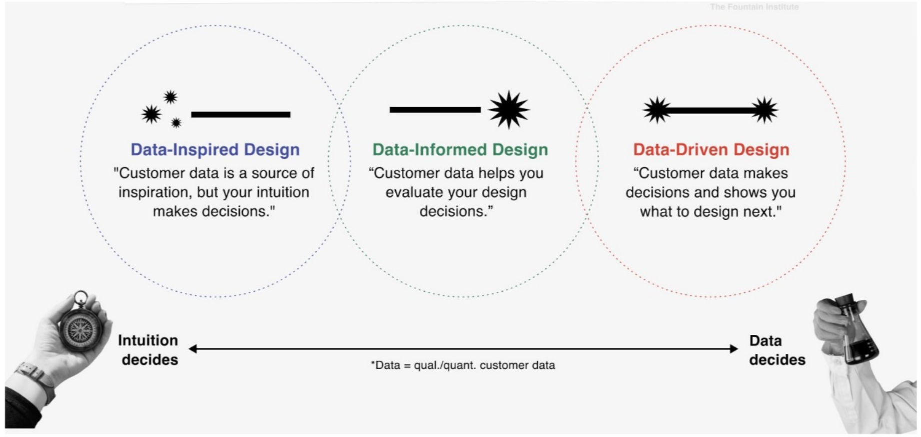
The Influence of Data on Design

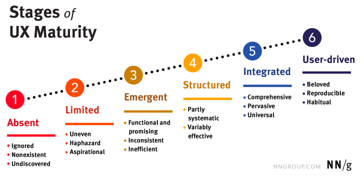

The Fountain Institute
Data-Inspired Design
"Customer data is a source of inspiration, but your intuition makes decisions."
Data-Informed Design
"Customer data helps you evaluate your design decisions."
Data-Driven Design
"Customer data makes decisions and shows you what to design next."
Intuition decides
Data decides
*Data = qual./quant. customer data

Source

Data Driven Design

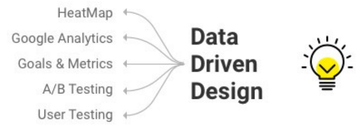
Pros:

- Human-centered since they let the customer directly influence what gets built.
- Customers decide through their behavior in various research and testing activities.
- Intuition can still play a role in interpreting the data, but you should submit your decisions to the results of customer behavior.



Cons:

- The iterative nature of letting data make decisions can mean you sometimes miss out on the big-picture view and trends that also have a part to play in design.
- Sometimes customers don't show their ideal behaviors in their actions. When we base our decisions on behaviors, we can end up with products that provide what customers want but don't need.



What is A/B Testing?

- **Compare two or more variations** of an experience to see which one performs the best for a single / multiple objective measure(s)
- **Test on a small fraction** of the entire user population (in order to minimize the side-effect of contaminating users with experimental designs)
- **Causal analysis** of the variations to user behavior; to understand how your design will affect user experience if launched

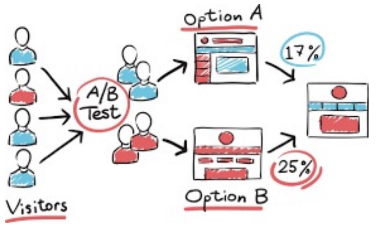
✓ A
✕ B
VS.
25%
Conversion
5%
Conversion

Why do we study A/B testing?

- Empirical evidences are more persuasive than design authority
- Testing improves your understanding of the users
- Data-driven decision making can become a part of the product

Ronald Fisher
It's me again. Back in 1920s, I created **randomized controlled experiment**, the core concept of A/B testing. I wish I could run A/B testing for the smoking-cancer causality.

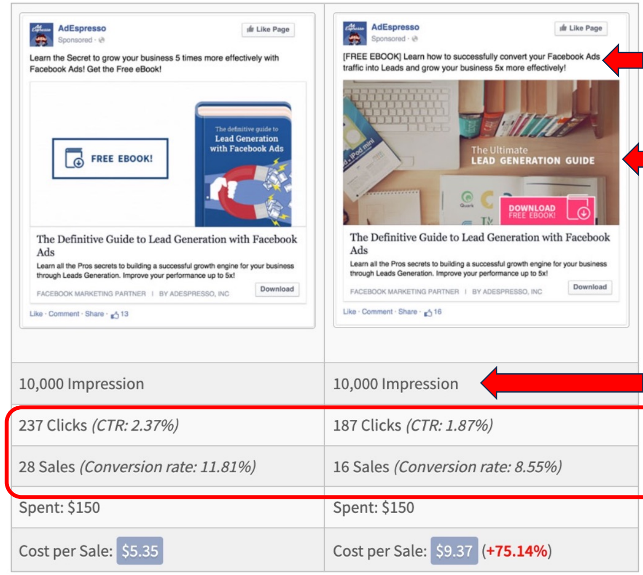
1950s- clinical trials adopted A/B testing
1960s- markerters evaluate whether postcard or letter would gather more customers?
1990s- online experiments
2020s- AI-driven, continuous and highly-automated experiments



https://www.diggintravel.com/airline-ab-testing/

Examples of A/B Testing

Advertisement lead (title and thumbnail) on Facebook



AdEspresso
Sponsored
Like Page
FREE EBOOK!
The definitive guide to Lead Generation with Facebook Ads
The Ultimate LEAD GENERATION GUIDE
DOWNLOAD FREE EBOOK!
Download
FACEBOOK MARKETING PARTNER | BY ADESPRESSO, INC
Like · Comment · Share · 13
Like · Comment · Share · 16

Learn the Secret to grow your business 5 times more effectively with Facebook Ads! Get the Free eBook!

The Definitive Guide to Lead Generation with Facebook Ads

Learn all the Pros secrets to building a successful growth engine for your business through Leads Generation. Improve your performance up to 5x!

[FREE EBOOK] Learn how to successfully convert your Facebook Ads traffic into Leads and grow your business 5x more effectively!

The Definitive Guide to Lead Generation with Facebook Ads

Learn all the Pros secrets to building a successful growth engine for your business through Leads Generation. Improve your performance up to 5x!

10,000 Impression	10,000 Impression
237 Clicks *(CTR: 2.37%)*	187 Clicks *(CTR: 1.87%)*
28 Sales *(Conversion rate: 11.81%)*	16 Sales *(Conversion rate: 8.55%)*
Spent: $150	Spent: $150
Cost per Sale: $5.35	Cost per Sale: $9.37 (+75.14%)

Titles

Thumbnail images

Same # observations

Two stage results

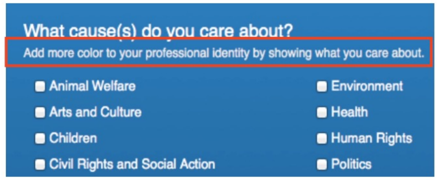
1. Click-through rate
2. Conversion rate

Test 1 has almost double ROI per dollar

Examples of A/B Testing

- (At LinkedIn) we have hundreds of experiments running in parallel

LinkedIn's Profile Edit



What cause(s) do you care about?

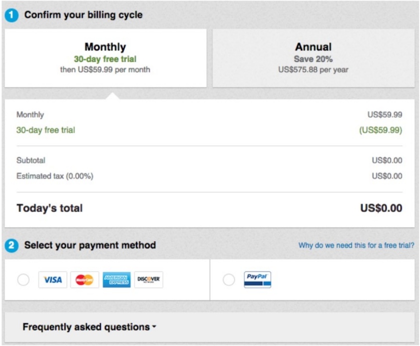
Add more color to your professional identity by showing what you care about.

□ Animal Welfare
□ Arts and Culture
□ Children
□ Civil Rights and Social Action
□ Environment
□ Health
□ Human Rights
□ Politics

The **movitational text** boosted up 14% of the profile edits

https://content.linkedin.com/content/dam/engineering/site- assets/pdfs/ABTestingSocialNetwork_share.pdf

LinkedIn's Premium Subscription Payment flow



1 Confirm your billing cycle

Monthly 30-day free trial then US$59.99 per month	Annual Save 20% US$575.88 per year

Monthly	US$59.99
30-day free trial	(US$59.99)
Subtotal	US$0.00
Estimated tax (0.00%)	US$0.00
Today's total	**US$0.00**

2 Select your payment method — Why do we need this for a free trial?

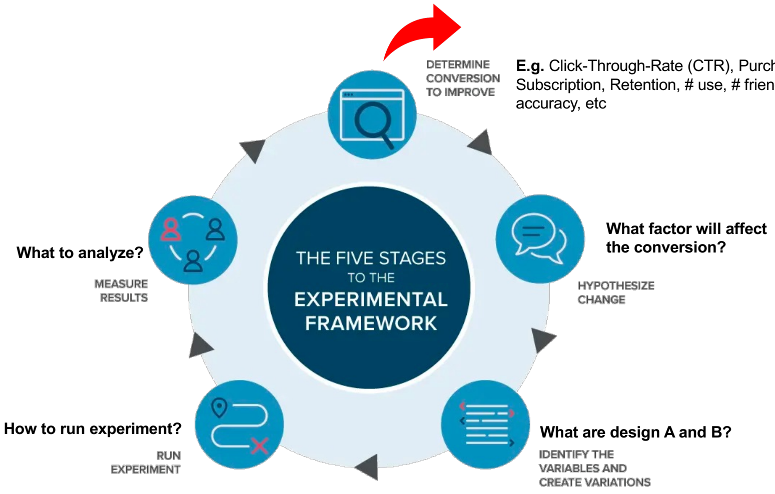
○ VISA, MasterCard, American Express, Discover ○ PayPal

Frequently asked questions ▾

The new payment flow merges two pages into one; and provides FAQs → 30% less refunding orders and 10% more free trial orders

A/B testing process

Interative cycles



E.g. Click-Through-Rate (CTR), Purchase, Subscription, Retention, # use, # friends, # likes, accuracy, etc

What factor will affect the conversion?

What are design A and B?

How to run experiment?

What to analyze?

Source

A/B testing process

Step 1. Set the Goal

- A/B testing usually focuses on a single (or couple) quantifiable measures of user behavior. While the below examples are commonly used, you may create custom measures.

Examples of Common Measures to optimize

Click-Through Rates (CTR)
- Ratio of users who click on a specific link
- A KPI (Key Performance Indicator) of an online banner / email advertising campaign

Conversion Rates
- Ratio of users who did the desired action such as product purchase, subscription, signing up, sharing with others, and so on.

Retention Rates
- Ratio of users who keep using the service within a specific period of time (and other conditions such as minimum # of usage, etc)
- An inverse of retention rates would be unsubscription rates

Task Completion Rate
- Ratio of users who complete the given task (e.g. form fill-in, survey, tutorial)

Error Rate
- Ratio of users who see any error message or make mistakes
- Unlike the others, we optimize a system to lower the error rate

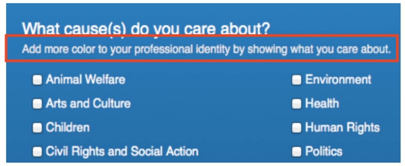
Satisfaction Rate
- Ratio of users who give positive answers via a survey or other communication channels

Cost
- Amount of money spent for specific returns (e.g. AD banner for clicks)
- businesses can find the option that offers better returns and get rid of the process that offers lower returns

Step 2. Construct Hypothesis

- Clarify the hypothesis to be tested.
- A hypothesis is usually a combination of cause(s) and the effect(s) that describes the current situation.
- More abstract than design variations and measurements.

E.g. LinkedIn Profile Edit Case

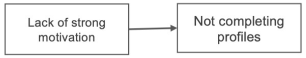


What cause(s) do you care about?

Add more color to your professional identity by showing what you care about.

- □ Animal Welfare
- □ Arts and Culture
- □ Children
- □ Civil Rights and Social Action
- □ Environment
- □ Health
- □ Human Rights
- □ Politics

H: Users are not completing their profiles because they do not have strong motivation



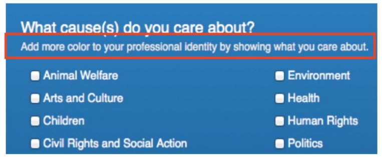
Lack of strong motivation → Not completing profiles

Note that the hypothesis is more abstract than the design variations and measurements in your mind. For instance, "lack of motivation" is not a concrete design yet. Adding a motivational sentence is one of many design choices. Also "not completing profiles" is not a specific measure yet. "# users completed their profiles" is one of many possible measurements.

Step 3. Create Design Variations

- It's time to create design variations to test your hypothesis. Variations may include a control and a single or multiple experimental condition(s). Variations must be evaluated via the measurement.

E.g. LinkedIn Profile Edit Case



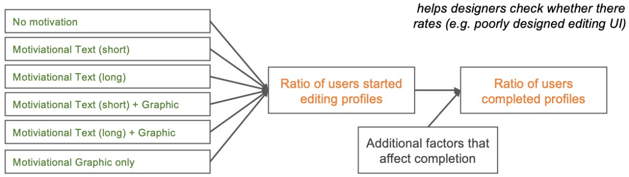
What cause(s) do you care about?

Add more color to your professional identity by showing what you care about.

- □ Animal Welfare
- □ Arts and Culture
- □ Children
- □ Civil Rights and Social Action
- □ Environment
- □ Health
- □ Human Rights
- □ Politics

Control. Showing no motivational text
Experimental. Showing "Add more color to …"
To get further insights, designers may create more than one experimental conditions. For instance, there could be variations of motivational text content, graphical and/or textual motivation, and so on.

Measurement. Ratio of users completed profiles
To get more detailed insights, designers are advised to create additional measurements. In this example, measuring "ratio of users start editing their profiles helps designers check whether there are additional factors affecting the completion rates (e.g. poorly designed editing UI)



No motivation

Motiviational Text (short)

Motiviational Text (long)

Motiviational Text (short) + Graphic

Motiviational Text (long) + Graphic

Motiviational Graphic only

Ratio of users started editing profiles

Ratio of users completed profiles

Additional factors that affect completion

Step 4. Running Experiment

- A/B testing is usually conducted with a small fraction of randomly-selected users. There is no fixed sample size or percentage, but you should consider the following rule-of-thumbs.

1. <u>Equally allocate users to design variations (if possible)</u>

Don't assign majority of users to specific conditions. It will slow down the experiment by increasing the minimum sample sizes to get statistical significance.

2. <u>Get sufficient samples to get statistical significance of</u>

<u>at least 95% (p<0.05)</u>

300-400 samples per variation is usually considered enough. If you cannot reach the 95% (p=0.05) with 300-400 samples * # variations, it is very unlikely that the variations have strong impact on the measured goal.

3. <u>Run the experiment for multiple behavioral cycles</u>

Lots of user behaviors tend to have weekly patterns (e.g. on Monday, people tend to be quite busy). To minimize the risk of weekly biases, run the experiment for at least one week (preferably two weeks or a month). However, it depends on what the task is and who the users are.

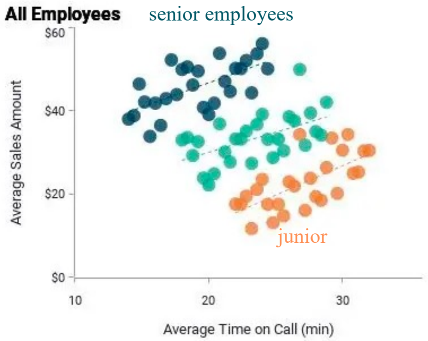
4. <u>Pay attention to external factors</u>

If it is Christmas or Valentine's day, an online marketing campaign might be highly affected by the external factor. While it's impossible to get a perfectly clean environment, consider external factors in your interpretation.

Step 5. Analysis

- Once you get the numbers, the first thing to do is EDA focusing on the validity of the experiment. Then you find the winner by comparing the goal measurements. Finally, you test the hypotheses that you set beforehand.

1. **Perform EDA for validating the**
   - Did you get the planned # samples per
   - Is every data point complete and accu
   - Is there any unexpected bias or confou



2. **Finding the winning variation**
   - Which variation achieved the highest
   - Drill down to smaller subgroups of pe also outperform for more specific user *<u>Simpson's paradox</u>*.
   - If you set multiple measurements, try to characterize pros and cons of each variation. For instance, a variation that achieved the highest conversion ratio may have a high ratio of refunding too.

Step 5. Analysis

3. Test Hypothesis

- After finding a winner you should investigate further – in order to gain generalizable design knowledge, and to get ideas for follow-up tests.
- E.g. (For the LinkedIn Payment Flow case) Were users not ordering the premium package because of the overly complicated payment process?
- Yes, by applying both of the 1-page flow and the FAQ link, we significantly ($p<0.03$) increased # free trial orders and decreased # refunding orders.

4. Plan for the next A/B test

- Results from A/B tests cannot guarantee reproducibility (i.e. you may not get the same result the next time), and it is likely that the first test will give more questions than concrete answers. (e.g. “Why did the winner perform so well?”)
- Repeat the same study. You may refine the test plan though (e.g. smaller # variations; more measurements; etc).
- Actual goals of A/B testing is not just finding winners but also developing a platform for continuously gaining insights from the real users.

Summary

- **A/B testing is a special kind of quantitative experiments**
  - where “A” representing the old control, and “B” representing a single / multiple experimental changes
  - where the goal is to maximize single/multiple measurements of user behavior
  - A/B testing is most useful when combined with qualitative methods (e.g. observation, survey, interview)
- **A/B testing has multiple goals as listed below**
  - Finding a winner among many design variations
  - Testing hypotheses of cause (design) and effect (user behavior)
  - Gaining generalizable understanding of the task and the user population
  - Developing a reusable platform for continuously running a series of A/B tests
- **A/B testing consists of five big steps**
  1. Choose a single (or couple) quantifiable measure(s) of user behavior
  2. Construct hypotheses
  3. Create design variations
  4. Run the experiment
  5. Analyze the results

Summary

- A/B testing is not silver bullet. It has a lot of limitations and pitfalls.

When A/B testing is not worth	What’s the problem? How can we make it worth again?
You don’t have meaningful traffic (i.e. # users)	• Without meaningful traffic you won’t be able to tell anything about statistical significance. • Reduce # variations • Wait until you have a big enough population • Construct a hypothesis that you can indirectly test on a similar platform (e.g. running A/B test of your streaming platform on YouTube users), crowdsourcing platform (e.g. Amazon Mechanical Turk), or as a lab experiment
You don’t have enough resources for running A/B test.	• You will regret, “We should’ve finished our design first.” • Wait until you have enough time • Consider using existing tools for A/B testing. They are not mature yet, but still useful for marketing.
You don’t have an informed hypothesis.	• Gather more information and perform EDA • Treat A/B test like real science. Learn from lectures, case studies, and tutorials. Ask psychologists and data scientists who have conducted randomized controlled experiments what hypothesis you can test for the given situation

Summary

- A/B testing is not silver bullet. It has a lot of limitations and pitfalls.

When A/B testing is not worth	What’s the problem? How can we make it worth again?
You have an obvious winner.	• It’s waste of resources to find an obvious winner. • Diversify the winner into multiple variations as the LinkedIn purchase flow case. You will learn a lot more than “who’s the winner”.
Your design variations have serious side-effects on user’s mindset.	• Users may get confused by inconsistent usages, and even lose their trust on your service. For instance, if a company assigns different prices randomly assigned for each customer, customers may become angry. • Test with a small fraction of users, or on a closed-beta system. Some companies test price variations not on their official market but through a 3rd party discount platforms (e.g. limited-time deal)
Your variations are not suitable for fair comparison	• A common scenario is UI redesign where users are already familiar with the original design. Your new design variation is unfairly undervalued by its learning curve. • Construct a hypothesis that does not involve familiarity. I.e. make all the variations familiar / unfamiliar. • Run A/B test on newbies only, or even create a new service.