

ECON7960 User Experience and A/B Test

Hong Kong Baptist University

Topic 1-6: Revision on Simple A/B Test and Understanding UX Experience

A QUIZ on Simple A/B Test

What did we learn from 4 workshops

- Please use your phone to download an apps “SOCRATIVE” student version, and open it, you should see

Enter the Room Name “**HUNG5085**”

Sign your student number

The quiz will start at 6:30pm and lasted for 40 minutes.

Keep the “SOCRATIVE” apps open after you finish the quiz

Student Login

Room Name

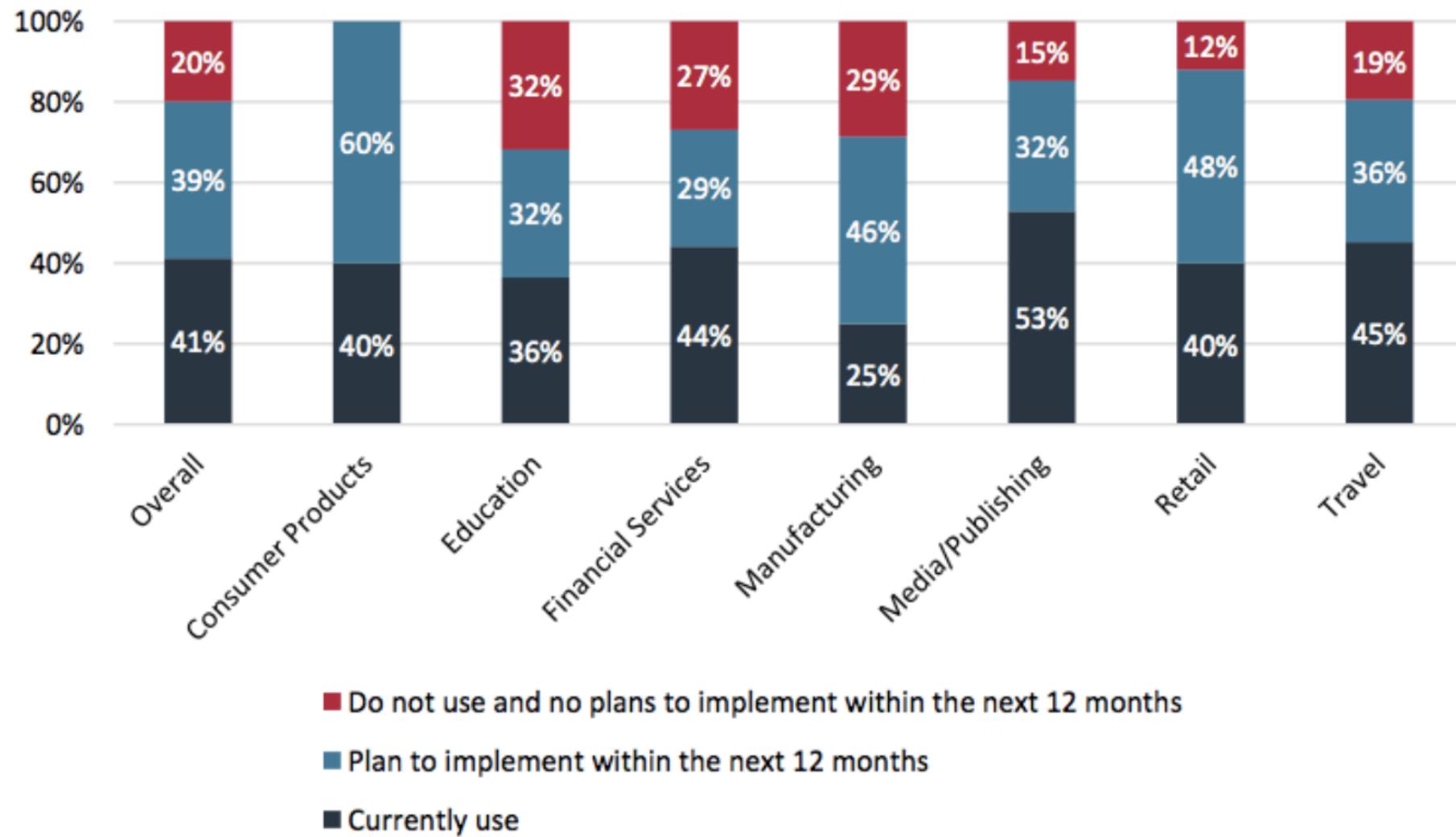
HUNG5085

JOIN

 English ▾

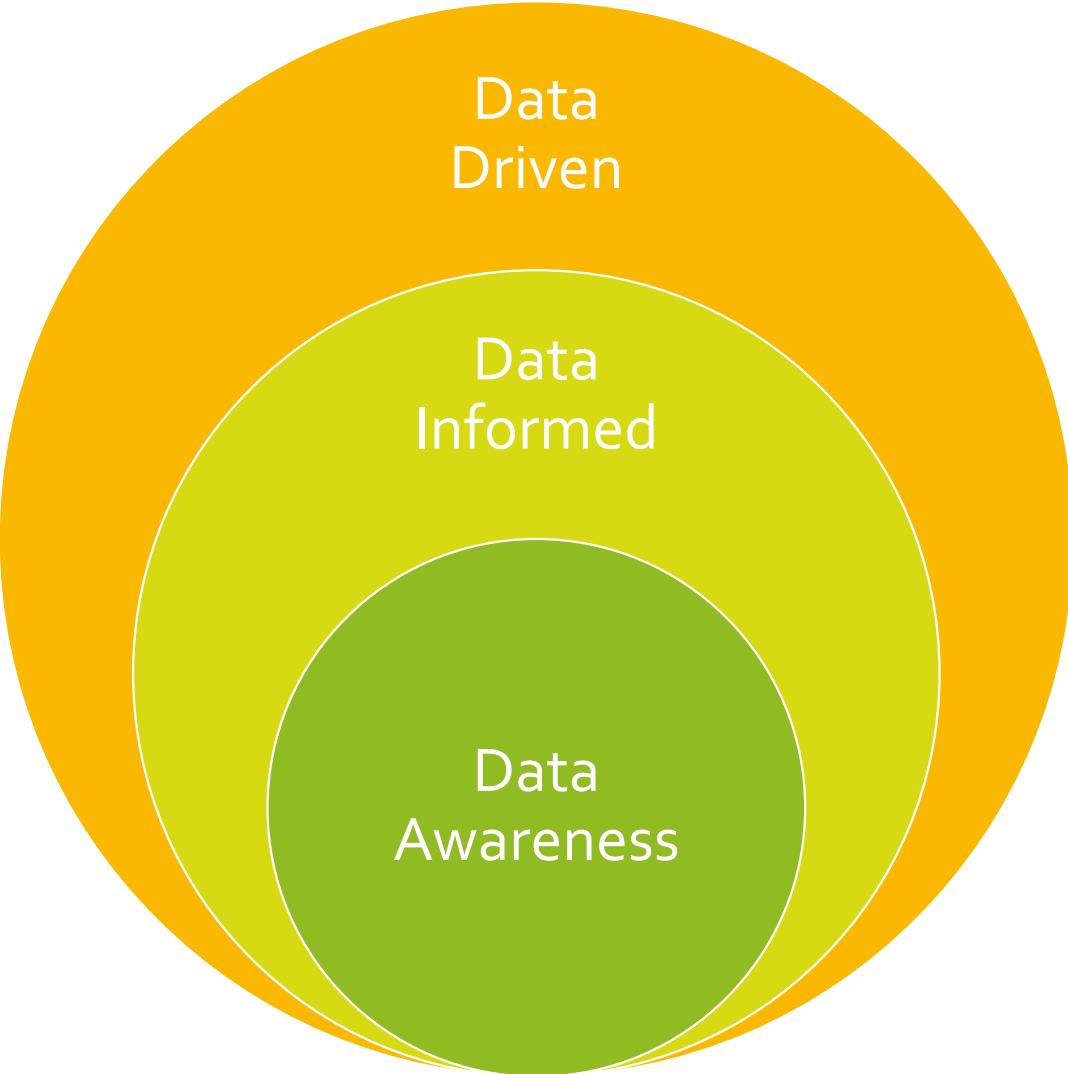
A 2017 study estimated that nearly half of Fortune 500 companies were already licensing a Data Marketing Platform (DMP) to manage their customer data, and another half plan on implementing one. Many enterprises developed data-driven marketing strategies

Figure 6 – Adoption of DMPs by Industry

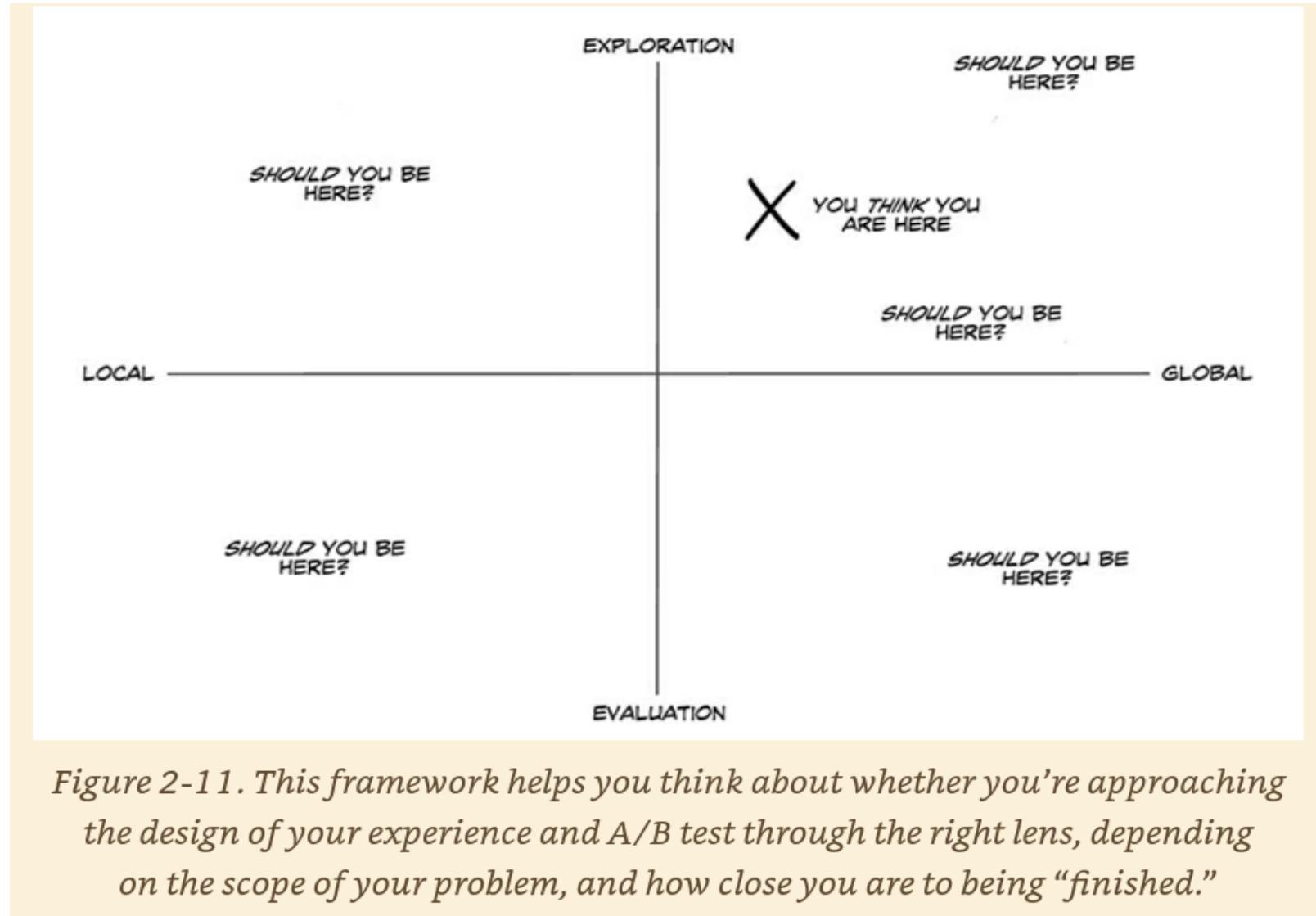




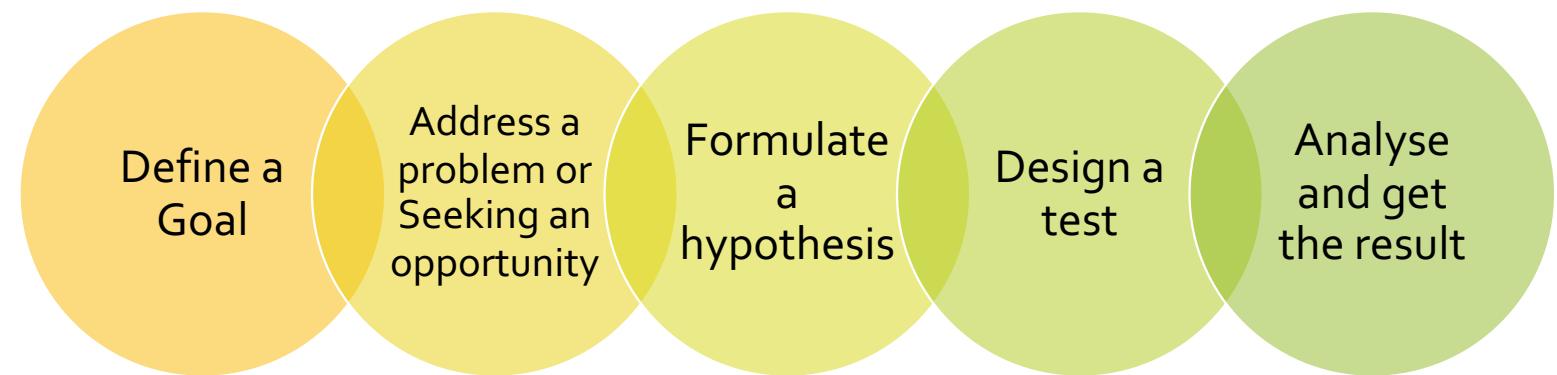
Three Ways to Think about Behaviour towards Data



A/B Testing: Online Experiment - Learning Dynamics and Outcome

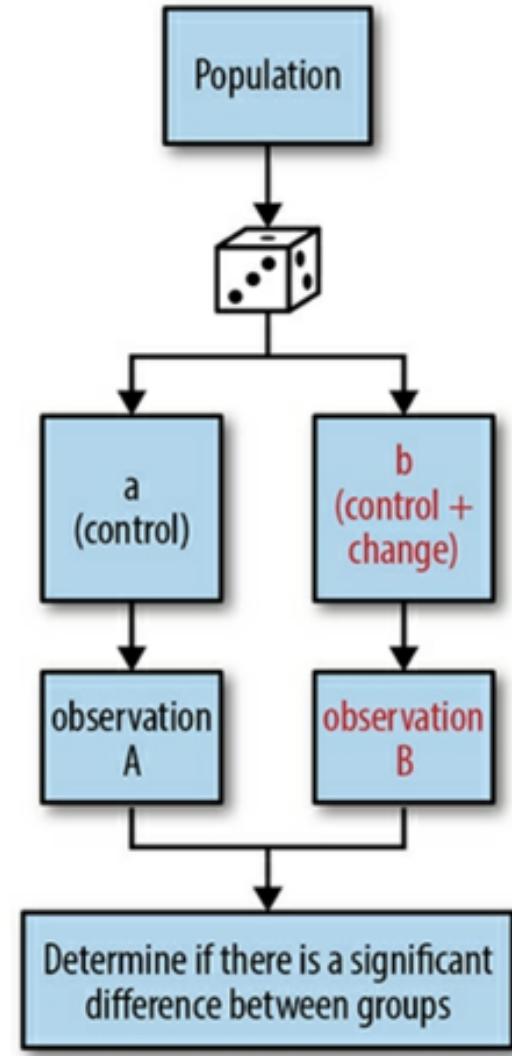


5 Ways to Think about Data



Two Characters: Exploration and Exploitation

- Let think about the workshop, what is the treatment, e.g. the new color of the new logo advertised in social media.
- Once the CEO is convinced of the merits of A/B testing, one will start to contemplate much larger scale experiments: instead of running an A/B test, comparing old black logo with 6 other colors, including some quirky colors like purple and chartreuse. From A/B test to A/B/C/D/E/F/G test (MVT).



Behavior Economics

The Deliberative and Intuitive Mind – A broad understanding of human

Personas

A name and picture

Demographics

Job title and major responsibilities

Goals and task in relations to your site

Environment

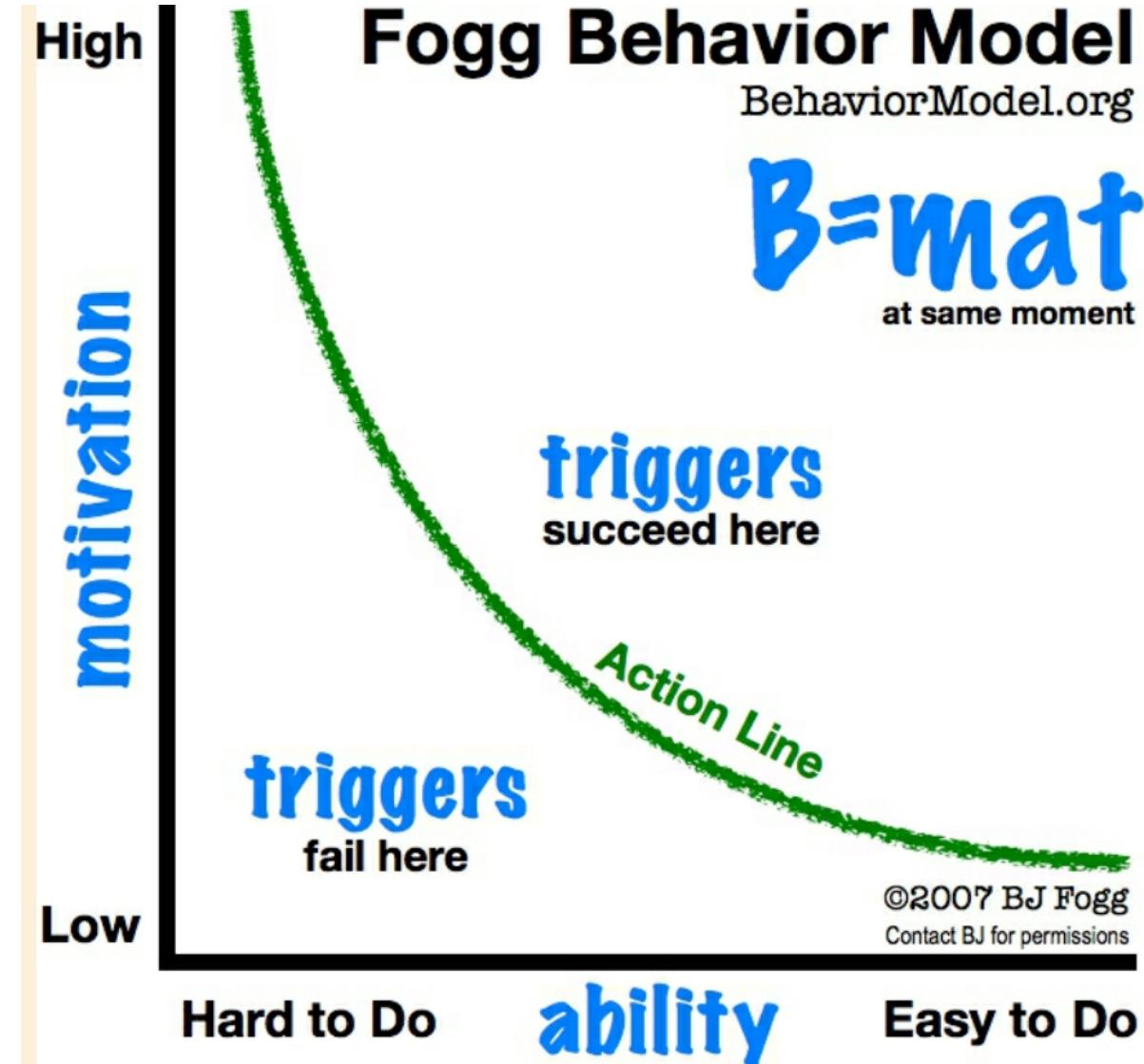
Behavior Pattern



6 Useful of Rules of Thumb

- Easier really is better
- Familiar really is better
- Beauty really is better
- Rewarding experiences really do make one want to come back
- People really do not want to fail
- People do urgent things first

The person responds to a **cue** that starts her thinking about the action. Her **intuitive mind** automatically reacts at an intuitive level to the idea. Her **conscious mind** evaluates the idea, especially in terms of costs and benefits. She checks if she has the **ability** to act—if she knows what to do, has what she needs, and believes she can succeed. She determines if the **timing** is right for action—especially whether or not the action is urgent.



Behavioral Decisions

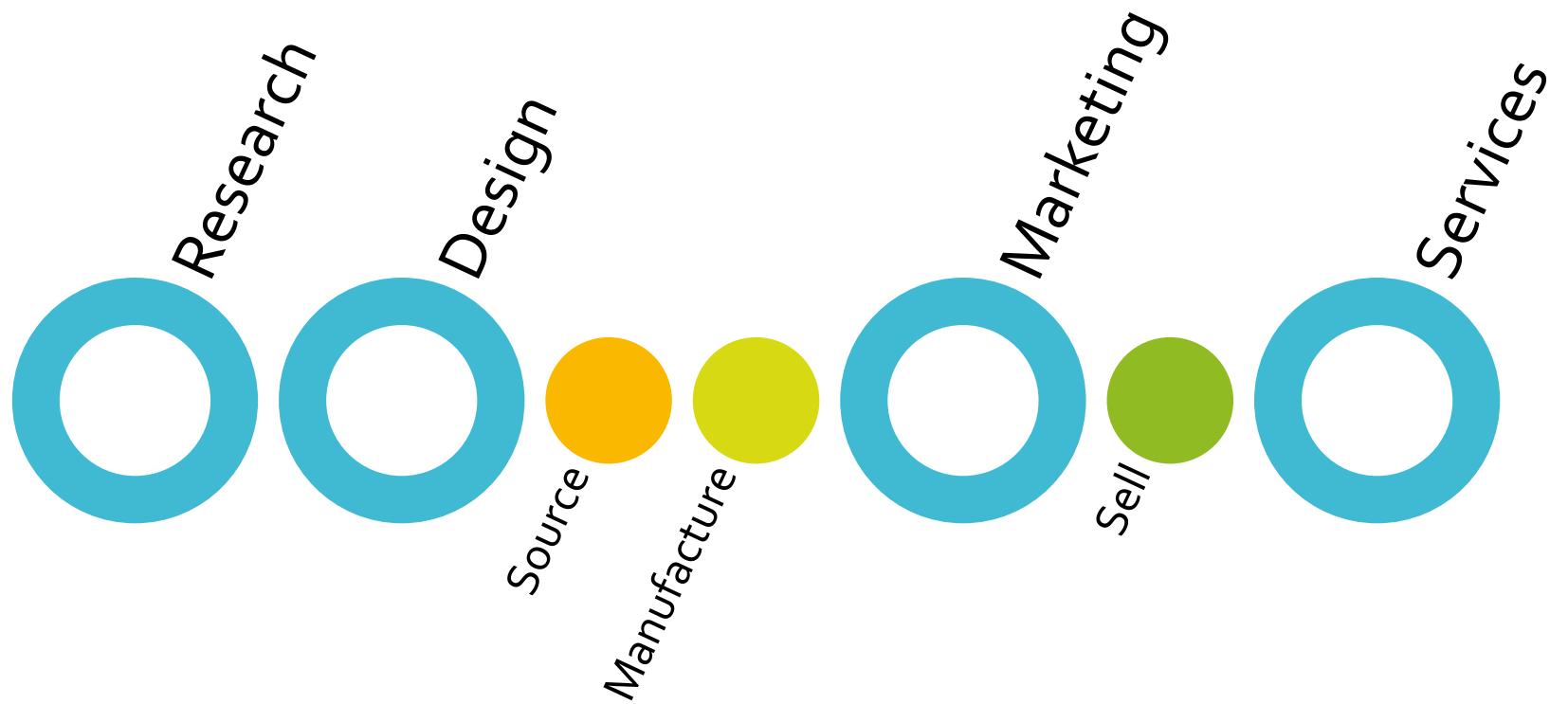
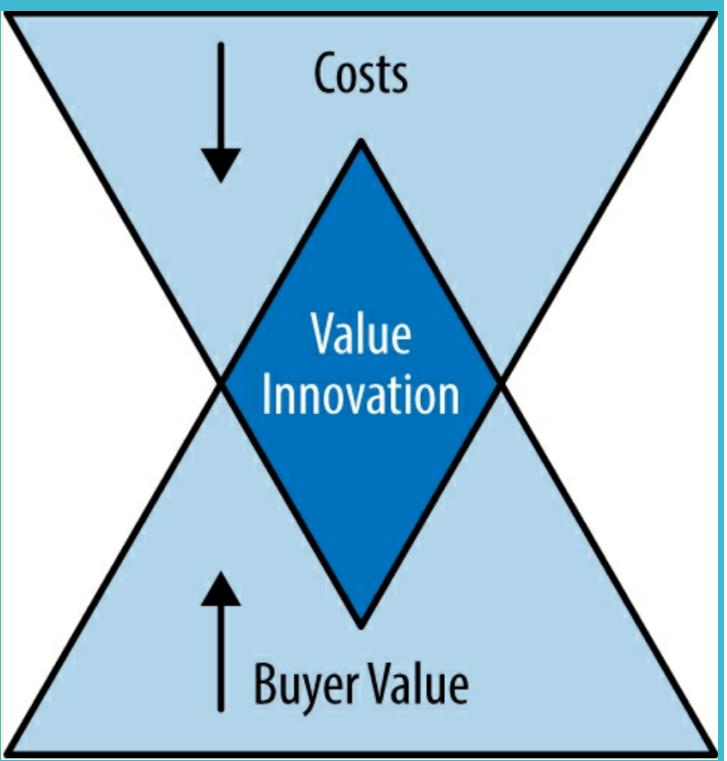


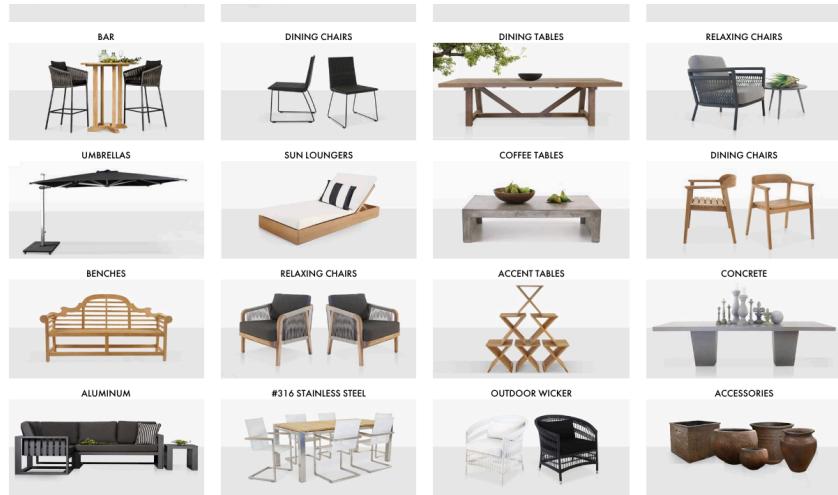
Why a UX Strategy is Crucial



- Human perceive the world by a mental model, the conceptual model in a person's mind about how a thing works.
- Usually most of the start-up today is try to change the mental model of you and me. But it is hard to get people to change their behavior.
- It is fun and a challenge to life, make the world an easier place to live in that makes for game-changing products.
- Make magic happen and destroy outdated mental models.

Value Innovation





SHOP OUTDOOR FURNITURE



NO ASSEMBLY REQUIRED
BUILT IN THE U.S.A.



ALL IN STOCK
AVAILABLE NOW



IMMEDIATE WHITE
GLOVE DELIVERY



DEEP
SEATING
CUSHIONS

An Illustration

SEO Integrate with UX Design

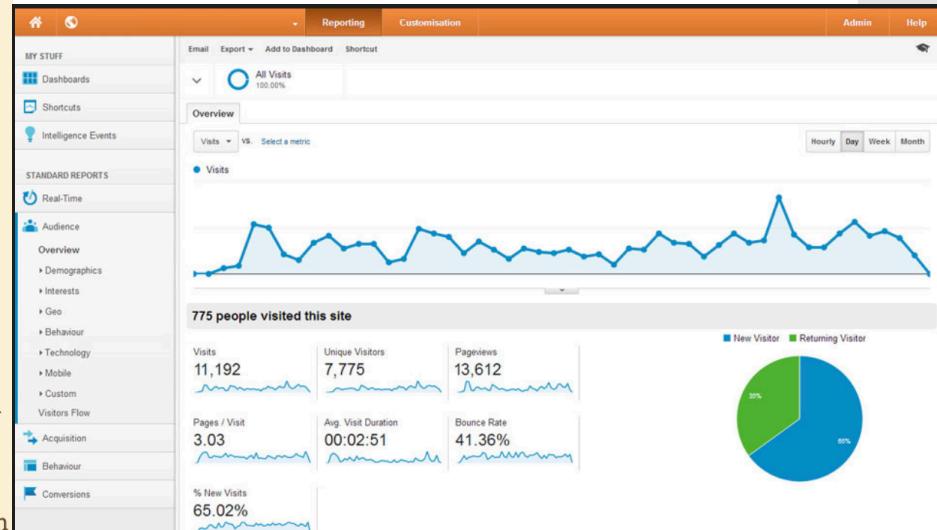


GA and SEO

Terminology

There are many elements in Analytics. These definitions will help you better understand the reports.

- **User** is a visitor who has had at least one Session.
- **Session** is a period of time that one user is engaged.
- **Pageviews** is page views by all Users and their Sessions.
- **A Bounce** is a single-page Session. A Bounce Rate is how many sessions only had one page view. 100% is bad. 60% or less is good.
- **A Goal** is a specific action that you designate as a conversion.
- **Source** refers to where the traffic of your website came from. There are many sources:
 - **Organic Search**—Visitors who come to your website after searching Google.com and other search engines
 - **Paid Search**—Visitors who come to your website from an AdWords or other paid search ad
 - **Direct**—Visitors who come to your website without a traceable referral source, such as typing your URL into their address bar or using a bookmark on their browser
 - **Social**—Visitors who come to your website from a social network
 - **Referral**—Visitors who come to your website from another website by clicking on a link; Some ads may fall into this category if they are not identified properly
 - **Other**—If you use custom parameters for custom campaign tracking, the traffic linked to those campaigns is listed here
 - **Affiliate**—If you have affiliate programs, there's a way to track it via this metric
 - **Display**—Banner or flash ad traffic



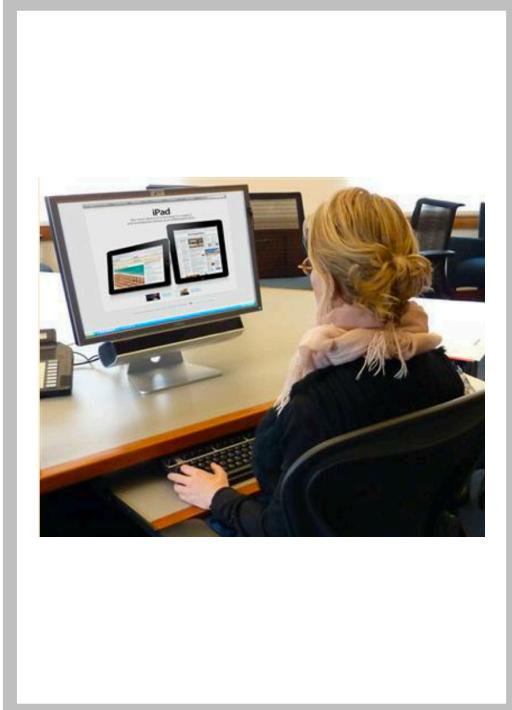
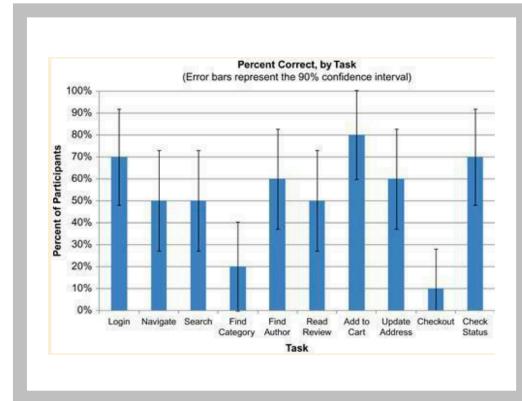
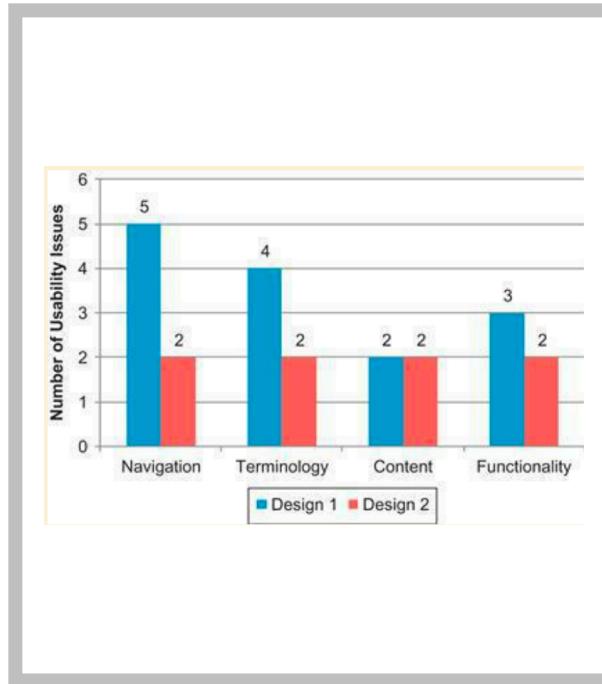
Types of UX Metrics

Performance Metrics

Issue-Based Metrics

Self-Reported Metrics

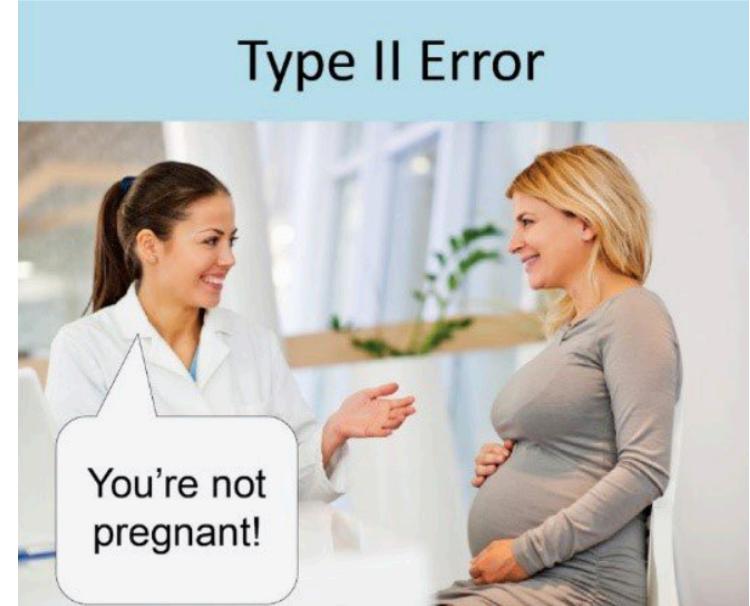
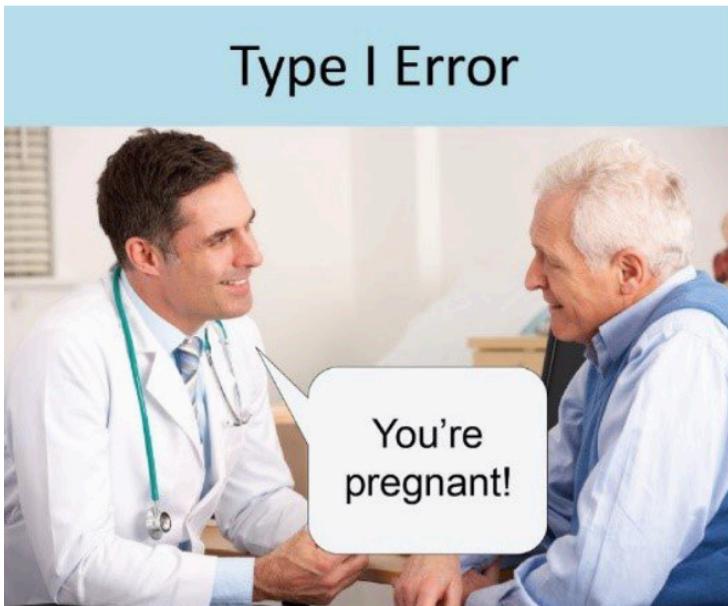
Behavioral and Physiological Metrics



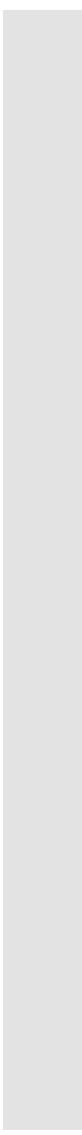
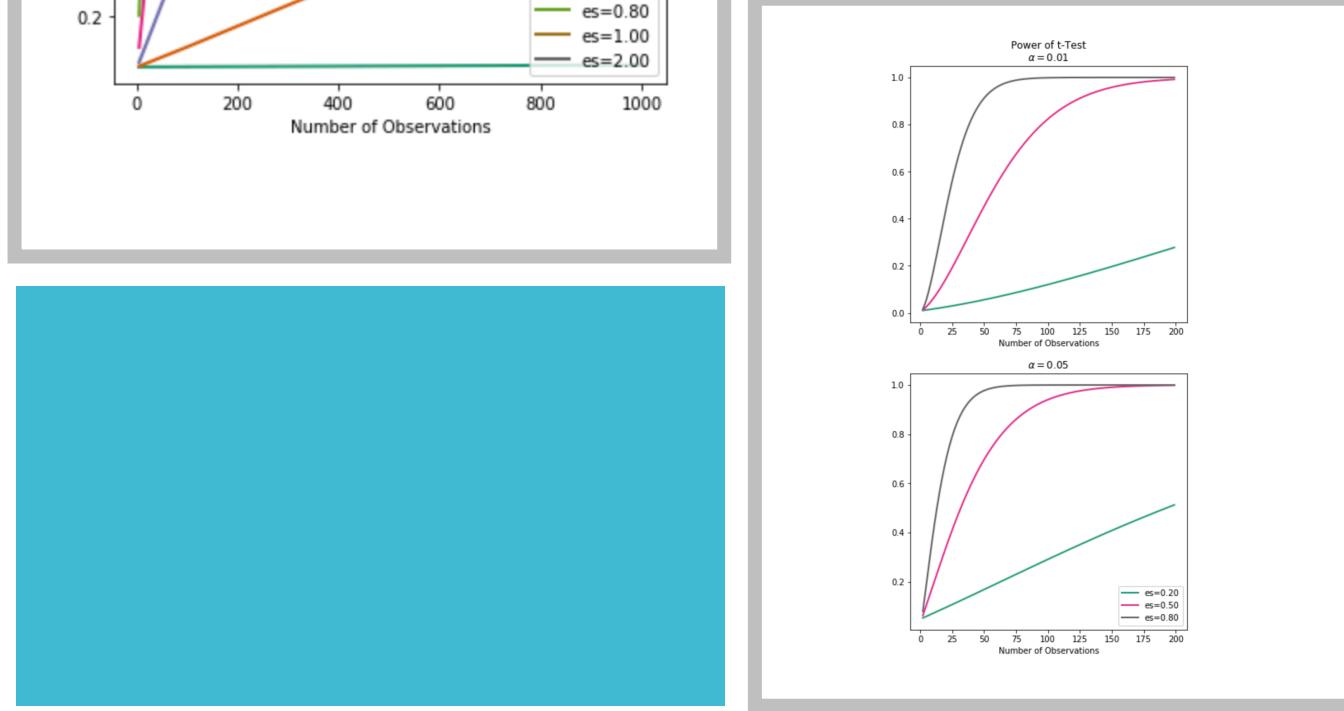
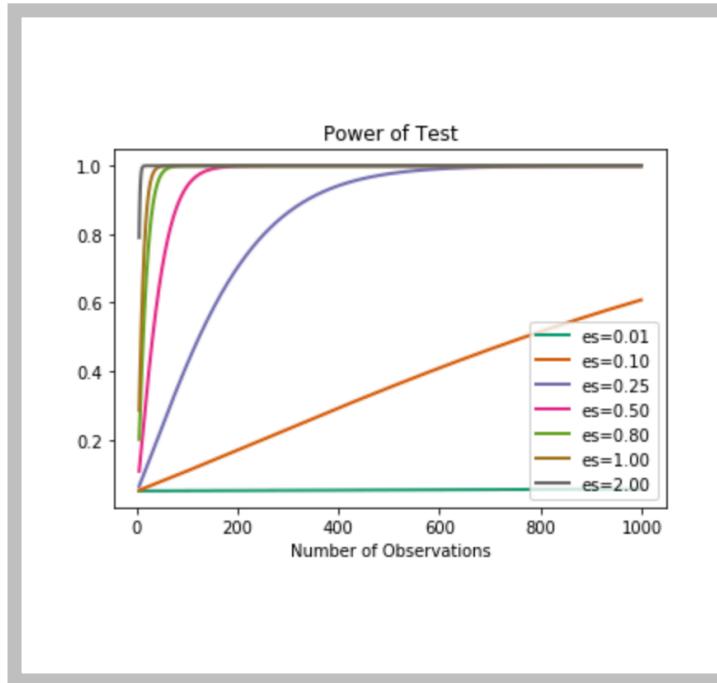
Recommended Steps in Hypothesis- Testing Analysis

1. Identify the parameter of interest and describe it in the context of the problem situation.
2. Determine the null value and state the null hypothesis.
3. State the alternative hypothesis.
4. Give the formula for the computed value of the test statistic.
5. State the rejection region for the selected significance level
6. Compute any necessary sample quantities, substitute into the formula for the test statistic value, and compute that value.
7. Decide whether H_0 should be rejected and state this conclusion in the problem context.

What is Type I and Type II Errors



Minimum Sample size and Power (1-Type II Error)



Effects and Rationale: Effect on your business, Netflix's Radical Change

Netflix wanted to understand the extent to which **the tab placement** could both affect people's impression of the company and the number of movies people streamed.



A hypothesis was put forward to change the **first tab** from being "**Browse DVDs**" to "**Watch Instantly**". At that time the company's core metrics were focused on both DVD consumption and streaming consumption — that is, how many DVDs were added to a user's queue as well as how much they streamed.



Broaden your user base by targeting new users (e.g., by opening new markets, expanding advertising efforts to new populations, etc.) or **Provide a better experience to an existing set of users** (e.g., to improve retention of those cohorts or segments)

Random Control Trials

- Randomized controlled trials (RCTs), in which researchers randomly assign the receipt of treatment. An RCT is often regarded as the **gold standard** for establishing **causality in many scientific disciplines** because it enables researchers to **isolate the effects of a treatment variable and quantify uncertainty**.
- Randomization of treatment assignment enables the estimation of average treatment effect, which averages the treatment effect over a group of individuals. Suppose that we are interested in estimating the sample average treatment effect (SATE), which is defined as the average of individual-level treatment effects in the sample.

Imai, Kosuke. Quantitative Social Science: An Introduction (p. 49). Princeton University Press.
Kindle Edition.

The **sample average treatment effect** (SATE) is defined as the sample average of individual-level causal effects (i.e., $Y_i(1) - Y_i(0)$):

$$\text{SATE} = \frac{1}{n} \sum_{i=1}^n \{Y_i(1) - Y_i(0)\},$$

where n is the sample size, and $\sum_{i=1}^n$ denotes the summation operator from the first observation, $i = 1$, to the last, $i = n$.

Test for Randomness:
Run test of randomness is a statistical test that is used to know the randomness in data.
A run is a sequence of similar or like events.
Sometimes called the Geary test, and it is a nonparametric test.

- The test applies to test the **random control trials** of an experiment. Is each unit randomly assigned either to the treatment or control group.
 - **Null Hypothesis:** The order of the Treatment and the Control is random.
 - **Alternative Hypothesis:** The order of the Treatment and the Control is not random.
- Example: On an experiment, the test wants to see whether the subjects being treated or not treated in a random manner. He observes the first 25 people, with the following sequence of Treatment (T) and Control (C).
 - TTT CC TTTT CT CCC TTTT CC TTT CC
 - Test for randomness at $\alpha = 0.05$, Z is a normal

$$Z = \sqrt{\frac{2n_1 n_2 (2n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2 (n_1 + n_2 - 1)}}$$

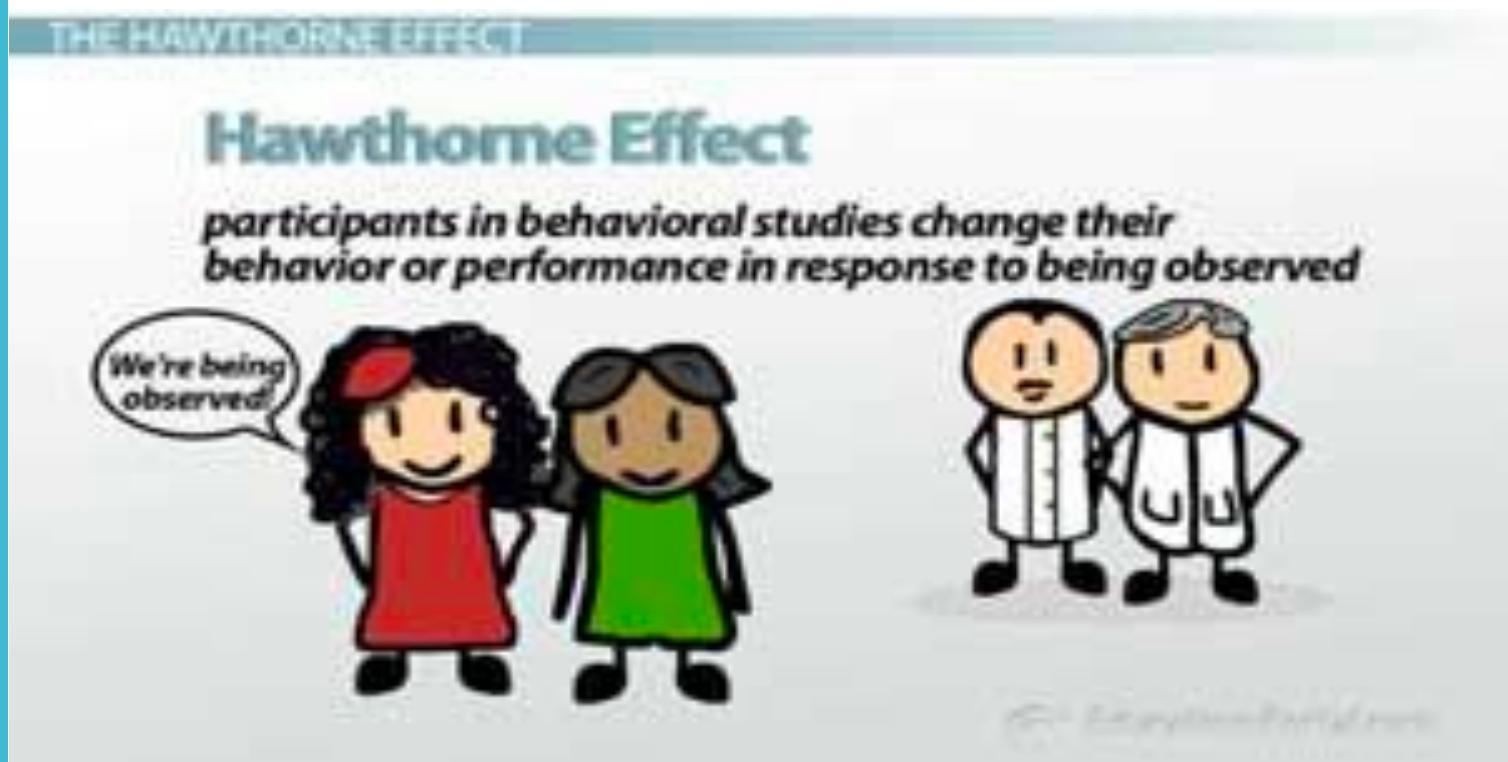
- **Test Steps**
- **Step 1:** State the null and alternate hypothesis
 H_0 : The pattern of occurrence of treatments and controls enter the experiment is random
 H_1 : The pattern of occurrence of treatments and controls entering the train is not random
- **Step 2:** Find the test statistic (number of runs)
You can easily get this by grouping each run as shown below:
 - FFF MM FFFF M F MMM FFFF MM FFF MM
- Test statistic, $r = 10$
 n_1 = number of treatments = 15
 n_2 = number of controls = 10
- **Step 3:** Find the critical value
We can find the lower and upper critical value from statistical run table
 $n_1 = 15, n_2 = 10$
Lower critical value = 7
Upper critical value = 18
- **Step 4:** Make your decision
Since $r = 10$ which is between 7 and 18, we accept the null hypothesis (we fail to reject it)
- **Step 5:** Draw a Conclusion
There are not enough evidence to reject the claim that the pattern of occurrence of treatments and controls join the experiment is determined by a random process

The Hawthorne effect refers to the phenomenon where study subjects behave differently because they know they are being observed by researchers.

<https://www.sciencedirect.com/science/article/pii/B9780128002322000134>

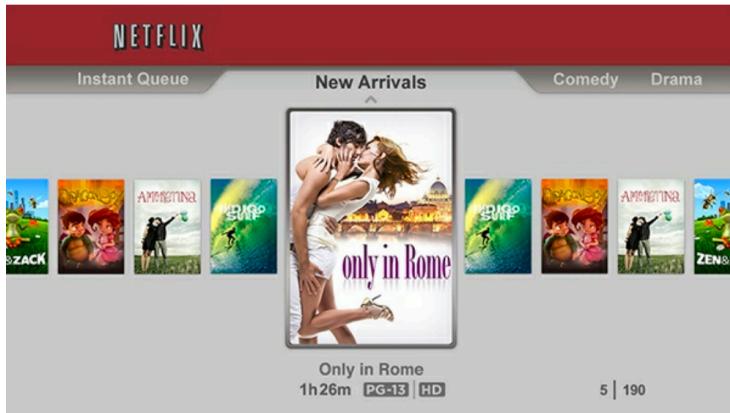
The Effect of Being Observed

Participants behave differently when observed; this is known as the **Hawthorne effect** (Landsberger, 1958). They will likely be on their best behaviour (e.g., observing standard operating procedures rather than using their usual shortcuts). It can take some time for users to feel comfortable with you and reveal their “true behaviour.” Users cannot keep up a façade for long, so you can expect aspects of the Hawthorne effect to diminish over time; the longer you spend observing participants and the better rapport you are able to establish with them, the lesser this effect will be.



Case Study: Netflix on the PlayStation 3

The Original Netflix design on
the Sony PlayStation Platform



This is the front page of Netflix
web site



Case Study: Netflix on the PlayStation 3

- Hypothesis 1 :By clearly giving users easier access to the entire catalog, they will be more likely to explore the breadth of the catalog and find a movie that they like, which will result in more hours of content consumption.” In the original experience, many users expressed a concern that they were only getting access to a limited part of the catalog. The basis for Hypothesis 1 was to ensure that the user felt confident that they had access to the entire catalog. The downside of allowing users to browse the entire catalog might result in a more complicated experience.
- Hypothesis 2 : By providing users with a simple interface that replicates the website experience, users will apply their existing understanding of how Netflix works, which will result in more hours of content consumption.” In the second hypothesis, the team strove to focus on simplicity as the core basis of the hypothesis. Rather than focusing on providing users with depth of access to the catalog, they thought about how to make things easier. One aspect of this hypothesis was that the website was an understood paradigm for Netflix users and replicating that interface on the TV would result in more streaming due to the familiarity and ease of navigating through the experience.

Case Study: Netflix on the PlayStation 3

- Hypothesis 3 : By separating the act of navigation from the act of browsing for a movie or TV show, the interface and experience will be simpler, which will result in more hours of content consumption." the driving concept was that you could simplify the experience more by separating the navigation from the content itself. Assuming know that people know and want to watch a "drama" before they even launch Netflix. By simplifying the available choices based on this knowledge and guiding them through a step-by-step process to select the movie, the team assumed the process would be faster and easier. The team took away as many distractions of choice as they could at each step along the way.
- Hypothesis 4 : By replicating a TV-like experience where discovery is based on video streaming, users will serendipitously find more things to watch, which will result in more hours of content consumption." In this final hypothesis, assumming people watch TV today. In other contexts, people are used to turning on the TV, sitting back, and just flipping through channels to serendipitously find something to watch. This hypothesis aimed to replicate that experience within Netflix, by allowing customers to decide what to watch through the act of watching.

Project ONE

- Testing your understanding about the customers
- Workshop: Course Project One (deadline 31st March 2020)
 - <https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data>