Whose Data Traces, Whose Voices? Inequality in Online Participation and Why it Matters for Recommender Systems Research

Eszter Hargittai
Professor and Chair of Internet Use and Society
Institute of Communication and Media Research
University of Zurich

www.eszter.com

@eszter
FIELDS ARRANGED BY PURITY

MORE PURE

Sociology is just applied psychology
Psychology is just applied biology.
Biology is just applied chemistry
Which is just applied physics.
It’s nice to be on top.
Oh, hey, I didn’t see you guys all the way over there.

Sociologists  Psychologists  Biologists  Chemists  Physicists  Mathematicians

Me

xkcd.com/435
FIELDS ARRANGED BY PRIORITY

MORE PURE

Sociology is just applied psychology.
Psychology is just applied biology.
Biology is just applied chemistry.
Which is just applied... physics.
It's nice to be on top.
Oh, hey, I didn't see you guys all the way over there.

MAKE POOR ASSUMPTIONS ABOUT YOUR DATA AT YOUR PERIL!

Me
Skills
Who is most likely to benefit from their digital media uses?

Eszter Hargittai
Social mobility vs social reproduction?
Why should you care?
Is bigger always better?

You're so big
Digital divide

Mere connectivity ≠ Effective, efficient uses
Overall Framework

Context
- Socio-Economic Status
- Academic Achievement
- Political Participation

Types of Uses
- Creative Output
- General Well-Being
- Health
- Job Performance & Productivity

Skills

Eszter Hargittai
Internet skills

- Awareness of what is possible
- Effective ways of communicating with others
- The ability to find and evaluate information
- Knowledge about seeking assistance
- The ability to create and share content
- Understanding and managing privacy
- Knowing how to stay safe and secure
- Managing one’s online reputation
- Handling information overload
- Understanding how algorithms influence what people see

Data sources for studying Internet skills?

Surveys

In-person observations and interviews
Observations of search and information seeking

Measures of actual skill → survey instruments

Eszter Hargittai
How familiar are you with the following computer and Internet-related items? Please choose a number between 1 and 5 where 1 represents “no understanding” and 5 represents “full understanding” of the item.

<table>
<thead>
<tr>
<th>Item</th>
<th>None</th>
<th>Little</th>
<th>Some</th>
<th>Good</th>
<th>Full</th>
</tr>
</thead>
<tbody>
<tr>
<td>JPEG</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Frames</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Preference settings</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Newsgroups</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>PDF</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Refresh/Reload</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Advanced search</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Blog</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>
Attention verification

The purpose of this question is to assess your attentiveness to question wording. For this question please mark the "Very often" response.

Never
Rarely
Sometimes
Often
Very Often
The purpose of this question is to assess your attentiveness to question wording. For this question please mark the “Very often” response.

- Never
- Rarely
- Sometimes
- Often
- Very Often
Are certain types of people more skilled than others?

Eszter Hargittai
Generational myths

A) All young people are digitally-savvy

B) Young people are savvier than older people

Eszter Hargittai
Data about young adults

• First-year students enrolled in the one required course
• Pencil and paper

Wave 1 – in class
  ➢ Spring, 2009
  ➢ 1,115 first-year students
  ➢ 80.5% response rate

Wave 3 – via postal mail
  ➢ Summer, 2016
  ➢ 387
  ➢ 70% response rate
## Sample descriptives

<table>
<thead>
<tr>
<th>Category</th>
<th>%</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Women</td>
<td>60</td>
<td>227</td>
</tr>
<tr>
<td>Men</td>
<td>40</td>
<td>154</td>
</tr>
<tr>
<td>Parental education</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school or less</td>
<td>25</td>
<td>95</td>
</tr>
<tr>
<td>Some college</td>
<td>26</td>
<td>98</td>
</tr>
<tr>
<td>College degree or more</td>
<td>49</td>
<td>188</td>
</tr>
<tr>
<td>Internet experiences</td>
<td>Mean (St.D.)</td>
<td></td>
</tr>
<tr>
<td>Years of use (in 2009)</td>
<td>5.6 (2.2)</td>
<td>382</td>
</tr>
<tr>
<td>Number of access locations</td>
<td>7.8 (1.7)</td>
<td>385</td>
</tr>
<tr>
<td>Weekly Web hours</td>
<td>21.4 (10.5)</td>
<td>385</td>
</tr>
<tr>
<td>Internet skills</td>
<td>3.6 (0.8)</td>
<td>385</td>
</tr>
</tbody>
</table>
If all young people were digitally savvy...
In reality: Internet skills of young adults
Understanding URLs

Which of the following sites is most likely to be the Web site of a bank called “Bankomat”? 

- www3.da-us.bankomat.com/cgi-bin/citifili/portal/I/I.do 11%
- www.bankomat.ve/rify.com 35%
- www.us.bankomat.businessportal.ru/bankomat/index.php 50%
- www.krezmin.nu/bankomat.com 4%
Generational myths

A) All young people are digitally-savvy

B) Young people are savvier than older people

n=1,512

Eszter Hargittai
### Sample descriptives

<table>
<thead>
<tr>
<th>Category</th>
<th>%</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td></td>
<td>48.9</td>
</tr>
<tr>
<td>Female</td>
<td>50.7</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>71.6</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>11.5</td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>11.6</td>
<td></td>
</tr>
<tr>
<td>Asian American</td>
<td>2.9</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>Income</td>
<td></td>
<td>$72,033</td>
</tr>
<tr>
<td>Education, HS or less</td>
<td>25.2</td>
<td></td>
</tr>
<tr>
<td>Education, some college</td>
<td>31.5</td>
<td></td>
</tr>
<tr>
<td>Education, college or more</td>
<td>43.3</td>
<td></td>
</tr>
<tr>
<td>Employed</td>
<td>62.6</td>
<td></td>
</tr>
<tr>
<td>Lives w/partner/spouse</td>
<td>61.2</td>
<td></td>
</tr>
<tr>
<td>Rural</td>
<td>13.4</td>
<td></td>
</tr>
</tbody>
</table>
If young people were generally more skilled than older people...

Made-up data!
The relationship of age and Internet skills

Eszter Hargittai
The relationship of age and Internet skills (age=<40)

Eszter Hargittai
The relationship of age and Internet skills
(age=<40)
Generational myths

A) All young people are digitally-savvy

B) Young people are savvier than older people
Internet skills by socioeconomic status

Eszter Hargittai
Use of social network sites, young adults

<table>
<thead>
<tr>
<th>Platform</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facebook</td>
<td>85%</td>
</tr>
<tr>
<td>Snapchat</td>
<td>73%</td>
</tr>
<tr>
<td>Instagram</td>
<td>70%</td>
</tr>
<tr>
<td>Pinterest</td>
<td>58%</td>
</tr>
<tr>
<td>LinkedIn</td>
<td>57%</td>
</tr>
<tr>
<td>Twitter</td>
<td>42%</td>
</tr>
<tr>
<td>Reddit</td>
<td>42%</td>
</tr>
</tbody>
</table>
Use of social network sites by socioeconomic status, young adults
Use of social network sites, US adults

- Facebook: 80
- Pinterest: 42
- Instagram: 34
- LinkedIn: 33
- Twitter: 27
- Snapchat: 20
- Reddit: 12
Use of social network sites, US adults

Eszter Hargittai
Use of social network sites by Internet skills, US adults

- Facebook
- Pinterest
- Instagram
- LinkedIn
- Twitter
- Snapchat
- Reddit

Eszter Hargittai
Amazon Mechanical Turkers vs AmeriSpeak sample

<table>
<thead>
<tr>
<th></th>
<th>NORC</th>
<th>AMT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%</td>
<td>Mean</td>
</tr>
<tr>
<td><strong>Background</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (18-94) ***</td>
<td>48.7</td>
<td>16.9</td>
</tr>
<tr>
<td>Income in U.S. $1,000s (2.5-225) ***</td>
<td>71.5</td>
<td>54.4</td>
</tr>
<tr>
<td>Female</td>
<td>51</td>
<td>1512</td>
</tr>
<tr>
<td>Employed</td>
<td>62</td>
<td>1512</td>
</tr>
<tr>
<td>Rural resident **</td>
<td>13</td>
<td>1512</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school or less ***</td>
<td>26</td>
<td>1512</td>
</tr>
<tr>
<td>Some college</td>
<td>32</td>
<td>1512</td>
</tr>
<tr>
<td>Bachelor's or higher ***</td>
<td>43</td>
<td>1512</td>
</tr>
<tr>
<td><strong>Race &amp; Ethnicity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>71</td>
<td>1511</td>
</tr>
<tr>
<td>Hispanic **</td>
<td>12</td>
<td>1511</td>
</tr>
<tr>
<td>Black *</td>
<td>11</td>
<td>1511</td>
</tr>
<tr>
<td>Asian ***</td>
<td>3</td>
<td>1511</td>
</tr>
<tr>
<td>Native American</td>
<td>2</td>
<td>1511</td>
</tr>
<tr>
<td>Other **</td>
<td>1</td>
<td>1511</td>
</tr>
</tbody>
</table>

*p<.05, **p<.01, ***p<.001
Amazon Mechanical Turkers vs AmeriSpeak sample

- Weekly Web hours
- Internet skills
- Online participation
- Social media use

Collaboration with Aaron Shaw Northwestern U.
Is Bigger Always Better? Potential Biases of Big Data Derived from Social Network Sites

By ESZTER HARGITTAI

Potential Biases in Big Data: Omitted Voices on Social Media

Eszter Hargittai
Collaboration with Aaron Shaw
Northwestern U.

Eszter Hargittai
One of the most popular Web sites
Popular for numerous topics
Largest free source of information
Billions of monthly page views
Millions of volunteer labor hours

Eszter Hargittai
There is a gender gap

Women make up an estimated:

- 16% of global WP editors
- 23% of US adult WP editors

Eszter Hargittai
Most existing work has...

Looked for answers *within* the community analyzing

- Community culture
- Dynamics of contribution

Selection on the dependent variable!

Eszter Hargittai
Looking at the Wikipedia gender gap by including data on:

- Non-editors
- User skills

Eszter Hargittai
The pipeline of online participation

Internet users
Has heard of site
Has visited site
Knows it's possible to contribute to site
Has contributed

Eszter Hargittai
Wikipedia contributions


Eszter Hargittai
97%
Predicted probability of having heard of Wikipedia

Eszter Hargittai
83%
Predicted probability of having visited Wikipedia

Internet Skills Score

Eszter Hargittai
18. Wikipedia is a site that:
(Select all that apply.)
- Is written by professional encyclopedia writers
- Can be edited by you
- Is organized by certified experts
- None of the above
Predicted probability of knowing Wikipedia can be edited
22. Have you ever edited a Wikipedia page by fixing a mistake or adding new material?

☐ No
☐ Yes
Predicted probability of contributing to Wikipedia

Internet Skills Score

Eszter Hargittai
- Education, Internet skills, & age → all outcomes.
- Income and employment status → earlier steps.
- Internet experiences → intermediate steps.
- Gender and race/ethnicity → later steps.
Takeaways

More privileged tend to be represented on platforms more than less privileged

Internet skills vary considerably and are important correlates of use

Need to know what biases are present in data sets we use
The Skilled Among You.

Yes, you.
You control the Information Age.
Welcome to your world.
Thanks to funders:

- MacArthur Foundation
- Google Research
- Northwestern School of Communication
- National Science Foundation
- Facebook
- Nokia Research Center
- Merck
- MSD
- Alfred P. Sloan Foundation
- Microsoft Research
- University of Zurich

Thanks to students/former students & staff:

[Images of people and logos]
Thank you!
Questions?
Comments?

@eszter

webuse.org

Eszter Hargittai