

Digital Experimentation Methods: AB Testing

An Overview

Shan Huang, HKU

About this course

- MSBA7025
- 10 minutes break after each hour
- Instructor: Shan Huang @ KKL 1229
- My Office Hours: by appointment
- TA: Chen Wang



**HKU
BUSINESS
SCHOOL**
港大經管學院

Shan Huang, HKU

About me

- Before HKU
 - Assistant Professor @ University of Washington, Seattle 2018-2020
 - PhD @ MIT Sloan School of Management, 2018
- Digital Fellow @ Stanford Digital Economy Lab
- Consultant @ WeChat, Tencent



About me

- My Research Areas: Digital Economy, Social Media, Social Networks, Business Analytics (e.g., A/B testing)
- *MIT Conference on Digital Experimentation (CODE) since 2014*



CODE
@MIT

About this course

- Applications (industry examples + recent developments) + Statistics
 - Everything will be practical, relevant, and cutting-edge.
 - What you will learn is being used at Tencent, ByteDance, Alibaba, Facebook, Google, and LinkedIn daily
- 4 Individual Assignments ($4 * 10\% = 40\%$ with iPython Notebook)
 - The assignments will be posted on Fridays and due the next Wednesday
- 1 Group Project (40%) - to be posted in the third week.
- Class Participation (20%) - Be prepared for the cold calls!

Course Outline

- Overview
 - Statistics Foundation of Experiments
 - Experimental Design
 - Analyze Experiments (Data)
 - New Developments
- * Guest lectures (industry speakers) - in-class or off-class time

Trustworthy, Fast, and Low-cost
(e.g., less user traffic)

Participation

- Class Attendance (including guest lectures) 10%
- In-Class Participation 10%

Group Project 40%

- Analyze data from an experiment to inform organizational decision makings
- Open questions
- Six people/group (Max 6, 5 is ok)
- Groups will be fixed next Thursday

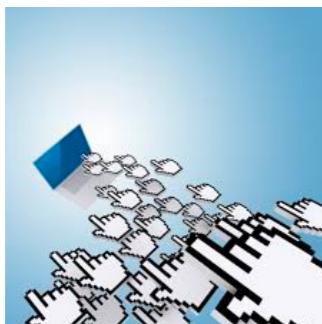
Overview: A/B Testing

These terms are all about Experiments!

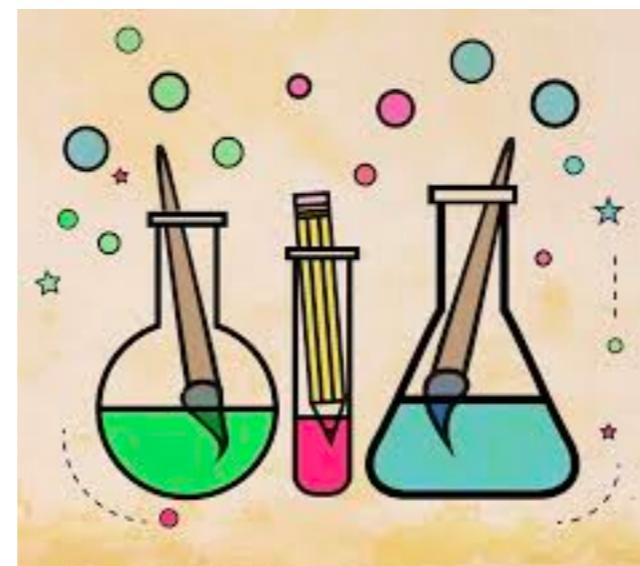
- Digital Experiments
 - Online Controlled Experiments
 - A/B Tests, A/B/n Tests (to emphasize multiple Treatments)
 - Flights (Microsoft), 1% Tests (Google), Bucket tests (Yahoo!)
- Online!**
-
- Field Experiments
 - Randomized Experiments
 - Randomized Controlled Trials (RCT)
- Academic Research**

What is Experiment?

- A procedure carried out to validate a **hypothesis**.
- Provides insight into **cause-and-effect** by demonstrating what **outcomes** occur when a particular factor(s) is manipulated.



Scientific Experiments

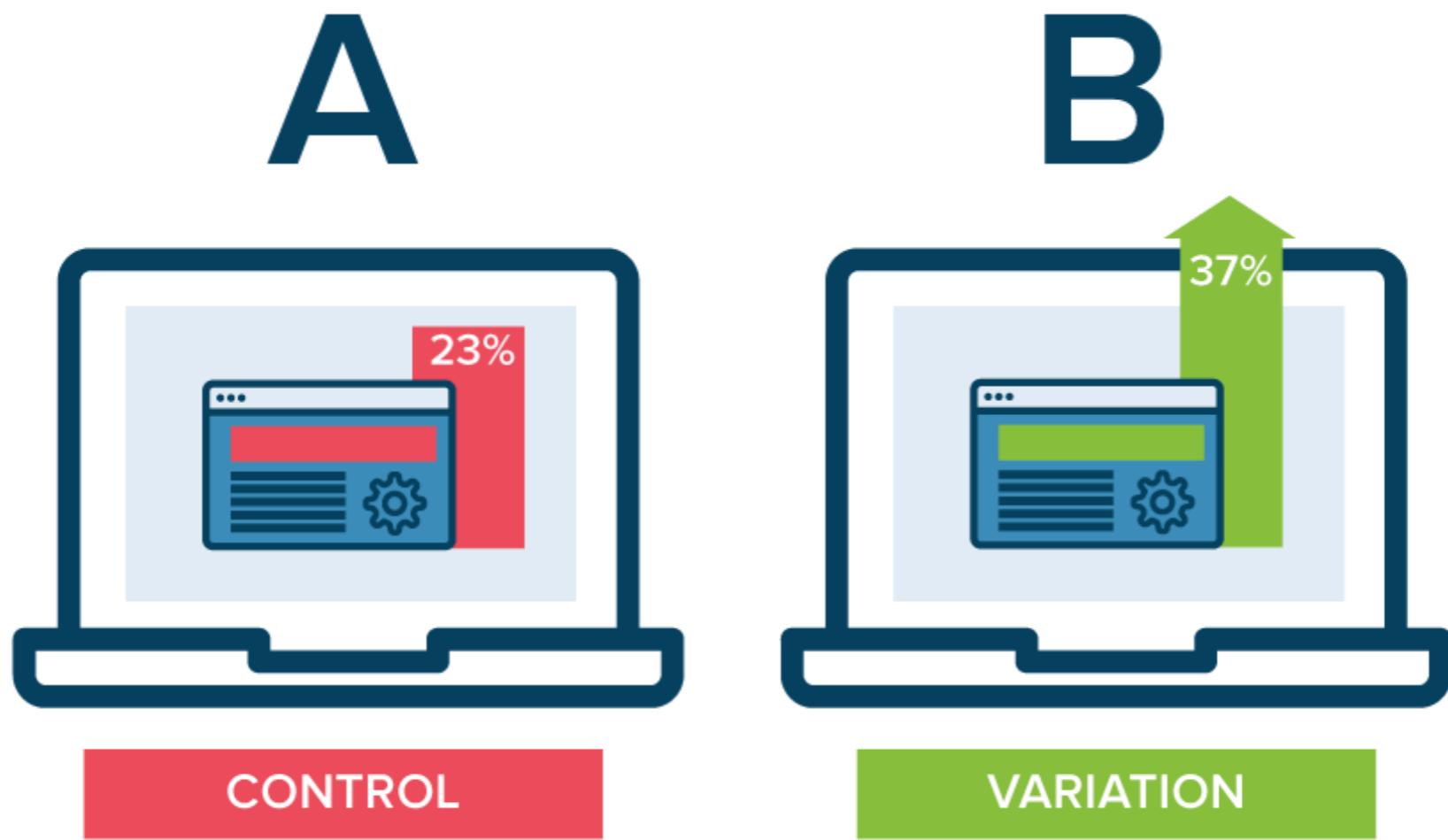


Experiments with Human Subjects



What is A/B Testing?

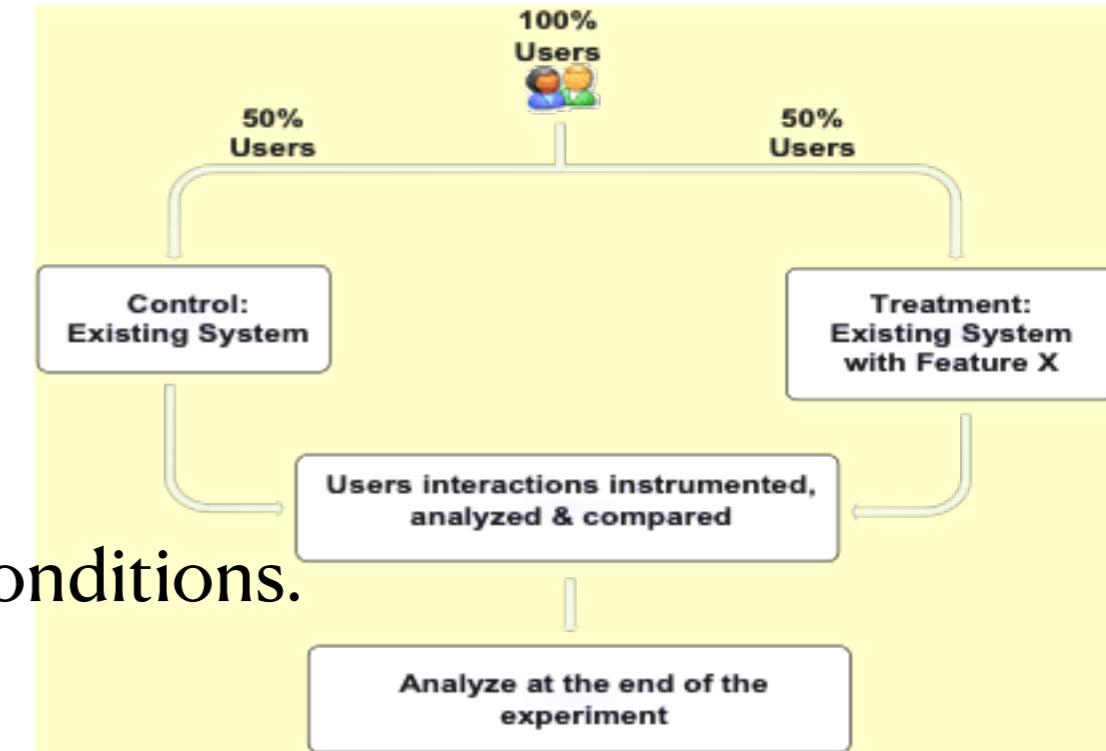
- Industry applications of online controlled experiments
 - Experiments on big user pool and industry system
 - To test different product strategies in daily operations.
 - UI, Algorithms (recommendations, ranking, pricing, etc..), Functions
- Applications: IT industry, Marketing, Fintech, Retailing
- Data Scientists, Product Managers, Engineers, Marketers...
- Social scientists, computer scientists, statisticians....
 - are working together
 - to provide **faster, cheaper, and more trustworthy** experiments.
- **A big step towards digitization!**
 - The machine replaces humans in making *some* decisions
 - Automate *some* decision makings in orgnizations with data.



A/B Testing: Two (Multiple/Many) Treatments

What are A/B Tests?

- Randomly split traffic between two (or more) conditions.
 - A (Control, typically existing system)
 - B (Treatment)
 - A/B/N (multiple treatments) - could be hundreds of conditions
- Collect metrics of interests
- Compare metrics between A and B (N)
- Must run statistical tests to confirm that differences in the sample **do not occur due to chance.**
- **The best** scientific way to prove **causality**, i.e., the changes in metrics are **caused** by changes introduced in the treatment(s)



The Earliest and Simplest A/B Test @ Bing

- An idea was proposed to change the way the ad titles were displayed on Bing
- Moved ad text to the title line to make it longer

Control - Existing Display

WEB IMAGES VIDEOS MAPS SHOPPING LOCAL NEWS MORE

bing MS Beta

flowers

358,000,000 RESULTS

Flowers at 1-800-FLOWERS® 1800Flowers.com Fresh Flowers & Gifts at 1-800-FLOWERS. 100% Smile Guarantee. Shop Now

FTD® - Flowers www.FTD.com Get Same Day Flowers in Hours! Buy Now for 25% Off Best Sellers.

Send Flowers from \$19.99 www.ProFlowers.com Send Roses, Tulips & Other Flowers. "Best Value" -Wall Street Journal. proflowers.com is rated ★★★★☆ on Bizrate (1307 reviews)

50% Off All Flowers www.BloomsToday.com All Flowers on the Site are 50% Off. Take Advantage and Buy Today!

Treatment - New idea called Long Ad Titles

WEB IMAGES VIDEOS MAPS SHOPPING LOCAL NEWS MORE

bing MS Beta

flowers

358,000,000 RESULTS

FTD® - Flowers Get Same Day Flowers in Hours! www.FTD.com Buy Now for 25% Off Best Sellers.

Flowers at 1-800-FLOWERS® | 1800flowers.com 1800Flowers.com Fresh Flowers & Gifts at 1-800-FLOWERS. 100% Smile Guarantee. Shop Now

Send Flowers from \$19.99 - Send Roses, Tulips & Other Flowers. www.ProFlowers.com "Best Value" -Wall Street Journal. proflowers.com is rated ★★★★☆ on Bizrate (1307 reviews)

\$19.99 - Cheap Flowers - Delivery Today By A Local Florist! www.FromYouFlowers.com Shop Now & Save \$5 Instantly.

True Bing Story

- It's one of hundreds of ideas proposed in an org and seems.
 - The implementation was delayed in: Feb, March, April....
- Without clearly knowing its benefit, engineers may delay it forever.
- An engineer thought: this is trivial to implement. He started a controlled experiment (A/B test).
- An alert fired that something is wrong with revenue, as Bing was making too much money.
- But there was no bug. The idea increased Bing's revenue by 12% (over \$120M at the time).

Unexpected Changes

- It is common for a change to have unexpected consequences
- For example, Bing has a related search block for peripheral products
- A small change to the block (say bolding of terms)
 - Changes the click rate to the block (intended effect)
 - Will it cause the distribution of queries to change?
 - Will some queries monetize better/worse than others? (so revenue is impacted)
- These consequences are hard to be predicted before experiments!

Related searches for **harry potter**

[harry potter cast](#)

[harry potter games](#)

[harry potter store](#)

[harry potter books](#)

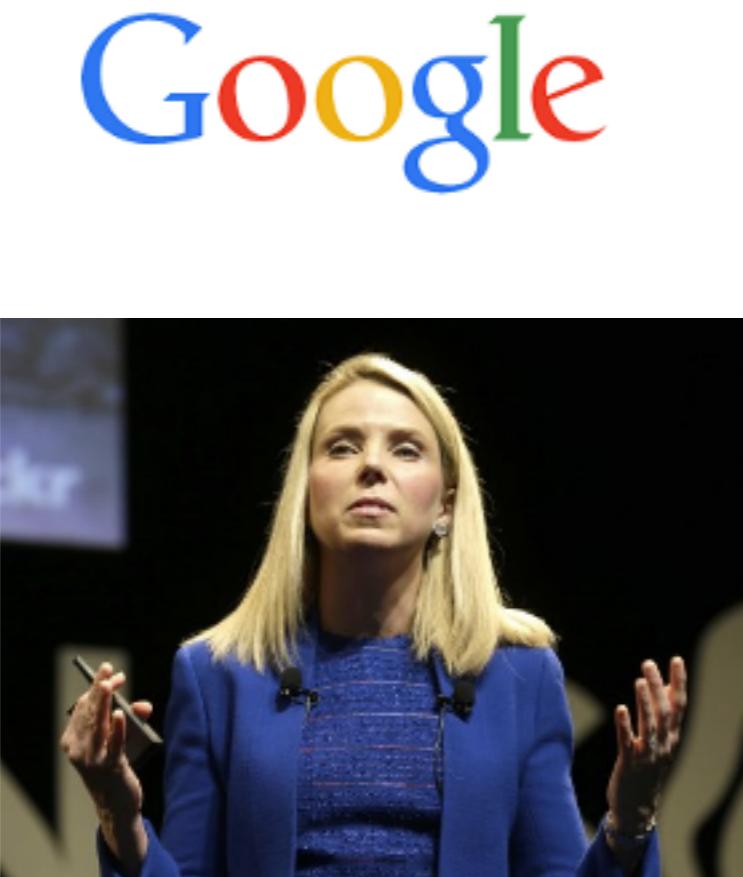
[harry potter characters](#)

[harry potter 8th book](#)

[harry potter movies](#)

[harry potter imdb](#)

TESTING 41 SHADES OF BLUE of Google Search Results



Google (@Google) · Twitter

<https://twitter.com/Google> 

How an inventor in Sweden is using #StreetView paired with a stationary bike to help dementia patients →
goo.gl/avs4sr #SearchOn
pic.twitter.com/jJvOnCE...

4 hours ago · Twitter

Whether a daily walk or once-a-year moment you just can't forget, reminders on Google Home from your #GoogleAssistant are here to help.
pic.twitter.com/WpsI2TN...

7 hours ago · Twitter

To inspire teens to code future, @madewithcode teaming up with @Snapc for a first-of-its-kind proj
goo.gl/JnShsU
pic.twitter.com/2RYpeft...

9 hours ago · Twitter

Facebook, Google probes proliferating in Washington - CNBC.com

<https://www.cnbc.com/.../facebook-google-probes-proliferating-in-washington.html> ▾

2 hours ago - Facebook and Google are facing their day of reckoning in Washington – here's a rundown ... And the growing number of probes portend trouble not only for Facebook but also for its chief rival, Alphabet's Google unit, which previously has also benefited from the disclosure waiver for ...

Google - YouTube

<https://www.youtube.com/user/Google> ▾

Experience the world of Google on our official YouTube channel. Watch videos about our products, technology, company happenings and more. Subscribe to ...

Google - Home | Facebook

<https://www.facebook.com/Google/> ▾

41 Shades of Blue Test

Google would note which color earned more clicks.
And that was how this blue color was chosen.

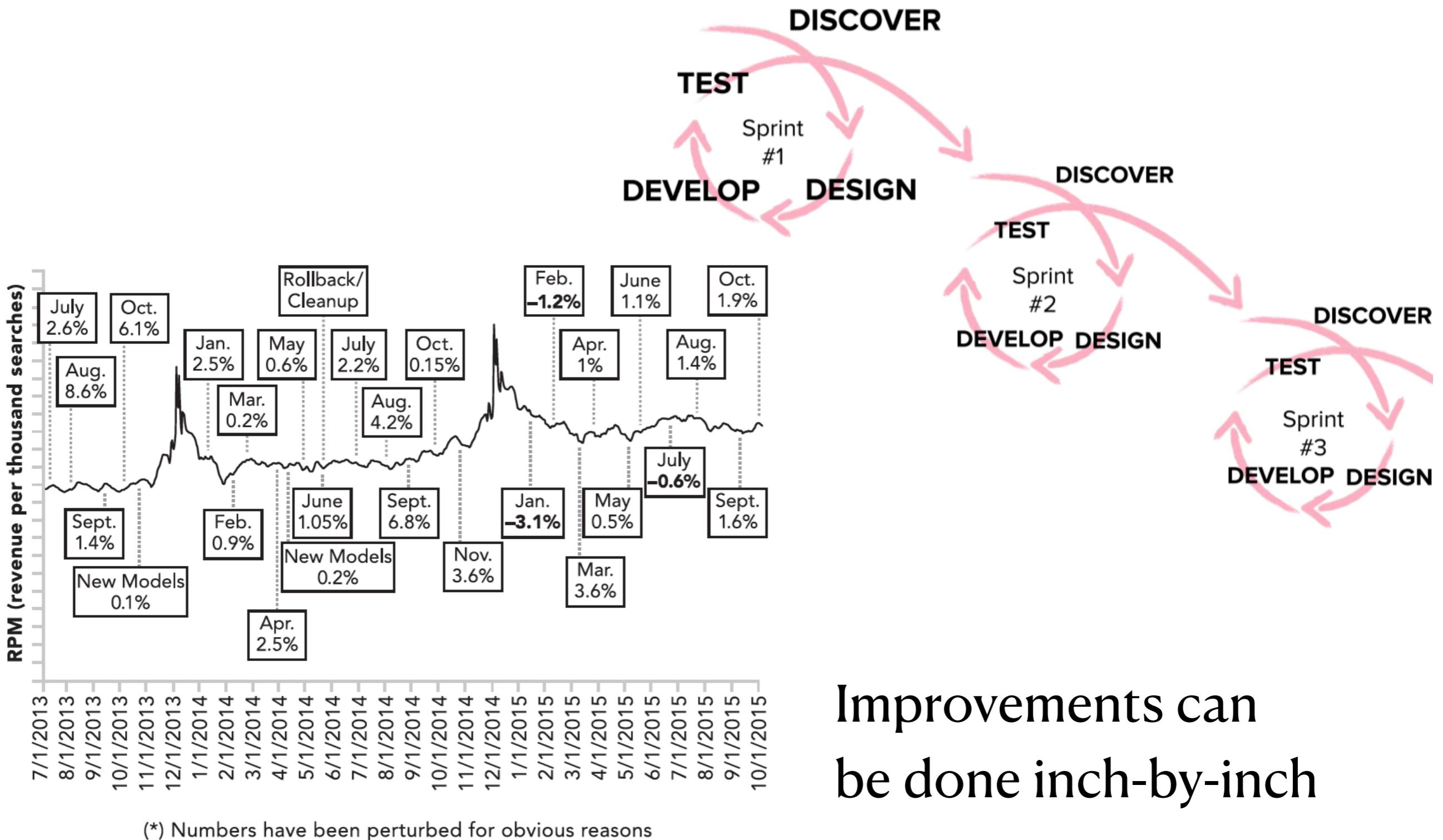


- Randomly assign users into 41 groups and stay for 2 weeks.
- Compare the clickthrough rate across the groups

Key Themes of A/B Tests

- It's hard to assess the value of an idea without experiments, especially for critical product updates!
- Small changes can have a big impact (Bing story)
- *BUT* experiments with big impact are rare!
 - Usually, it will be no effect or small effects.
 - A big impact may happen rarely.
- Small increases can mean a lot for big companies with a large user pool.
- The overhead of running an experiment must be small.
 - Build an Experimentation System (Infrastructure) to automate experiments
- The metrics must be clear.
 - What you aim to improve with the new feature(s) must be *very clear!*

Improvements over Time



Improvements can
be done inch-by-inch

Figure 1.4 Bing Ad Revenue over Time (y-axis represents about 20% growth/year). The specific numbers are not important

The HiPPO



- **HiPPO:** Highest Paid Person's Opinion.
- We want a Data-driven HiPPO
- Listen to the customers, and don't let the HiPPO kill good ideas.
- **The most reliable data for *causality* comes from controlled experiments!**

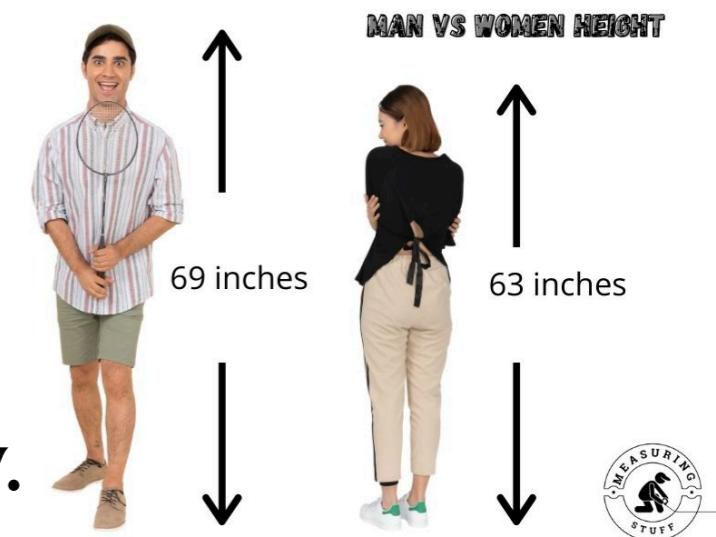
Why Best Data Come from Experiments

Observational Studies

- An observational study is not controlled or with no randomization.
- Controlled: units are in different conditions.
- Randomization: Randomly assignment of units (e.g., users) into control and treatment groups.
- *Question:*
 - Can experiments be randomized but not controlled?

Observation

- The shorter a person is, the longer they will live.
- You wouldn't have believed that people's height is causal for life expectancy.
 - Women are often shorter and live longer.
 - Gender is a common cause of height and life expectancy.
 - Height correlates with your life expectancy.



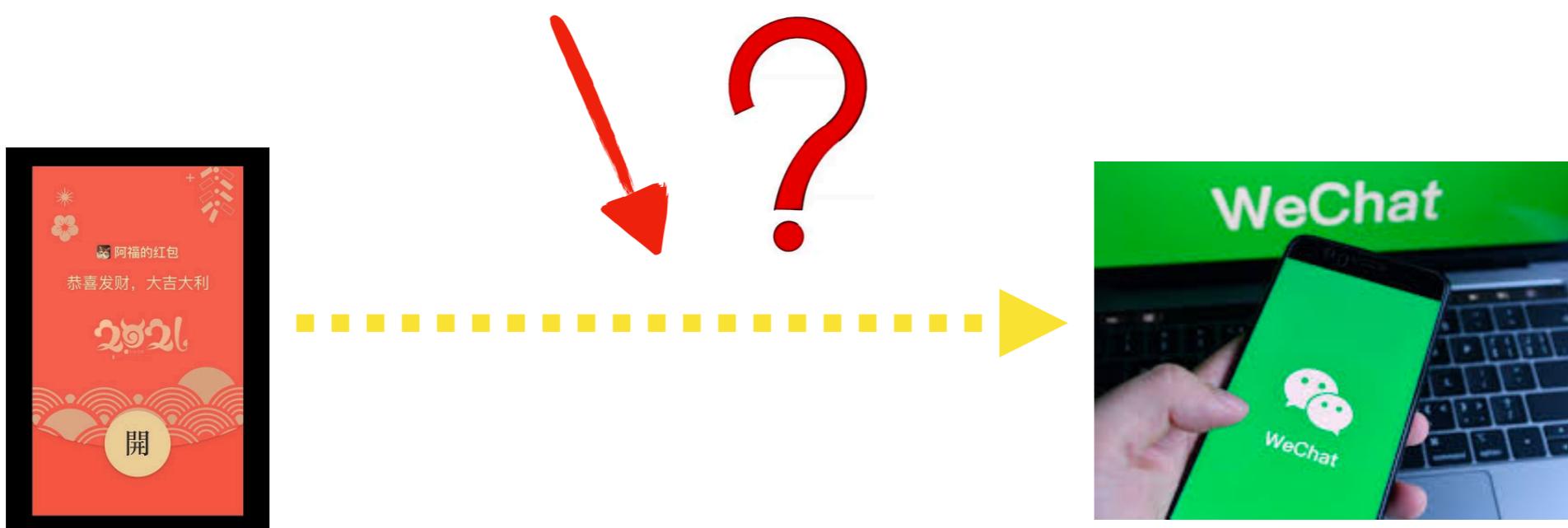
You cannot change the outcome (B) through
manipulating A,

If you only know A **correlates** B.

For example, you cannot increase your life
expectancy by becoming shorter.

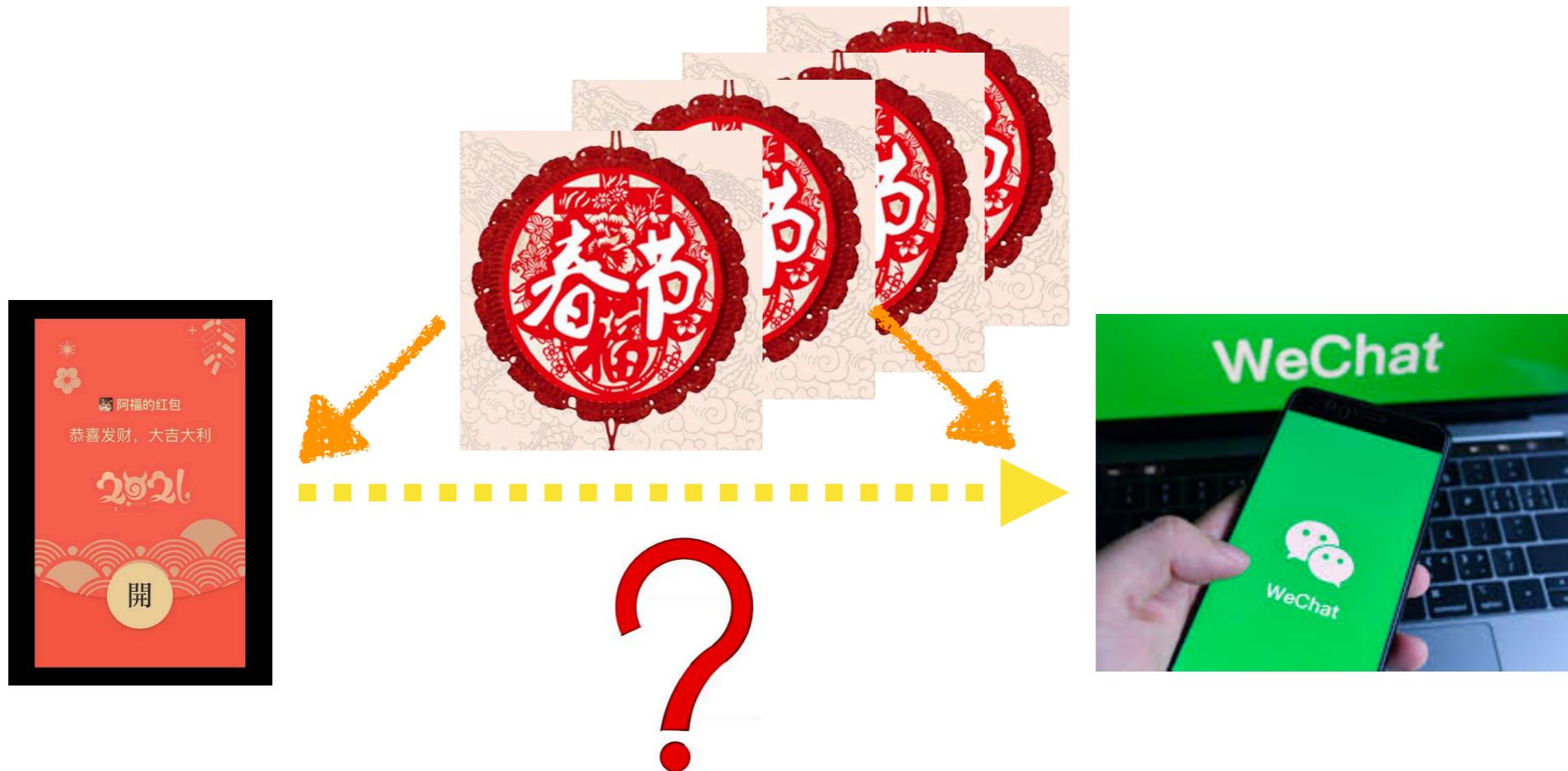
Compare before/after launching a new feature

- New Feature: WeChat Red Pocket
- If we compare #Daily Active Users before and after launching WeChat Red Pocket in Jan 2014
- # DAU: The total number of people who open and engage with a mobile app, a web product, or a feature in a given day.
- Difference = #DAU (after Red Pocket) - #DAU (before Red Pocket) ??



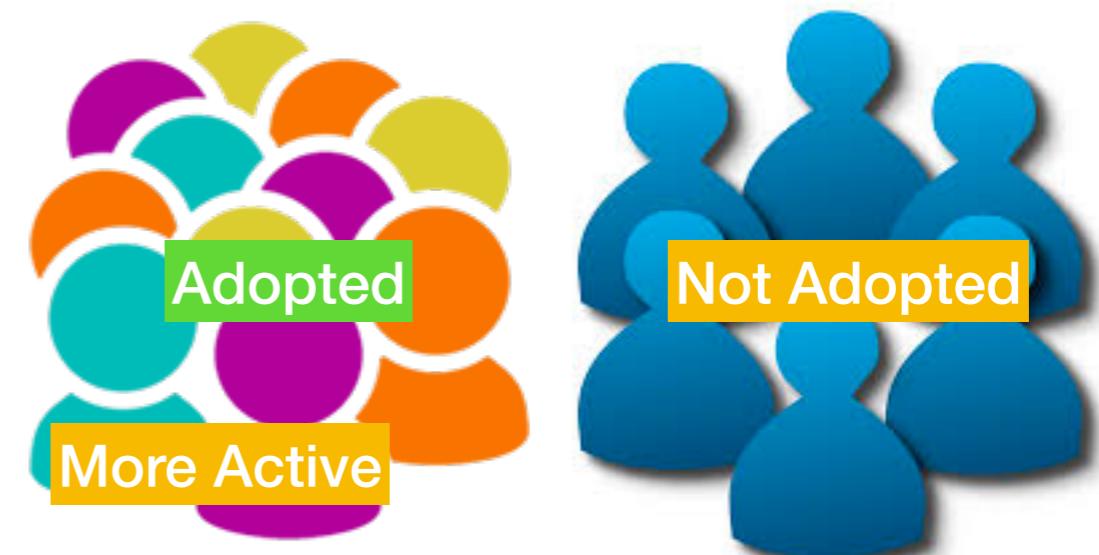
Compare before/after launching a new feature

- Can you conclude that launching Red Pocket increase #DAU on WeChat?
- Is this observational study controlled?
- OR with randomization?



Compare users who adopted/not new feature

- Compare users who adopted vs. who did not adopt.
 - If we find that users who adopted the new feature used more of the product
- Can we conclude that
 - New feature improved usage?

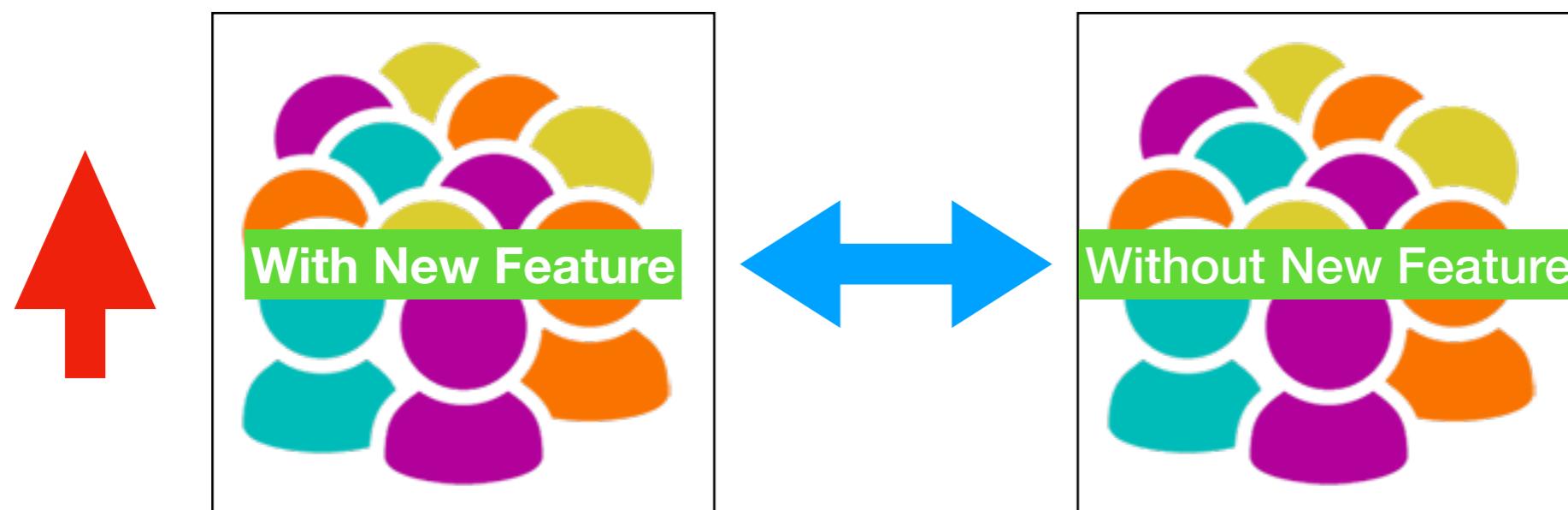


**Is this observational study controlled OR with randomization?
You did not randomize; these two groups are not homogenous.**

If we ***randomly assign*** users to different conditions, this will be a **randomized experiment** rather than an observational study,

& we will be able to apply the word “***causes***” rather than “correlates” to any statistically significant difference.

Randomly Split into Two Groups



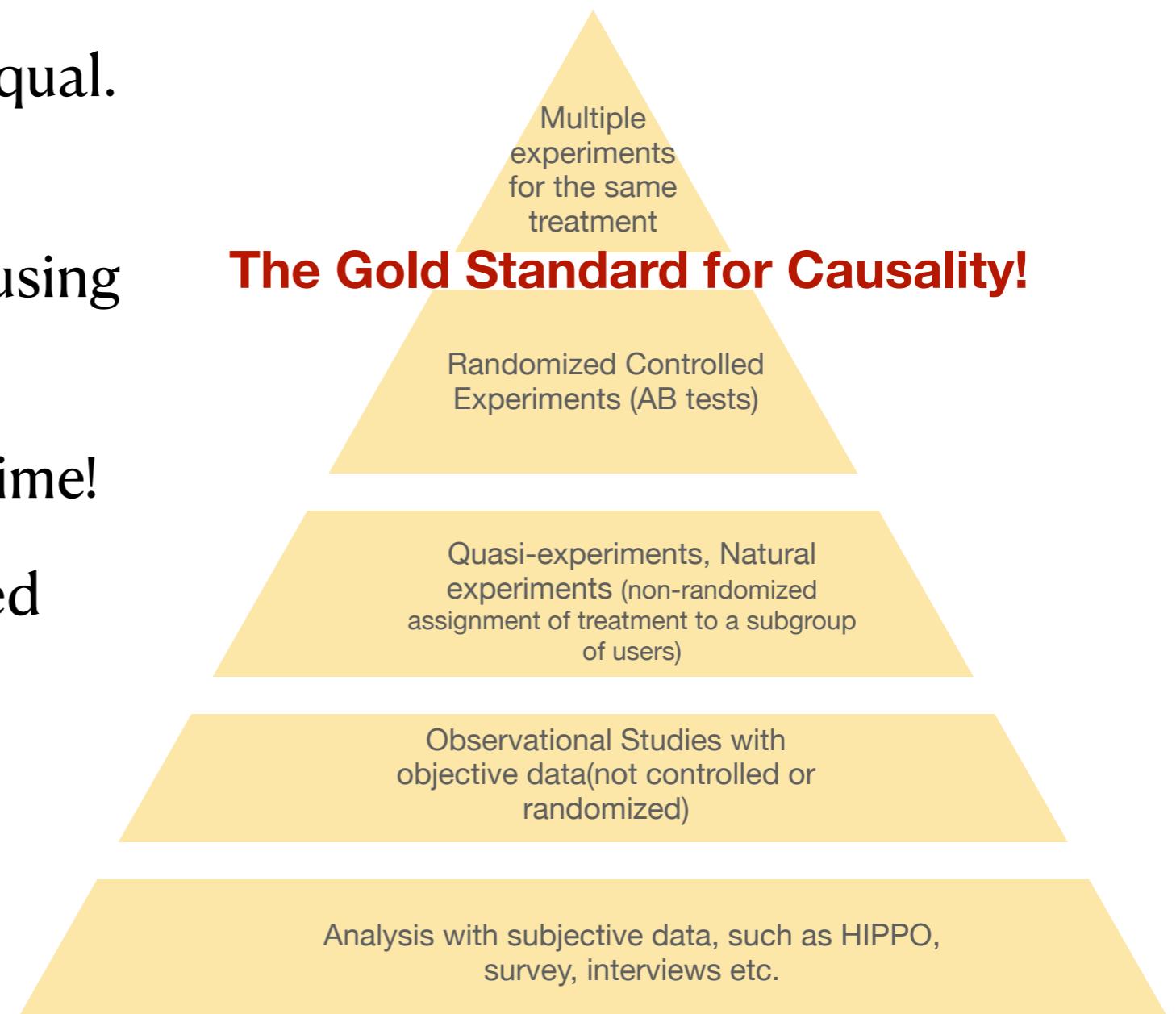
Randomization creates homogeneous groups.

You can change the outcome (B) through manipulating A, only if you know A causes B.
For example, you can increase user engagement by adding a new feature.

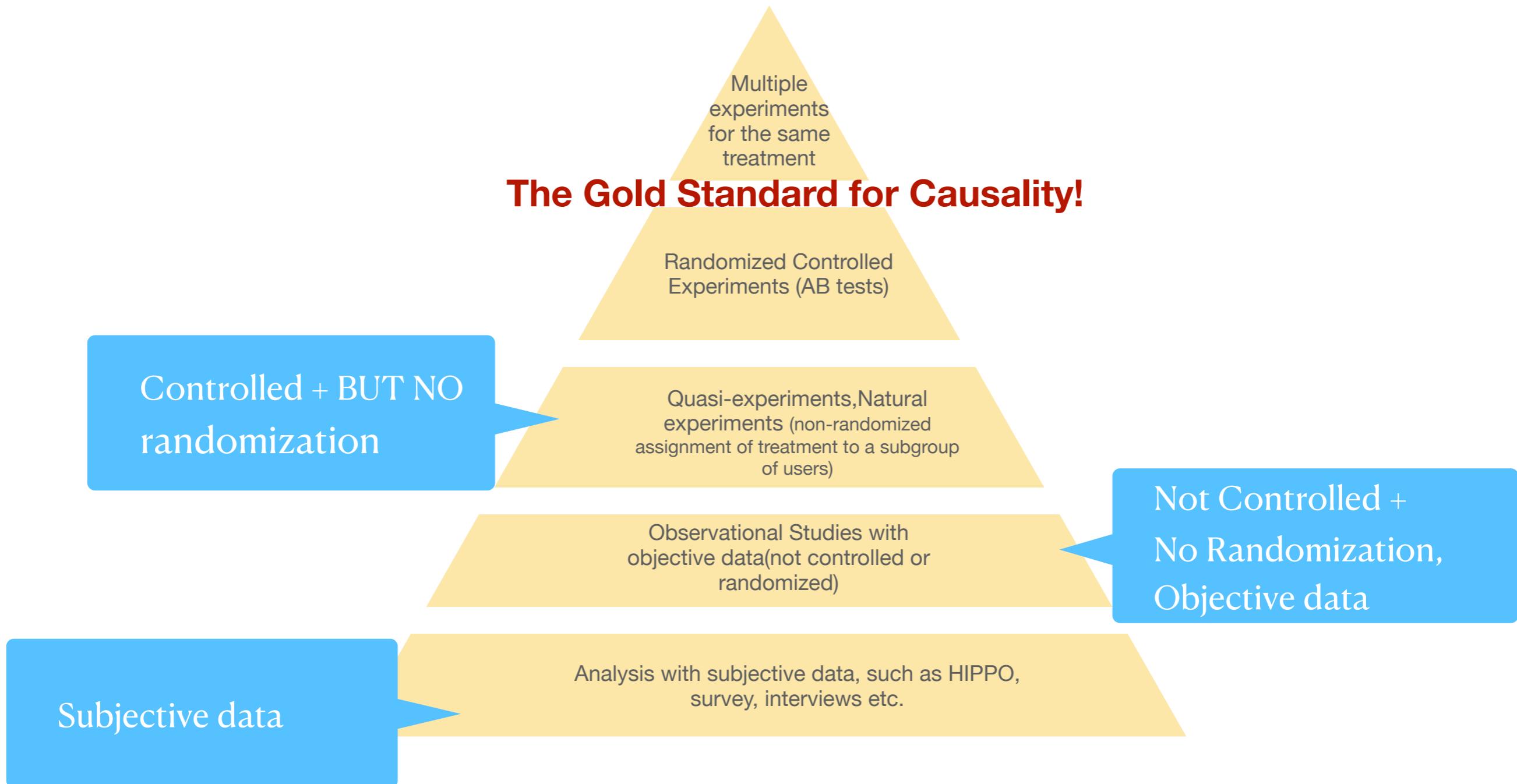
You need causal evidence,
if you want to decide on a product
update/strategy to achieve better
performance.

Hierarchy of Causal Evidence

- All evidence is not created equal.
- You should be very skeptical about concluding causality using observational data.
 - Wrong for most of the time!
- **Gold Standard:** Randomized Controlled Experiments
- Even higher: Multiple Experiments with the same treatments



Hierarchy of Evidence



Terminology for Experiment

- Metrics (OEC)
- Parameter
- Conditions (control and treatment conditions/groups)
- Randomization Unit

Metrics

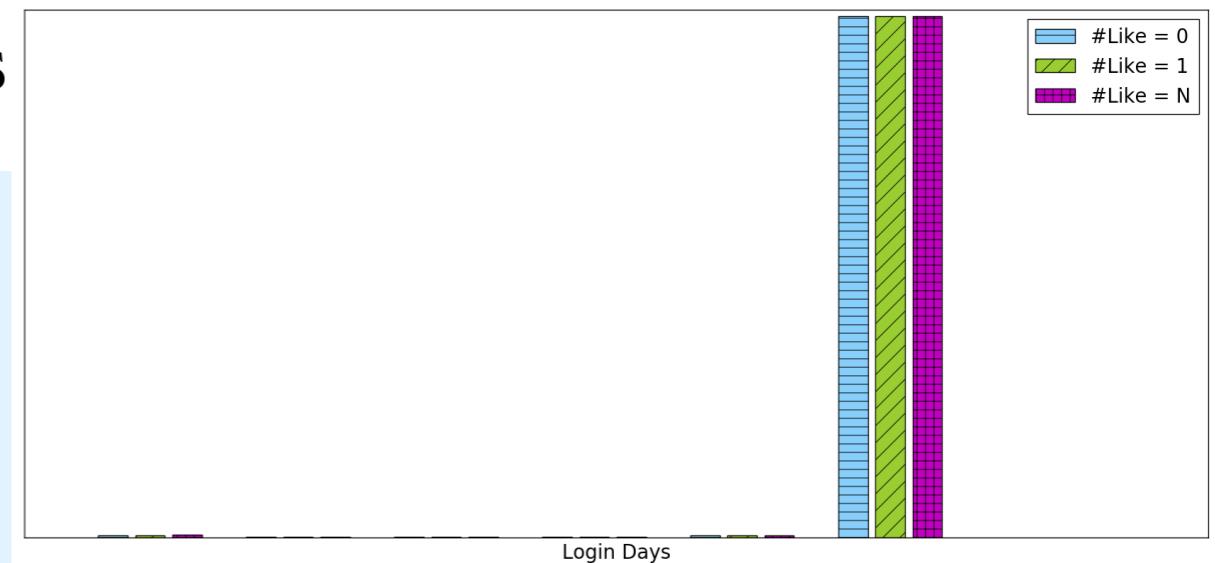
- Overall Evaluation Criterion (OEC), Outcome, Dependent Variable, Metrics
- What *outcomes* do you aim to improve through the new feature?
 - Ad click-through rate, CTR, Purchase rate, #Daily Active User (DAU), #Monthly Active User (MAU)
 - *Quantitative* measures of the experiment's objective.
 - How the treatment will make a difference on the metrics.
- Getting agreement on Good Metrics in the org is a huge first step forward.
- Orgs should accumulate their understanding of Metrics and use AB tests to improve them.

Good Metrics

1. Measurable in the short-term (during experiments)
2. Sensitive to show differences (bad example: WeChat log-in days)
3. Drive long-term strategic objectives

- e.g. Online Advertisements OECs
 - Click-through Rate
 - $CTR = \# \text{ clicks}/\text{impression}$
 - pay-per-click (PPC)
 - Dislike (Hide) Ads

Measurable during experiments



Can maximizing the metrics improve long-term goals?

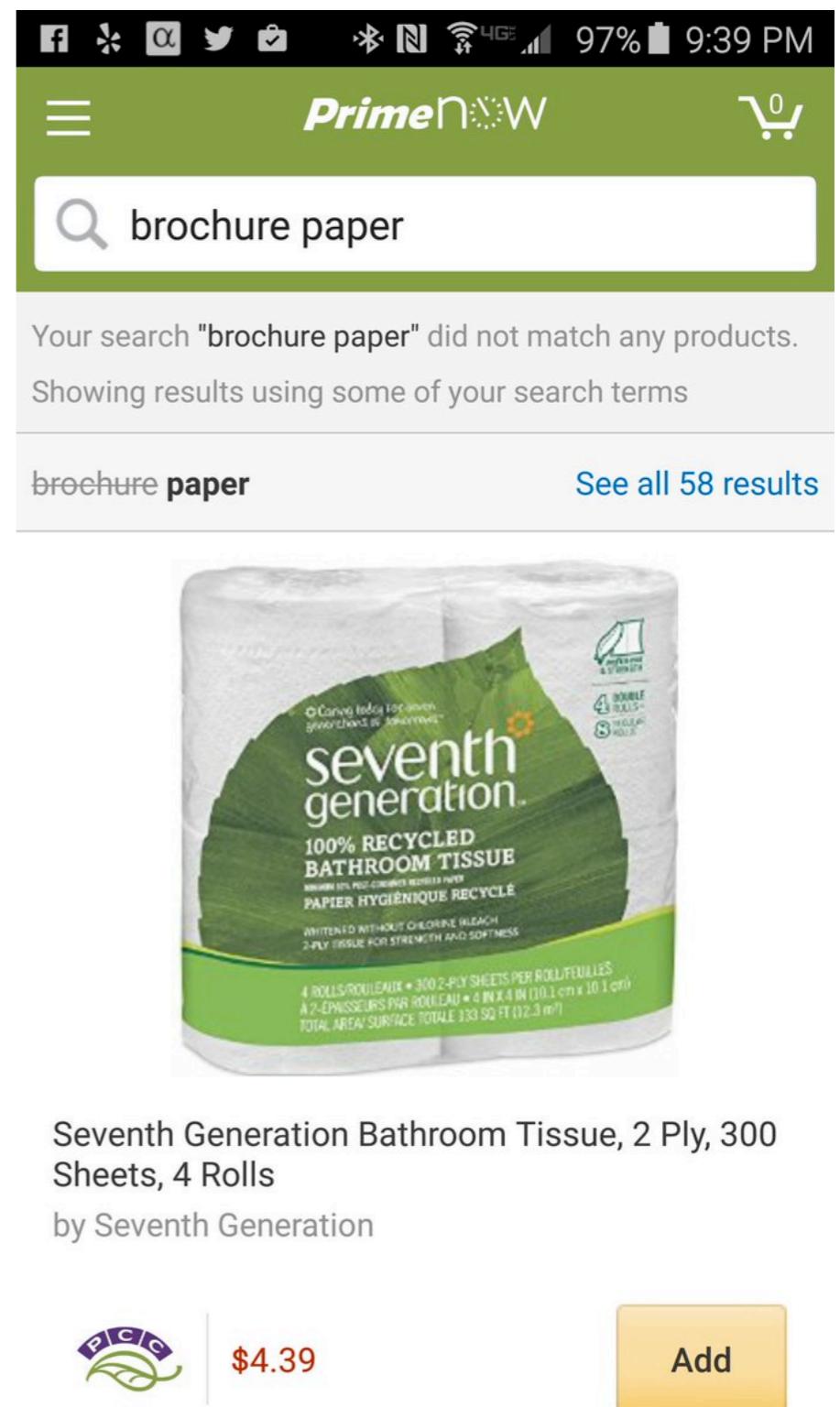
- Platform User Experience
- Long-term Ads Revenue

Bad Metrics Example

- Increase the number of ads/user:
 - It will increase total #clicks and thus short-term ads revenue.
 - **However**, users will likely be overwhelmed with ads
 - Hurt user experience (user activeness)
 - Hurt long-term ads revenue (why?)

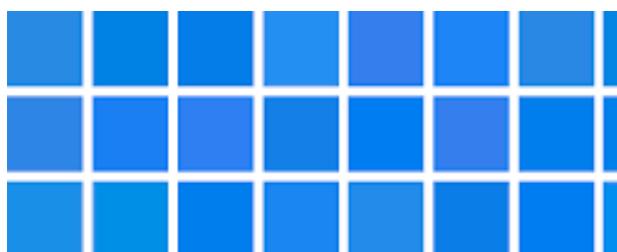
Bad OEC Example

- Your data scientists makes an observation:
2% of queries end up with “No results.”
- Manager: must reduce.
 - Assigns a team to minimize “no result”
- Metric improves, but results for query for brochure paper are crap
- Sometimes it **is** better to show “No Results.”
- **The objective of this experiment does not serve the long-term goal of user experience.**



Parameters

- **Factors or Variables** that you want to decide on and will manipulate in the experiment
 - e.g., you want to decide on the color of the searching results.
- You will assign **values** to the **Parameter: Color**
- **Values** = {blueo, blue1, blue2, blue3...blue_n}
- You want to find which color results in the highest OEC (e.g., Clicks, #DAU)
- e.g., Parameter: font size
- values = {16, 22, 28, 40}



Font 10
Font 16
Font 22
Font 28
Font 34
Font 40

Google

HKU MSBA programs

All Images News Maps Videos More Settings

About 2,930 results (0.57 seconds)

[www.fbe.hku.hk › msba](#)

HKU Master of Science in Business Analytics

The HKU MSc in Business Analytics programme is designed for individuals who are interested in theories and methods of analysing large volumes of data to gain competitive advantages and strategic insights in today's business world.

[Why Master of Science in...](#) · [About The Programme](#) · [Application Procedures](#)

Parameters

- Parameters: the parameters for algorithms
- Example: An experiment to decide on the parameters (k_1, k_2, b) for a algorithm that optimizes the content relevance to a query
 - Variants = different combinations of (k_1, k_2, b)

Google search results for "HKU MSBA programs". The search bar shows "HKU MSBA programs". Below it are filters for "All", "Images", and "News". The snippet from the HKU website includes the formula for relevance $R(q_i, d) = \frac{f_i \cdot (k_1 + 1)}{f_i + K} \cdot \frac{qf_i \cdot (k_2 + 1)}{qf_i + k_2}$ and the formula for $K = k_1 \cdot (1 - b + b \cdot \frac{dl}{avgdl})$.

HKU MSBA programs

All Images News

About 2,930 results (0.57 seconds)

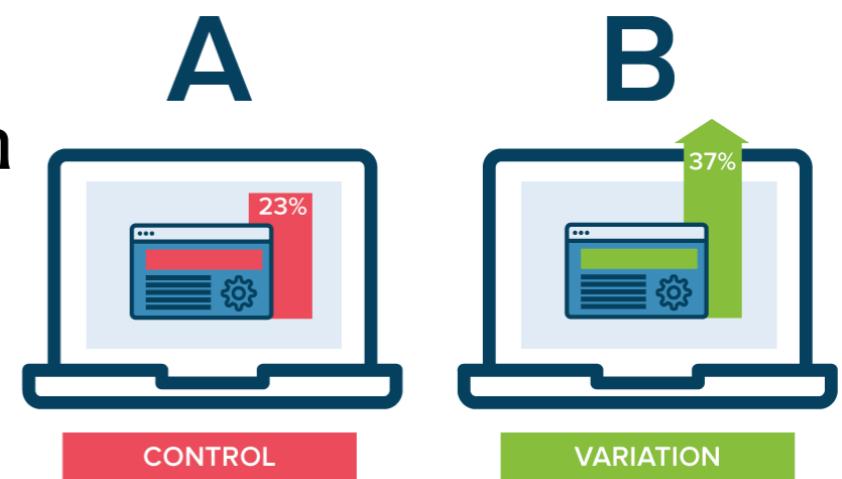
[www.fbe.hku.hk › msba](http://www.fbe.hku.hk/msba)

HKU Master of Science in E
The HKU MSc in Business Analytics | interested in theories and methods of ai advantages and strategic insights in tod...
[Why Master of Science in...](#) · [About The Programme](#) · [Application Procedures](#)

$$R(q_i, d) = \frac{f_i \cdot (k_1 + 1)}{f_i + K} \cdot \frac{qf_i \cdot (k_2 + 1)}{qf_i + k_2}$$
$$K = k_1 \cdot (1 - b + b \cdot \frac{dl}{avgdl})$$

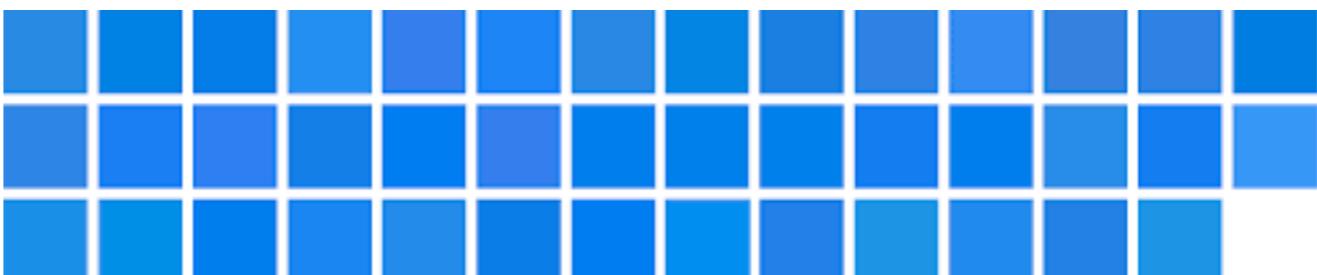
Conditions

- Variants, Conditions, Groups, e.g., control, treatment groups/conditions
- A condition is a group with a specific value of **one** parameter
 - # Values = # Conditions (for the parameter)
- In a simple A/B test, A and B are the two variants,
 - A is called Control Group/Condition
 - B is called Treatment Group/Condition
 - Multiple Treatments: A/B/n Tests
 - Multiple Controls: A/A Tests



Conditions : Example

- e.g., the experiment is to optimize the searching results' color
 - Searching results are shown in 41 shades of blue in the experiment
 - 41 variants with values = {blue0, blue1, blue2, blue3...blue_40}
 - Control - A special variant - e.g., blue0
 - The existing value of the parameter
 - Treatments - Variants other than the control
 - The new blues



Bad Example

- Some conditions have different values for another parameter (e.g., different color + different font).
- *Muti-variate Experiment* (multiple parameters) should be applied here.
 - Two values for colour and three values for font.
 - How many conditions will be involved in the experiment?

	Blue	Red
12	A	B
14	C	D
16	E	F

Randomization Unit

A

bing MS Beta

WEB IMAGES VIDEOS MAPS SHOPPING LOCAL NEWS MORE

flowers

358,000,000 RESULTS

Flowers at 1-800-FLOWERS® 1800Flowers.com Fresh Flowers & Gifts at 1-800-FLOWERS. 100% Smile Guarantee. Shop Now

FTD® - Flowers www.FTD.com Get Same Day Flowers in Hours! Buy Now for 25% Off Best Sellers.

Send Flowers from \$19.99 www.ProFlowers.com Send Roses, Tulips & Other Flowers. "Best Value" -Wall Street Journal. proflowers.com is rated ★★★★★ on Bizrate (1307 reviews)

50% Off All Flowers www.BloomsToday.com All Flowers on the Site are 50% Off. Take Advantage and Buy Today!

B

bing MS Beta

WEB IMAGES VIDEOS MAPS SHOPPING LOCAL NEWS MORE

flowers

358,000,000 RESULTS

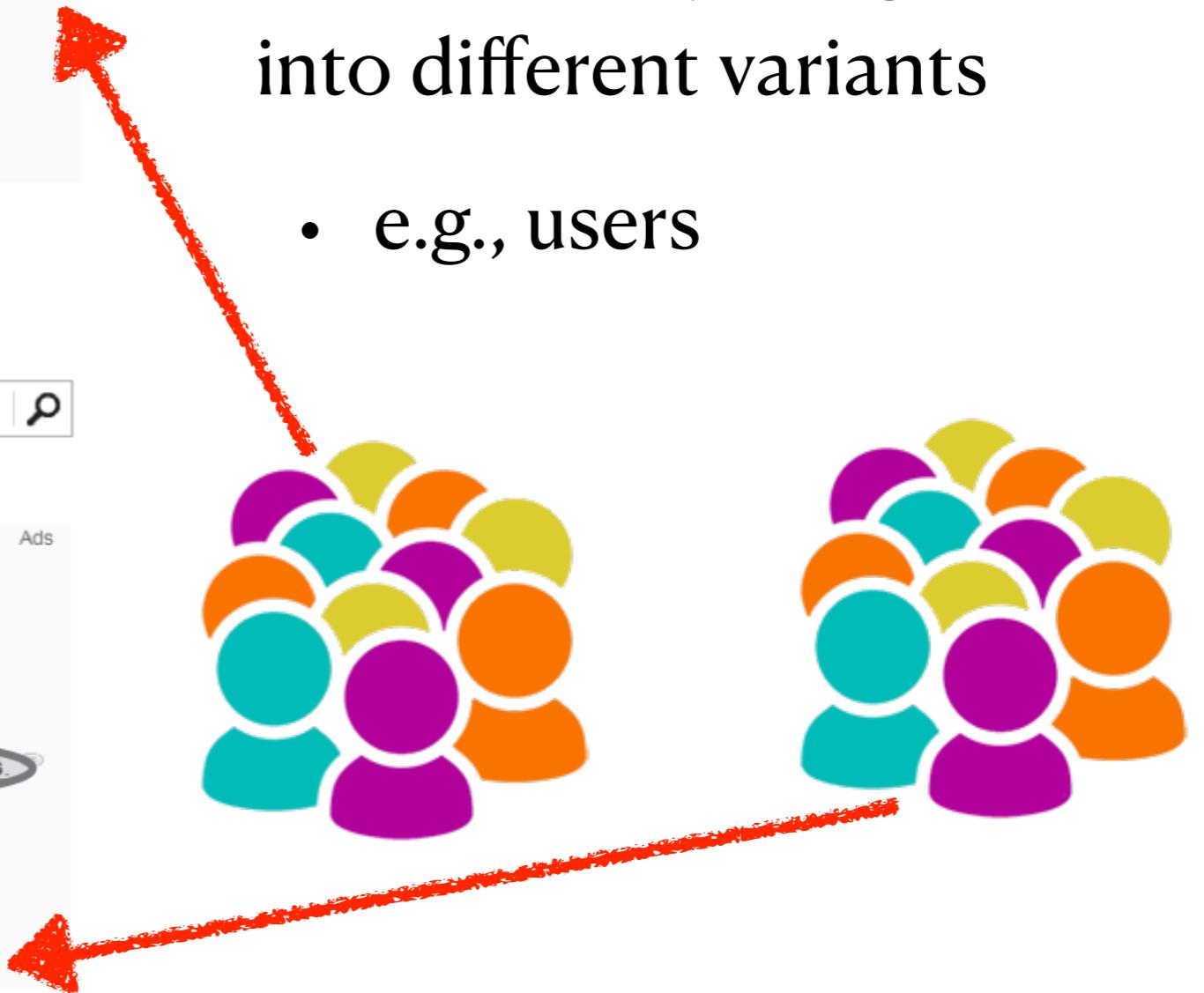
FTD® - Flowers Get Same Day Flowers in Hours! www.FTD.com Buy Now for 25% Off Best Sellers.

Flowers at 1-800-FLOWERS® | 1800flowers.com 1800Flowers.com Fresh Flowers & Gifts at 1-800-FLOWERS. 100% Smile Guarantee. Shop Now

Send Flowers from \$19.99 Send Roses, Tulips & Other Flowers! www.ProFlowers.com "Best Value" -Wall Street Journal. proflowers.com is rated ★★★★★ on Bizrate (1307 reviews)

\$19.99 - Cheap Flowers - Delivery Today By A Local Florist! www.FromYouFlowers.com Shop Now & Save \$5 Instantly.

- A randomization procedure will be applied to units
- You randomly assign units into different variants
- e.g., users



Two Principles

- **Consistent**
 - A unit should consistently experience the same condition
 - Otherwise, a unit experience an inconsistent condition.
 - We cannot associate the outcomes with the condition
- **Independent**
 - A unit's experience should not be affected by the other units' conditions (variants)
 - Otherwise, a unit experiences mixed conditions.
 - We cannot associate the outcomes with the variant.

Which is a good Randomization Unit?

WEB IMAGES VIDEOS MAPS SHOPPING LOCAL NEWS MORE

bing
MS Beta

flowers

358,000,000 RESULTS

Flowers at 1-800-FLOWERS®  Ads
[1800Flowers.com](http://www.1800Flowers.com)
Fresh Flowers & Gifts at 1-800-FLOWERS. 100% Smile Guarantee. Shop Now

FTD® - Flowers 
www.FTD.com
Get Same Day Flowers in Hours! Buy Now for 25% Off Best Sellers.

Send Flowers from \$19.99 
www.ProFlowers.com
Send Roses, Tulips & Other Flowers. "Best Value" -Wall Street Journal.
proflowers.com is rated ★★★★ on Bizrate (1307 reviews)

50% Off All Flowers 
www.BloomsToday.com
All Flowers on the Site are 50% Off. Take Advantage and Buy Today!

- Users vs. Pages - the most common two!
 - Users are assigned to different conditions see different headlines



WEB IMAGES VIDEOS MAPS SHOPPING LOCAL NEWS MORE

bing®
MS Beta

flowers

358,000,000 RESULTS

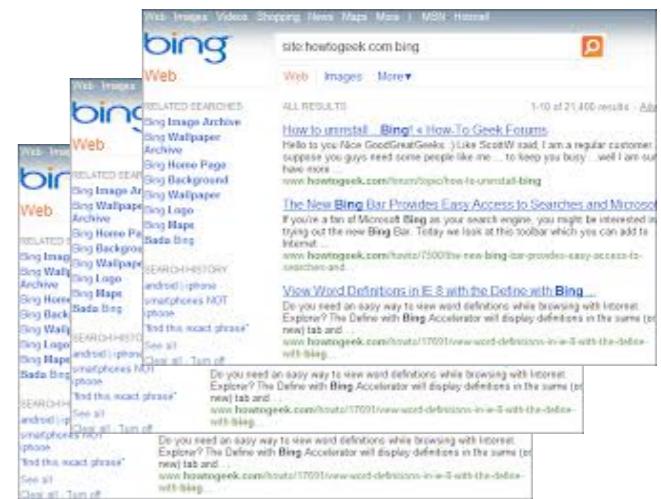
FTD® - Flowers Get Same Day Flowers in Hours! Ad
www.FTD.com
Buy Now for 25% Off Best Sellers.

Flowers at 1-800-FLOWERS® | 1800flowers.com 
1800Flowers.com
Fresh Flowers & Gifts at 1-800-FLOWERS. 100% Smile Guarantee. Shop Now

Send **Flowers** from \$19.99 - Send Roses, Tulips & Other **Flowers**
www.ProFlowers.com
"Best Value" -Wall Street Journal.
proflowers.com is rated ★★★★☆ on Bizrate (1307 reviews)

\$19.99 - Cheap **Flowers** - Delivery Today By A Local Florist! 
www.FromYouFlowers.com
Shop Now & Save \$5 Instantly.

- Pages are assigned to different conditions with different headlines



User

WEB IMAGES VIDEOS MAPS SHOPPING LOCAL NEWS MORE



flowers



358,000,000 RESULTS

Flowers at 1-800-FLOWERS®

1800Flowers.com

Fresh Flowers & Gifts at 1-800-FLOWERS. 100% Smile Guarantee. Shop Now

A

FTD® - Flowers

www.FTD.com

Get Same Day Flowers in Hours! Buy Now for 25% Off Best Sellers.

Send Flowers from \$19.99

www.ProFlowers.com

Send Roses, Tulips & Other Flowers. "Best Value" -Wall Street Journal.
proflowers.com is rated ★★★★ on Bizrate (1307 reviews)

50% Off All Flowers

www.BloomsToday.com

All Flowers on the Site are 50% Off. Take Advantage and Buy Today!



flowers



358,000,000 RESULTS

FTD® - Flowers

www.FTD.com

Buy Now for 25% Off Best Sellers.

B

Flowers at 1-800-FLOWERS® | 1800flowers.com

1800Flowers.com

Fresh Flowers & Gifts at 1-800-FLOWERS. 100% Smile Guarantee. Shop Now

Send Flowers from \$19.99

Send Roses, Tulips & Other Flowers.

www.ProFlowers.com

"Best Value" -Wall Street Journal.

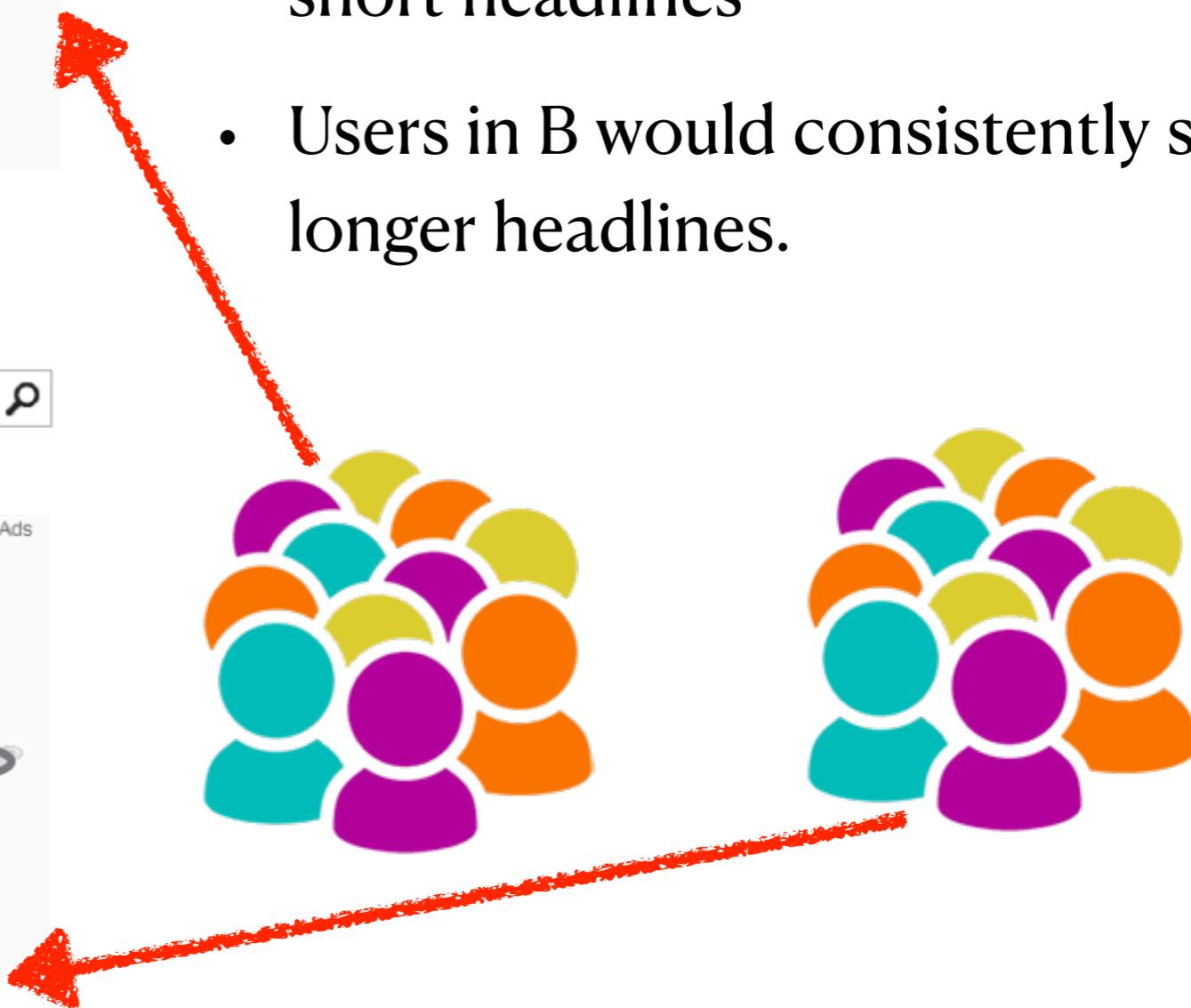
proflowers.com is rated ★★★★ on Bizrate (1307 reviews)

\$19.99 - Cheap Flowers - Delivery Today By A Local Florist!

www.FromYouFlowers.com

Shop Now & Save \$5 Instantly.

- Randomization Unit = User
- Assign users to A and B conditions
- Users in A would consistently see short headlines
- Users in B would consistently see longer headlines.



Page

WEB IMAGES VIDEOS MAPS SHOPPING LOCAL NEWS MORE



flowers



358,000,000 RESULTS

Flowers at 1-800-FLOWERS®

1800Flowers.com

Fresh Flowers & Gifts at 1-800-FLOWERS. 100% Smile Guarantee. Shop Now

FTD® - Flowers

www.FTD.com

Get Same Day Flowers in Hours! Buy Now for 25% Off Best Sellers.

Send Flowers from \$19.99

www.ProFlowers.com

Send Roses, Tulips & Other Flowers. "Best Value" -Wall Street Journal.
proflowers.com is rated ★★★★ on Bizrate (1307 reviews)

50% Off All Flowers

www.BloomsToday.com

All Flowers on the Site are 50% Off. Take Advantage and Buy Today!

A



flowers



358,000,000 RESULTS

FTD® - Flowers

Get Same Day Flowers in Hours!

www.FTD.com

Buy Now for 25% Off Best Sellers.

Flowers at 1-800-FLOWERS® | 1800flowers.com

1800Flowers.com

Fresh Flowers & Gifts at 1-800-FLOWERS. 100% Smile Guarantee. Shop Now

Send Flowers from \$19.99 - Send Roses, Tulips & Other Flowers.

www.ProFlowers.com

"Best Value" -Wall Street Journal.

proflowers.com is rated ★★★★ on Bizrate (1307 reviews)

\$19.99 - Cheap Flowers - Delivery Today By A Local Florist!

www.FromYouFlowers.com

Shop Now & Save \$5 Instantly.

B

- Randomization Unit = Page

- Pages would be assigned to A or B every time when a user opens a page (search)

- Users see different headlines for different pages .



site:howtogeek.com bing

ALL RESULTS 1-10 of 21,400 results - Ads

[How to Uninstall Bing](#) > How To Geek Forums

Hello to you nice GoodGreatGeeks :) like ScottW said I am a regular customer suppose you guys need some people like me ... to keep you busy ... well I am sure www.howtogeek.com/intro/bing/how-to-uninstall-bing

[The New Bing Bar Provides Easy Access to Searches and Microsoft](#)

If you're a fan of Microsoft Bing as your search engine, you might be interested in trying out the new Bing Bar. Today we look at this toolbar which you can add to Internet... www.howtogeek.com/intro/1500/the-new-bing-bar-provides-easy-access-to-searches-and...

[View Word Definitions in IE 8 with the Define with Bing...](#)

Do you need an easy way to view word definitions while browsing with Internet Explorer? The Define with Bing Accelerator will display definitions in the same (ie: new) tab and... www.howtogeek.com/intro/17091/view-word-definitions-in-ie-with-the-define-with-bing...

SEARCH HISTORY

See all Clear all... Turn off

Smartphones NOT phone

Find this exact phrase

See all Clear all... Turn off

Do you need an easy way to view word definitions while browsing with Internet Explorer? The Define with Bing Accelerator will display definitions in the same (ie: new) tab and... www.howtogeek.com/intro/17091/view-word-definitions-in-ie-with-the-define-with-bing...

SEARCH HISTORY

See all Clear all... Turn off

Smartphones NOT phone

Find this exact phrase

See all Clear all... Turn off

Do you need an easy way to view word definitions while browsing with Internet Explorer? The Define with Bing Accelerator will display definitions in the same (ie: new) tab and... www.howtogeek.com/intro/17091/view-word-definitions-in-ie-with-the-define-with-bing...

site:howtogeek.com bing

ALL RESULTS 1-10 of 21,400 results - Ads

[How to Uninstall Bing](#) > How To Geek Forums

Hello to you nice GoodGreatGeeks :) like ScottW said I am a regular customer suppose you guys need some people like me ... to keep you busy ... well I am sure www.howtogeek.com/intro/bing/how-to-uninstall-bing

[The New Bing Bar Provides Easy Access to Searches and Microsoft](#)

If you're a fan of Microsoft Bing as your search engine, you might be interested in trying out the new Bing Bar. Today we look at this toolbar which you can add to Internet... www.howtogeek.com/intro/1500/the-new-bing-bar-provides-easy-access-to-searches-and...

[View Word Definitions in IE 8 with the Define with Bing...](#)

Do you need an easy way to view word definitions while browsing with Internet Explorer? The Define with Bing Accelerator will display definitions in the same (ie: new) tab and... www.howtogeek.com/intro/17091/view-word-definitions-in-ie-with-the-define-with-bing...

SEARCH HISTORY

See all Clear all... Turn off

Smartphones NOT phone

Find this exact phrase

See all Clear all... Turn off

Do you need an easy way to view word definitions while browsing with Internet Explorer? The Define with Bing Accelerator will display definitions in the same (ie: new) tab and... www.howtogeek.com/intro/17091/view-word-definitions-in-ie-with-the-define-with-bing...

SEARCH HISTORY

See all Clear all... Turn off

Smartphones NOT phone

Find this exact phrase

See all Clear all... Turn off

Do you need an easy way to view word definitions while browsing with Internet Explorer? The Define with Bing Accelerator will display definitions in the same (ie: new) tab and... www.howtogeek.com/intro/17091/view-word-definitions-in-ie-with-the-define-with-bing...

User



- **Consistent**

- A unit should consistently experience the same condition (variant)

- Otherwise, a unit experience an inconsistent variant.

- We cannot associate the outcomes with the variant

- **Independent**

- A unit's experience should not be affected by the other units' conditions (variants)

- Otherwise, a unit experience a mixed variant.

- We cannot associate the outcomes with the variant.



Page



Can the two principles be satisfied, if the randomization unit = page?

- **Consistent**

- A unit should consistently experience the same condition (variant)

- Otherwise, a unit experience an inconsistent variant.

- We cannot associate the outcomes with the variant

- **Independent**

- A unit's experience should not be affected by the other units' conditions (variants)

- Otherwise, a unit experience a mixed variant.

- We cannot associate the outcomes with the variant.



Page



Can the two principles be satisfied, if the randomization unit = page?

- **Consistent**

- A unit should consistently experience the same condition (variant)

- Otherwise, a unit experience an inconsistent variant.

- We cannot associate the outcomes with the variant

- **Independent**

- A unit's experience should not be affected by the other units' conditions (variants)

- Otherwise, a unit experience a mixed variant.

- We cannot associate the outcomes with the variant.



Digital Experimentation Methods

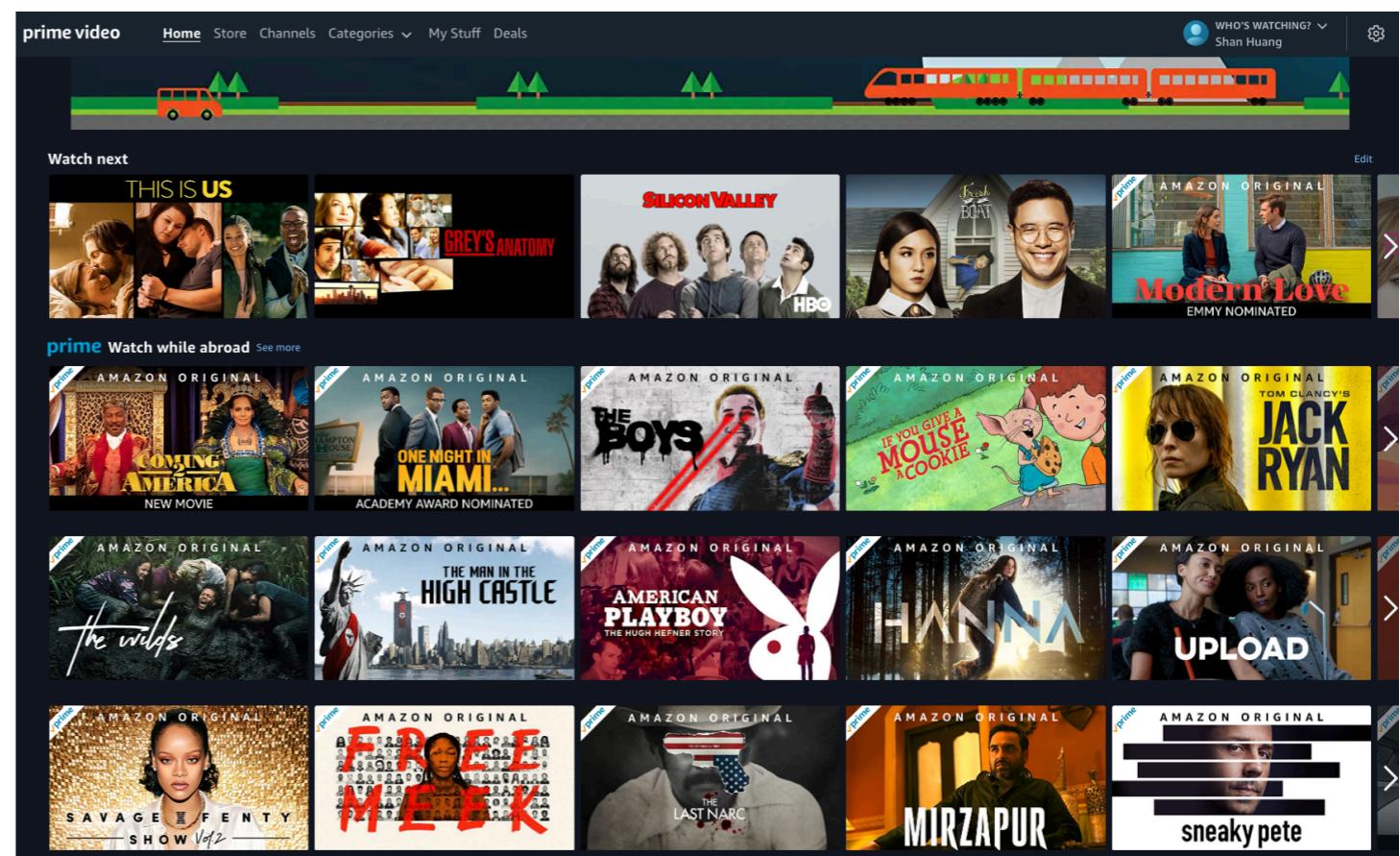
AB Testing

Session 1b: An Overview

Shan Huang, HKU

Class Exercise

- As an engineer, you want to find the best parameters for your algorithm to optimize the user clicks on the movies recommended on Amazon Prime.
- Assume that there are 8 combinations of the parameters (k, b) .
- Could you design an experiment to find the (k, b) which optimizes user clicks?



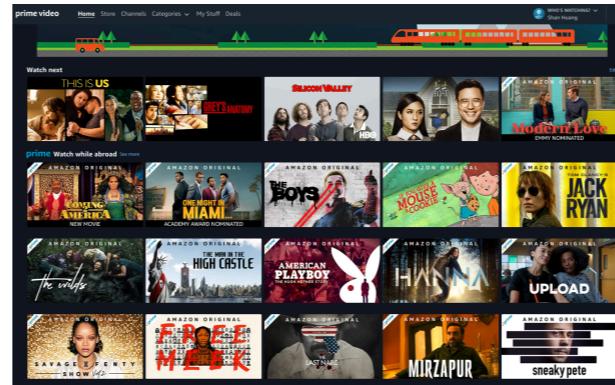
Class Exercise

- Metrics (OEC)
- Parameters
- Conditions (Variants)
- Randomization Unit

Q1: Which one should you consider/decide first?

Randomization Unit

- Users or Pages?



- Users

- Randomization occurs, when a new user opens the page.
- Different users see the movies recommended by the algorithm with a specific (k,b)

- Pages Larger Sample Size (N), Why?

- Randomization occurs every time when a new page is initiated.
- Every time when a user opens the page, he or she will see the movies recommended by the algorithm with a specific (k,b)

Terminology

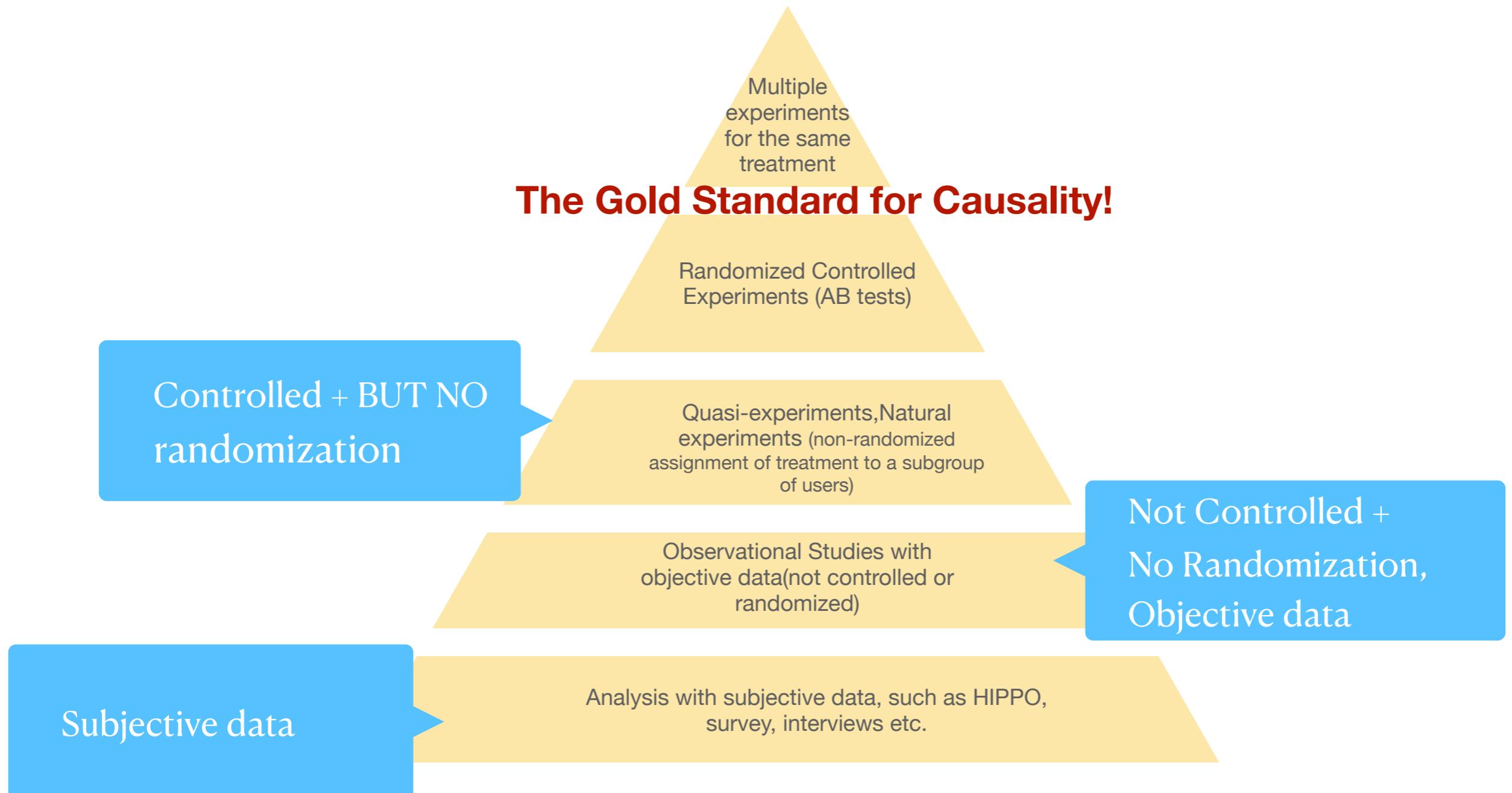
- Randomization Unit = pages
- Metric: clickthrough rate/page, #clicks/page ...
- Parameter = (k,b)
- Conditions = eight conditions with eight different values of (k,b)

The unit of metrics is generally consistent with that of the randomization unit
- *The unit of metrics should be consistent or smaller than randomization units*

Review: Terminology for Experiment

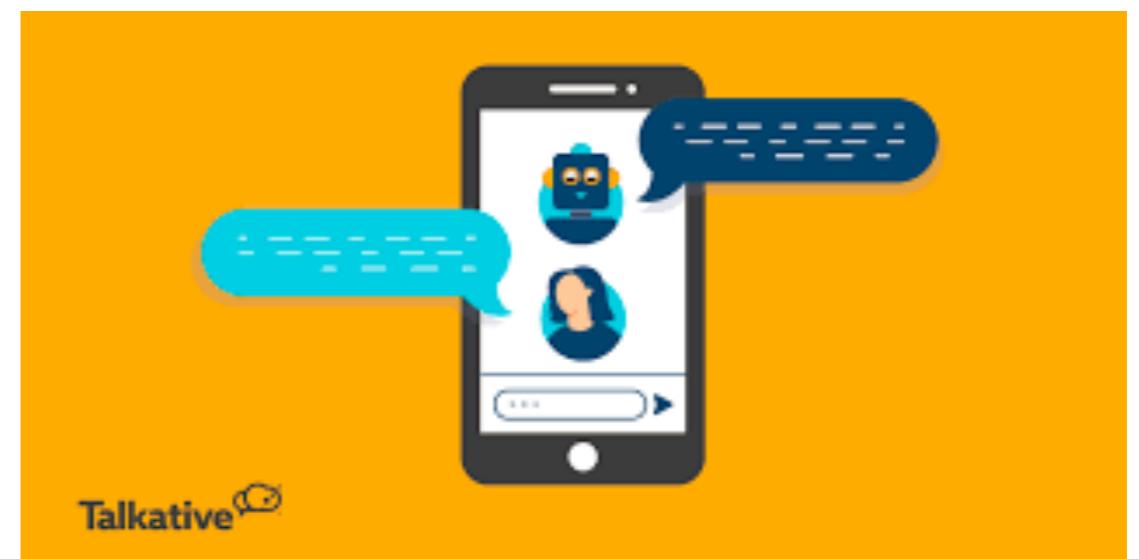
- Metrics (OEC)
- Parameter
- Conditions (control and treatment conditions/groups)
- Randomization Unit

Review: Hierarchy of Evidence



Class Exercise

- Suppose Taobao is launching chatbots to serve its customers.
- You want to test the effectiveness of this function.
- How will you design the AB test?
 - Metrics (OEC)
 - Parameters
 - Conditions (Variants)
 - Randomization Unit



Class Exercise

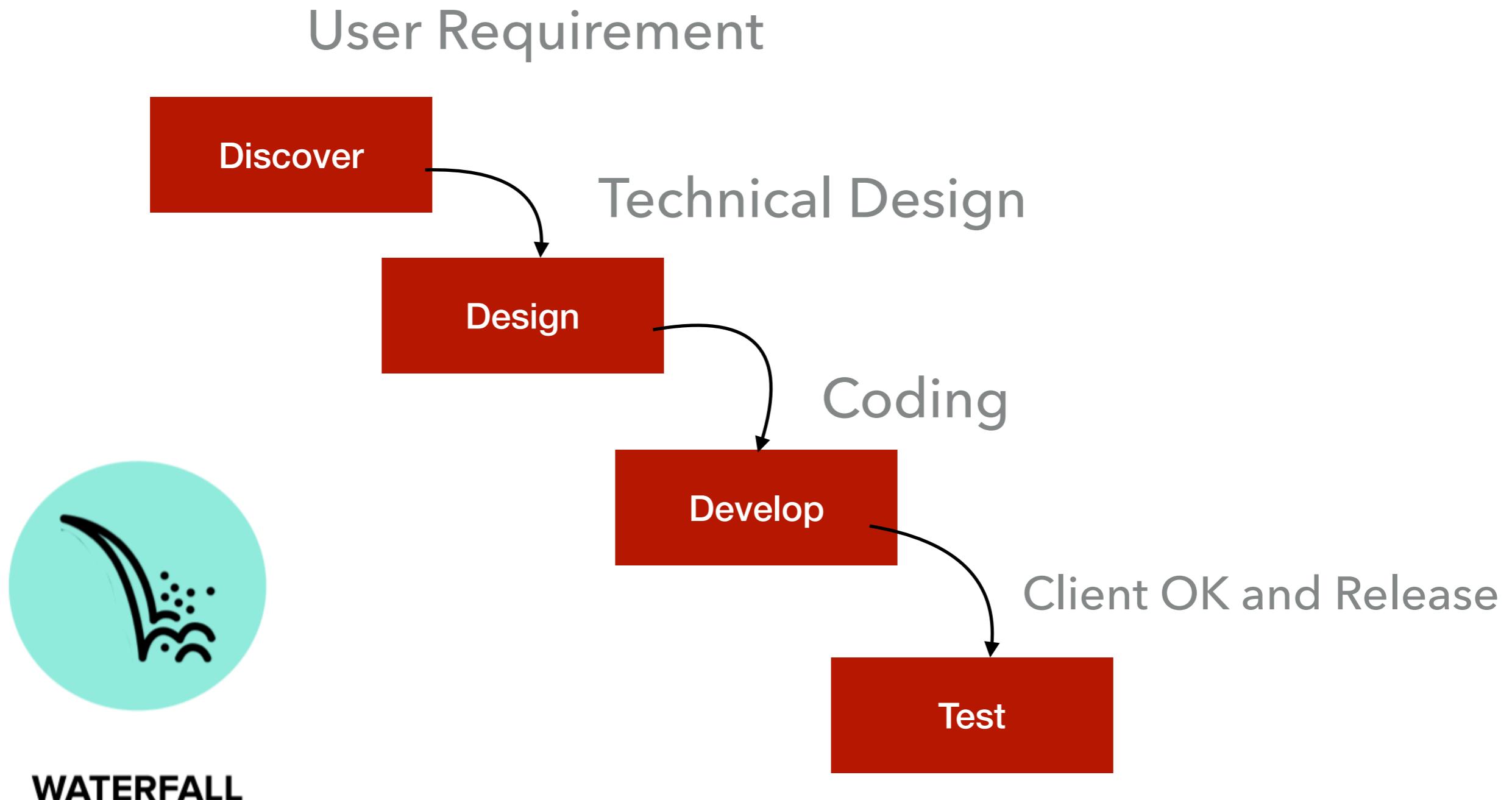
- Suppose Taobao is launching chatbots to serve its customers.
- You want to test the effectiveness of this function.
- How will you design the AB test?
 - Parameters - new function (yes, no)
 - Conditions (Variants) - with/out chatbots
 - Randomization Unit - chat/user ?
 - Metrics (OEC) - sales, revenue, #chat / user

Talkative

Why are experiments so popular today?

TRADITIONAL PRODUCT DEVELOPMENT

TRADITIONALLY, WE BUILD PRODUCT LIKE THIS!



TRADITIONAL: WATERFALL MODEL

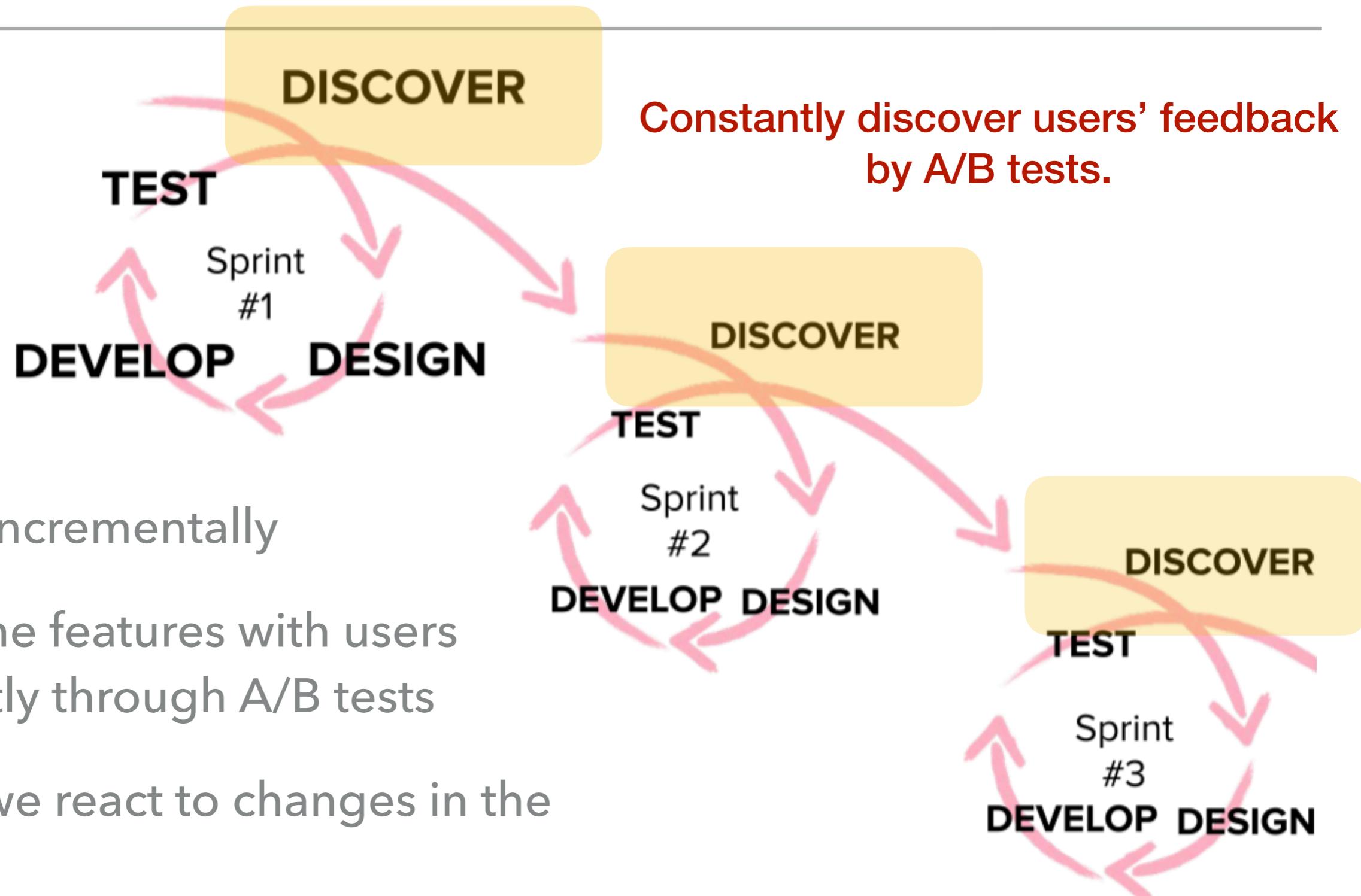


- ▶ Completely planned requirements
- ▶ Get client approval of the contract
- ▶ Complete design document before implementation
- ▶ Coding & Building after design documents get approved

WATERFALL

Get client's feedback at the *very end!*

AGILE



NOW WE BUILD PRODUCT LIKE THIS!

Start from Minimum Viable Product (MVP)

- **MVP** - Use simple prototypes wherever possible to learn from users
 - Keep MVPs small
 - Agile product development
- Iterate based on learnings
- Fast, cheap experiments
- Set metrics, learn from users by A/B tests, and iterate
 - Algorithms (e.g., ranking, recommender systems) - backend
 - UI (e.g., color, fontsize) - frontend
 - New functions (e.g., coupons)
 - Combination of algorithms and UI

“Organizations will recognize maximal benefits from experimentation when it is used in conjunction with an “innovation system”.

–Stefan Thomke (HBS Professor)

AB Testing is An Emerging Field!

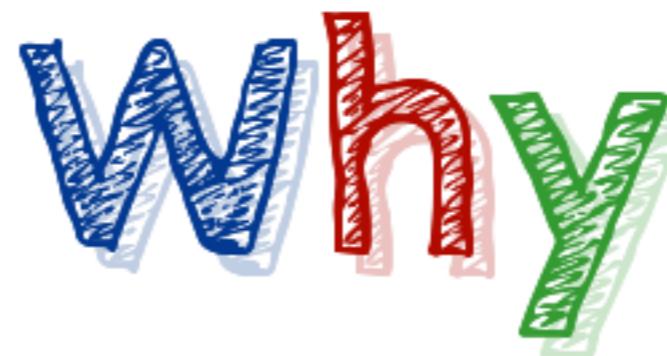
- It started @ Bing (Microsoft) **about 10+ years ago.**
- All the big IT firms in China have intensively invested in the experimentation system and building the A/B testing culture **during the past 3 - 5 years.**
- N*5000+ / day @ Tencent, Alibaba, Bytedance, and more in the future!



What should organizations do to embrace the A/B tests?

Tenets for Organizations

- Recognize they are poor at assessing the value of ideas.
 - Run experiments
- However, experiments often conclude insignificant results.
 - Hypothesis: the new feature would improve OEC.
 - BUT the new features DO NOT show positive effects most of the time:
 - Microsoft 70%
 - Google 80%
 - Netflix 90%



Tenets for Organizations

- Willing to invest in the infrastructure (experimental system)
 - to decrease the variable costs
 - to ensure that the results are trustworthy.
- Automate the process of experiments
 - Randomize units into conditions
 - Collect data
 - Make treatments (changes) easily
 - Run large-scale experiments (statistics on large N)
 - Run concurrent (overlapping) experiments

**Getting Numbers is easy, getting numbers you can trust is
HARD!**

Mission of A/B Testing Team

What roles do engineers, PMs, and data scientists play in AB testing?

- Accelerate innovation through trustworthy analysis and experimentation.
- Engineers
 - build experimentation systems.
- Product managers
 - initiate new product (features) ideas (hypotheses).
- Data Scientists
 - run experiments and collect and analyze data.
- Engineers and Data Scientists can also initiate experiments
 - Back-end experiments to optimize algorithms.
 - Test new experimentation methods.

Decision Makers (e.g.,HIPPO)

- Decision makers should be willing to make data-driven decisions and formalize metrics to evaluate their products.
- Decisions are NOT purely driven by experiments
 - Data-informed or Data-aware
- Decisions should be made with many resources of relevant data
 - Controlled experiments, surveys, costs estimates, experts' opinions

**What are Interesting
experiments?**

Interesting Experiments

- How do you define interesting experiments?
 - The absolute difference between the expected outcome and the actual results is large.
 - You learned from the experiments!
- If you thought something was going to happen and it didn't, then you've learned something important.
- If you thought something minor was going to happen, and the results are a major surprise and lead to a breakthrough, you've learned something highly valuable.
- If you thought something was going to happen, and it happened, then you confirmed your existing knowledge but have *not* learned much

Interesting Examples

Guess Whether Treatments or Control Wins or All the Same

SERP Truncation

- SERP is a Search Engine Result Page
- OEC is Clickthrough Rate (CTR) on 1st SERP per query (ignore issues with click/back, page 2, etc.)
- Which condition gives the higher OEC?
 - A. Control: Show 10 algorithmic results
 - B. Treatment: Show 8 algorithmic results by removing the last two results.
 - C. All the same.

The screenshot shows a Google search results page for the query "A/B testing". The top result is from Optimizely, followed by several other links from various websites like VWO, Wikipedia, Harvard Business Review, Hubspot, Oracle, Kameleoon, Crazy Egg, and AdEspresso. A sidebar on the right lists "People also ask" questions: "Why do we do AB testing?" and "What is AB sample?". The bottom of the page shows a snippet of the first result from Optimizely, which discusses the rules of thumb for web site experimenters.

Google AB testing https://www.optimizely.com › optimization-glossary › ab... A/B testing - Optimizely

People also ask : Why do we do AB testing? What is AB sample?

https://vwo.com › ab-testing What is A/B Testing? A Practical Guide With Examples | VWO A/B testing, also known as split testing, refers to a randomized experimentation process wherein two or more versions of a variable (web page, page element, ... Frequentist Approach: Bayesian Approach Testing · A/B testing tools · Mobile App A/B Testing · Conversion Rate Optimization

https://en.wikipedia.org › wiki › B_testing A/B testing - Wikipedia A/B testing (also known as bucket testing or split-run testing) is a user experience research methodology. ... A/B tests consist of a randomized experiment with ... Overview · Challenges · History · Examples You've visited this page 4 times. Last visit: 3/19/21

https://hbr.org › 2017/06 › a-refresher-on-ab-testing A Refresher on A/B Testing - Harvard Business Review 28 Jun 2017 — A/B testing, at its most basic, is a way to compare two versions of something to figure out which performs better. While it's most often ...

https://blog.hubspot.com › marketing › how-to-do-a-b... How to Do A/B Testing: 15 Steps for the Perfect Split Test 24 Aug 2021 — A/B testing, also known as split testing, is a marketing experiment wherein you split your audience to test a number of variations of a campaign ...

https://www.oracle.com › ... › Marketing What is A/B testing? | Oracle Hong Kong SAR, PRC A/B testing and multivariate testing can show you what is and isn't working in your email marketing campaigns, your website, and even your mobile apps.

https://www.kameleoon.com › ab-testing What is A/B testing? | Kameleoon A/B testing is an online experiment conducted on a website, mobile application or ad, to test potential improvements in comparison to a control (or original) ...

https://www.crazyegg.com › Blog › Testing How to do A/B Testing and Improve Your Conversions Quickly 29 Jun 2021 — A/B testing (also known as split testing) is the process of comparing two versions of a web page, email, or other marketing asset and ...

https://adespresso.com › facebook-ads-optimization › a... A/B Testing | A Beginners Guide to Split Testing Facebook Ads in 2017 A/B testing or split testing Facebook ads is the most effective way to create successful campaigns. Learn how to effectively test your ads in this guide!

https://www.dynamicyield.com › lesson › introduction-t... A/B testing guide by CRO experts, with examples - Dynamic ... A/B testing is a method of comparing two versions of a webpage or app against each other to determine which one performs better against a specific objective. It ...

SERP Truncation

- SERP is a Search Engine Result Page
- OEC is Clickthrough Rate (CTR) on 1st SERP per query (ignore issues with click/back, page 2, etc.)
- Which condition gives the higher OEC?
 - A. Control: Show 10 algorithmic results
 - B. Treatment: Show 8 algorithmic results by removing the last two results.
 - C. All the same.

The screenshot shows a Google search results page for the query "A/B testing". The top result is from Optimizely, followed by links from VWO, Wikipedia, Harvard Business Review, Hubspot, Oracle, Kameleoon, Crazy Egg, and AdEspresso. A sidebar on the right lists "People also ask" questions: "Why do we do AB testing?" and "What is AB sample?".

Google search results for "A/B testing":

- <https://www.optimizely.com/optimization-glossary/ab-testing/> A/B testing - Optimizely
- <https://vwo.com/ab-testing/> What is A/B Testing? A Practical Guide With Examples | VWO
- https://en.wikipedia.org/wiki/B_testing A/B testing - Wikipedia
- <https://hbr.org/2017/06/a-refresher-on-ab-testing> A Refresher on A/B Testing - Harvard Business Review
- <https://blog.hubspot.com/marketing/how-to-do-a-b-test> How to Do A/B Testing: 15 Steps for the Perfect Split Test
- <https://www.oracle.com/.../Marketing> What is A/B testing? | Oracle Hong Kong SAR, PRC
- <https://www.kameleoon.com/ab-testing> What is A/B testing? | Kameleoon
- <https://www.crazyegg.com/Blog/Testing> How to do A/B Testing and Improve Your Conversions Quickly
- <https://adespresso.com/facebook-ads-optimization/a/b-testing> A/B Testing | A Beginners Guide to Split Testing Facebook Ads in 2017
- <https://www.dynamicyield.com/lesson/introduction-to-ab-testing> A/B testing guide by CRO experts, with examples - Dynamic ...

SERP Truncation

- Rule of Thumb:
 - Reducing the click-through rate is hard.
 - Shifting clicks is easy.
- While there are obviously exceptions, most of the time users click at the same rate.
 - Users simply shifted the clicks from the last two algorithmic results to other elements of the page.
 - Show 7,6 - No difference!

Experiments cannot do...

The Gold Standard for Causality!

Multiple experiments for the same treatment

Randomized Controlled Experiments (AB tests)

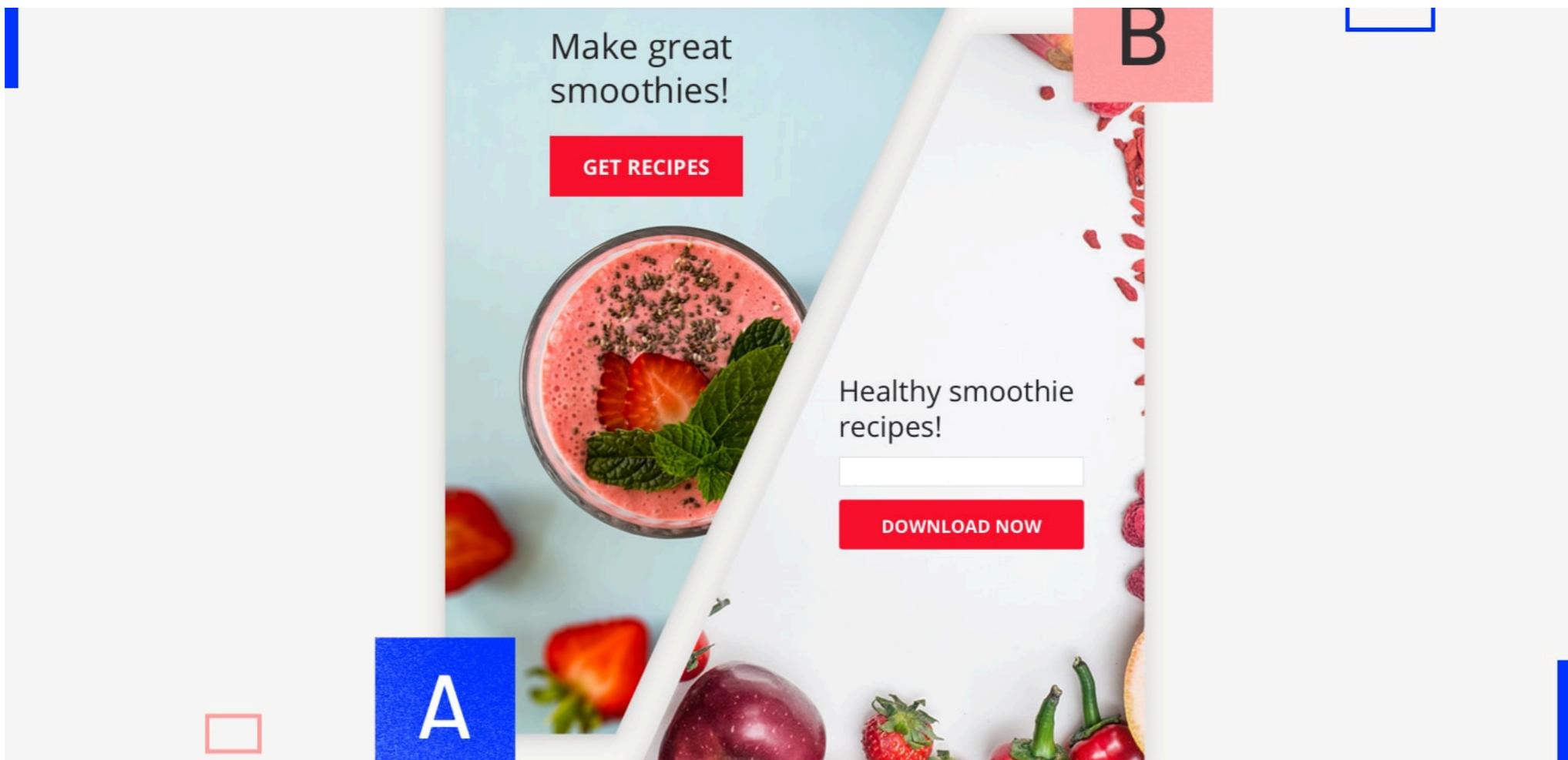
Quasi-experiments, Natural experiments (non-randomized assignment of treatment to a subgroup of users)

Observational Studies with objective data(not controlled or randomized)

Analysis with subjective data, such as HIPPO, survey, interviews etc.

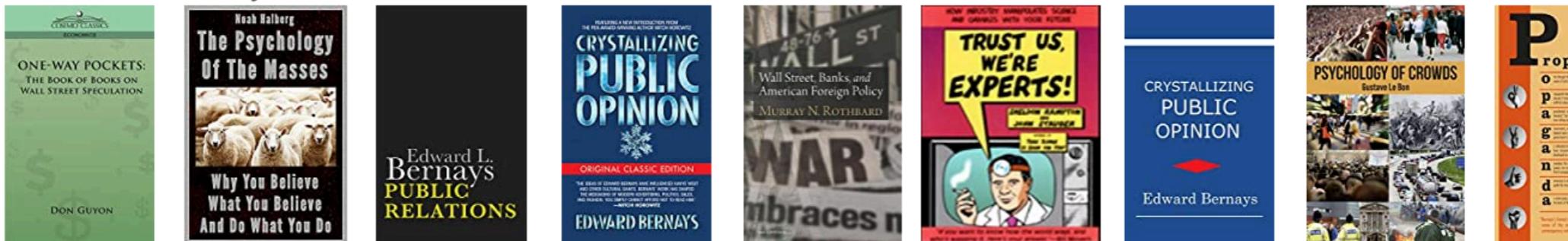
Not every decision can be made with experiments!

- A new product idea with no users
- COVID-19 Impact on revenue
- Dramatic changes to products
- Hard to randomize - e.g., in social networks

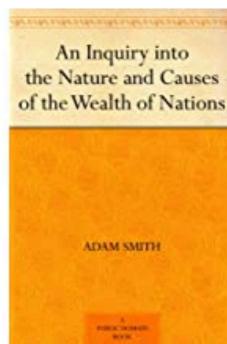


An End-to-End Example

Online Commerce Website



Review your purchase



[See more products to review](#)

Deals on overstock shoes in Outlet



[See more](#)

Small space hacks



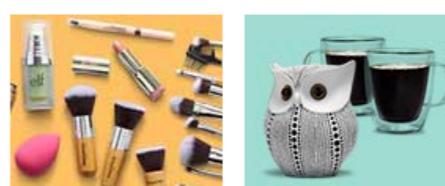
[Shop solutions for every room](#)

Customers' Most-Loved



Women's

Men's



Beauty

Home

[See more](#)

Inspired by your purchases



Online Commerce Platform

- Marketing teams want to increase sales by sending coupons to users - **Add a coupon system**
- Test new features for the platform
 - including changes to both UI & back-end changes, such as recommendation algorithms

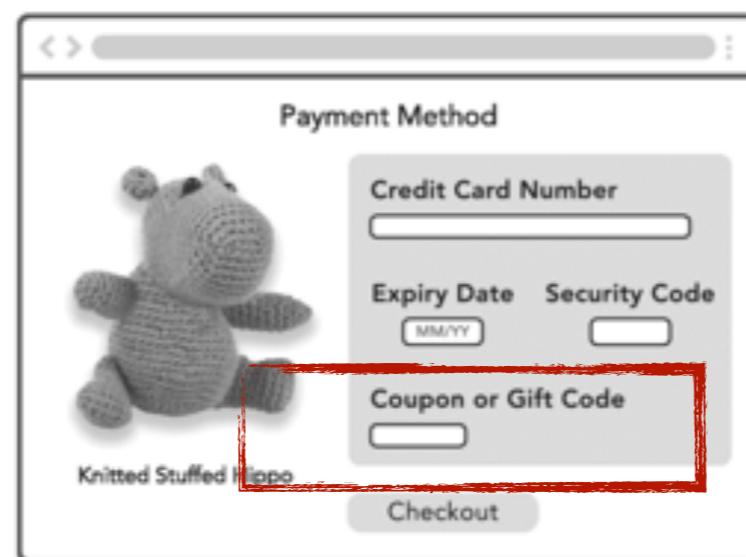


Experiments

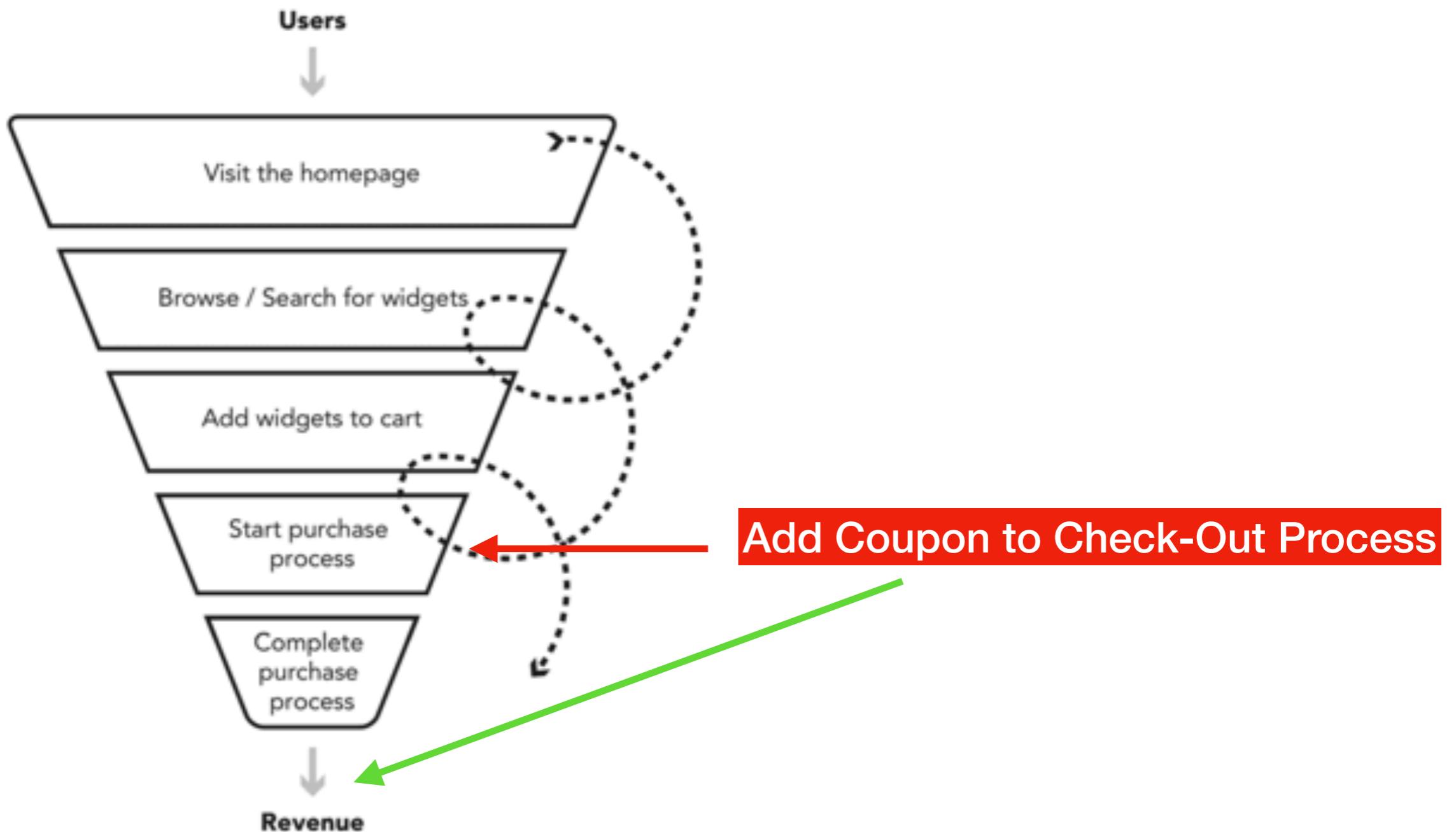
- Hypothesis: Adding a coupon system will generally increase sales.
- Add Coupons
 - UI Changes - Add coupons in promotion emails and the check-out process & in different ways.
 - Recommender Systems - Different algorithms to recommend coupons to users
- To test one hypothesis may need *multiple* experiments
 - Test different UIs
 - Test different recommendation algorithms...

Setting up the Experiment

- Hypothesis - This will distract users from checking out and thus degrade revenue.
 - Theory - Users seeing this field will slow them down, and cause them to search for codes, or even abandon.



Online Shopping Funnel



- However, implement coupon systems is very costly.
- How could we test it fast and at a low cost?

Fake Door Approach

- Keep the change simple and test it fast in an experiment - MVP.
- Do not implement a true coupon system.
- BUT paint a *fake door* on the wall with only UI change without any coupon issued.
- Theory- This will distract users from checking out and thus degrade revenue.



User enter a code

Shows: Invalid
Coupon Code!

Setting up the Experiment

- Metric (OEC)
 - Revenue per user
 - Why not total revenue
- Conditions (Groups, Variants)
 - Control & Treatments
 - UI changes

Control

Payment Method
Credit Card Number
Expiry Date Security Code
MM/YY
Knitted Stuffed Hippo
Checkout

(1) Control: the old checkout page.

Treatment 1

Payment Method
Credit Card Number
Expiry Date Security Code
MM/YY
Coupon or Gift Code
Knitted Stuffed Hippo
Checkout

(2) Treatment one: coupon or gift code field below credit card information

Treatment 2

Payment Method
Credit Card Number
Expiry Date Security Code
MM/YY
Coupon Code
Knitted Stuffed Hippo
Checkout

Enter Coupon Code
Apply

(3) Treatment two: coupon or gift code as a popup
Even more distracting!

Compare Means

- Basic Statistics behind A/B tests
 1. Calculate means and standard errors for different groups (Variants)
 2. Compare the means (or other summary statistics, e.g., median)
 3. Decide whether the differences are **statistically and practically** significant
- Large Enough **Statistical Power** (80%+)
 - Detect the difference when there is a difference
 - Detect the difference = the difference is statistically significant

Hypothesis Testing

- Null Hypothesis: (population) means of the OEC are the same between Treatment and Control
 - Compare two groups at a time (usually)
 - Through observing sample means, we can decide how unlikely Null is True.
 - **Reject the Null if it is unlikely.**

Control

The screenshot shows a 'Payment Method' form. It features a small image of a knitted stuffed hippo labeled 'Knitted Stuffed Hippo'. Below the image are fields for 'Credit Card Number', 'Expiry Date' (with a 'MM/YY' placeholder), and 'Security Code'. At the bottom is a 'Checkout' button.

(1) Control: the old checkout page.

Treatment 1

The screenshot shows a 'Payment Method' form. It features a small image of a knitted stuffed hippo labeled 'Knitted Stuffed Hippo'. Below the image are fields for 'Credit Card Number', 'Expiry Date' (with a 'MM/YY' placeholder), 'Security Code', and 'Coupon or Gift Code'. At the bottom is a 'Checkout' button.

(2) Treatment one: coupon or gift code field below credit card information

Setting up the Experiment

1. What is the randomization unit?
2. What population of randomization units do we want to target?
3. How large (sample size) does our experiment need to be?
 - How long will the experiment take?
4. How large the effects do you want to detect/care about?

Setting up the Experiment

- Randomization Unit - Users or Pages?

Control

A screenshot of a web browser window titled "Payment Method". Inside, there's a small image of a knitted stuffed hippo. Below it is a form with three input fields: "Credit Card Number", "Expiry Date" (with a "MM/YY" placeholder), and "Security Code" (with a placeholder). At the bottom is a "Checkout" button.

(1) Control: the old checkout page.

Treatment 1

A screenshot of a web browser window titled "Payment Method". It features the same knitted hippo image and basic form as the control version. However, it includes an additional section labeled "Coupon or Gift Code" with an input field. At the bottom is a "Checkout" button.

(2) Treatment one: coupon or gift code field below credit card information

Setting up the Experiment

- The population of randomization units to target
 - A. All users who visited the site
 - B. Only users who complete the purchase process
 - C. Only users who start the purchase processes - Trigger Experiment!

Control

The Control payment method page features a placeholder for a coupon or gift code below the credit card information fields. The page includes a 'Payment Method' header, a 'Knitted Stuffed Hippo' image, and a 'Checkout' button.

(1) Control: the old checkout page.

Treatment 1

The Treatment 1 payment method page has a separate field for a coupon or gift code located below the credit card information fields. The page includes a 'Payment Method' header, a 'Knitted Stuffed Hippo' image, and a 'Checkout' button.

(2) Treatment one: coupon or gift code field below credit card information

Setting up the Experiment

- The population of randomization units to target
 - A. All users who visited the site
 - B. Only users who complete the purchase process
 - C. Only users who start the purchase processes - Trigger Experiment!

Control

The Control payment method page features a placeholder for a coupon or gift code below the credit card information fields. The page includes a 'Payment Method' header, a 'Knitted Stuffed Hippo' image, and a 'Checkout' button.

(1) Control: the old checkout page.

Treatment 1

The Treatment 1 payment method page has a visible 'Coupon or Gift Code' field positioned between the credit card number and expiry date fields. The page includes a 'Payment Method' header, a 'Knitted Stuffed Hippo' image, and a 'Checkout' button.

(2) Treatment one: coupon or gift code field below credit card information

Sample

- A large sample is (almost) always beneficial
- Allocate more traffic (% users) to the experiment
 - Why not 100% traffic to an experiment
- Run the experiment longer
 - More **unique** users
 - New users, but not repeat users
- Examine effects on user segments
 - Different ages, cities, days, etc.
- **Is the new feature 100% positive for user experience?**
 - Test on a smaller portion of users and run shorter to prevent big damage to user experience.

RULE OF THUMB for Sample Size: the larger, the better!

Setting up the Experiment

1. What is the randomization unit? **user**
2. What population of randomization units do we want to target?
 - Target the users who start checkout page
3. How large (size) does our experiment need to be?
4. How large the effects do you want to detect/care about?
Power Analysis + Practical/Contextual knowledge!

Sanity Checks

- More ways than you expect for bugs that invalidate experiments
 - Randomization is correct
 - SRM tests, AA tests, Guardrail metrics...
 - Data collection is correct
 - Data analysis is correct
 - Make “right/wise” decisions

Interpreting the Results

Revenue-per-user	Treatments	Control	Difference	p-value
Treatment 1 vs. Control	3.10	3.21	-0.11	0.0003
Treatment 2 vs. Control	2.96	3.21	-0.25	1.50E-23

- Treatment 1 < Control
- Treatment 2 < Control

- Are the differences large enough *statistically and practically?*

From Results to Decisions

- A/B tests are for decision-making on products and business strategies.
- Translate the results to a launch/no-launch decision
- Launch the new features if the improvements are both statistically and practically significant.
- Do you need to make tradeoffs between different metrics?
 - User engagement increases, but revenue decreases
 - Is that too costly to launch the new feature?