Exploring Public Perception of Algorithmic Unfairness

Researchers
Allison Woodruff (Security & Privacy UX)
Jeff Warshaw (Security & Privacy UX)
Sarah Fox (Security & Privacy UX Intern)
Steven Rousso-Schindler (Visual Ethnography Contractor)

Sponsors
Sunny Consolvo (Security & Privacy UX Manager)
Lea Kissner (Director, Security & Privacy Nightwatch)
Lawrence You (Director of Privacy)

March 2017

Talk Outline

1. Background
2. Methodology
3. Findings

Imagine that a search engine company launches a maps application that directs users away from neighborhoods with high crime rates, and hides businesses located in those neighborhoods

— The Good Wife, Episode "Discovery" (November 29 2015)

algorithmic unfairness—unjust or prejudicial treatment of people that is related to sensitive characteristics such as race, income, sexual orientation, or gender, through algorithmic systems or algorithmically aided decision-making

Source: A. Woodruff & A. Schou. Definition of Algorithmic Unfairness. March 2017.
go/algorithmic-unfairness-definition

Mounting External Public & Regulatory Pressure

Google Photos labeled black people 'gorillas'

Jessica Guynn , USA TODAY 2:10 p.m. EDT July 1, 2015

theguardian

home › tech US politics world opinion sports soccer arts lifestyle fa al

Google

Women less likely to be shown ads for high-paid jobs on Google, study shows

US THE HUFFINGTON POST

WS POLITICS ENTERTAINMENT WELLNESS WHAT'S WORKING VOICES VIDEO ALL SE

BUSINESS

Google Image Search Has A Gender Bias Problem

04/10/2015 05:20 pm ET | Updated Apr 10, 2015

Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights

Executive Office of the President

May 2016

GDPR

EU General Data Protection Regulation

Efforts Across Google

example: go/ml-fairness

- Machine Intelligence
- Trust & Safety
- Public Policy
- Security & Privacy

... and many more!

Talk Outline

1. Background
2. Methodology
3. Findings

Overview

Participatory design workshops +
1:1 interviews

44 participants (Black, Hispanic, or low SES)

July-September 2016

Participants

Workshop 1	Low SES
Workshop 2	Black women
Workshop 3	Hispanic
Workshop 4	Black
Workshop 5	Low SES

SES based on an approximation of Glasmeier's Living Wage Model (livingwage.mit.edu)

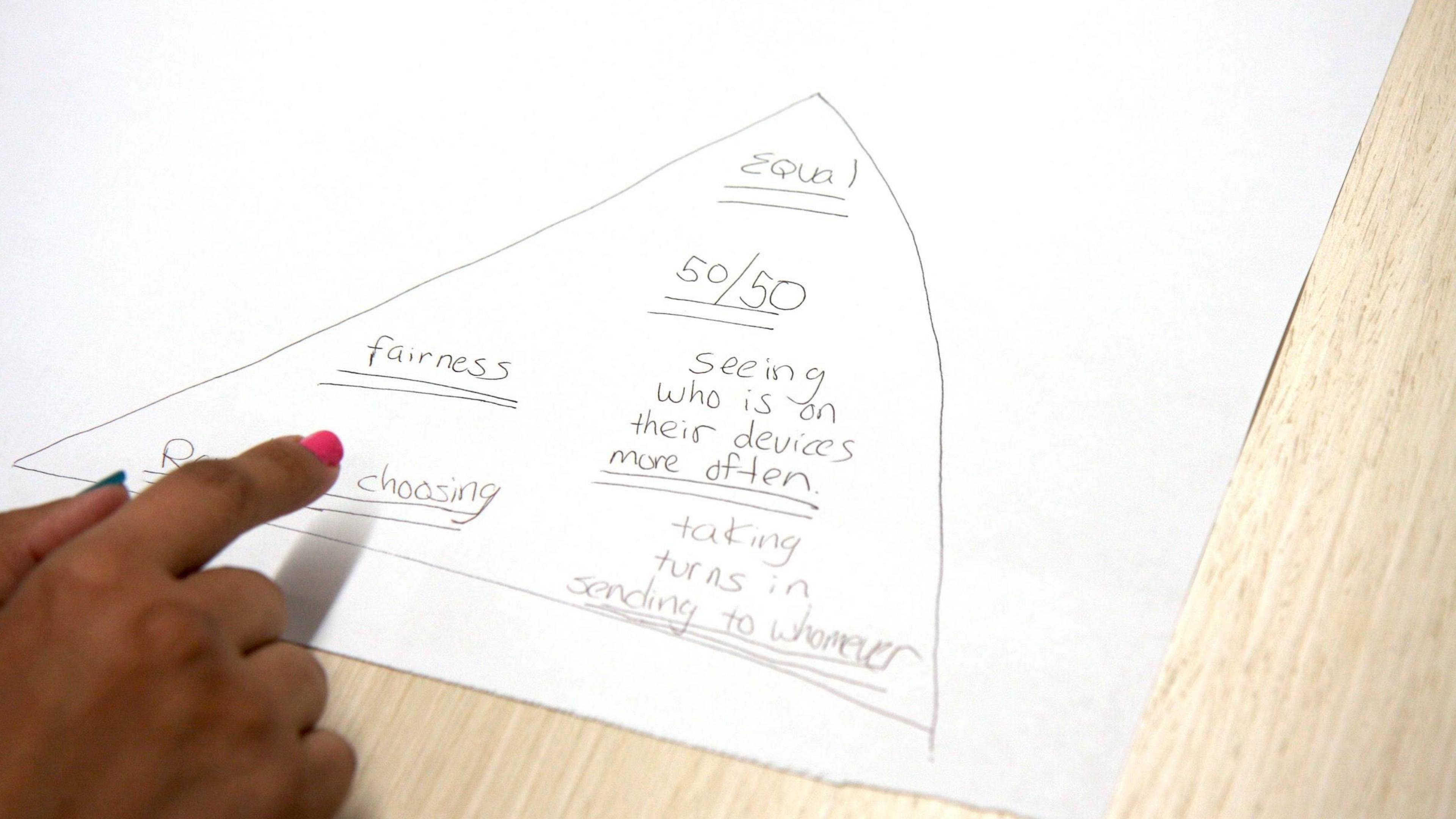
Location:
San Francisco Bay Area (East Bay, San Francisco)

Occupations:
Varied (e.g. teacher, public transportation driver, retail manager, tasker, line cook)

Ages:
18-65

Workshop Structure (5 hours)

- Ice breaker (Peggy McIntosh's "Invisible Knapsack")
- Experiences with discrimination
- Discussion of algorithmic discrimination
- Meal
- For each of three scenarios
  - Brief reactions
  - Design activity (working independently)
  - Share ideas & group discussion
- Concluding discussion



Equal
50/50
fairness
seeing who is on their devices more often.
choosing
taking turns in sending to whomever


Teacher.
Any Gender
Salary
Relating message
Starts From
Diversity in the workplace

Scenarios

Scenario 1: High-Paying Job Ads Inspired by A. Datta et al. Automated Experiments on Ad Privacy Settings. PETS 2015, pp. 92-112, June 2015.	A man visits a newspaper website and sees ads for high-paying jobs, while a woman visiting the same website sees ads for low-paying jobs
Scenario 2: Trayvon Martin Autocomplete Inspired by S.U. Noble (2014). Trayvon, Race, Media and the Politics of Spectacle. *The Black Scholar* 44(1).	Autocomplete results for Trayvon Martin are negative, but those for George Zimmerman are positive
Scenario 3: Restaurant Finder Inspired by *The Good Wife*, Episode “Discovery”. November 29 2015.	A restaurant review app doesn’t show businesses in neighborhoods with a high crime rate

Analytic Approach

- Visual ethnography
- Transcripts
- Open coding
- Affinity clustering

Talk Outline

1. Background
2. Methodology
3. Findings

Unfamiliar but not Unfathomable

Most participants were not aware of algorithmic discrimination before the study, although...

- Occasional concerns about being targeted for low-income ads
- Some participants reported turning off location history to avoid racial profiling

Learning about it often elicited strong negative feelings, and evoked broader experiences with stereotyping...

- “It’s totally unfair, because not every woman’s the same” - P33
- “For me, it’s a negative, because they didn’t even base it on what I’ve done in the past, they’re just basing it on what they think I am” - P23

...but overall, unfamiliarity meant their perspectives were still malleable. Opinions shifted during the workshops as participants discussed them

Misunderstanding Scale and Impact of Algorithmic Systems

Few participants demonstrated understanding of how pervasive, autonomous, and dynamic algorithms are in everyday life

Echoing our prior Inference Literacy work, most participants presumed algorithms are:

- Small-scale
- Calculator-like tools that help human employees make inferences and decisions
- Based on simple rules

Reference: go/inference-literacy

Misunderstanding Scale and Impact of Algorithmic Systems, cont.

Even small statistical disparities in algorithmic decisions can perpetuate or increase inequalities in different groups' life choices

- example: credit scoring (Fourcade & Healy, 2013)

Participants mostly disregarded this as a point of concern, describing small statistical inequalities as:

- Natural,
- Inevitable, and
- Impossible to fix

"It sounds fine to me…
I don't expect perfection, of course."
— P43

Failure to appreciate the scale, impact, and nature of algorithmic unfairness is a major barrier for change: advocating for an issue requires acknowledging that it not only matters, but also that it can and should change

… but failure to appreciate scale and impact of predictive systems is coupled with a deep appreciation of the importance of representation

High Salience of Representational Consequences

Participants were aware and concerned about skewed portrayals of marginalized groups

- “If you really type in ‘two black teenagers’, you will see all mugshots of black boys. But with white teenagers, you will see them playing basketball, boy scout...It was crazy.” — P29
- “I already see it when I cut on the TV and see the way people are portrayed in the media. When I get on the computer, through searching Google I shouldn’t have to be subjected to racial stereotypes.” — P11

They felt popularity algorithms are not benign mirrors of the world: they amplify societal biases and increase the reach of stereotyping messages

- “Feeding into that stuff, to me, is going backwards. Even encouraging people to read about that stuff and feeding into those thoughts, there’s no need to feed.” - P22

Accountability

Many participants held the programmer accountable for an algorithm's discrimination, even if the programmer had honest intentions

- "When you lack that diversity, they may not be able to input certain things into that equation...because they don't know that reality." — P20
- "People create the technology to do these things, so that's why I say it stems from the writer." — P29

They also often called out the role society played in creating the problem

Belief that companies could resolve the problem if they were motivated

*“**It's not really like a company being racist… it’s really just a machine, it's stats**… It’s counting numbers, it's counting what we are all looking at. It's based on what we're looking at, not what Google wants you to look at… **the problem is us, and what we have in our minds**, so we can't really turn around and be like, ‘oh, Google did it.’” — P02*

*"I think that people that work for these companies... **they can make the change tonight if they wanted to**. It's just a matter of how are they going to meticulously put everything so it will still benefit them in some aspect." — P29*

Journalistic Standards

Many said that Google has a responsibility to not knowingly present biased or opinion-based content when there are facts we could present instead

In-product information processed by algorithms can give the impression that Google endorses a message

*"I know it's the popular searches, but still... **it just seems like Google is... saying it themselves**." — P24*

*“**It looks like Google’s the one that’s putting this out**, and that’s what people would think. You know, that’s what I would think.”* — P17

Algorithmic Unfairness Can Damage User Trust

Google is viewed a trusted source for information, but inaction risks appearing to endorse others' discrimination by signal boosting it...

- "You guys are pretty much promoting this hate and promoting this deceit...that's not doing nothing but making everybody mad." - P04

...or behaving beneath what users expect of Google

- "I've used Google a lot, it's been my lifeline almost... maybe that's why I'm even more offended that this is what was suggested. It's like, come on, Google. I thought we were better than that." - P24

Acknowledgements

PUBLIC POLICY
Alex McPhillips

LEGAL
Neal Cohen
Will DeVries
Brad Krueger
Erin Simon

STUDY PROGRAM MGMT
Irene Tang

STUDY DESIGN FEEDBACK
Tara Matthews

A/V EQUIPMENT
Mark Chow
Tyler Watkins

UX INFRASTRUCTURE
Justin Shaw
Gina DeNatale
Cindy Yepez

SFO SECURITY & FACILITIES
Christine DeLeonardis
Leo Pailano

WORKSHOP TRANSCRIPTION (WITH SARAH FOX)
Rena Coen

VIDEO 1 EDITING (WITH S. ROUSSO-SCHINDLER)
Chris Sotelo

VIDEO 1 CRITIQUE
Anne Halkedis
Manya Sleeper
Anna Turner

Learn more, or get involved!

CONTACTS

woodruff@

ml-fairness-leads@

LINKS

go/algorithmic-unfairness-definition

go/allegations-of-algorithmic-bias

go/discrimination-and-stress

go/inference-literacy

go/ml-fairness