The Aesthetics of Knowledge Consumption:

Investigating Stylistics and Representation in Online Science Communication

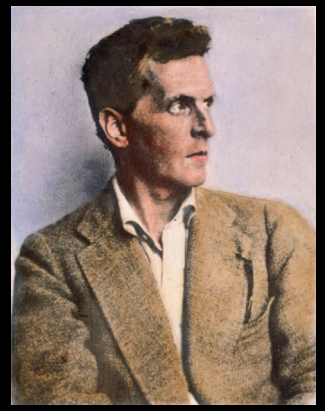
Leoson Hoay

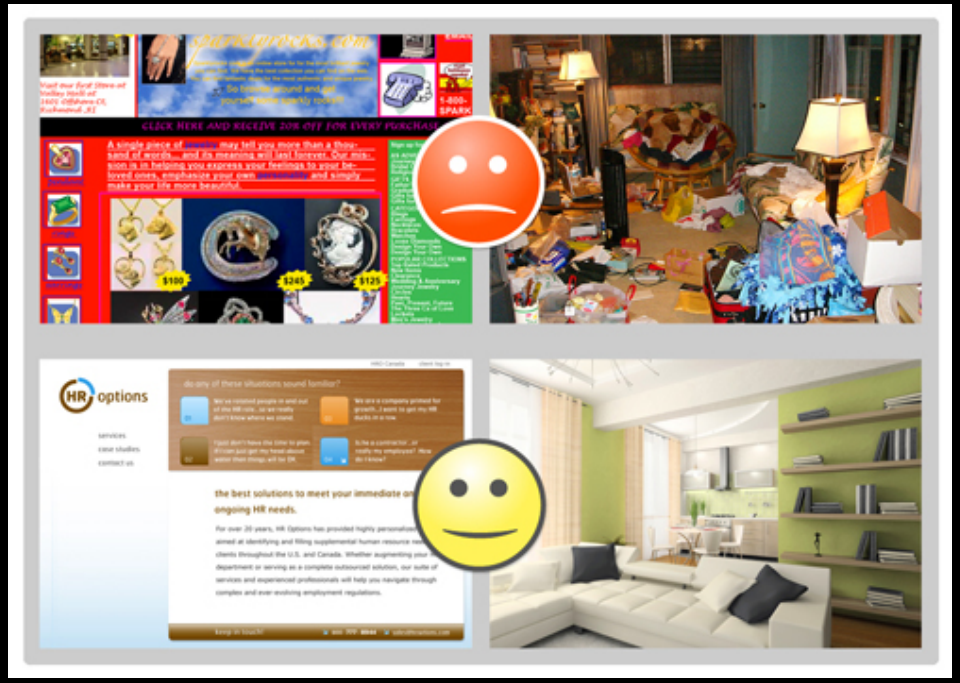
Advisor: A/P Hoyt Long - Knowledge Lab | Department of East Asian Languages and Civilizations

Additional Thanks: A/P Patrick Jagoda - Department of English | Department of Cinema and Media Studies

FOUNDATIONS AND RESEARCH QUESTION

- "Knowledge is in the end based on acknowledgement."/"Ethics and Aesthetics are one." - Wittgenstein
- Value and Aesthetics are inextricable (Gombrich, 1960)
  - **Especially in communicative acts – science communication is not spared from this!**





Talking about Aesthetics in Science

- The effects of aesthetics and writing style has been widely studied in marketing/commerce/design
  - Well-designed visuals promote attention (Markovic, 2012)
  - And increase perceived reliability (Robins & Holmes 2008, Alsudani & Casey 2009, Goering et. al 2011)
- In education
  - Well-written, readable texts result in better motivation and problem-solving (Walkington et. al)
- Science Communication – to advance the knowledge society, to increase interest in science and encourage knowledge sharing
  - General public outside of scientific circles consume scientific content **through media**
  - The "network" (Jagoda, 2016)

Science Communication

- Studies in science communication have been traditionally focused on contextual/social factors, and public literacy (Nisbet and Scheufele 2009 and others)
  - Less emphasis on actual technique and quantifiable metrics
  - Calls by sociologists over recent years to develop more formalized examinations of hermeneutics and knowledge diffusion (Declich & d'Andrea 2005, Leydesdorff 2009, Nielsen 2013)

- **Are computational measures of aesthetics and style robust enough in predicting human readers' perceptions along similar measures?**

- **Can we 'open the door', so to speak, to more empirical pathways through which science communication can be assessed?**

Defining ‘Aesthetics’ and Stylistics

- **Visual Aesthetics**
  - Colorfulness (Reinecke et. al 2013)
  - Screen Balance (Ngo et. al 2000, Altaboli & Lin 2011)
- **Textual Style**
  - Readability (De-jargonizer, Flesch-Kincaid)
  - Uncertainty/Hedging (Vincze et. al 2014)
- **How do these correlate with readers’ perception of:**
  - Aesthetic Design
  - Colorfulness
  - Tidiness
  - Reliability
  - Readability
  - Enjoyment

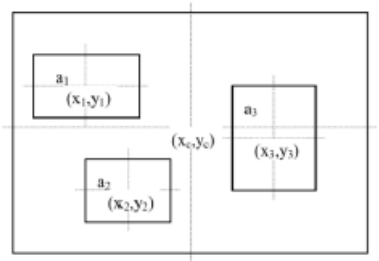


Figure 5: A screen in equilibrium (Ngo et. al, 2000).

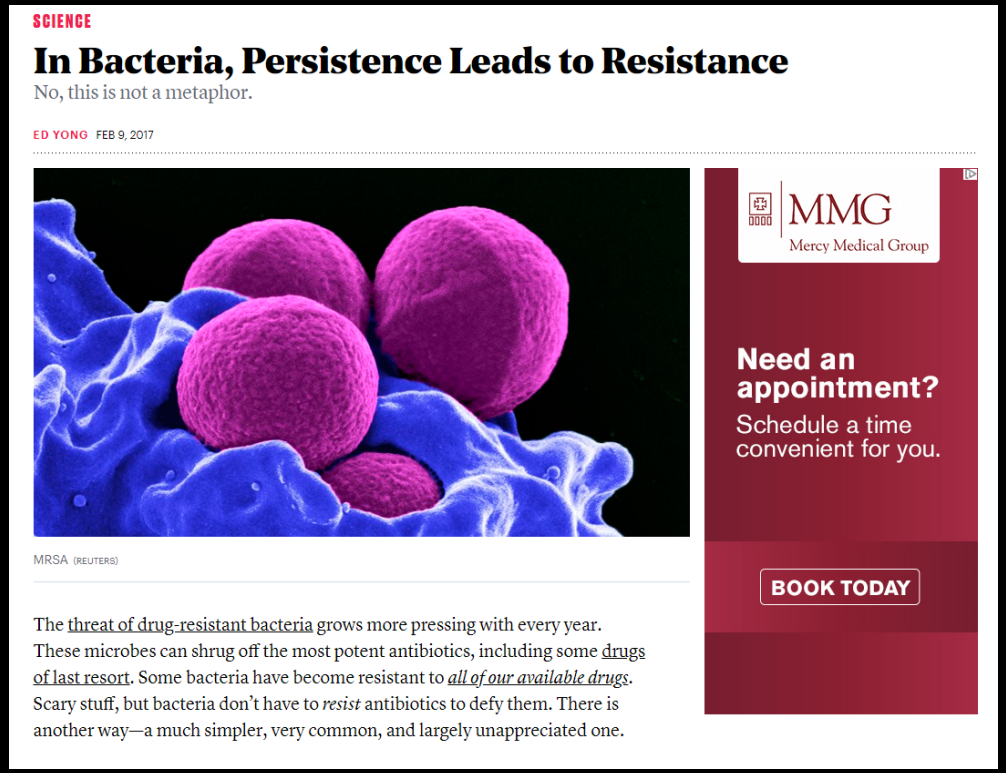
might not be
wonder trying to say
maybe
could be perhaps
probably
to some degree
some confidence
probably true

‘Hedge words’

Study Design

- Science Article extracted from The Atlantic
  - Manipulated on Colorfulness (High/Med/Low)
  - Screen Balance (Balanced/Not Balanced)
  - Readability (High/Low)
  - Level of Uncertainty (High/Low)

- 250 MTurk Participants
  - With demographic controls



SCIENCE

In Bacteria, Persistence Leads to Resistance

No, this is not a metaphor.

ED YONG FEB 9, 2017

MRSA (REUTERS)

The threat of drug-resistant bacteria grows more pressing with every year. These microbes can shrug off the most potent antibiotics, including some drugs of last resort. Some bacteria have become resistant to *all of our available drugs*. Scary stuff, but bacteria don't have to *resist* antibiotics to defy them. There is another way—a much simpler, very common, and largely unappreciated one.

Study Design (Tools)

- **EBImage (R)** to assess and manipulate visual aesthetics
  - aPixel Manipulation, Element Distance
- **De-jargonizer** (Rakedzon et. al, 2017)
  - Readability model trained on 250k BBC articles
- Flesch-Kincaid Reading Ease
- Uncertainty Classifier (Vincze et. al, 2014)

```
In [51]: cls = Classifier(binary = True)
         pred_list = cls.predict("The scientists may have found something that seems to hint at the existence of dark matter.")
         print(pred_list)
         print("")

         total = 0
         for j in pred_list:
                 if j == 'U':
                     total+=1
         print("Uncertainty score: ", total/len(pred_list))

         ['C', 'C', 'U', 'C', 'C', 'C', 'C', 'U', 'C', 'C', 'C', 'C', 'C', 'C', 'C', 'C', 'C']

         Uncertainty score:  0.11764705882352941
```

Uncertainty Scoring: Hedge Words ÷ Total Words

Colorfulness is operationalized by first defining the opposing color spaces:

$$rg = R - G \tag{2}$$

$$yb = \frac{R - G}{2} - B \tag{3}$$

Then define the standard deviation(μ) and mean(σ), before computing the colorfulness metric C (Hassler and Susstrunk, 2003):

$$\mu_{rgyb} = \sqrt{\mu_{rg}^2 + \mu_{yb}^2} \tag{4}$$

$$\sigma_{rgyb} = \sqrt{\sigma_{rg}^2 + \sigma_{yb}^2} \tag{5}$$

$$C = \sigma_{rgyb} + 0.3 \times \mu_{rgyb} \tag{6}$$

Operationalizing Colorfulness

Synonym Replacement

- Using Wordnet's **synset.lemma_names()** to obtain synonym lemmas of a word, then checking it against the word rarity model borrowed from Rakedzon et. al 2017 -> replace rare words with more common words

In [94]:
```
def replacer(sent, ref):
    ls_rep = sent.split(" ")
    new_sent = []
    for i in ls_rep:
        if i in list(ref[0]):
            if(int(ref.loc[ref[0] == i][1]) < 600):
                syn = wn.synsets(i)[0]
                for j in syn.lemma_names():
                    if j in list(ref[0]):
                        if(int(ref.loc[ref[0] == j][1]) < 600):
                            pass
                        else:
                            new_sent.append(j)
                            break
                    else:
                        new_sent.append(j)
                        break
            else:
                new_sent.append(i)
        else:
            new_sent.append(i)

    return new_sent
```

In [98]:
```
sent1 = "The apoptosis observed in the sample was unusual."
print("Original sentence: " + sent1)
print("")
print("New sentence: " + " ".join(replacer(sent1, bbc)))
```

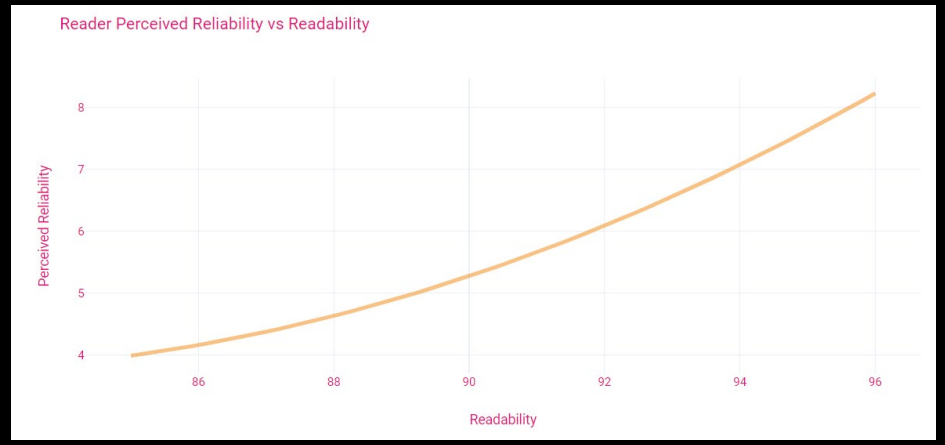
```
Original sentence: The apoptosis observed in the sample was unusual.

New sentence: The programmed_cell_death observed in the sample was unusual.
```

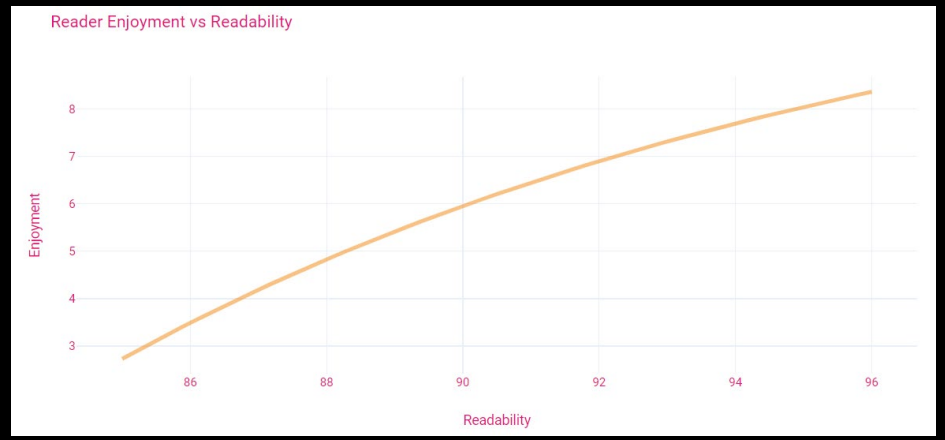
Study Design

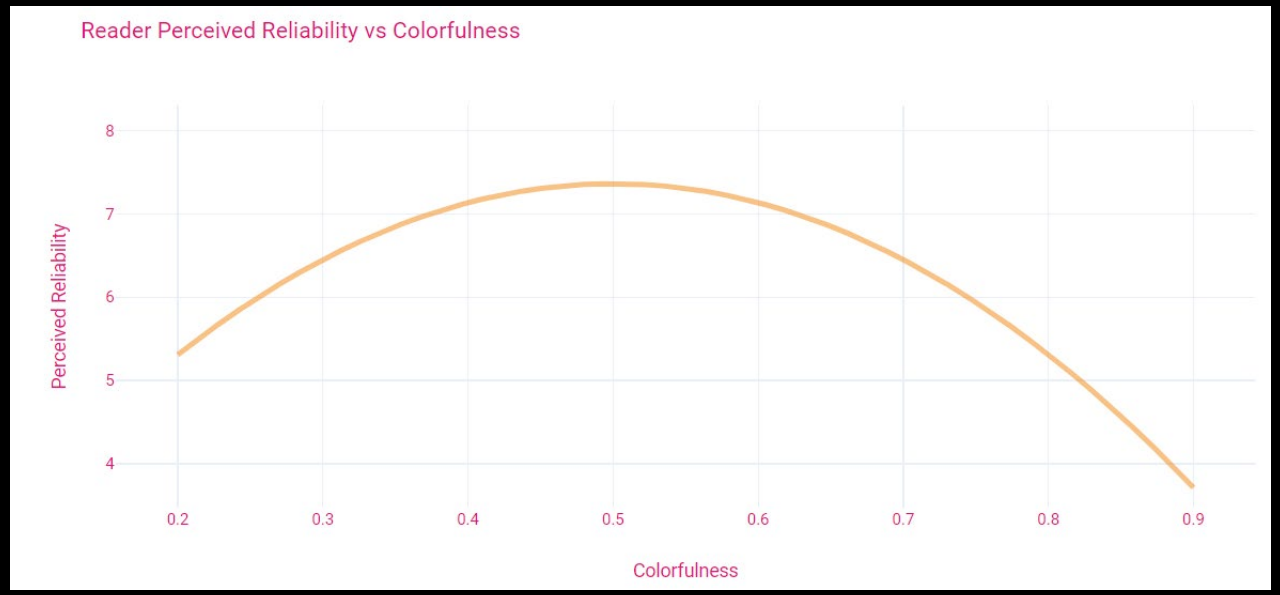
- On a scale of 1-9, how would you **rate the design** of the webpage?
- On a scale of 1-9, how **colorful** was the webpage?
- On a scale of 1-9, how **tidy** was the webpage layout?
- On a scale of 1-9, how **difficult** was the article to read?
- On a scale of 1-9, how **reliable** do you think was the information in the article you just read?
- On a scale of 1-9, how much did you **enjoy reading** the article?

* Participants' 'linger time' were also recorded.

Results







NS relationship between linger time and readability. Possible hypothesis: If it is easy to read, then it will take less time. If it is difficult to read, participants may give up on digesting the obscure portions?