CSE 291/DSC 291
Information Manipulation: Trustworthiness of Information in Cyberspace

Molly Roberts & Stefan Savage, Winter 2023

Social Media, Algorithms, and Misinformation
Zeynep Tufekci

- YouTube the great radicalizer
- “Videos about vegetarianism led to videos about veganism. Videos about jogging led to videos about running ultramarathons.”
- What is Tufekci’s explanation for why this happens?
- Overall, this explanation has been put forward by a lot of media: https://www.youtube.com/watch?v=VvvfktYlj1w
- It seems like common sense?
What does the data show us?

- Today: three studies on this
What is an audit study?

- Traditional audit studies typically measure discrimination.
- Discrimination in housing market, labor market, by government agencies, etc.
- E.g. randomly submit resumes with names associated with two different ethnicities, assess calls for interviews

SEX DISCRIMINATION IN RESTAURANT HIRING: AN AUDIT STUDY*

DAVID NEUMARK
WITH THE ASSISTANCE OF ROY J. BANK AND KYLE D. VAN NORT

In an audit study of sex discrimination in hiring, comparably matched pairs of men and women applied for jobs as waiters and waitresses at restaurants in Philadelphia. In high-price restaurants (where earnings are higher), job applications from women had an estimated probability of receiving a job offer that was lower by about 0.4, and an estimated probability of receiving an interview that was lower by about 0.36. Both estimated differentials are statistically significant. Additional evidence suggests that customer discrimination partly underlies the hiring discrimination.

I. INTRODUCTION

The overall sex gap in wages can be broken into an across-occupation and a within-occupation component. The sex gap related to across-occupation segregation has alternatively been attributed to human capital investment [Polachek 1981], preferences [Daymont and Andrisani 1984], and employment dis-
What is an audit study?

- Algorithmic audits: act like an audit study, but the thing being audited is an algorithm.

- Audits of:
  - Pricing
  - Online advertising
  - Facial/speech recognition
  - Search engines, recommendation algorithms
  - Admissions, hiring, etc
Hussien et al.

- Research question: What is the effect of personalization on the amount of misinformation presented to users on YouTube?
  - Demographics: age, gender, geo-location
  - Watch history

- Topics: 9/11 conspiracy theories, chemtrail conspiracy theory, flat earth, moon landing conspiracy theories and vaccine controversies

- Outcomes:
  - Search results for new accounts
  - Search results, Up-Next, Top-5 for accounts with watch history
  - For each video: classified into promoting, debunking, or neutral toward the conspiracy
## Experimental Conditions

<table>
<thead>
<tr>
<th>Experiment #</th>
<th>Category</th>
<th>Feature</th>
<th>Tested Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search (Exp 1)</td>
<td>Demographics</td>
<td>Age</td>
<td>&lt;18, 18-34, 35-50, &gt;50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Gender</td>
<td>Male, Female</td>
</tr>
<tr>
<td>Search (Exp 2)</td>
<td>Geolocation</td>
<td>IP Address</td>
<td>Georgia, Montana, New Jersey, Ohio, South Carolina</td>
</tr>
<tr>
<td>Watch (Exp 3)</td>
<td>Demographics</td>
<td>Age</td>
<td>&lt;18, 18-34, 35-50, &gt;50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Gender</td>
<td>Male, Female</td>
</tr>
<tr>
<td>Watch (Exp 4)</td>
<td>Geolocation</td>
<td>IP Address</td>
<td>Georgia, Montana, New Jersey, Ohio, South Carolina</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Watch history</td>
<td>Promoting, Neutral, Debunking</td>
</tr>
</tbody>
</table>

**Table 2.** List of user features for our audit experiments.
Search Experiment

**Fig. 4.** Steps performed in *Search* experiments 1 and 2.
# Search queries

<table>
<thead>
<tr>
<th>Search Topic</th>
<th>Seed Query</th>
<th>Hot</th>
<th>Cold</th>
<th>Sample Search Query</th>
</tr>
</thead>
<tbody>
<tr>
<td>9/11 conspiracy theories</td>
<td>9/11 and 9/11 conspiracy</td>
<td>Maryland</td>
<td>Ohio</td>
<td>9/11 inside job 9/11 tribute 9/11 conspiracy</td>
</tr>
<tr>
<td>Chemtrail conspiracy theory</td>
<td>chemtrail</td>
<td>Montana</td>
<td>New Jersey</td>
<td>chemtrail chemtrail flu chemtrail pilot</td>
</tr>
<tr>
<td>Flat Earth</td>
<td>flat earth</td>
<td>Montana</td>
<td>New Jersey</td>
<td>flat earth proof is the earth flat</td>
</tr>
<tr>
<td>Moon landing conspiracy theories</td>
<td>moon landing</td>
<td>Ohio</td>
<td>Georgia</td>
<td>moon moon hoax moon landing china</td>
</tr>
<tr>
<td>Vaccine controversies</td>
<td>vaccines</td>
<td>Montana</td>
<td>South Carolina</td>
<td>anti vaccine vaccines vaccines revealed</td>
</tr>
</tbody>
</table>

**Table 1.** Seed query, hot & cold regions, and sample search queries for the five misinformation search topics.
Watch Experiments

Fig. 5. Steps performed in Watch experiments 3 & 4. These experiments have two phases: (1) watch phase (denoted by →), (2) search phase (denoted by ←).

<table>
<thead>
<tr>
<th>Stance</th>
<th>No. of accounts (Demographics)</th>
<th>No. of accounts (Geolocation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Debunking (-1)</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>Neutral (0)</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>Promoting (1)</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td><strong>Total accounts</strong></td>
<td><strong>24</strong></td>
<td><strong>6</strong></td>
</tr>
</tbody>
</table>

Table 3. Accounts created to execute Watch experiments for each misinformative topic. In total, we created 120 (24X5) accounts to run experiment 3 and 30 (6X5) accounts for experiment 4. Here 5 denotes the number of topics.
Qualitative Coding Scheme -- >2,000 videos

<table>
<thead>
<tr>
<th>Annotation Value</th>
<th>Stance Description</th>
<th>Annotation Heuristics</th>
<th>No.of videos</th>
<th>Normalized Score</th>
<th>Sample Videos (Video Title (Video URL, youtu.be))</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>debunking, mocking, disproving related misinformation</td>
<td>narrative of video disputes, mocks or provides authoritative evidence against conspiracy theories related to the topic under audit</td>
<td>430</td>
<td>0 (D)</td>
<td>Bill Maher Throws Out 9/11 Conspiracy Theories On Live TV (p8RxXaMqQgU)</td>
</tr>
<tr>
<td>0</td>
<td>neutral &amp; related to misinformation</td>
<td>narrative of the video does not take any stance on conspiracy theories related to the topic under audit</td>
<td>258</td>
<td>0 (N)</td>
<td>The Howard Stern Show and WCBS-2 On Sept. 11 (O3LT9FMf2i8)</td>
</tr>
<tr>
<td>1</td>
<td>promoting, supporting, justifying, explaining related misinformation</td>
<td>narrative of video promotes, supports or substantiates any conspiratorial views related to the topic under audit</td>
<td>374</td>
<td>0 (P)</td>
<td>9/11 truthers attend treason in America (2-TUCe-2M2g)</td>
</tr>
<tr>
<td>2</td>
<td>debunking, mocking, disproving unrelated misinformation</td>
<td>narrative of video denies, mocks or provides evidence against a conspiratorial view related to a topic different than the one under audit</td>
<td>64</td>
<td>0 (D)</td>
<td>Did the Titanic Really Sink? The Olympic Switch Theory Debunked (_mplIRCQQB20)</td>
</tr>
<tr>
<td>3</td>
<td>neutral &amp; related to another misinformation</td>
<td>narrative of the video does not take any stance on conspiracy theories unrelated to the topic under audit</td>
<td>25</td>
<td>0 (N)</td>
<td>JFK coverage 12:30pm-1:40pm 11/22/63 (pXOo1jgg206)</td>
</tr>
<tr>
<td>4</td>
<td>promoting, supporting, justifying, explaining unrelated misinformation</td>
<td>narrative of the video promotes, supports, justifies or explains any conspiratorial view unrelated to the topic under audit</td>
<td>66</td>
<td>0 (P)</td>
<td>Mafia Boss Tells All - Jimmy Hoffa, JFK Assassination and Much More (_LwmaAfALb)</td>
</tr>
<tr>
<td>5</td>
<td>not about misinformation</td>
<td>video content does not contain any conspiratorial views</td>
<td>1667</td>
<td>0 (N)</td>
<td>Former Abortionist Dr. Levalino At Virginia Tech (d6Rcwe-o560U)</td>
</tr>
<tr>
<td>6</td>
<td>foreign language</td>
<td>video content in non-English language</td>
<td>35</td>
<td>0 (N)</td>
<td>Las voces del 11S, documental en Español del Canal Nacional Geográfico (7vMA02R_3wU)</td>
</tr>
<tr>
<td>7</td>
<td>undefined/unknown</td>
<td>annotators were unable to assign any of the above annotation values to the video</td>
<td>9</td>
<td>0 (N)</td>
<td>Ahmed Mohammed's Dad Pushes 9/11 Conspiracy Theories Online (CTKeOEknc)</td>
</tr>
<tr>
<td>8</td>
<td>removed</td>
<td>video removed from the platform at the time of annotation</td>
<td>35</td>
<td>0 (N)</td>
<td>n/a (pSOTJiOLH5e)</td>
</tr>
</tbody>
</table>

Table 4. Description of the annotation scale and heuristics along with sample YouTube videos corresponding to each annotation value. We map our 9-point annotation scale to 3-point normalized scores with values -1 (Promoting, (′′′)), 0 (Neutral, (′′)) and 1 (Debunking, (′′)). We have shared the list of 2,943 unique videos along with their annotation values in our online dataset.7
Results

• No effect of demographics only on search
• But when including watch history, then recommendations seem to differ by demographics and topics
  - For most topics, men are recommended more misinfo than women (two exceptions)
  - Geolocation and age doesn’t seem to matter much
• Watch history affects search results only in one topic (vaccines)
• Watch history affects Top-5 and Up Next videos for all topics (except for vaccines)
• Lots of variation across topics generally
Discussion

– Why do you think there is so much variation over topics?
– How does this study answer or not answer the question: “Is YouTube a Great Radicalizer?”
– How might the design be improved?
Hosseinmardi et al.

- Dataset: 309,813 people representative of the U.S. population + their online browsing histories from January 2016-December 2019
- Where did this data come from? Is this an audit study?
- Nielsen’s nationally representative desktop web panel, subset to users who have at least one YouTube view
- 9,863,964 unique video IDs
Hosseinmardi et al.

- How do they label?
- Channel ID, categorize political leaning of channel, all videos labeled with that channel
- In the end cover 974 channels, corresponding to 523,242 videos
- Imputed missing data (from deplatformed channels and videos not in the API)
Time trends

Fig. 2. Breakdown of percent of (A) users and (B) consumption falling into the six political channel categories, per month, January 2016 to December 2019. A is the percent of users falling into each community, and B presents the percentage of viewership duration from each channel category. Solid lines show the fitted linear models and the shading shows the 95% confidence intervals.
Bursts of consumption correlated with future consumption

Fig. 5. Difference in means of daily consumption change, in the event of bursty consumption from a specific political category. Individuals are assigned either to bursty consumption group in the event of watching at least $M^k_v$ videos from category $k$ ($k \in \{L, C, AW, R, fR\}$) within a session, or to a control group, if none of their sessions has more than one video from the same category with at least $M^k_v$ videos in their lifetime. We run three experiments with different values of $M^k_v$, where $\blacktriangledown: M^k_v = 2$, $\bullet: M^k_v = 3^*$, $\star: M^k_v = 4$. Markers show the difference in means, and the vertical lines present the 95% CI. The exposure can be driven by user, recommendation, or external sources. Difference in change of daily consumption, after bursty consumption, is almost twice as large for AW compared to the other political categories, when controlled for other covariates.
Is the YouTube algorithm creating this?

- Does YouTube look different than off platform content?
- How did users get to the video?
- Was political content more likely to be consumed later in the session?
Discussion

▪ What are the limitations of this study?
▪ Which study do you find more convincing?
▪ How would you design a study that built on the strengths of each of the studies we read?