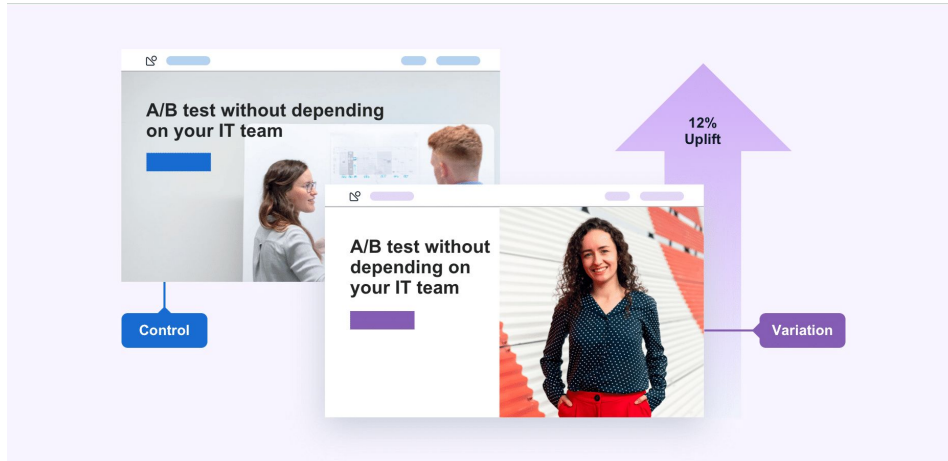




A/B Testing

Nathan HAGEL, Polina ROHOZA

Objectives of A/B Testing



[3, 8]

- Direct Comparison of different designs
- Decide based on data
- Improvement of UI, Reducing Errors, Improving time goals, sales etc.
- Extensively used by companies and research

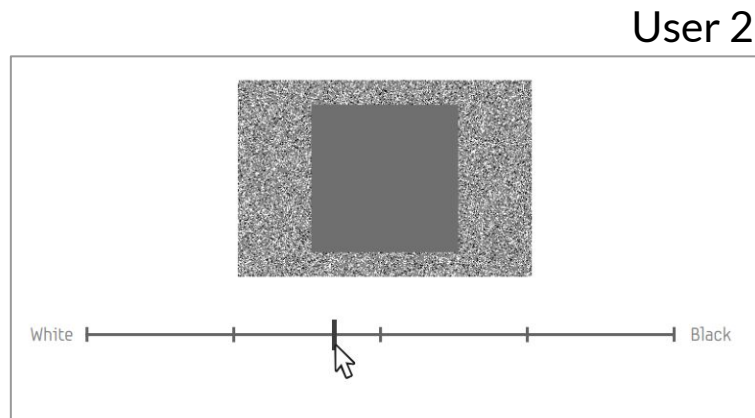
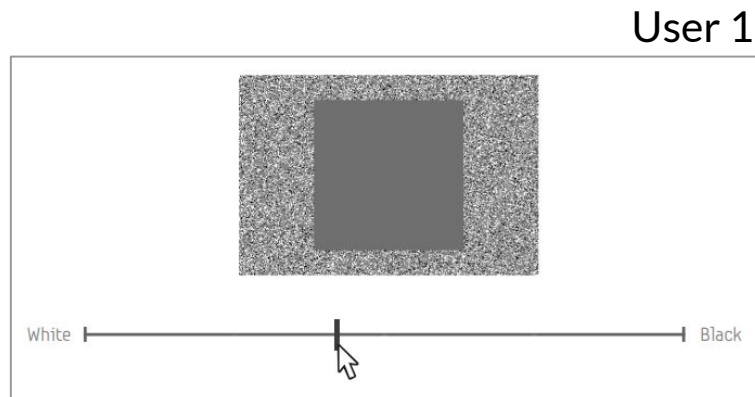
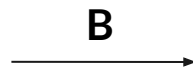
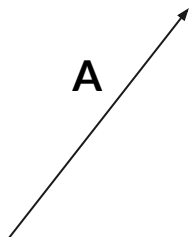
How A/B-testing is done?



Task for user: to rate the “blackness” of a shade of grey.



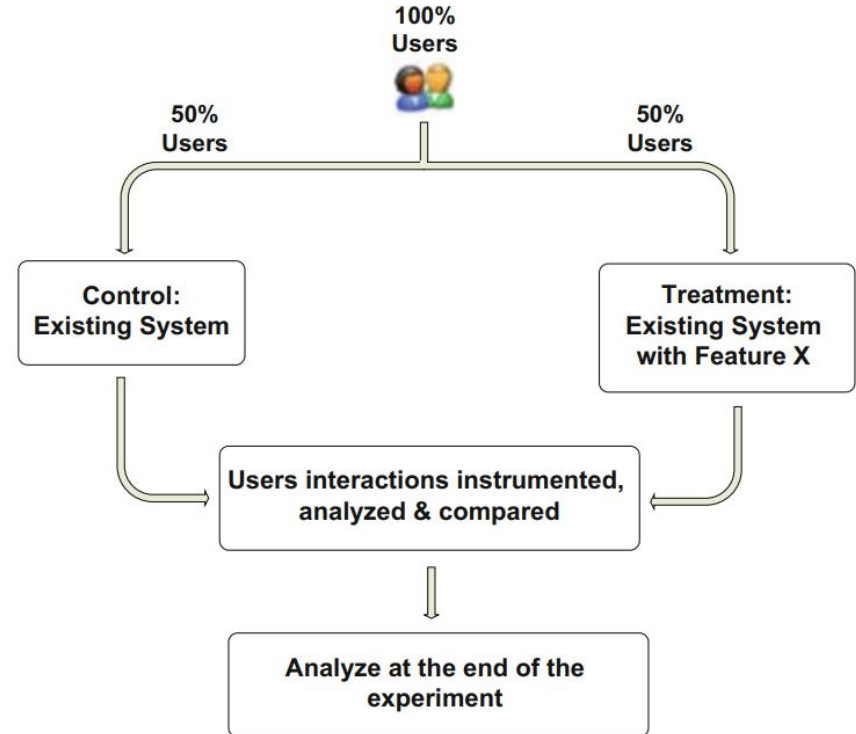
[1]



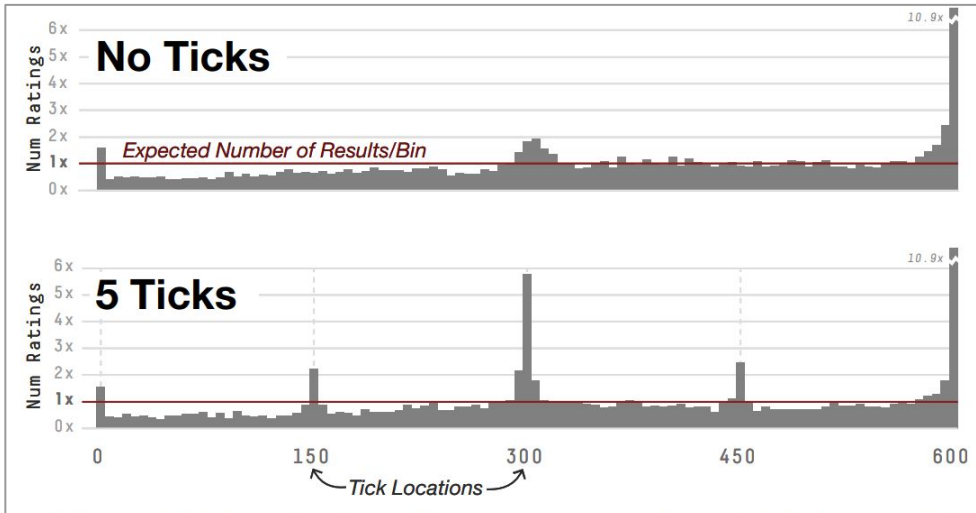
A/B Experiment Structure

Platform: Amazon's Mechanical Turk

Organization: “Once a participant completed one condition (A), they were excluded from participating in any other conditions (B, C...) of the experiment.” **and** “...the 50 shades of gray (were) presented in a randomized order.”



Do Ticks Matter?



Results: Tick marks bias the results towards the location of the ticks [1].

Distribution of users' responses

Other Studies

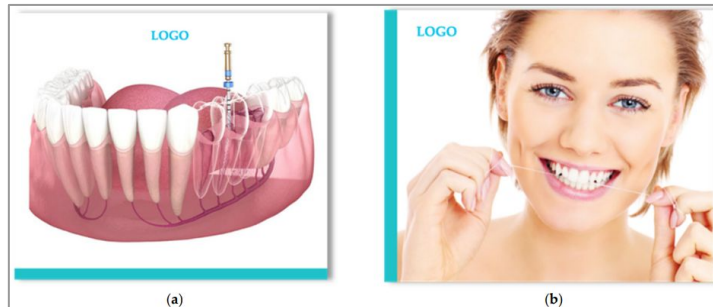
Control

[Esurance® Auto Insurance - You Could Save 28% with Esurance.](#) Ads
www.esurance.com/California
 Get Your Free Online Quote Today!

Treatment

[Esurance® Auto Insurance - You Could Save 28% with Esurance.](#) Ads
www.esurance.com/California
 Get Your Free Online Quote Today!
[Get a Quote](#) · [Find Discounts](#) · [An Allstate Company](#) · [Compare Rates](#)

2: Add campaign in Google Search [11]



3: Facebook posts [12]

Control

Find a new home or apartment

Existing Homes from REALTOR.com®
 Foreclosures from RealtyTrac.com™
 Price Range: \$0 - No Maximum
 Enter City Select a State
 Or Enter ZIP Go

• Senior Living • Home Plans

Treatment 1

Existing Homes Foreclosures New Construction Rentals

Find Existing Homes for Sale

Enter City State
 or
 Enter Zip
 Find homes

Treatment 2

Existing Homes Foreclosures New Construction Rentals

Find Existing Homes for Sale

Enter City State
 or
 Enter Zip
 Find homes

Treatment 3

What are you looking for?

Existing Homes
 New Construction
 Rentals
 Foreclosures
 Senior Living
 Home Valuation
 Professional Services

Enter City State
 Enter Zip
 \$0 to No Max
 Condos/Townhouse Single Family Home
 Find homes

Treatment 4

Find a new Home or Apartment

Existing Homes New Construction Foreclosures Rentals

Enter Zip or Enter City State Search listings

Treatment 5

Find Your Dream Home or Apartment

City, State or ZIP

Existing homes New construction
 Foreclosures Rentals
 Search listings

1: A/B/C/D... test MSN [5]

Benefits and Drawbacks



Benefits:

- + **Data-Driven Decisions**
- + **Reduces Risk of Change**
- + **Isolated Feedback on design changes**
- + **Easy to implement**
- + **User Behaviour Insights**

Drawbacks:

- **Explainability -> Only quantitative metrics, no explanation**
- **Primacy and Newness Effect**

[3, 9, 10]

Best Practices for A/B Testing Pt. 1



**Define
Overall
Evaluation
Criteria
(OEC)**

**Use
Random
Groups**

**Establish
Controlled
Environment**

Be Humble

Best Practices for A/B Testing Pt. 2



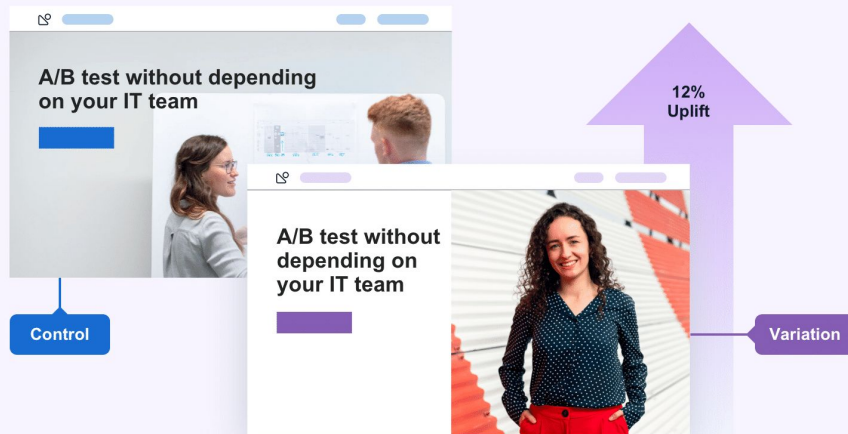
**Keep
Primacy and
Newness
Effect in
mind!**

**Testing only
one Change
at a time
may be too
slow**

**Control the
Setup with
A/A tests**

**Logging of
user
interactions
can save
time**

Conclusion



- Helpful method for evaluating design changes
- Data-Driven decisions
- Consider best practices
- Extensively used in industry

Sources



1. Justin Matejka, Michael Glueck, Tovi Grossman, and George Fitzmaurice. 2016. The Effect of Visual Appearance on the Performance of Continuous Sliders and Visual Analogue Scales. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16). Association for Computing Machinery, New York, NY, USA, 5421–5432. <https://doi.org/10.1145/2858036.2858063>
2. Alix Goguey, Deepak Ranjan Sahoo, Simon Robinson, Jennifer Pearson, and Matt Jones. 2019. Pulp Friction: Exploring the Finger Pad Periphery for Subtle Haptic Feedback. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19). Association for Computing Machinery, New York, NY, USA, Paper 641, 1–13. <https://doi.org/10.1145/3290605.3300871>
3. Lazar, Jonathan, Jinjuan Heidi Feng, and Harry Hochheiser. *Research methods in human-computer interaction*. Morgan Kaufmann, 2017.
4. Michael Toomim, Travis Kriplean, Claus Pörtner, and James Landay. 2011. Utility of human-computer interactions: toward a science of preference measurement. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11). Association for Computing Machinery, New York, NY, USA, 2275–2284. <https://doi.org/10.1145/1978942.1979277>
5. Kohavi, R., Crook, T., Longbotham, R., Frasca, B., Henne, R., Ferres, J. L., & Melamed, T. (2009). Online experimentation at Microsoft. *Data Mining Case Studies*, 11(2009), 39.

Sources



6. Ron Kohavi, Alex Deng, Brian Frasca, Roger Longbotham, Toby Walker, and Ya Xu. 2012. Trustworthy online controlled experiments: five puzzling outcomes explained. In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '12). Association for Computing Machinery, New York, NY, USA, 786–794. <https://doi.org/10.1145/2339530.2339653>
7. Ron Kohavi, Alex Deng, Brian Frasca, Toby Walker, Ya Xu, and Nils Pohlmann. 2013. Online controlled experiments at large scale. In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '13). Association for Computing Machinery, New York, NY, USA, 1168–1176. <https://doi.org/10.1145/2487575.2488217>
8. Picture A/B Testing Websites: <https://static.wingify.com/gcp/uploads/2021/05/1.png>
9. Diane Kelly and Karl Gyllstrom. 2011. An examination of two delivery modes for interactive search system experiments: remote and laboratory. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11). Association for Computing Machinery, New York, NY, USA, 1531–1540. <https://doi.org/10.1145/1978942.1979166>
10. Kohavi, R., Longbotham, R., Sommerfield, D. *et al.* Controlled experiments on the web: survey and practical guide. *Data Min Knowl Disc* **18**, 140–181 (2009). <https://doi.org/10.1007/s10618-008-0114-1>

Sources



11. Kohavi, Ron, and Roger Longbotham. "Online controlled experiments and A/B tests." *Encyclopedia of machine learning and data mining* (2015): 1-11.
12. Nichifor, Eliza, et al. "Eye tracking and an a/b split test for social media marketing optimisation: The connection between the user profile and ad creative components." *Journal of Theoretical and Applied Electronic Commerce Research* 16.6 (2021): 2319-2340.