

---

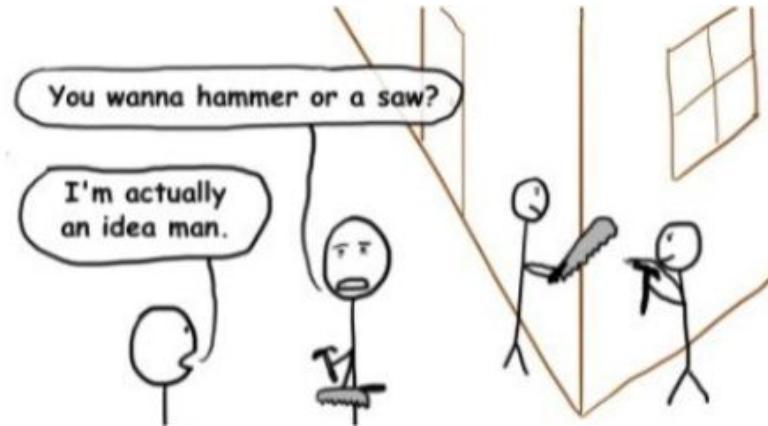
# Customer Analytics Intro

**Reid Bryant**  
Sr Principal Data Product Manager at Red Hat  
Email: [reidbryant@gmail.com](mailto:reidbryant@gmail.com)



# My background

- UNC undergrad, class of '05
- 8 years of real estate + finance stuff
- IAA grad, class of '14
- 4 years @ Brooks Bell
  - experimentation analytics
- 1.5 years @ Blue Acorn iCi
  - marketing analytics
- ~5 years @ Red Hat
  - marketing + product analytics
- Still learning daily :)



# WHY SHOULD YOU CARE ABOUT CUSTOMER ANALYTICS?

**“Half the money I spend on advertising  
is wasted; the trouble is I don't know  
which half.”**

- John Wanamaker

# What's your goal for today's lecture?

My napkin math says you will spend ~500 hours listening to lectures here at the IAA

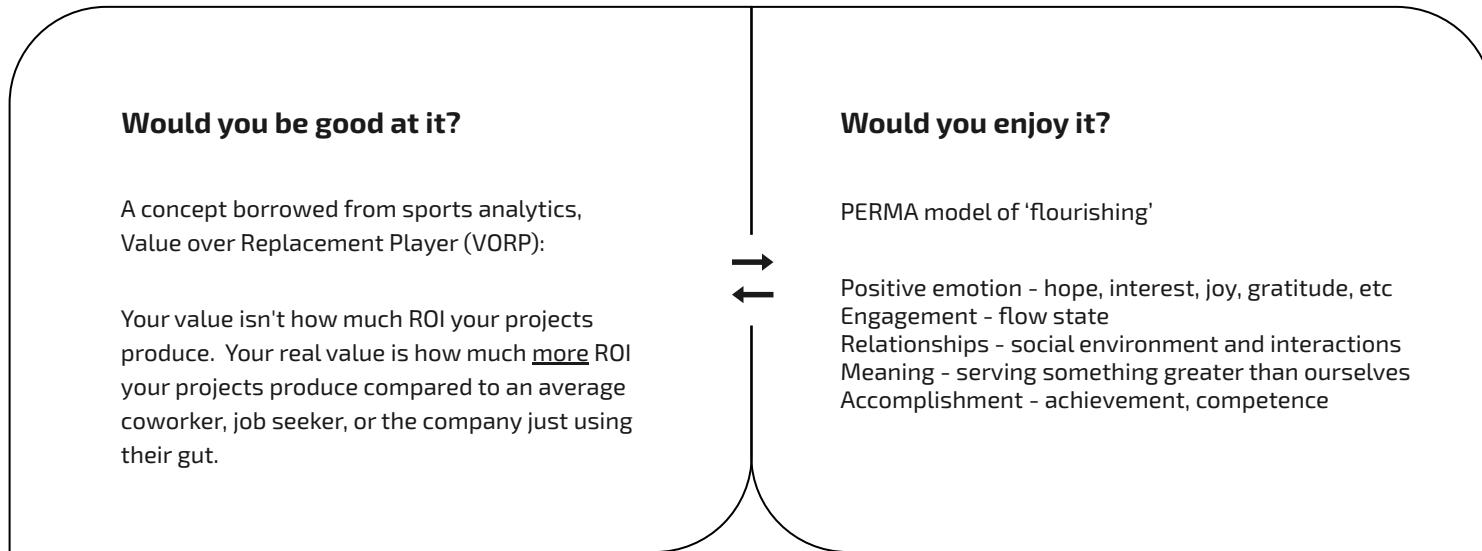
Two truths:

- 1) That's a lot of time, but not enough to be 'great' at anything
- 2) What you pay attention to grows; you can be 'great' if you want to be

One question:

- 1) Is customer/marketing/product analytics something that you want to be great at?

# How to think about 'is this for me?'



# Would you enjoy it compared to other options?

From my experience on the marketing side...

Shorter, Iterative ----- Longer, Singular

Creative ----- Defined

Collab w/ the biz ----- Collab w/ data peeps

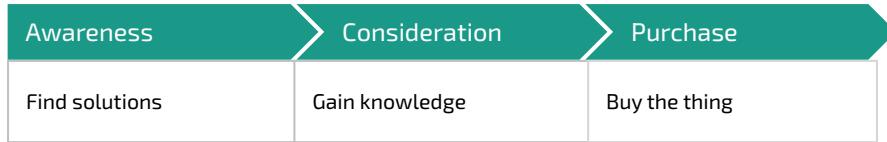
Industry tool heavy ----- Data tool heavy

Generate insights ----- Build production models

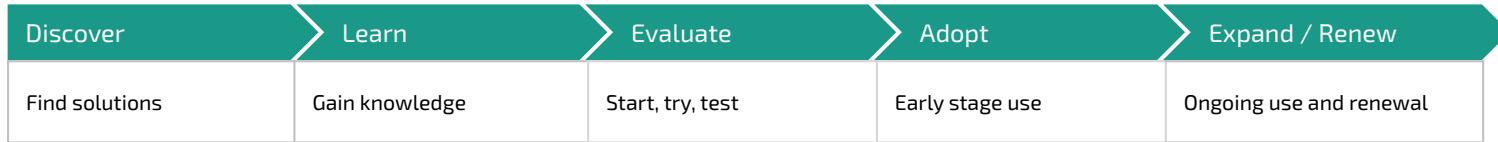
# PRIMER ON BUYER JOURNEYS

# The Buying Journey

Can be as simple as a few mins in B2C / eCommerce



Or as long as a year in complex B2B software scenarios



Huge differences in terms of the type/amount of data and effectiveness of attribution

# Channels / Touchpoints

## **How the user/account is reached:**

- Organic - do we have the right content to drive users/accounts to our website?
- Paid - are we targeting net new via paid social, search engine, display, etc?
- Email - are we growing our contact database and nurturing effectively?
- Events - are our in person events and webinars converting?
- Trials - are our freemium offerings leading to paid subscriptions?

# B2C v B2B - Traditional Models

## Business to Consumer (B2C)

- Selling to a user / customer
- Short(er) sales cycles
- Goals are to build brand awareness and convert as quickly as possible, ideally within a single browsing session
- Conversion occurring through the site system, a third party retailer, or Amazon.

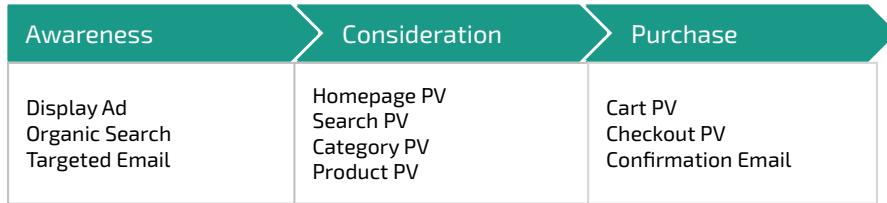
## Business to Business (B2B)

- Selling to an account which consists of many buying personas
- Long(er) sales cycles
- Early funnel goals are to build awareness, grow contact database, land inquiries
- Later funnel goals are to build 'opportunity value' in the customer relationship management and convert that to 'closed won' revenue
- Conversion usually occurring via a sales team to close the deal

# Enterprise Sales Buyer Journey Example

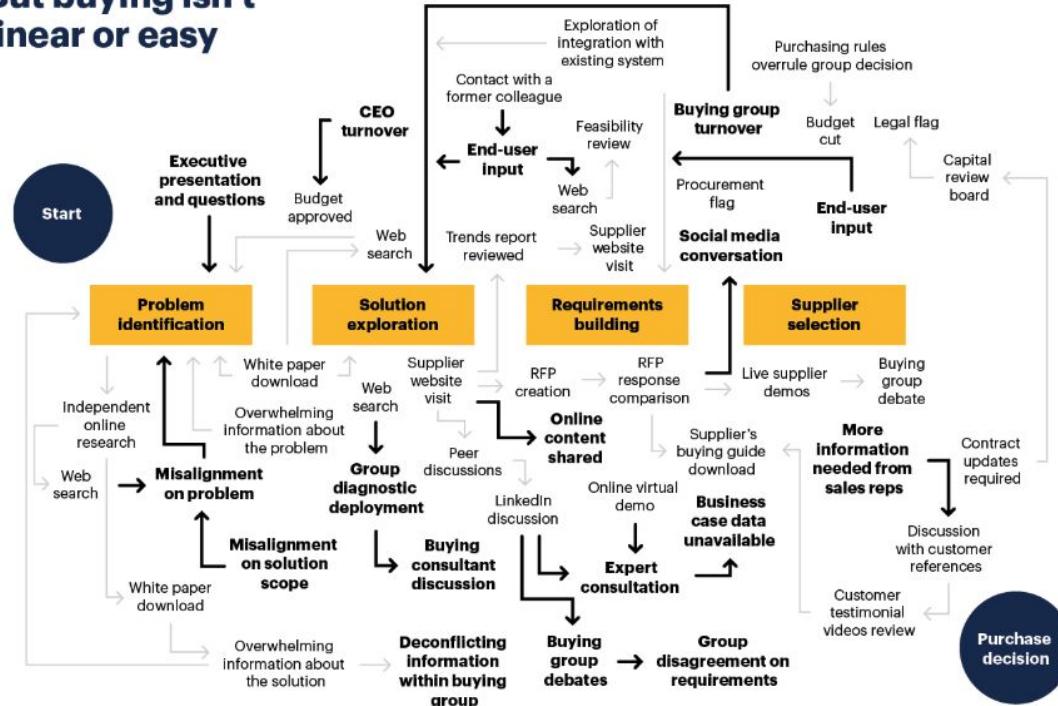
Discover	Learn	Evaluate	Adopt	Expand / Renew
General Website PV Specific Product PV	Form submit Whitepaper download On demand webinar view Account created Product Doc views Learning Path view Email Open Event Registration Event Attendance	Trial Started Trial Operator Detected Trial Project Created Trial Server Started Trial Daily Usage	Opportunity Created Opportunity Won Entitlement Issued Product Started Operator Detected Project Created Server Started Daily Usage	Same Product Renewal Opp Different Product Opp

# Ecommerce Buyer Journey Example



# Every buyer journey is unique

But buying isn't  
linear or easy



# Outcomes to Optimize

**What do we care about related to events / outcomes across the journey?**

**Optimizing the 4 V's, often against finance targets or period over period targets:**

- Volume - Target the right audience to generate qualified, early stage traffic
- Velocity - Move these users/accounts through the funnel at speed
- ConVersion - Maximize conversion through each step of the funnel
- Value - Increase revenue to return business value

# Convergence with Product Analytics

Distinction between B2B and B2C is narrowing, from both sides:

- More marketers are focusing on 'Product Outcomes'
- More product teams care about early funnel demand generation
- Has resulted in Product Led Growth models

Product Led Growth (PLG):

- PLG uses product as acquisition vehicle rather than a traditional sales funnel, typically using a trial or freemium to get users in the product ASAP
- Experiences start unauthenticated on websites, but often end authenticated in-product / in-app
- Marketing teams want to understand how their efforts against product outcomes (usually activation, onboarding, usage, shorter term retention)
- Product teams want to understand the correlation between features added and product outcomes (usually usage, longer term retention)

# WHERE CAN ORGANIZATIONS BENEFIT?

# Specific Areas Analytic can Provide Value

**1** Targeting

Which user do I match to this content?  
What content do I surface to this user?

**2** Attribution

Is this working holistically?

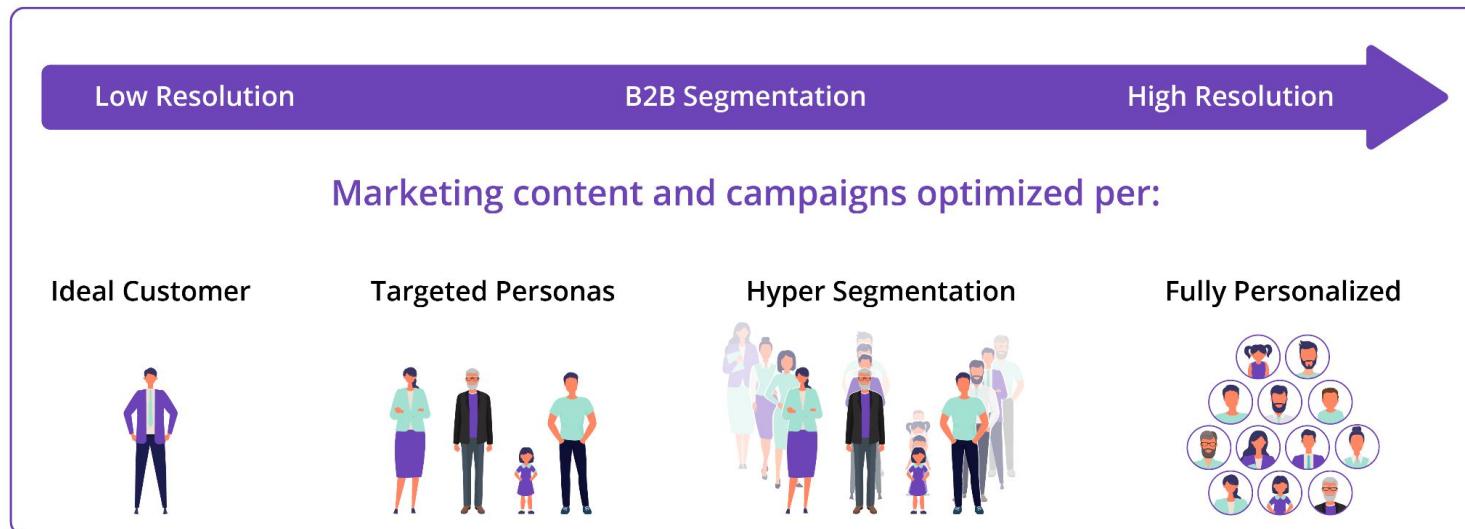
**3** Experimentation

Is this working at a micro level?

# **SEGMENTATION & TARGETING**

# Segmentation and Targeting

- Segmentation: Collecting 1st and 3rd party data attributes to segment the overall audience into smaller subgroups
  - Targeting: defining rules to match each segment to specific marketing content that has been tailored to the needs of each group



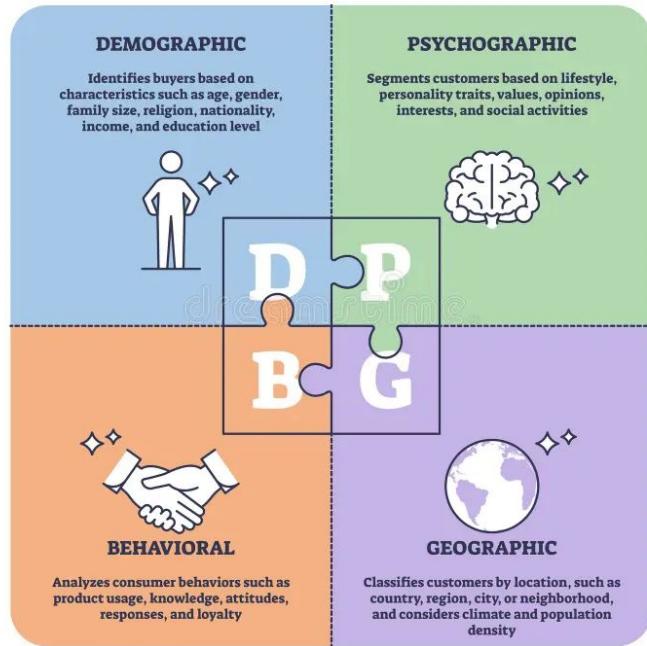
# Segmentation Factors

- Geographic
  - zip, city, country, timezone, language
- Demographic
  - age, gender, income, occupation, race
- Psychographic
  - values, goals, needs, interests
- Behavioral
  - browsing, spending, usage

AND FOR B2B ...

- Firmographics
  - characteristics that describe a company

## MARKET SEGMENTATION



# 1st, 2nd and 3rd Party Data

- 1st Party - data attributes your company has collected on its customers/prospects
  - form submits, web traffic, purchase history, etc
- 2nd Party - second hand first party data explicitly shared
  - companies & indirect sellers sharing data
- 3rd Party - purchased from a data aggregator, available on the market to anyone

ZERO-PARTY	FIRST-PARTY	SECOND-PARTY	THIRD-PARTY
 Consumer intentionally provides	 Business collects via consumer's interaction with its properties	 Another business's first-party data you use with formal permission.	 Aggregated info strung together from public and non-public sources.
Self-reported			
	Direct relationship, not shared		Indirect, shared
		Individual data, consented, low volume, high accuracy	Aggregated, consent varies, high volume, low accuracy
Product preferences	Purchase history	Fashion brand purchase data sold to a CC company.	Purchases from sites across the web.

# Persona Development

What are they and how are they developed?

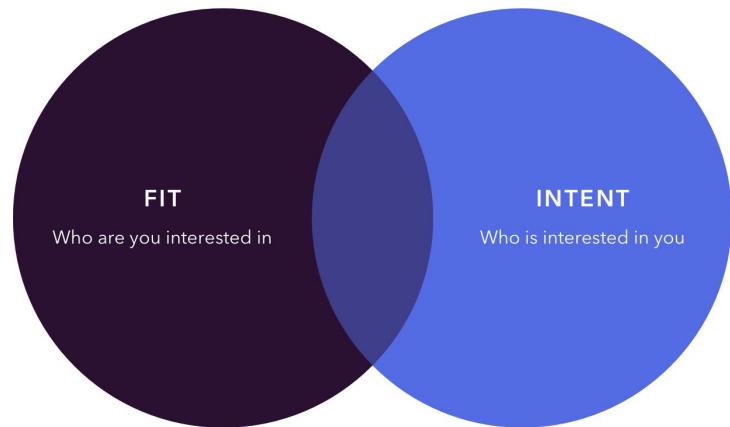
- Fictitious representations of a companies ideal customer
- Most companies have multiple buying personas they market and sell to
- Usually developed by sales and marketing, utilizing audience and market research
  - Best personas are data driven - based on demographics, psychographic, behavior, interests, challenges, etc
  - Opportunity to apply clustering or rules based definitions

What are they used for?

- Creating content that is personalized to needs of each group
- Developing campaigns to organize delivery of that content over time

# Lead Scoring - Finding the right prospects

- Practice of categorizing marketing or sales leads according to their propensity to become customer
  - Consider both volume and conversion rate
- Can be rule based, banded solutions
  - Ex: simple fit & intent models
- Can be algorithmic
  - Based on weighting of multiple factors
- Most effective model is the one that is used & integrated into workflow
- Typical factors:
  - Person's title, company size, previous relationship, lead source (ex: contact sales v whitepaper download)



# Product Recs - Surfacing the right products

- More easily shows customer of what relevant items and helps in cross-sell
- Simple stuff like Trending, Top Sellers, etc
  - By category
  - By location
- Probability based association Rules
  - Co-browse
  - Co-buy
  - Evaluated by:
    - Support, Confidence, Lift
- Collaborative Filtering
  - User-Based: how similar a user is to other users who have already bought
  - Item-Based: how similar an item is to other known items
  - Evaluated by:
    - Cosine similarity, Pearson's correlation, euclidean, etc

# ATTRIBUTION

# Attribution

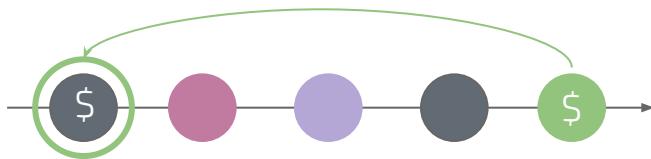
- Refers to allocating credit for an outcome to the events that preceded the outcome
- An 'outcome' could be:
  - Sale / Order
  - Subscription
  - An arrival or leave
- An 'event' could be:
  - Marketing promotion
  - Website landing
  - Website activity
  - Engaging with a representative
- Attribution is important because the relative value produced by the event is often assessed against the cost of providing that experience (Return on Investment)

# The paradox of attribution

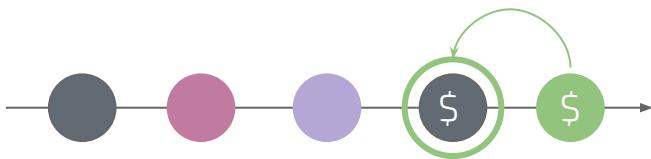
- Attribution is observational (not causal)
  - there are many ways to allocate how much of the outcome is associated with the events in consideration
- Attribution is important to understanding Return on Investment
  - the relative value provided by the event is often assessed against the cost of providing that experience

# Winner Take All - 1st Touch v Last Touch

- First Touch - the first channel within a converting path receives full credit



- Last Touch - the last channel within a converting path receives full credit



# More Nuanced models

- Linear - All events within the converting path equally distribute credit
- Position-based (U & W) - Credit is usually distributed
  - U could be something like 40% first, 40% last, and 20% equally between
  - W could be something like 30% first, 30% last, 30% to some (perceived important) mid funnel event, and 10% equally between
- Time Decay - All events receive credit as a function of proximity to the outcome
- Data Driven - Similar to position based, but weights allocated based on your companies analysis that associates your events to your outcomes

# Marketing Sourced v Marketing Influenced

- For B2B businesses where a sales opportunity is created, attribution modeling can be run on two different scenarios ...
- Marketing Influenced - allocates credit to outcomes where a marketing event has been consumed by anyone in the organization who bought
- Marketing Sourced - allocates credit to outcomes where a marketing event has been consumed by a specific buyer that sales sold directly to

# EXPERIMENTATION

# Quick note

- The remainder will focus on digital experimentation on the marketing side because those use cases are simple and common in practice
- Experimentation is also very popular in product applications, medical trials, etc so these concepts will be equally valuable elsewhere

# How did they go from one iteration to the next?

The image shows two versions of the Barnes & Noble website side-by-side, illustrating the changes in design and layout over time.

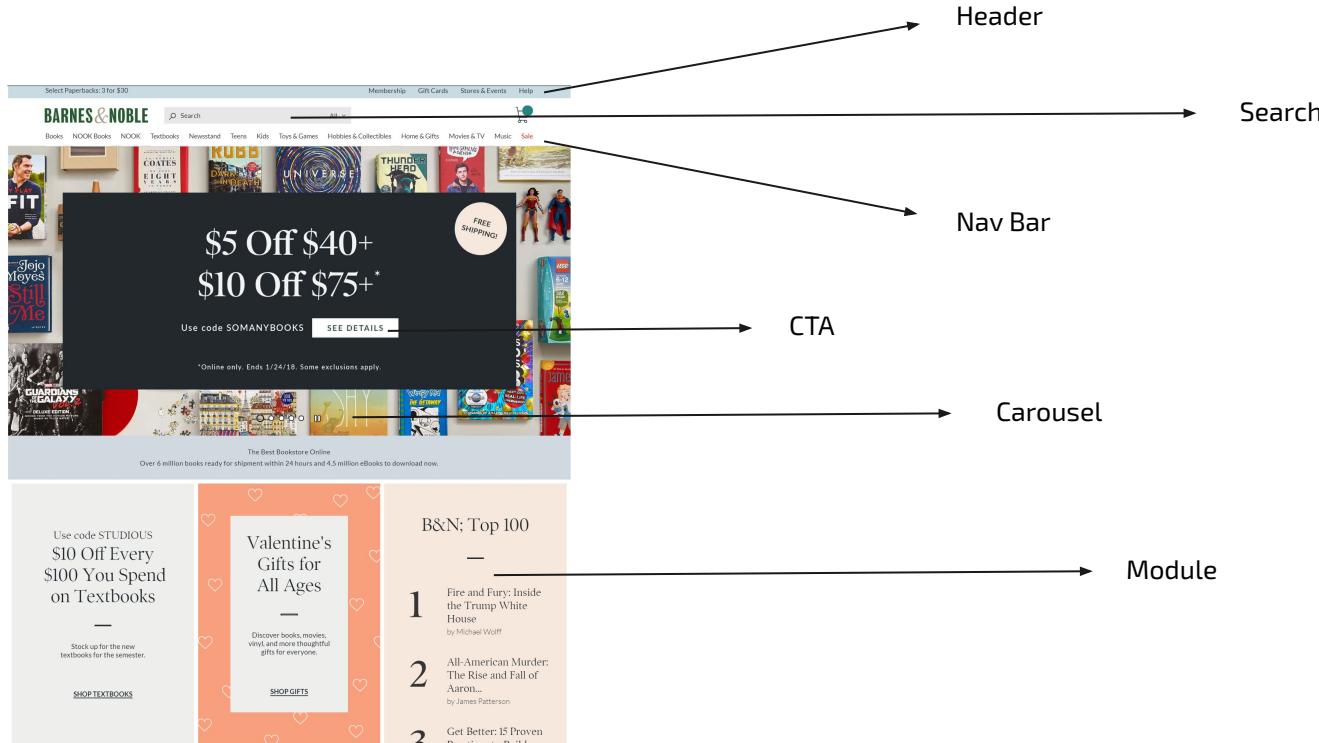
**Left Side (2007 Version):**

- Header:** Features a "FAST & FREE DELIVERY" banner with a "SEE DETAILS" link, a "CART" button with "0 Items", and a "CHECK OUT" button.
- Navigation:** Includes links for "HOME", "AFTER HOLIDAY SALE", "BOOKS", "USED & OUT OF PRINT", "NEW & USED TEXTBOOKS", "DVD", "MUSIC", "PC & VIDEO GAMES", "CHILDREN'S", "TODDLERS & GAMES", "GIFTS & CALENDARS", "GIFT CARDS", and "B&N MEMBER PROGRAM".
- Content:** Promotions for "Members Save 10% Every Day!" and "After Holiday Sale Save Up to 80%". A sidebar highlights "FIND YOUR LOCAL B&N STORE" and "BROWSE B&N'S UNIQUE OFFERINGS".
- Bottom:** A "TOP BOOK SUBJECTS" section and a "TOP 10 BESTSELLERS" list.

**Right Side (2013 Version):**

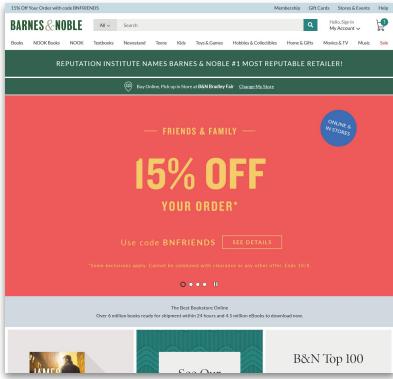
- Header:** Shows a "Sign In" link, "My Account", "Order Status", "My NOOK", "Stores & Events", and "Help". A banner for "After Holiday Clearance Sale: Save 50% or More".
- Navigation:** Includes "All Products" and "Search" buttons, along with links for "Books", "NOOK Books", "NOOK", "Textbooks", "Newsstand", "Trends", "Kids", "Toys & Games", "Home & Gifts", "Movies & TV", "Music", and "Gift Cards".
- Content:** Promotions for "Up to 90% Off Textbooks", "50% Off Bestselling NOOK Books", and "Buy 2, Get a 3rd Free". A large "Book of the Year" banner for "The Heaven & Earth Grocery Store" by James McBride.
- Bottom:** A "Only at Barnes & Noble" section featuring a "Command Authority" book and a "Author of the Year" banner for "The Wager" by David Grann.

# They experimented with each site component...



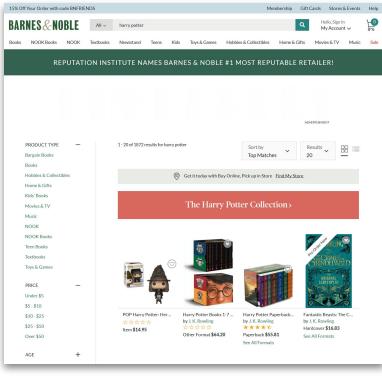
# ... and each page type

Home



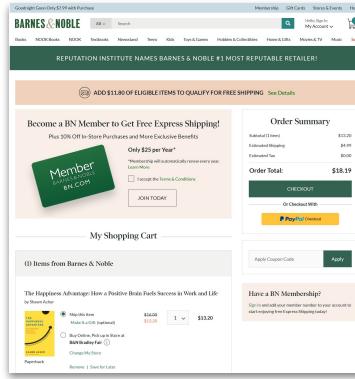
The Barnes & Noble Home page features a prominent red banner at the top offering a 15% discount for Friends & Family. Below the banner, there's a search bar and a navigation menu. A large central image shows a person reading a book, with text overlaying it. The bottom section includes a "Top 100" list and a "Books & More" sidebar.

Search Results



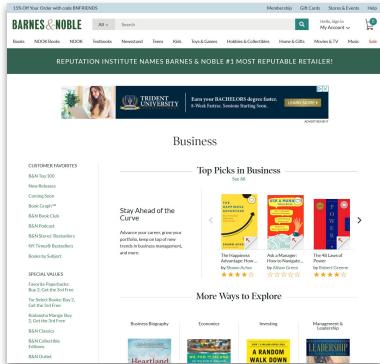
The search results page for "Harry Potter" shows a grid of book covers. A sidebar on the left lists "PRODUCT TYPE" (e.g., Books, DVDs) and "PRICE" (Under \$5, \$5-\$10, \$10-\$20, \$20-\$50, Over \$50). A sidebar on the right offers "Get It Online, Pick Up In Store" options.

Cart



The cart page displays a single item: "The Harry Potter Collection" by J.K. Rowling, priced at \$18.19. It includes a "Member BN.COM" badge, a "JOIN TODAY!" button, and a "CHECKOUT" button.

Category Page



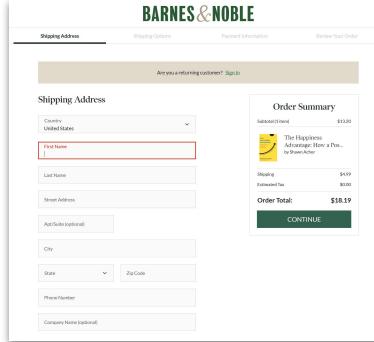
The category page for "Business" shows a sidebar with "CUSTOMER FAVORITES" and "Top Picks in Business". It features a "Stay Ahead of the Curve" section with book covers and descriptions. The main content area displays a grid of business-related books.

Product Page



The product page for "The Fifth Risk" by Michael Lewis includes the book cover, author information, and purchase options like "ADD TO CART" and "Instant Purchase". It also features a "Buy Online, Pick up in Store" section and a "Check Availability at Your Local Store" button.

Checkout



The checkout page shows the shipping address section with fields for "First Name" (filled with "Michael"), "Last Name", "Street Address", "City", "State", "Zip Code", and "Phone Number". The "Order Summary" on the right shows a total of \$18.19.

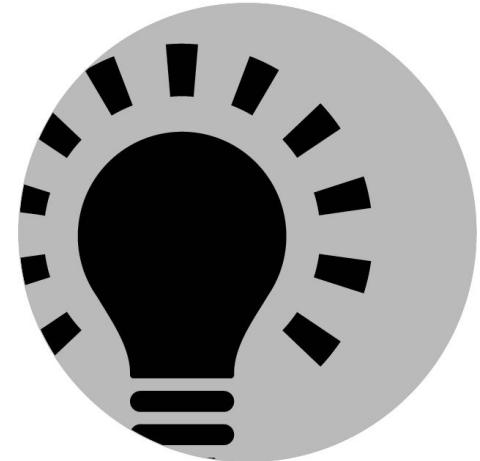
# Each test contains a hypothesis

Hypothesis Statement:

Based on [analytics + research] we believe that [user problem] exists.  
We believe [solution] will result in [this KPI outcome].

Variation Creation:

- Each variation could potentially have its own hypothesis
- Increasing variations increases test duration



# What benefit does experimentation provide?

Experimentation allows us to continually **optimize the customer's experience** resulting in positive impact to the business.

Most importantly, experimentation helps build a better understanding of customers than typical descriptive analysis alone, **by isolating variables and identifying causation**.

In this way, experimentation brings **an iterative and data-driven** approach to business decisions - that ideally replaces slow waterfall projects and/or gut-feel decisions.

Experimentation is practiced across almost all domains, and is particularly powerful in:

- 1) high volume, low conversion B2C marketing applications such as websites and paid
- 2) lower volume, high conversion marketing applications like email, product trials, etc

Note: test will struggle to reach statistical significance in low volume, low conversion scenarios

# Why does the business care?

Experimentation is often the mechanism to **drive incremental revenue** and ensures new experiences benefit the customer.

In this example a 10% lift to conversion rate generates \$300k of annual incremental revenue for the business.

	Version A	Version B
Sessions	10M	10M
Conversion Rate	1.0%	1.1%
Orders	100k	110k
Average Order Value	\$30	\$30
<b>Revenue</b>	<b>\$3.0M</b>	<b>\$3.3M</b>

# How sure are we about that additional \$300k?

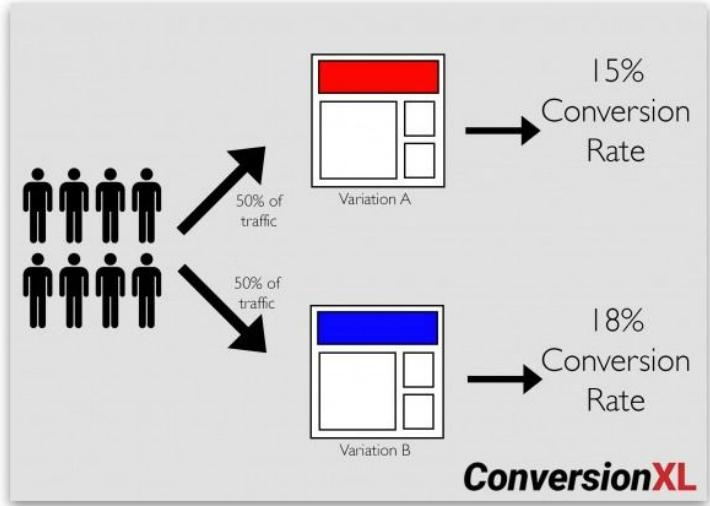
- Experimentation is industry speak for a Randomized Control Trial (RCT)
- Each user is randomly assigned to control or variation
- Random assignment guards against bias and spreads the variance of uncontrollable factors
- Allows us to establish causation when tested against a control during the same time period.
- No experiment will provide a positive return forever, different user expectations and future interactions necessitate continual testing to maximize revenue → yay, job security :)

# How does it work?

- 1) Select of page and site component
- 2) Set target audience
- 3) Holds variation(s) that modifies the control using front-end JavaScript or a WYSIWYG
- 4) Randomizes traffic assignment to the control and each variation
- 5) Summarizes the statistical results of each test

## Marketing Technology vendors

- Adobe Target
- Optimizely
- Google Optimize
- Maxymizer
- Sitespect
- Many more!



# Types of Experiments, aka RCTs

A/B Testing



A single 'factor' is changed  
in each variation

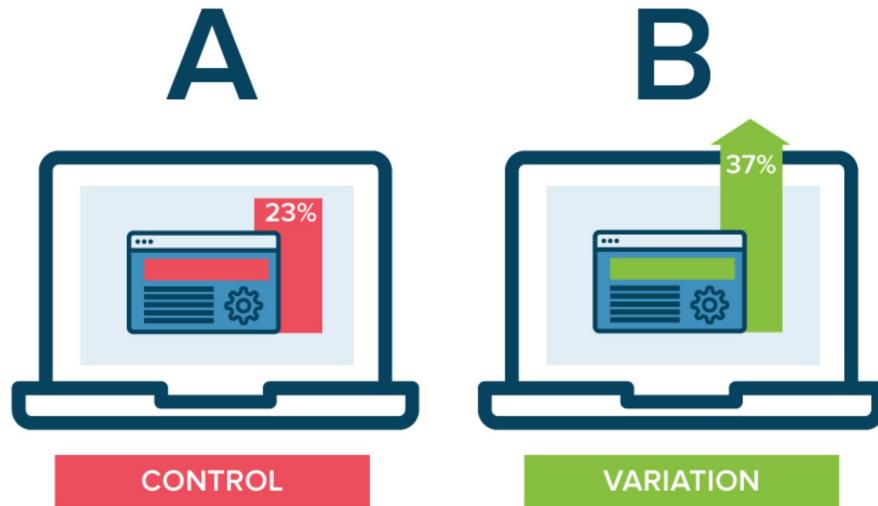
Multivariate Testing (MVT)



Multiple 'factors' are identified, each  
factor with  $\geq 2$  levels; each variation  
changes different permutations of factors  
and factor levels.

# A/B/n Testing

- Also known as “split testing”
- A/B testing refers to the method of randomly dividing site traffic between a control and one or more variations
- Compares the performance of the each variation against the control to inform decisions on which variation is best



# Multivariate Testing

- A method of testing your online campaign using many variations of a design, usually organized by "factor" with each "factor" having two or more "levels"
- You could live test changes to a headline, an image, and new copy ... all at once, in various permutations
- Also known as MVT



	Image	Headline
VERSION 1		+ "ACME WIDGETS"
VERSION 2		+ "ACME WIDGETS"
VERSION 3		+ "THE ONE AND ONLY ACME WIDGETS"
VERSION 4		+ "THE ONE AND ONLY ACME WIDGETS"

# Multivariate Testing - Full v Partial Factorial

## Full

- All permutations of factor levels are formed into variations and all are selected for the test
- So many permutations lead to long durations

Treatments:	
1. A1, B1, C1	✓
2. A1, B1, C2	✓
3. A1, B2, C1	✓
4. A1, B2, C2	✓
5. A2, B1, C1	✓
6. A2, B1, C2	✓
7. A2, B2, C1	✓
8. A2, B2, C2	✓

$2^3$  variations tested = 8

## Partial

- Only a certain number (percentage) of permutations are selected for the test
- Testing tools leverage an algorithm to choose the best variations of the different elements based on which combinations provide sufficient information to determine element effects
- Design Resolution is the measure of a tradeoff between speed and certainty, higher resolution provides more certainty at the cost of speed

Treatments:	
1. A1, B1, C1	
2. A1, B1, C2	
3. A1, B2, C1	
4. A1, B2, C2	
5. A2, B1, C1	
6. A2, B1, C2	
7. A2, B2, C1	
8. A2, B2, C2	

Total variations tested = 4

# Setting Sample Sizes / Stopping Rules

Stopping Rules (Frequentist, Fixed Time Horizon)

- A stopping rule is the criteria at which the test should be stopped and results collected
- Establishing these criteria will ensure your tests are not concluded with bias and that the results will be repeatable
- Test margins will fluctuate greatly over the course of the test duration but predetermined sample sizes are calculated to ensure that the test will approach the true difference in populations instead of "p-hacking" to find a winner

*Caution: peeking into the test results with the intention of stopping a test early or extend the duration can inflate Type 1 errors*



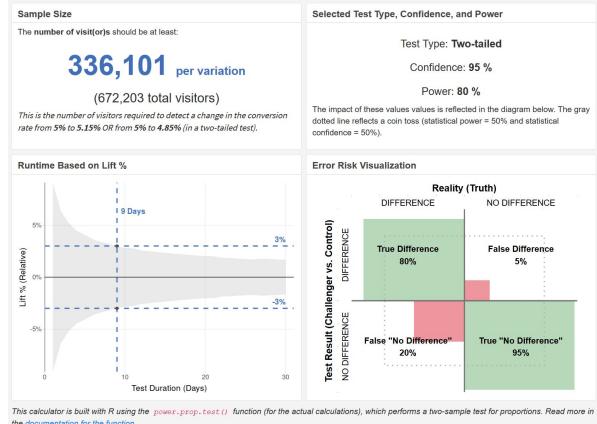
# Design of Experiments - Frequentist

## Fixed Time Horizon Inputs

- Confidence - controls Type I errors
- Power - controls Type II errors
- Effect Size - the min. detectable lift
- Primary KPI info - response rate for binary,
- Daily Traffic - to convert to duration
- Number of Variations - to calc duration
- Tails - one or two tailed statistical test

Ex:

.05 Alpha  
.80 Beta  
3% MDE  
5% Order Rate  
75k Daily Visitors  
2 total  
2 tails



# Possible Outcomes and Business Cost

- Perform significance testing on the Primary KPI
- Include observed values, statistical significance, and lift percentage w/ confidence intervals

Truth (can never know)	Decision of 'No difference'	Decision of 'Difference'
No difference	True Negative (1-alpha) <span style="color: green;">★</span>	False Positive (alpha) <span style="color: red;">✖</span>
Difference	False Negative (beta) <span style="color: red;">✖</span>	True Positive (1-beta) <span style="color: green;">★</span>

- False Positive - we recommend change that actually won't make a difference
  - We think +\$ when it's actually \$0 or -\$
  - Generally thought to be worse, and thus tighter control (5%)
- False Negative - we fail to recommend a change that would have made a difference
  - We think it's \$0 or -\$ when it's actually +\$
  - Generally of less concern, as thus less control (20%), but specific to your experiment
- Always consider the costs of each error when assigning alpha and beta

# More in the spring on ...

**1** Targeting

**2** Attribution

**3** Experimentation



---

# Questions?

**Reid Bryant**  
[reidbryant@gmail.com](mailto:reidbryant@gmail.com)

