

Marketing & Data Analysis

Experiments

Santiago Alonso-Díaz

Tecnológico de Monterrey
EGADE, Business School

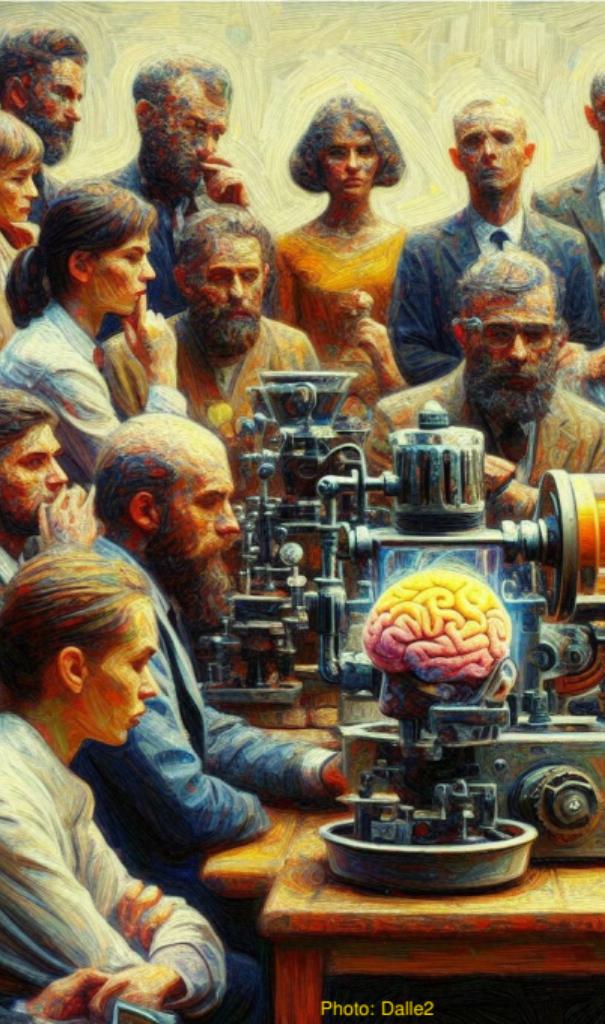
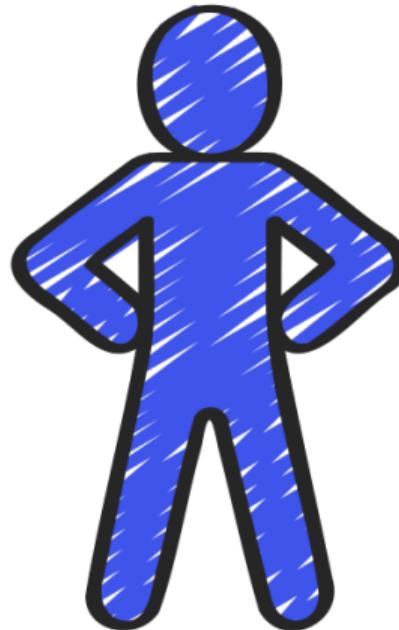


Photo: Dalle2

Business Problem

We want azulito to engage more, that is to enjoy and stay more in my streaming service.



Business Problem

Potential plans. Which one would you pick (only one):

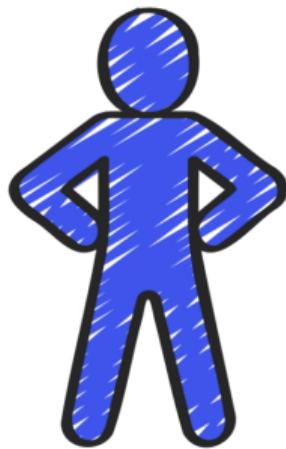
- Economic (reduce prices)
- Marketing (more offers)
- Publicity (enticing ads)
- Content (more content)
- Cognitive (reduce cognitive load)

Business Problem

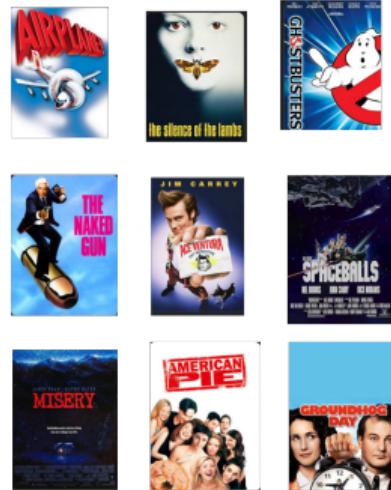
Let's test the cognitive-load theory. With an A/B test.

- Design?
- Outcome measure?
- Analyses plan?
- Follow-up?

Business Problem



A
Choice Overload



B
No Choice Overload



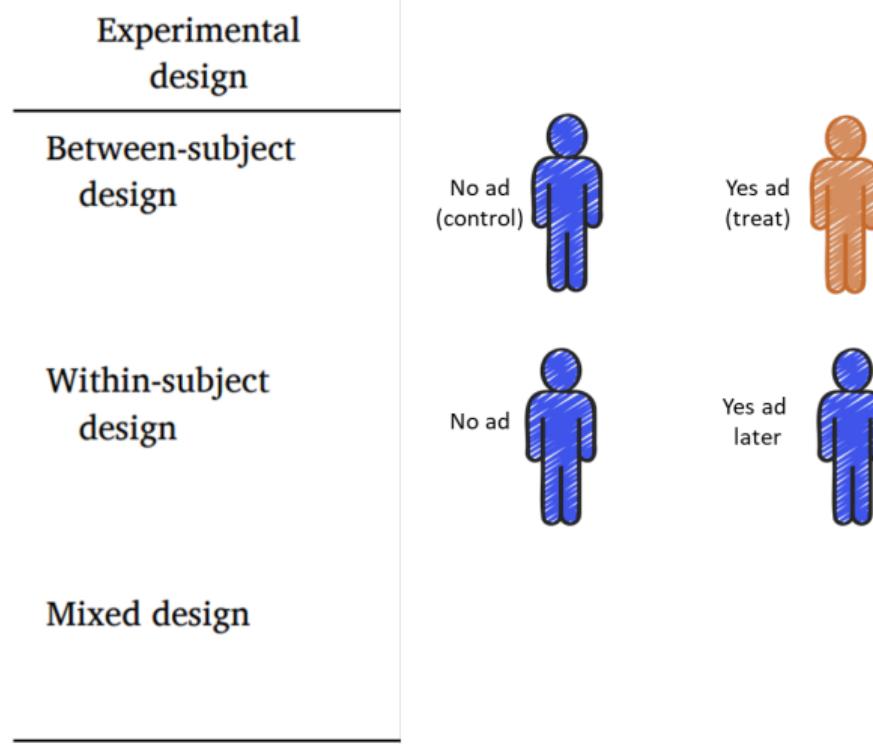
Figure: Engagement predictions? A or B more?

Typology of Experiments

Internal Validity	Type of experiment	
	Conventional laboratory experiment	Wine tasting and price levels in an fMRI machine
	Experiment with increased behavioral realism	Auctions to buy wines at the end of an fMRI session
	Field experiment	Wine tasting and price levels in a winery
	Quasi experiments/Natural data	Change in legislation and wine consumption
	Conjoint analysis	Survey with different wine characteristics and prices

Figure: Viglia et al., 2021

Typology of Experiments



Typology of Experiments

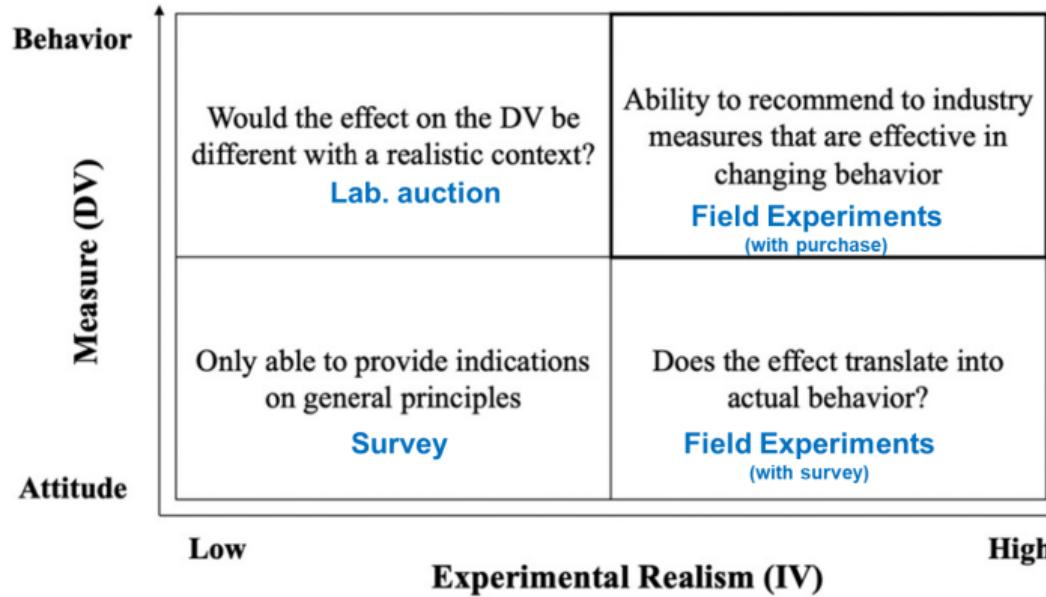


Figure: Viglia et al., 2021

Table of contents

1 Overview

2 A/B Testing

- Examples
- Implementation (Siroker & Koomen, 2015)
- Business Requirements (BR) (Siroker & Koomen, 2015)
- Challenges

3 Discrete Choice Experiments

- Examples
- Theory
- Implementation
- Challenges

4 References

A/B Testing

Prof. Zoom toilet surveys

Which A/B option would work best? Results in zoom

Why? Emotion analysis (e.g. ask chatGPT for emotion label of a paragraph)

Examples of A/B testing

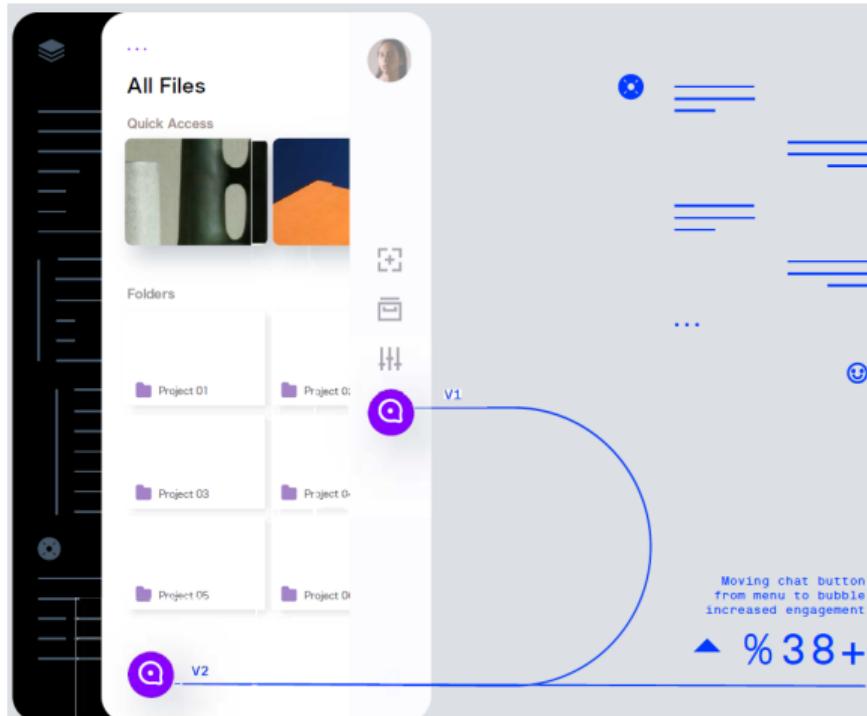


Figure: Position matters? Source: ebook Optimizely (2022)

Examples of A/B testing

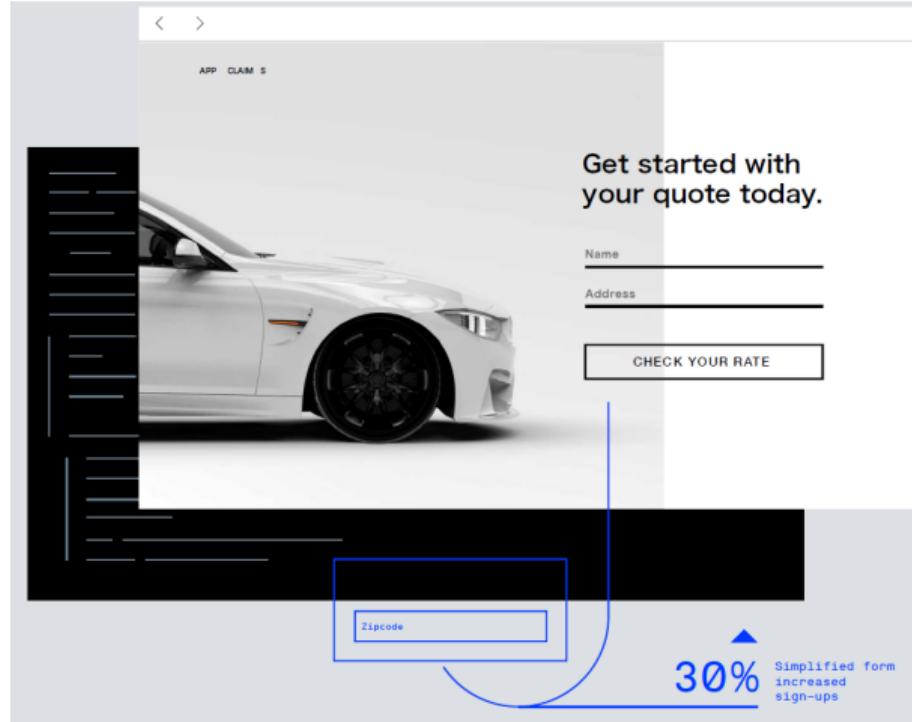


Figure: Simplicity matters? Source: ebook Optimizely (2022)

Examples of A/B testing

The screenshot shows a fashion website layout. On the left, there is a large image of a woman wearing a white crop top and a pink floral lace midi skirt. To the right of the image, the product details are listed:

pink floral lace midi skirt
£10.00
size
[add to bag](#) [add to wishlist](#) [view full details](#)

white basic satin bralet
£12.00
size
[add to bag](#) [add to wishlist](#) [view full details](#)

Below these products is a promotional banner with a light green background and a small unicorn icon:

unicorn class premier delivery is here,
get unlimited next day delivery for a whole year,
for £9.99

next day delivery has never looked so fly

A blue curved arrow points from the text "Testing a new premium service, 'Unicorn Delivery' with a small group of customers" at the bottom of the banner to the "next day delivery has never looked so fly" text.

Testing a new premium service, 'Unicorn Delivery' with a small group of customers

Figure: Preferential treatment matters? Source: ebook Optimizely (2022)

Examples of A/B testing

Donate to charity: water

100% of your money will fund clean water projects for people in need.

GIVE MONTHLY
\$30 can bring clean water to one person every month.

\$ 60 USD/MONTH

Give in honor of someone

Give by Link to Bank

Give by PayPal

GIVE ONCE
Help fund water projects for communities around the world.

\$ 60 USD

Give in honor of someone

Give by Give by

Give by check or stock

Sponsor an entire project for \$10,000+

Testing of donation forms increased average donation size

Figure: Not just for-profit marketing. Source: ebook Optimizely (2022)

Examples of A/B testing

The screenshot shows the Saks Fifth Avenue checkout process. It includes fields for Shipping Address (Last Name: Smith, Street: 123 Main St, Apt 1A, Email: customer@example.com), Delivery & Gift Options (Delivery Method: Standard - 3-5 business days), Payment Method (Credit Card: Visa **** 2880), and a Shopping Bag containing 2 items. At the bottom is a large "PLACE ORDER" button.

Control

Place Order CTA in line with Order summary

The screenshot shows the Saks Fifth Avenue checkout process. It includes fields for Shipping Address (Last Name: Smith, Street: 123 Main St, Apt 1A, Email: customer@example.com), Delivery Method (Delivery Method: Standard - 3-5 business days), Payment Method (Credit Card: Visa **** 2880), and a Shopping Bag containing 2 items. Below these steps is a "PLACE ORDER" button.

Variation B

Place Order CTA in line with steps

The screenshot shows the Saks Fifth Avenue checkout process. It includes fields for Shipping Address (Last Name: Smith, Street: 123 Main St, Apt 1A, Email: customer@example.com), Payment Method (Credit Card: Visa **** 2880), and a Shopping Bag containing 2 items. Below these steps is a "PLACE ORDER" button. A note at the bottom states: "Did you know? Spend \$600 today for FREE 1-day Shipping & FREE Returns, delivered to Westgate CT! Learn More | See R&B".

Variation C

Dual Place Order CTA

Figure: A/B does not mean only two options. Source: ebook Optimizely (2022)

Examples of A/B testing

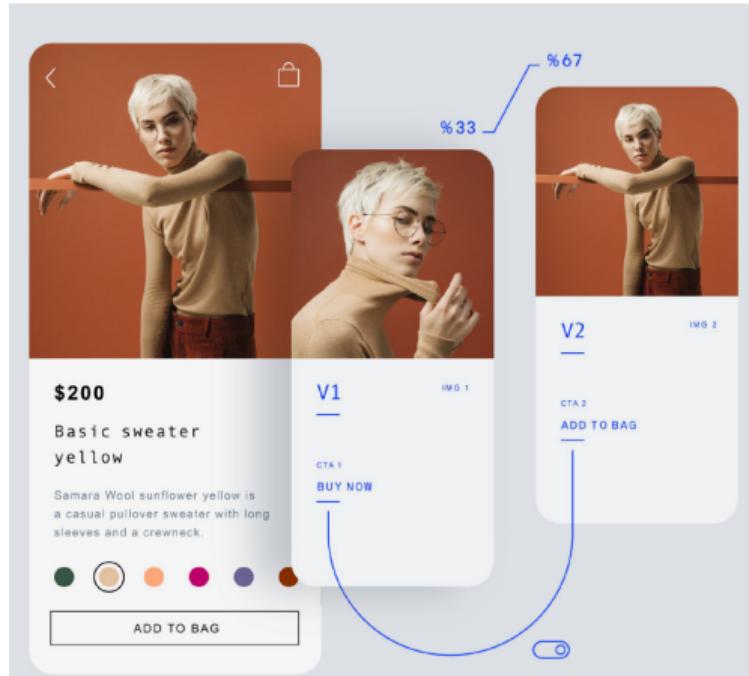


Figure: V1 and V2 differ and the option to buy or add. Is the double variation good or bad or irrelevant?. Source: ebook Optimizely (2022)

Some facts

Table 6

Application domain ×A/B target.

<i>Application domain</i>	<i>A/B target</i>	Algorithm	Visual elements	Workflow /process	Back-end	New app. func.	Other
Web	17	6	8	1	3	0	
Search engine	17	16	3	7	2	0	
E-commerce	10	2	7	0	0	1	
Interaction	5	6	2	2	1	0	
Finances	7	2	4	0	1	0	
Transportation	2	0	0	1	1	0	
Other	2	1	3	0	0	2	

Figure: In the online world, most applications test algorithms or visual elements (Quin et al., 2024)

Some facts

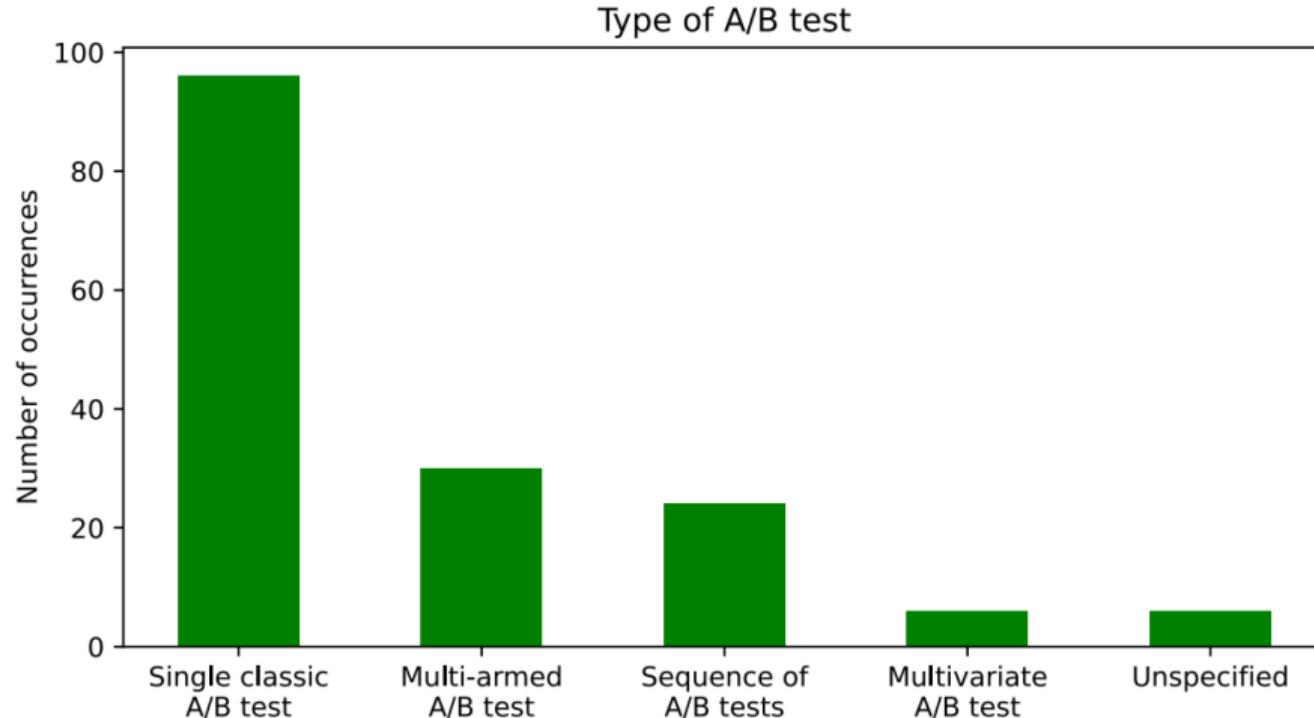


Figure: Most use simply A/B tests (Quin et al., 2024)

Some facts

Table 7
Identified A/B metrics.

A/B metric	Number of occurrences
Engagement metrics	226
Click metrics	83
Monetary metrics	64
Performance metrics	50
Negative metrics	34
View metrics	21
Feedback metrics	17

Figure: Some tested metrics (Quin et al., 2024)

Some facts

Table 8

Statistical methods employed during A/B testing.

Statistical methods employed	Number of occurrences
Hypothesis - equality	57
Hypothesis - equality (concrete method unspecified)	39
Bootstrapping	11
Hypothesis - inference	8
Goodness of fit	8
Estimator	8
Correction method	7
Hypothesis - independence	5
Regression method	2

Figure: Popular data analyses approaches (Quin et al., 2024)

Some facts

Table 12

Data collected for the A/B tests.

Data collected	Number of occurrences
Product/system data	49
User-centric data	26
Spatial-temporal data	20
Secondary data	6

[Figure](#): Data types (Quin et al., 2024)

Implementation

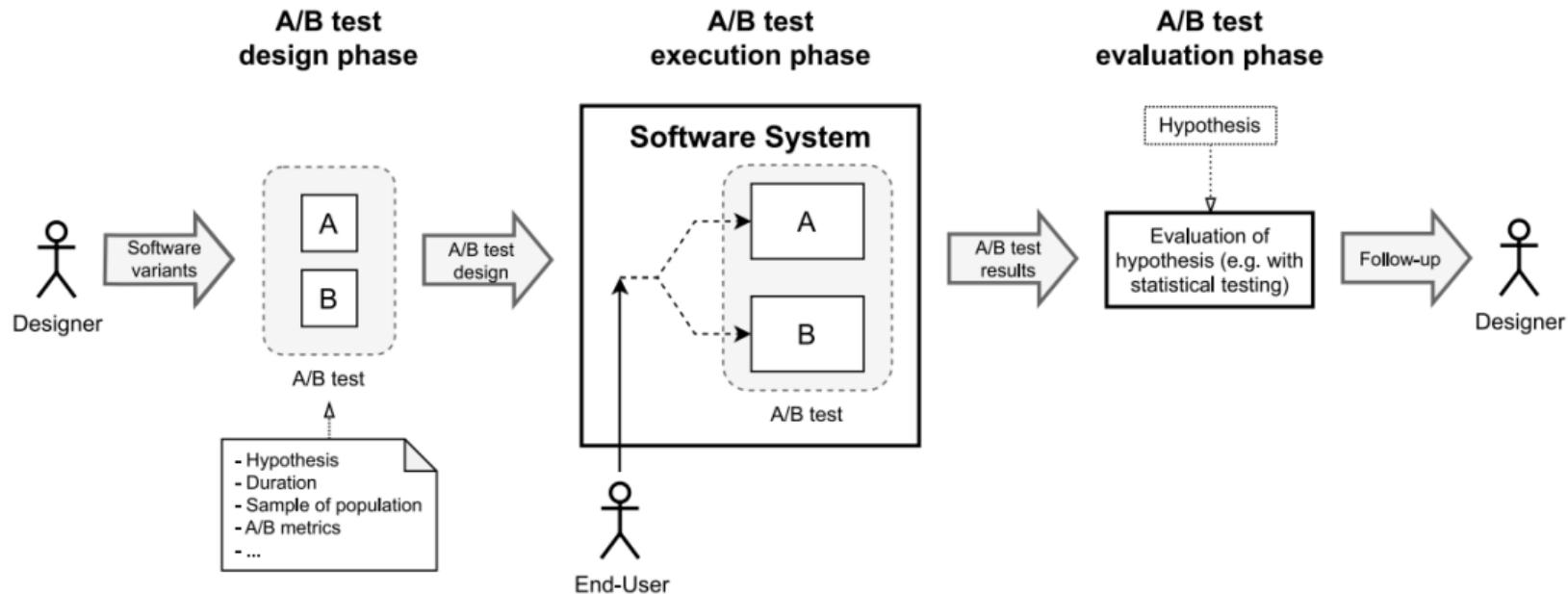


Figure: Quin et al., 2024

Implementation

- Build proprietary A/B testing tools
- Buy commercial A/B testing tools. Or open source.
- Hire consultants to run and analyze A/B testing.

Table 20
Environments and tools used for A/B testing.

Environment	Number of occurrences
In-house experimentation system	21
Research tool or prototype	13
Commercial A/B testing tool	10
Commercial non A/B testing tool	7
User survey	1

Figure: Quin et al., 2024

Implementation

Build

Costly, it requires engineering and data science team.

Perhaps necessary, due to proprietary internal systems.

Think, for instance:

- Amazon
- Google
- Mexican government
- Hospital

Implementation

Buy

There are companies and open source alternatives that offer A/B testing (Google Optimize, AB Tasty, Optimizely, Adobe Target, Oracle Maxymiser).

Advantages:

- Pay as you need
- Expertise and customer service
- Community

Implementation

Hire

Useful if there is a weak engineering and data science know how. Also, if A/B testing is too sporadic.

Think, for instance:

- New product launch
- Change in logistics but expect to last years

BR: Culture

- Cure the HiPPO syndrome: highest paid person opinion (rather than data)
- Fight risk aversion (e.g. by presenting a clear win case.)
- Make evidence-based decisions (frequentist and Bayesian)
- Execute fewer and shorter meetings (around the evidence).
- IMPORTANT: Build a testing culture

BR: Stakeholders

Some are concept designer, experiment architect, setup technician, management

- Do not anger key people (e.g. by rejecting their old ideas) .
- Keep it simple or go slow with most areas (specially technical ones e.g. web keepers) .
- Communicate results constantly to stakeholders (e.g. commercial VPs) .
- Stakeholders could get tired of testing. Be creative (e.g. IGN gamified testing by betting which option wins, many times people lost, data surprises you) .

BR: Team

- Centralized (Testing department e.g. Staples)
- Decentralized (Each department has a testing team e.g. Netflix)

BR: Scientific Logic

- Identify problem
- Come up with an hypothesis
- Test the hypothesis with A/B
- Commit to a sample size and an analyses.
- Pilot. Do A/A tests first to test the system e.g. the same page to both groups.
Results should be the same.
- Run the A/B test.
- Analyze (with the committed analyses + exploratory ones).
- Accumulate and transmit knowledge.
- REPEAT

Challenges

- Small sample sizes
- Naive scientific practices

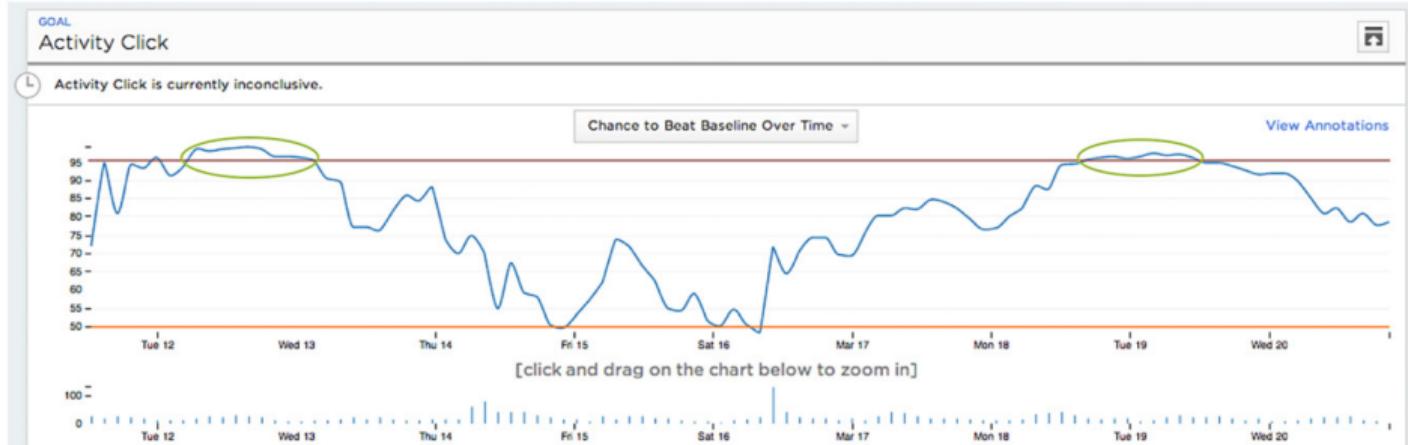


Figure: Probs. better than null (blue trace). The temptation is to stop when a test goes above 95% (green ovals). Problem: under the null hypothesis p_{vals} are uniform, so stopping may lead us to a false positive (i.e. that green oval is plausible under the null). (Johari et al., 2022)

Challenges

- Small sample sizes
- Naive scientific practices
- Other

Table 18

List of identified open problems.

Open problem category	Open problem sub-category	Number of occurrences
Evaluation-related	Extend the evaluation	21
	Provide thorough analysis of approach	16
	Other evaluation-related	36
Process-related	Add process guidelines	9
	Automate process	7
Quality-related	Enhance scalability	7
	Enhance applicability	6

Figure: Quin et al., 2024

Discrete Choice Experiments

Examples

Simulator

Examples

Table 5. Description of Attributes of Laptop Data.

Attributes	Levels
Price in EUR	600, 1,000, 1,400, 1,800, 2,200, 2,600, 3,000, 3,500, 4,000
Brand	Dell, Lenovo, Hewlett-Packard (HP), MaxData, Acer, Apple, Asus
Memory	8 GB, 16 GB, 32 GB, 64 GB
Screen size (in inches)	12, 13, 14, 15, 17
Resolution	1,280 × 1,024, 1,600 × 1,200, 1,920 × 1,080, 1,920 × 1,200, 3,840 × 2,160
Processor	AMD Athlon, AMD Ryzen 3, AMD Ryzen 5, Intel Core i3, Intel Core i5, Intel Core i7, Apple M1
Hard disk	HDD (.3 GB/sec), SSD (.6 GB/sec), SSD (1 GB/sec), SSD (2.5 GB/sec), SSD (3.1 GB/sec)
Size of hard disk	250 GB, 512 GB, 1 TB, 2 TB

Figure: Pachali et al., 2023

Examples

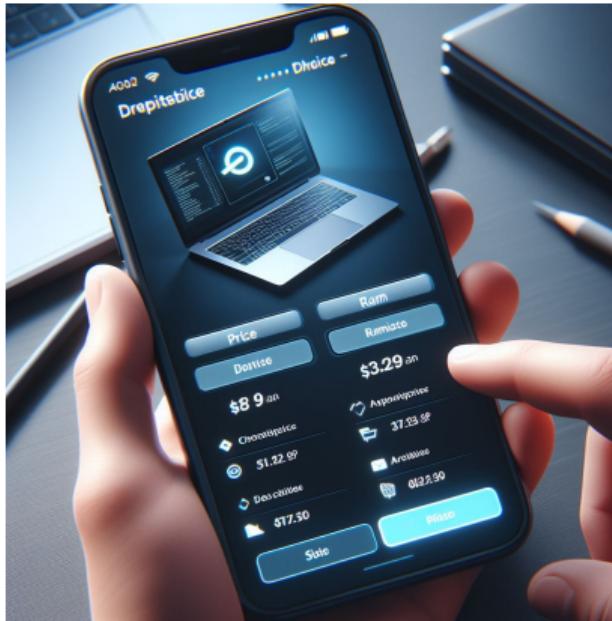


Figure: Depiction of data capturing (done with Copilot Designer)

Examples

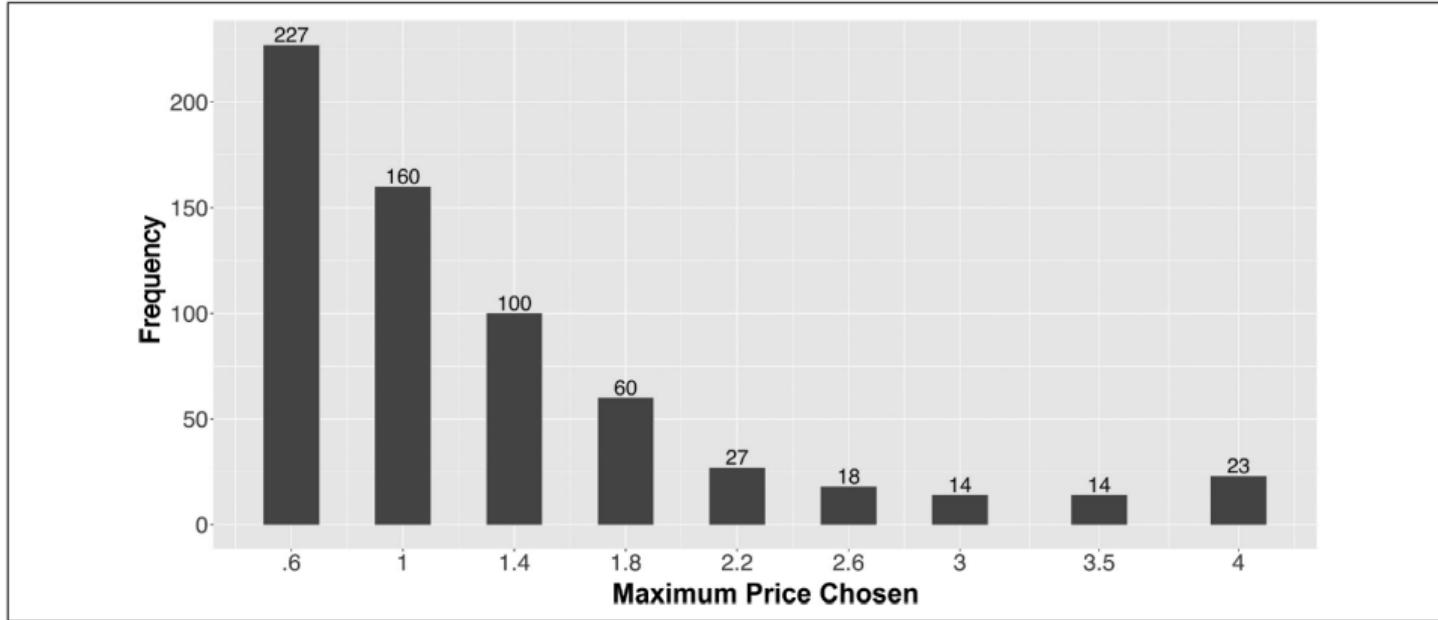


Figure 5. Empirical Frequency Distribution of the Maximum Price Chosen.

Notes: The distribution is across the 643 respondents considered in the analysis. Numeric values are included on top of the bars. Price axis is in 1,000 EUR.

Figure: Stated preferences Pachali et al., 2023

Examples

Table 11. Equilibrium Prices (in EUR).

	Max. Price	Standard Model
Dell	4,000.00	2,230.01
Lenovo	4,000.00	2,242.47
Hewlett-Packard (HP)	4,000.00	2,629.69
MaxData	4,000.00	716.18
Acer	4,000.00	750.33
Apple	4,000.00	3,236.18
Asus	4,000.00	732.13

Figure: With the experiment data and a model, we can calculate equilibrium prices by brand
Pachali et al., 2023

Theory

- People generate a noisy internal utility from observed attributes
- Choice probability depends on the utility
- Stated preferences are comparable to revealed/actual preferences.

Theory

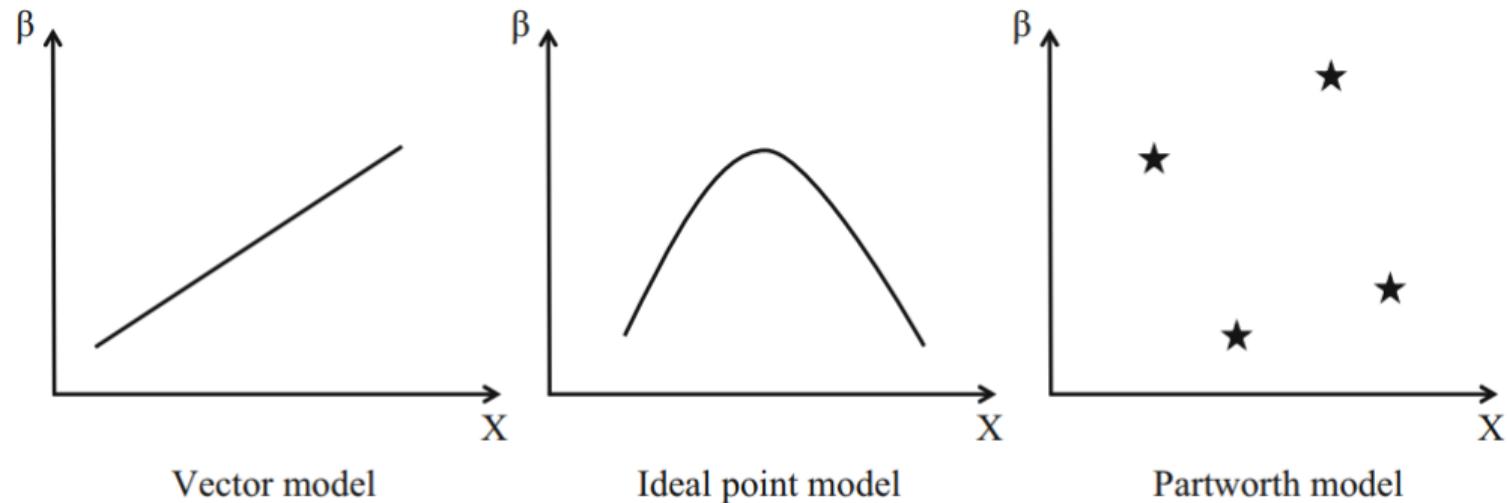


Fig. 4 Alternative functional forms for the evaluation of attribute levels

Figure: Partworth is for qualitative attributes e.g. color. y axis: utility, x axis: attribute level value (Eggers et al., 2021).

Theory

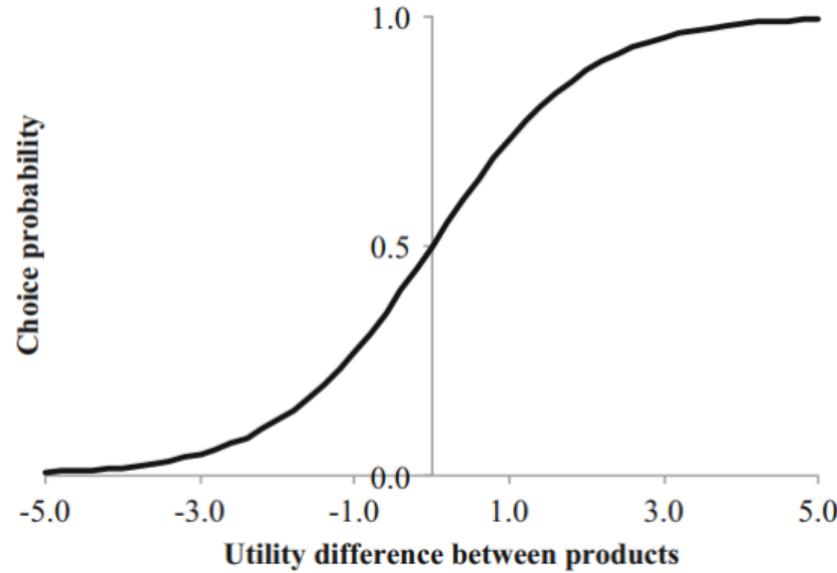


Figure: Choice probability increase with larger utility differences between options (Eggers et al., 2021).

Theory

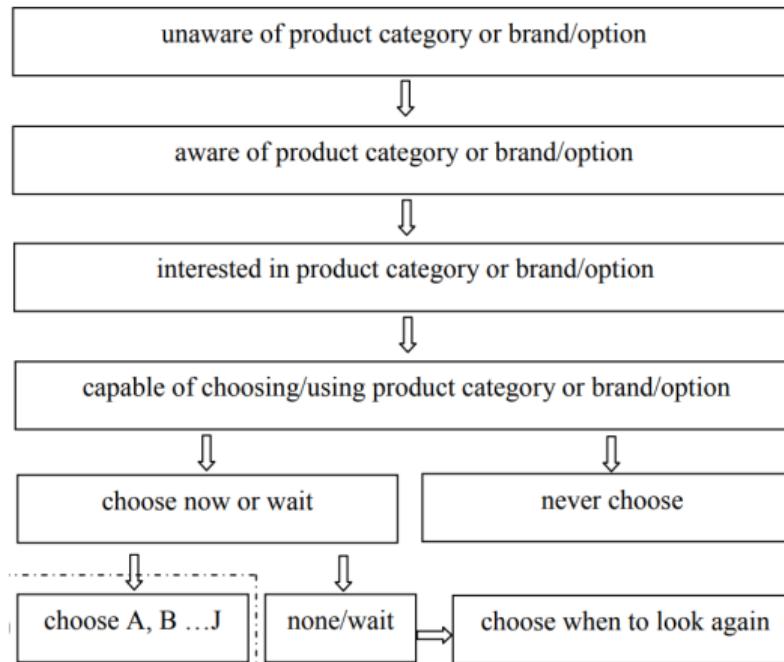


Figure: Consumer choice process (Louviere et al., 2010). if time allows, present some behavioral/neuroeconomics

Implementation (Louviere et al., 2010)

- Identify attributes to test (e.g. Price)
- Assign levels to those attributes (e.g. Price with three levels)
- Decide which combination of attributes and levels to present
- Design a way to present the selected combination of attributes and levels (e.g. via app)
- Select a decision mechanism (e.g. yes/no, auctions, rank-order)
- Sample selection
- Analyze/Model the data

Challenges

- Combinatory explosion of attributes and levels
- Computability of multi-attribute utilities
- Bounded rationality of respondents (e.g. satisficing)
- Complex analysis (e.g. not everyone in a company is familiar with utility theory or multinomial models)

References

-  **Eggers, F., Sattler, H., Teichert, T., & Völckner, F. (2021).** Choice-based conjoint analysis. In *Handbook of market research* (pp. 781–819). Springer.
-  **Johari, R., Koomen, P., Pekelis, L., & Walsh, D. (2022).** Always valid inference: Continuous monitoring of a/b tests. *Operations Research*, 70(3), 1806–1821.
-  **Louviere, J. J., Flynn, T. N., & Carson, R. T. (2010).** Discrete choice experiments are not conjoint analysis. *Journal of choice modelling*, 3(3), 57–72.
-  **Pachali, M. J., Kurz, P., & Otter, T. (2023).** Omitted budget constraint bias and implications for competitive pricing. *Journal of Marketing Research*, 60(5), 968–986.
-  **Quin, F., Weyns, D., Galster, M., & Silva, C. C. (2024).** A/b testing: A systematic literature review. *Journal of Systems and Software*, 112011.
-  **Siroker, D., & Koomen, P. (2015).** *A/b testing: The most powerful way to turn clicks into customers*. John Wiley & Sons.
-  **Viglia, G., Zaefarian, G., & Ulqinaku, A. (2021).** How to design good experiments in marketing: Types, examples, and methods. *Industrial marketing management*, 98, 193–206.