

18734: Foundations of Privacy

# Information Flow Experiments

Giulia Fanti

Based on slides by Anupam Datta

CMU

Fall 2019

# Administrative

- Grading HW1 almost done
- Project proposals due on **Friday Sept 20**
  - Submit **one** per group on Gradescope
- HW2 has been released
  - Due **Friday, Sept. 27** at 5 pm ET/2 pm PT
  - Please start early!
- **No recitation** on Friday Sept 20
- **Time change** for my office hours on Friday Sept 20
  - 11.30 am-1 pm ET/8.30 am-10 am PT
  - PGH: my office, CIC 2118
  - SV: Google Hangouts Link on the calendar

# In-Class Quiz



- 10 minutes

Today:  
Information Flow Experiments Methodology




Michael Tschantz, ICSI  
Amit Datta, CMU  
Anupam Datta, CMU  
Jeannette Wing, MSR


IEEE Computer Security Foundations Symposium, 2015

# Personalized Web Advertising

Indiatimes | The Times of India | The Economic Times | More ▾ Sign In / Create Account  Like 6.7m  Follow 3.2M

ADVERTISEMENT

 RBC Royal Bank 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](#)

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

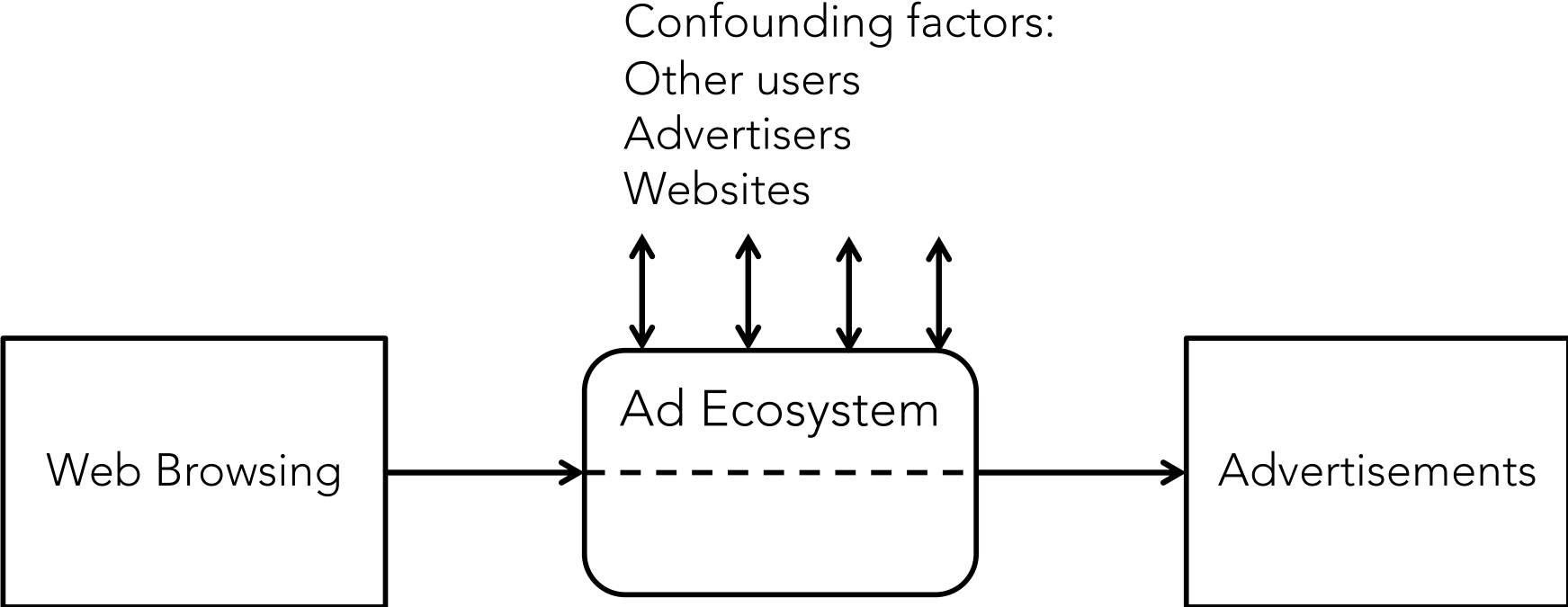
Reuters | Dec 12, 2014, 08:39 AM IST

**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

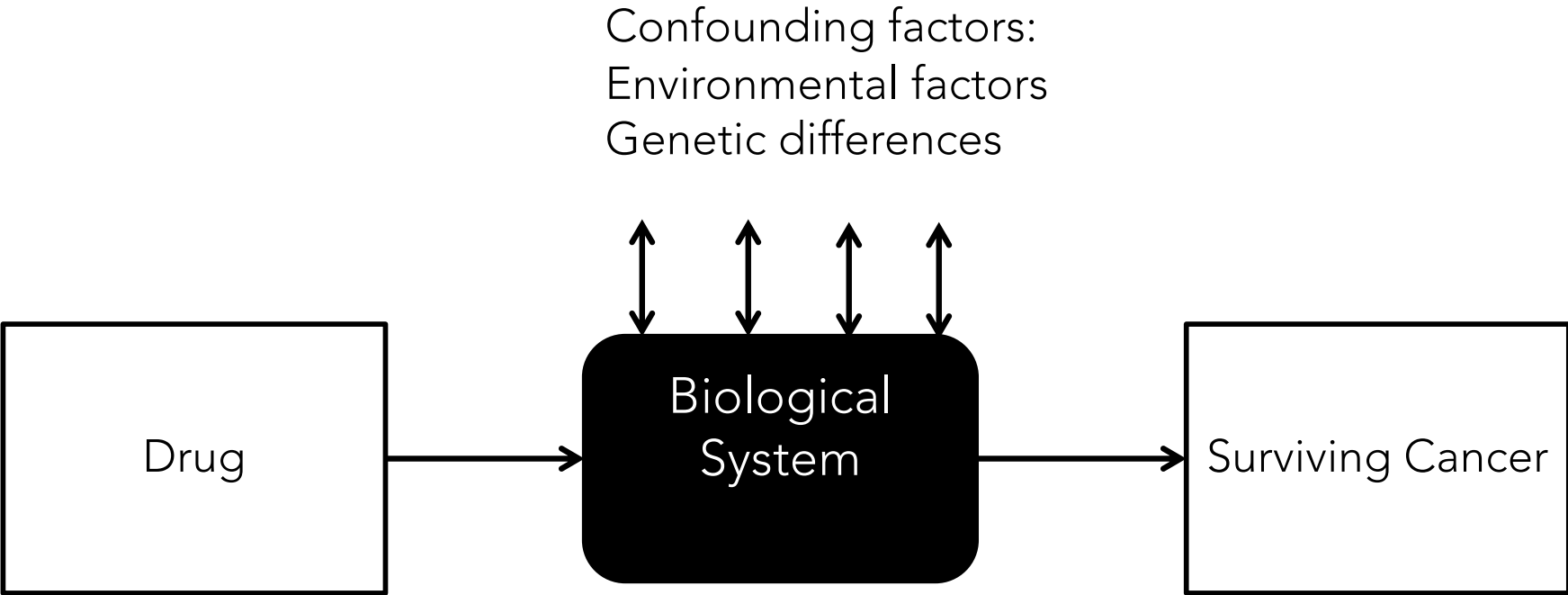
READ MORE » [Hong Kong Protesters](#) | ['We'll Be Back'](#) | [Hong Kong](#) | [CY Leung](#)

 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,

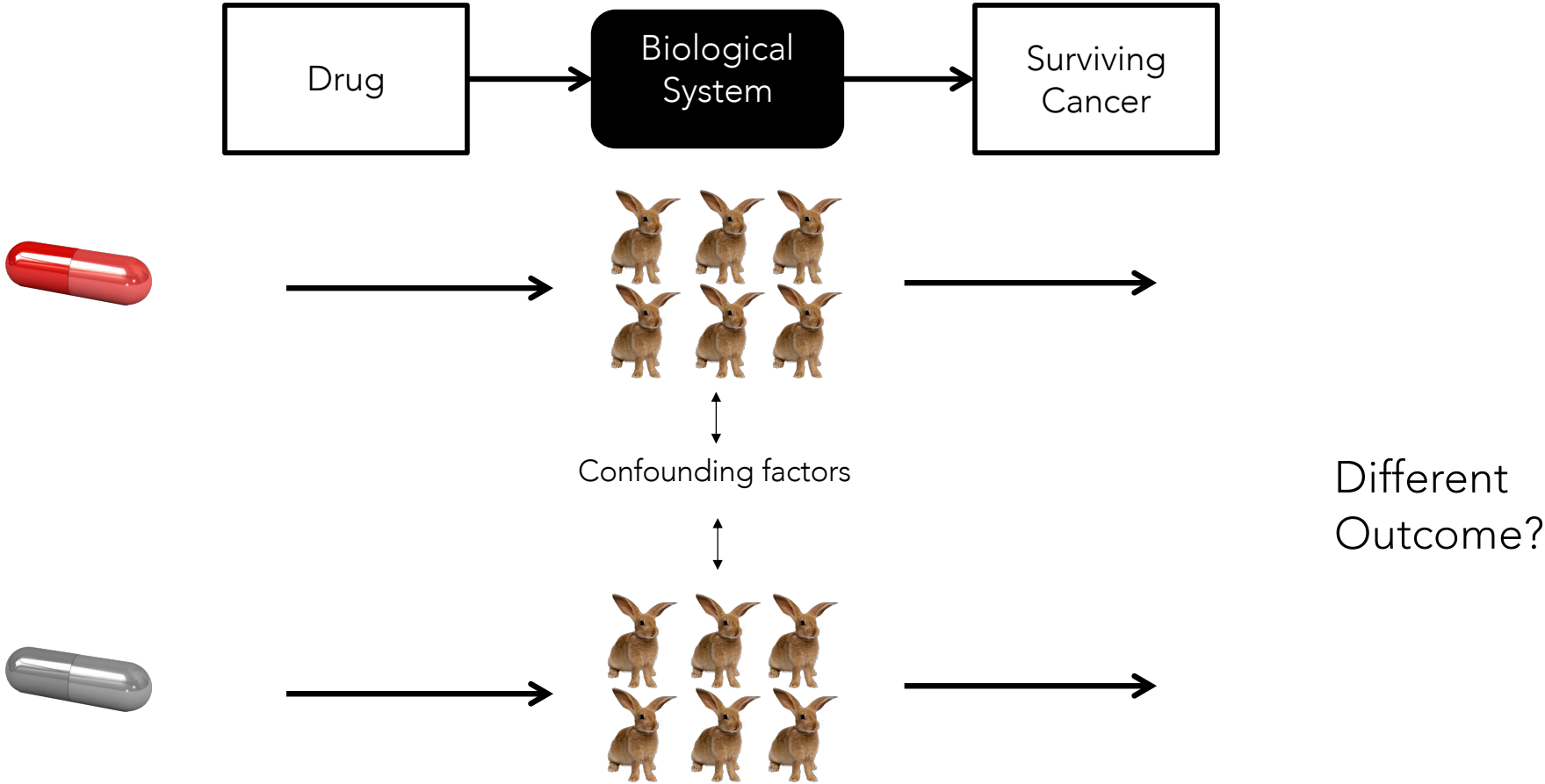
# Personalized Web Advertising



# Experimental Design

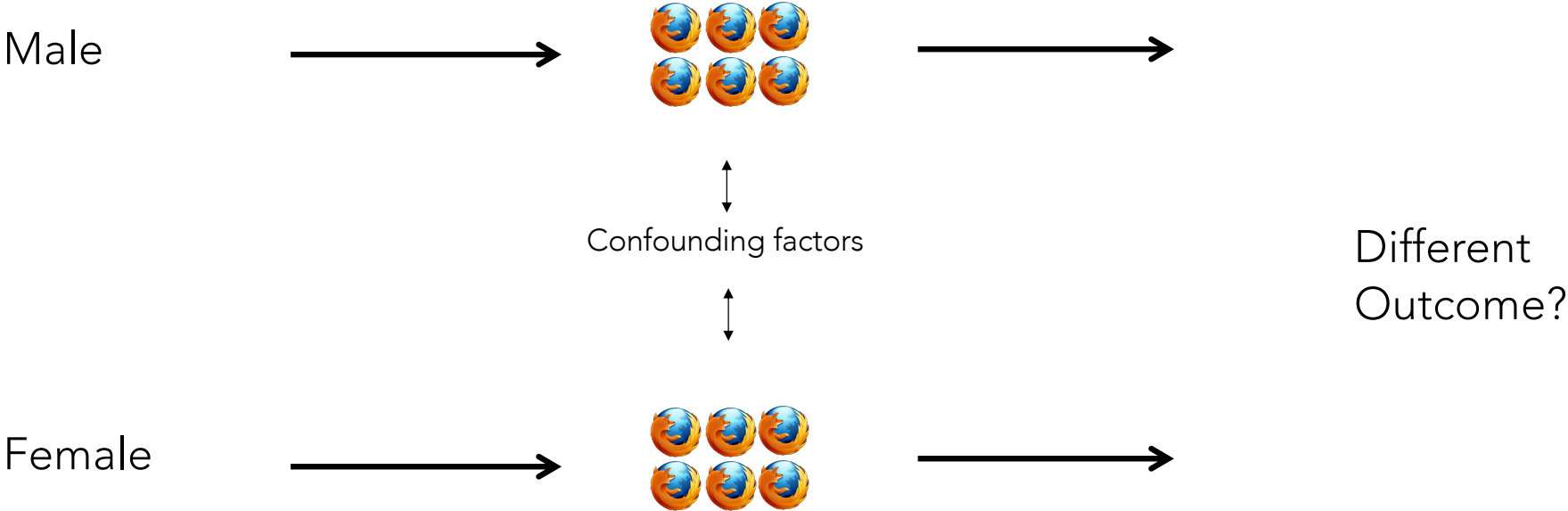


# Experimental Design





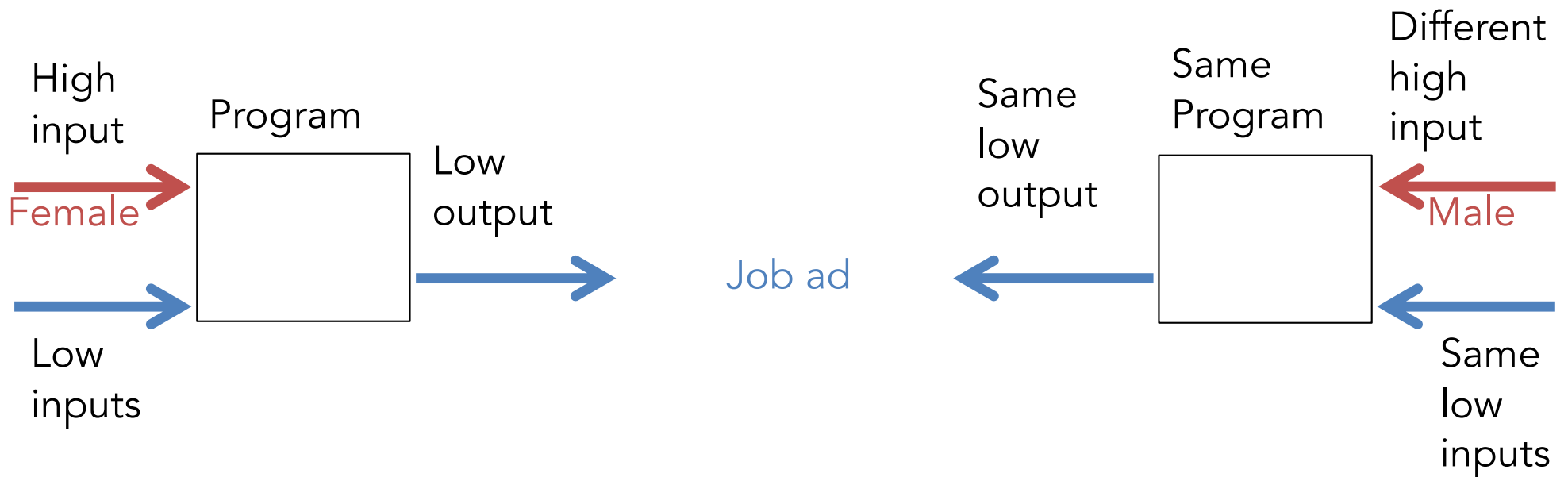
# Information Flow Experiments



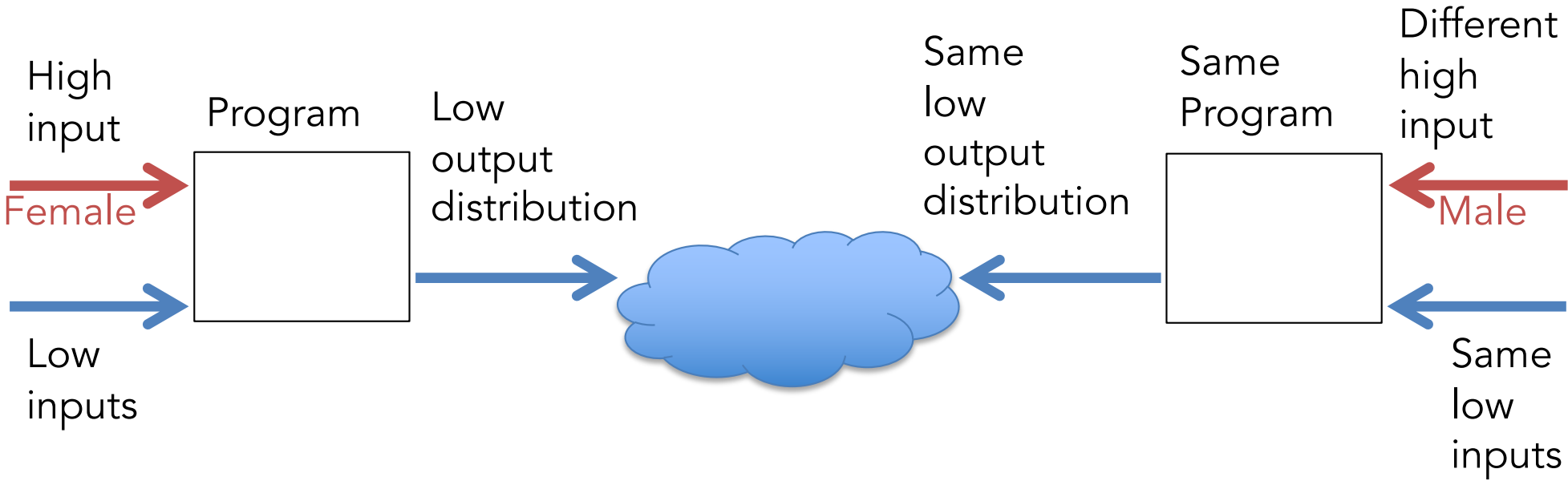
# Information Flow Experiments as Science

Experimental Science	Information Flow
Natural process	System in question
Population of units	Subset of interactions
Treatments	Inputs
Responses	Outputs
...	...
Causation	Information flow

# Noninterference



# Probabilistic Noninterference



# We would like to test the following

- Null hypothesis
  - The outputs from the experimental and control units are drawn from the same distribution
- Apply a significance test
  - On the outputs from the two groups
  - Returns a p-value
  - Reject the null hypothesis if  $p\text{-value} < 0.05$

# First try: Kolmogorov-Smirnov Test

- (on document cam)

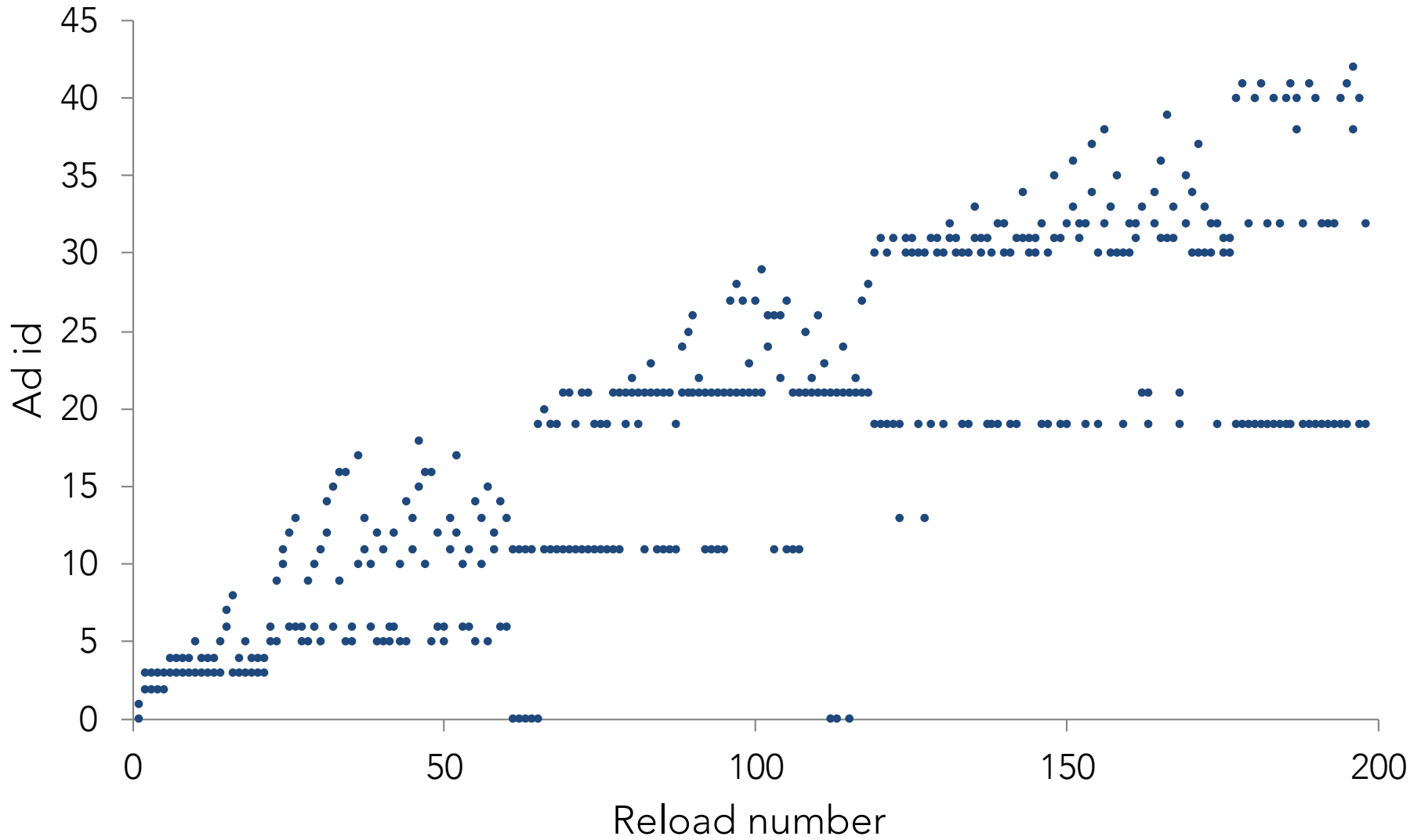
## Things to look out for

### **Beware the Kolmogorov-Smirnov test!**

This article is by Eric Feigelson and G. Jogesh Babu, Center for Astrostatistics, Penn State University.

- Low sensitivity
  - Only considers **maximum** deviation
  - Anderson-Darling Test, Cramer von Mises test
- Do not use to compare with  $F(x)$  derived from your data
- Cannot be applied on higher-dimensional data
- Assumes IID data points

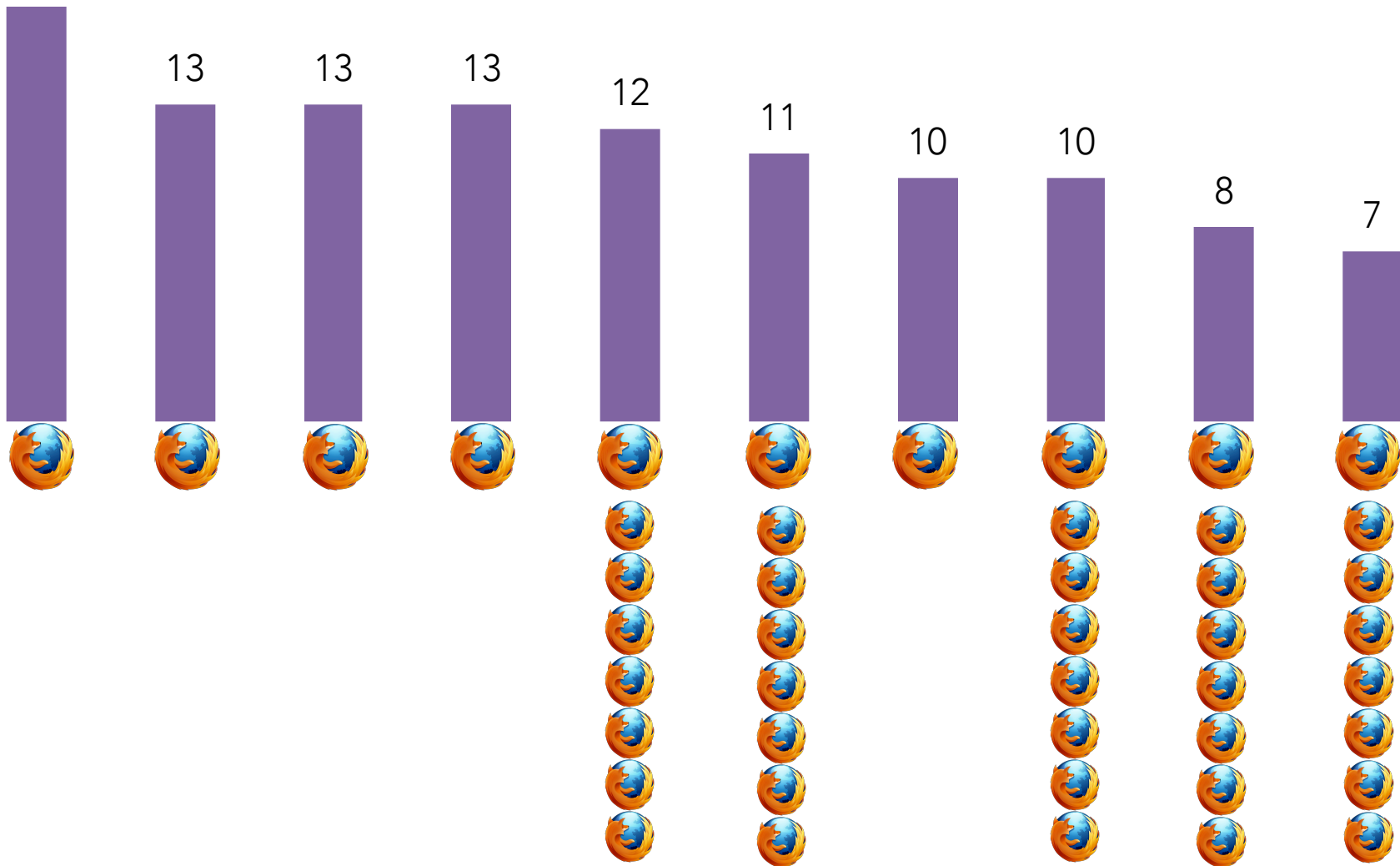
# Mechanism of ad delivery is complex





# Browser agents may not be independent

17

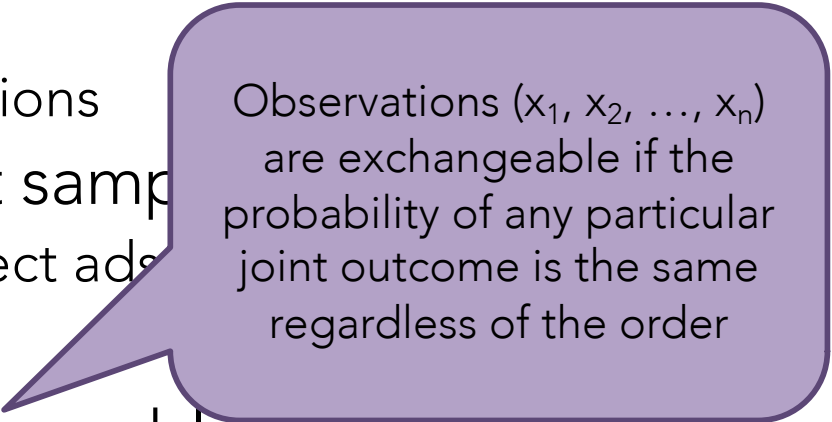


# Key Idea:

- Use a non-parametric test
  - Does not require a model for Google
- Specifically, a permutation test
  - Does not require independence among browser instances or assumption that ads are independent and identically distributed

# Permutation Test [Good'05]

- It is a non-parametric test
  - No assumptions about ad distributions
- It does not require independent samples
  - Ads served to one browser can affect ads served to another
- Assumption: Samples are exchangeable under the null hypothesis
- A statistic that discriminates between the null and alternate hypotheses

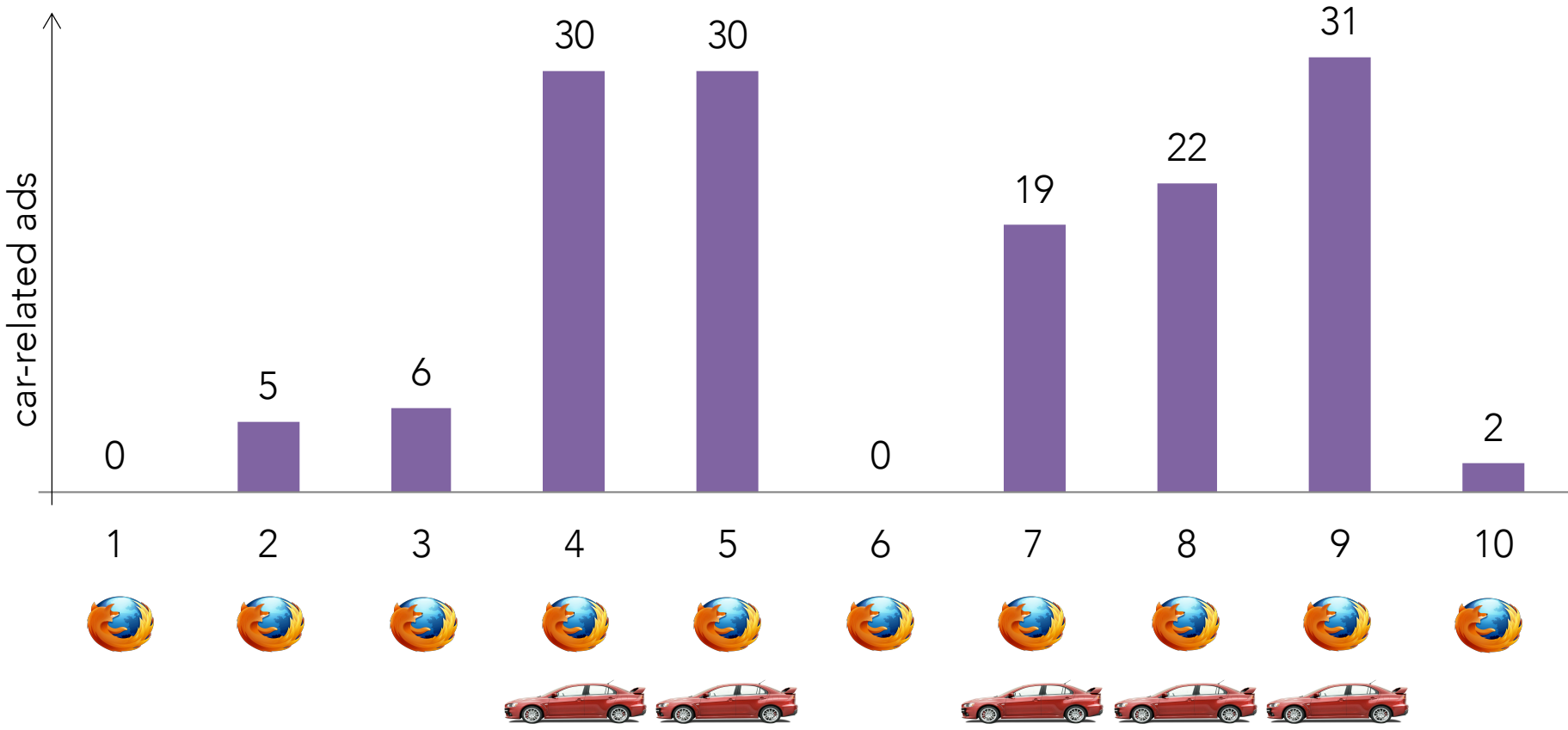


Observations  $(x_1, x_2, \dots, x_n)$  are exchangeable if the probability of any particular joint outcome is the same regardless of the order

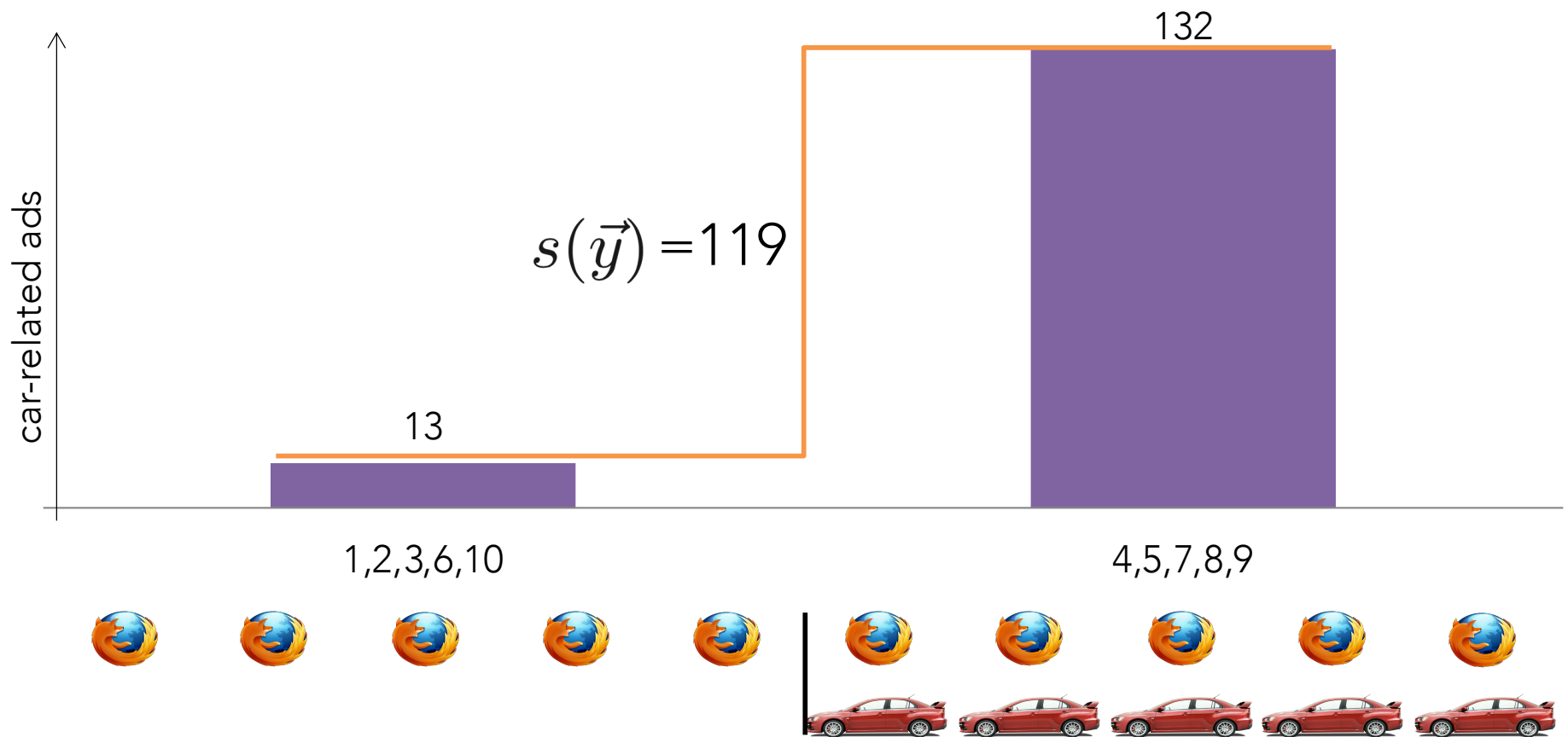
P. Good.

Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses. Springer, 2005

# Permutation Test: Example



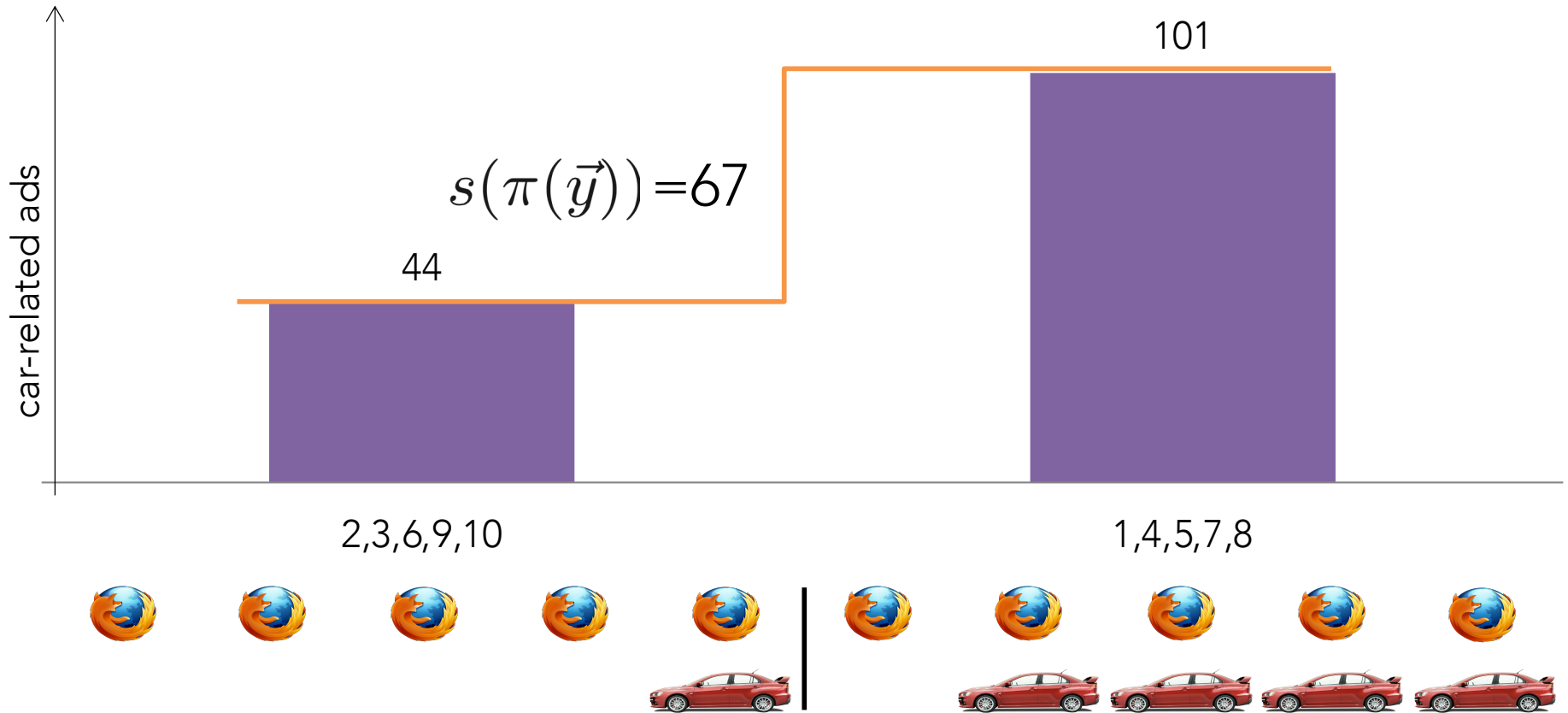
# Permutation Test: Example



$\vec{y}$  is the measurement vector

$s(\vec{y})$  is the statistic computed over  $\vec{y}$

# Permutation Test: Example



$\pi(\vec{y})$  is a permutation of  $\vec{y}$

$s(\vec{y})$  is the statistic computed over  $\vec{y}$

# Permutation Test: Example



$$s(\pi(\vec{y})) = 67$$



$$s(\pi(\vec{y})) = 119$$



$$s(\pi(\vec{y})) = 7$$

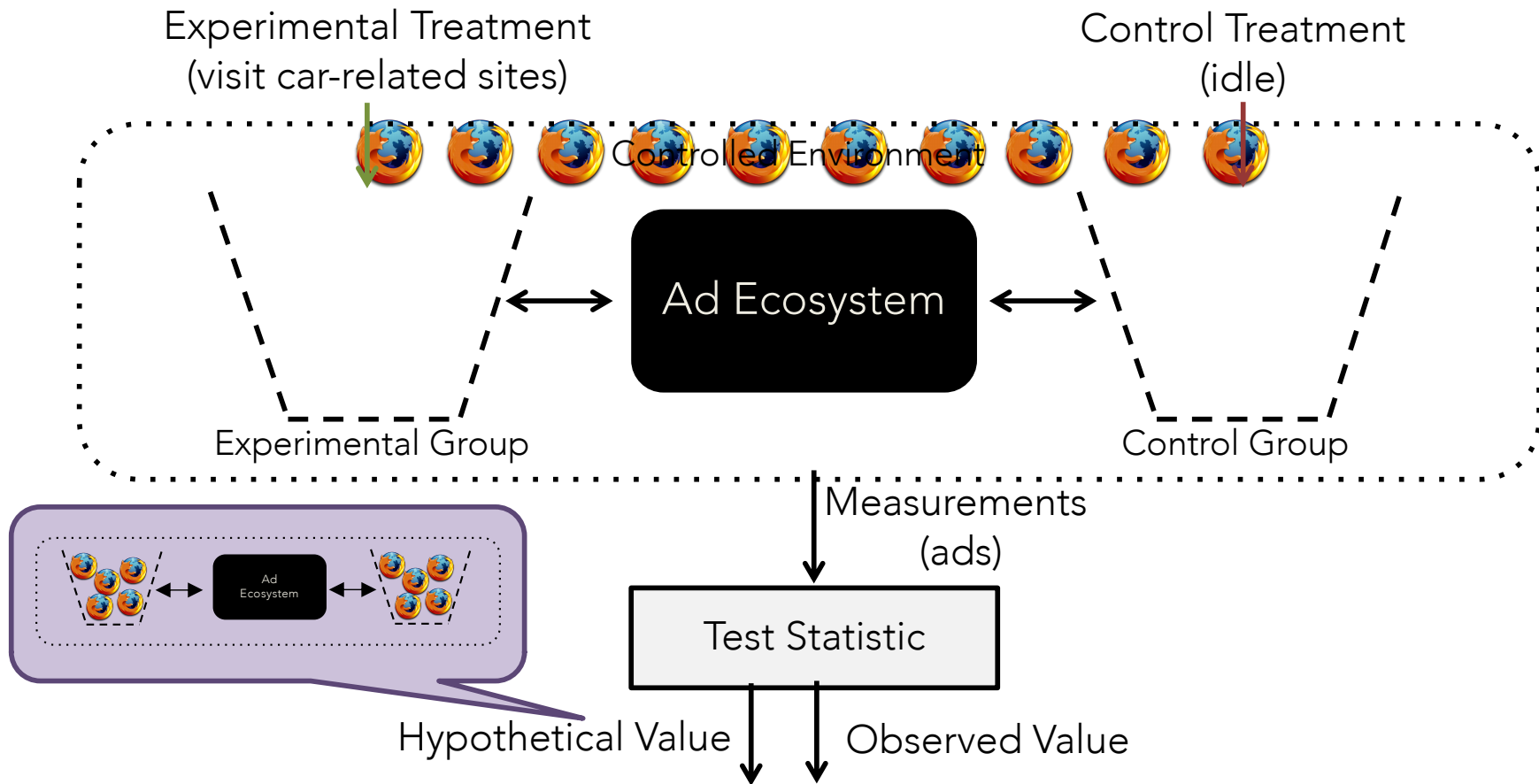
$$\text{p-value} = \frac{\text{count}[s(\vec{y}) \leq s(\pi(\vec{y}))]}{\text{number of unique permutations}} = \frac{1}{{}^{10}C_5} = 0.004$$

$$s(\vec{y}) = 119$$

$\pi(\vec{y})$  is a permutation of  $\vec{y}$

Reject null hypothesis

# Information Flow Experiments





# A rigorous methodology for information flow experiments

- Connection between Information Flow and Causal Experiments
- Statistical principles for designing Information Flow Experiments
  - Control for known confounders
  - Randomize to break unknown confounders
- Significance testing with non-parametric statistical tests