

CS 4530 & CS 5500

Software Engineering

Lecture 11.3: Acceptance and Inclusivity Testing

Jonathan Bell, John Boyland, Mitch Wand
Khoury College of Computer Sciences
© 2021, released under [CC BY-SA](#)

Learning Objectives for this Lesson

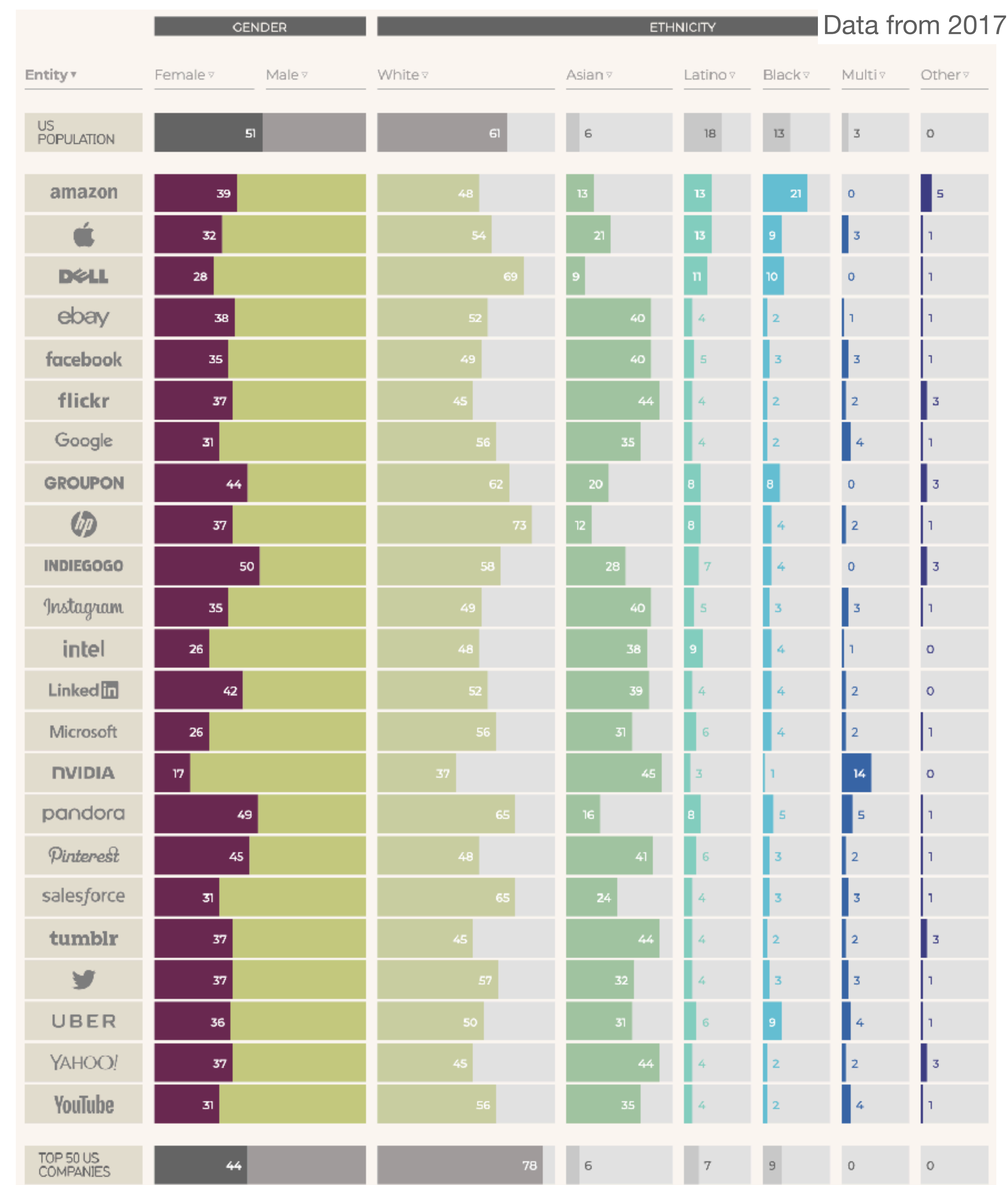
By the end of this lesson, you should be able to...

- Explain ways in which developers of software often differ from the users of that software, introducing potential inclusivity bugs
- Recognize persona-based cognitive walk-throughs as an approach to help put yourself in someone else's shoes

Bias is the Default

We are not our users

- Creating inclusive software requires us to acknowledge that we differ from our users
- Our quality assurance is only as good as we can understand our users



Bias is the Default

Example: Google Photos auto-tagging (2015)



THE WALL STREET JOURNAL



DIGITS

Google Mistakenly Tags Black People as ‘Gorillas,’ Showing Limits of Algorithms

By [Alistair Barr](#)

Updated July 1, 2015 3:41 pm ET

SHARE TEXT

Google is a leader in artificial intelligence and machine learning. But the company’s computers still have a lot to learn, judging by a major blunder by its Photos app this week.

The app tagged two black people as “Gorillas,” according to Jacky Alciné, a Web developer who spotted the error and tweeted a photo of it.

“Google Photos, y’all f**ked up. My friend’s not a gorilla,” [he wrote on Twitter](#).

Google apologized and said it’s tweaking its algorithms to fix the problem.

“We’re appalled and genuinely sorry that this happened,” a company

<https://www.wsj.com/articles/BL-DGB-42522>

<https://www.wired.com/story/when-it-comes-to-gorillas-google-photos-remains-blind/>

WIRED

BACKCHANNEL BUSINESS CULTURE GEAR IDEAS MORE

SIGN IN



TOM SIMONITE

BUSINESS 01.11.2018 07:00 AM

When It Comes to Gorillas, Google Photos Remains Blind

Google promised a fix after its photo-categorization software labeled black people as gorillas in 2015. More than two years later, it hasn’t found one.



Unconscious Bias in Software

Does your software support a variety of users?

- Aside from gender, race and ethnicity, *how* people interact with software varies, research has shown key inclusiveness *facets*:
 - Motivations
 - Information processing style
 - Computer self-efficacy
 - Risk averseness
 - Tech learning style
- Idea: Perform *cognitive walkthrough* of our software, through the lens of someone else



GenderMag: Gender Inclusiveness Magnifier

Persona-based evaluation

Abby Jones¹



scanning all her emails first to get an overall picture before answering any of them.

Background and skills

their software systems are new to her.

“numbers person”

She likes Math and knows how to think with numbers

she also enjoys working with numbers and logic.

Motivations and Attitudes

- **Motivations:** Abby uses technologies to accomplish her tasks. She learns new technologies if and when she needs to, but prefers to use methods she is already familiar and comfortable with, to keep her focus on the tasks she cares about.
- **Computer Self-Efficacy:** Abby has low confidence about doing unfamiliar computing tasks. If problems arise with her technology, she often blames herself for these problems. This affects whether and how she will persevere with a task if technology problems have arisen.
- **Attitude toward Risk:** Abby's life is a little complicated and she rarely has spare time. So she is risk averse about using unfamiliar technologies that might need her to spend extra time on them, even if the new features might be relevant. She instead performs tasks using familiar features, because they're more predictable about what she will get from them and how much time they will take.

How Abby Works with Information and Learns:






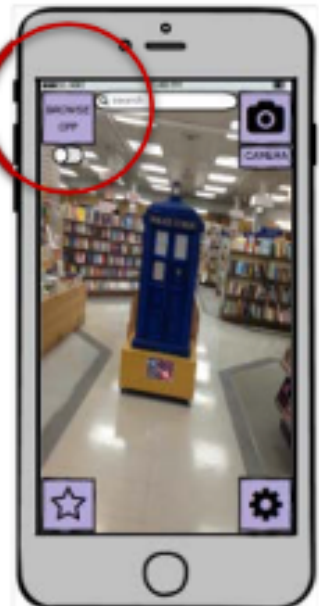


- **Information Processing Style:** Abby tends towards a *comprehensive information processing style* when she needs to more information. So, instead of acting upon the first option that seems promising, she gathers information comprehensively to try to form a complete understanding of the problem before trying to solve it. Thus, her style is “burst-y”; first she reads a lot, then she acts on it in a batch of activity.
- **Learning: by Process vs. by Tinkering:** When learning new technology, Abby leans toward process-oriented learning, e.g., tutorials, step-by-step processes, wizards, online how-to videos, etc. She doesn't particularly like learning by tinkering with software (i.e., just trying out new features or commands to see what they do), but when she does tinker, it has positive effects on her understanding of the software.

¹Abby represents users with motivations/attitudes and information/learning styles similar to hers. For data on females and males similar to and different from Abby, see <http://eusesconsortium.org/gender/gender.php>

GenderMag: Gender Inclusiveness Magnifier

Persona-based evaluation

- Step through a use case for your tool, acting as the persona
- Avoid jumping to conclusions - work in a group with multiple evaluators, take notes of issues as they occur
- Compare to heuristic evaluation (week 6)

<ul style="list-style-type: none"> • 1. Pick a persona. eg: Abby  • 2. Pick a use case/scenario in your tool, eg: <ul style="list-style-type: none"> – in Book Store Navigator app... – “Find science fiction books” 	<ul style="list-style-type: none"> • 3a-b. Pick a Subgoal for that scenario. eg:  Subgoal #1: “See bookstore map”. Q: Will Abby have formed this sub-goal...? <ul style="list-style-type: none"> • Yes/no/<i>maybe</i>. Why? <i>Consider Abby's Motivations...</i> 
<ul style="list-style-type: none"> • 3c-d. Pick an Action for that subgoal.  Action #1: “Tap ‘Browse Off’”: – Q1. Will Abby know what to do? <ul style="list-style-type: none"> • Yes/no/<i>maybe</i>. Why? <i>Consider Abby's, ... Tinkering</i>  <p>→ First answer Q1. After answering it, then perform the action.</p>	<ul style="list-style-type: none"> – 3e. Q2. If she performs the action, producing   <p>will Abby see progress toward the subgoal?</p> <ul style="list-style-type: none"> • Yes/no/<i>maybe</i>. Why? <i>Consider Abby's Self-Efficacy & ...</i>

The Curb Cut Effect



Usability Testing

Directly measuring the usability and inclusivity of our software

- Observe real users interacting with your software - provide each user with a task, monitor their progress towards completing that task
- Consider a diverse set of users that represent those who will use your software
- Validate problems (and fixes) that you identify in cognitive walkthroughs
- Example: applying GenderMag + usability testing for Microsoft Academic

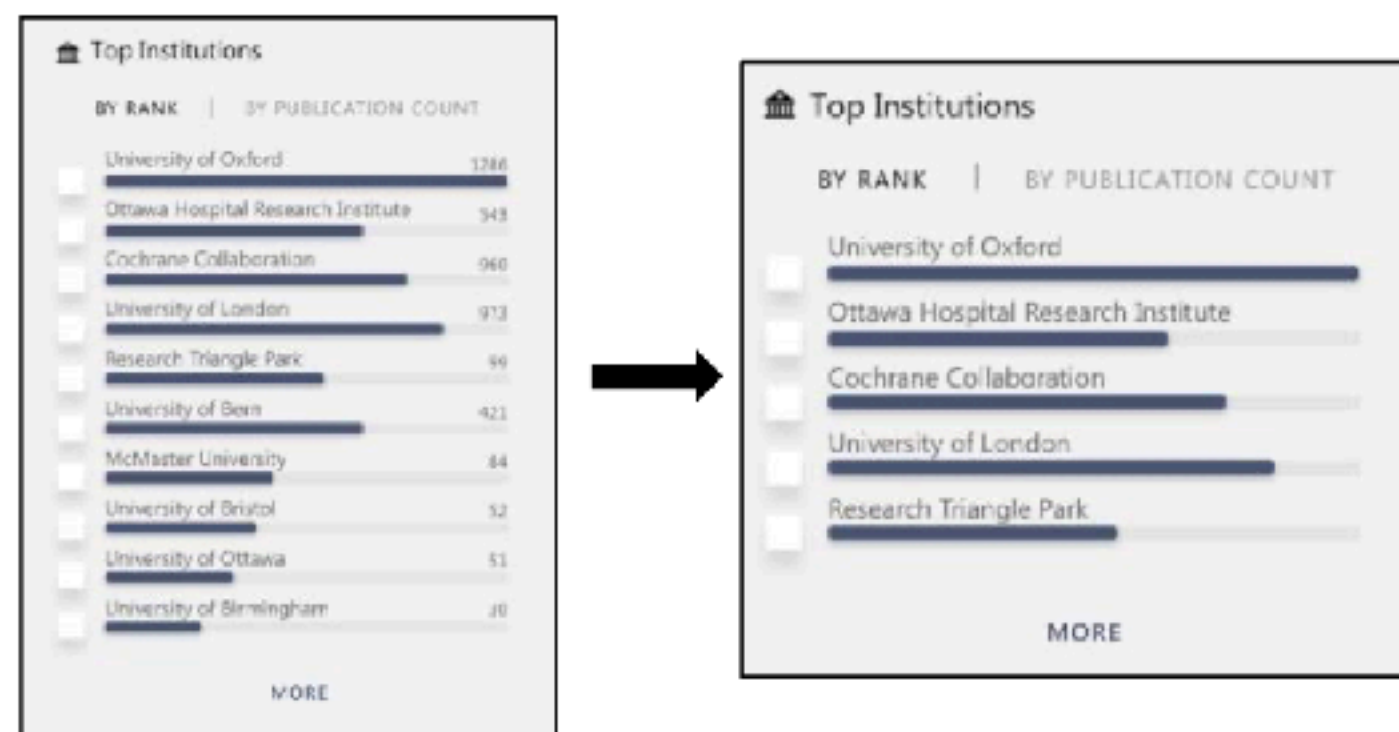


Figure 5. Issues 1 & 2 filtering redesign. (Left) Original: List of institutions with publication counts for each. (Right) post-GenderMag: Shorter list of institutions and removed the publication counts that drew attention away from the checkbox actionability.

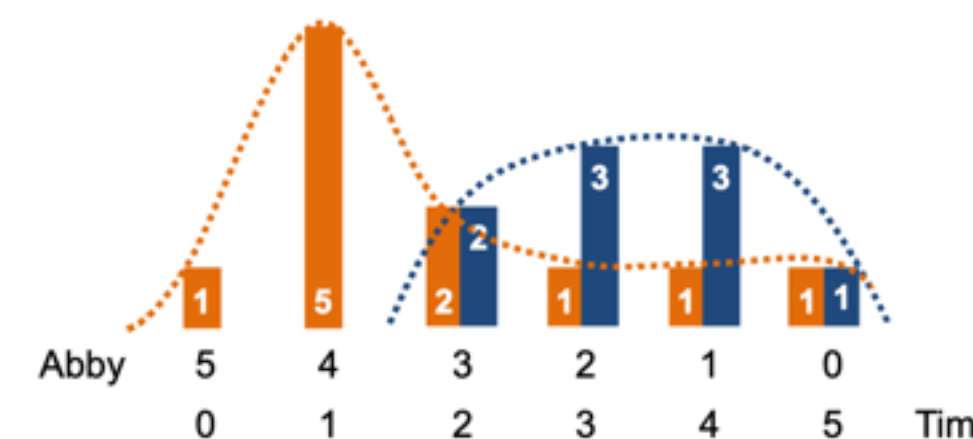


Figure 13. Y-axis: Counts of the 20 men and women participants by their facet values. (Same as Figure 2 but broken out by gender.) Orange: women, blue: men. X-axis: Abby=Abby Facets, Tim=Tim Facets. Example: the left bar says that the only participant with 5 Abby facets (0 Tim facets) was a woman; the right pair of bars says that one man and one woman had 5 Tim facets (0 Abby facets).

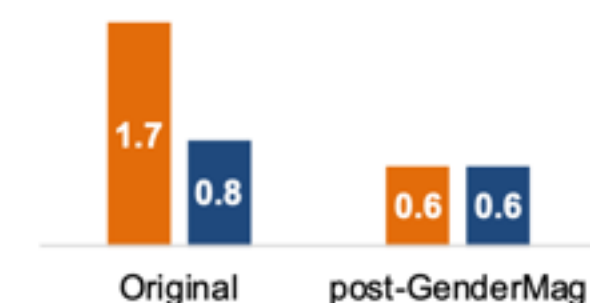
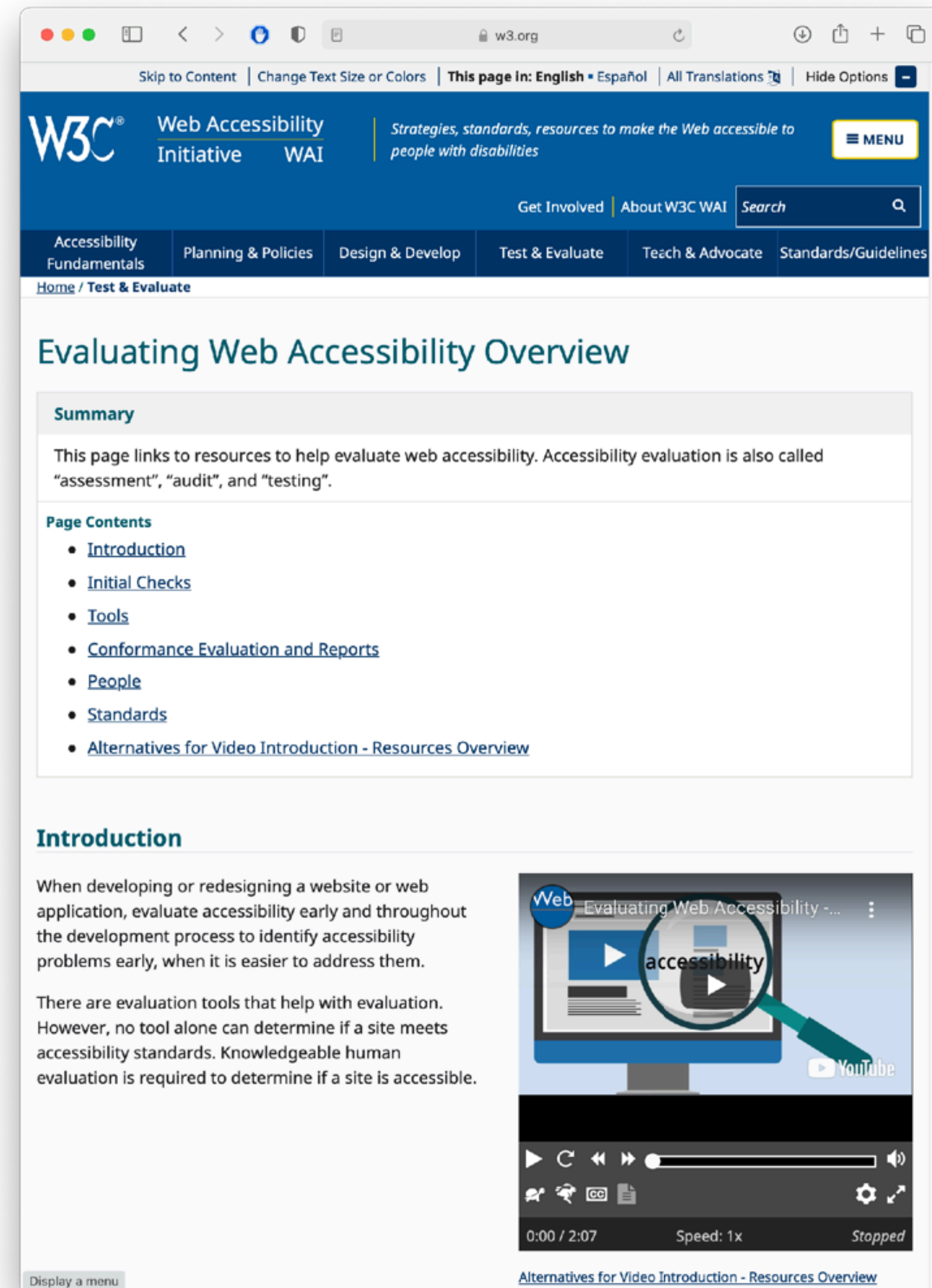


Figure 14. Average number of action failures per person by gender identification (orange: women, blue: men). In the Original version, women's action failure rates were over twice as high as men's; with the post-GenderMag redesign, all failure rates went down, and the gender gap disappeared.

Usability Testing

Evaluating Accessibility

- Check for conformance with requirements of standards
- Involve users in your evaluation - simply “meeting a standard” does not guarantee accessibility



The screenshot shows a web browser displaying the W3C website. The page title is "Evaluating Web Accessibility Overview". The page content includes a summary, a table of contents, and an introduction section. A video player is embedded in the introduction section, showing a video titled "Evaluating Web Accessibility - ...". The video player is currently stopped at 0:00 / 2:07. The page also features a navigation menu and a search bar.

Summary

This page links to resources to help evaluate web accessibility. Accessibility evaluation is also called “assessment”, “audit”, and “testing”.

Page Contents

- [Introduction](#)
- [Initial Checks](#)
- [Tools](#)
- [Conformance Evaluation and Reports](#)
- [People](#)
- [Standards](#)
- [Alternatives for Video Introduction - Resources Overview](#)

Introduction

When developing or redesigning a website or web application, evaluate accessibility early and throughout the development process to identify accessibility problems early, when it is easier to address them.

There are evaluation tools that help with evaluation. However, no tool alone can determine if a site meets accessibility standards. Knowledgeable human evaluation is required to determine if a site is accessible.

[Alternatives for Video Introduction - Resources Overview](#)

<https://www.w3.org/WAI/test-evaluate/>

Usability Testing

For some software, we are *nothing* like our users

ENERGY 14 SYSTEM

COMMODITY SWAP 2 ADDITIONAL INFO +

Reval ^ Run PDC ^

Currency Pair: USD/CAD

Settlement: Physical NDF

GENERAL

Book: Search Books by code or name...

Counterparty: Search Counterparties by code or name...

Trade Date: [Calendar Icon]

Execution Time: 12 : []

UNDERLYING

Swap: PJABF00-CLc1

Instrument: PJABF00

Weight: 1

Unit Conversion: ~

ENERGY 14 SYSTEM

COMMODITY SWAP

Instrument: PJABF00

Weight: 1

Unit Conversion: ~

Swap: PJABF00-CLc1

Instrument: PJABF00

Weight: 1

Unit Conversion: ~

Period: 12M

From: 01-Jan-2016

To: 31-Dec-2016

Frequency: Monthly

Alignment: Calendar

Side: Buy

Quantity: 10,000

Unit: BBL

Per: per Day

Total: 2,520,000

Amount: US\$

Currency: USD

Period: 5

Decision Funding: SP

Calculation Method: Standard

Business Calendar: PLD & NYMEX

Counter Pricing: Yes

Decision Order: After period end

Payment: US 5bd

From: After period end

Business Calendar: NY

Day Type: Business

From: After period end

Business Calendar: NY

Day Type: Business

ENERGY 14 SYSTEM Commodity Swap BPI_1200667-0 cal'15 10K BBL/day (2.52M total) US 5bd BBL

Commodity Swap History BPI_1200667-0

Book: PANATSILOLO Counterparty: RMSFEN01 Executed: 11-Oct-2015 11:10

Underlying: PJABF00 - CLc1 Details Schedule

Instrument	Weight	Unit Conversion
PJABF00	1	~
CLc1	-1	0.274

Swap Period: cal'16

From	To	Frequency	Alignment
01-Jan-2016	31-Dec-2016	Monthly	Calendar

Volume

Unit	per	Total
Buy 10,000	BBL per Day	2,520,000

Price

Currency	per Unit	Calculation	Method
0.86	USD per BBL	3DP	Standard

Observation Dates

Business Calendar	Counter Pricing	Calculation Order
PLD & NYMEX	Yes	Avg at End

Payment: US 5bd

From	Day Type	From	Business Cal	Roll
5	Business	After period end	NY	MFEC

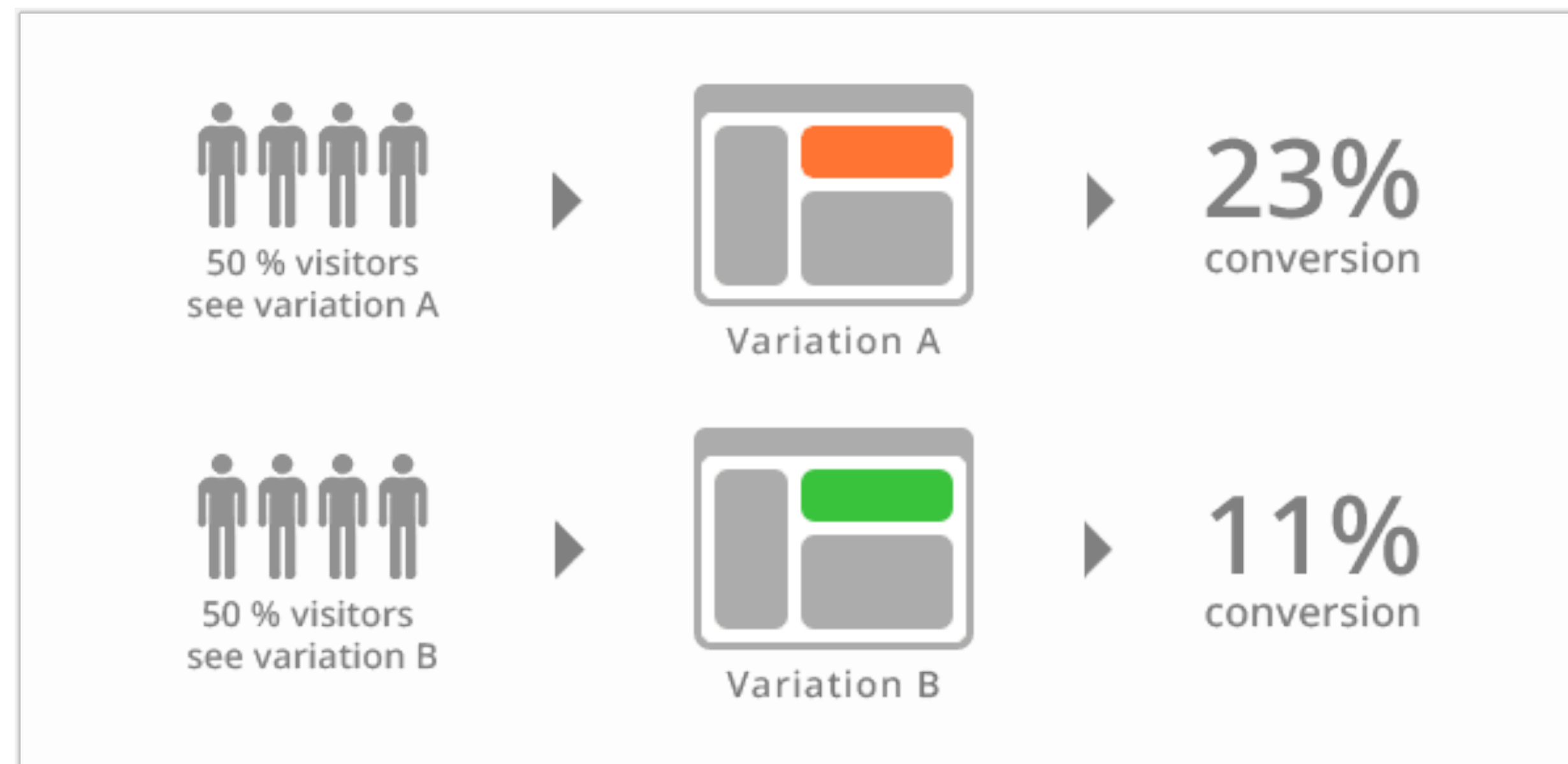
1,000,000 * 0.104 Reval PDC BOOK as Marketer BOOK as Trader

Commodity swap screens & story © 2016, Brad Paley
<http://didi.co/consulting/>

Usability Testing in Continuous Development

A/B Testing

- Ways to test new features for usability, popularity, performance without a focus group
- Show 50% of your site visitors version A, 50% version B, collect metrics on each, decide which is better



Usability Testing in Continuous Development

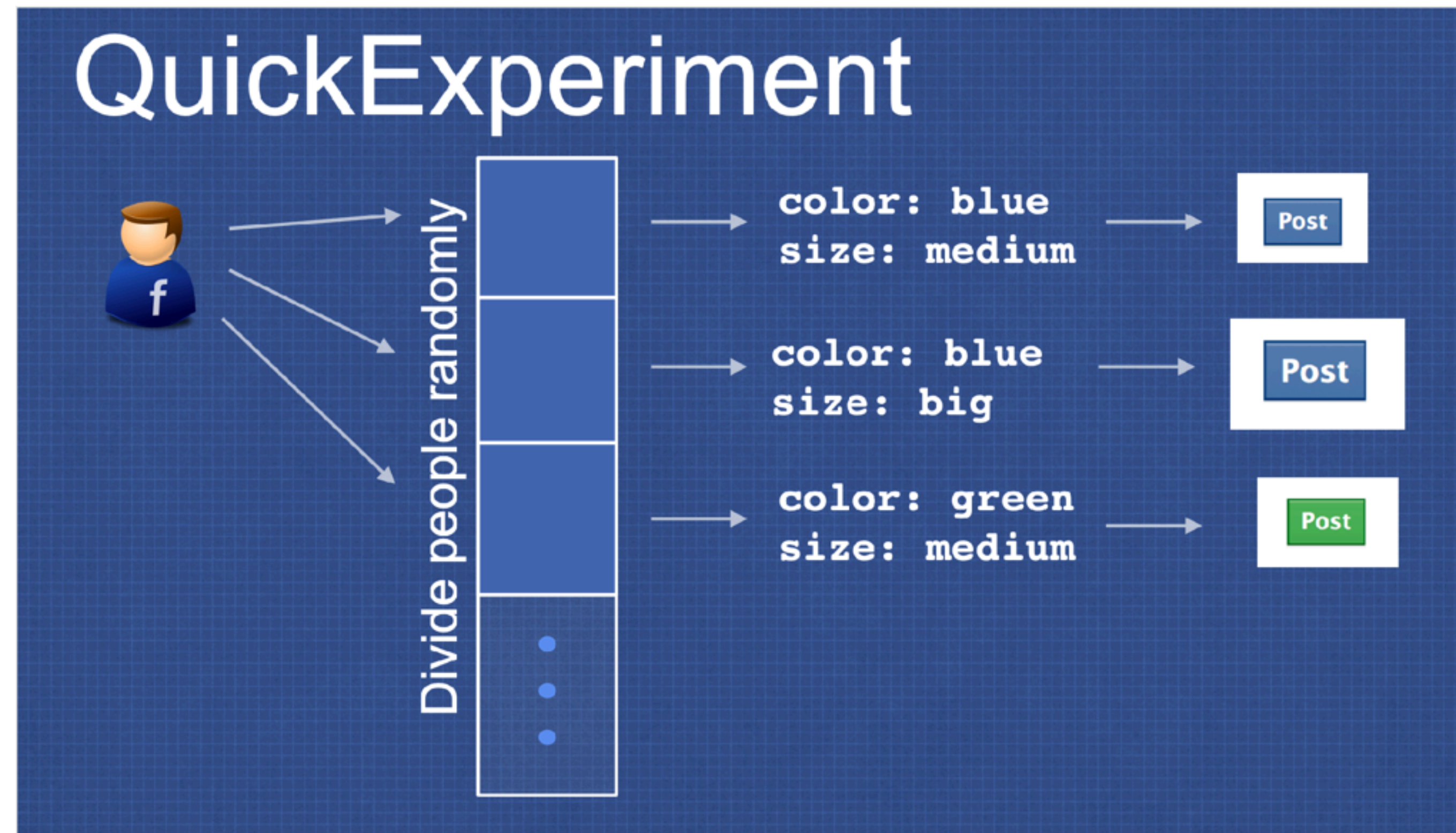
A/B Testing: PlanOut from Facebook (“N=10⁹ user study”)

- Used to test advertising strategies (and Facebook functionality)
- Segment audience and define KPIs, collect results



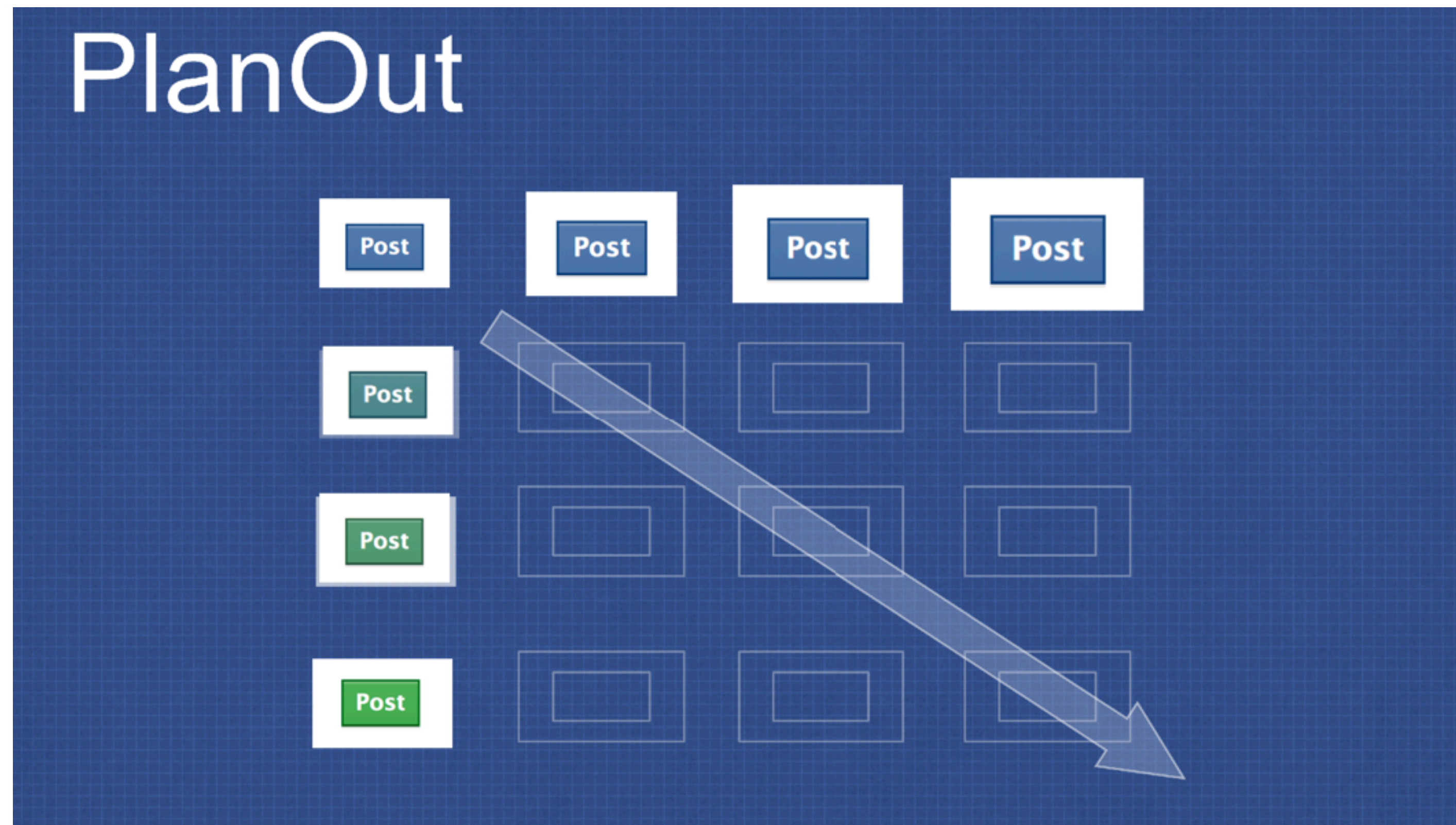
Usability Testing in Continuous Development

A/B Testing: PlanOut from Facebook (“N=10⁹ user study”)



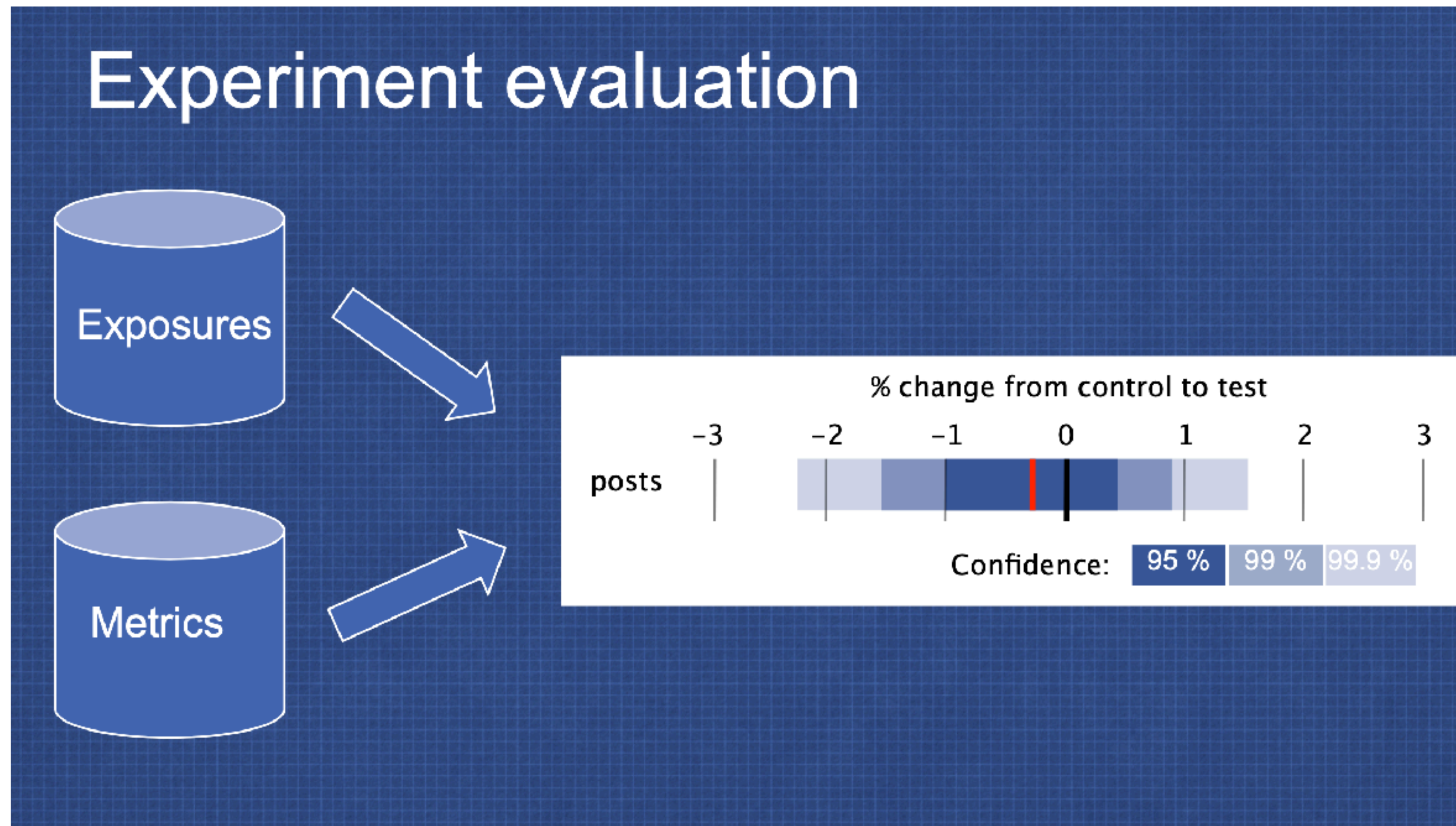
Usability Testing in Continuous Development

A/B Testing: PlanOut from Facebook (“N=10⁹ user study”)



Usability Testing in Continuous Development

A/B Testing: PlanOut from Facebook (“N=10⁹ user study”)



Engineering Equitable Software

Key takeaways

- With great power comes great responsibility
- Anticipate the unanticipated consequences of your software
- Limiting the accessibility of software might save money in the short term, but cost much more in the long term
- Form a diverse development team, and involve a diverse group of users to validate your software

This work is licensed under a Creative Commons Attribution-ShareAlike license

- This work is licensed under the Creative Commons Attribution-ShareAlike 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-sa/4.0/>
- You are free to:
 - Share — copy and redistribute the material in any medium or format
 - Adapt — remix, transform, and build upon the material
 - for any purpose, even commercially.
- Under the following terms:
 - Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.
 - ShareAlike — If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.
 - No additional restrictions — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.