

# AdFisher

## Information Flow Experiments on Ad Privacy Settings

Michael Carl Tschantz  
International Computer Science Institute

Amit Datta, CMU  
Anupam Datta, CMU  
(sometimes) Jeannette M. Wing, MSR

[www.cs.cmu.edu/~mtschant/ife](http://www.cs.cmu.edu/~mtschant/ife)

ADVERTISEMENT



Enter the DEBIT TO WIN IT™ contest.

[Learn More >](#)

# THE TIMES OF INDIA China



The Times of India

Search

Advanced Search »

[Home](#) [World](#) [US](#) [Pakistan](#) [South Asia](#) [UK](#) [Europe](#) **China** [Middle East](#) [Rest of World](#) [Mad, Mad World](#) [Videos](#)

You are here: Home » World » China

4

comments

1 like

Like

Share

77

Tweet

0

g+1

1

Share

Share More

Email

Print

AA



*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.*

RELATED

another day."

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

Reuters | Dec 12, 2014, 08.39 AM IST

### Time to Hug\* by Huggies®

**Parenting info, Prizes and Offers! To Meet new Moms like You. : [www.facebook.com/TimetoHug](http://www.facebook.com/TimetoHug)**

Ads by Google

[READ MORE »](#)
[Hong Kong Protesters | 'We'll Be Back'](#) |
 [Hong Kong](#) |
 [CY Leung](#)

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

Connect with us



LIVE  
1ST TEST

AUS 517/7 dec 57.2 0v & 207/3 444 IND

Day 4: 3rd Session - Australia lead by 280 runs

Subscribe to our Newsletters



- Top News (Daily)
- Tech News (Daily)

2

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

[knowmydepression.com/antidepressant](http://knowmydepression.com/antidepressant) ▾

Visit For Treatment Info & Facts.



health insurance

Web

Images

Maps

Shopping

More ▾

Search tools

About 674,000,000 results (0.29 seconds)

Ads related to **health insurance** ⓘ

[\*\*Health Insurance Quotes\*\*](#) 1 (877) 688 2154

[www.ehealthinsurance.com/](http://www.ehealthinsurance.com/)

Nation's #1 **Health Insurance** Site. Compare the Best Prices Instantly!

eHealthInsurance has 170 followers on Google+

[Get California Health Insurance](#)

[Apply For Medical Insurance In CA](#)

[Shop Individual & Family Plans](#)

[Compare Blue Shield of California](#)

[\*\*Low Cost CA Health Plans\*\*](#) 1 (800) 446 8171

[california.healthcare.com/](http://california.healthcare.com/)

Blue Cross, Kaiser, CIGNA, United. Search Top CA Plans, from \$49/month

[Individual Plans](#) - [Family Plans](#) - [Catastrophic Coverage](#) - [Self Employed](#)

[\*\*Low Cost Health Insurance\*\*](#) 1 (877) 485 3896

[www.gohealthinsurance.com/](http://www.gohealthinsurance.com/)

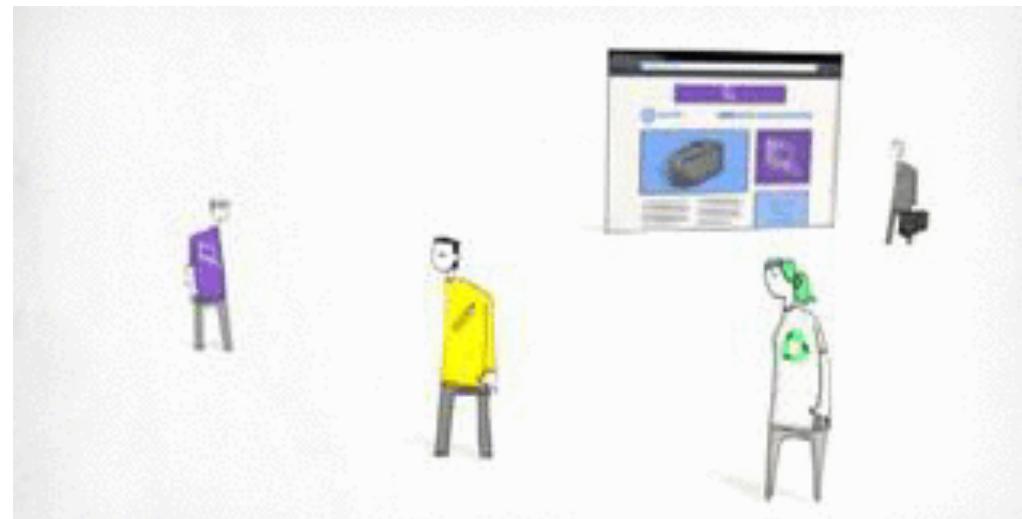
Blue Cross, United, Aetna, & CIGNA. See Rates by Your State or Zip!

[View Quotes for Health Insurance](#) - [Compare Health Insurance Rates](#)

[\*\*Health Insurance\*\* - Find Affordable \*\*Health Insurance\*\* Plans and Buy ...](#)

# Google AdSense

- Website owner (Publisher) wants to earn money
- Registers website with Google AdSense
- Google runs word analysis on website and figures out keywords related to page



# Google AdWords

## Getting started



1. Choose your budget



2. Create your ads



3. Select keywords that match your ads to potential customers



4. Enter your billing information.

### Keywords

[Select keywords](#)

Enter one keyword per line.

dog toy  
puppy toy  
pup

Sample keywords.  
No sample keywords



» [Advanced option: match types](#)

**Important note:** We cannot guarantee that these keywords will improve your campaign or ad. You are responsible for the keywords you select and for ensuring that your use of them complies with our [AdWords terms of service](#).

# Ad Targeting

Google automatically delivers ads that are **targeted** to your content or audience. We do this in several ways:

- **Contextual targeting**

Our technology uses such factors as keyword analysis, word frequency, font size, and the overall link structure of the web, in order to determine what a webpage is about and precisely match Google ads to each page.

- **Placement targeting**

With placement targeting, advertisers choose specific **ad placements**, or subsections of publisher websites, on which to run their ads. Ads that are placement-targeted may not be precisely related to the content of a page, but are hand-picked by advertisers who've determined a match between what your users are interested in and what they have to offer.

- **Interest-based advertising**

~~Interest-based advertising~~ enables advertisers to reach users based on their interests and demographics (e.g. 'sports enthusiasts'), and allows them to show ads based on a user's previous interactions with them, such as visits to advertiser websites. To complement interest-based advertising, [Ads Settings](#) lets users view and edit their interest categories.

# Google's Privacy Policy

When showing you tailored ads, we will not associate a cookie or anonymous identifier with sensitive categories, such as those based on race, religion, sexual orientation or health.

# 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



Search

## Google ads across the web ?



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

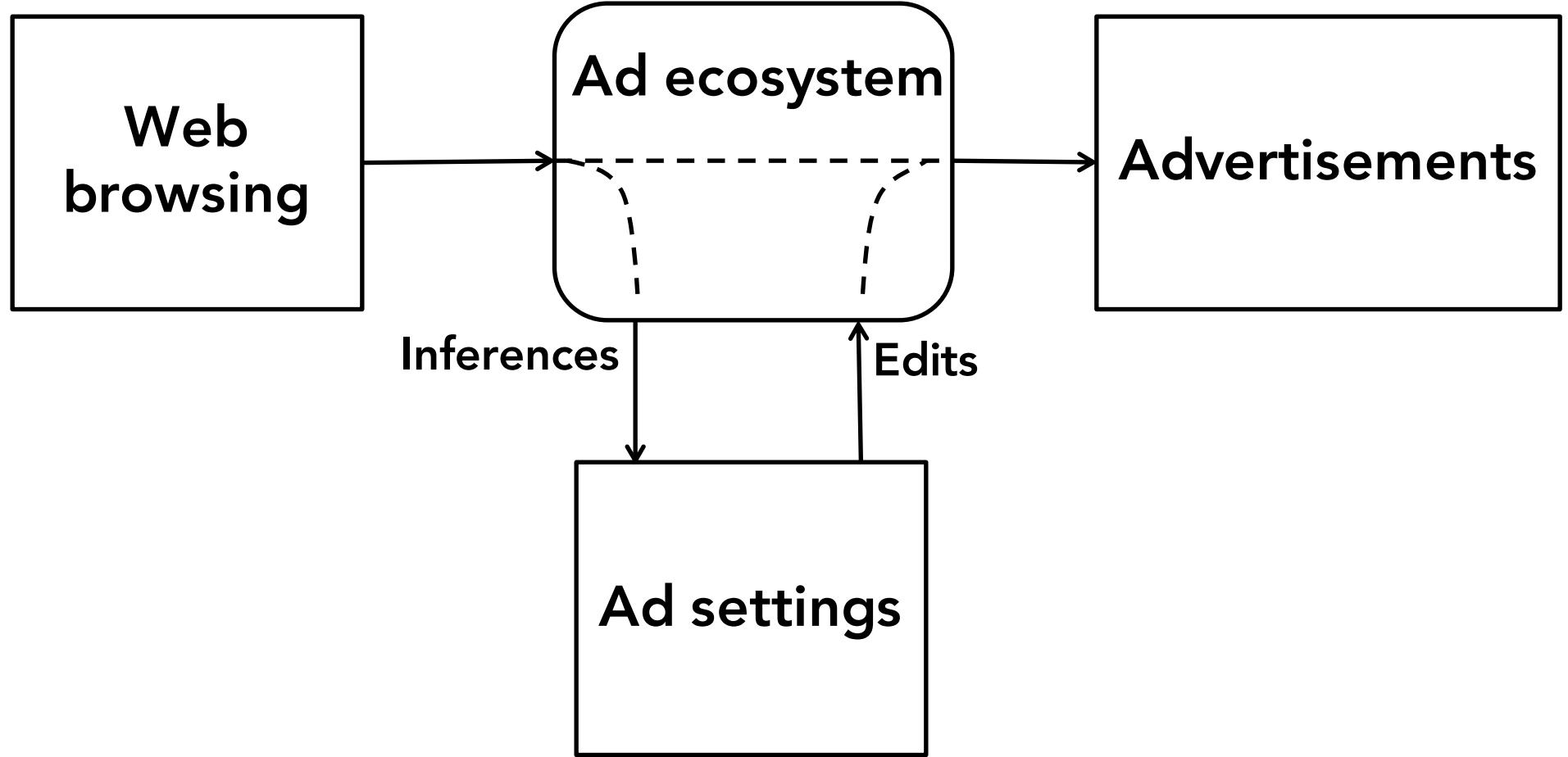
Air Travel, and 30 more [Edit](#)

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

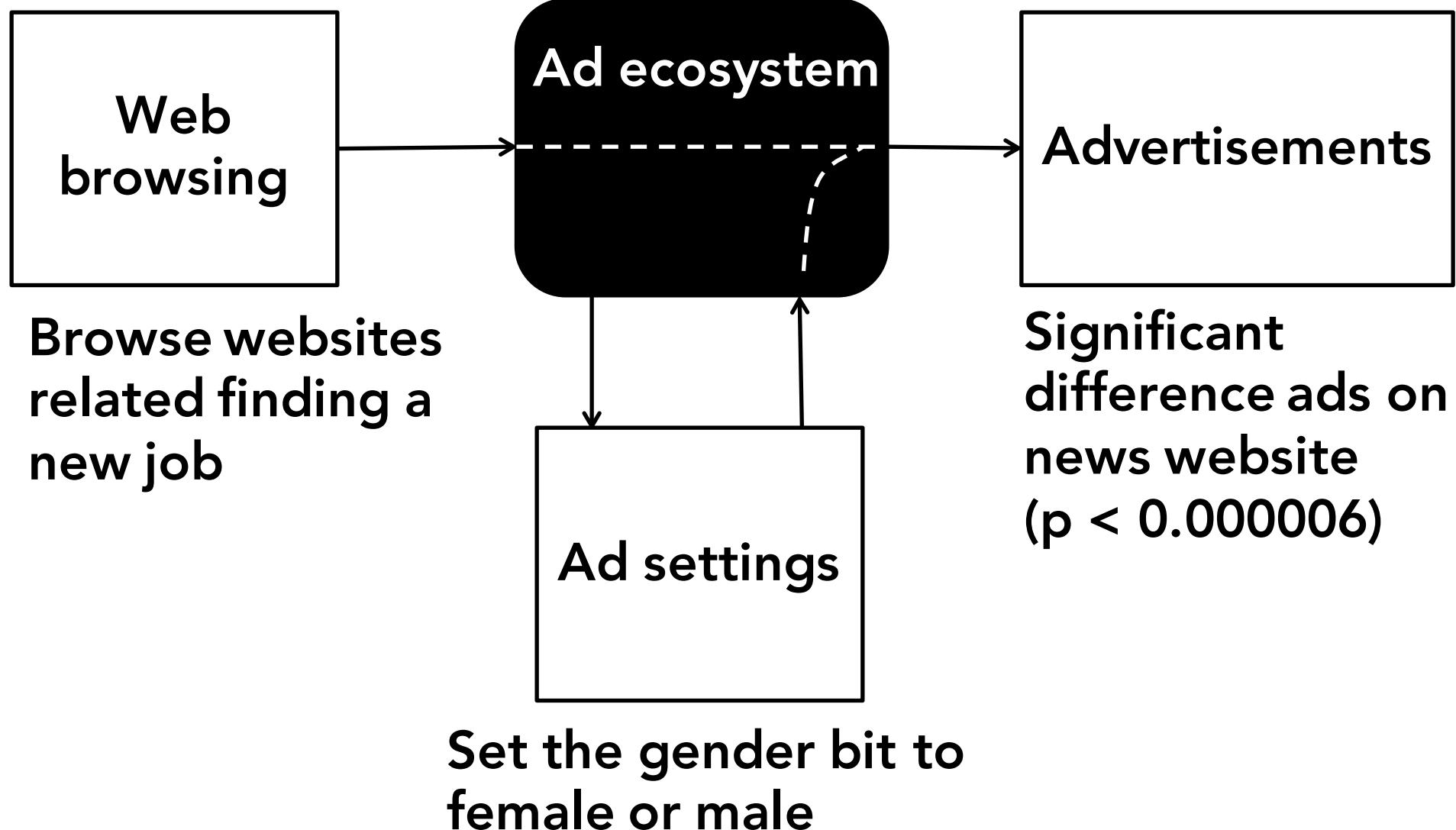
[Opt out](#) of *interest-based* Google ads across the web



# AdFisher

- Emulates users with fresh browser instances
- Randomized assignment
- Statistical analysis to find causal relations
- Open source: [github.com/tadatitam/info-flow-experiments](https://github.com/tadatitam/info-flow-experiments)

# Discrimination



# Discrimination Explanation

## Female Group

### Jobs (Hiring Now)

[www.jobsinyourarea.co](http://www.jobsinyourarea.co)

45 vs. 8

### 4Runner Parts Service

[www.westernpatoyotaservice.com](http://www.westernpatoyotaservice.com)

36 vs. 5

### Criminal Justice Program

[www3.mc3.edu/Criminal+Justice](http://www3.mc3.edu/Criminal+Justice)

29 vs. 1

## Male Group

### \$200k+ Jobs - Execs Only

[careerchange.com](http://careerchange.com)

311 vs. 1816

### Find Next \$200k+ Job

[careerchange.com](http://careerchange.com)

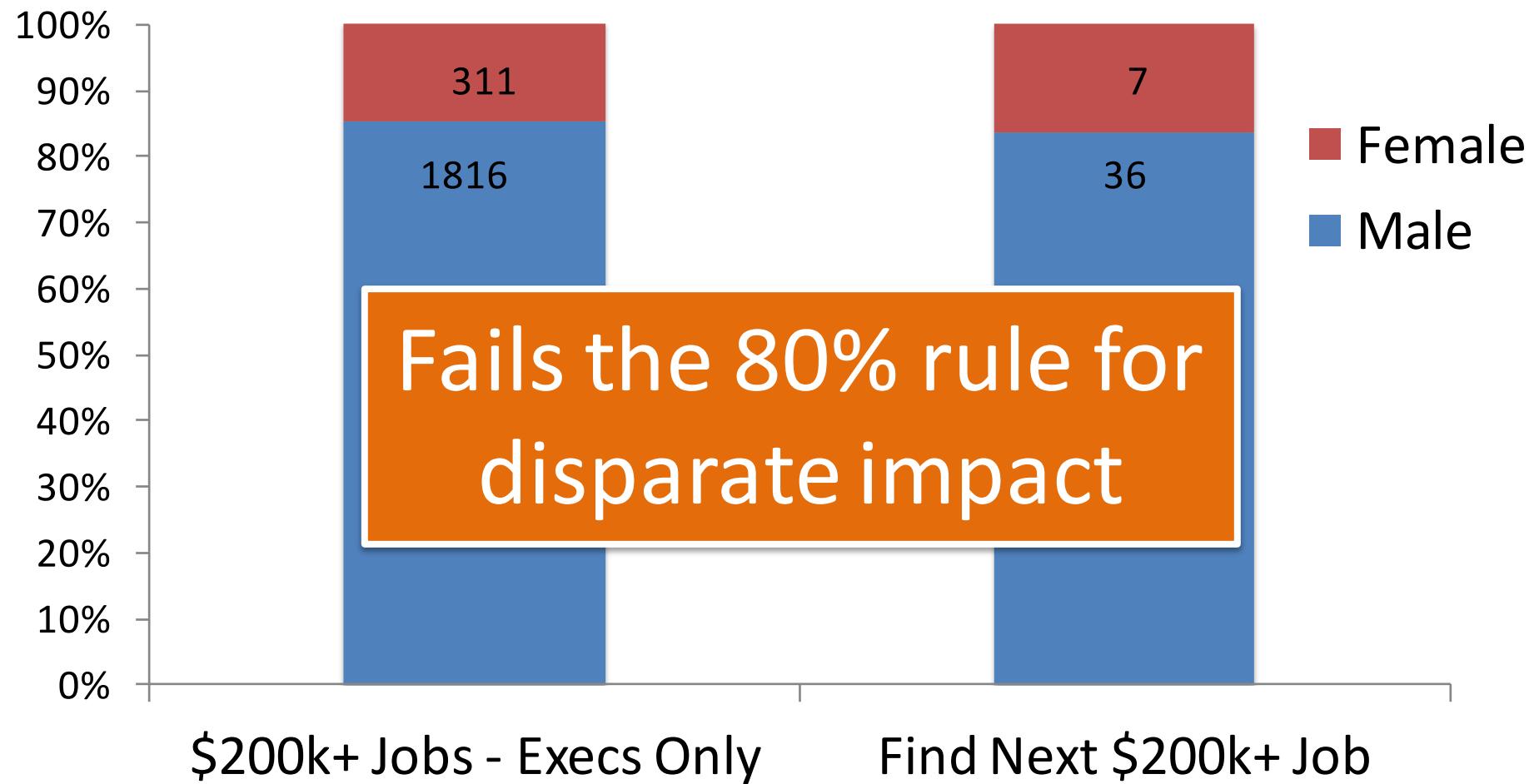
7 vs. 36

### Become a Youth Counselor

[www.youthcounseling.degreetleap.com](http://www.youthcounseling.degreetleap.com)

0 vs. 310

# Discrimination Explanation



# Findings and Non-Findings

- An instance of discrimination
- ~~Policy Violations~~
  - No claims that Google or anyone else violated any policies
- ~~Generalization~~
  - These results might not generalize to other settings
- ~~Blame~~
  - We do not assign blame

```
graph LR; User[User] -- Input --> Google[Google]; Google -- Output --> Advertisers[Advertisers]; Advertisers --- Barrett[The Barrett Group]; Advertisers --- Websites[Websites]; Advertisers --- Male[Male users]; Advertisers --- Female[Female users];
```

The Barrett Group

Other  
advertisers



**Input** →  
← **Output**

Google

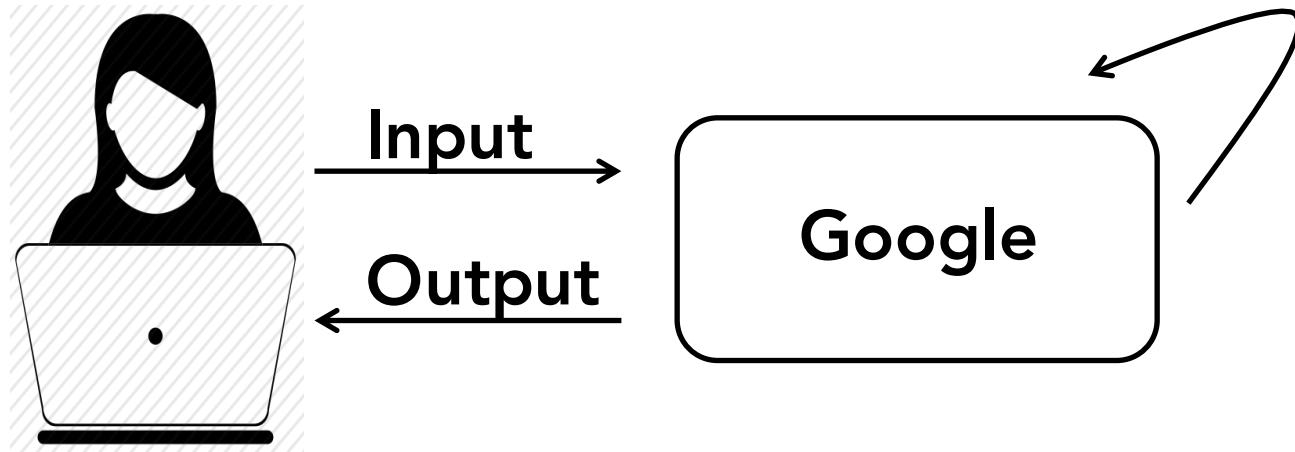
Websites

Male  
users

Female  
users

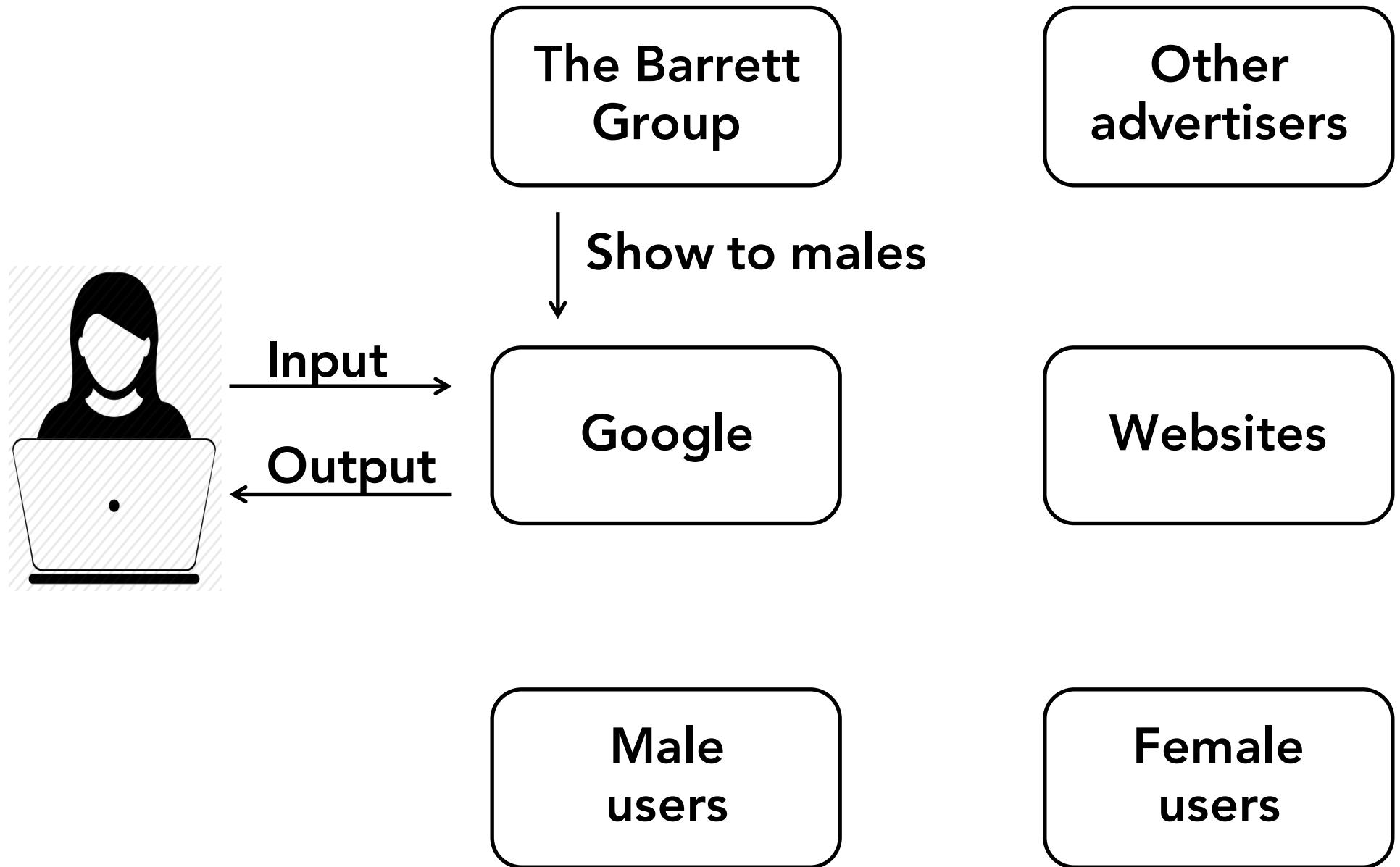
The Barrett  
Group

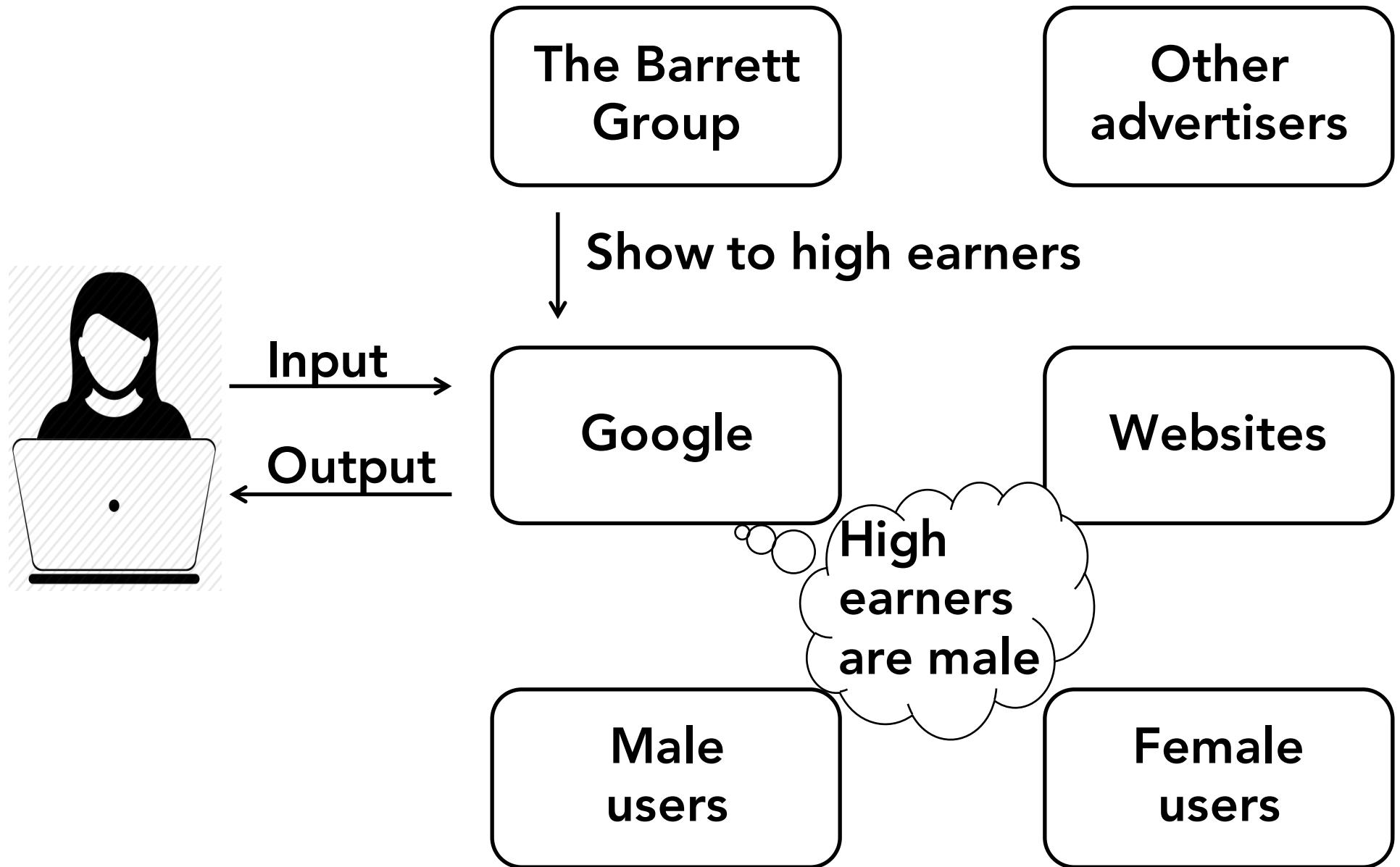
Other  
advertisers

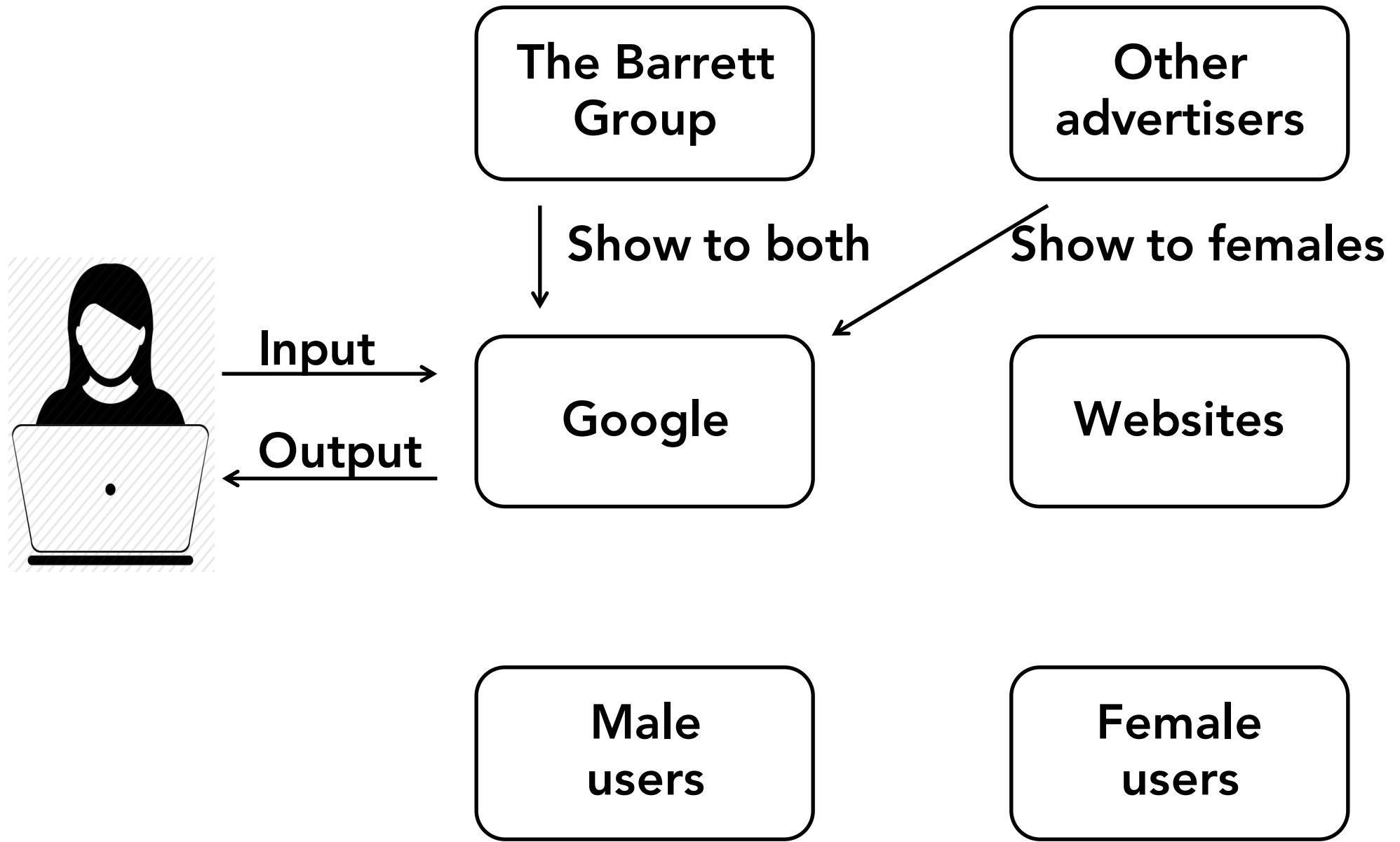


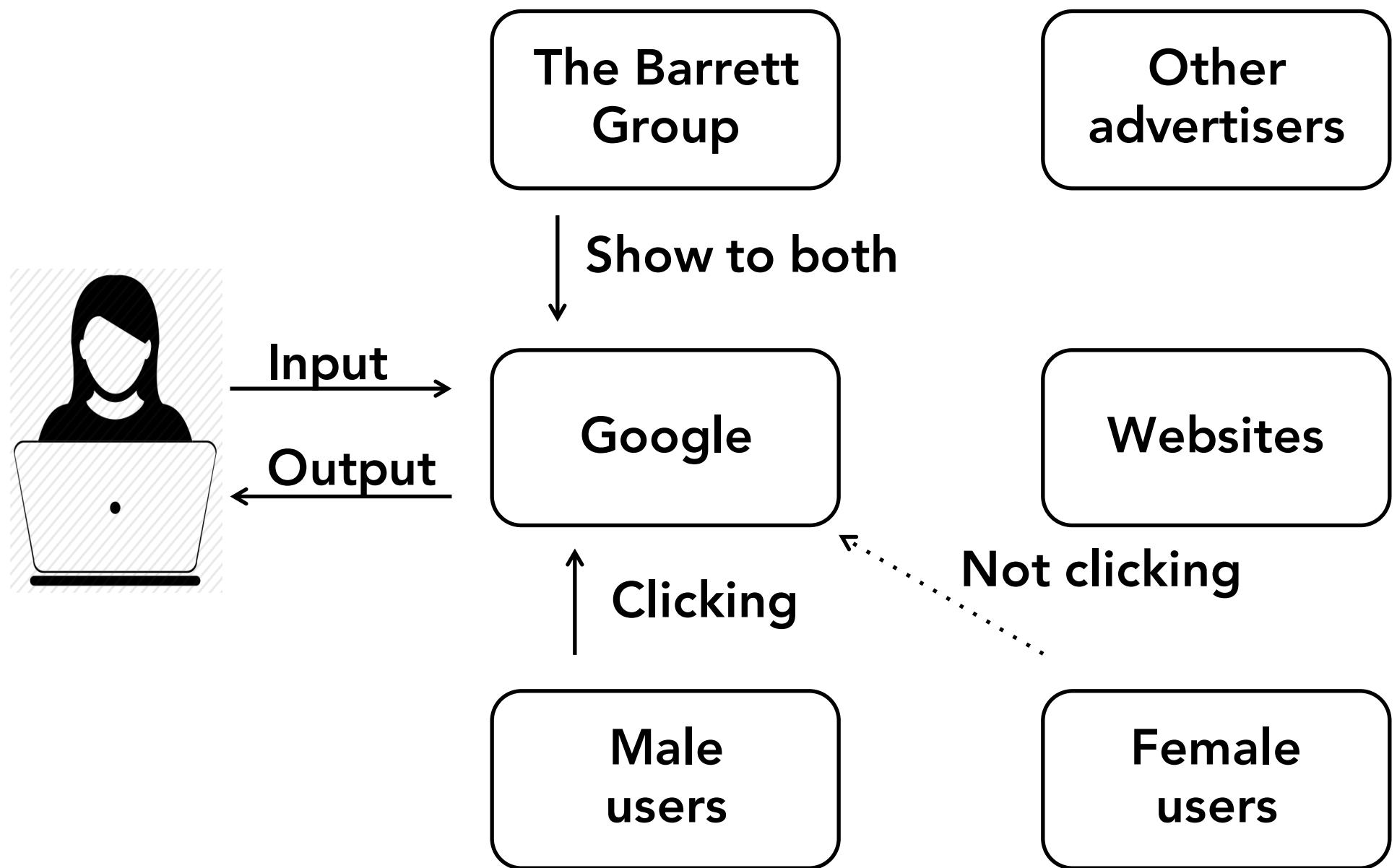
Male  
users

Female  
users









```
graph LR; User[User] -- Input --> Google[Google]; Google -- Output --> Male[Male users]; Google -- Output --> Female[Female users]; Google -- Output --> Websites[Websites]; Google -- "Not clicking other ads" --> Male; Google -- "Clicking other ads" --> Websites
```

The Bartlett Group

Other advertisers



**Input** →  
← **Output**

Google

Websites

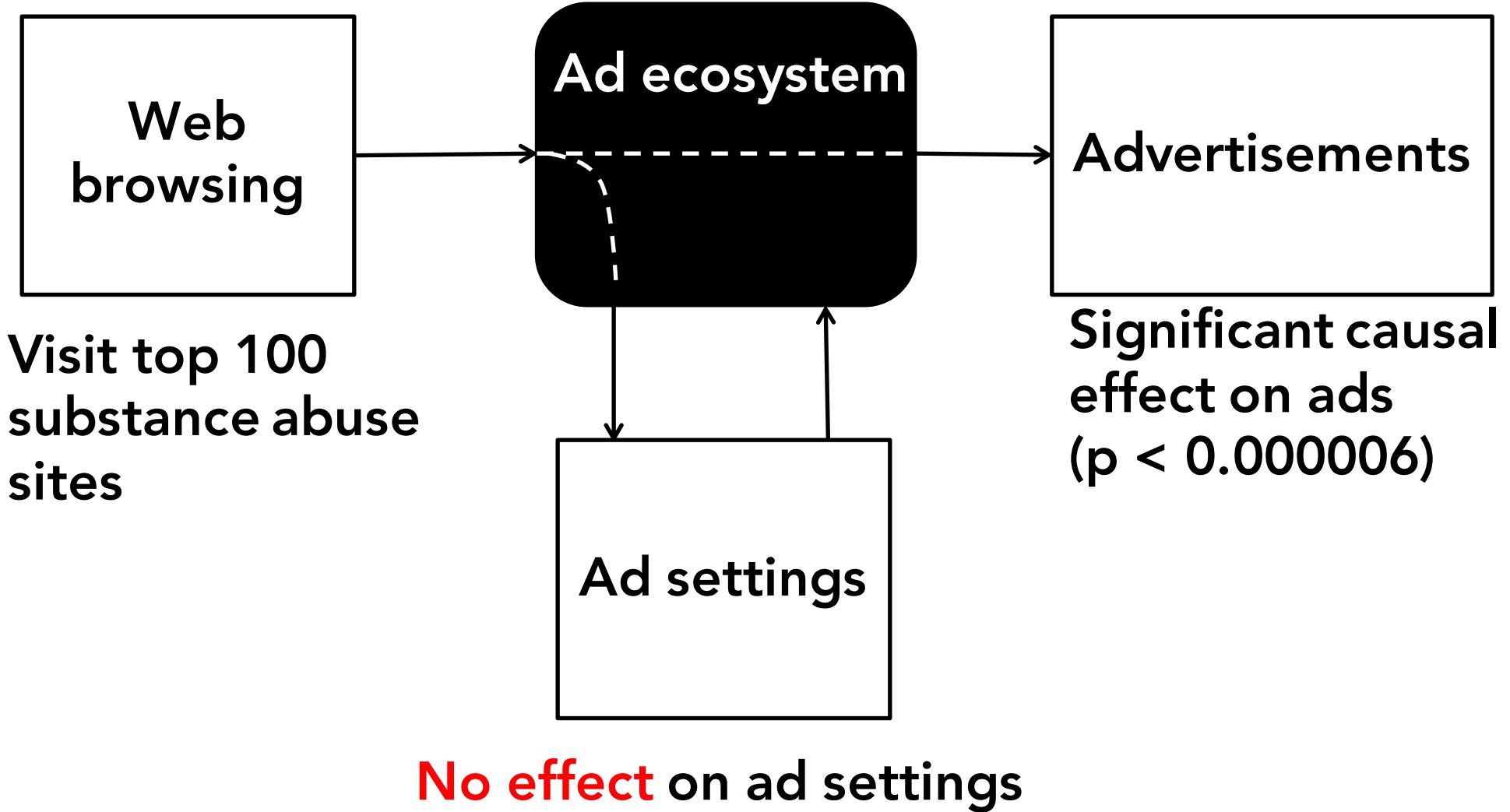
↑  
Not clicking  
other ads

Clicking other ads

Male  
users

Female  
users

# Transparency



# Transparency Explanations

## Substance Abuse Visitors

**The Watershed Rehab**

[www.thewatershed.com/Help](http://www.thewatershed.com/Help)

**2276 vs. 0**

**Watershed Rehab**

[www.thewatershed.com/Rehab](http://www.thewatershed.com/Rehab)

**362 vs. 0**

**The Watershed Rehab**

**(none)**

**771 vs. 0**

## Control Group

**Alluria Alert**

[www.bestbeautybrand.com](http://www.bestbeautybrand.com)

**0 vs. 9**

**Best Dividend Stocks**

[dividends.wyattresearch.com](http://dividends.wyattresearch.com)

**24 vs. 54**

**10 Stocks to Hold Forever**

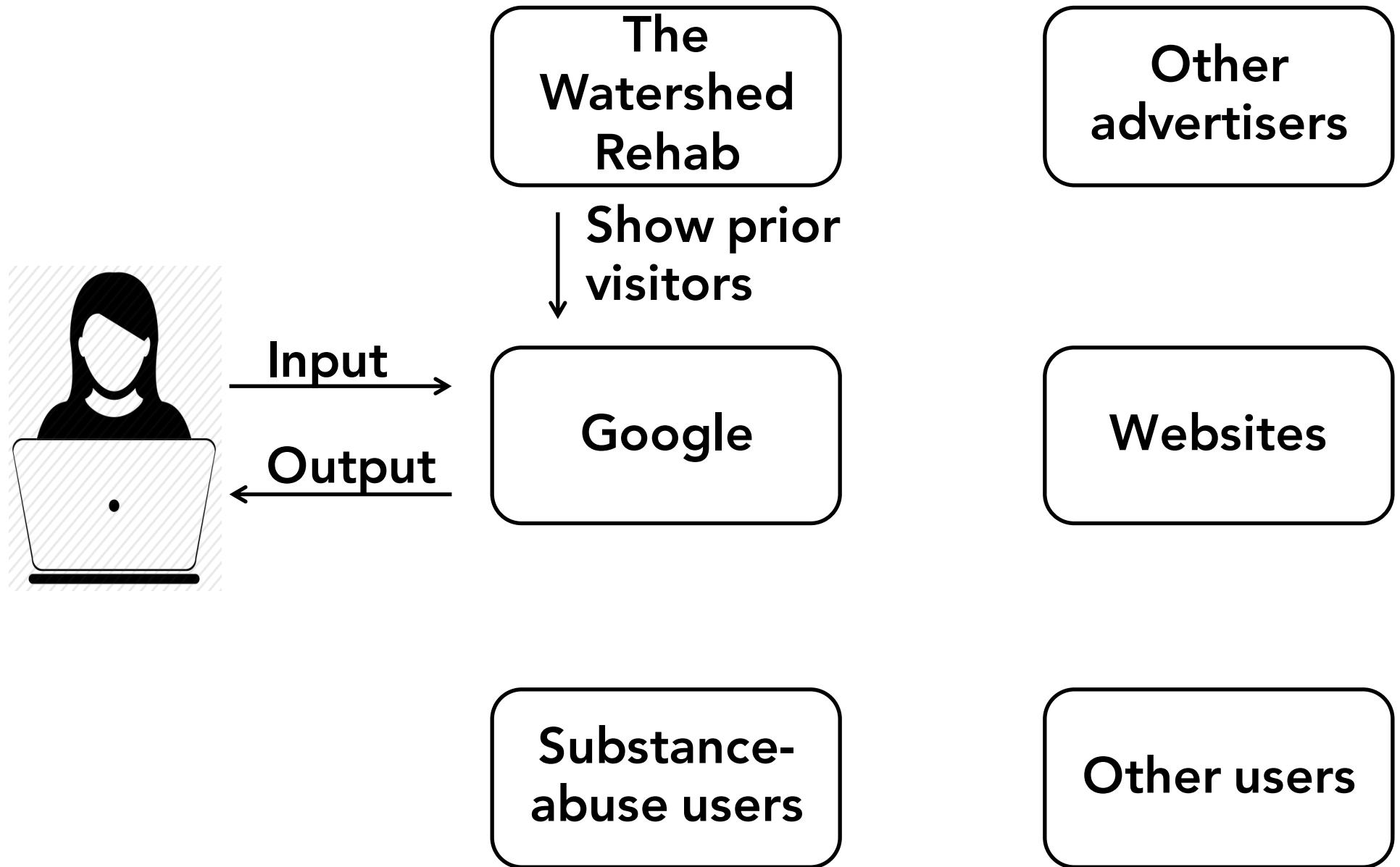
[www.streetauthority.com](http://www.streetauthority.com)

**76 vs. 118**

**The Watershed Rehab**

[www.thewatershed.com/Help](http://www.thewatershed.com/Help) - Drug & Alcohol Rehabilitation Call Today For Help Now!

Ads by Google



# Additional Notice on Ad Settings

## Ads Settings

Interests

Shooter Games [Edit](#)  
From your previous searches

Adventure Games, and 13 more.  
Based on the websites you've visited

We use interests from your activity on websites to tailor ads to you. [How it works](#)

Before

Interests

Air Travel, and 87 more [Edit](#)  
From your previous activity on Google

Adventure Games, and 11 more  
Based on the websites you've visited

Your activity on websites is used to serve interest based ads.

These categories can only control some of the Google ads you see. For instance, the interests listed here do not reflect ads selected based on a visit to a specific advertiser's page (remarketing) or ads selected by other advertising providers and delivered by Google's ad platforms. [How it works](#).

After

# Additional Notice on Ad Settings

## Interests

Air Travel, and 87 more [Edit](#)

From your previous activity on Google

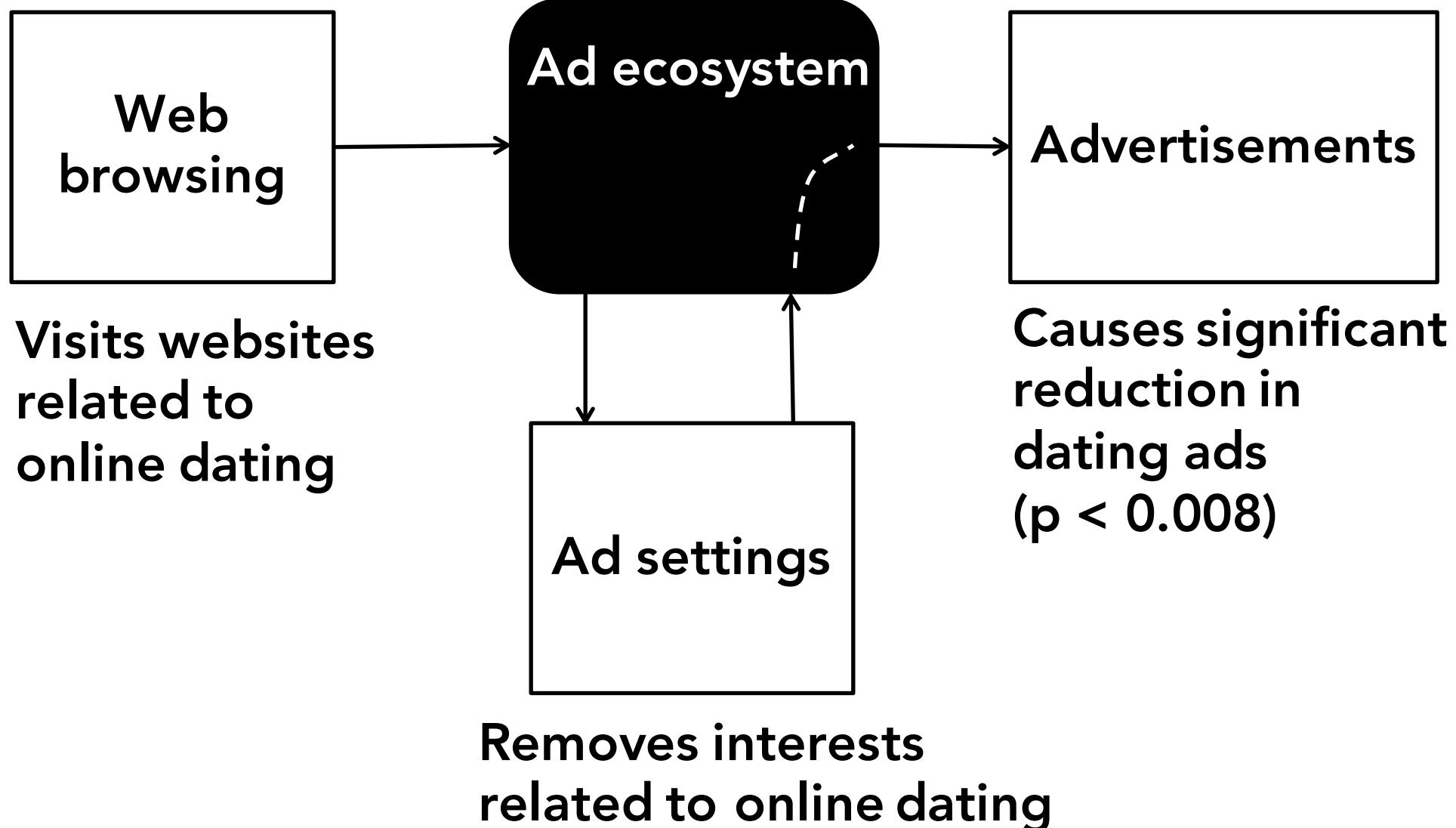
Adventure Games, and 11 more

Based on the websites you've visited

Your activity on websites is used to serve interest based ads.

These categories can only control some of the Google ads you see. For instance, the interests listed here do not reflect ads selected based on a visit to a specific advertiser's page (remarketing) or ads selected by other advertising providers and delivered by Google's ad platforms. [How it works.](#)

# Choice



# Choice Explanation

## Keeping Dating Interest

**Are You Single?**

[www.zoosk.com/Dating](http://www.zoosk.com/Dating)  
2433 vs. 78

**Top 5 Online Dating Sites**

[www.consumer-rankings.com/Dating](http://www.consumer-rankings.com/Dating)  
408 vs. 13

**Why can't I find a date?**

[www.gk2gk.com](http://www.gk2gk.com)  
51 vs. 5

## Removing Dating Interest

**Car Loans w/ Bad Credit**

[www.car.com/Bad-Credit-Car-Loan](http://www.car.com/Bad-Credit-Car-Loan)  
8 vs. 37

**Individual Health Plans**

[www.individualhealthquotes.com](http://www.individualhealthquotes.com)  
21 vs. 46

**Crazy New Obama Tax**

[www.endofamerica.com](http://www.endofamerica.com)  
22 vs. 51

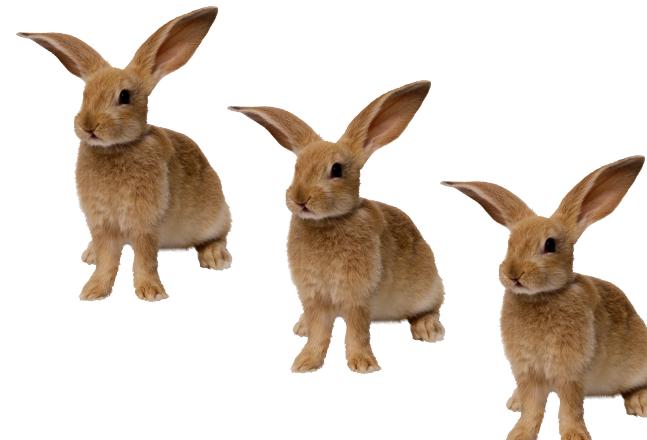
# Experimental Design

Treatment 1 (Drug)



Experimental Group

Treatment 2 (Placebo)



Control Group

# Experimental Design

Treatment 1 (Drug)



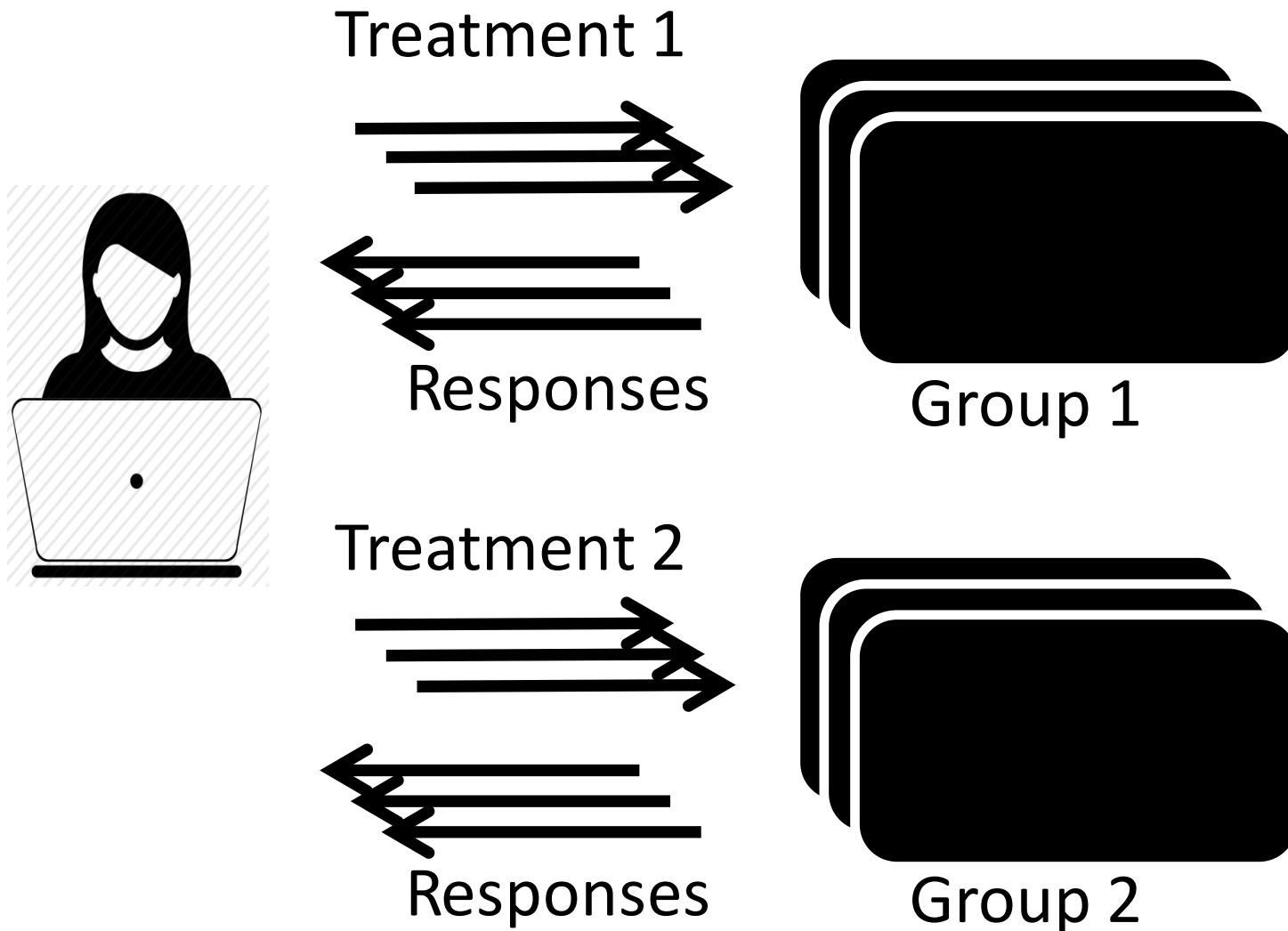
Experimental Group

Treatment 2 (Placebo)



Control Group

# Information Flow Experiment



# Experimental Design

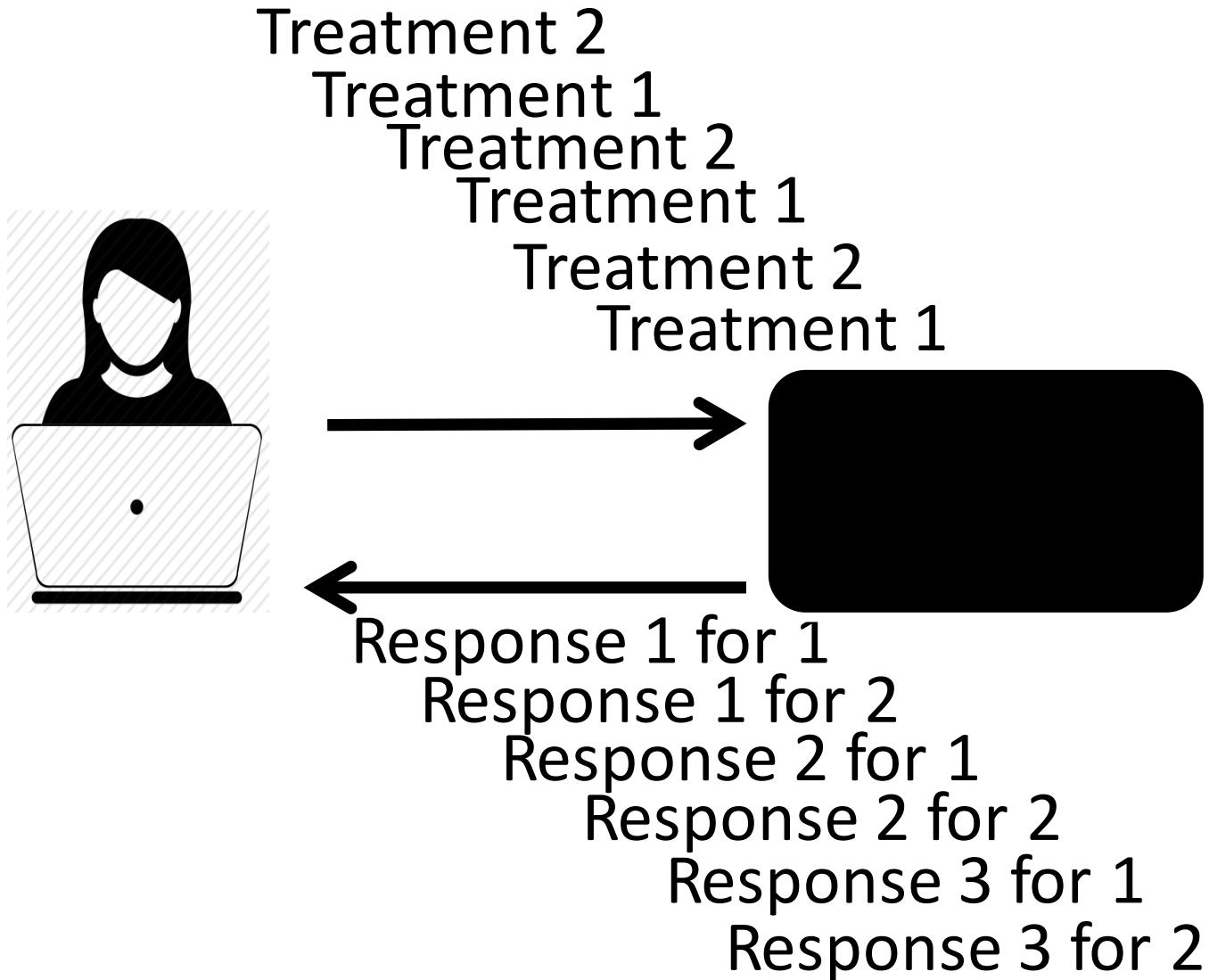
Drug



Placebo

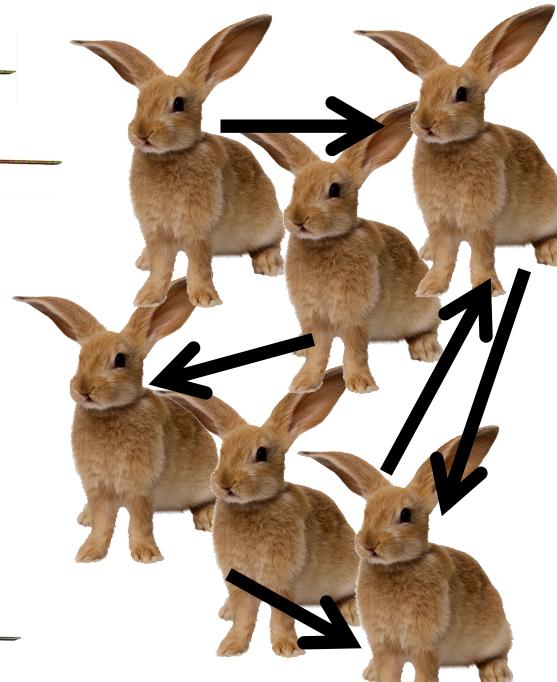
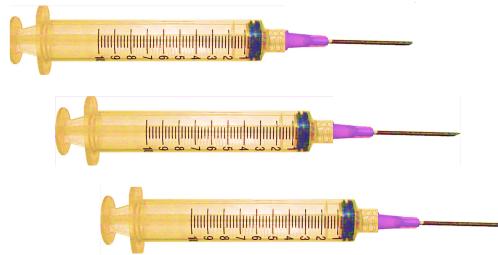


# Information Flow Experiment



# Cross-Unit Effects

Drug



Placebo





Treatment 1

↔



Treatment 1

↔



Treatment 1

↔



Treatment 2

↔



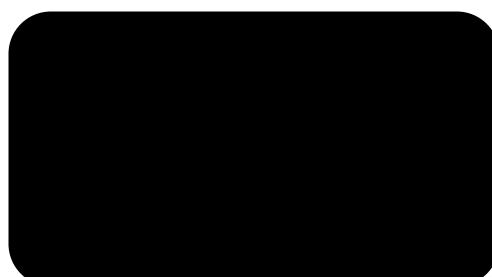
Treatment 2

↔



Treatment 2

↔



# Information Flow Experiments

## Natural Sciences

Natural process

Population of units

Treatments

Responses

...

Causation

## Information Flow

System in question

Subset of interactions

Inputs

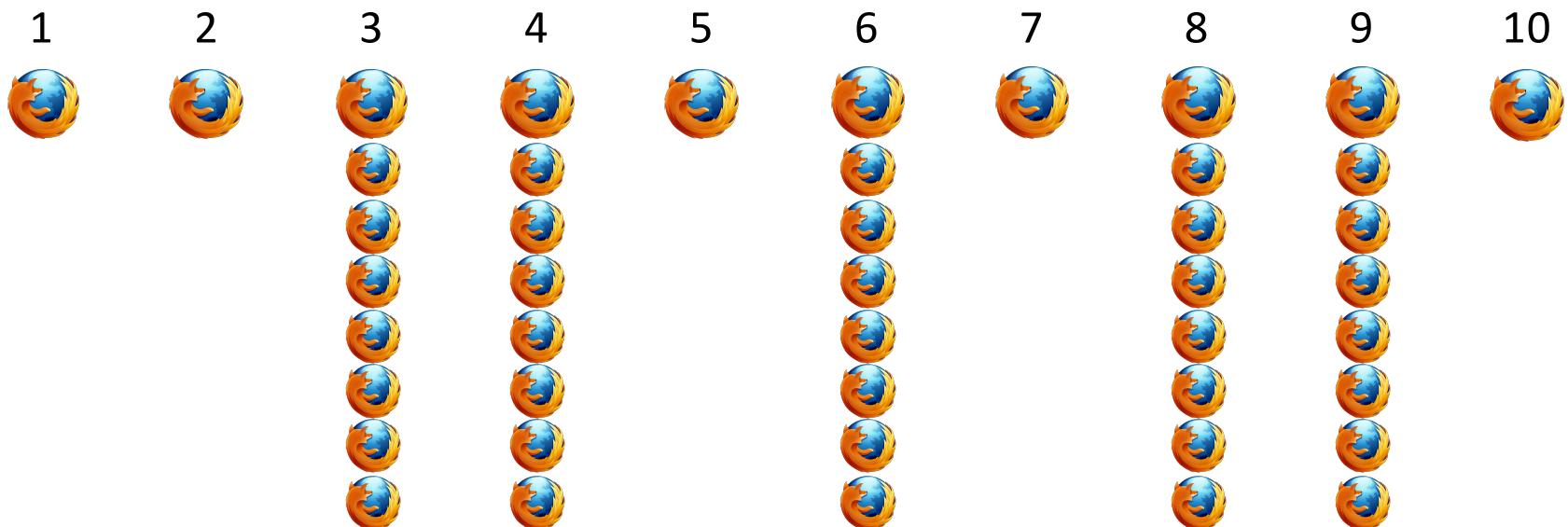
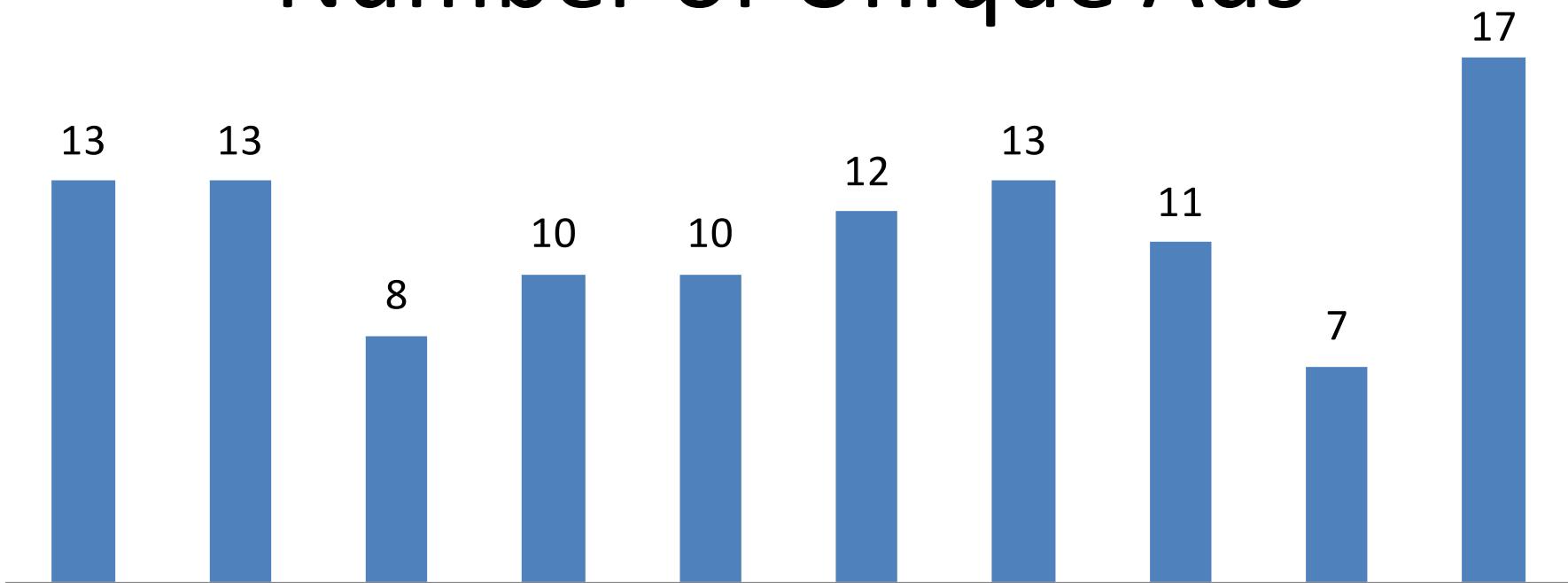
Outputs

...

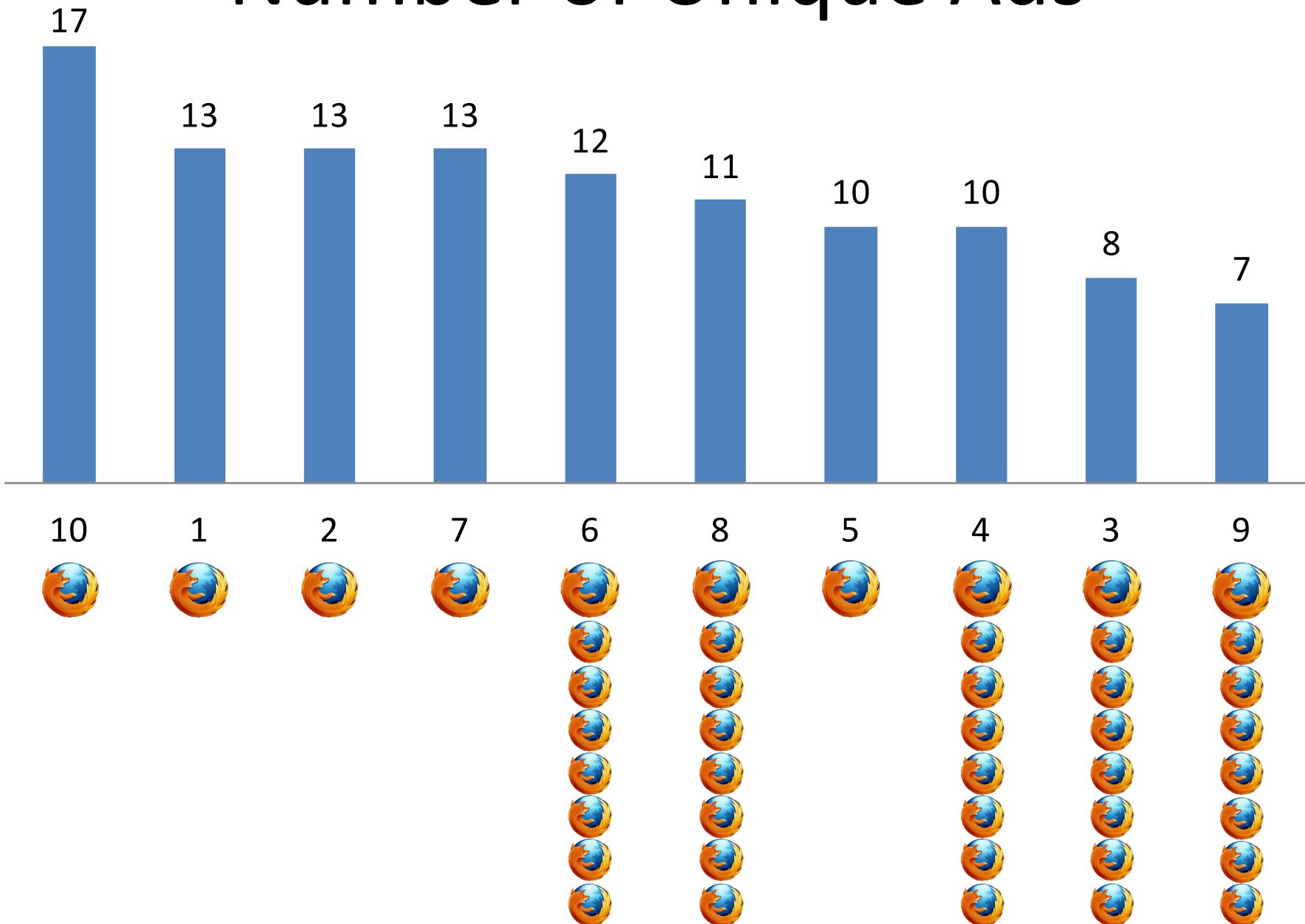
Information flow

Theorem

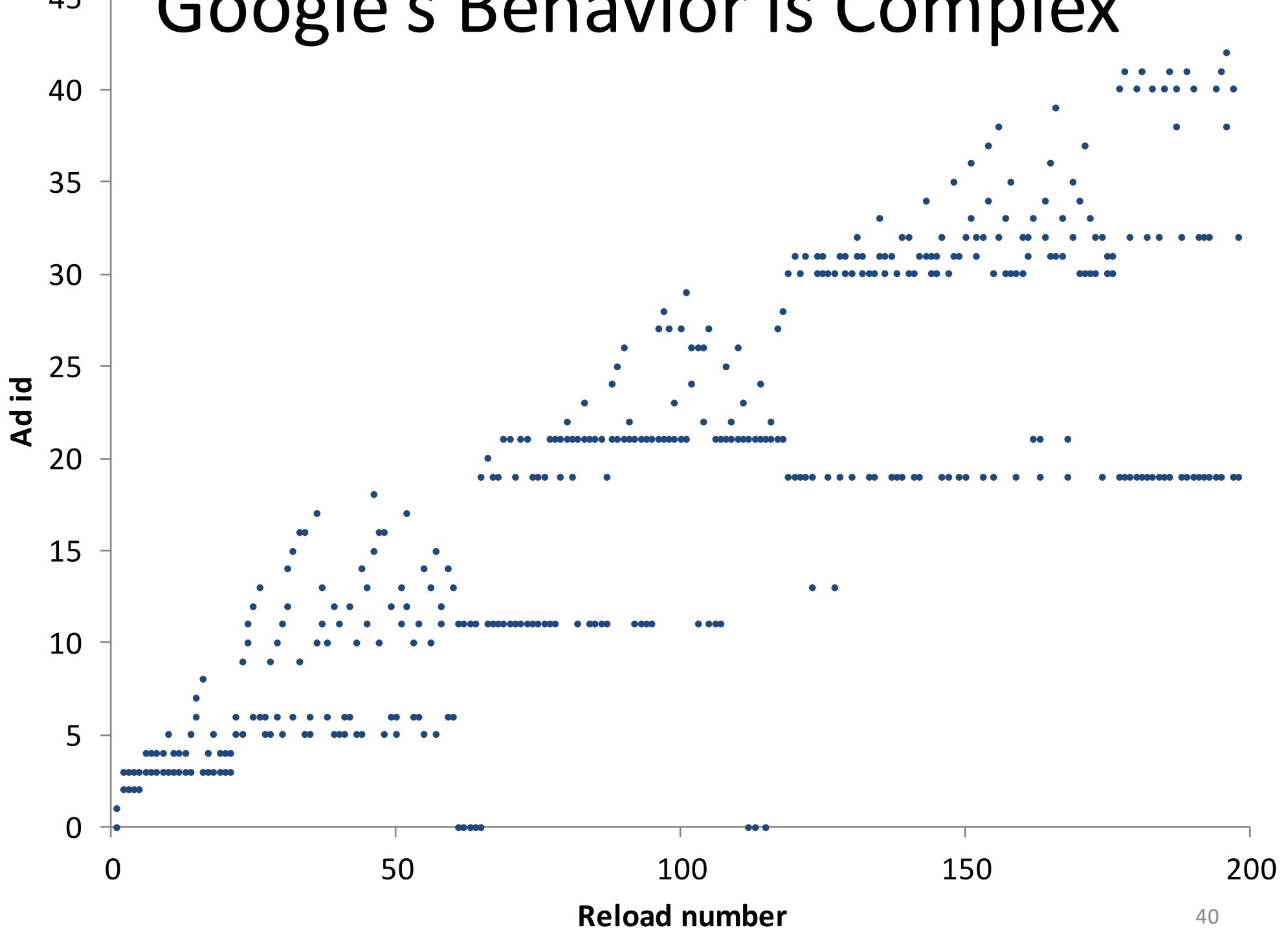
# Number of Unique Ads



# Number of Unique Ads



# Google's Behavior is Complex



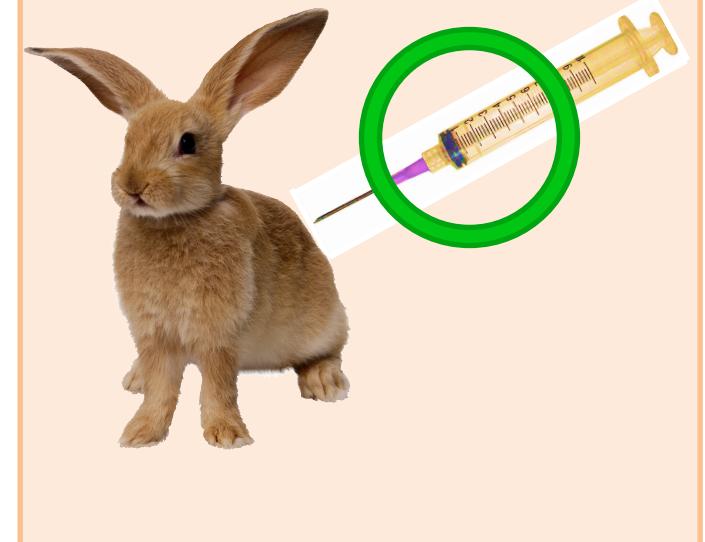
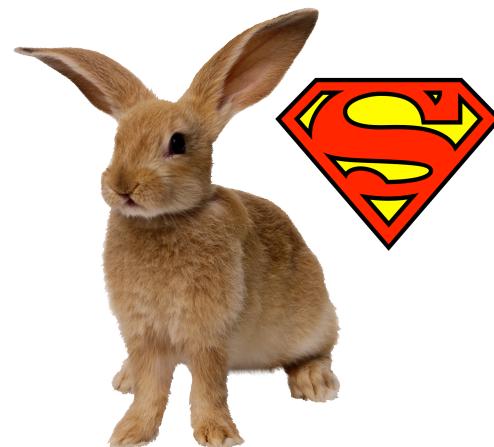
# Prior Work on Behavioral Marketing

- No statistical test
  - Guha<sup>+</sup> 10
  - Balebako<sup>+</sup> 12
  - Wills & Tatar 12
  - Liu<sup>+</sup> 13              AdReveal
- Assumptions about ads
  - Barford<sup>+</sup> 14        AdScape
  - Lecuyer<sup>+</sup> 14        XRay
  - Englehardt<sup>+</sup> 14    OpenWPM

Lives under  
Drug

Lives under  
Placebo

Dies under  
Placebo



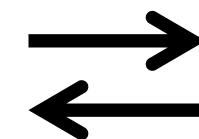
Dies under  
Drug



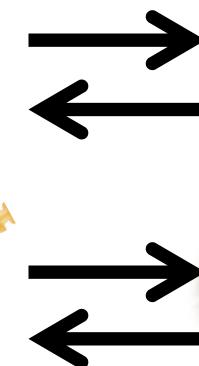
Drug



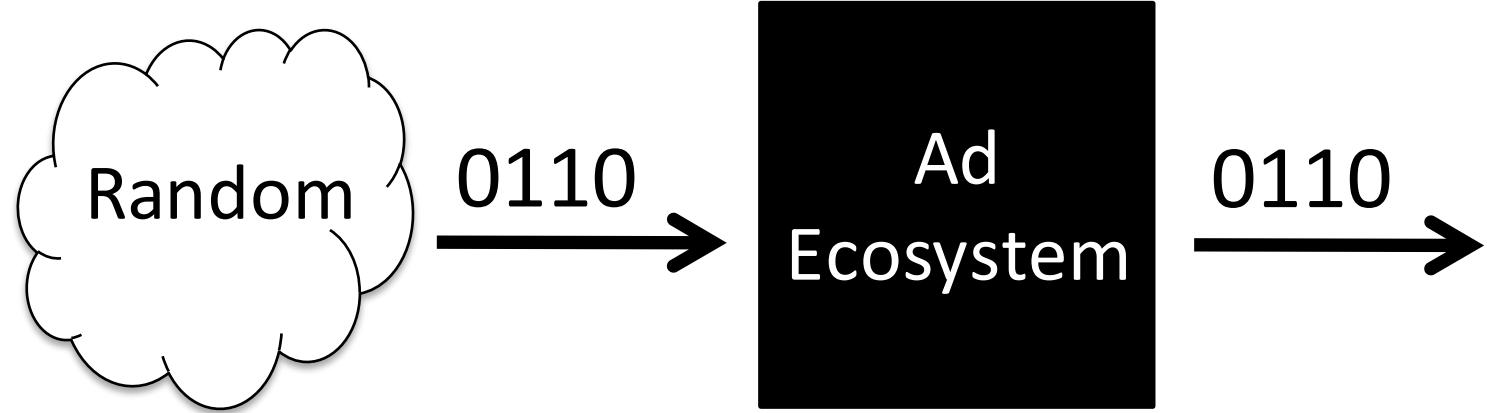
Experimental Group

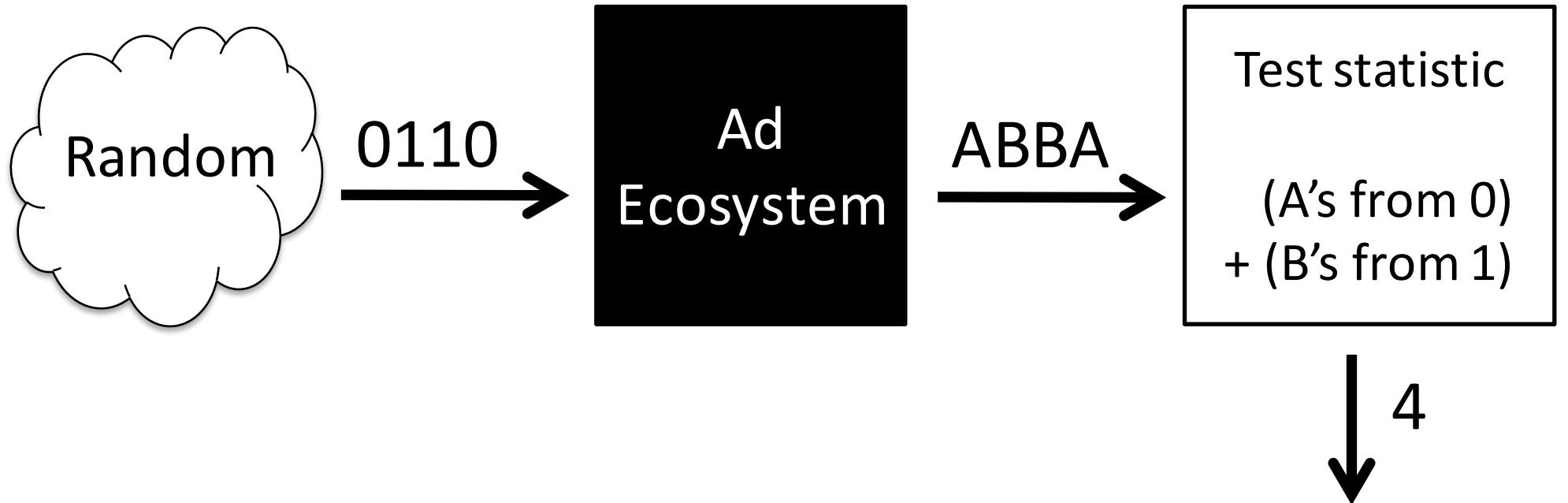


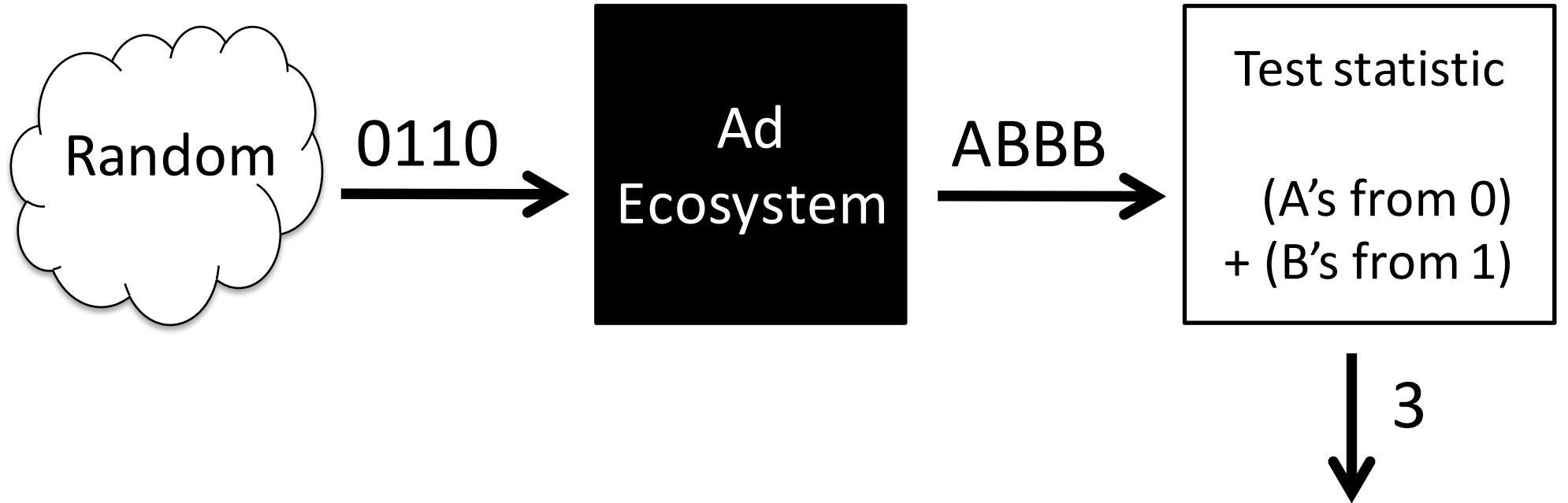
Placebo



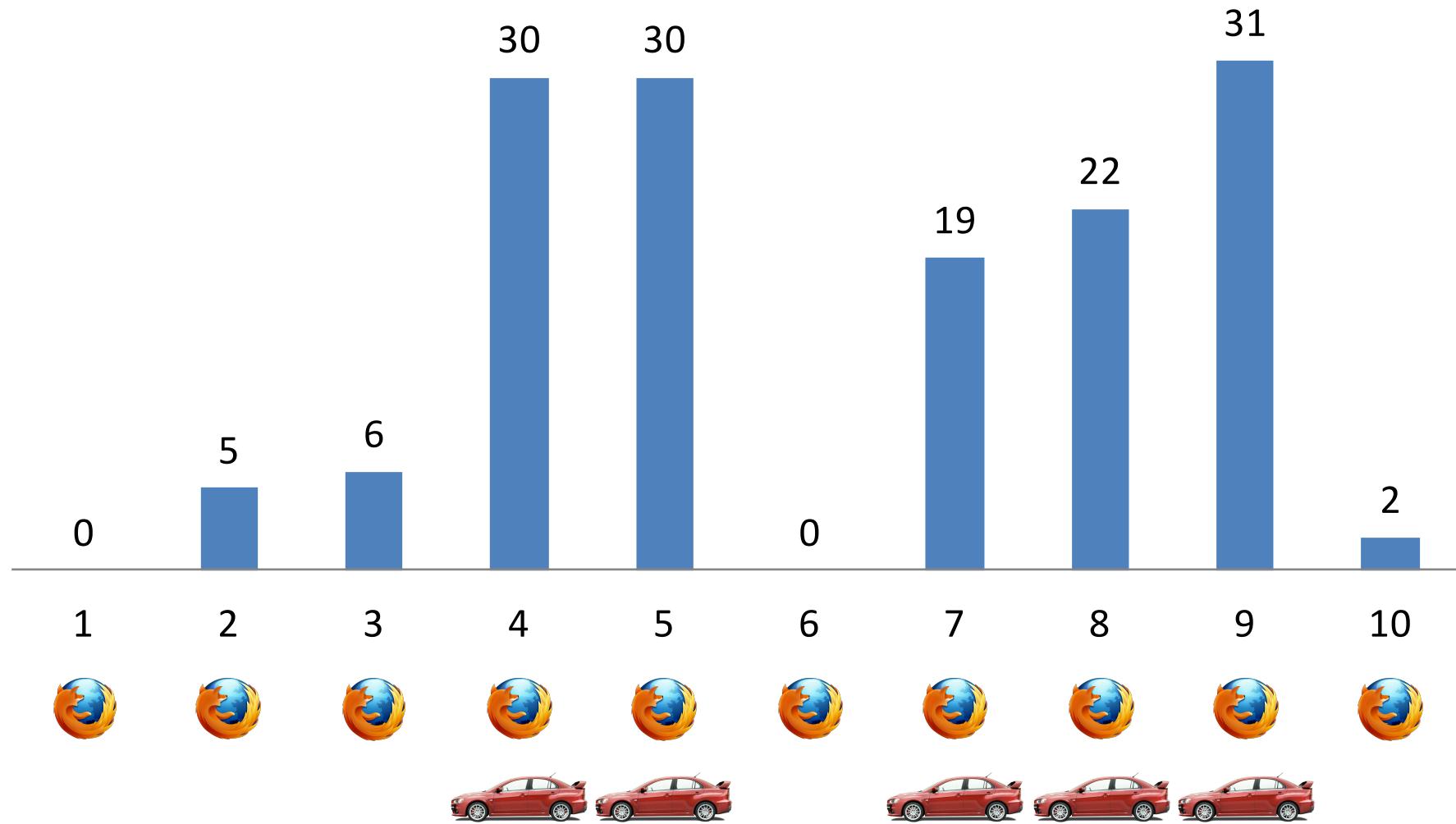
Control Group



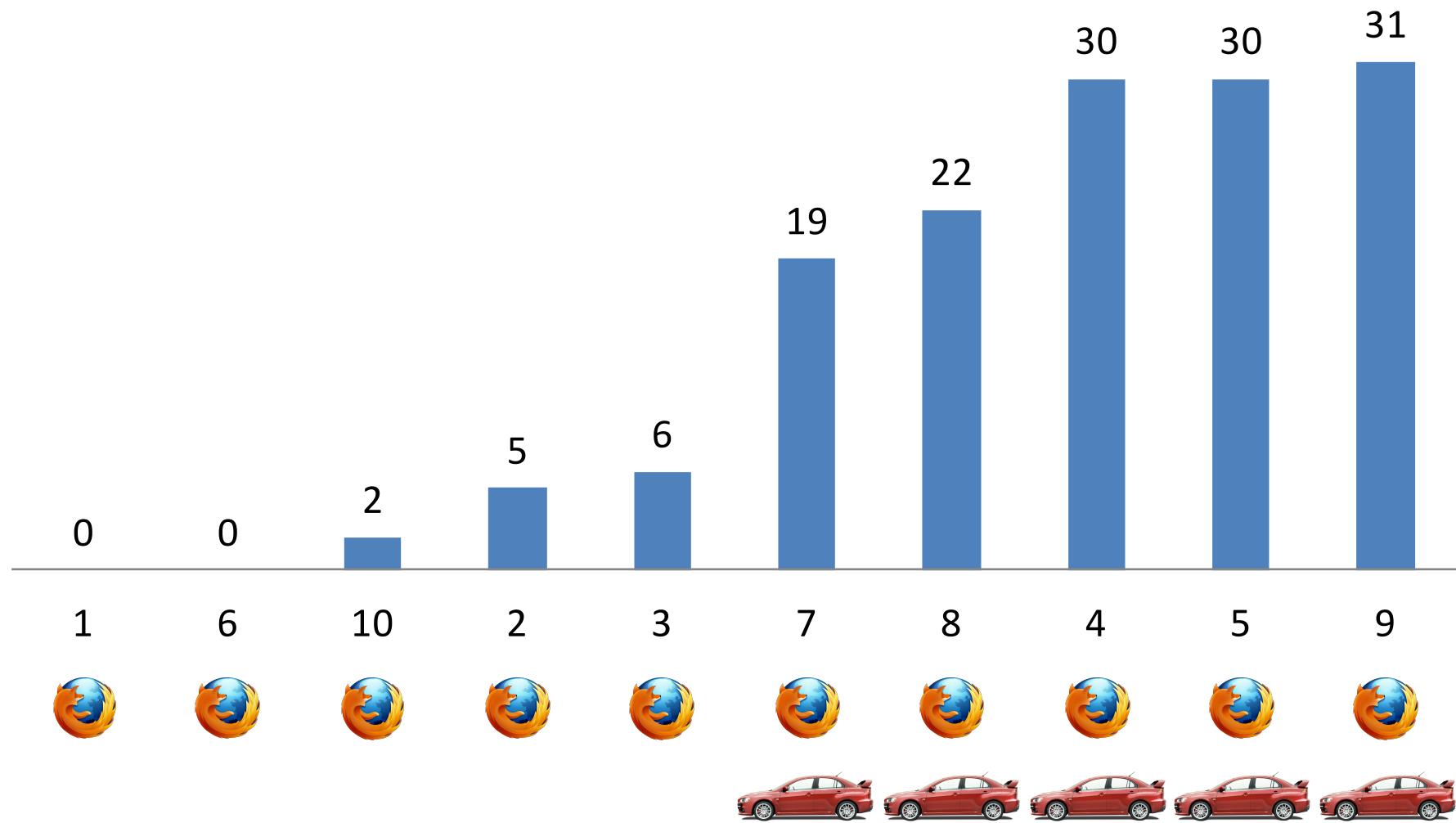




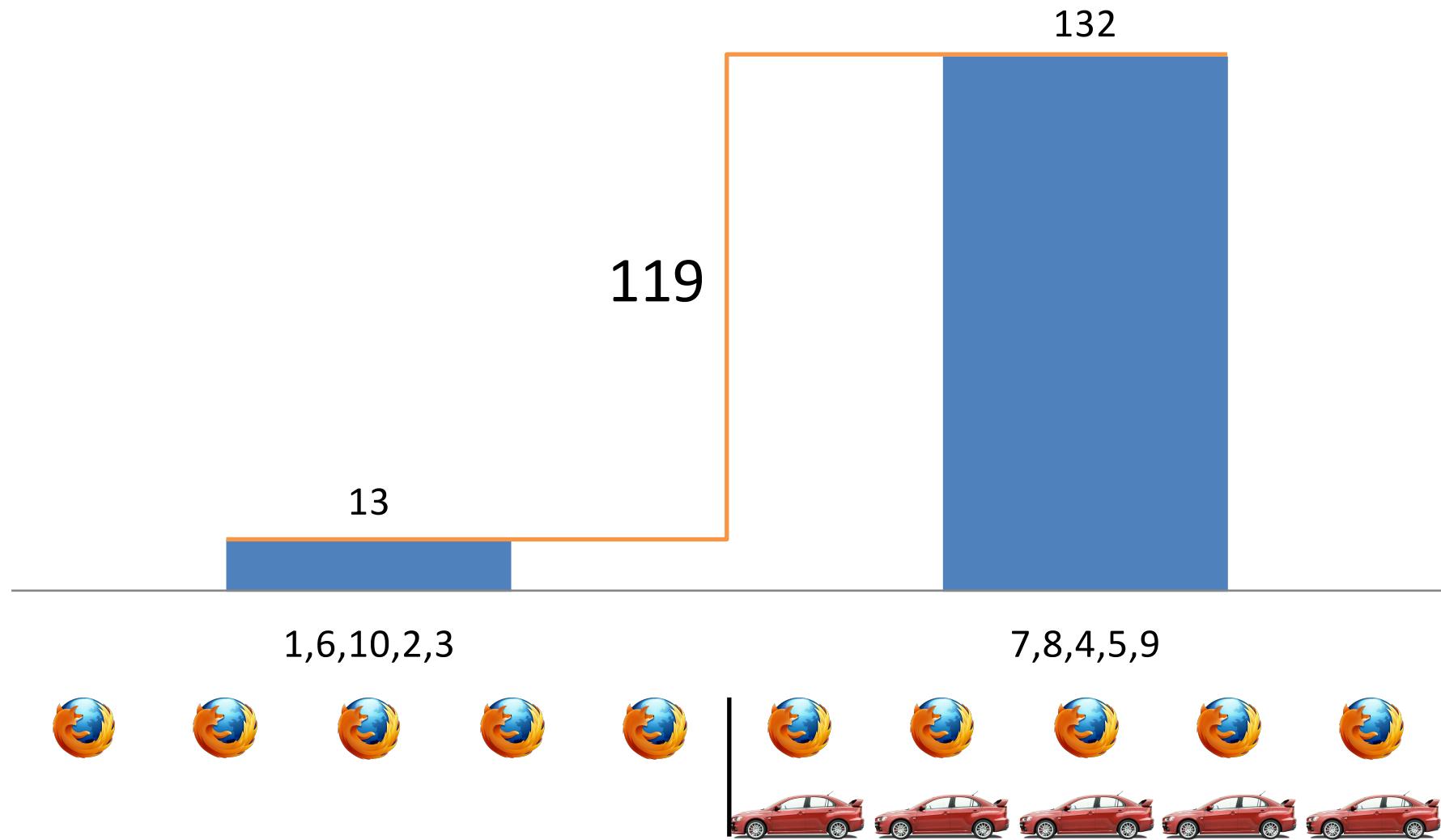
# Permutation Test over Keywords



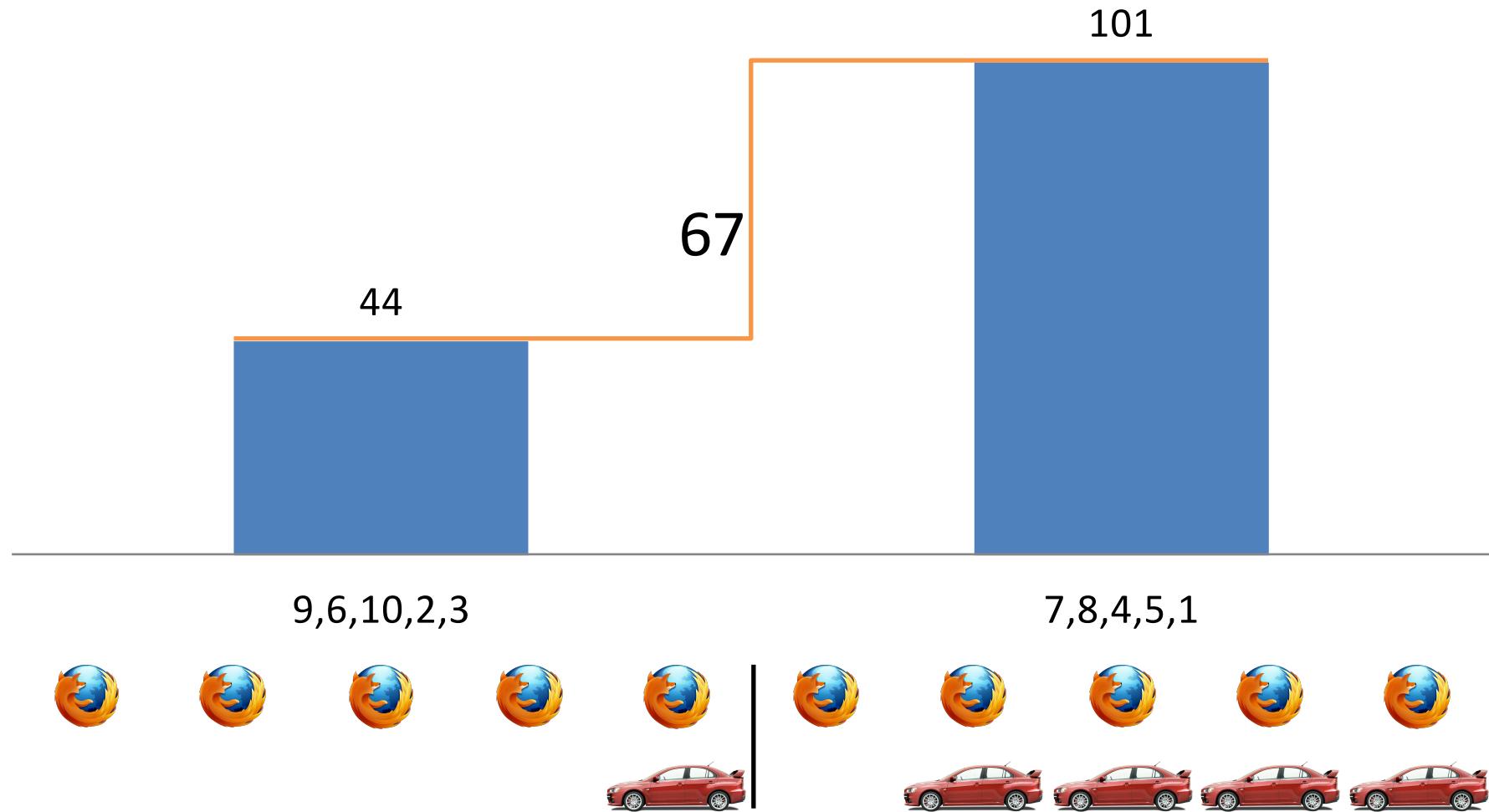
# Permutation Test over Keywords



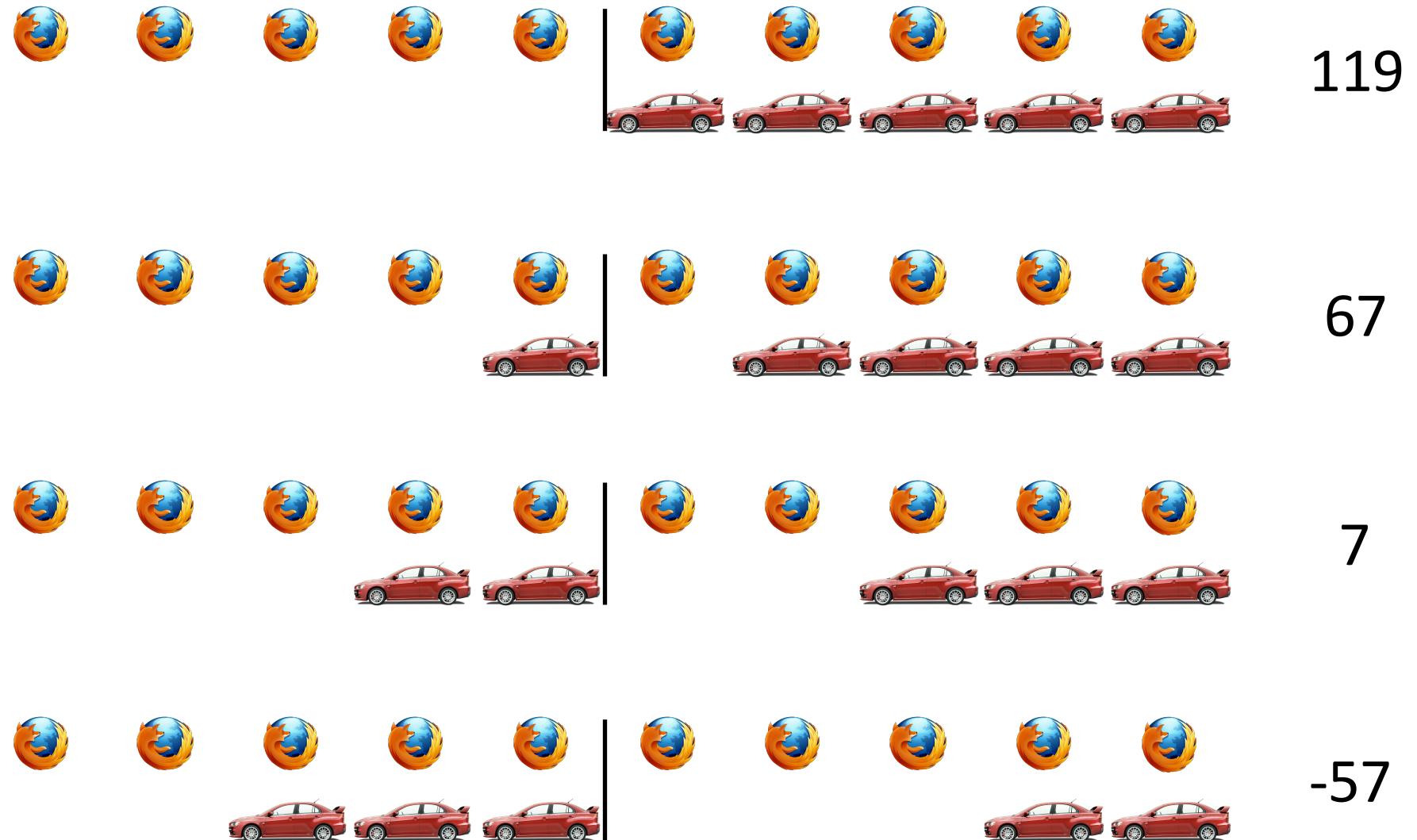
# Permutation Test over Keywords



# Permutation Test over Keywords

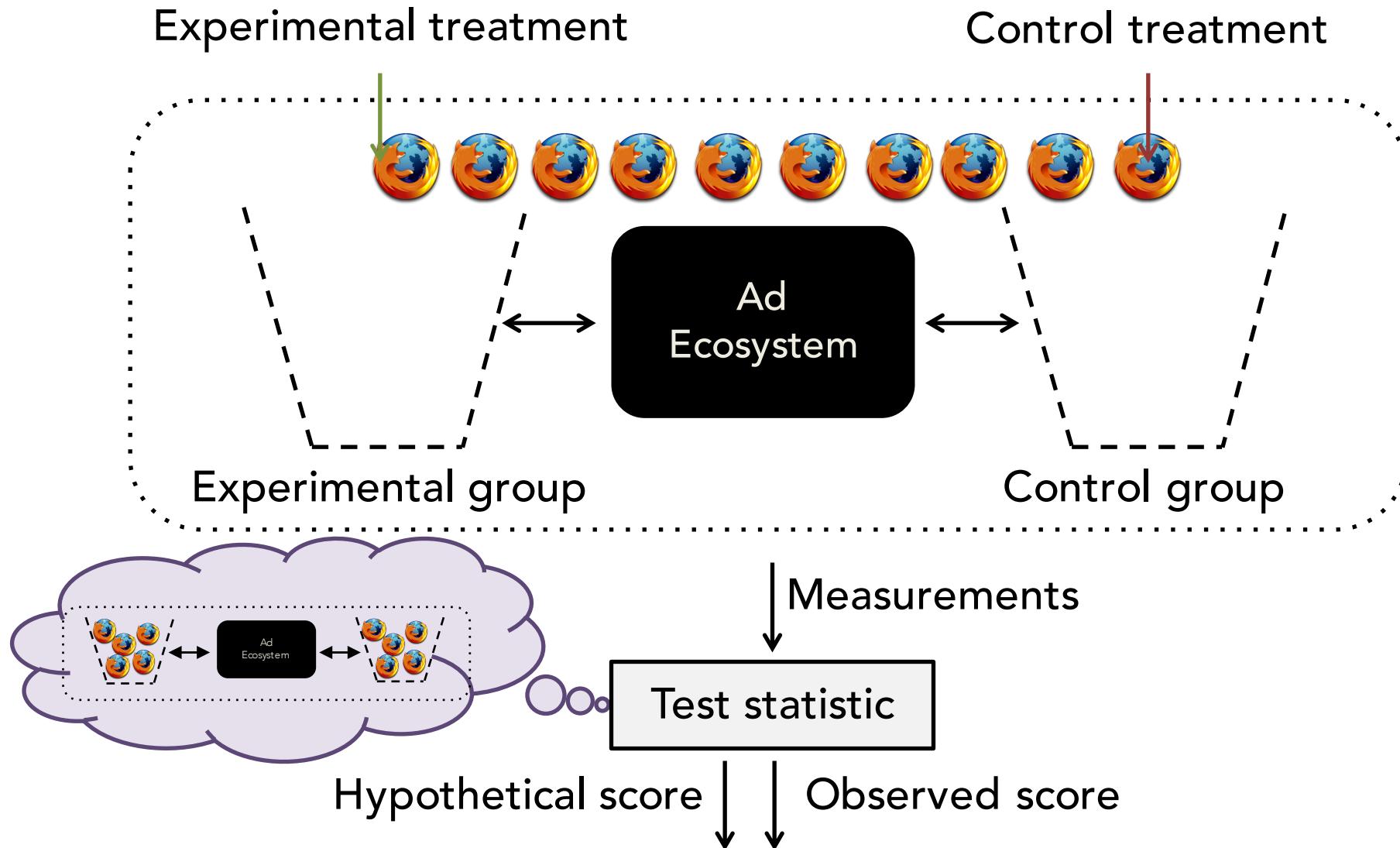


# Look at all permutations

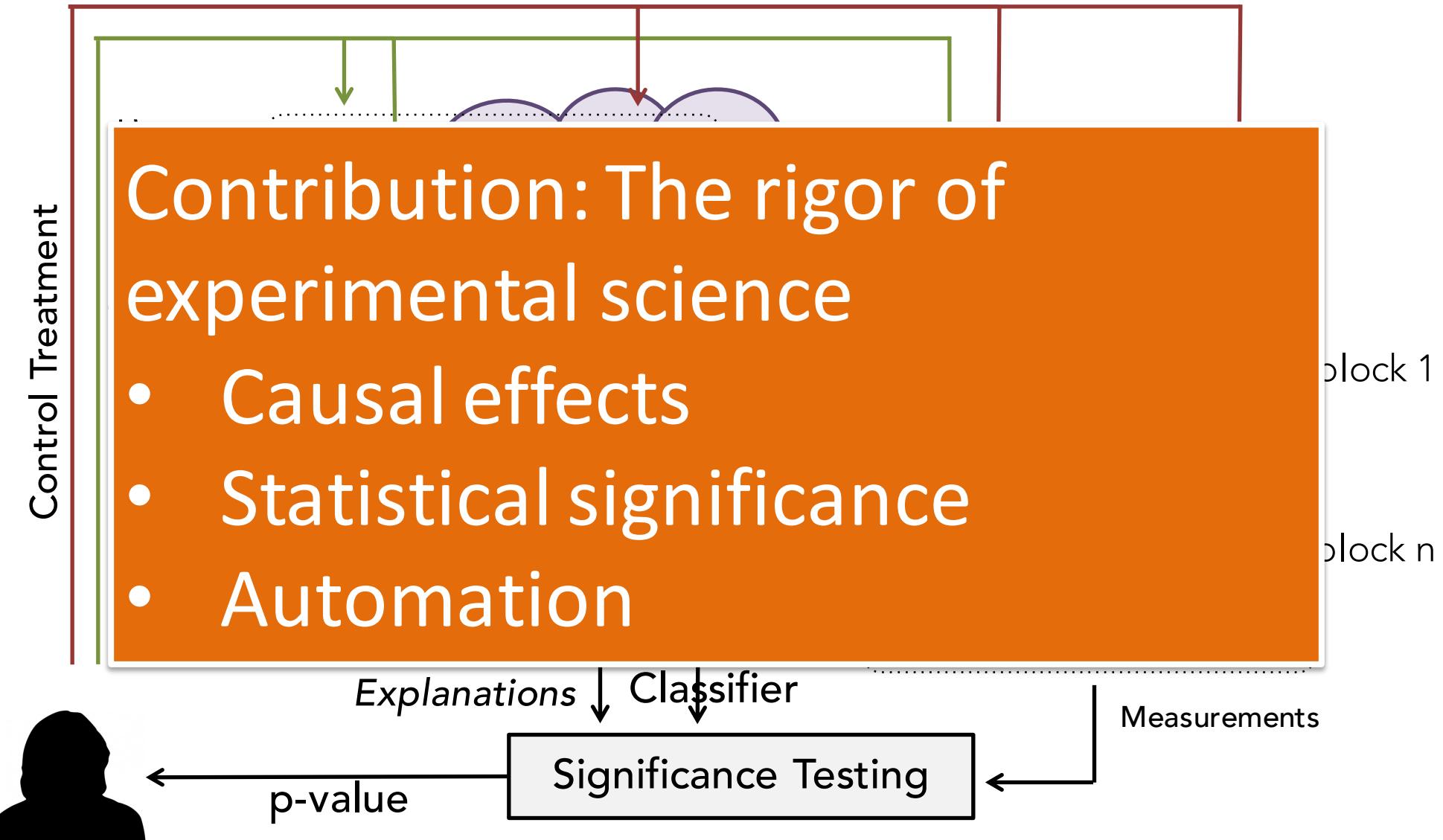


And so forth...

# Randomized Controlled Trials



# Our Methodology



# Interference

System has *interference* from High inputs to Low outputs iff there exists  $H_1$ ,  $H_2$ ,  $O_1$ ,  $O_2$ , and  $L$  such that  $O_1 \neq O_2$  and:



and



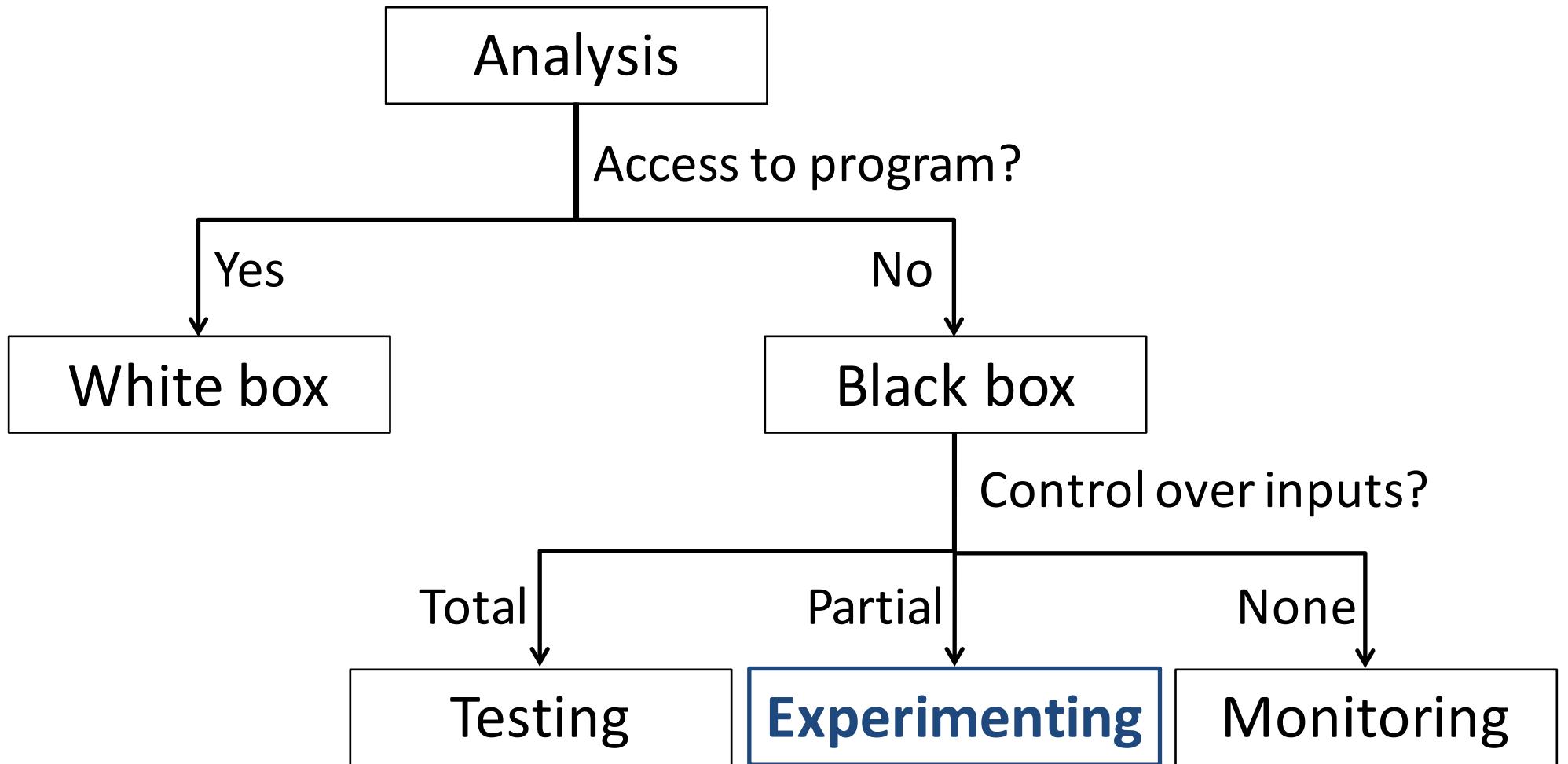
# Definition of Interference

High has *probabilistic interference* with Low  
iff

there exists  $H_1$ ,  $H_2$ , and  $L$  such that  
the distribution of Low's outputs in  
 $Q(H_1, L)$   
equals not the distribution of Low's outputs in  
 $Q(H_2, L)$

# Information Flow Problems

	<b>Analyst</b>	<b>Adversary</b>	<b>Protects</b>
Confidentiality	Has program	Low user	High input
Taint analysis	Has program	Low user	High output
Information usage	Outside access	System	High input



# Impossibility Results

- No sound experiments for interference
- No sound experiments for noninterference  
(except degenerate ones)

# Structural Equation Models

Let  $M$  be a Structural Equation Model defined by

- $\text{State}_0 := s_0$
- $\text{Input}_1 := i_1$
- $\text{State}_1 := \text{transition\_function}(\text{State}_0, \text{Input}_1)$
- $\text{Output}_1 := \text{output\_function}(\text{State}_0, \text{Input}_1)$

# Structural Equation Models

Let  $M$  be a Structural Equation Model defined by

- $\text{State}_0 := s_0$
- $\text{Input}_1 := i_1$
- $\text{State}_1 := \text{transition\_function}(\text{State}_0, \text{Input}_1)$
- $\text{Output}_1 := \text{output\_function}(\text{State}_0, \text{Input}_1)$

$M[\text{Input}_1 := i']$  is  $M$  except with

- $\text{Input}_1 := i'$

# Definition of Effect

The factor X has an *effect* on Y given Z:=z  
iff

there exists  $x_1$  and  $x_2$  such that  
the distribution of Y in

$M[X:=x_1][Z:=z]$

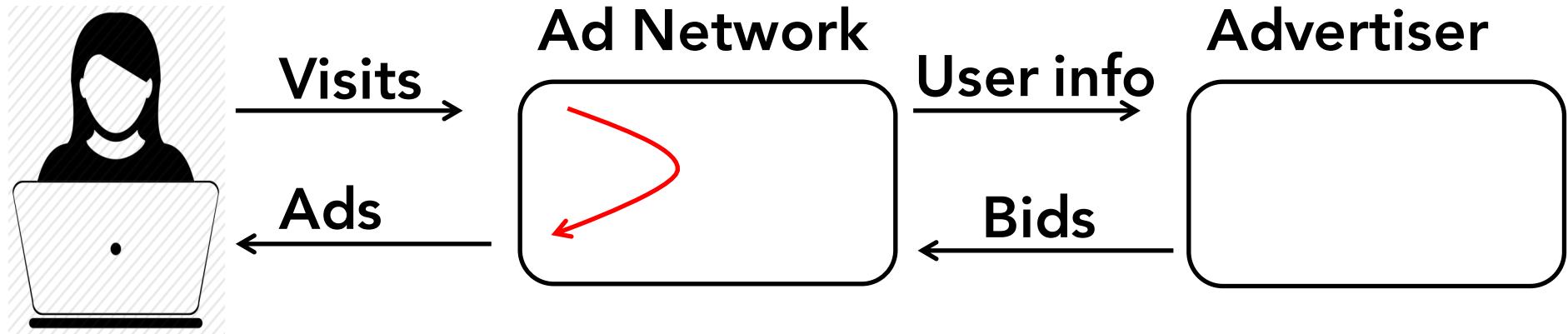
equals not the distribution of Y in

$M[X:=x_2][Z:=z]$

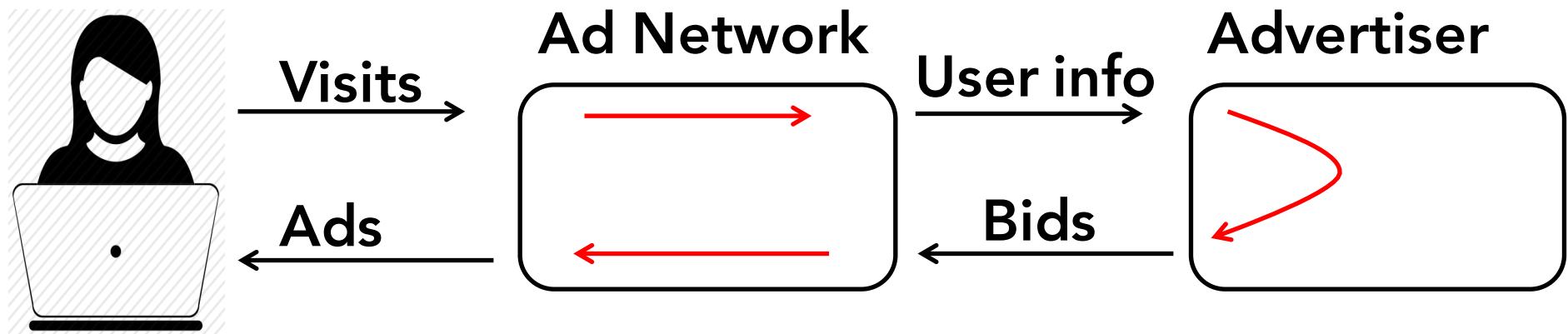
# Theorem

- Let  $M_Q$  be an SEM representation of an automation  $Q$
- $Q$  has probabilistic interference iff there exists a value  $l_i$  for low inputs such that High\_inputs have an effect on low\_outputs given Low\_inputs :=  $l_i$  in  $M_Q$

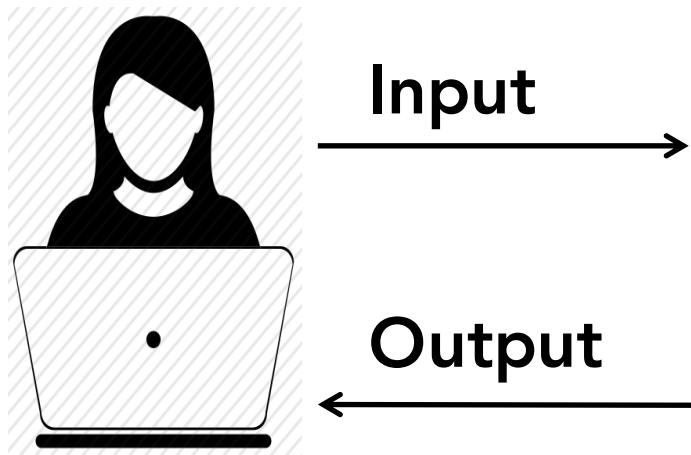
## Ad Network has interference



## Ad Network has noninterference



# Ad ecosystem



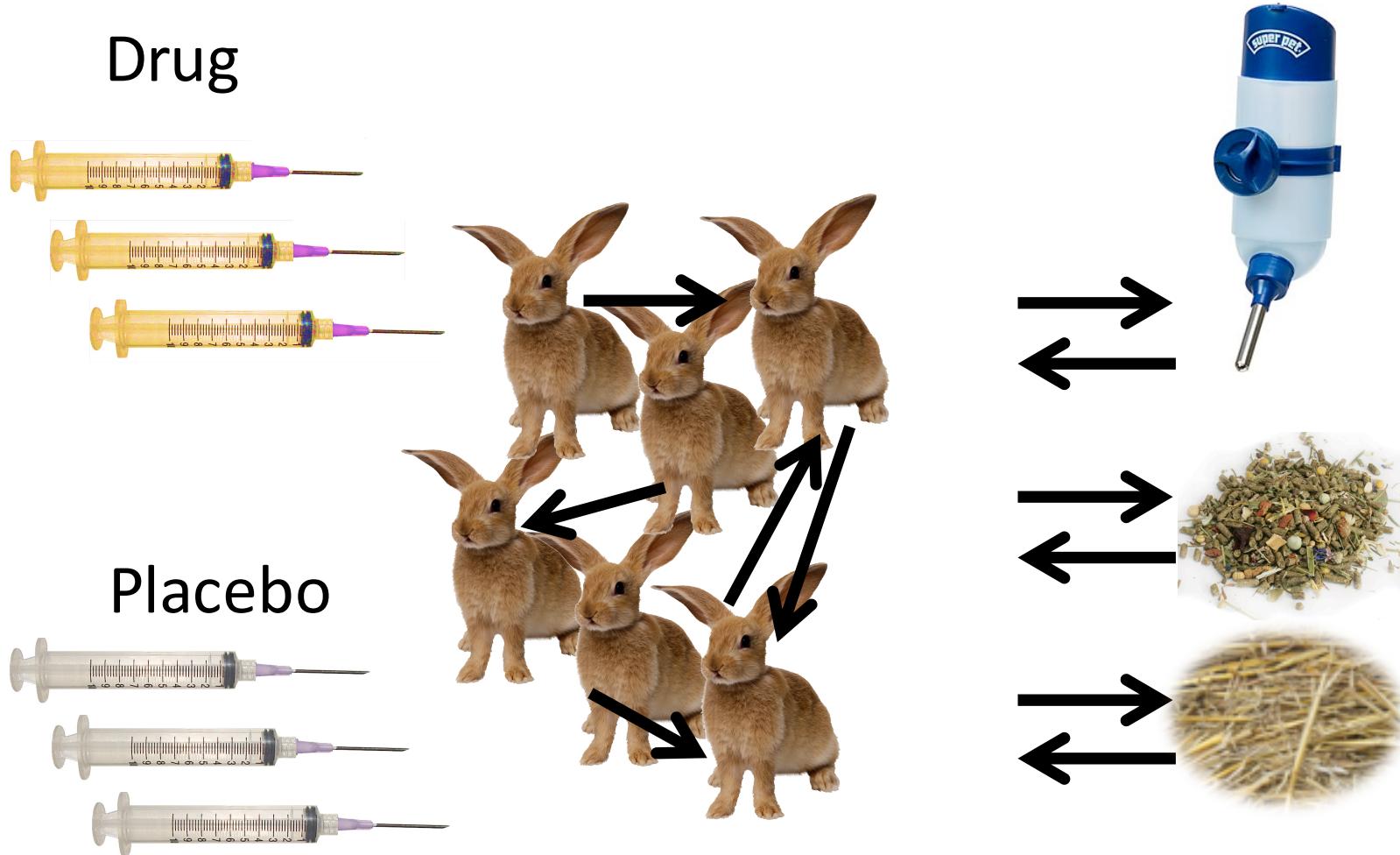
Ad  
networks

Advertisers

Websites

Other users

# Cross-Unit Effects with Environment



# Rigorous Information Flow Experiments

1. Problem: blackbox information flow analysis
2. Interference = Pearl's causation
3. Experimental design and statistical analysis  
for adversarial setting based on permutation  
testing
4. Experimental results showing uses of  
information by Google

# Future Work

- Effect size
- Generalization
- Assigning blame
- Finding noninterference

Ads related to **Michael Tschantz** ⓘ

## We Found **Michael Tschantz**

[www.whitepages.com/Michael+Tschantz](http://www.whitepages.com/Michael+Tschantz)

Get Phone, Address & More for **Michael Tschantz**, Try Free Now!

WhitePages has 1,603 followers on Google+

[Name Popularity & Facts](#) - [Neighbor Search](#) - [Reverse Phone Lookup](#)

## We Found **Michael Tschantz** - Search Arrests, Address, Phone, etc

[www.instantcheckmate.com/](http://www.instantcheckmate.com/)

Search records for **Michael Tschantz**.

## Searches related to **Michael Tschantz**

[michael tschantz salon](#)

[michael tschantz oh](#)

[michael tschantz cmu](#)

[michael tschantz hair salon](#)

[michael tschantz facebook](#)

[michael tschanz](#)

< **Goooooooooooooogle** >

[Previous](#)

2 3 4 5 6 7 8 9 10 11

[Next](#)

[Advanced search](#)

[Search Help](#)

[Send feedback](#)

# Multiple Treatments

Visit jobs website

		Yes	No
		Group Y, Y	Group Y, N
Visit cooking website	Yes	Group Y, Y	Group Y, N
	No	Group N, Y	Group N, N

# Multiple Treatments

Visit jobs website

		Yes	No
		Group Y, Y	Group Y, N
Visit cooking website	Yes	Group Y, Y	Group Y, N
	No	Group N, Y	Group N, N

# Multiple Treatments

Visit jobs website

		Yes	No
		Group Y, Y	Group Y, N
Visit cooking website	Yes	Group N, Y	Group N, N
	No		

# XRay

- Reduces number of needed tests by assuming
  - Monotonicity: visiting more websites only increases the number of ads targeted to you
  - Identically distributed ads

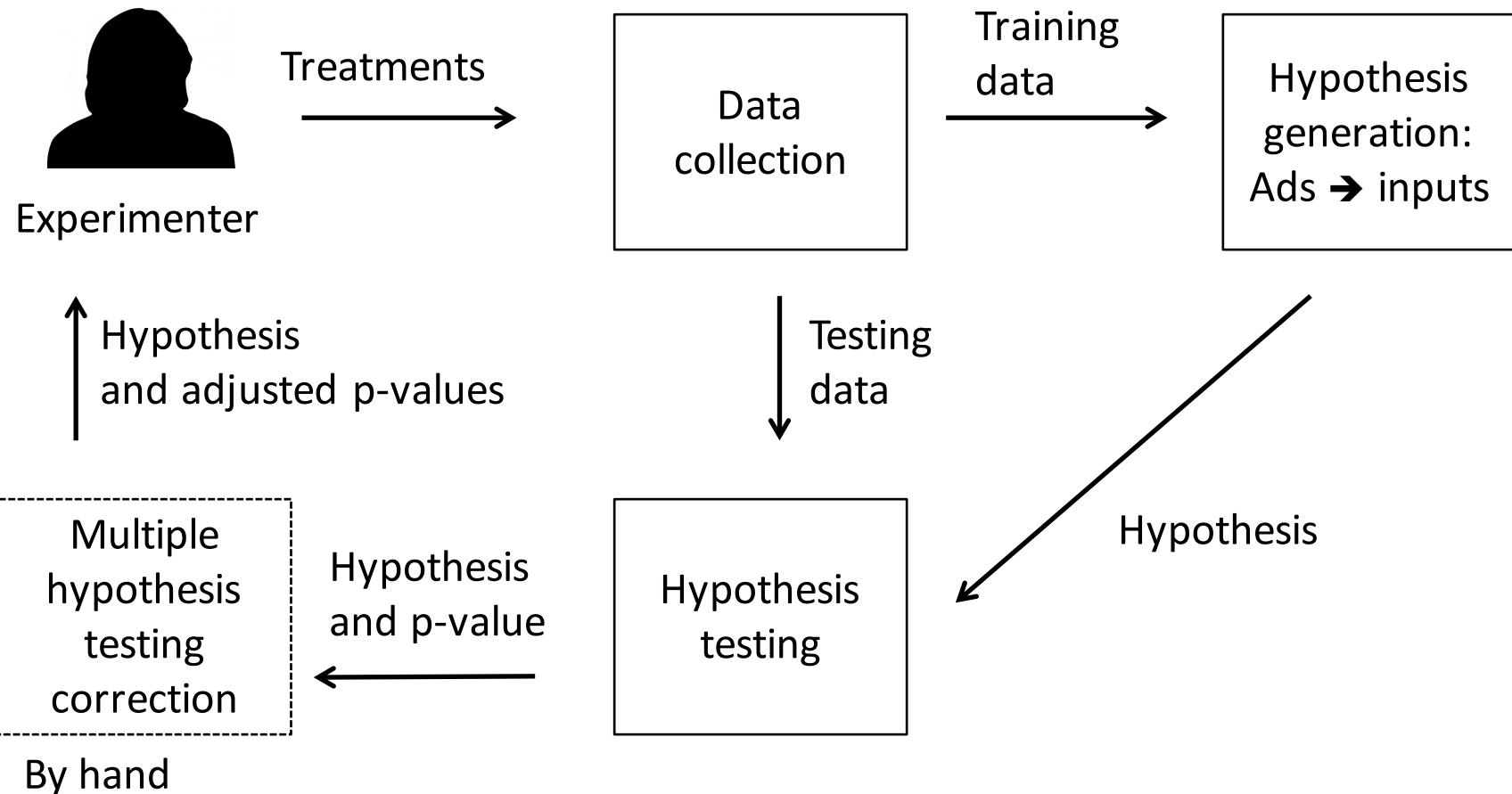
Supported:

		Visit jobs website	
		Yes	No
Visit cooking website	Yes	Group Y, Y	Group Y, N
	No	Group N, Y	Group N, N

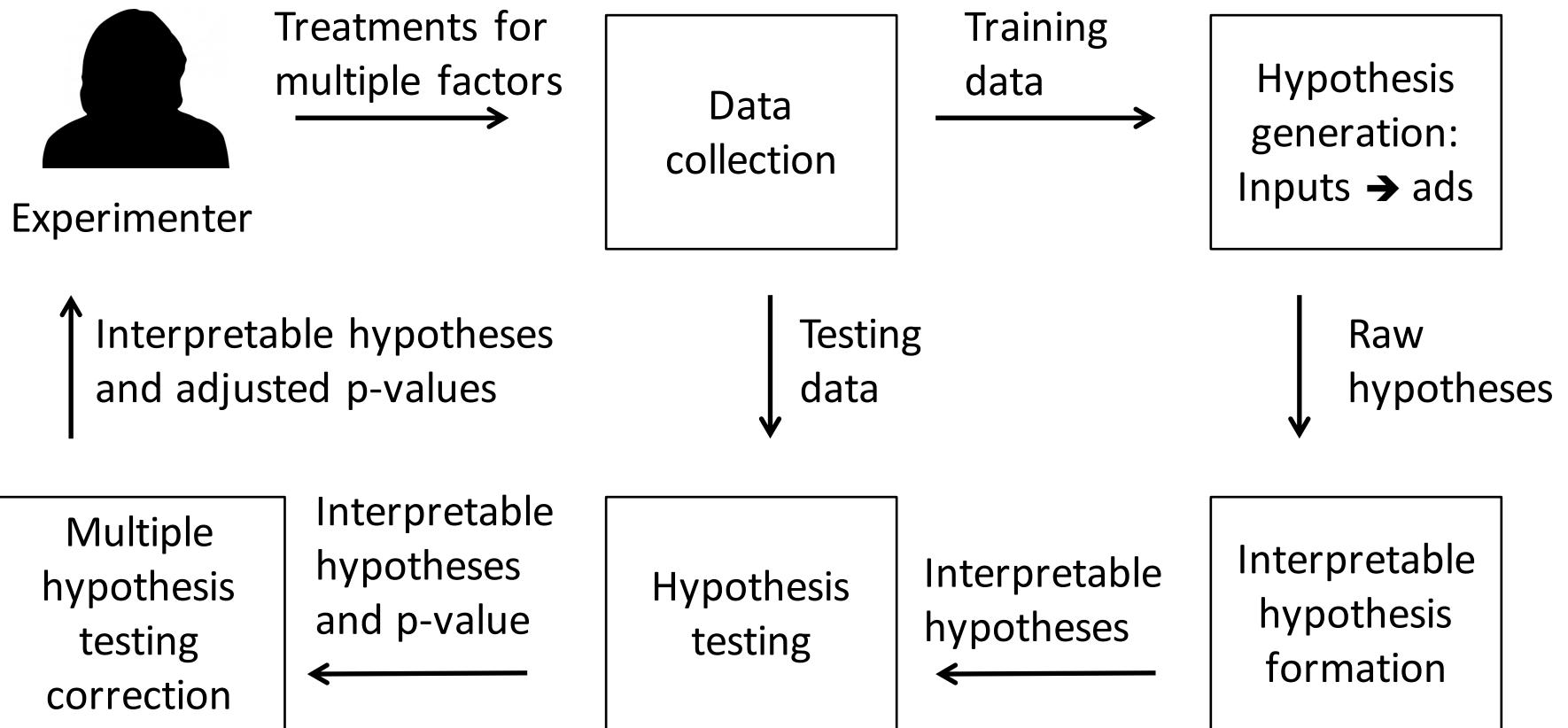
Not supported:

		Visit jobs website	
		Yes	No
Visit cooking website	Yes	Group Y, Y	Group Y, N
	No	Group N, Y	Group N, N

# AdFisher



# Sunlight



Adapted from Lecuyer et al.

# More Information

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



**W I R E D**

**MIT  
Technology  
Review**

The New York Times  
**The Upshot**

**Pittsburgh  
Post-Gazette®**  
post-gazette.com

The Washington Post  
**The Intersect**

**YAHOO!  
TECH**

**THE  
WALL STREET  
JOURNAL**

# AdFisher

## Information Flow Experiments on Ad Privacy Settings

Michael Carl Tschantz  
International Computer Science Institute

Amit Datta, CMU  
Anupam Datta, CMU  
(sometimes) Jeannette M. Wing, MSR

[www.cs.cmu.edu/~mtschant/ife](http://www.cs.cmu.edu/~mtschant/ife)