Designing Field Experiments at Scale (SOC 412)

Feb 5, 2019

Sherrerd Hall 306

J. Nathan Matias

@natematias

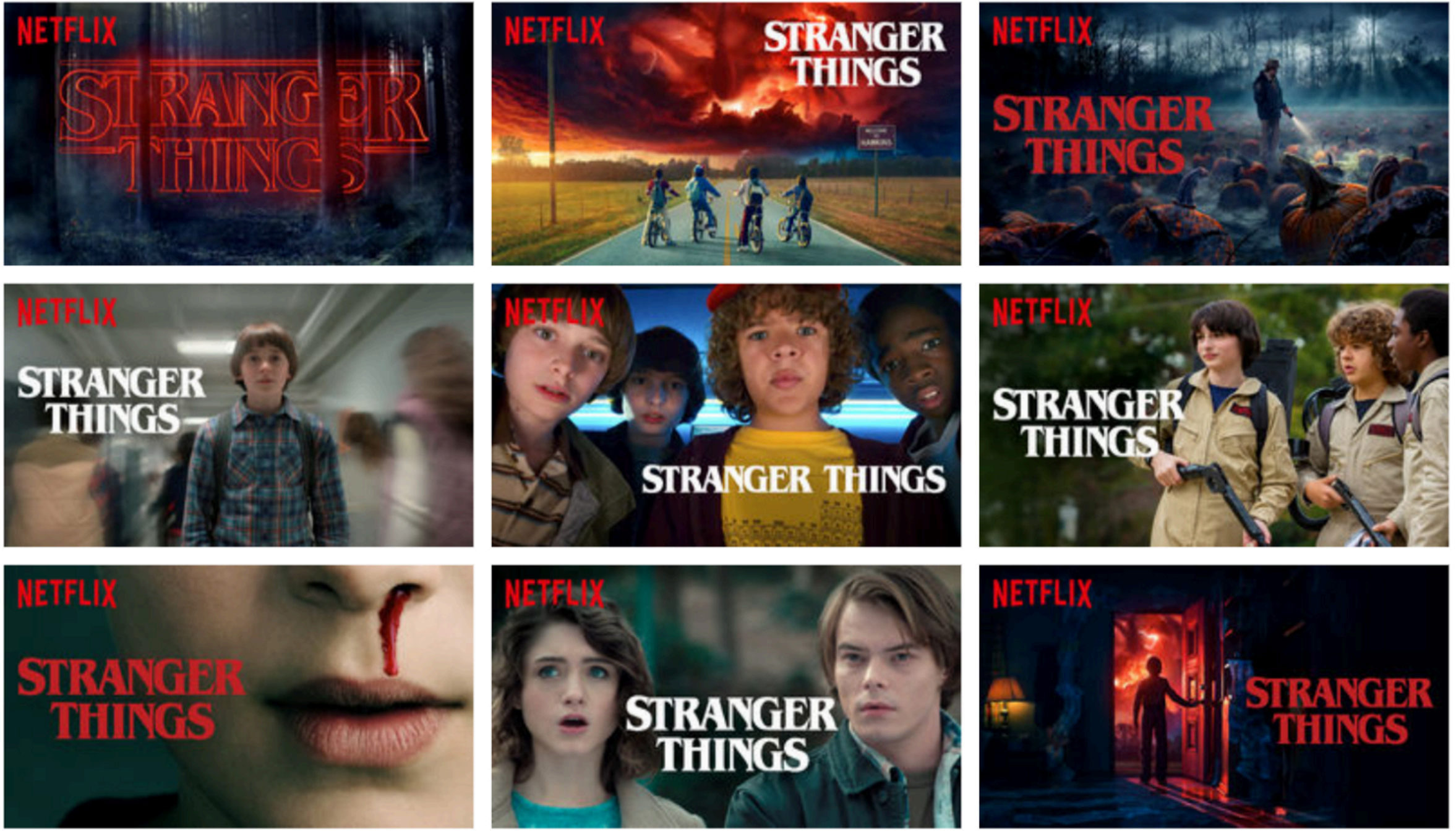
civilservant.io

jmatias@princeton.edu

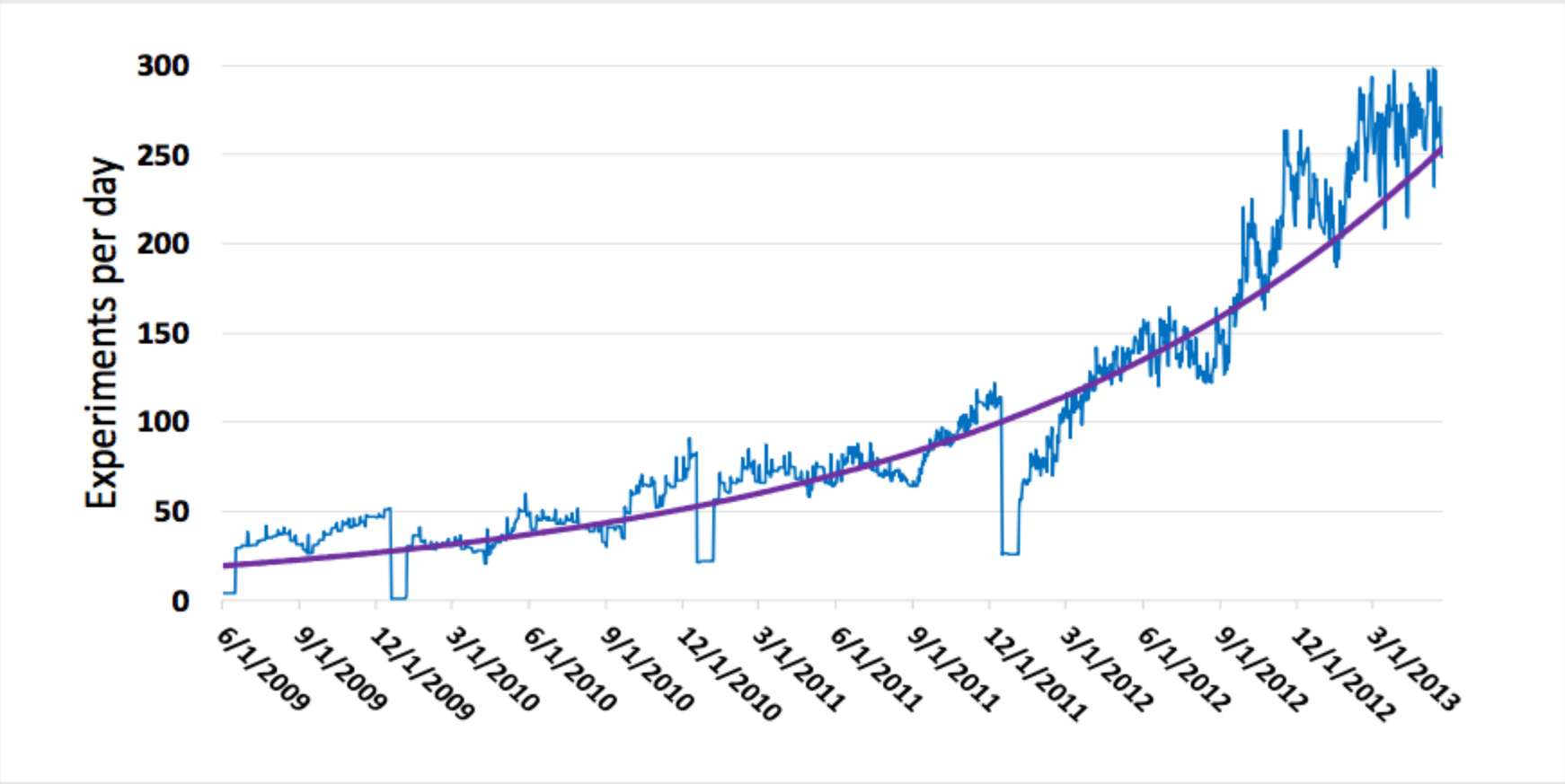
Department of Psychology
PRINCETON UNIVERSITY

CITP

mit media lab



Chandrashekar, A., Amat, F., Bailico, J., Jebara, T, Artwork Personalization at Netflix. medium.com



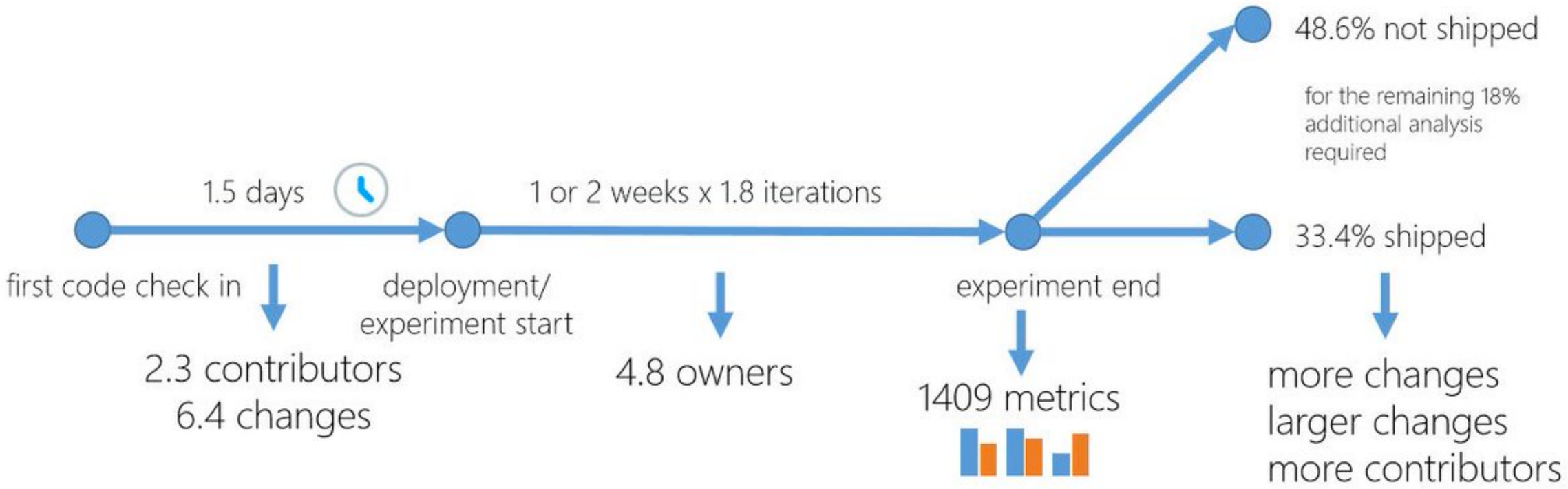
Experiments Per Day on bing.com

Kohavi, R., Deng, A., Frasca, B., Walker, T., Xu, Y., & Pohlmann, N. (2013, August). **Online controlled experiments at large scale**. In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 1168-1176). ACM.

bing

> the number of concurrent experiments running in ERF has grown from a few dozen (in 2014) to **about 500 concurrent experiments** [May 2017]
>
> Today we compute ~2500 distinct metrics per day and roughly **50k distinct experiment/ metric combinations**.

Parks, Jonathan. Scaling Airbnb's Experimentation Platform. Airbnb Engineering & Data Science. May 10, 2017.



Kevic, K., Murphy, B., Williams, L., & Beckmann, J. (2017, May). **Characterizing experimentation in continuous deployment: a case study on bing.** In Proceedings of the 39th International Conference on Software Engineering: Software Engineering in Practice Track (pp. 123-132). IEEE Press.

bing

> [in] an entire redesign of the Premium Subscription's **payment flow** …
>
> The experiment showed **an increase of millions of dollars in annualized bookings**, about 30% reduction in refund orders and over 10% lift in free trial orders.

Xu, Y., Chen, N., Fernandez, A., Sinno, O., and Bhasin, A. **From infrastructure to culture: A/b testing challenges in large scale social networks.** In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM (2015), 2227–2236.

How Uber Uses Psychological Tricks to Push Its Drivers' Buttons

The company has undertaken an extraordinary experiment in behavioral science to subtly entice an independent work force to maximize its growth.

By NOAM SCHEIBER and graphics by JON HUANG | APRIL 2, 2017

> Employing **hundreds of social scientists and data scientists**,
>
> Uber has **experimented** with video game techniques, graphics and noncash rewards of little value that can **prod drivers into working longer and harder** — and sometimes at **hours** and **locations** that are **less lucrative** for them.

Scheiber, N., Huang, J. (2017) **How Uber Uses Psychological Tricks to Push Its Drivers' Buttons**. New York Times Magazine. April 2, 2017.

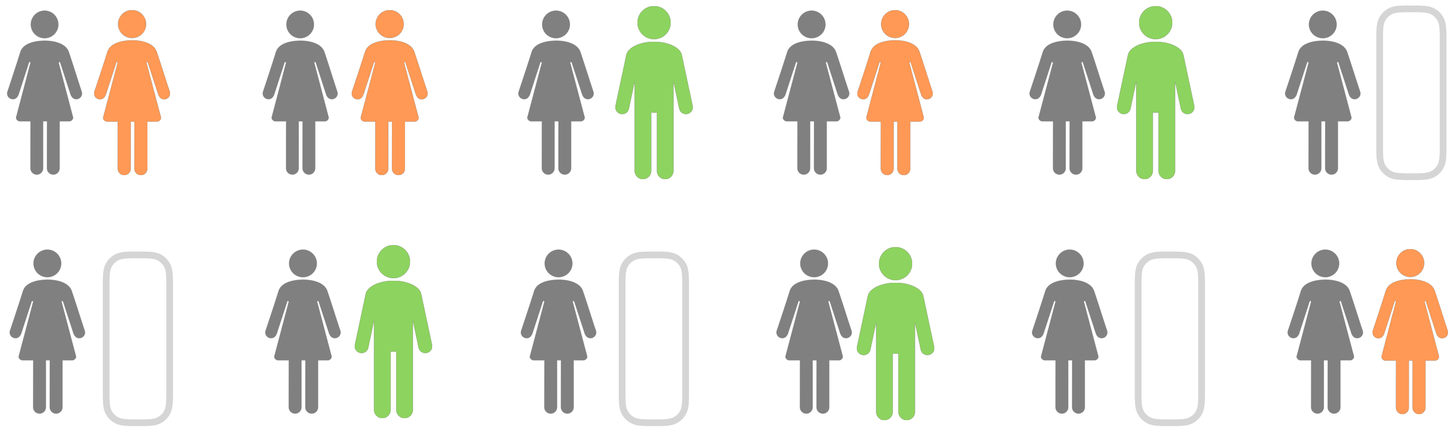
- A **20 per cent reduction in speeding** over six months
- A **34 per cent increase in acceptances** by students from under-represented schools to top universities.
- An **8% reduction in annual household gas consumption** following installation of smart heating controls
- A **38 per cent reduction in patient referrals** to over-booked hospitals

Behavioral Insights Team. The Behavioral Insights Team Update Report 2016-2017.

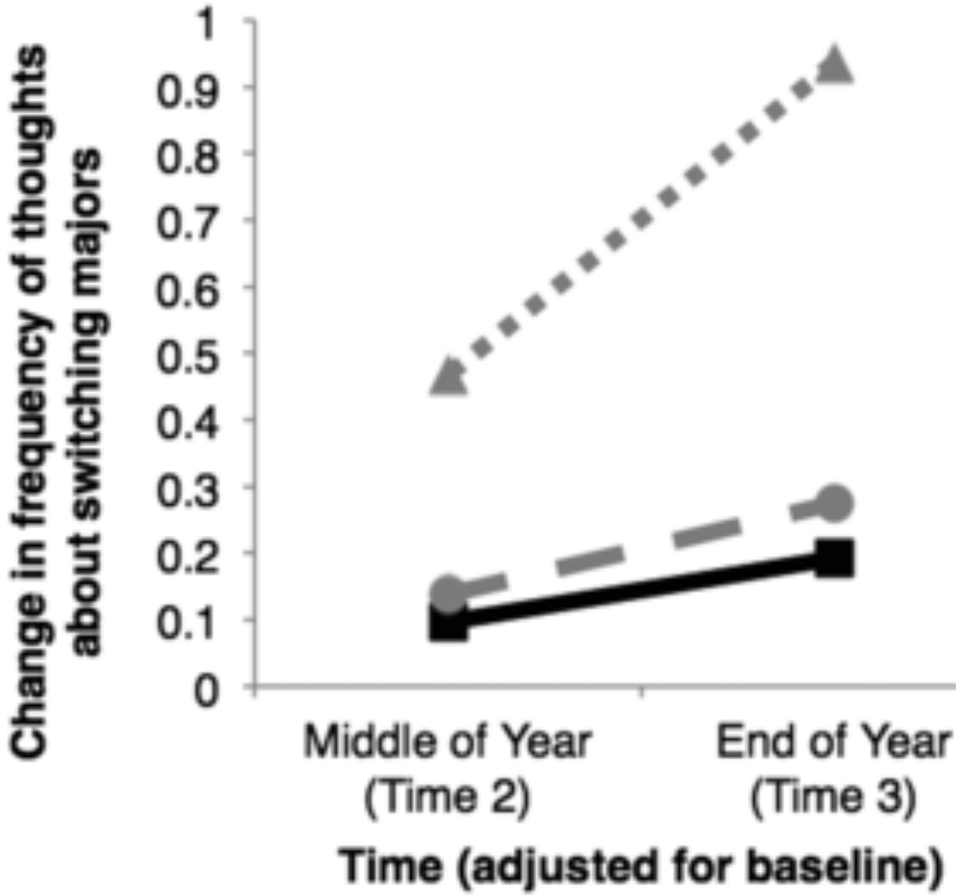
THE BEHAVIORAL INSIGHTS TEAM

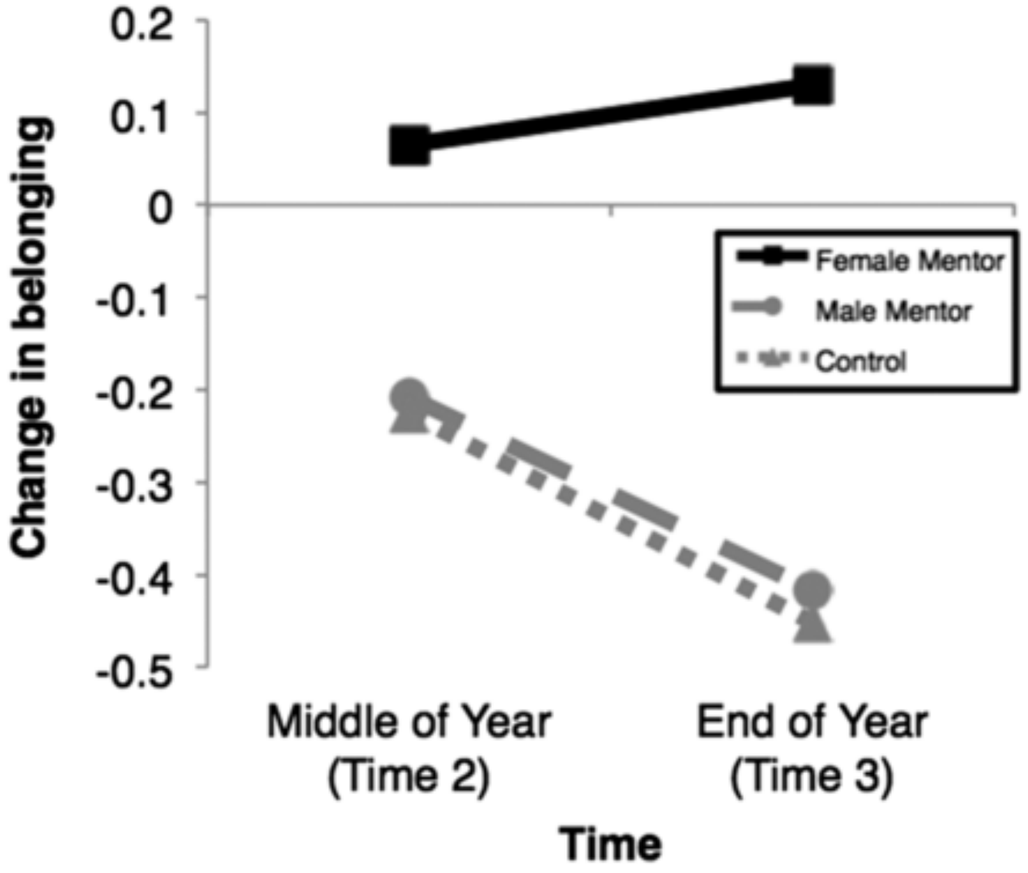
Does mentoring have any real benefits for women in engineering?

(emails to 150 incoming engineering majors)

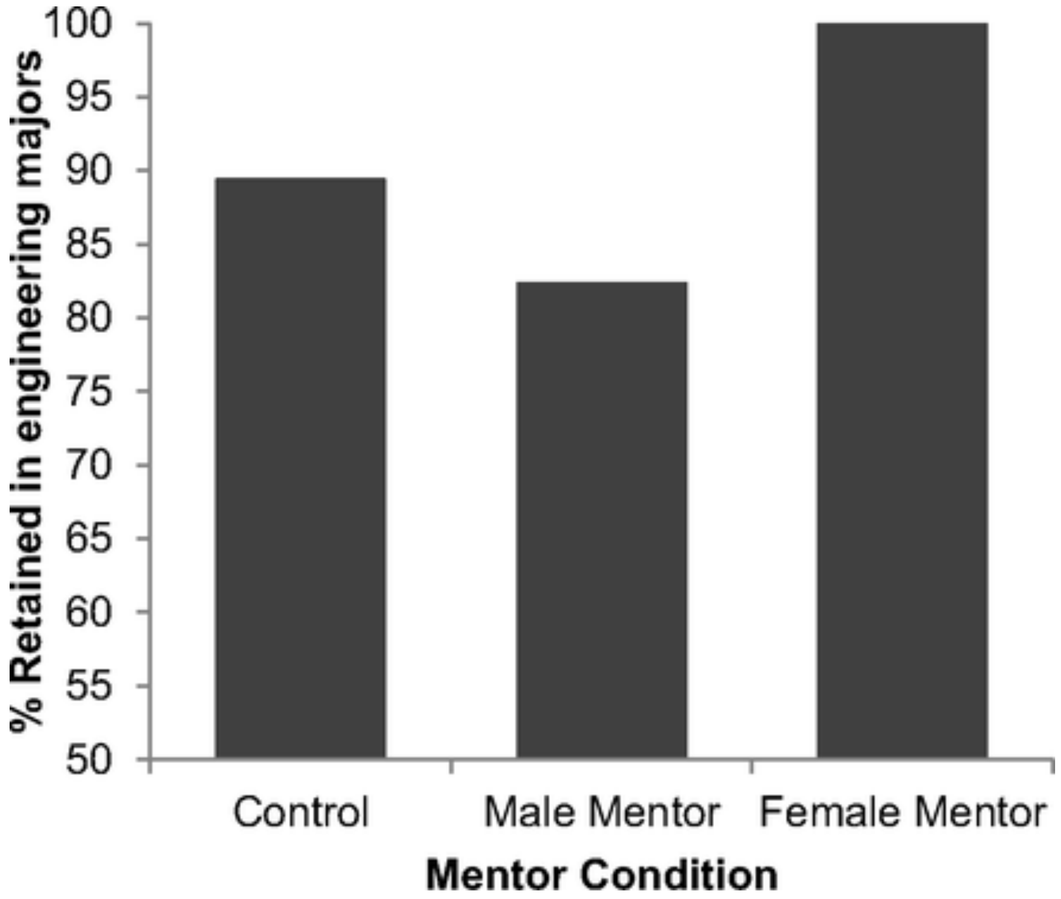


Dennehy, T. C., & Dasgupta, N. (2017). Female peer mentors early in college increase women's positive academic experiences and retention in engineering. *Proceedings of the National Academy of Sciences*, *114*(23), 5964-5969.



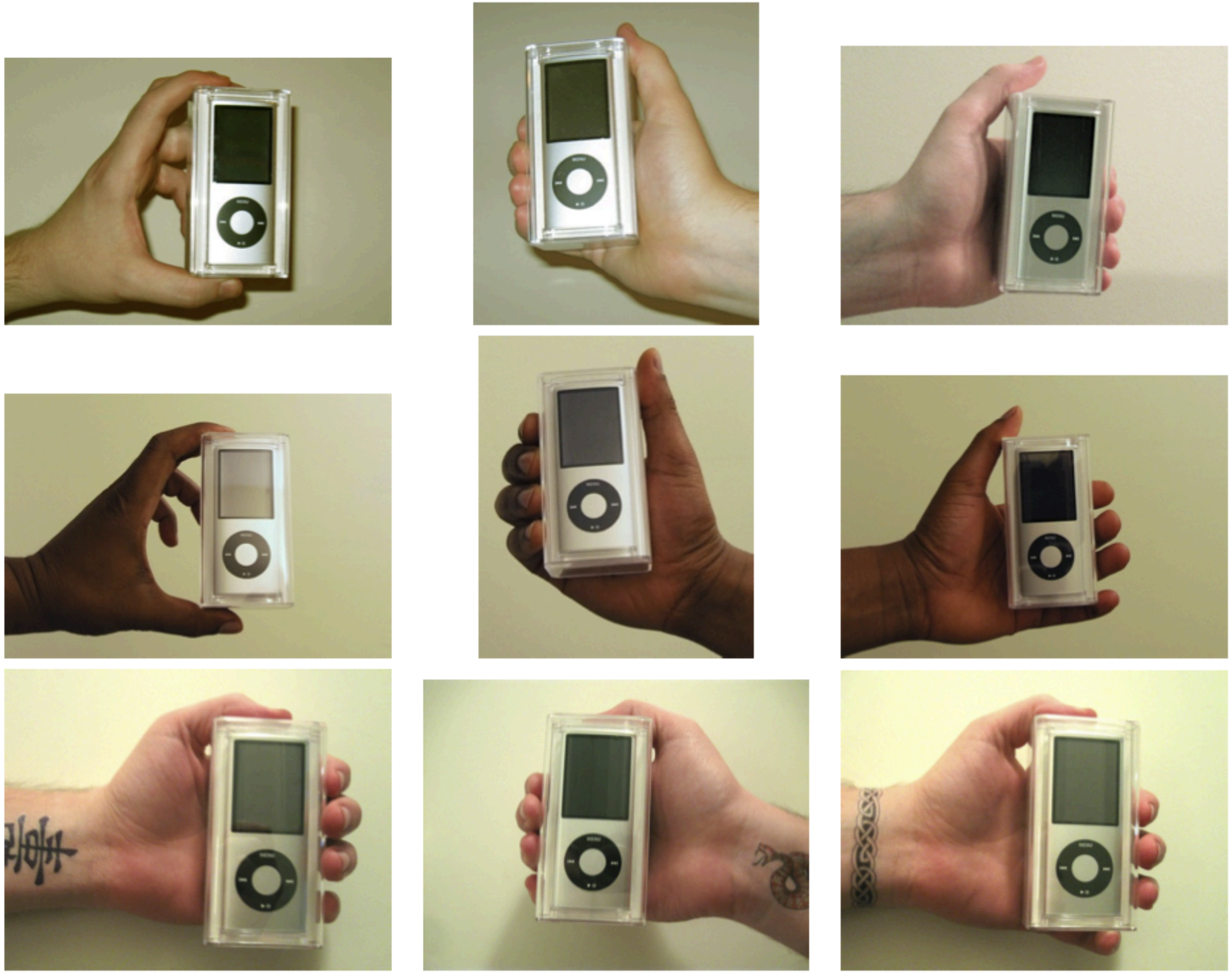


Dennehy, T. C., & Dasgupta, N. (2017). Female peer mentors early in college increase women's positive academic experiences and retention in engineering. *Proceedings of the National Academy of Sciences*, *114*(23), 5964-5969.



"women in engineering who were **assigned a female** (but not male) **peer mentor** experienced more **belonging**, **motivation**, and **confidence** in engineering, better **retention** in engineering majors, and greater engineering career **aspirations**."

Dennehy, T. C., & Dasgupta, N. (2017). Female peer mentors early in college increase women's positive academic experiences and retention in engineering. *Proceedings of the National Academy of Sciences*, *114*(23), 5964-5969.



Doleac, J. L., & Stein, L. C. (2013). ***The visible hand: Race and online market outcomes.*** The Economic Journal, 123(572), F469-F492.

In local online classified advertisements throughout the US… **Black sellers do worse than white sellers** on a variety of market outcome measures:

- They receive 13% **fewer responses**
- 17% **fewer offers**
- Conditional on receiving at least one offer,
  - black sellers also receive 2-4% **lower offers**
- buyers corresponding with black sellers exhibit **lower trust**:

Doleac, J. L., & Stein, L. C. (2013). ***The visible hand: Race and online market outcomes.*** The Economic Journal, 123(572), F469-F492.

J. Nathan Matias

@natematias

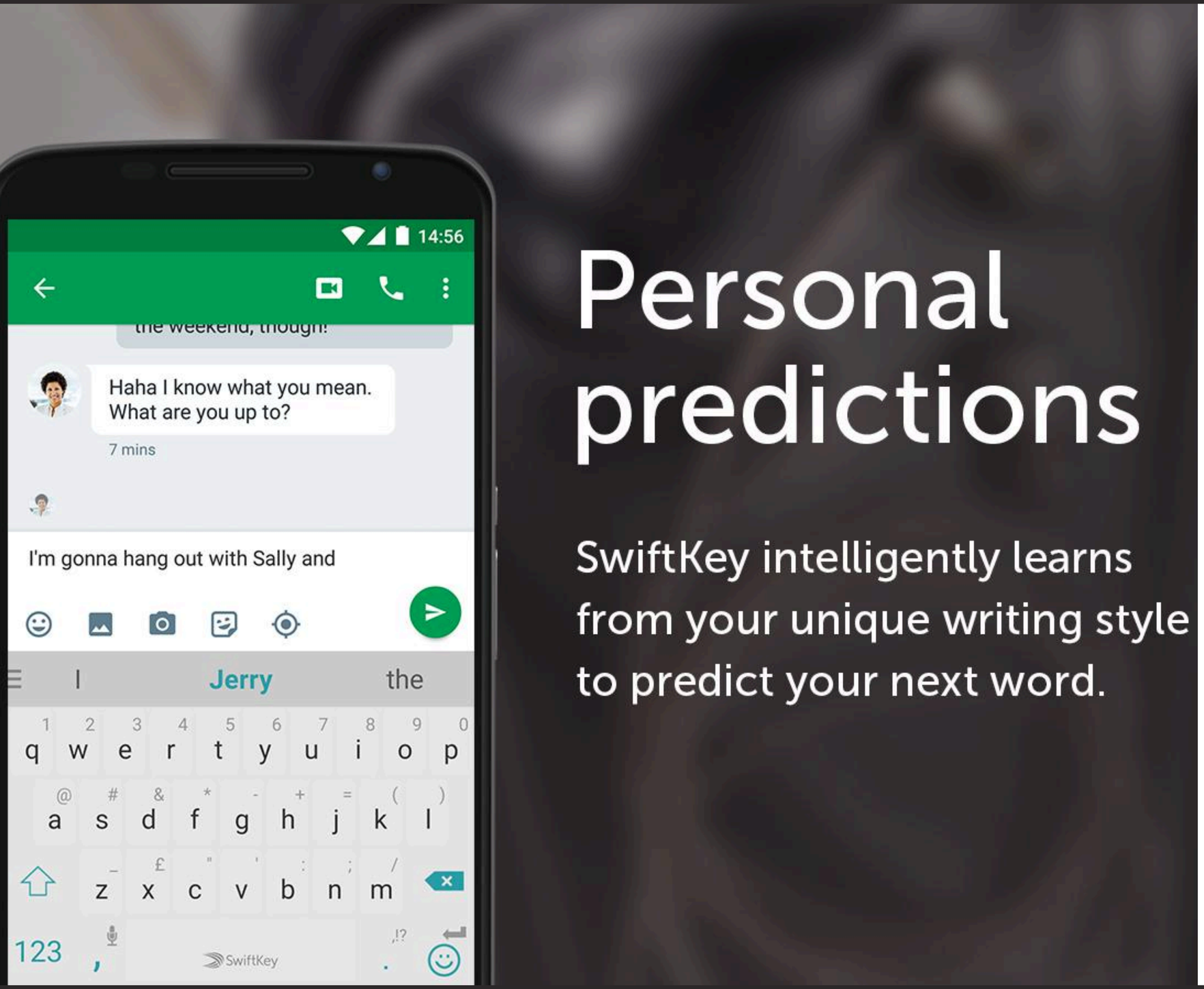
civilservant.io

jmatias@princeton.edu

Department of Psychology
PRINCETON UNIVERSITY

CITP

mit media lab



Personal predictions

SwiftKey intelligently learns from your unique writing style to predict your next word.

J. NATHAN MATIAS

Connect

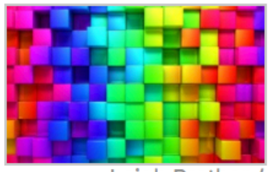
J. Nathan Matias researches technology for cooperation and civic life at the MIT Media Lab and Center for Civic Media. He also facilitated @1book140, *The Atlantic*'s Twitter book club from 2012 – 2014.

Reuters

Who Gets to Use Facebook's Rainbow 'Pride' Reaction?

An analysis to see if the algorithms behind the new emoji contribute to political bubbles in America

J. NATHAN MATIAS, AIMEE RICKMAN, AND MEGAN STEINER JUN 26, 2017



Leigh Prather / Shutterstock

Were All Those Rainbow Profile Photos Another Facebook Study?

The social network learns more about its users than they might realize.

J. NATHAN MATIAS JUN 28, 2015

Randall Munroe

The Tragedy of the Digital Commons

Advocates for fairer, safer online spaces are turning to the conservation movement for inspiration.

J. NATHAN MATIAS JUN 8, 2015

Houghton Mifflin Company

Coming Up: A Live Twitter Chat With 1book140 Author Alison Bechdel

The creator of *Fun Home* will take reader questions on Monday at 7 p.m. Eastern.

J. NATHAN MATIAS OCT 16, 2014

CivilServant

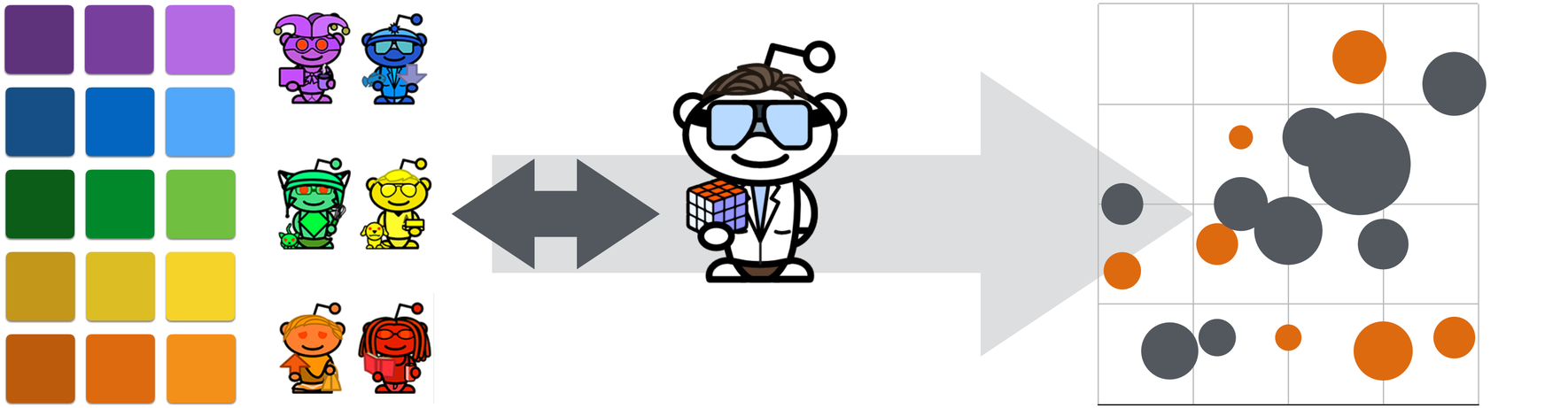
Citizen Behavioral Science for a fairer, safer, more understanding internet

civilservant.io

GlobalVoices

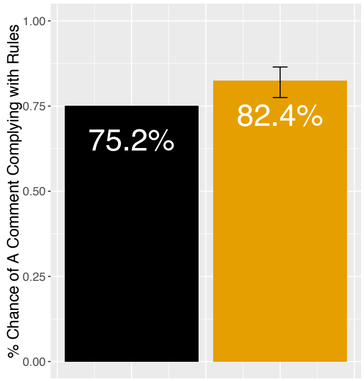
Civil Servant

Community-Led Field Experiments in Governing Human & Machine Behavior

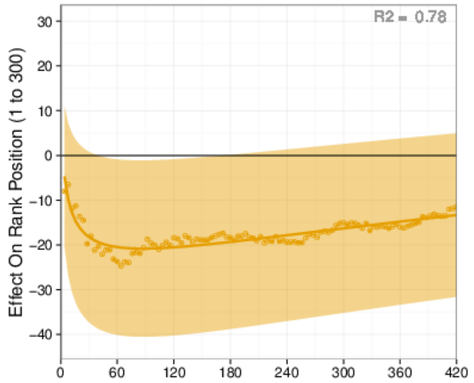


Community Experiments

Open Archive of Experiment Results



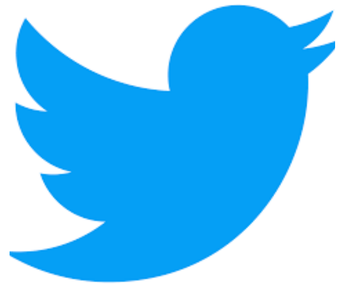
Preventing Harassment



Limiting Misinformation

Managing Conflict

Banning Accounts

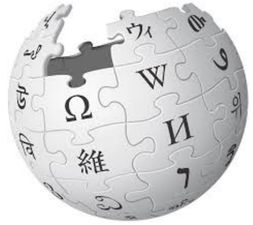


Responding to Harassment

Auditing Algorithms

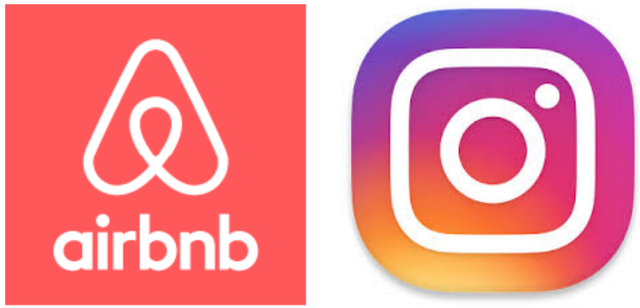
Auditing AI Law Enforcement

Accountability & Ethics



Checkbox image CC-by-2.0 by Weltenraser on The Noun project

THE BEHAVIORAL INSIGHTS TEAM



Science

Design, Conduct, Interpret Field Experiments

Design Experiments & Replications at Scale

Read, **Interpret**, and **Plan Replications** of Other Research

Write & **Critique** an Article Reporting Experiment Results

Understand what Experiments Bring to **Policy**, Social **Science** & **Business**

Participate in Debates on **Ethics** and **Politics** of Experiments

What You Need to Know

- **Statistics**
  - Linear regression (including multivariate)
  - Maximum likelihood models (logistic regression, for ex)
  - Some experience with R

February

Part I: Understanding Field Experiments

- Experience running 1+ experiments
- Experience analyzing experiment results

March into April

Part II: Planning Your Field Experiment

- Experiment Plan
- Community Relationship

April - May

Part III: Deploying & Reporting Your Field Experiment

- Initial Results
- Initial Analysis
- Paper Draft (gradstudents)
- May 7 Final Presentations & Feedback

What to Expect Each Week (more or less)

Homework

- Analysis/writing assignment (**pairs**)
- Progress on an experiment (**pairs**)
- Readings & reflection on slack (**individual**)

Precepts

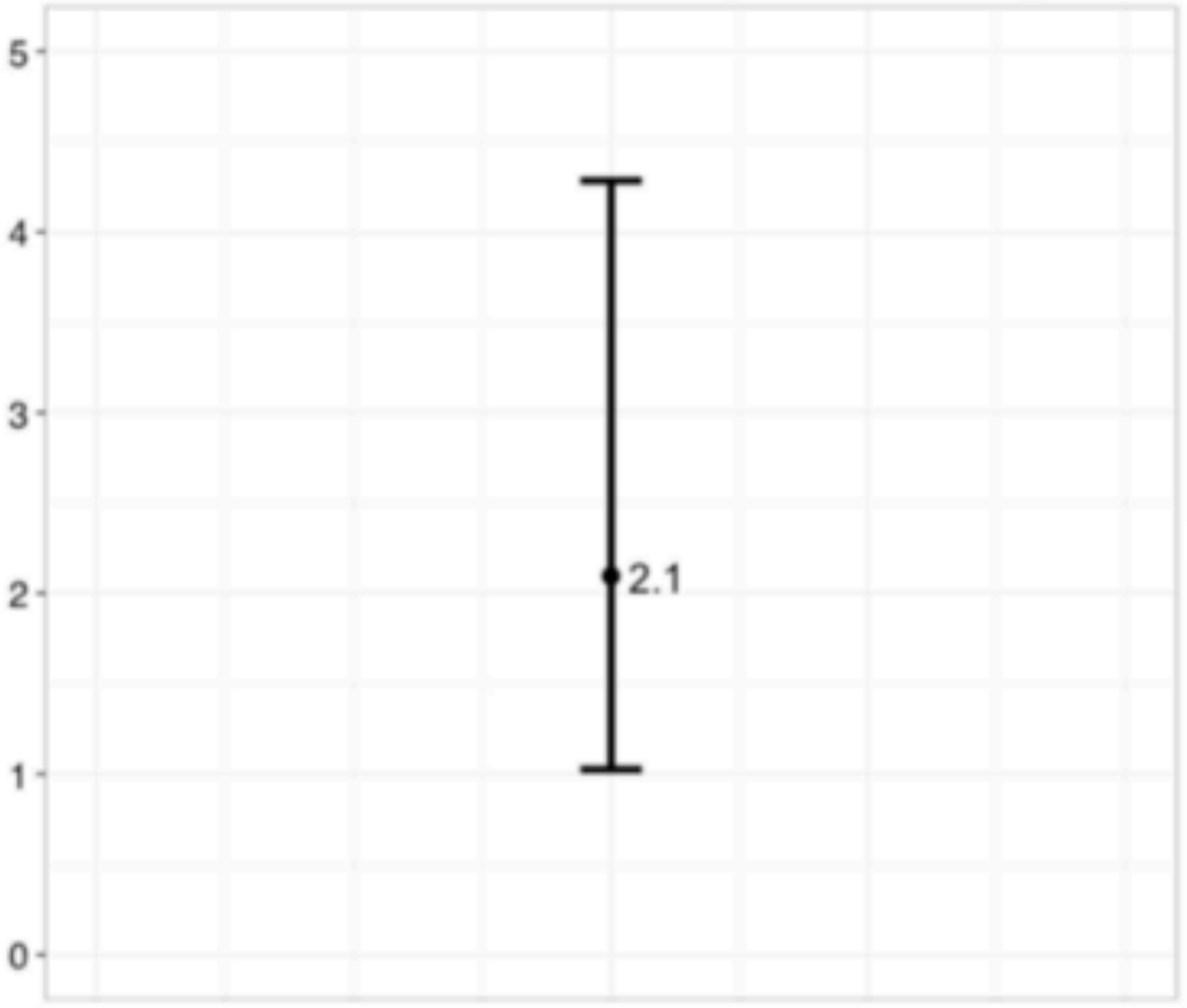
- Discuss last week's assignment
- Discuss context readings (presentation in **pairs**)

Grading Rubric

- **Lectures & Precepts** (20%)
  - Did you attend?
  - Did you speak in class?
  - Did you add meaningfully to the conversation?
- **Weekly Assignments** (20%)
  - Split among Implementation, Analysis, and Writing
- **Midterm Experiment Design** (20%)
- **Final Project** (40%)


UPDATED
MAP OF ONLINE COMMUNITIES
SIZE ON MAP REPRESENTS VOLUME OF DAILY SOCIAL ACTIVITY (POSTS, CHAT, ETC). BASED ON DATA GATHERED OVER THE
SPRING AND SUMMER OF 2010
SPOKEN LANGUAGE
INTERNET
CELL PHONES
PLAINS OF AWKWARDLY PUBLIC FAMILY INTERACTIONS
BLATHERSKITE RIVER
JUNGLE-BAY MOUNTAINS OF "IT'S COMPLICATED"
LAMEBOOK BAY
PEOPLE YOU CAN'T UNFRIEND
DATA MINES
FARM
DOPAMINE SEA
HABBO
MAPLE STORY
MMO ISLE
MOUNTAINS OF STEAM
RUNESCAPE (FARMLAND)
RIVER GRIND
WoW
GULF OF LAG
CDC GAMES
EVE ONLINE
CHARRED WASTELAND OF ABANDONED SOCIAL NETWORKS
NICHE MARKET MOUNTAINS
NING
EURO GULF
SKYROCK
BADOO
MYSPACE
ORKUT
VKONTAKTE
LINKEDIN
HI5
OZONE
CORPORATE BAY
REGIONAL BAY
MIXI
BUZZWORD BAY
FLICKR
YOUTUBE
VIRAL SHORES
MUSIC VIDEO BAY
SNOB SOUND
DEVIANTART
LATITUDE
FOURSQUARE
GEOTAGGED BAY
YELP
GOOGLE BUZZ
SOCIAL MEDIA CONSULTANT CHANNEL
BIEBER BAY
BIT.LY MOUNTAINS
TWITTER
DESERT OF FOOD UPDATES
SEA OF OPINIONS
TUMBLR
PHOTO BLOGS
BAY OF DRAMA
DIARY
GOSSIP BLOGS
BLOGOSPHERE (CORE)
DIGG
REDDIT
STUMBLEUPON
TROLL BAY
SEA OF MEMES
FORUMS
4CHAN
2CHANNEL
CRAIGSLIST
GREAT FIREWALL
GULF OF CHINA
GRASS MUD HORSE BAY
AIM
WINDOWS LIVE
IRC ISLES
PROTOCOL

Using Facebook's **color backgrounds** on **poems** **increases the rate** of **likes** & **comments** by **2.1x**



Short poems (n=22) posted Oct-Nov 2017 by J. Nathan Matias (@natematias)
Results of a log-transformed linear regression, adjusted for day of week (R2=0.54)

The Cornhole Experiment

Final Project

The final project will involve designing, conducting, and analyzing early results from a field experiment, writing for a policy audience.

Graduate students will be expected to write up the field experiment for a policy audience, and also for academic publication.

The final can be a project of your own, or you can choose from an upcoming CivilServant study.

Auditing How Tech Companies Enforce Political Ad Policies

Austin Hounsel

J. Nathan Matias

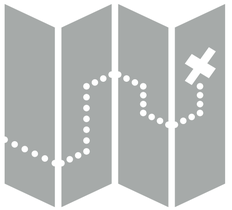
Nick Feamster

Volunteer Investigators

+

Generated Ads

+



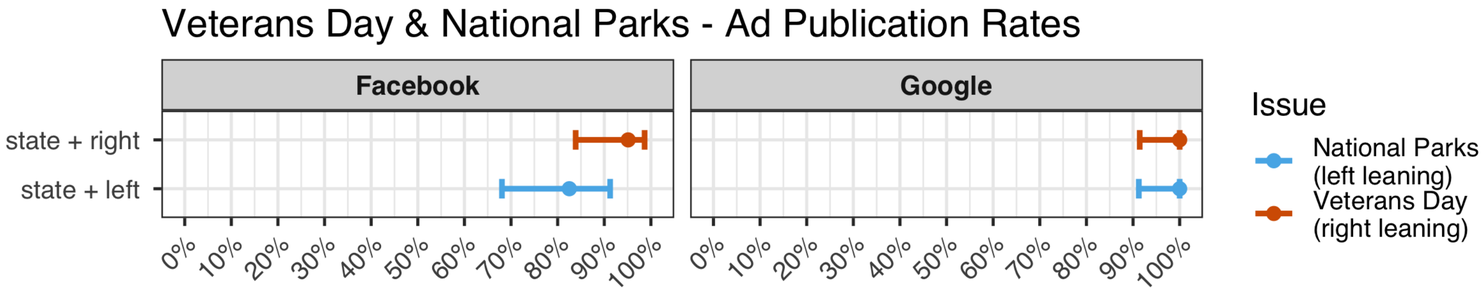
Targeting Criteria

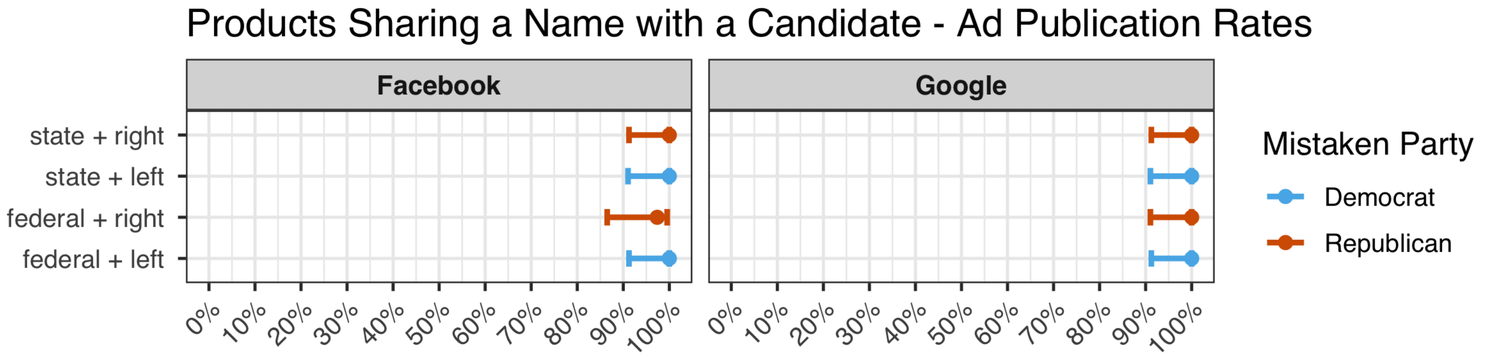
Auditing How Tech Companies Enforce Political Ad Policies

Austin Hounsel

J. Nathan Matias

Nick Feamster





Estimated chance of publication for a given ad combination(election, leaning). Ad placement attempts: 477 ads placed by 7 people from 2018-09-17 to 2018-10-10. Product ads are music albums that share a word with a candidate name. Veterans Day & National Park ads are about events and places that could be be mistaken by platform policy enforcers as election-related ads of national importance. 95% confidence intervals use the Wilson method. Code & data at: https://github.com/citp/mistaken-ad-enforcement
Data and analysis by J. Nathan Matias & Austin Hounsel of Princeton University,
with Melissa Hopkins, Ben Werdmuller, Jason Griffey, Chris Peterson, Scott Hale, and Nick Feamster

The Gray Phone Challenge

Zenobia Chan

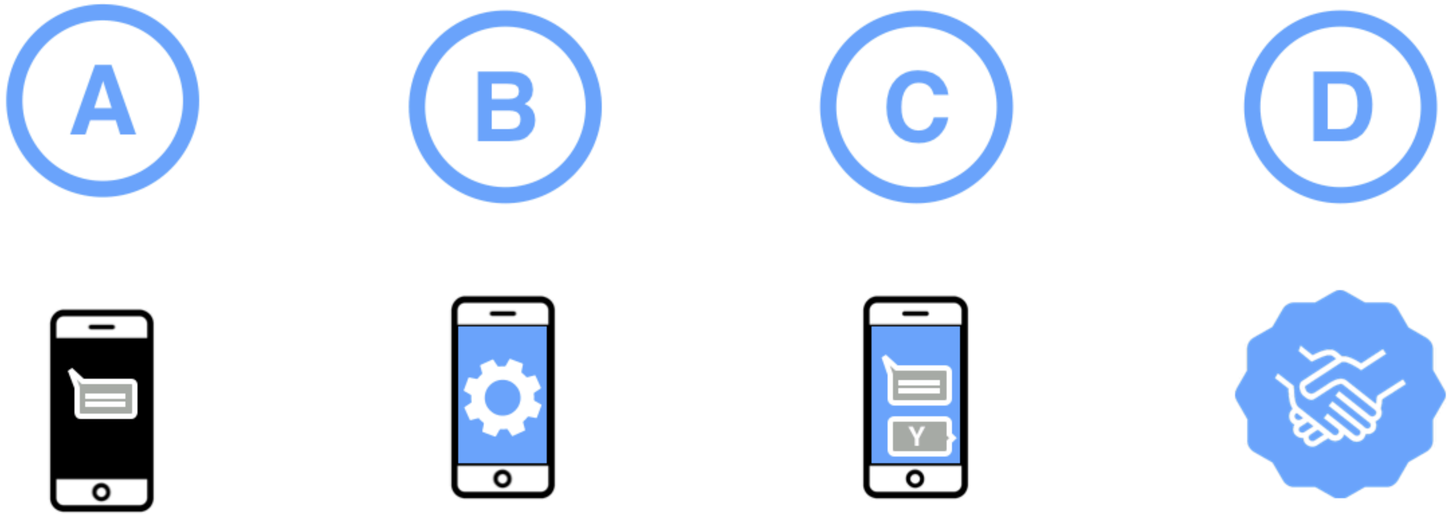
J. Nathan Matias

Gray Phone Challenge Daily Routine

Checking-in just takes a minute before bed or in the morning.
Time-tracking resets at midnight.

Zenobia Chan

J. Nathan Matias



Receive Daily Message

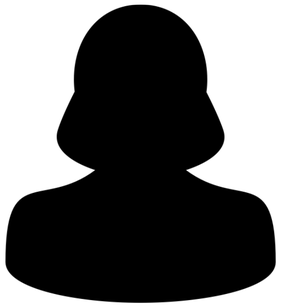
Adjust Color

Confirm Daily Status

Stick with this all day :-)

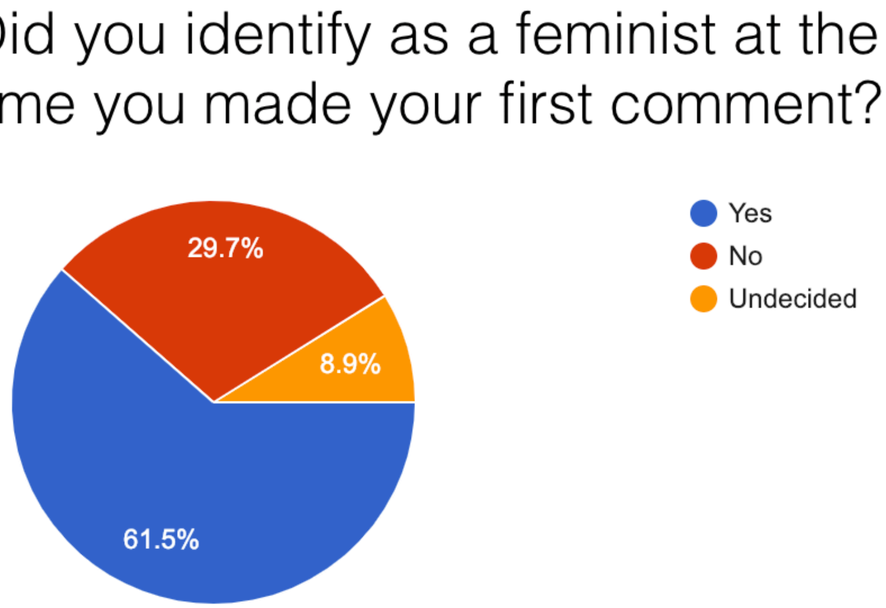
Promoting Inclusion and Participation in an Online Gender-Related Discussion Community

Tyler Simko



Emily Hedlund

J. Nathan Matias



Final Project Opportunities

- Growing Diverse Knowledge Online with WikiLovesAfrica
- Gray Phone Challenge v2
- reddit study TBD
- (analysis) Analyzing Thousands of Historical Experiments
- (design) Survival Models as Average Treatment Effects on Wikipedia
- (Your Project Here)

Office Hours

Tuesday, 11am - 12am (15 & 30 minutes)

Thursday, 11am - 12am (15 & 30 minutes)

Sherrerd 3rd floor lounge
(or Sherrerd 313 if privacy is needed)

https://meetme.so/natematias-soc412

IRB Training

Essential. If you do not have certification for Human Subjects Research, please take this course ASAP, as your participation in the class will depend on this. It may take ~3-4 hours.

“Social and Behavioral Researchers”

https://www.princeton.edu/ria/

Getting Further Support

- **Statistics Support**
  - Data & Statistical Services (dss.princeton.edu)
  - Center for Statistics & Machine Learning (csml.princeton.edu)
- **Writing Support**
  - Princeton Writing Center (writing.princeton.edu/center)

Next Steps

- **Complete the Survey** at: http://bit.ly/2MQqDWe
- **Join Slack**
- **Read Shadish, Cook, and Campbell**
- **Download assignment data:** https://github.com/natematias/SOC412/
- If necessary, complete Princeton IRB training **by Tuesday Feb 12**