
PRIVACY CON
FEDERAL TRADE COMMISSION
DC // 1.14.16

Remarks

Commissioner Julie Brill

PRIVACY CON

FEDERAL TRADE COMMISSION

Session 3:
Big Data and Algorithms: Transparency Tools Revealing Data Discrimination

PRIVACY CON
FEDERAL TRADE COMMISSION

Michael Carl Tschantz
University of California, Berkeley
Anupam Datta
Carnegie Mellon University

Automated Experiments on Ad Privacy Settings

Co-author: Amit Datta (Carnegie Mellon University)

AdFisher

Information Flow Experiments on Ad Privacy Settings

Michael Carl Tschantz
International Computer Science Institute

Anupam Datta
Carnegie Mellon University

Joint work with Amit Datta, CMU

'We'll be back': Hong Kong protesters chant as camp site dismantled

Reuters | Dec 12, 2014, 08.39 AM IST

READ MORE »Hong Kong Protesters | 'We'll Be Back' | Hong Kong | CY Leung

Police officers stand guard before they move on to remove protesters from a road written 'We Will Be Back' with tarps at an occupied area outside government headquarters in Hong Kong.

HONG KONG: Hong Kong police arrested pro-democracy activists and cleared most of the main protest site on Thursday, marking an end to more than two months of street demonstrations in the Chinese-controlled city, but many chanted: "We will be back".

Most activists chose to leave the Admiralty site, next to the Central business area, peacefully, despite their demands for a free vote not being met. But the overall mood remained defiant.

Hong Kong Federation of Students leader Alex Chow said: "You might have the clearance today but people will come back on to the streets another day."

RELATED

Antidepressant Medication - Info On An Rx **Antidepressant** Drug

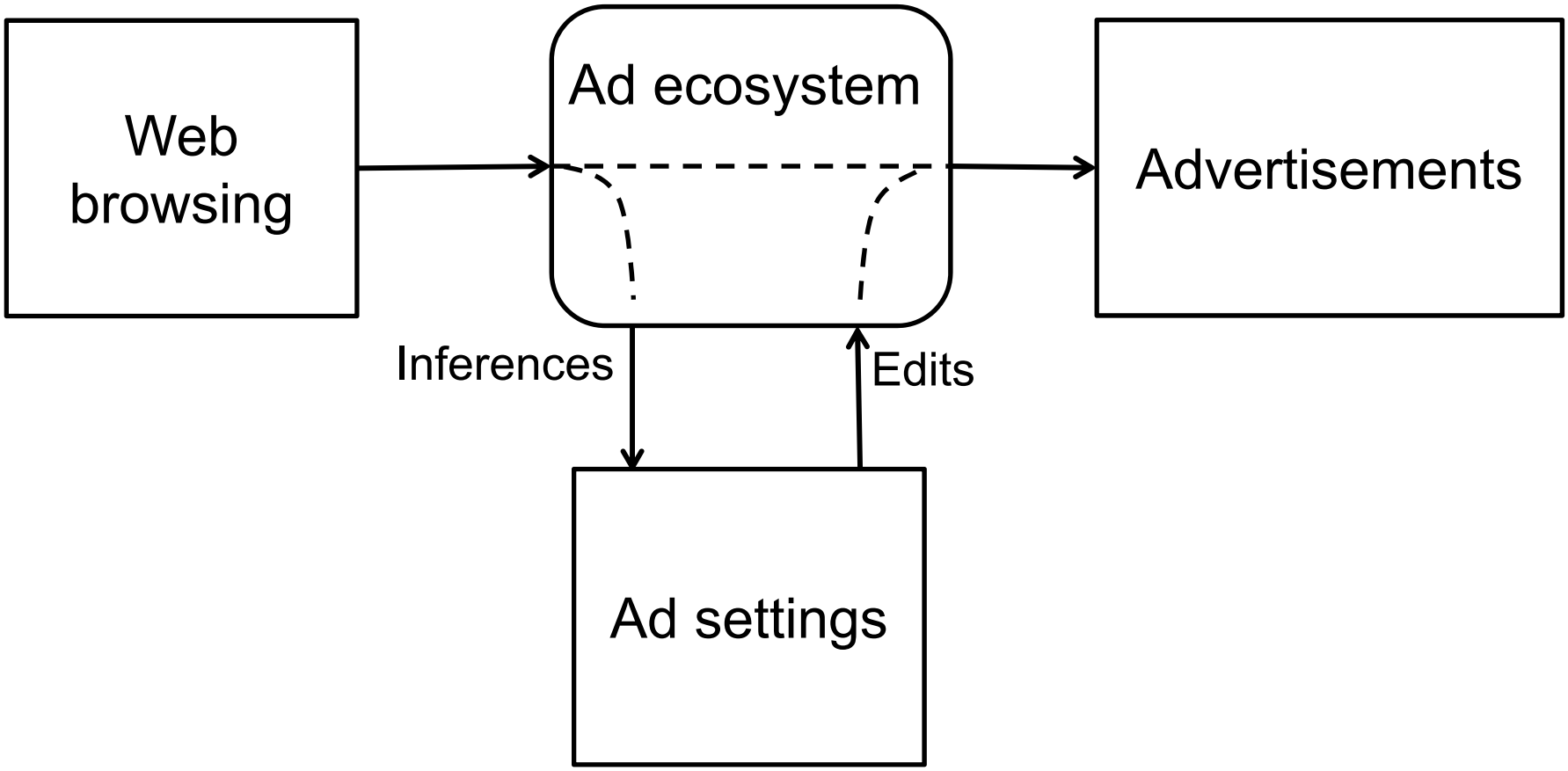
knowmydepression.com/**antidepressant**

Visit For Treatment Info & Facts.

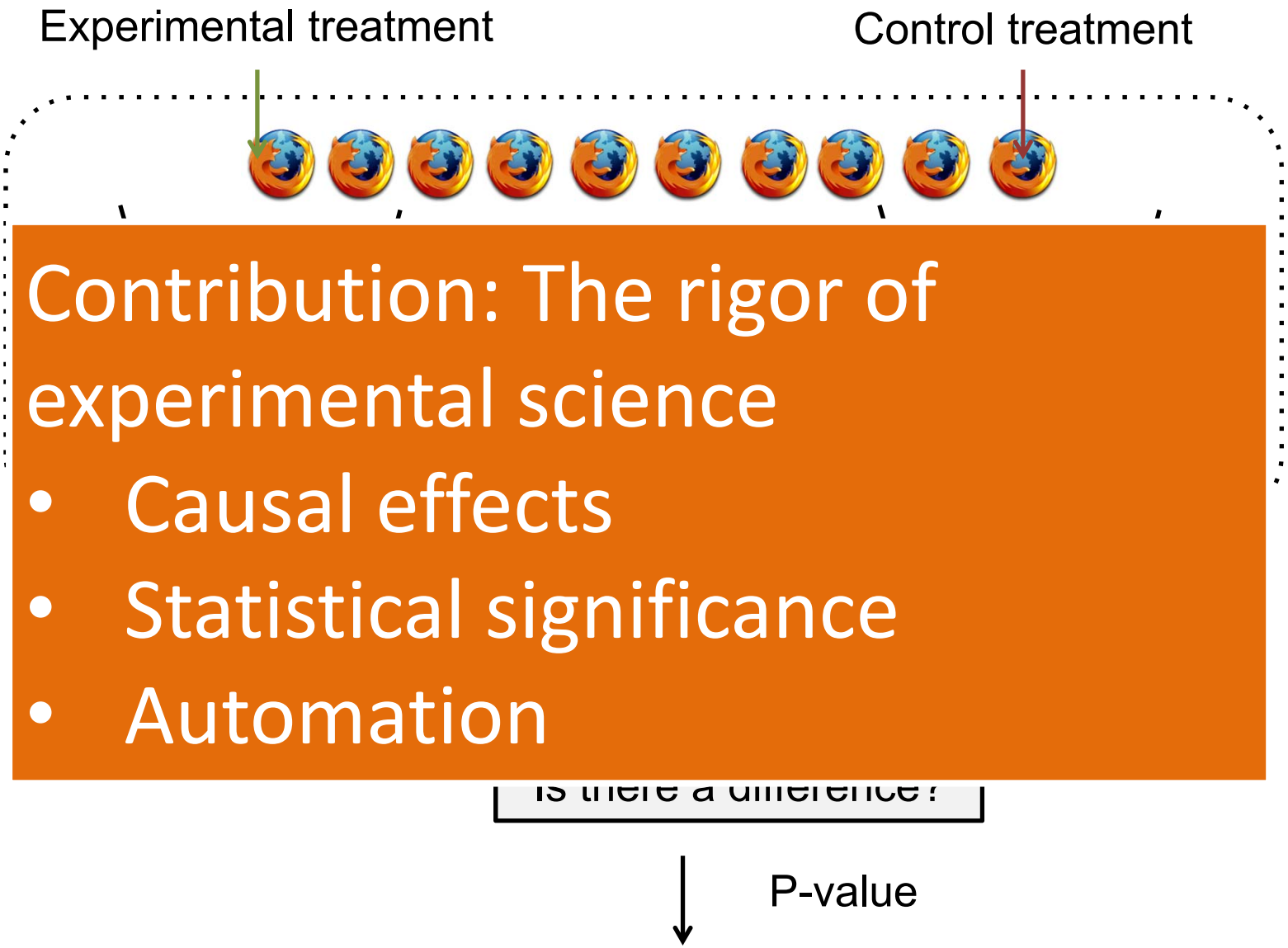
Settings for Google ads

Ads enable free web services and content. These settings help control the types of Google ads you see.

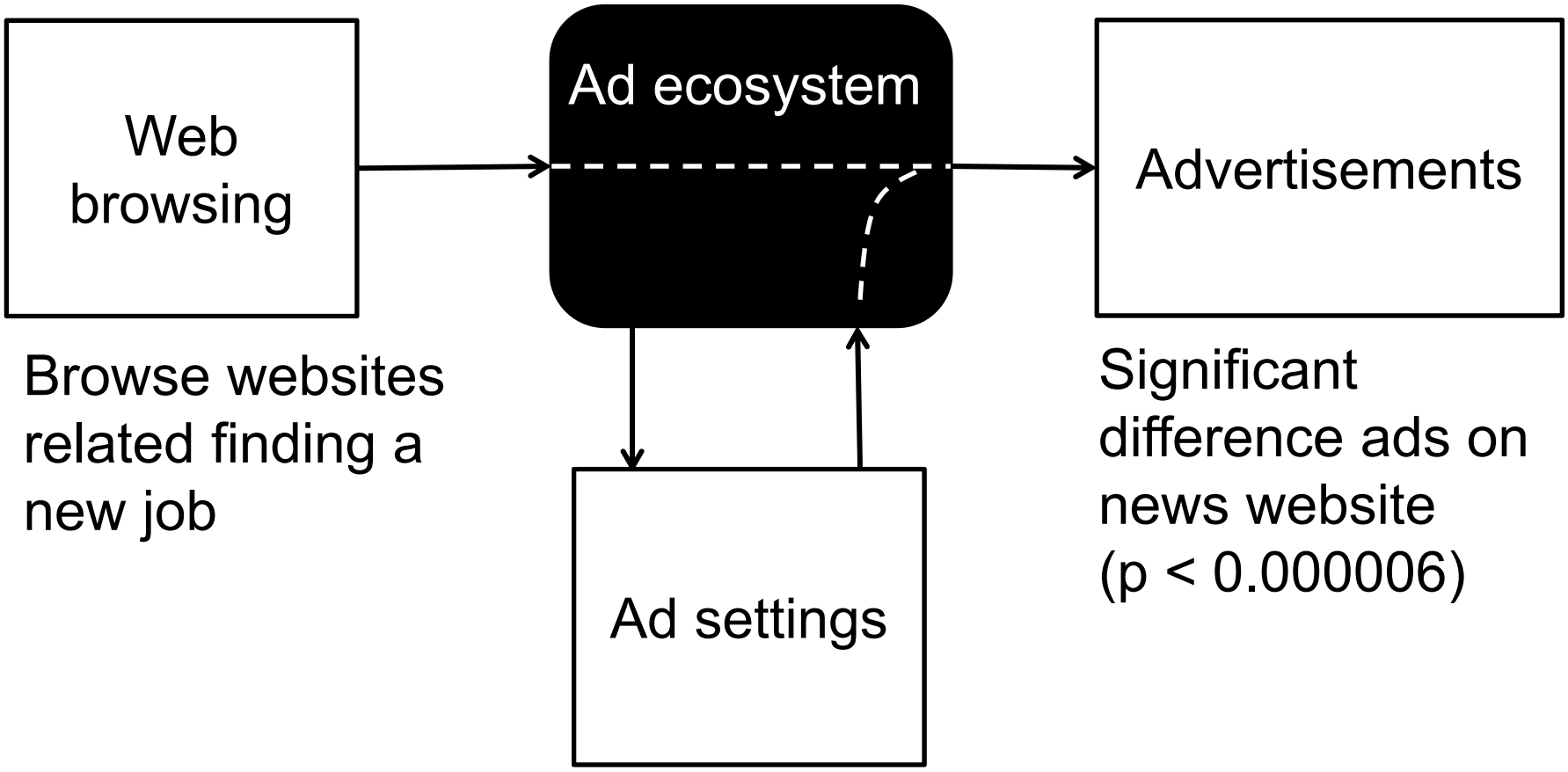
	Ads on Google	Google ads across the web
	 Search	 Google ads across the web  YouTube
Gender	N/A	Female Edit Based on the websites you've visited
Age	N/A	25-34 Edit Based on the websites you've visited
Languages	N/A	English Edit Based on the websites you've visited
Interests	N/A	Air Travel, and 30 more Edit Based on the websites you've visited
Opt-out settings	You've opted out of *interest-based* ads on Google. Opt in to *interest-based* ads on Google	Opt out of *interest-based* Google ads across the web


Web browsing
Ad ecosystem
Advertisements
Inferences
Edits
Ad settings

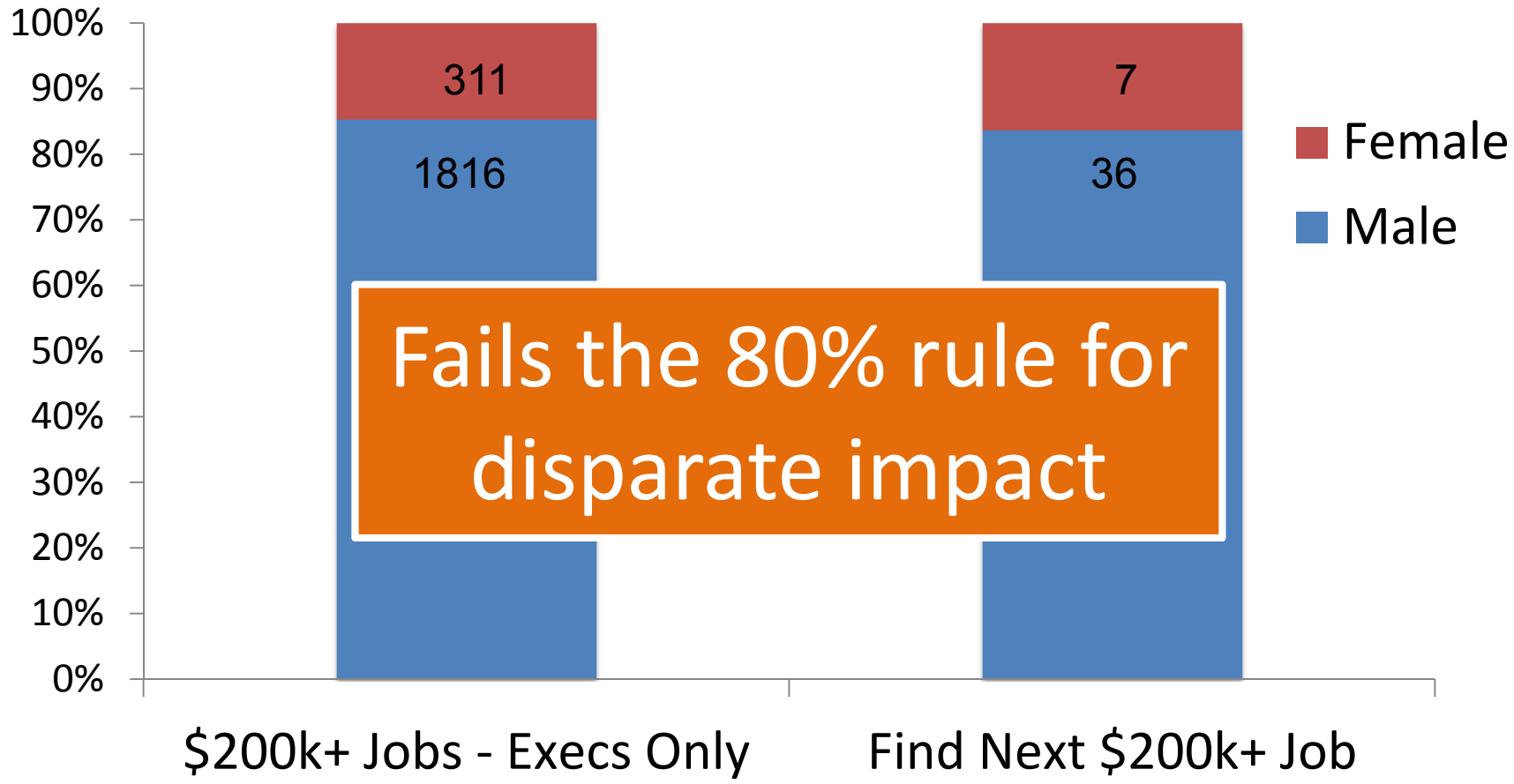
AdFisher


Experimental treatment
Control treatment
Contribution: The rigor of experimental science
• Causal effects
• Statistical significance
• Automation
P-value

Discrimination



Discrimination Explanation



Open questions

MIT Technology Review

Pittsburgh Post-Gazette
post-gazette.com

YAHOO! TECH

WIRED

The New York Times
TheUpshot

The Washington Post
The Intersect

THE WALL STREET JOURNAL

- How widespread?
- Who is responsible?

The Barrett Group

Other advertisers

Input

Output

Google

Websites

Male users

Female users

Input

Output

The Barrett Group

Other advertisers

Google

Websites

Male users

Female users

Input

Output

The Barrett Group

Show to males

Google

Male users

Other advertisers

Websites

Female users

Input

Output

The Barrett Group

Other advertisers

Show to high earners

Google

Websites

High earners are male

Male users

Female users

Input

Output

The Barrett Group

Other advertisers

Show to both

Show to females

Google

Websites

Male users

Female users

Input

Output

The Barrett Group

Other advertisers

Show to both

Google

Websites

Not clicking

Clicking

Male users

Female users

Summary

- AdFisher: Rigorous experimental design
  - Causal effects
  - Statistical significance
  - Automation
- Found gender-based discrimination
- Open questions:
  - How widespread?
  - How to assign responsibility?

More Information

- http://www.cs.cmu.edu/~mtschant/ife/

- M.C. Tschantz, A. Datta, A. Datta, and J.M. Wing.
  A methodology for information flow experiments.
  CSF 2015.

- A. Datta, M.C. Tschantz, and A. Datta.
  Automated Experiments on Ad Privacy Settings:
  A Tale of Opacity, Choice, and Discrimination.
  PETS 2015

Roxana Geambasu

Columbia University

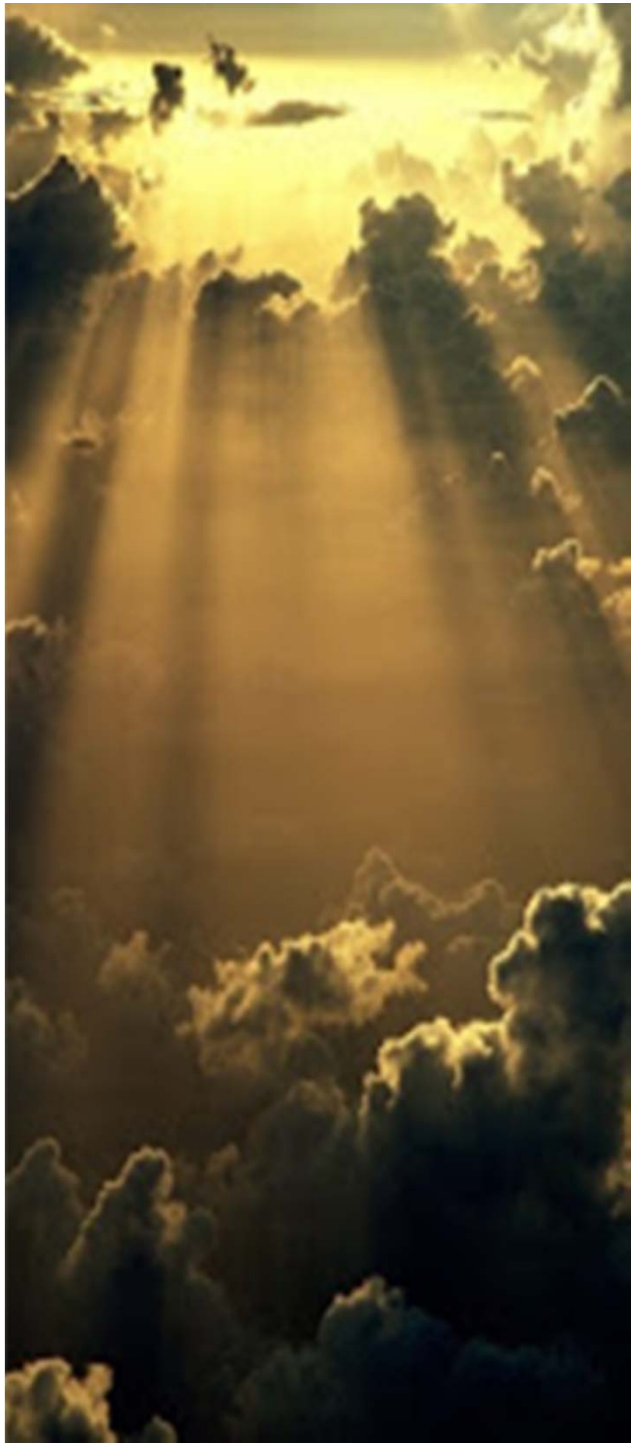
Sunlight: Fine-grained Targeting Detection at Scale with Statistical Confidence

Co-authors: Mathias Lecuyer, Riley Spahn, Yannis Spiliopoulos, Augustin Chaintreau, Daniel Hsu (Columbia University)

Sunlight:
Web Transparency at Scale.

Mathias Lecuyer, Riley Spahn, Yannis Spiliopoulos,
Augustin Chaintreau, Roxana Geambasu, and Daniel Hsu
Columbia University

http://columbia.github.io/sunlight/



Example: Gmail Ads

Example: Gmail Ads

	email subject & text
E1	**Vacation** I'm going on vacation to travel.
E2	**Homosexual** Gay, lesbian, homosexual.
E3	**Pregnant** I'm pregnant. I'm having a baby.
E4	**Unemployed** I'm unemployed.
E5	**Ford** I want to buy a car, maybe a Ford.

ad title, url & text	
Ralph Lauren Online Shop www.ralphlauren.com The official Site for Ralph Lauren Apparel, Acccessories & More	Ad1
Cedars Hotel Loughborough www.thecedarshotel.com 36 Bedrooms, Restaurant, Bar Free WiFi, Parking, Best Rates	Ad2

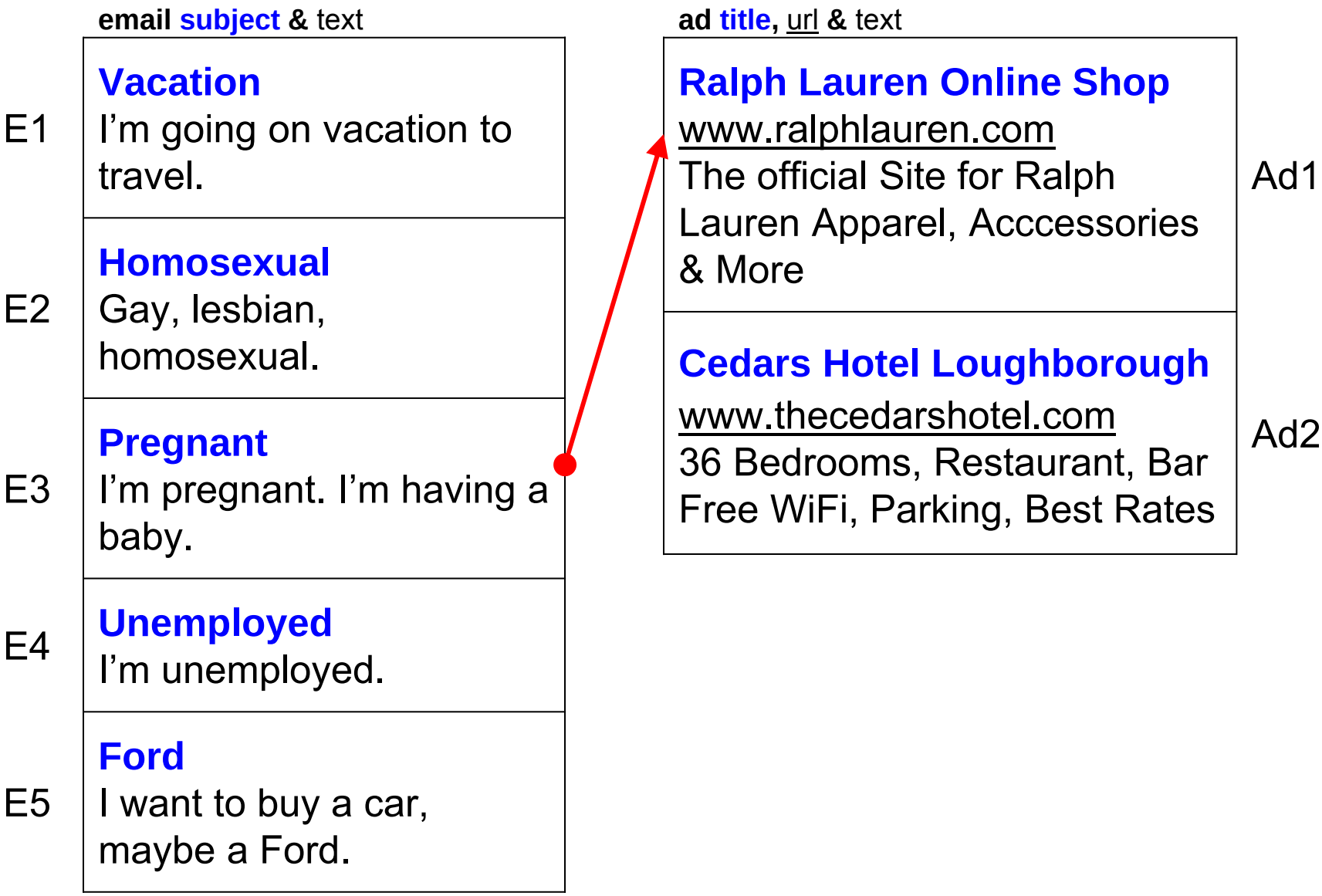
Example: Gmail Ads

	email subject & text
E1	**Vacation** I'm going on vacation to travel.
E2	**Homosexual** Gay, lesbian, homosexual.
E3	**Pregnant** I'm pregnant. I'm having a baby.
E4	**Unemployed** I'm unemployed.
E5	**Ford** I want to buy a car, maybe a Ford.

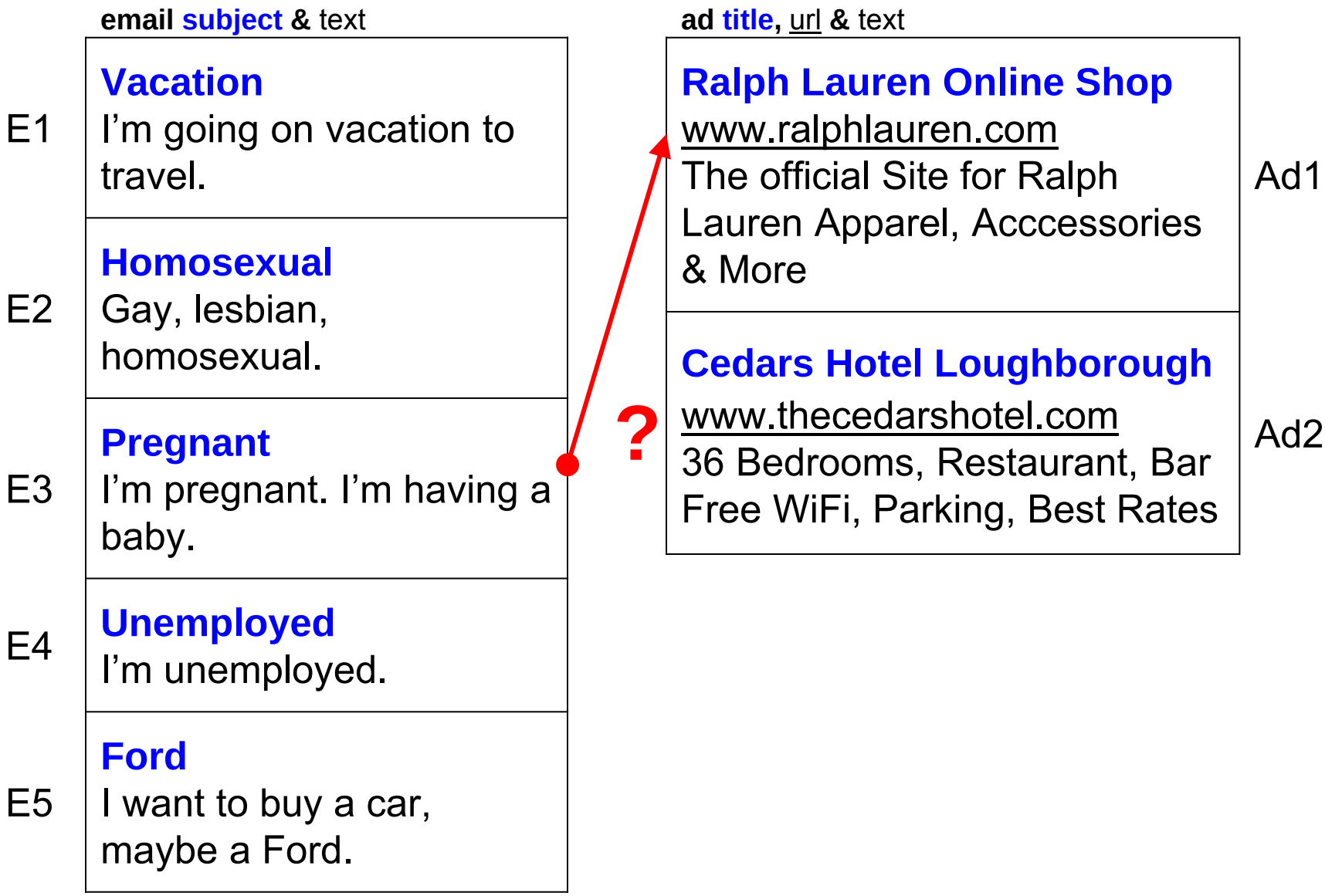
?

ad title, url & text	
Ralph Lauren Online Shop www.ralphlauren.com The official Site for Ralph Lauren Apparel, Acccessories & More	Ad1
Cedars Hotel Loughborough www.thecedarshotel.com 36 Bedrooms, Restaurant, Bar Free WiFi, Parking, Best Rates	Ad2

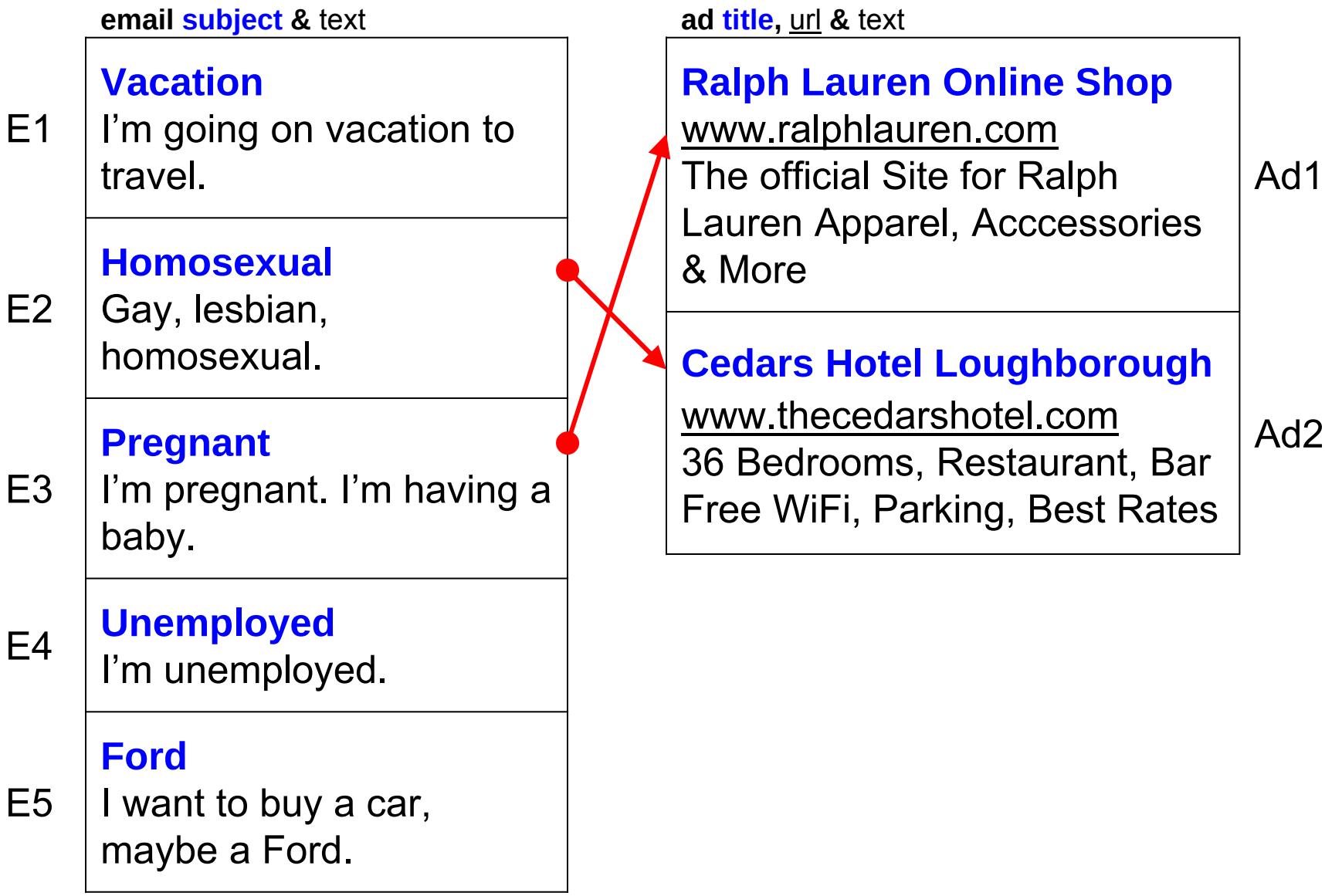
Example: Gmail Ads



Example: Gmail Ads



Example: Gmail Ads



It’s not just Gmail...

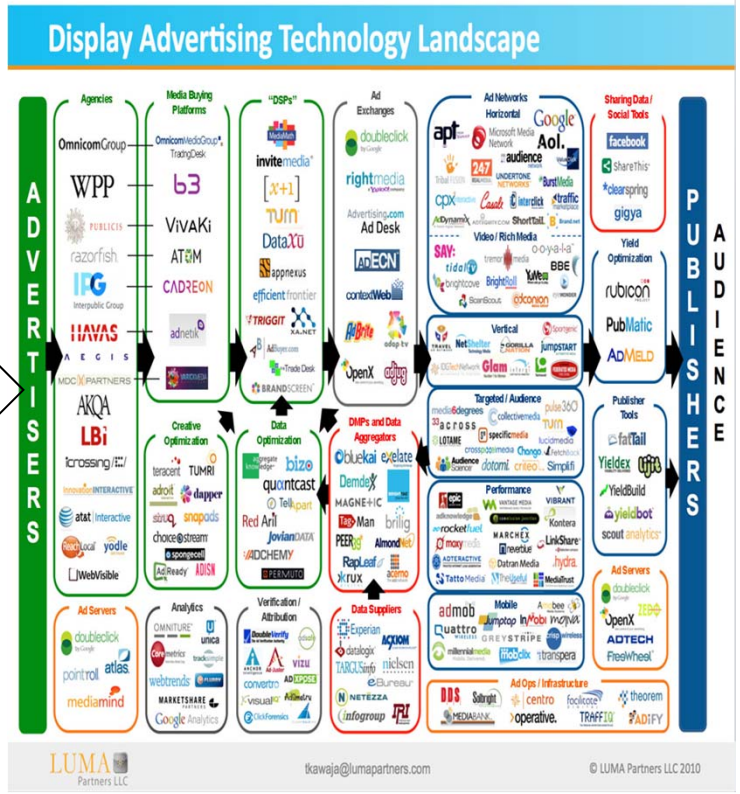
Did you know?

- Data brokers can tell when you're sick, tired and depressed -- and sell this information. [CNN ’14]
- Google Apps for Ed used institutional emails to target ads in personal accounts. [SafeGov’14]
- Credit companies are looking into using Facebook data to decide loans. [CNN’13]

The data-driven web

- The web is a complex and opaque ecosystem driven by massive collection and monetization of personal data.

data



- Who has what data?
- What's it used for?
- Are the uses good or bad for us?

- End-users, privacy watchdogs (eg, FTC) are equally blind.

Our research

- Build transparency and oversight tools that increase users' awareness and society's oversight over web services' use of personal data.

- Timeline:
  - 2014: **XRay,** the first targeting detection tool; it reveals targeting through correlation [USENIX Security'14].
  - 2015: **Sunlight**, second-generation, more robust tool; it reveals the causes of targeting at scale and with statistical justification [CCS'15].
  - Ongoing: **DataObservatory**, the first tool to reveal personalization on arbitrary web pages.
  - Ongoing: **Hubble**, transparency tool based on end-user information.

Ph.D. students:

Mathias Lecuyer

Riley Spahn

Yannis Spiliopoulos

Faculty:

Augustin Chaintreau

Roxana Geambasu

Daniel Hsu

Arvind Narayanan

Sunlight

Generic and broadly applicable system that detects personal data use for targeting and personalization.

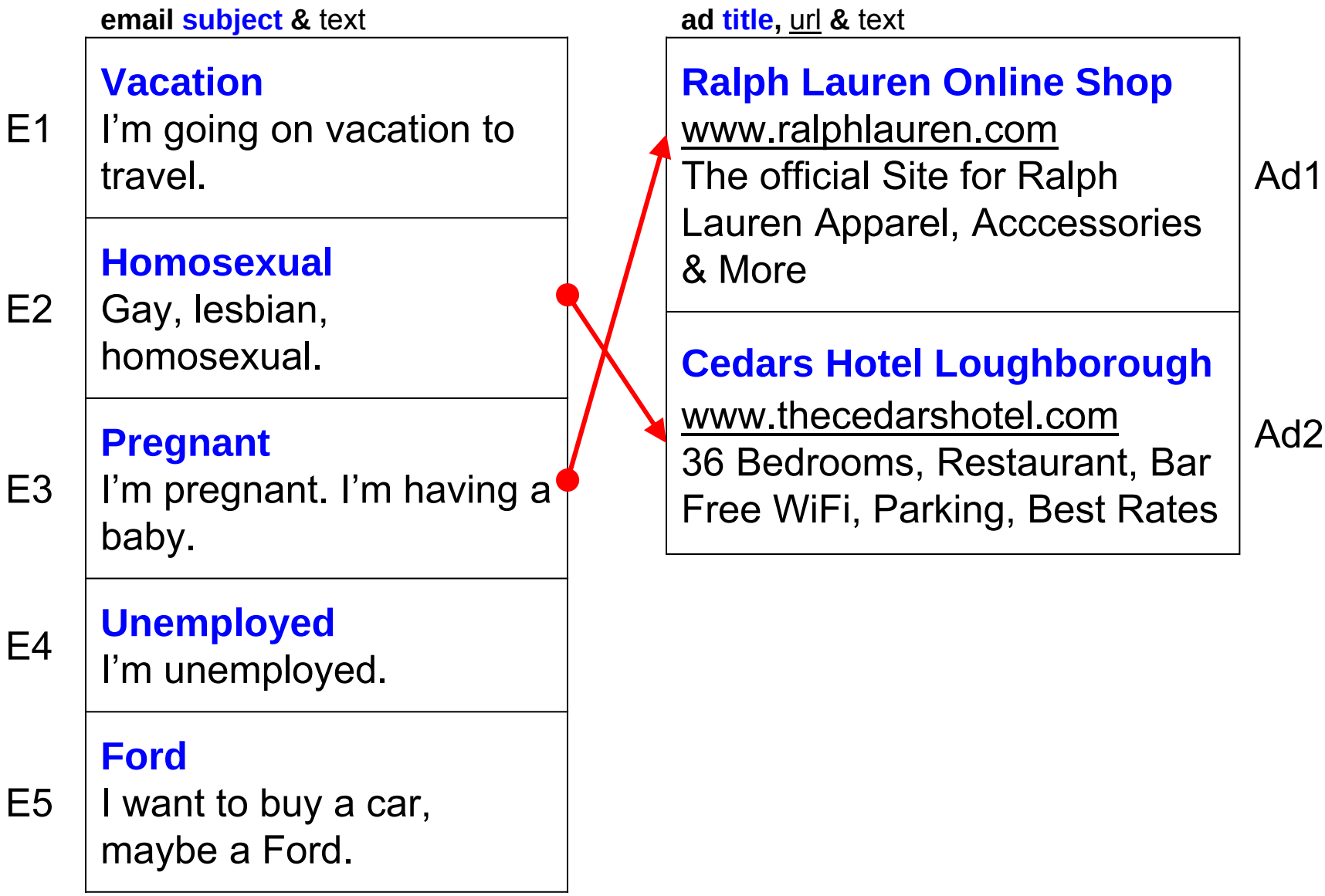
Reveals which data (e.g., emails) triggers which outputs (e.g., ads).

- Key idea: correlate inputs with outputs based on observations from profiles with differentiated inputs.

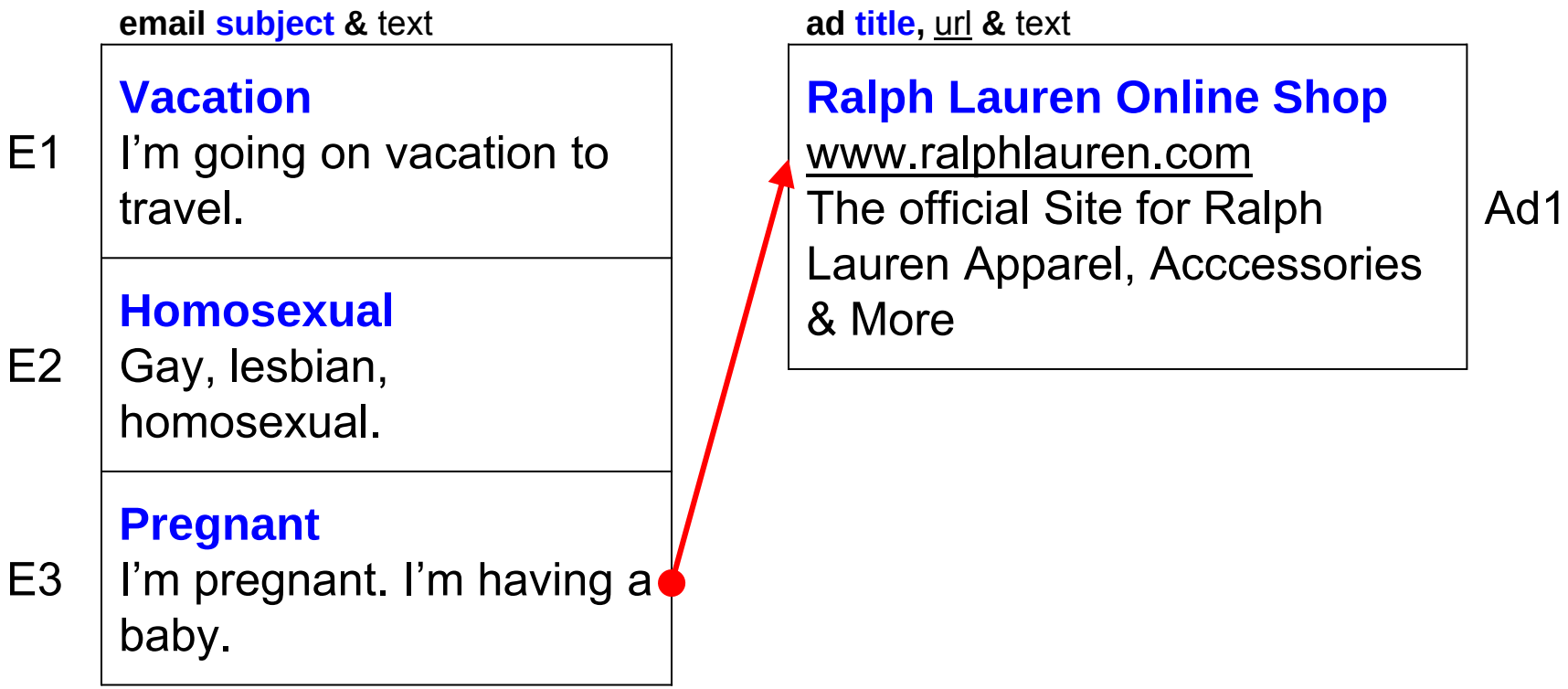
Sunlight is precise, scalable, and works with many services.

We tested it for Gmail ads, ads on arbitrary websites, recommendations on Amazon & YouTube, prices in travel websites.

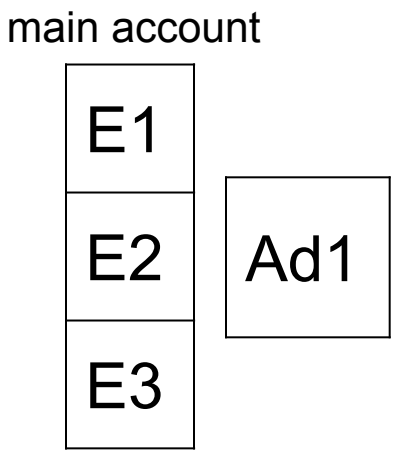
Example


email subject & text
E1
Vacation
I'm going on vacation to travel.
E2
Homosexual
Gay, lesbian, homosexual.
E3
Pregnant
I'm pregnant. I'm having a baby.
E4
Unemployed
I'm unemployed.
E5
Ford
I want to buy a car, maybe a Ford.
ad title, url & text
Ralph Lauren Online Shop
www.ralphlauren.com
The official Site for Ralph Lauren Apparel, Acccessories & More
Ad1
Cedars Hotel Loughborough
www.thecedarshotel.com
36 Bedrooms, Restaurant, Bar Free WiFi, Parking, Best Rates
Ad2

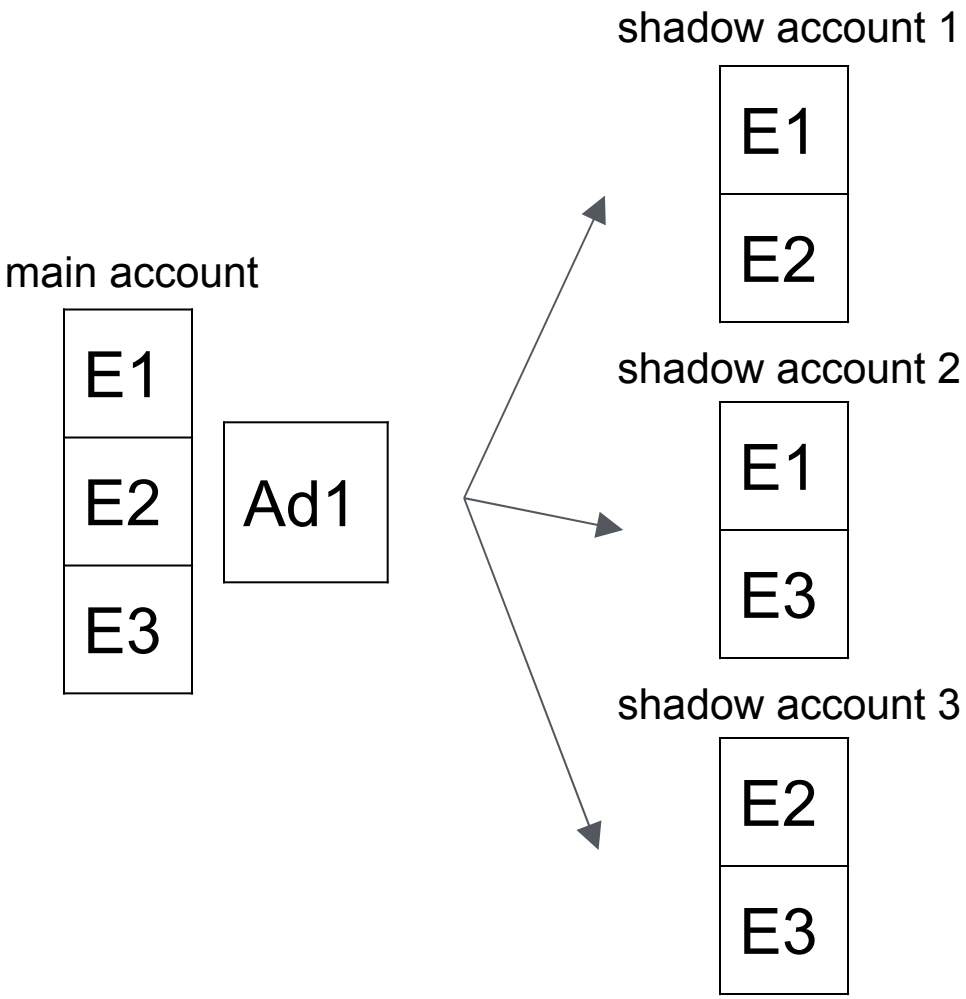
Example


email subject & text
E1
Vacation
I'm going on vacation to travel.
E2
Homosexual
Gay, lesbian, homosexual.
E3
Pregnant
I'm pregnant. I'm having a baby.
ad title, url & text
Ralph Lauren Online Shop
www.ralphlauren.com
The official Site for Ralph Lauren Apparel, Acccessories & More
Ad1

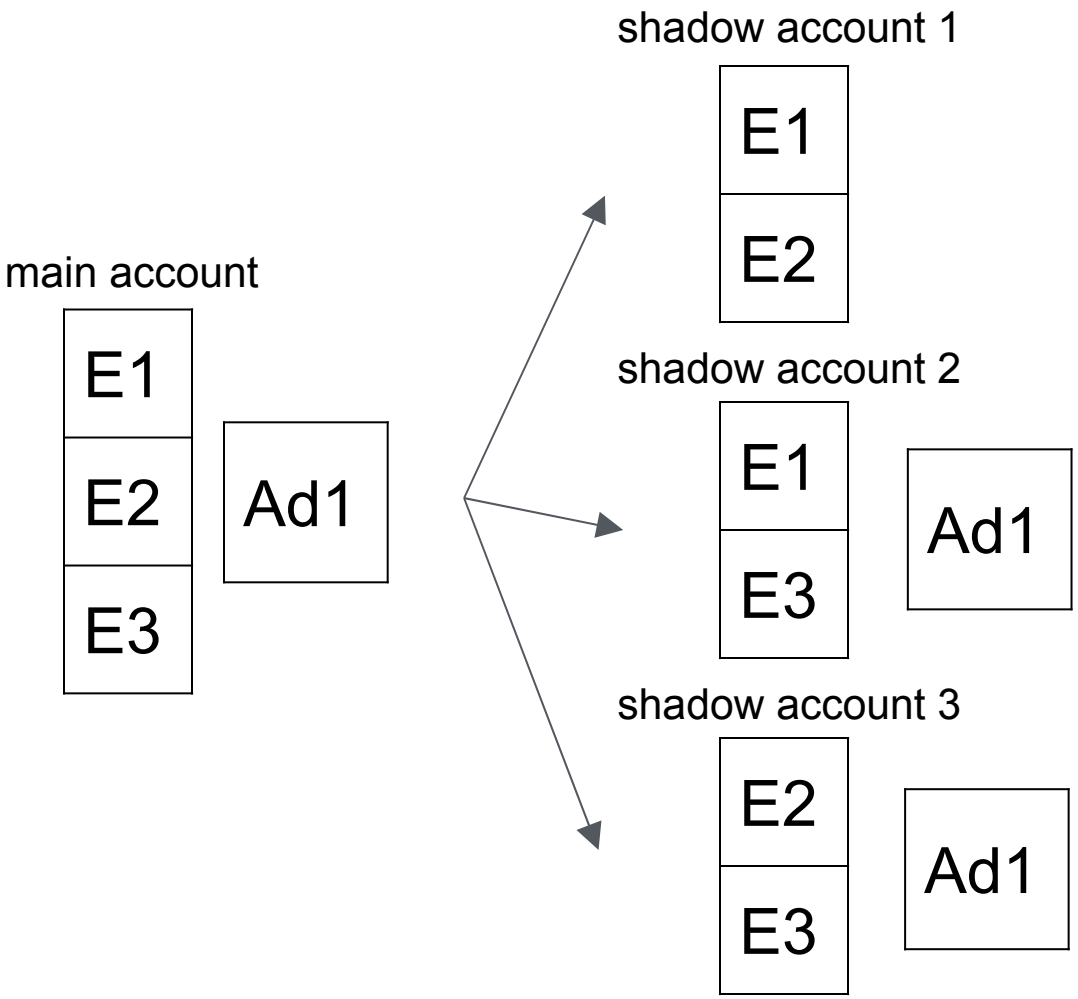
Example



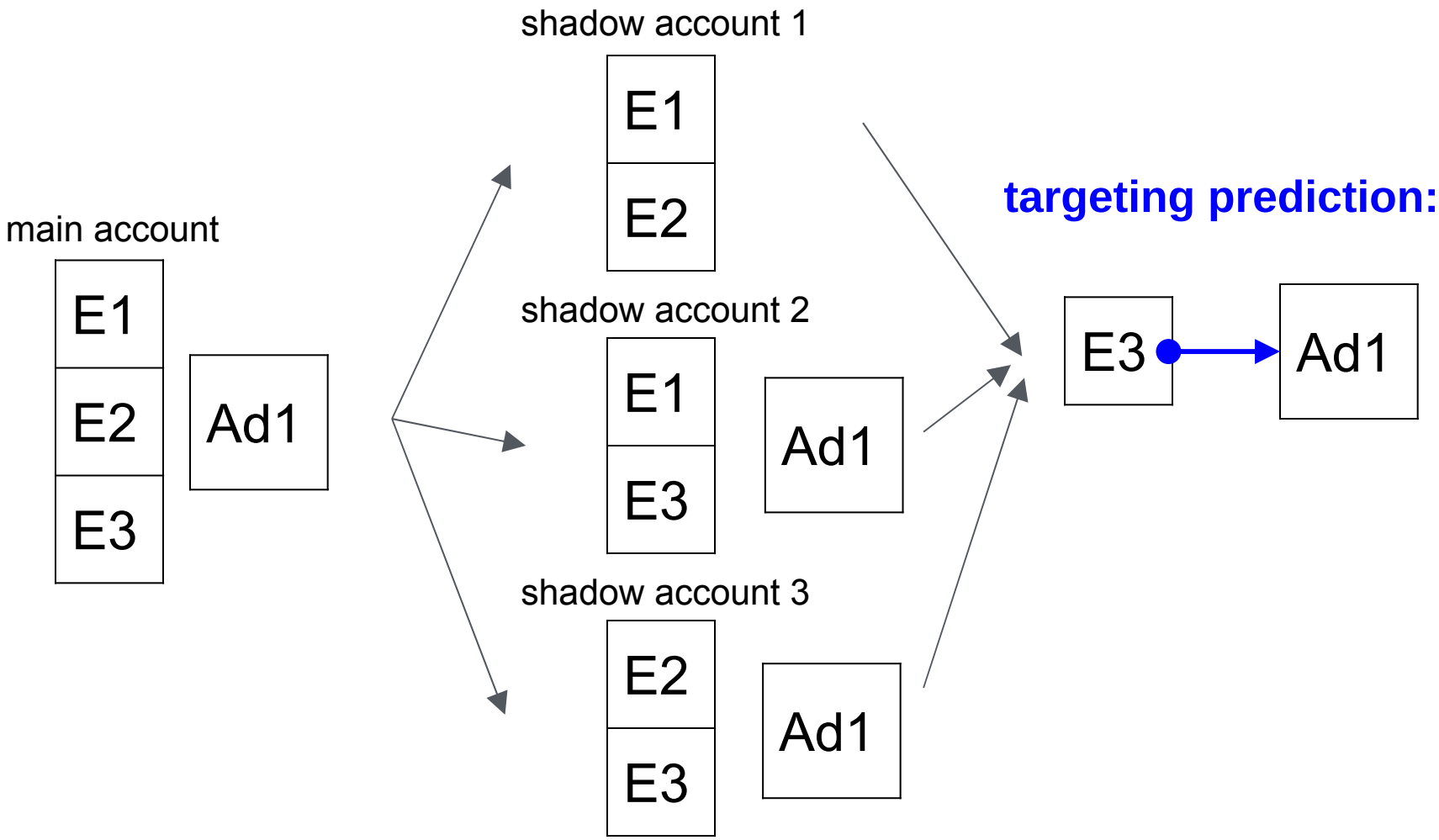
Example

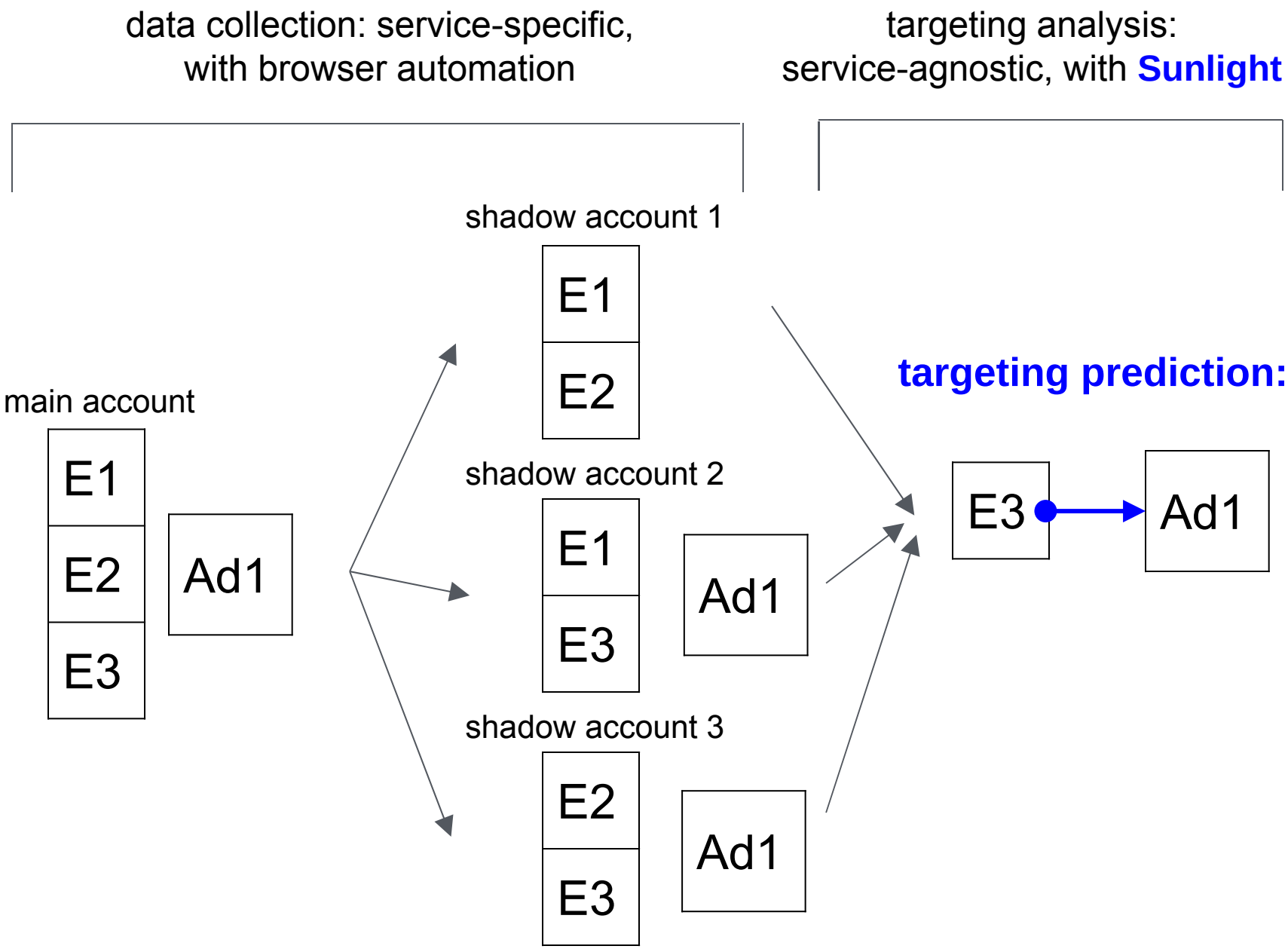

main account
E1
E2
E3
Ad1
shadow account 1
E1
E2
shadow account 2
E1
E3
shadow account 3
E2
E3

Example

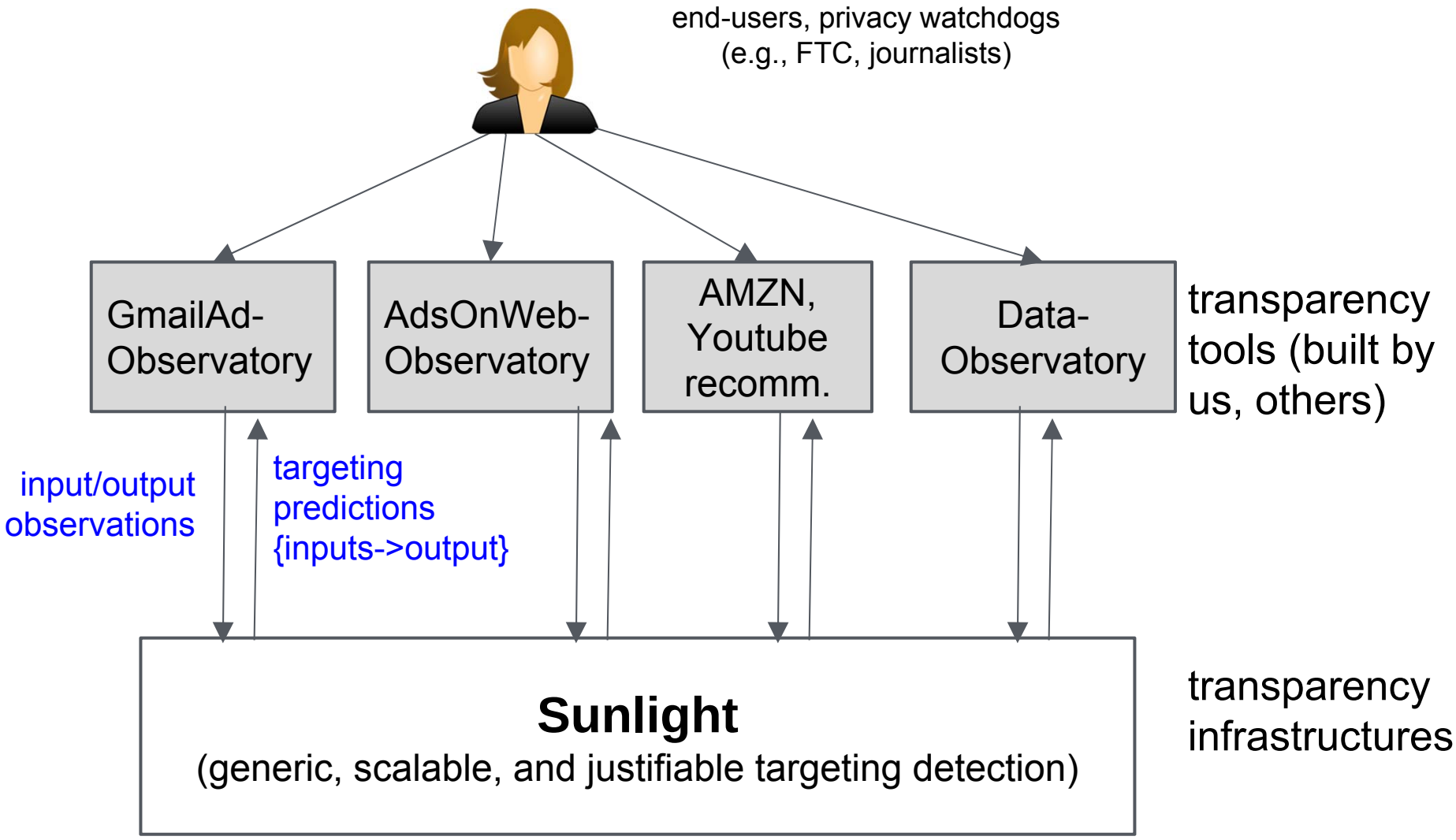

main account
E1
E2
E3
Ad1
shadow account 1
E1
E2
shadow account 2
E1
E3
Ad1
shadow account 3
E2
E3
Ad1

Example


main account
E1
E2
E3
Ad1
shadow account 1
E1
E2
shadow account 2
E1
E3
Ad1
shadow account 3
E2
E3
Ad1
targeting prediction:
E3
Ad1


data collection: service-specific,
with browser automation
targeting analysis:
service-agnostic, with Sunlight
main account
E1
E2
E3
Ad1
shadow account 1
E1
E2
shadow account 2
E1
E3
Ad1
shadow account 3
E2
E3
Ad1
targeting prediction:
E3
Ad1

Transparency solutions



Sunlight goals

Genericity

We assume that a small set of inputs is used to produce each output. Our goal is to discover the *correct* input combination.

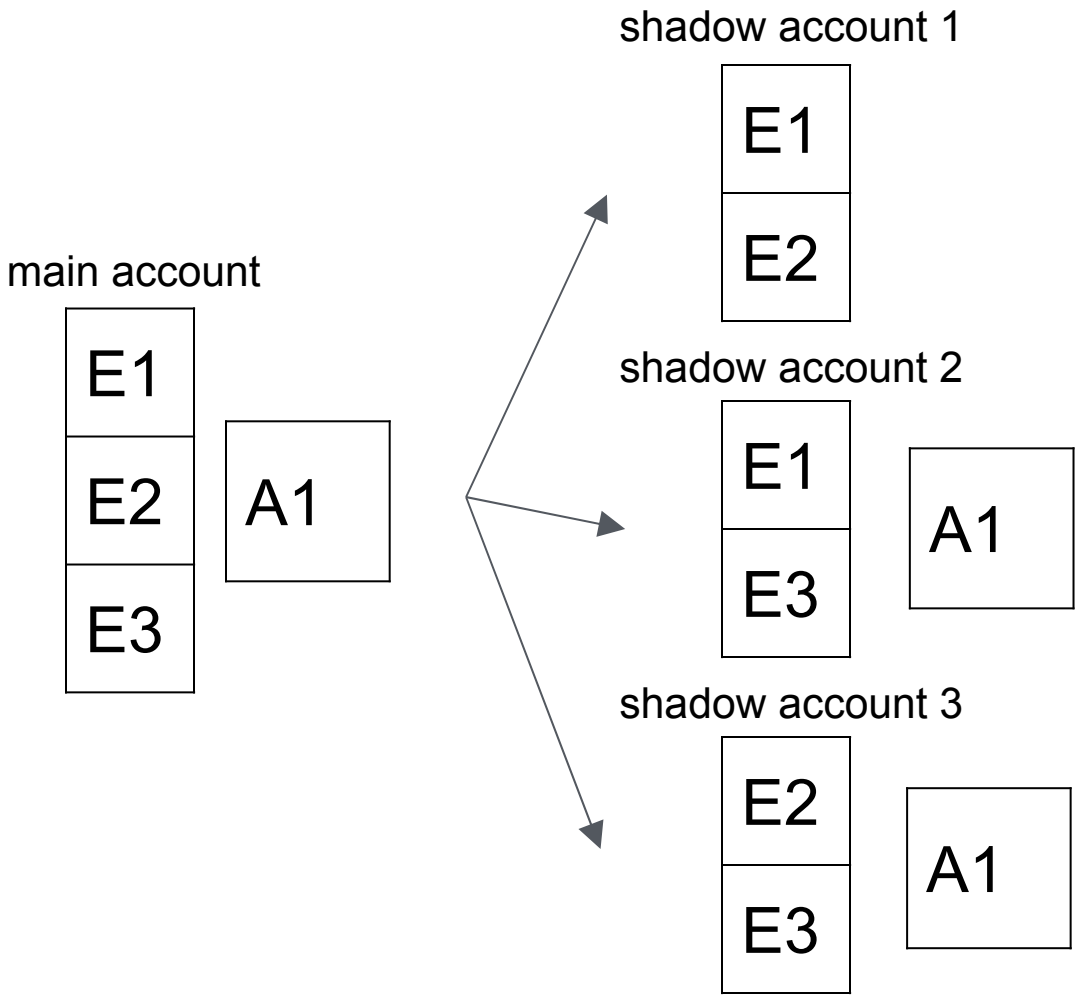
Scalability

Detect targeting of many outputs on many inputs w/ limited resources.

Precision

Targeting predictions must be statistically justified. Our goal is to detect as many *true* predictions as possible.

The scalability challenge



- To detect targeting on combinations of the inputs, will we need shadow profiles for all combinations???

Scalable targeting detection

- **Theorem:** *Under sparsity assumptions, for any ε > 0 there exists an algorithm that requires C x log(N) accounts to correctly identify the inputs of a targeted output with probability (1 − ε). N is the number of inputs.*

- Key insight: rely on sparsity properties (like compressed sensing).

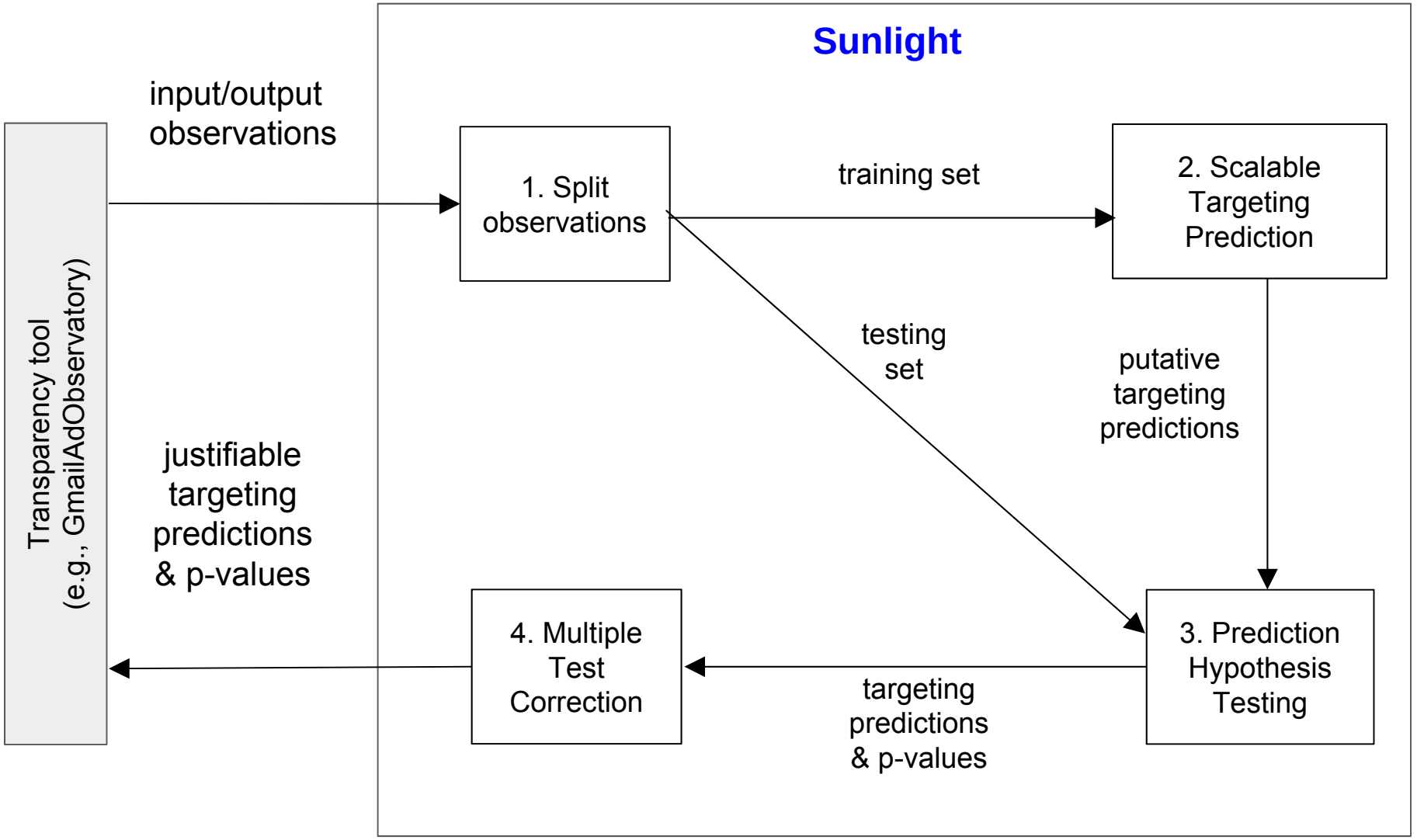
- Sunlight supports several sparse detection algorithms, including sparse regressions with Lasso.

Justifiable targeting predictions

- Sparse algorithms only guarantee asymptotic correctness of the targeting predictions.
- We need correctness assessment for each targeting prediction.

- Solution: hypothesis testing.
  - Provides quantification of statistical significance of each targeting association (a p-value).
  - p-value gives knob for precision/recall tradeoff.

Architecture



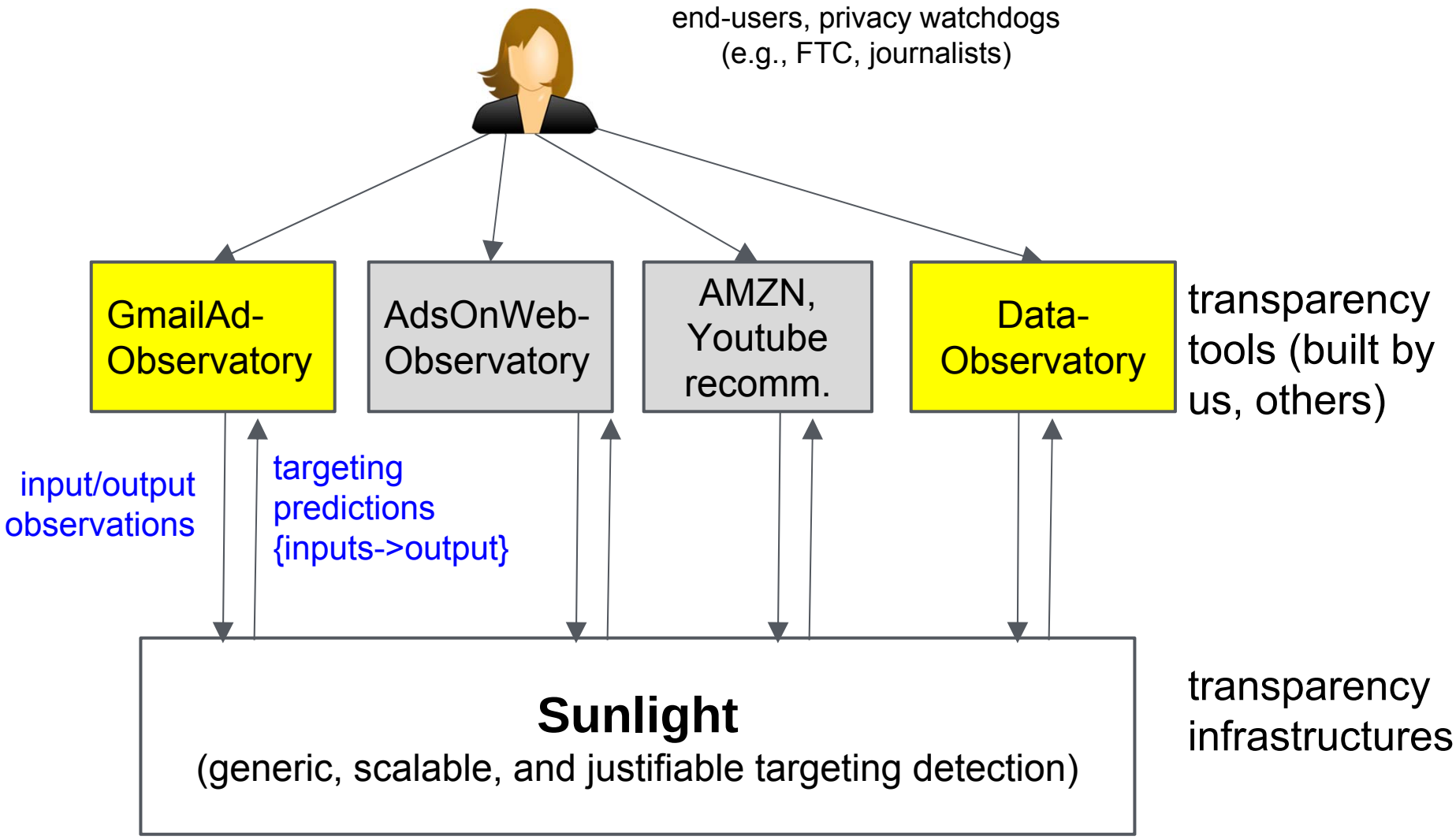
What we get in the end

If during data collection we randomly assign our inputs independently of any other variable, Sunlight's associations will have a causal interpretation (not just correlation).

However, Sunlight cannot explain how this targeting happens.

E.g.: What player in the ecosystem is responsible? Is it a human intervention or an algorithmic decision? Is it intentional or not?

Transparency tools



Tool 1: GmailAdObservatory

- Service to study targeting of Gmail ads on users' emails.
  - Meant for researchers and journalists.
- How it works:
  - Researcher supplies a set of emails.
  - GmailAdObservatory uses a set of Gmail accounts to send emails to a separate set of Gmail accounts (the shadows).
  - It then collects ads periodically.
  - Uses Sunlight to detect targeting for each collected ad.
- We ran a 33-day pilot study and we found violations of Google privacy statements.

Google privacy FAQ

http://support.google.com/mail/answer/6603

Privacy, Transparency and Choice

[...]

Only ads classified as Family-safe are displayed in Gmail. We are careful about the types of content we serve ads against. For example, Google may block certain ads from running next to an email about catastrophic news. We will also not target ads based on sensitive information, such as **race**, **religion**, **sexual orientation**, **health**, or **sensitive financial categories**.

"We will also not target ads based on sensitive information, such as race, religion, sexual orientation, **health**, or sensitive financial categories."

	email subject & text	ads Title, url & text	Results
General Health	**Affordable** afforable care [...] **(OR)** **Nursing** nursing home [...]	**Illinois Senior Living** www.cottagesofnewlenox.com Assisted Living for Seniors in New Lenox [...]	p-value = 0.03 103 impressions in 36 profiles 28% in context
	Alzheimer Alzheimer Alzheimer	**1/3 of Seniors 65+ Fall** jacuzzi-walk-in-tubs.com/Safety Help Eliminate the Fear of Falling in the Bathroom [...]	p-value = 0.01 21 impressions in 8 profiles 100% in context
	Depressed depression **(OR)** **Anxious** anxious anxiety	**Is He A Cheater?** spokeo.com/Cheating-Spouse-Search Enter His Email Address. Find Pics & Profiles From 70+ Social Networks.	p-value = 0.03 1179 impressions in 52 profiles 20% in context
	Cancer advice How did you cope with cancer in your familly? What an aweful disease!	**The Business of Wellness** healthmediagroup.blogspot.com What my doctor can learn from my Shoe Shine Man [...]	p-value = 0.04 380 impressions in 28 profiles 91% in context

"We will also not target ads based on sensitive information, such as race, religion, sexual orientation, health, or **sensitive financial categories**."

	email subject & text	ads Title, url & text	Results
Sensitive Financial	**Unemployed** lazy unemployed	**Easy Auto Financing** www.midsouthautoloans.com Need a quick car loan? We work with credit issues	p-value = 0.006 161 impressions in 24 profiles 8% in context
	Payday payday loan	**Fast Cash Loan Online.** www.checkintocash.com Apply Now. Takes Only 5 Minutes. It's as Easy as 1,2,3.	p-value = 0.007 198 impressions in 10 profiles 6% in context

Notice the extremely low in-context impressions -- the most obscure form of targeting.

Tool 2: DataObservatory
(work in progress)

- Discovers personalization on arbitrary websites *without* any a-priori specification of targeted outputs.

- How it works (in progress!):
  - Visits a website from the vantage point of multiple user profiles with differentiated inputs.
  - Compares various versions of each page by comparing DOM trees.
  - Uses Sunlight to detect how differences are targeted on the inputs.

Ex: Personalization on Booking.com

New York, NY

Booking.com

An outstanding value on these dates.

Save to a list 831 831 people added this property to their wish list

Santa Monica, Los Angeles, CA

There are 8 people looking at this hotel.

Good 7.5 702 reviews

Parking

Last booked: 43 minutes ago

people	Available room types	Availability			
Max people: 2	Double Room	We have	2	rooms left!	$194 $177

See all rooms available ▶

Select your room

Berlin, Germany

Booking.com

An outstanding value on these dates.

Save to a list 831 831 people added this property to their wish list

Santa Monica, Los Angeles

There are 8 people looking at this hotel.

Good 7.5 702 reviews

Last booked: 43 minutes ago

people	Available room types	Availability			
Max people: 2	Double Room	Only 2 rooms left on our site!	US$221	-	US$202

See all rooms available ▶

Select your room

or Save to a list You bookmarked this

Summary

clone me on github
(http://columbia.github.io/sunlight/)

- We are building the first generic and broadly applicable transparency tools that enable **oversight at scale.**
  - **Sunlight** reveals the causes of targeting from controlled experiments with many inputs.
  - **DataObservatory** reveals personalization on arbitrary pages.
- Tools can be used to **study complex targeting phenomena**.
  - E.g.: ad targeting, price tuning, personalization based on tracking, cross-device targeting, remote fingerprint-based tracking, how children are targeted, etc.
- **Open challenge**: avoid the pitfalls of controlled experiments.

Demo page

http://www.cs.columbia.edu/~yannis/stable/booking_com_us_ger_LA_feb01-feb02_exp/Visualization.html

NOTE: This is very much in-progress work, but the demo illustrates the kinds of functionality the DataObservatory will provide.

Daniel Hsu

Columbia University

Discovering Unwarranted Associations in Data-Driven Applications with the FairTest Testing Toolkit

Co-authors: Vaggelis Atlidakis, Roxana Geambasu (Columbia University); Florian Tramèr, Jean-Pierre Hubaux, Huang Lin (École Polytechnique Fédérale de Lausanne); Ari Juels (Cornell Tech)

FairTest:

discovering unwarranted associations in data-driven applications

Florian Tramèr[#], Vaggelis Atlidakis*, Roxana Geambasu*, Daniel Hsu*,
Jean-Pierre Hubaux[#], Mathias Humbert[#], Ari Juels[@], Huang Lin[#]

[#]École Polytechnique Fédérale de Lausanne, *Columbia University, [@]Cornell Tech

“Unfair” associations + consequences

Google Photos labeled black people 'gorillas'

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

SAN FRANCISCO — Google has apologized after its new Photos application identified black people as "gorillas."

On Sunday Brooklyn programmer Jacky Alciné tweeted a screenshot of photos he had uploaded in which the app had labeled Alcine and a friend, both African American, "gorillas."

Yontan Zunger, an engineer and the company's chief architect of Google+, responded swiftly to Alciné on Twitter: "This is 100% Not OK." And he promised that Google's Photos team was working on a fix.

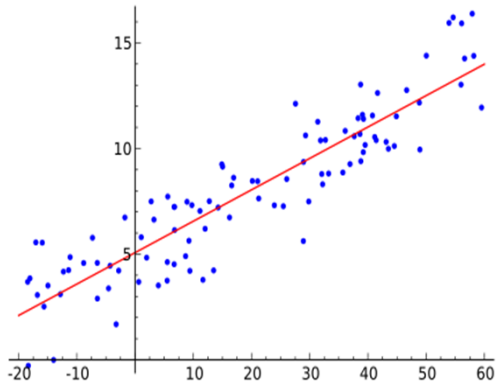
Journal's testing also showed that areas that tended to see the discounted prices had a higher average income than areas that tended to see higher prices.

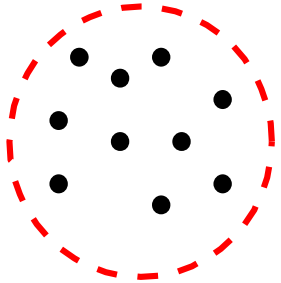
These are **software bugs**: need to *actively test for them* and *fix them (i.e., debug)* in data-driven applications…
just as with functionality, performance, and reliability bugs.

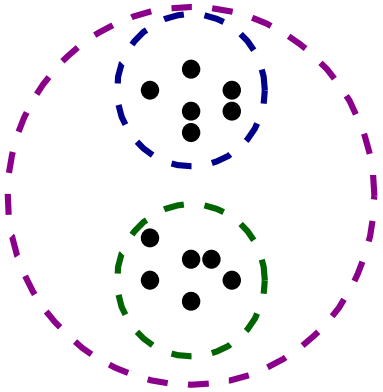
Limits of preventative measures

What doesn't work:

- Hide protected attributes from data-driven application.
- Aim for statistical parity w.r.t. protected classes and service output.



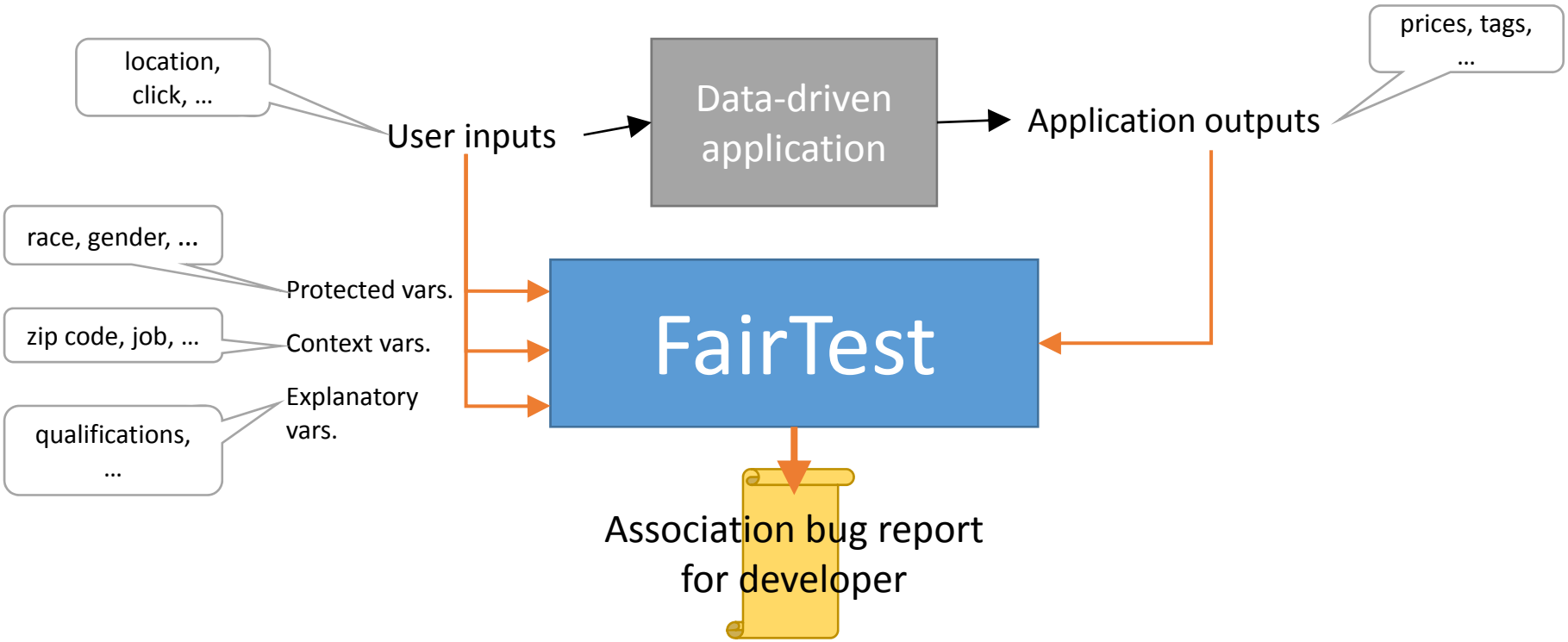




Foremost challenge is to even detect these unwarranted associations.

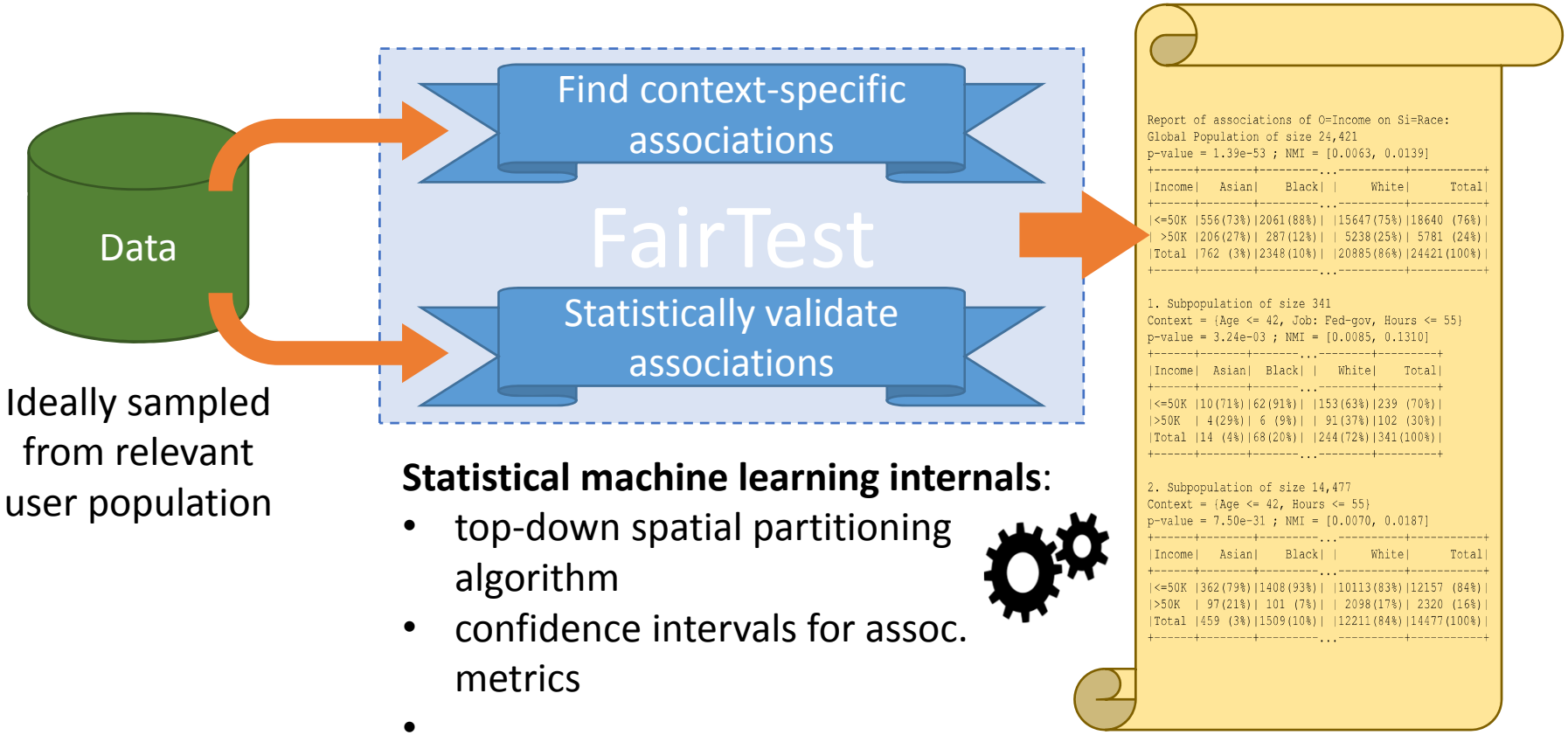
FairTest: a testing suite for data-driven apps

- Finds context-specific associations between protected variables and application outputs
- Bug report ranks findings by assoc. strength and affected pop. size



A data-driven approach

Core of FairTest is based on statistical machine learning



Ideally sampled from relevant user population

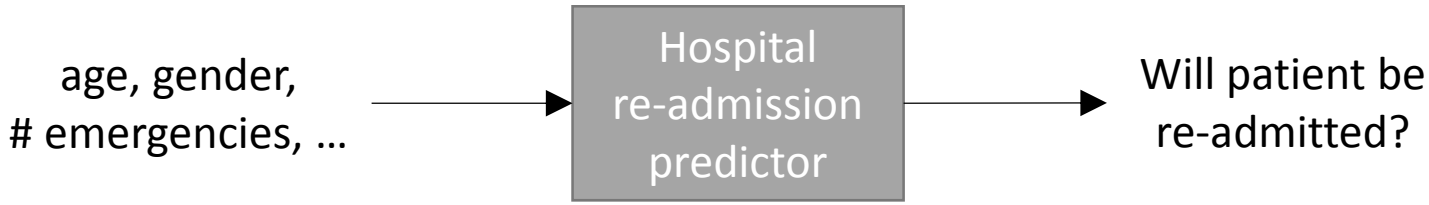
Statistical machine learning internals:

- top-down spatial partitioning algorithm
- confidence intervals for assoc. metrics
- …

Example: health care application

Predictor of whether patient will visit hospital again in next year
(from winner of 2012 Heritage Health Prize Competition)

FairTest's finding: significant contexts exhibiting strong association between age and prediction error rate.

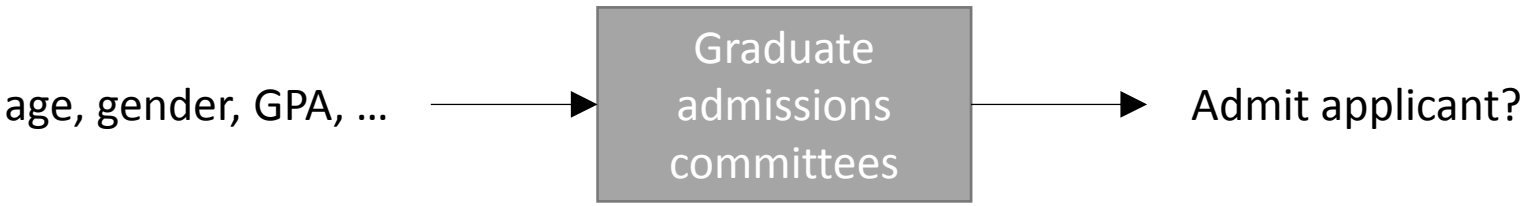


Association may translate to quantifiable harms (e.g., if app is used to adjust insurance premiums)!

Example: Berkeley graduate admissions

Admission into UC Berkeley graduate programs
(Bickel, Hammel, and O'Connell, 1975)

Bickel *et al*'s (and also FairTest's) findings: gender bias in admissions at university level, but mostly gone after conditioning on department



FairTest helps developers understand & evaluate potential association bugs.

Closing remarks

- **Other applications studied using FairTest** (http://arxiv.org/abs/1510.02377):
  - Image tagger based on deep learning (on ImageNet data)
  - Simple movie recommender system (on MovieLens data)
  - Simulation of Staple's pricing system
- **Other features in FairTest**:
  - Exploratory studies (e.g., find image tags with offensive associations)
  - Adaptive data analysis (preliminary) – i.e., statistical validity with data re-use
  - Integration with SciPy library

Developers need better statistical training and tools to make better statistical decisions and applications.

Thanks!

Discussion of Session 3

Discussants:

- **Dan Salsburg,** Federal Trade Commission
- **James C. Cooper,** George Mason University School of Law
- **Deirdre K. Mulligan,** University of California, Berkeley

Presenters:

- **Michael Carl Tschantz,** University of California, Berkeley **& Anupam Datta,** Carnegie Mellon University
- **Roxana Geambasu,** Columbia University
- **Daniel Hsu,** Columbia University