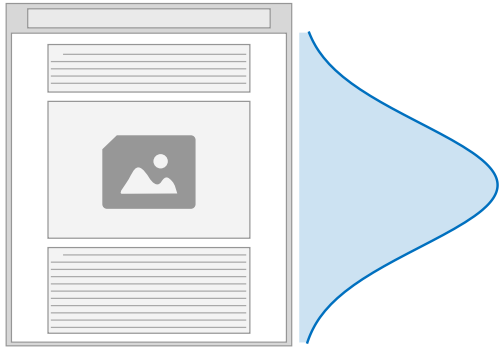


Modeling Sub-Document Attention Using Viewport Time

Max Grusky Jeiran Jahani Josh Schwartz Dan Valente Yoav Artzi Mor Naaman

(with support from Nir Grinberg)

Cornell University

CORNELL TECH

Chartbeat

FOOD

A Superior Chicken Soup

By JULIA MOSKIN NOV. 29, 2016

298

Chicken soup is one of the most painless and pleasing things to make in a home kitchen. But do modern cooks know that?

During America's inexorable march toward processed food, chicken soup became something to buy, not something to make — Campbell's alone produces more than 50 varieties — and many cooks simply don't know how satisfying a project it is.

"So what comes first, the chicken or the soup?" Ashley Aguilar, a radiology technician in San Jose, Calif., asked after spotting an online photo of my homemade chicken soup.

Current understanding of user engagement on the Web is limited mostly to the ***document level***.

OVERALL GOAL

Understand how users interact with documents online.

IN THIS WORK

Design and validate a method of measuring ***sub-document*** user attention.

OUR APPROACH

1. Develop a model of sub-document attention by building on results of prior small-scale lab studies.

2. Validate our sub-document attention model on ***large-scale Web data*** using a known ***user behavior metric***.

1.2 million reading sessions on a popular news site.

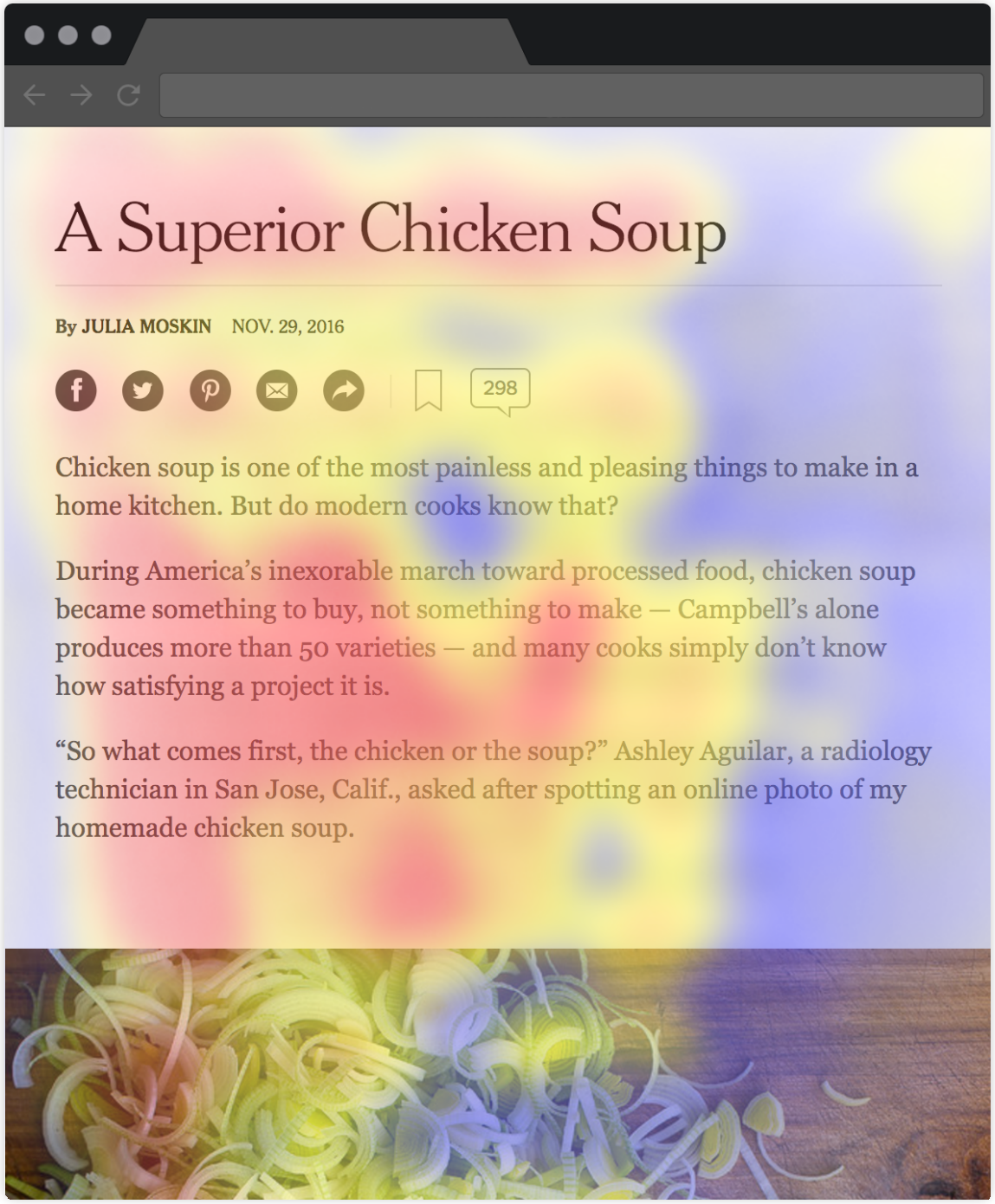
Cross-language reading rate.

Eye tracking to measure attention

+ Fine-grained measurement of attention
+ Great for understanding engagement patterns

– Expensive, dedicated hardware
– Calibration and lab setting
– Does not scale to millions of web users

Our approach: measure attention using only ***standard browser data***.



Uniform attention

- Divide total user attention ***uniformly*** across the page, based on page element size.
  - Does not take into account viewport information.
  - What can we learn about attention from lab studies?

FOOD

A Superior Chicken Soup

By JULIA MOSKIN NOV. 29, 2016

298

Chicken soup is one of the most painless and pleasing things to make in a home kitchen. But do modern cooks know that?

During America's inexorable march toward processed food, chicken soup became something to buy, not something to make — Campbell's alone produces more than 50 varieties — and many cooks simply don't know how satisfying a project it is.

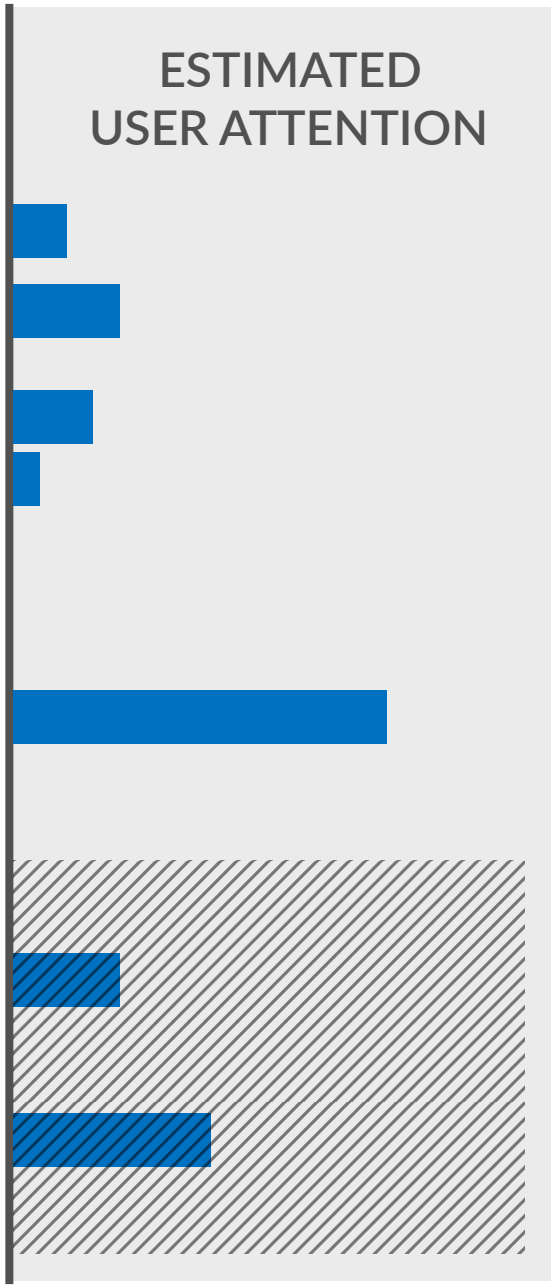
"So what comes first, the chicken or the soup?" Ashley Aguilar, a radiology technician in San Jose, Calif., asked after spotting an online photo of my homemade chicken soup.

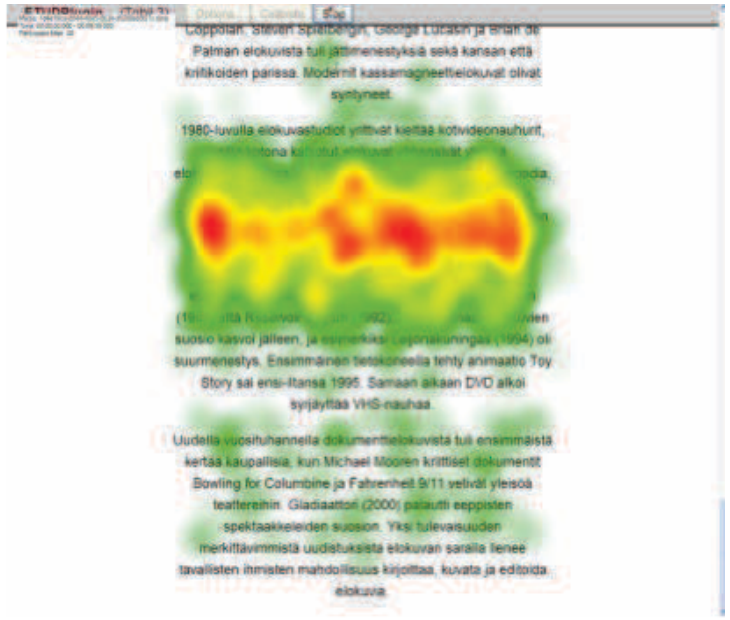
It is a very deep question, beyond the scope of this column.

Fresh leeks and carrots at the ready for finishing the soup. Andrew Scrivani for The New York Times

The chicken soup with root vegetables (carrot, onion, leeks, celery) that we recognize as the American classic was first a staple across Northern Europe. Egg noodles, the perfect filling addition, ranged from the thinnest of white threads to fat yellow twists.

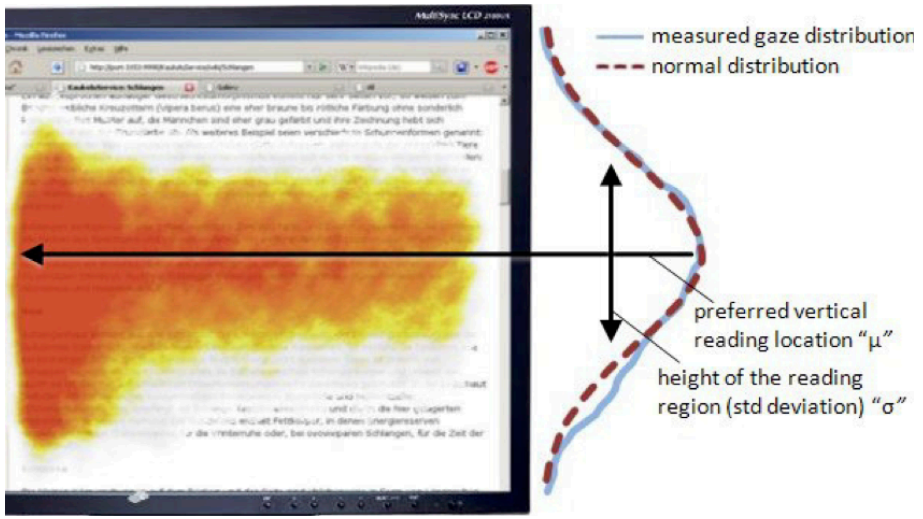
The formulas were carried to the United States by cooks from Scotland (cock-a-leekie), from Poland (rosol) and from all the places in between where Mennonites, Amish and Jews lived. Jewish families in villages across the region raised chickens instead of the more usual pigs, which may explain why Ashkenazi Jews are so connected to chicken soup here. In Yiddish, chicken broth is called goldene yoich, golden broth — much as America was called the goldene medina, the golden land — with all the same connotations of richness, sunshine and good fortune.

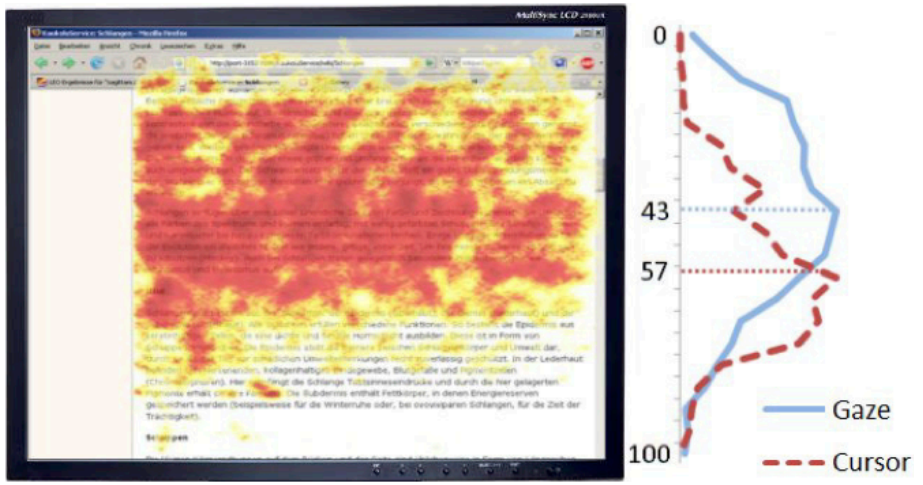




(Sharmin et al., 2013)

Viewport attention distribution is very predictable!

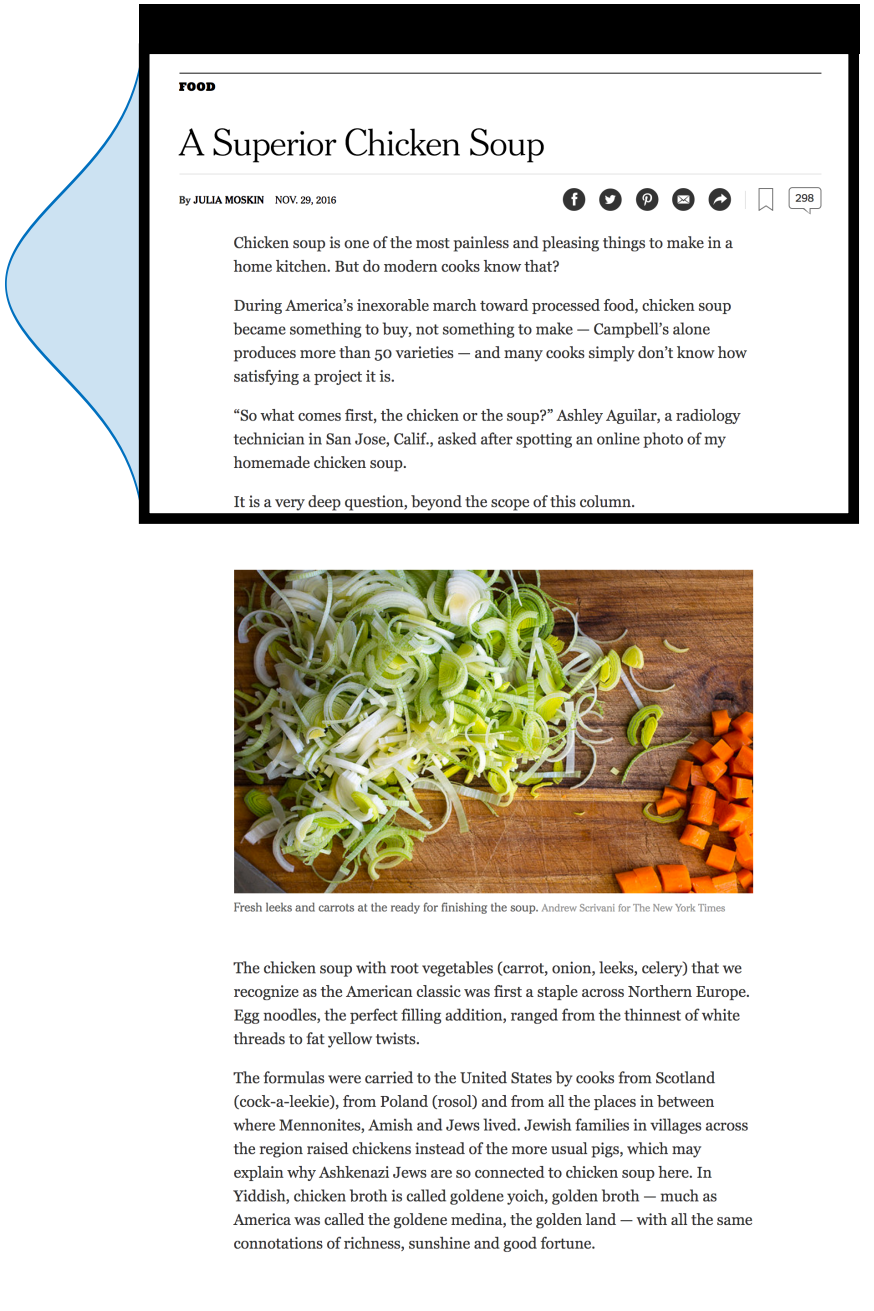


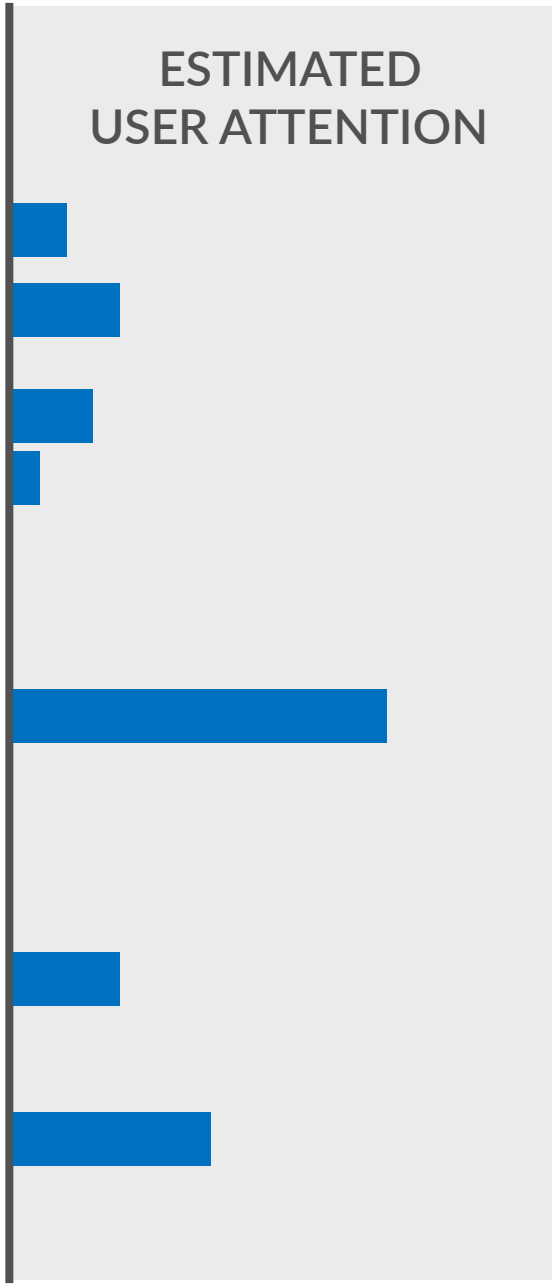


(Buscher et al., 2010)

Gaussian viewport attention

- Uses the user's viewport and attention distribution to assign attention to page elements.
  - + *Empirically motivated:* Builds on prior research in attention.
  - How do we validate it?





User engagement dataset

- 1.2 million reading sessions across a popular new website collected by Chartbeat, Inc.
- Each session consists of second-to-second viewport time data.

Cross-language reading rate

- Readers have predictable and measurable rates across nine languages.

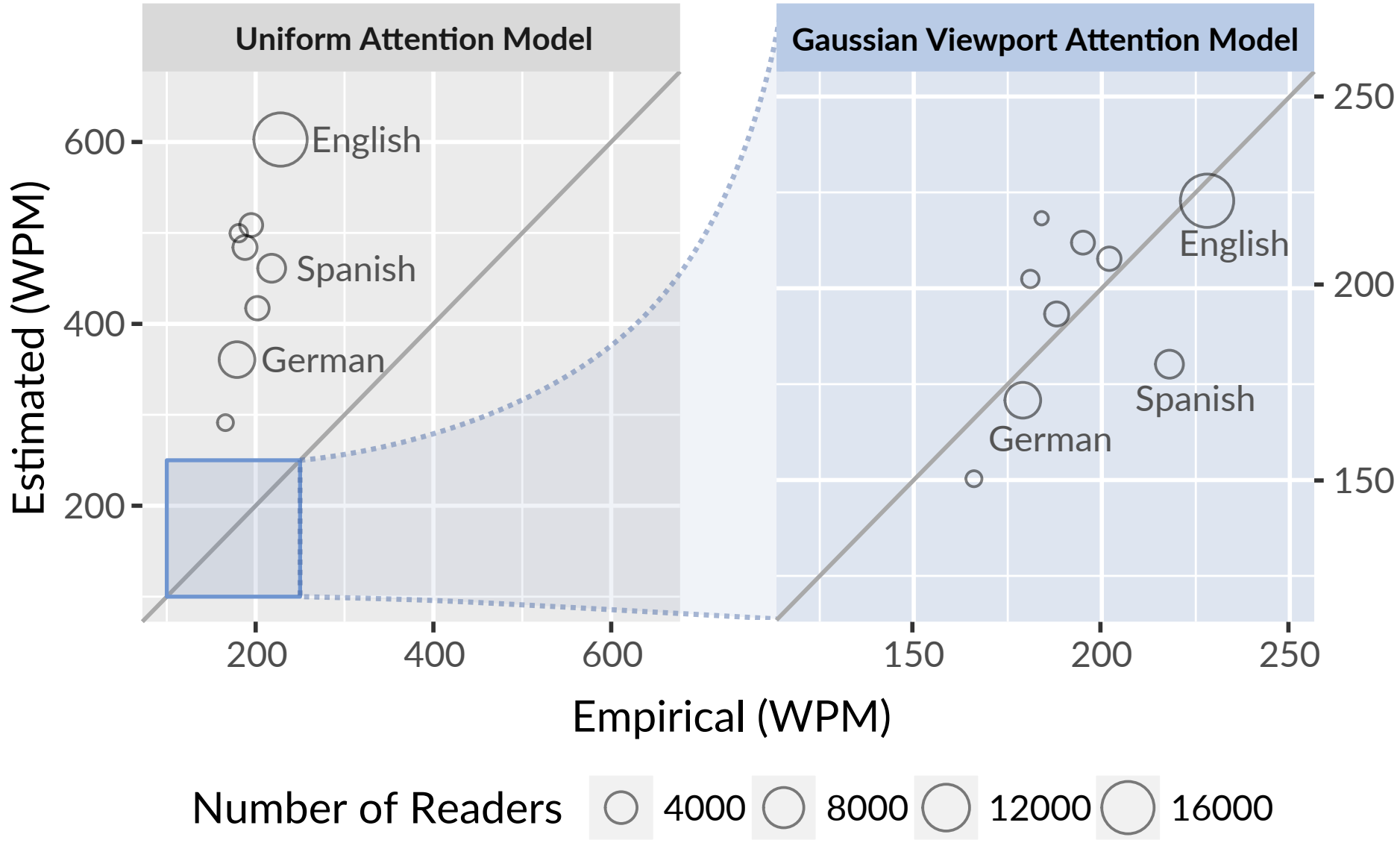
  (Susanne Trauzettel-Klosinski and Klaus Dietz, 2012)

Validation Approach: Reading rates estimated by the attention model should correspond with known reading rates for each language.

5 sec. → *240 WPM*

12 sec. → *220 WPM*

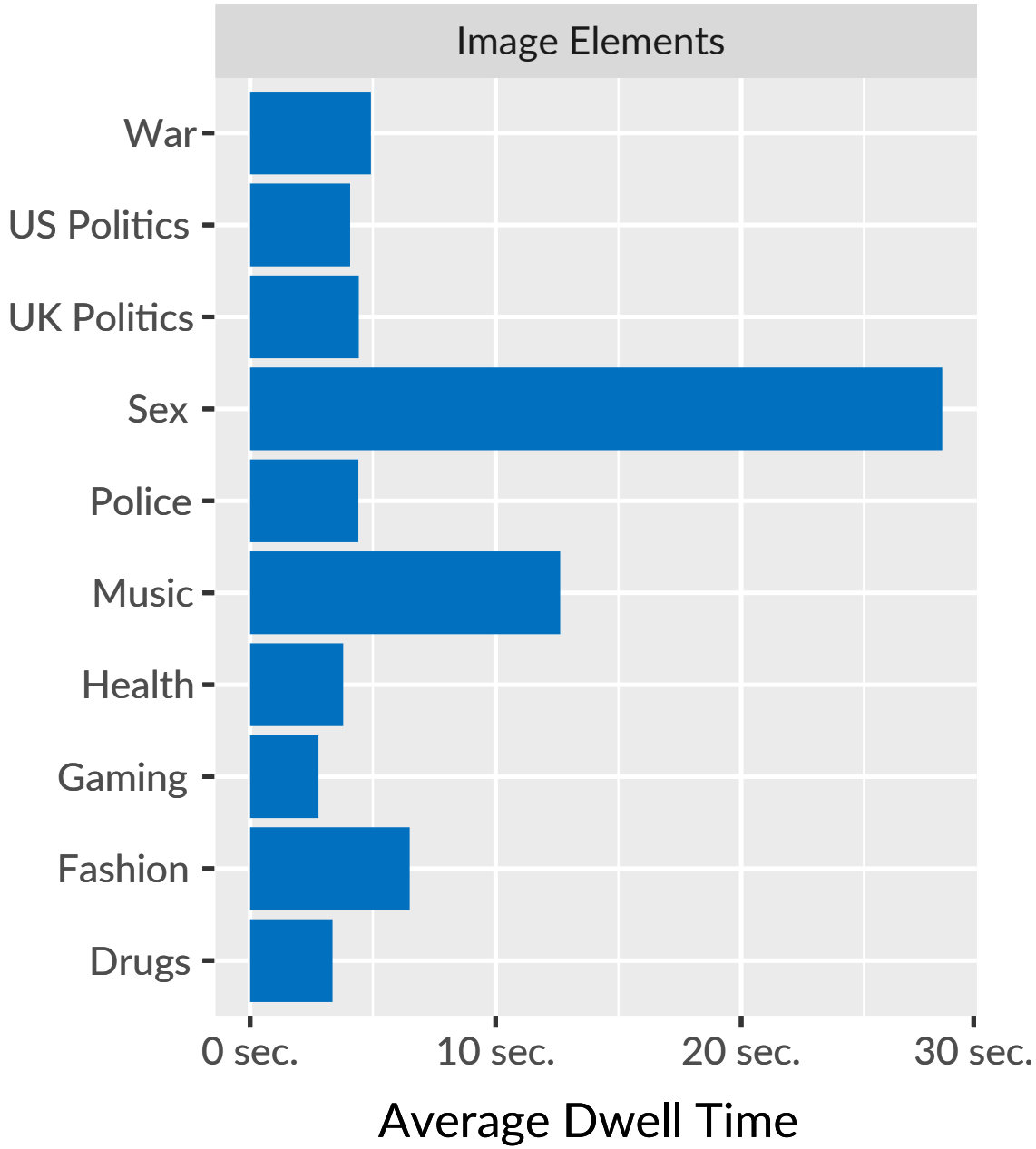
8 sec. → *100 WPM*


Uniform Attention Model
Gaussian Viewport Attention Model
Estimated (WPM)
Empirical (WPM)
600
400
200
250
200
150
200
400
600
150
200
250
English
Spanish
German
English
Spanish
German
Number of Readers
4000
8000
12000
16000

Applying our model

What about images? Which article topics use the ***most engaging*** images?

- Group articles by topic.
- Apply our viewport-based sub-document attention model to each session.



IMPLICATIONS & APPLICATIONS

Understanding language

- Write automatic summaries using sub-document attention distributions.

Understanding engagement

- Help identify and aid struggling readers.
- Better understand user preferences.

CONCLUSIONS

Attention measurement at scale

- We can reliably measure sub-document attention within the browser.
- Especially useful tool at scale.
  - Such as measuring attention of millions of users to thousands of images.

Modeling Sub-Document Attention Using Viewport Time

Max Grusky
Cornell University
grusky@cs.cornell.edu