ECON7960 User Experience and A/B Testing

Hong Kong Baptist University

Topic 2: Experience versus Experiment: Explore-Exploit Dilemma

Content

- Role: A Scientist or A Businessman
- Formulate a Framework of Experimentation
- Understand the Decision Mind
- Workshop: A/B Testing Application on Pricing
- Background: King Chapter 3. Wendel Chapter 1

A Quick Review

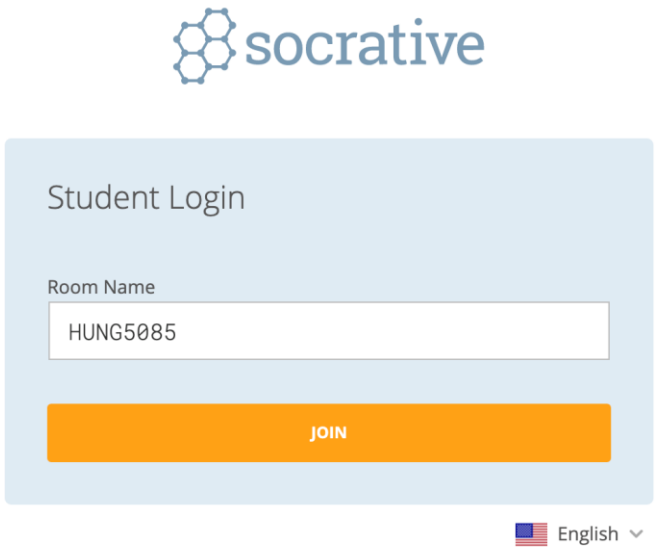
- What did we learn last week

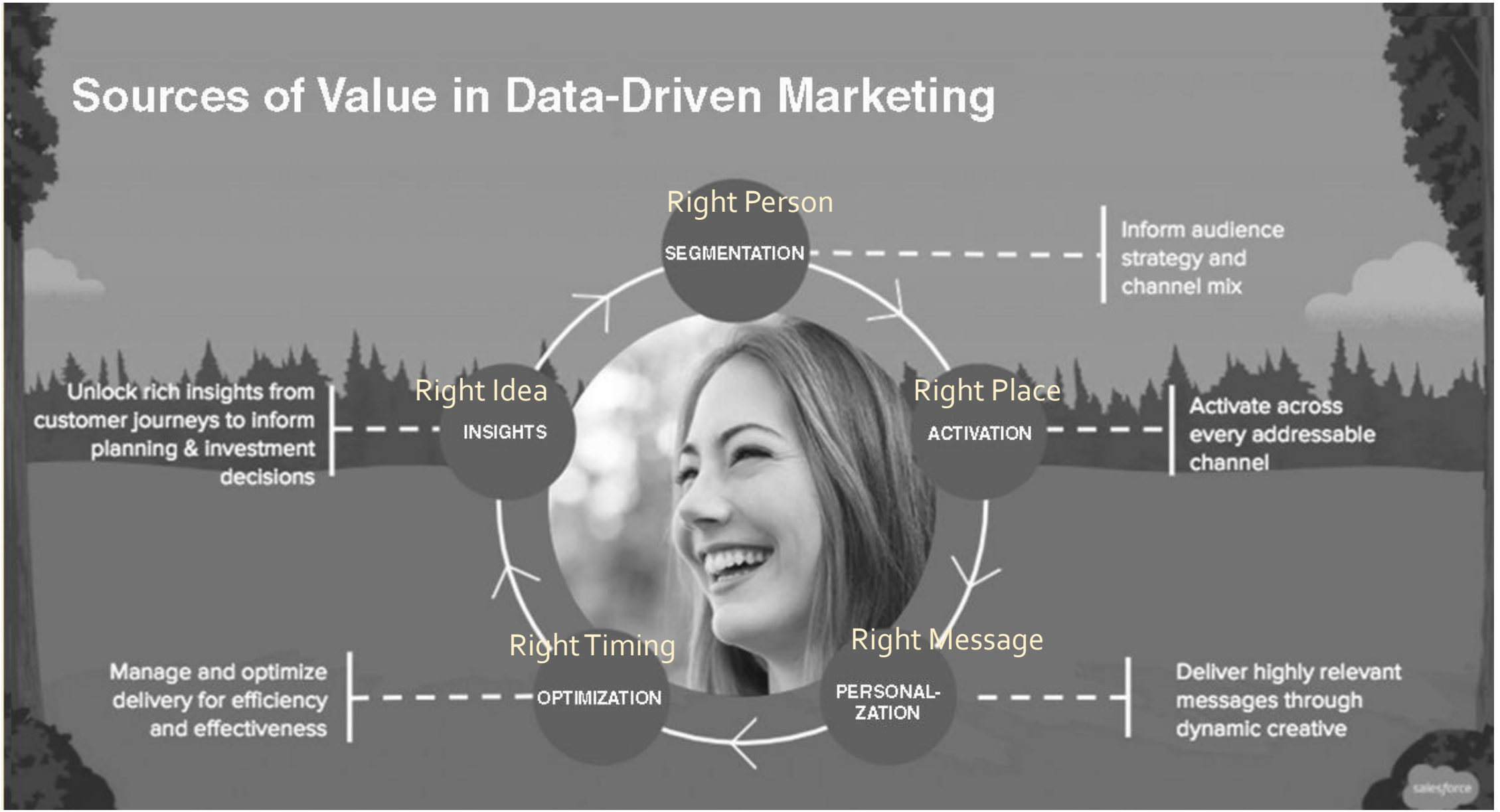
- Reference to: King, Rochelle. Designing with Data Chapter 1 and 2. O'Reilly Media. Kindle Edition.

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

Enter the Room Name "HUNG5085"

When most of you are ready, we can start together




Sources of Value in Data-Driven Marketing
Right Person
SEGMENTATION
Inform audience strategy and channel mix
Unlock rich insights from customer journeys to inform planning & investment decisions
Right Idea
INSIGHTS
Right Place
ACTIVATION
Activate across every addressable channel
Right Timing
OPTIMIZATION
Right Message
PERSONAL-ZATION
Manage and optimize delivery for efficiency and effectiveness
Deliver highly relevant messages through dynamic creative
salesforce

Last week workshop revisit

It is related to a test whether the new landing page attract more people converted

Workshop 1 Basic Python Code and Simple A/B Test on Landing Page

Introduction

This workshop reviews the concepts of Python code syntax and gets students/participants to familiar with Jupyter Notebook setting for the data analytics been used later for A/B Test applications.

The case here is about the project in the course https://www.udacity.com/course/ab-testing--ud257 which design to experiment on the response to the landing page of an online learning course

The following libraries are imported to do the calculation (A typical Python programme makes use of other people open source codes to develop the algorithm for particular application)

Jupyter notebook is an interface platform for Python high level programme language that one can insert explanations and the codes, along the context as a document so that any readers can follow through the usage of the code and understand the underlying procedures. Jupyter notebook is very common for python data scientist for mutual reference.

```
In [ ]: from scipy import stats
        import pandas as pd
```

Python libraries:

random, datetime for common language program function

pandas for data file or database manilpulation

seaborn for data visulation

statistics, numpy and scipy are for statistical and mathematical formula/function

```
In [ ]: import numpy as np
        import random
        import matplotlib.pyplot as plt
        %matplotlib inline
```

```
In [ ]: import statsmodels
```

```
In [ ]: import statsmodels.api as sm
```

Table of Contents

Part I - Probability Part II - A/B Test Part III - Regression

Introduction A/B tests are very commonly performed by data analysts and data scientists.

In this project I will go through the results of an A/B test run by an e-commerce website. The goal is to work through this notebook to help the company understand if they should implement the new page, keep the old page, or perhaps run the experiment longer to make their decision.

Part I - Probability First I will import the libraries.

```
In [ ]: df = pd.read_csv("ab_data.csv")
```

```
In [ ]: df.head()
```

Points to noted

Notice which version of python (2 or 3) is used. Remember version 2 will not be supported in future although it has lots of code libraries.

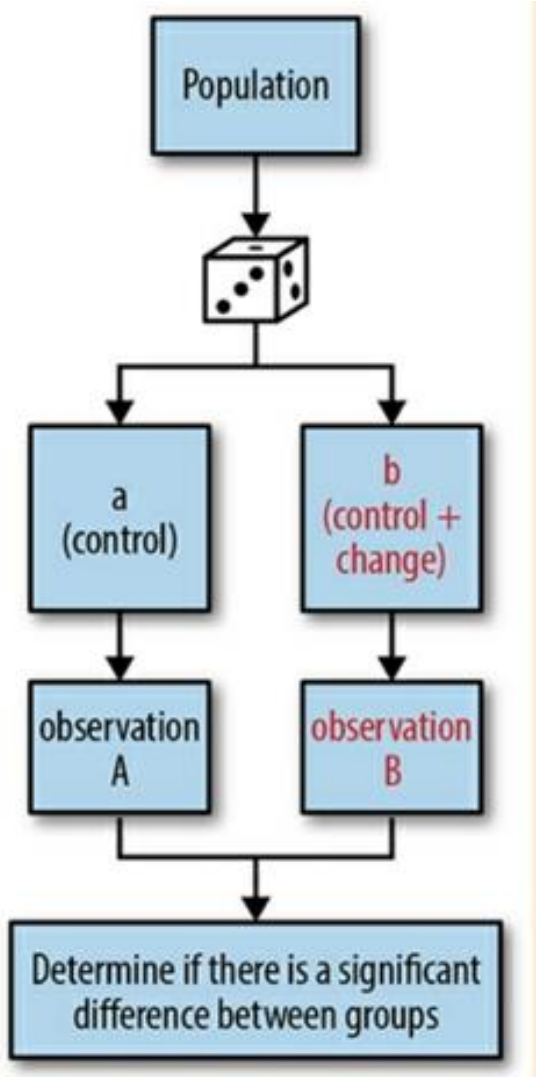
When timestamp is used, the format of the time is relevant. It may take more time in coding to get relevant information from the timestamp data

Floating point problem

Hypothesis testing depends on the formulation of the hypothesis. In here

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).

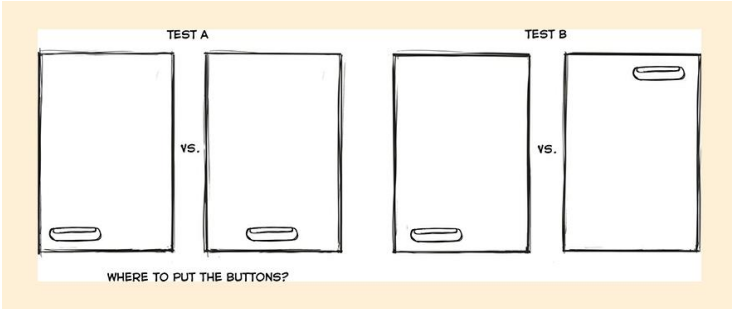


The Dilemma: Mind of a scientist and a businessman

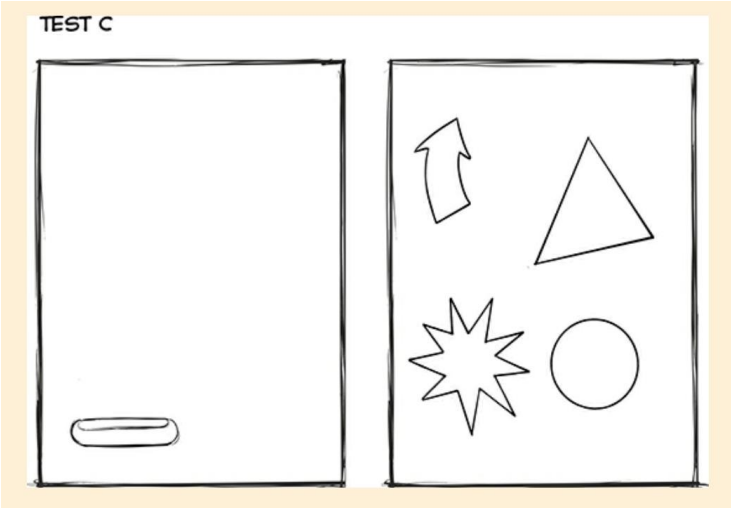
- A small business couldn't afford to behave like a scientist and spend money gaining knowledge for knowledge's sake
- To minimize the risk of business, one need to understand the need of customer by experiment to capture on their needs, as a good experiment, one need to make a proper apples to apples comparison
- One need to (A) learn about new ideas (which always call exploring from now on), while one also need to (B) take advantage of the best of one old ideas (which always call exploiting from now on)
- How to balance seeking the knowledge of your customers to capture them expected in the long run or seeking profit by just conjecture what is the best to customers without testing.

A/B Test: Local or Global (Minor or Major)

Minor Design Change



A Major Design Change



Esty's Story:

The first phase was to do some big redesigns of the page, and we spent the first half of the year doing big changes. …Search a variant that are happy with and was even remotely close in terms of performance to what had now.

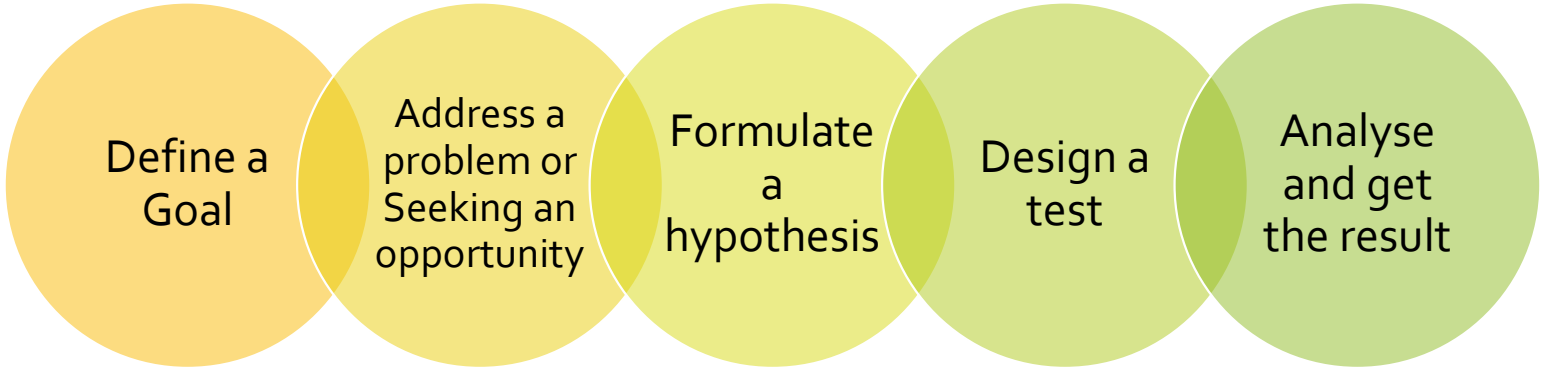
Decide to devote 5% of traffic to a new variant without making people too angry. Spent 2-week iterations from January until July, trying all sorts of different things. Found one which looks very close to what Etsy's listing page looks like today. "OK, optimize it until it's better."

Between July and October. Use smaller A/B tests on that variant until felt right.

When finally out of time for the year, had a variant that was 5% better in terms of conversions on the existing listing page. At the time, it was like 50 million dollars.

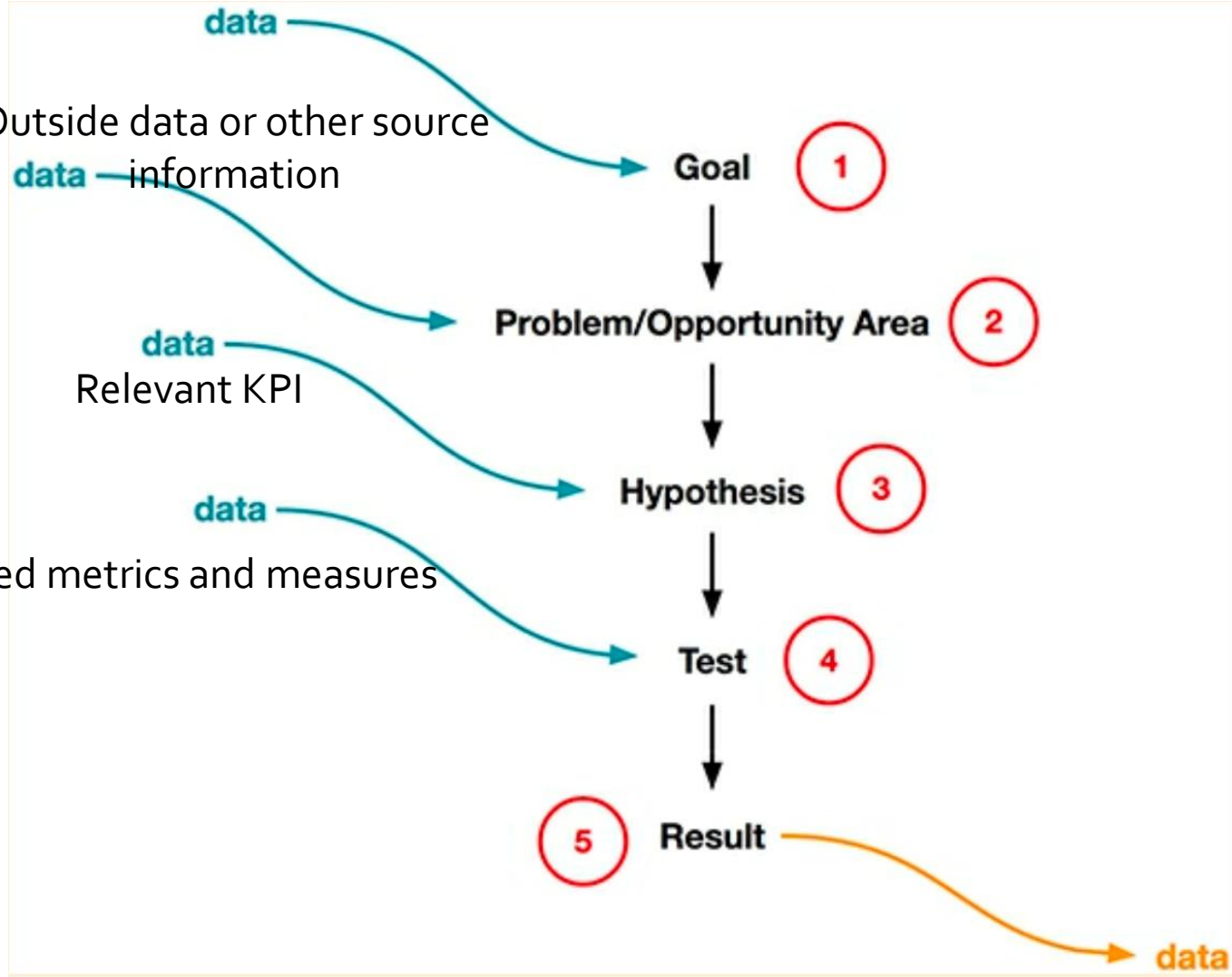
King, Rochelle. Designing with Data (pp. 94-95). O'Reilly Media. Kindle Edition.

An Experiment Framwork



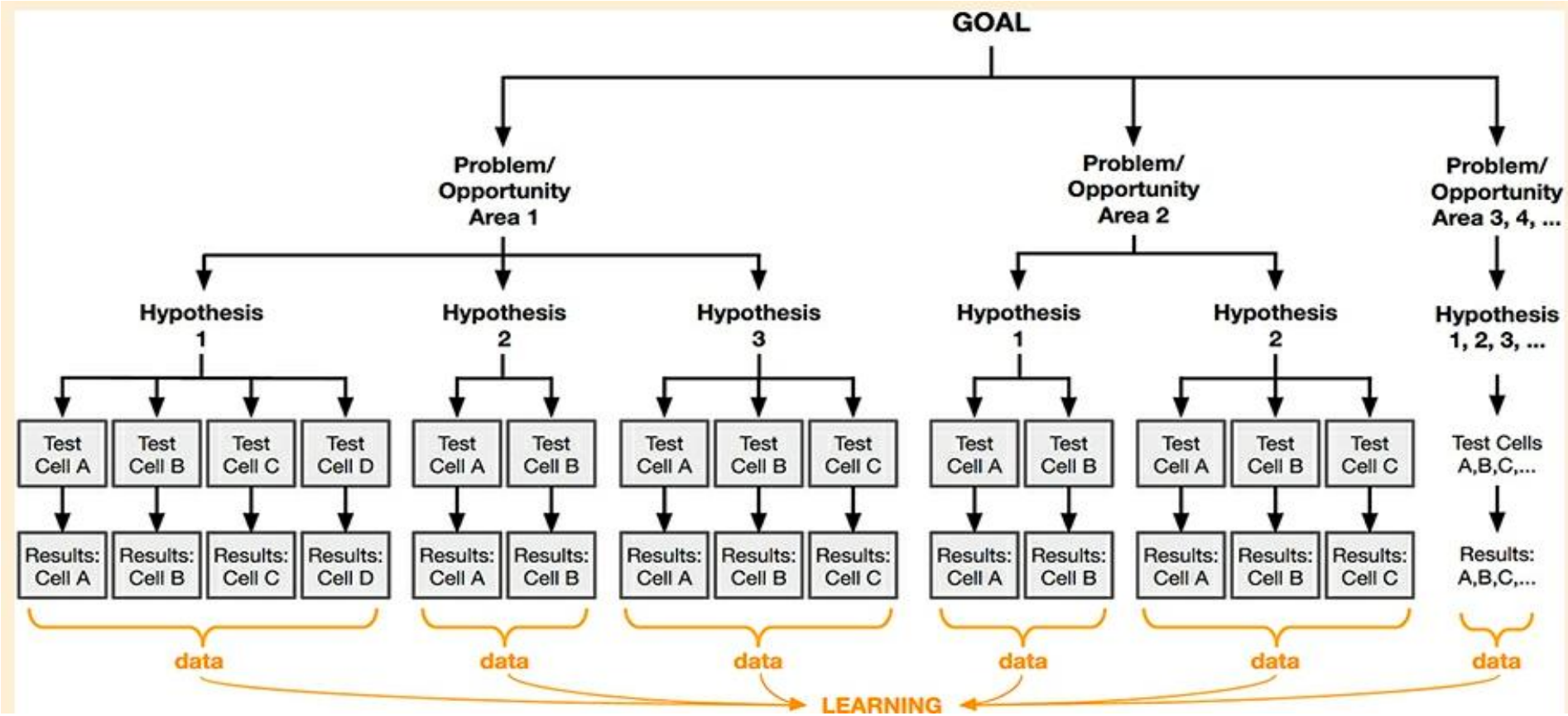
All Steps Related to Data but only One Step Test on the Measures or Metrics

Past data observed in the system



Working with Data = parts of {Art} {Craft} {Science}

- In a data aware design, one should always be thinking about the volume of ideas or potential solutions that can generate, the variation between those ideas or solutions and how one can learn the most about.
- Build an experimentation framework with the goals of generating and articulating multiple solutions that you can test and learn from along the way.
- Every experiment comes with the costs of designing and launching a test to users also takes time and energy, so ensuring that you are running meaningful learning tests rather than relying on A/B test as a "crutch"
- With a perspective of continual learning, it is important that avoid just think about each experiment conduct in isolation, rather think them about the series of experiments one can learn more about their users and customers

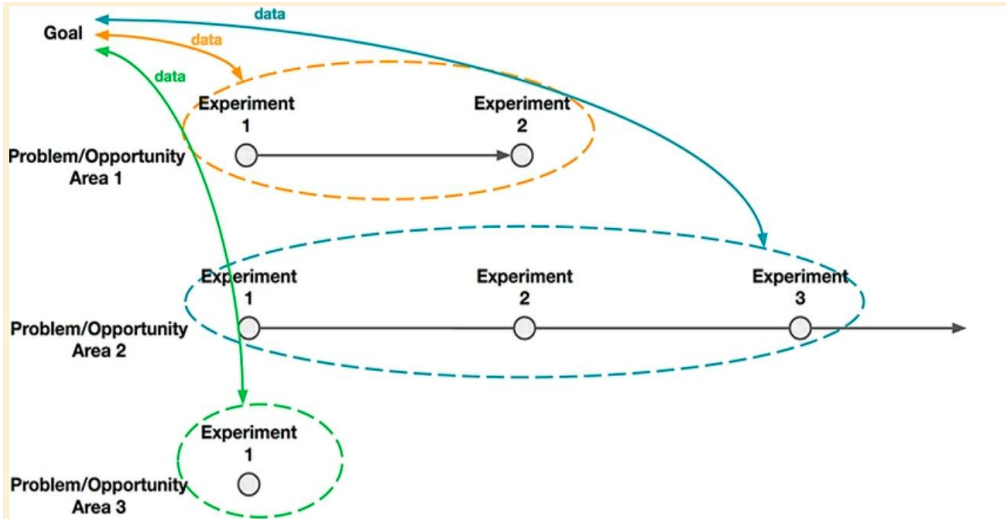


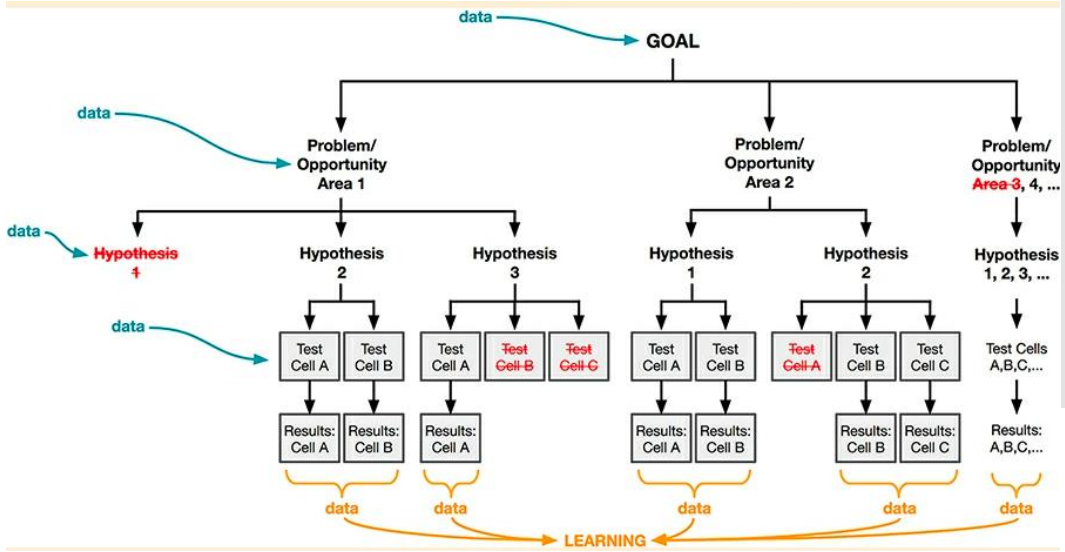
Identifying multiple problem areas or opportunities with multiple hypotheses and multiple tests

Here there are not only A/B test but also with multi-variate test. But how to proceed?

Three Phases

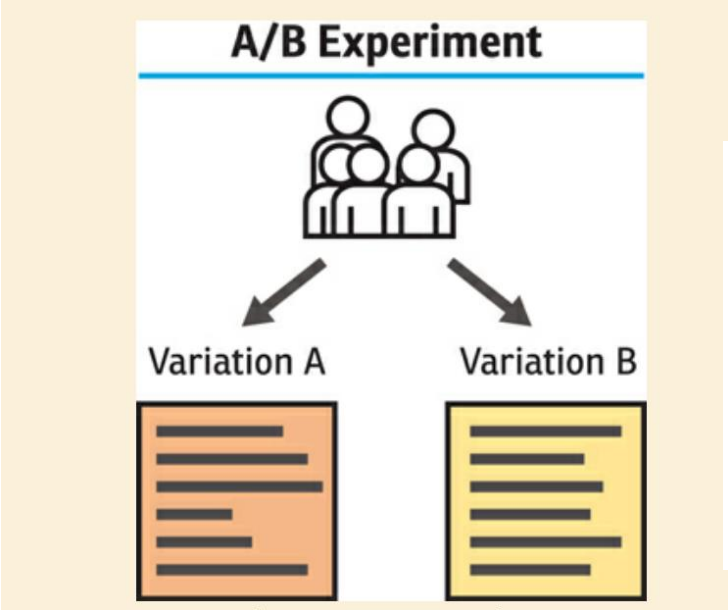
- Define the goal: making more money or other goals
- Decide which things to test for, try some of the ideas out by executing against them
- Once set up experiment, able to test, eliminate possible path by analyzing the result





Different Between A/B Test and Multivariate Test (MVT)

A/B Test
Simple Hypothesis Testing



MVT
Machine Learning Approach

	Condition 1	Condition 2
Factor A Tagline	The World's Best Sausages	The Sausage Factory of Seattle
Factor B Call to action	**Buy 2 Sausages, Get 1 Free**	**Free delivery on all sausages Buy now**
Factor C Image on site		

A Simple Test or a Complex Test

- Stage1: One are focused on testing fast and delivering results; A/B testing is the perfect method. Using a simple hypothesis, you can get fast results to improve ROI.
- Stage 2: One will continue to A/B test but introduce MVT to bring greater depth and learning where suitable.
- Stage 3: Experiment everything? One choose a method based on the complexity of the hypothesis you're going to experiment. One will be using both methods simultaneously.

McFarland, Colin. Experiment! (p. 48). Pearson Education. Kindle Edition.

Over the past decade, the power of A/B testing has become an open secret of high-stakes web development. It's now the standard (but seldom advertised) means through which Silicon Valley improves its online products. "The A/B Test: Inside the Technology That's Changing the Rules of Business," Brian Christian, Wired

Get to A/B Testing First

Merits of A/B Test

• A/B testing forces you to focus. MVT can tempt you to add unnecessary complexity to the experiment.

• A/B testing requires less traffic and takes less time to reach conclusive results.

• A/B testing is faster to develop and easier to submit to technical quality assurance than MVT.

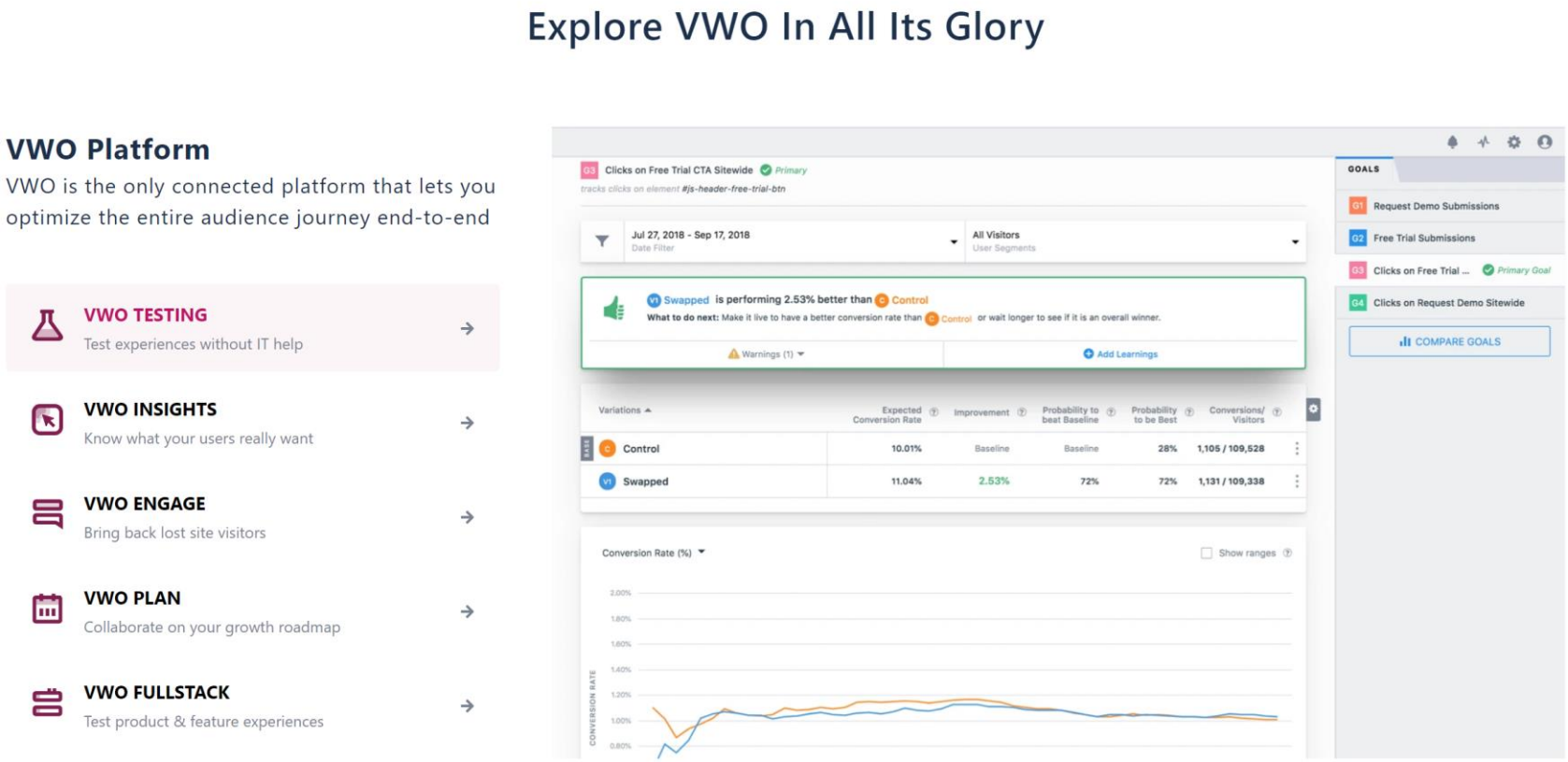
There are **obvious benefits** to running an **MVT test**.

• MVT lets you experiment with many independent factors at the same time, hugely increasing the ability to learn and discover.

• MVT tells you not only what the winners are, but also the most and least influential variables in your experiment whereas An A/B test will only tell you the winner, and won't capture interaction effects.

Tools and Vendors

- Visual Website Optimizer is an easy-to-use A/B and MVT tool that lets you create your first experiment in under 5 minutes.



https://conversion-rate-experts.com/split-testing-software/

Compare the cost different tools

Tools and Vendors

Plenty of supporting material exists online for developers building a custom platform. In fact, perhaps one doesn't need to start from scratch at all; there are prebuilt frameworks that they could build upon.

The right tools can give the flexibility and features to run on specific experiments, but they won't make one to be the world's best experimenter. Ultimately, one will still have to design and execute winning ideas, and can develop that with the simplest of tools.

Vanity

Experiment
Driven Development

Intro
Metrics
A/B Testing
Using with Rails
Testing emails
Managing Identity
Configuring
Adapters
Contributing
Experimental

build passing

Who's using it?

Welcome to Vanity

Vanity is an Experiment Driven Development framework for Rails.

Sidebar (A/B Test)

Sidebar, sidebar, who's got a sidebar!

Option A: `false`	**80.0% (60% better than option B)**	**showing**
Option B: `true`	50.0%	show

The best choice is option A: it converted at 80.0% (60% better than option B). With 90% probability this result is statistically significant. Option B converted at 50.0%. Option A selected as the best alternative.

Started Fri, Nov 13 2009

PlanOut

A FRAMEWORK FOR ONLINE FIELD EXPERIMENTS

GET STARTED GITHUB

A 15 minutes BREAK

Behavior Economics

The Deliberative and Intuitive Mind – A broad understanding of human

Making Sense of the Mind

- Ambiguity effect

One are intuitively uncomfortable with actions in which the potential effects have unknown probabilities. This makes people avoid otherwise preferred options when uncertainty is added

- Anchoring

One automatically use an initial reference point (anchor) as basis for estimates, even if the estimate is wrong. For example, the initial listing prices for houses, even if completely invalid, strongly affect how much buyers (and real estate agents) think the house is worth

- Attentional bias

One pay attention to particular cues in the environment based on the internal state. For example, people who are addicted to a drug are extra sensitive to cues related to their addiction. They effectively see things that relate to the drug more often than everyone else, whether they want to or not

Wendel, Stephen. Designing for Behavior Change: Applying Psychology and Behavioral Economics . O'Reilly Media. Kindle Edition.

Making Sense of the Mind

- Availability cascade

Incorrect (and correct) ideas can become increasingly believed and widespread because of (a) repetition by well-meaning people who don't want to appear wrong, and (b) manipulation from interested parties. [ref112] cite the example of the Love Canal toxic waste scare in New York — which, from expert accounts, was vastly overblown and was later discredited.

- Availability heuristic

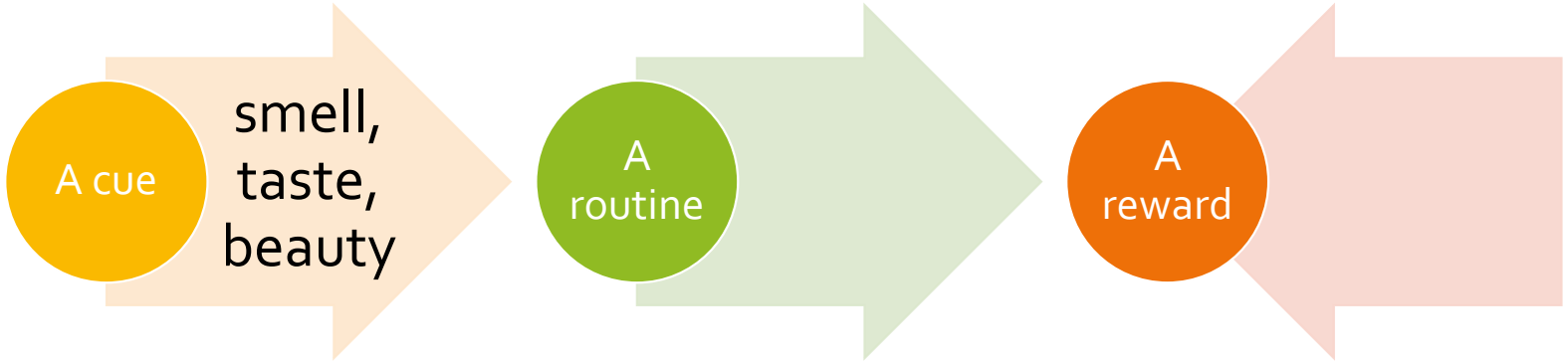
One estimates the likelihood of events based on how easy they are to remember. For example, people incorrectly believe that the names of famous people to be more common than normal names ([ref176]).

Wendel, Stephen. Designing for Behavior Change: Applying Psychology and Behavioral Economics . O'Reilly Media. Kindle Edition.

Most of the Time, people are not Actually "Choosing" What to DO Next

- Lessons for Behavioral Products: if one design a product to appeal to someone's conscious, rational decision-making process, one might educate the rational mind, but not resulted in actually affecting behavior (because human behavior is often intuitive or automatic).
- Be very clear about the type of behavior you are trying to encourage — a conscious choice or an intuitive response.

Habits Drive Intuitive Behaviors in Predictive Ways



Once the habit forms, the reward itself does not directly drive one behavior; the habit is automatic and outside of conscious control.

Inducing a Behavior: an Example

- In early 1900s, advertising man Claude C. Hopkins moved American society from being one in which very few people brushed their teeth to a majority brushing their teeth in the span of only 10 years. His advertisement helped Americans form the the habit of brushing
  - He taught people a cue – feeling for tooth film
  - When people felt tooth film, the response was a routine, brushing teeth
  - The reward was minty tingle in their mouths

Two Experiments about Cue

Case 1

- Cues "prime" people to act in a way that is appropriate for that frame of reference.
- A famous study about what happens when we're primed to think about stereotypes. Researchers divided a set of Asian American women into three groups, each of whom were asked a set of questions about their lives, and then subsequently took a math test.
  - The group that received questions relating to race later answered 54% of the math questions correctly;
  - The group that received questions relating to gender later answered only 42% correctly, and
  - those with generic questions were in between, with 49% correct

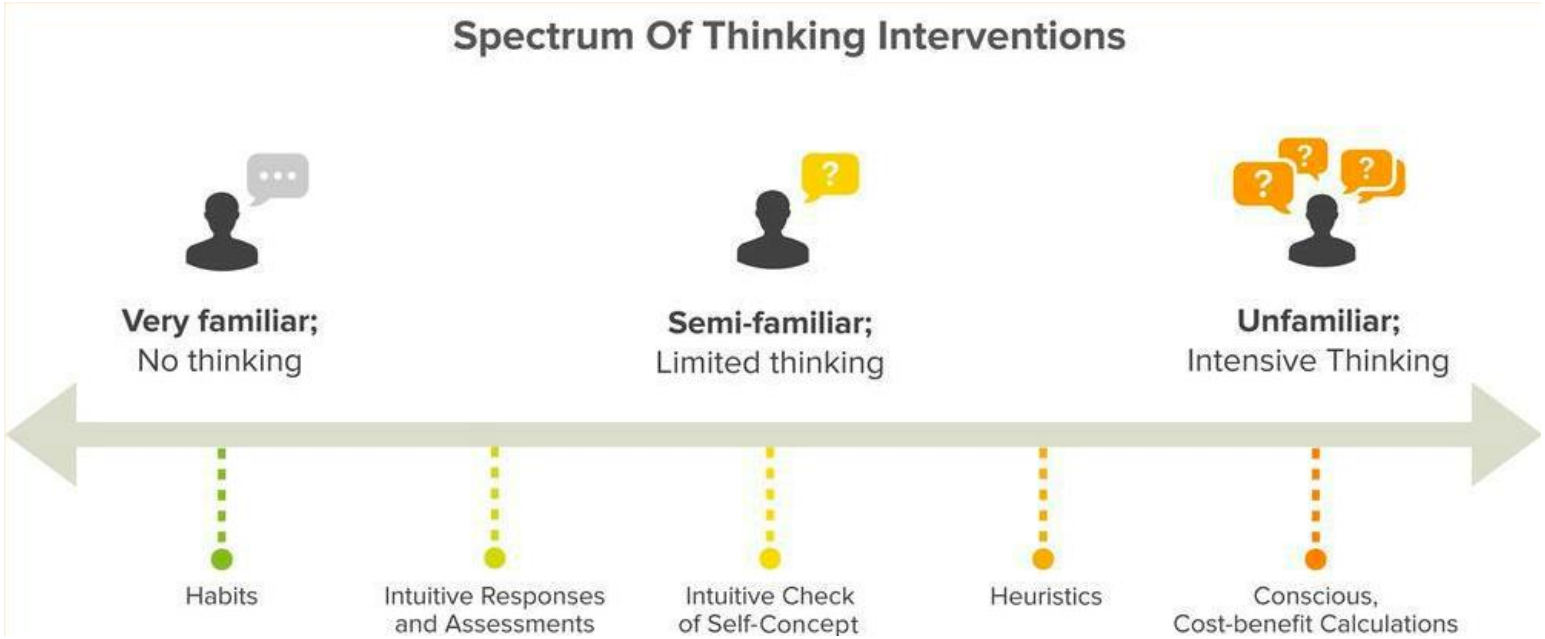
- Wendel, Stephen. Designing for Behavior Change: Applying Psychology and Behavioral Economics . O'Reilly Media. Kindle Edition.

Case 2

- Randomly selected students who were given a positive interpretation of their early problems in college came to see themselves as capable and performed better on tests than their randomly selected fellows

Even When We Exercise Our Brain, Our Mind Save Work

- We Find Easier Problems
- Our Peers Provide Answers
- Our Mental Resources Are Sorely Limited
  - Memory
  - Attention
  - Willpower and mental energy



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

5 Mechanism at where it is mostly likely to be used

- Habits – Familiar cues trigger a learned routine
- Other intuitive responses – Familiar and semi-familiar situations, with a reaction based on prior experiences
- Active mindset or self-concept – Ambiguous situations with a few possible simple interpretations
- Heuristics – Situations where conscious attention is required, but the choice can be implicitly simplified
- Focused, conscious calculation – Unfamiliar situations where a conscious choice is required or very important decisions that direct one attention toward

- The important question is what you want your customer purchase or usage pattern, choose the mechanism wisely