Massive Online Experiments

Joshua K. Hartshorne
Boston College
Cognitive Science Bottleneck

Brick-and-Mortar labs

Pros
- Control over environment
- Specialized equipment
- 100 yrs institutional knowledge

Cons
- Restricted populations
- Low power
- Constrained by lab
  - Num. experimenters
  - Num. subjects that fit
  - Num. rooms, computers, etc.
  - Usually one-off, 30-60 min. intervals
Cognitive Science Bottleneck

Brick-and-Mortar labs

Pros
- Control over environment
- Specialized equipment
- 100 yrs institutional knowledge

Cons
- Restricted populations
- Low power
- Constrained by lab
  - Num. experimenters
  - Num. subjects that fit
  - Num. rooms, computers, etc.
  - Usually one-off, 30-60 min. intervals
Cognitive Science Bottleneck

Brick-and-Mortar labs

Pros

- Control over environment
- Specialized equipment
- 100 yrs institutional knowledge

Cons

- Restricted populations
- Low power
- Constrained by lab
  - Num. experimenters
  - Num. subjects that fit
  - Num. rooms, computers, etc.
  - Usually one-off, 30-60 min. intervals
The Dream: Massive Online Experiments

- Many subjects, world-wide
- Many items
- Many contexts/manipulations
- Longitudinal
- Social networks
- "A whole literature in one experiment"
Lifespan Development

(3,224 subjects)  

(Hartshorne & Pinker, in prep)
Geography of Personality

(Bleidorn et al., 2016)
Music popularity vs. quality

(Salganik, Dodds, & Watts, 2006)
Demographics & Visual Preferences

Preferred by under-20s

(a) under 20=7.2, over 51=5.4
(b) under 20=3.0, over 51=4.8
(c) under 20=7.1, over 51=5.7
(d) under 20=3.5, over 51=5.0

Preferred by over-51s

Figure 3. Examples websites with some of the largest differences in mean ratings of appeal between two age groups. Websites preferred by under 20 year olds on the left and those preferred by participants over 50 years of age on the right. All standard errors  0.05.

According to our results, website designs that appeal to most have a medium to high colorfulness, but a low to medium visual complexity. A good example for this is the website shown in Figure 1(d) with it’s complexity level of 3.3, and higher colorfulness of 5.7. Seeing that saturation has a significant influence on the overall perceived colorfulness [25], the finding reaffirms that of Palmer and Schloss [23], who found that (Western) adults prefer colors of higher saturation.

Results on the Influence of Demographics

Our model suggests that preferences are simultaneously influenced by multiple aspects of our demographic backgrounds. In the following, we will attempt to disentangle these effects and point out specific trends within demographic subgroups in the order of importance they play in the model.

Age

Colorfulness significantly interacts with age (F(1, 198) = 1105, p < .001). Calculating the peak appeal per age group, we found that participants aged 31 to 40 years prefer a slightly lower colorfulness than others (peak appeal = 5.6, SE = 0.04, see also Figure 2(a)). Participants under 20 and those over 51 years of age gave highest ratings for websites with a colorfulness level of 6.5 (SE < 0.03). While these peak preferences for a medium to high colorfulness level only slightly differ between age groups (all means of peak appeal between 5.6 and 6.5), older participants find plain, colorless websites less visually appealing than any other age group (Cohen’s d between low colorfulness and peak appeal = 1.8 vs. 0.4–1.1 for other age groups), and are less negatively effected by a high colorfulness (cf. Figure 2(a)).

Participants’ age also significantly affected their preference for certain levels of visual complexity (F(1) = 1721.1, p < .001): The older someone is, the more complex they prefer websites to be. The difference in appeal between optimal and suboptimal complexity levels within age groups is large, suggesting that people are more negatively affected by suboptimal complexity levels than by suboptimal colorfulness levels. Participants between 12 and 40 do not strongly differ in their preference for a moderate complexity (peaks for the three different age groups between 4.1 and 4.2, SE = 0.1). However, as participants get older, the peak appeal occurs at an increasingly higher visual complexity: For the 41-50 year olds at 4.5 (SE = 0.1), and for the over 50 year olds at 4.7 (SE = 0.01). In other words, participants over 41 liked websites with a higher complexity than under 40 year olds. This is different from the results of [25] who reported that participants older than 45 years preferred a low visual complexity more than other age groups. We attribute the difference to

(Reinecke & Gajos, 2014)
Crowdsourcing Linguistic Judgments

Progress (old):

1,247 verbs
7 semantic features
~10,000 volunteers
~450,000 judgments
The Dream: Massive Online Experiments

- Many subjects, world-wide
- Many items
- Many contexts/manipulations
- Longitudinal
- Social networks
- “A whole literature in one experiment”

So are we living the dream?
In the study reported in this paper, EMS [Electronic Mail Service] was clearly the preferred data-collection method because it produced adequate data, response rates, and willingness for further participation, with little expenditure of researcher time or effort and a high degree of convenience for respondents – Sproull (1986), p. 167

27% of APA journals have published online studies.
In the study reported in this paper, EMS [Electronic Mail Service] was clearly the preferred data-collection method because it produced adequate data, response rates, and willingness for further participation, with little expenditure of researcher time or effort and a high degree of convenience for respondents – Sproull (1986), p. 167.

27% of APA journals have published online studies.
In the study reported in this paper, EMS [Electronic Mail Service] was clearly the preferred data-collection method because it produced adequate data, response rates, and willingness for further participation, with little expenditure of researcher time or effort and a high degree of convenience for respondents – Sproull (1986), p. 167

27% of APA journals have published online studies

BBC Lab UK

faceresearch.org

gameswithwords.org

testmybrain.org

webex list

Moral Sense Test

ProjectImplicit.com

WWW established

1st online psych study

outofservice.com


In the study reported in this paper, EMS [Electronic Mail Service] was clearly the preferred data-collection method because it produced adequate data, response rates, and willingness for further participation, with little expenditure of researcher time or effort and a high degree of convenience for respondents – Sproull (1986), p. 167
Cognitive Science Bottleneck

<table>
<thead>
<tr>
<th>Brick-and-Mortar labs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pros</strong></td>
</tr>
<tr>
<td>• Control over environment</td>
</tr>
<tr>
<td>• Specialized equipment</td>
</tr>
<tr>
<td>• 100 yrs institutional knowledge</td>
</tr>
<tr>
<td><strong>Cons</strong></td>
</tr>
<tr>
<td>• Restricted populations</td>
</tr>
<tr>
<td>• Low power</td>
</tr>
<tr>
<td>• Constrained by lab</td>
</tr>
<tr>
<td>• Num. experimenters</td>
</tr>
<tr>
<td>• Num. subjects that fit</td>
</tr>
<tr>
<td>• Num. rooms, computers, etc.</td>
</tr>
<tr>
<td>• Usually one-off, 30-60 min. intervals</td>
</tr>
</tbody>
</table>
Barriers to Use

• Paradigm shift
  • “I don’t need that many subjects”
  • “My studies take more than 10 minutes!”
• Recruitment: I can’t pay 1,000,000 subjects
• What studies aren’t we running?

• Software
  • Robust scaling
  • Dynamic experiment design
    • Optimal Experimental Design
    • Active Learning
    • “Pipeline” experiments
• Tracking repeat subjects / longitudinal data
• Recruitment
  • Feedback, social media integration, mailing lists
  • Citizen Science
  • Forum, badges, etc.

• Analysis
Pushkin
robust tools for massive online experiments

• **Completed**
  • Stimuli: Text, video, audio, images
  • Responses: Keyboard, RTs (within-subject), drag-and-drop
  • Mobile-friendly webpages
  • Stub website
  • Auto-scaling (mostly)
  • (Limited) dynamic stimulus selection

• **Available soon**
  • “Endless” quizzes
  • Support for robust dynamic stimulus section
  • Interactive forum
  • Badges & leaderboards
  • Social media *authentication*
  • Profile page
  • Simplified social media *integration*
  • Eyetracking / preferential looking

• **On deck**
  • OED & Active Learning with WebPPL
  • Experiment templates
  • “Pipeline” experiments
  • [Insert your ideas here]