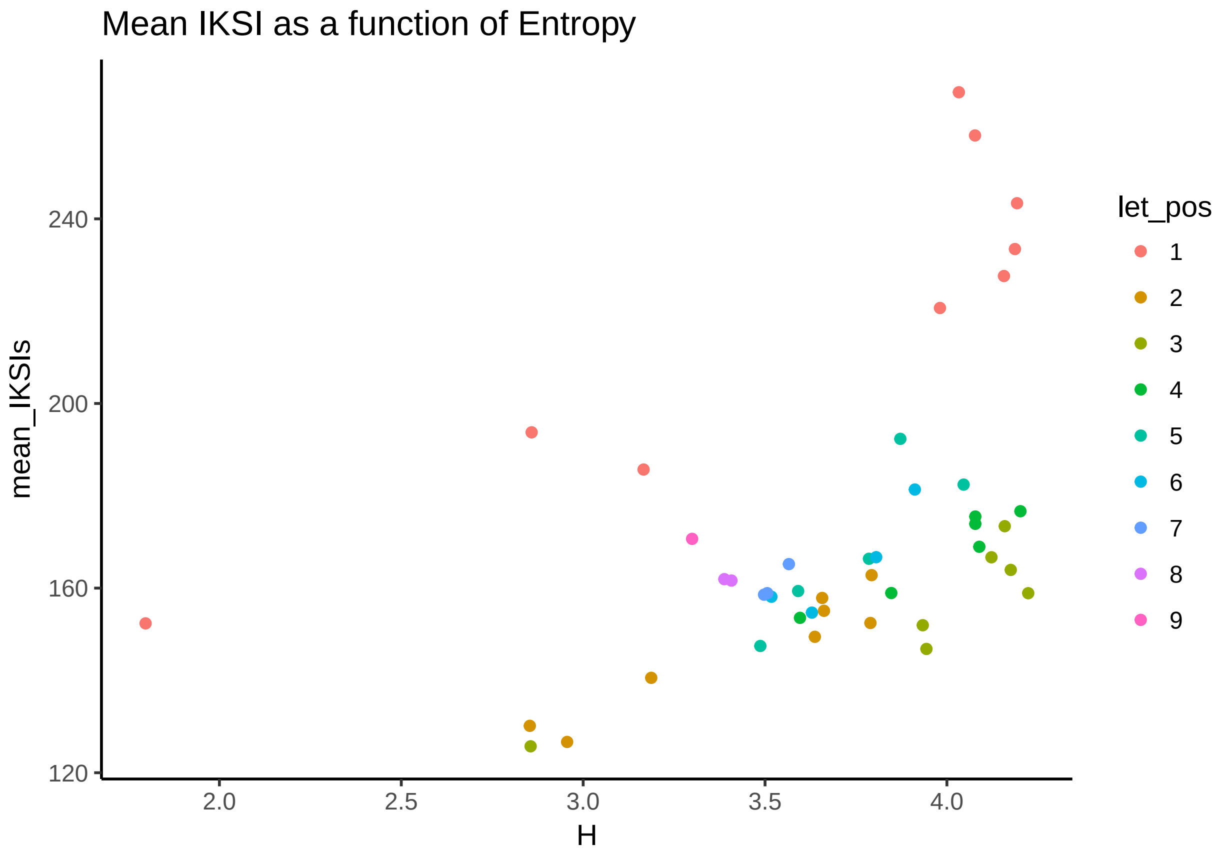

Figure 4

Discussion

The takeaway from this experiment was that a typist types a letter faster when the letters commonly found at the given letter position are few. These findings are consistent with Hick's Law. As the alternatives become equiprobable, as it is observed in the middle of words, the number of bits required to process the choices increases.

Conclusion

Future studies should investigate the role of probability of repetition in regulating response times.

Kornblum in 1969 found that with constant H, trials that had higher chances of sequentially repeating stimuli experienced faster response times.

References

Aust, F., & Barth, M. (2018). *papaja: Create APA manuscripts with R Markdown*. Retrieved from https://github.com/crsh/papaja

Crump, M. (2017). *Crump: Crump lab functions*.

Dowle, M., & Srinivasan, A. (2017). *Data.table: Extension of 'data.frame'*. Retrieved from https://CRAN.R-project.org/package=data.table

Eddelbuettel, D. (2017). *RcppZiggurat: 'Rcpp' integration of different "ziggurat" normal rng implementations*. Retrieved from https://CRAN.R-project.org/package=RcppZiggurat

Eddelbuettel, D., & Balamuta, J. J. (2017). Extending extitR with extitC++: A Brief Introduction to extitRcpp. *PeerJ Preprints*, *5*, e3188v1. doi:10.7287/peerj.preprints.3188v1

Eddelbuettel, D., & François, R. (2011). Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, *40*(8), 1–18. doi:10.18637/jss.v040.i08

Müller, K. (2018). *Bindrcpp: An 'rcpp' interface to active bindings*. Retrieved from https://CRAN.R-project.org/package=bindrcpp

Papadakis, M., Tsagris, M., Dimitriadis, M., Fafalios, S., Tsamardinos, I., Fasiolo, M., … Lakiotaki, K. (2018). *Rfast: A collection of efficient and extremely fast r functions*. Retrieved from https://CRAN.R-project.org/package=Rfast

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Wickham, H. (2009). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from http://ggplot2.org

Wickham, H., Francois, R., Henry, L., & Müller, K. (2017). *Dplyr: A grammar of data manipulation*. Retrieved from https://CRAN.R-project.org/package=dplyr

Xie, Y. (2015). *Dynamic documents with R and knitr* (2nd ed.). Boca Raton, Florida:

Chapman; Hall/CRC. Retrieved from https://yihui.name/knitr/