RESEARCH INSIGHT STRATEGIES FOR DESIGNERS AND DECISION-MAKERS

Strategies for Stakeholder Management of Research Insights

Christian P. Rohrer, PhD

I will make these slide available at the end of the talk.

My career at a glance

My Professional Purpose: I believe in business and technology that is human-centered. I empower, educate, and inspire creative, analytic and product professionals to design technology that succeeds for people and thereby excels for the business.

UC Santa Cruz
Computer & Information Sciences (Honors BA, Phi Beta Kappa)

Stanford University
Cognitive Science, HCI, Education (PhD)

Usability Engineer, Director of User Experience Research

Senior Director of User Experience Design

VP of Enterprise Design & Research

SurveyMonkey
Interim VP, Product Design

UserZoom
Industry Advisor

TD Bank
VP of Design

1987 — 1992 — 1998 — 2004 — 2008 — 2011 — 2016 — 2018 — 2022

The Santa Cruz Operation (aka Caldera)
Unix/Xenix Tech Support Engineer & Trainer

Network Computer, Inc. (aka Liberate)

Ethnographic Field Researcher

Director of UER and Product Research

McAfee
Vice President & Chief Design Officer

XD strategy
Founder & Principal

McAfee
Head of CXD

California State University MONTEREY BAY
Lecturer, School of Computing & Design

oscillo biosciences
UX Strategist & Advisor

Nielsen Norman Group Invited Industry Instructor

Tech | Research | Design

My contributions to CHI

SIGCHI

TorCHI

CHI'94 CHI'96 CHI 2002 CHI 2004 CHI 2008 CHI 2016

1987 1992 1998 2004 2008 2011 2016 2018 2022

CHI'95

BayCHI
Jan 2009

BayCHI
Jan 2019

TorCHI
Sep 2023

BayCHI

Student Volunteer

Author

Speaker

XD strategy

WHAT KIND OF RESEARCH INSIGHTS WORKS BEST FOR DESIGNERS?

WHY?

WHAT KIND OF RESEARCH INSIGHTS WORKS BEST FOR EXECUTIVES AND DECISION MAKERS?
WHY?

CAN WE SATISFY THE NEEDS OF BOTH?

XD strategy

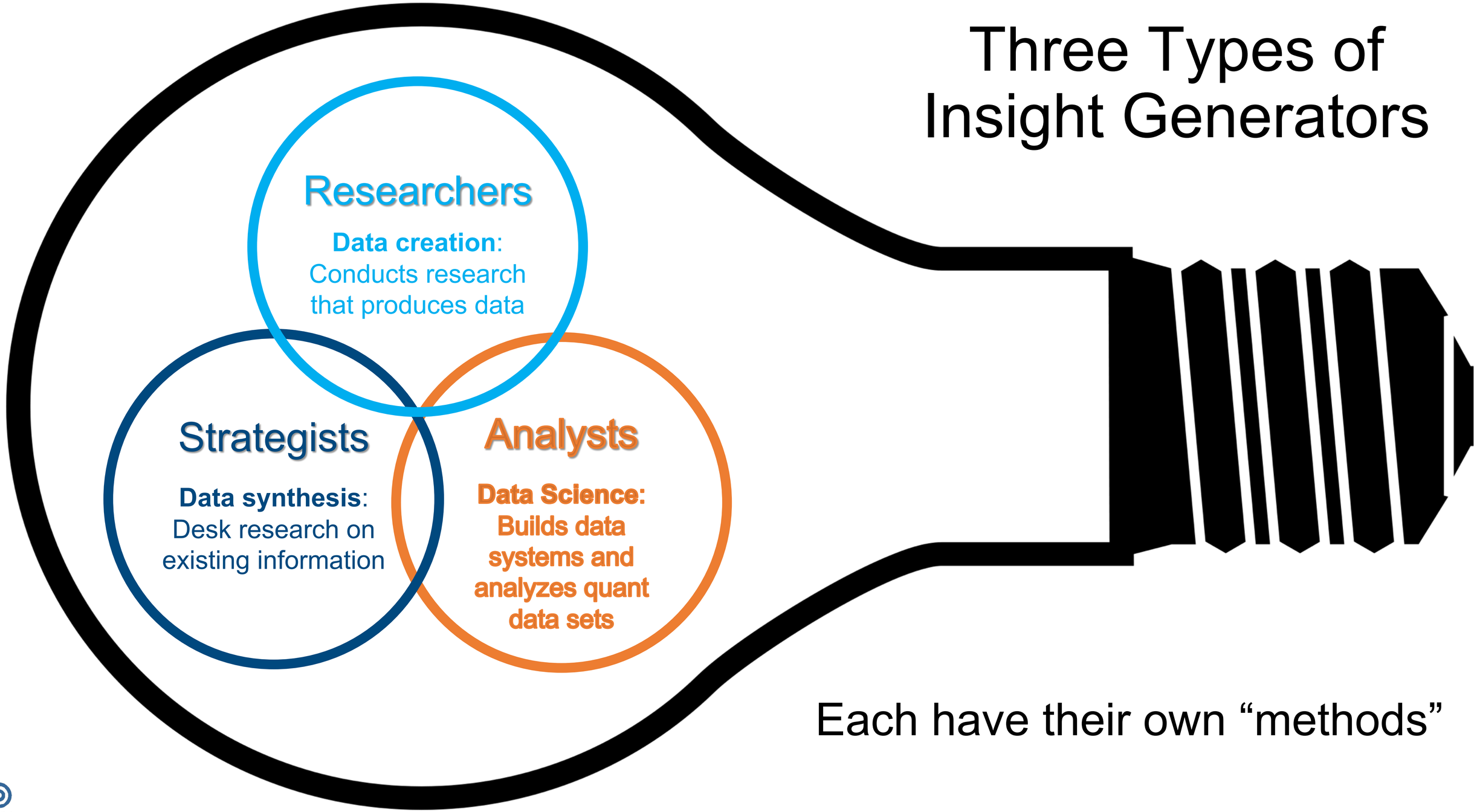
STRATEGIES I'VE USED IN THE PAST (IN THIS ORDER, FOR BETTER OR WORSE)

1. Teach Everyone About Research Methods
2. Sell the Benefits of Qualitative Research
3. Conduct both Qualitative and Quantitative Research
4. Develop Your Own Metrics
5. Lead With the Design Process

STRATEGIES I'VE USED IN THE PAST (IN THIS ORDER, FOR BETTER OR WORSE)

1. Teach Everyone About Research Methods
2. Sell the Benefits of Qualitative Research
3. Conduct both Qualitative and Quantitative Research
4. Develop Your Own Metrics
5. Lead With the Design Process

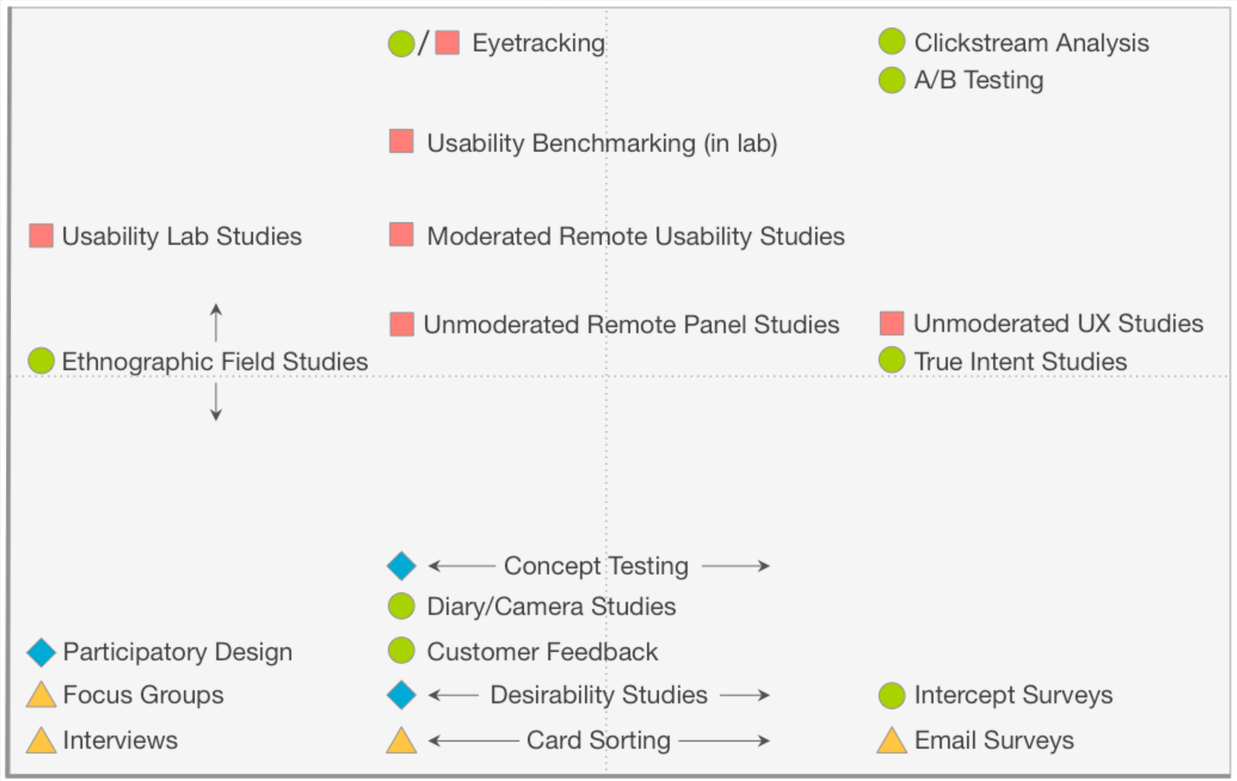
Three Types of Insight Generators



Each have their own “methods”

XD strategy

DATA & USER RESEARCH METHODS



strategy

Data Analysis & User Research Methods

- Many methods available
- Many pros and cons
- But what method should you to use?
- One way to know: Plot the methods on a 3D landscape

The Qualitative vs. Quantitative Dimension

Qualitative Research*

- Data typically gathered directly by observing the user
- Researcher can ask follow-up questions, probe on behavior, and possibly adjust the protocol as the study progresses
- Analysis of data is not mathematical

Quantitative Research

- Data typically gathered indirectly through a research instrument such as a survey or web server logs
- Large amounts of data that can be coded and analyzed mathematically to compare and measure

*Sometimes “qualitative” is used to refer to open-ended survey data; that is not what is meant by qualitative here, as it is not directly gathered by the researcher.

QUALITATIVE VS. QUANTITATIVE DATA

BEHAVIORAL

QUALITATIVE DATA IS GATHERED *DIRECTLY* BY THE RESEARCHER

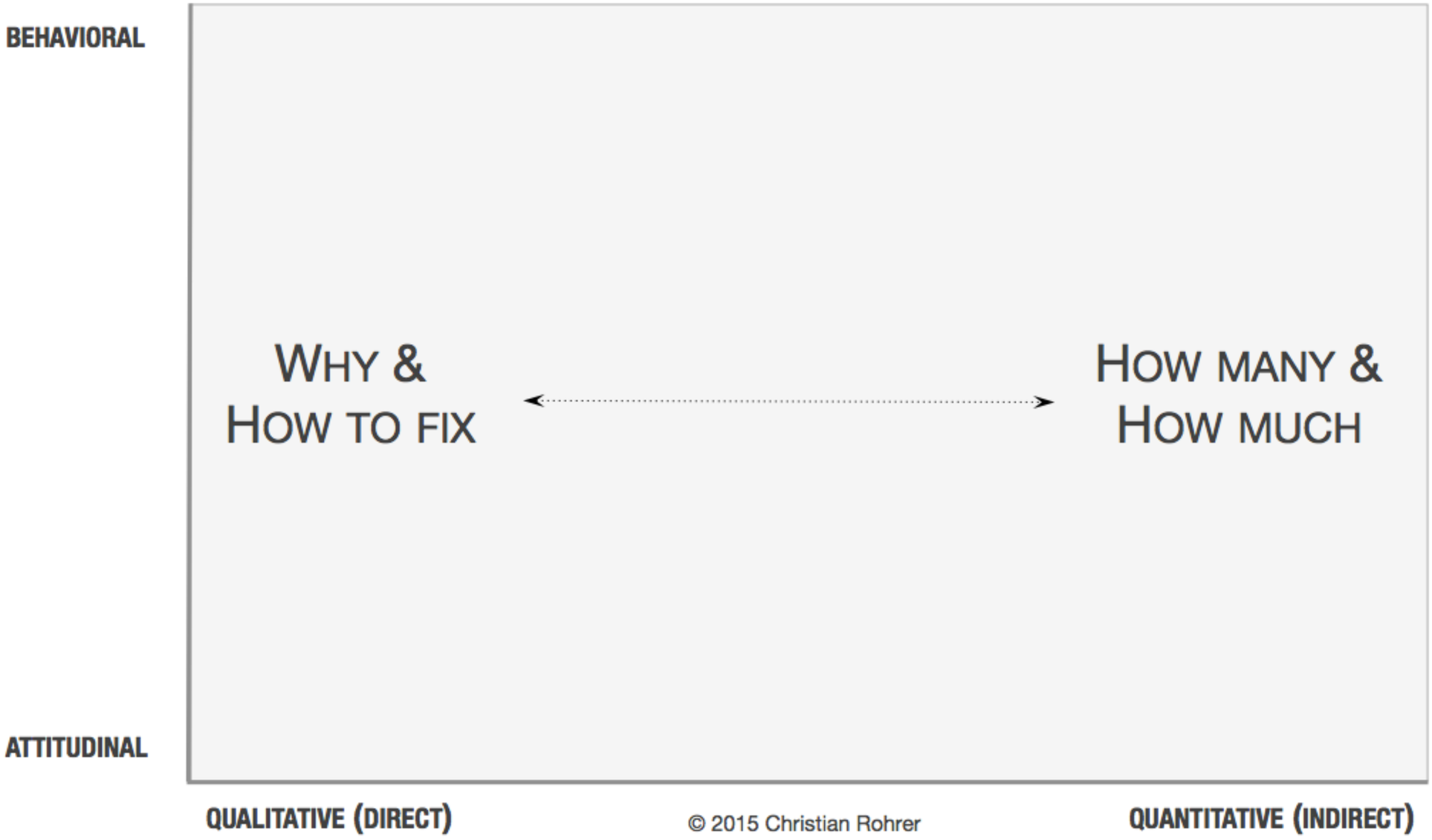
QUANTITATIVE DATA IS GATHERED *INDIRECTLY*, BY AN INSTRUMENT (WEB LOG OR SURVEY)

ATTITUDINAL

QUALITATIVE (DIRECT)

QUANTITATIVE (INDIRECT)

Qualitative vs. Quantitative Data



QUALITATIVE VS. QUANTITATIVE DATA

BEHAVIORAL

Describes and understands people

Inspires design ideas

WHY & HOW TO FIX

Useful to Designers

Findings are valid, even with small N

ATTITUDINAL

HOW MANY & HOW MUCH

QUALITATIVE (DIRECT)

QUANTITATIVE (INDIRECT)

Qualitative vs. Quantitative Data

BEHAVIORAL

Quantifies, often with precision

Measures which is “better”

Why & How to fix

How many & How much

Useful to Decision Makers

People can be blinded by or misuse large N

ATTITUDINAL

QUALITATIVE (DIRECT)

QUANTITATIVE (INDIRECT)

The Attitudinal vs. Behavioral Dimension

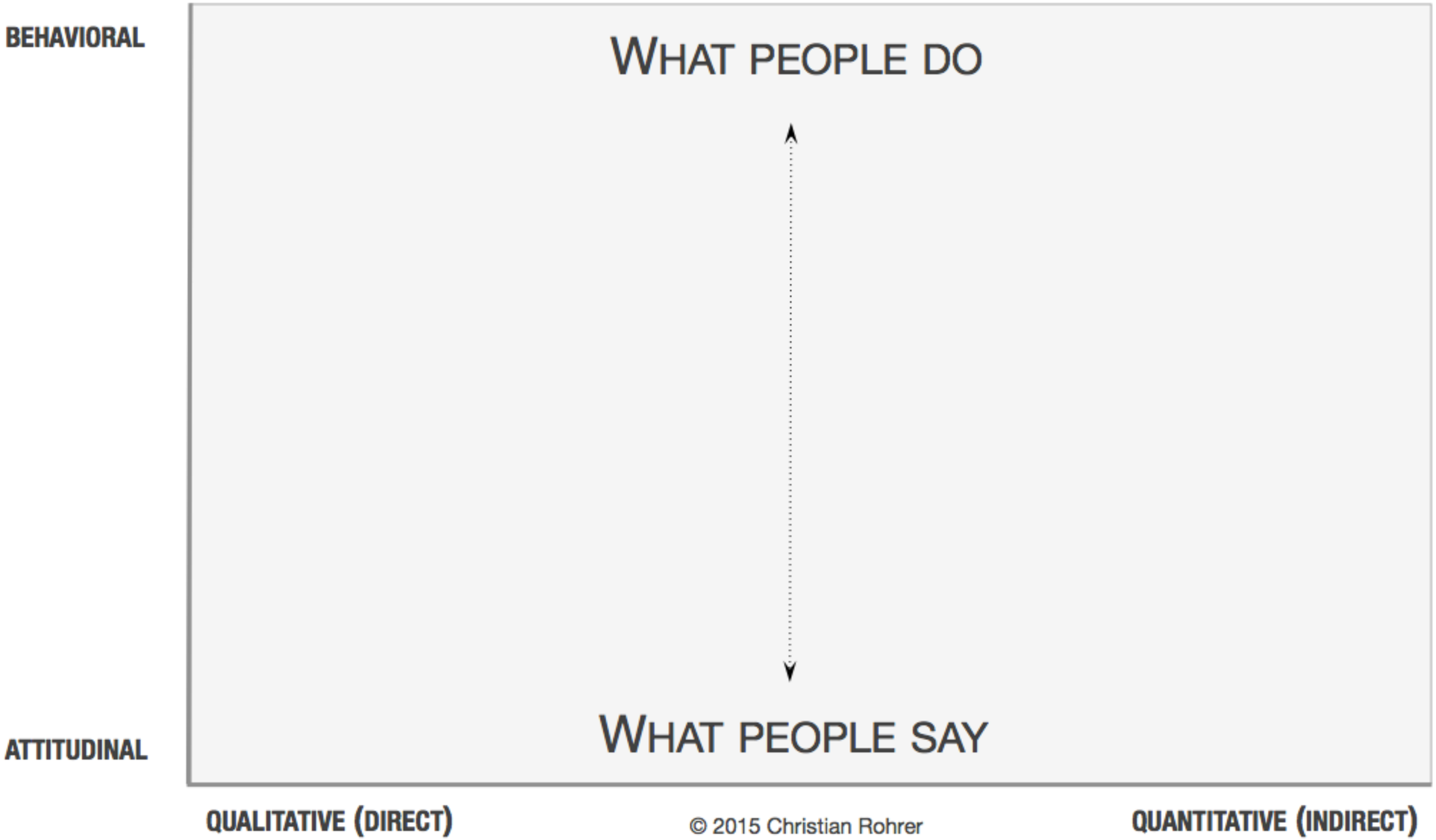
Attitudinal Research

- Understand, measure, or inform change of people's stated beliefs
- Often called "self-reported" data
- Often relied on heavily in marketing departments
- Example methods: Surveys, Focus Groups

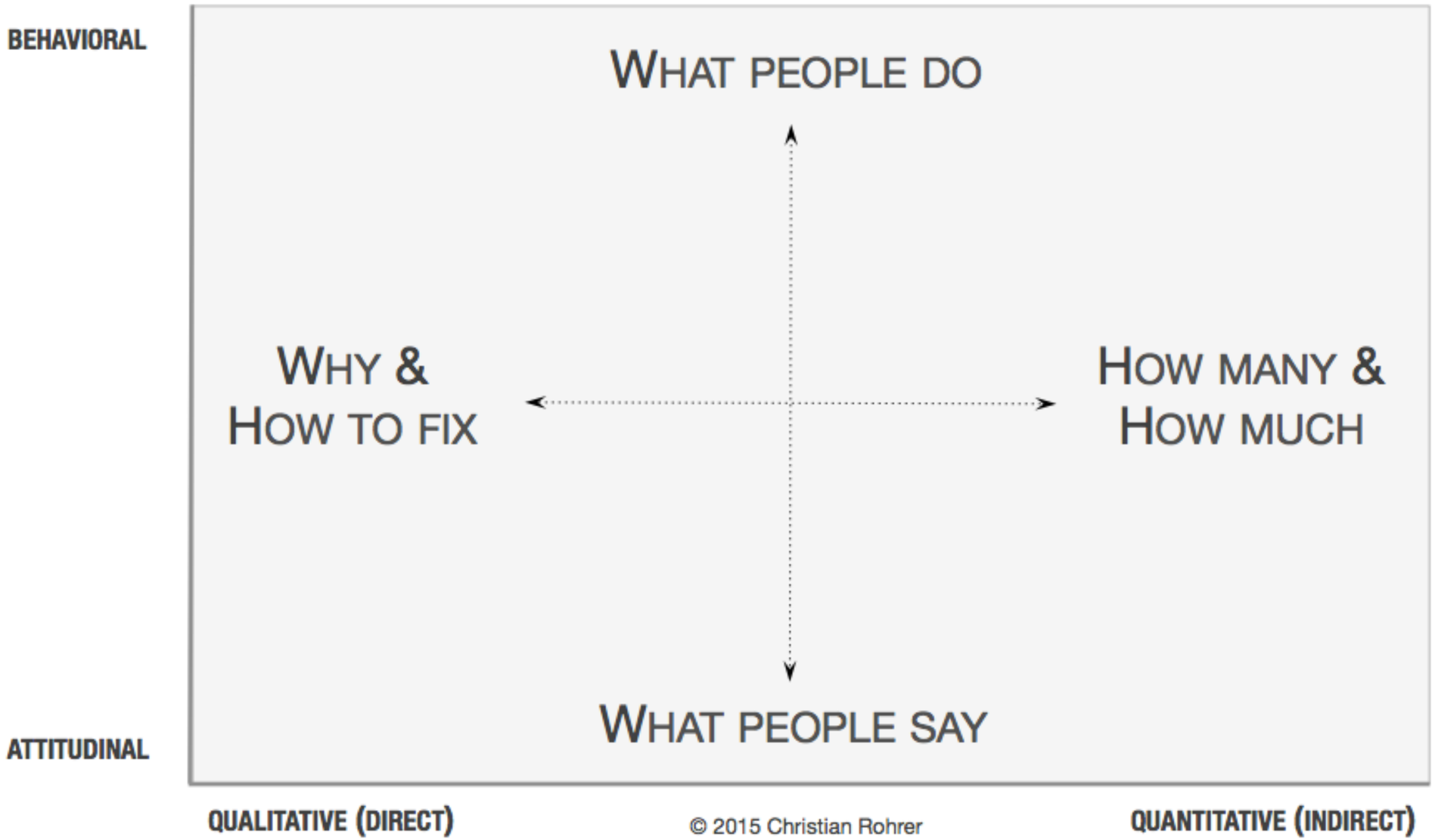
Behavioral Research

- Understand what people do with minimal interference from the method itself
- Example methods: Data Mining/Analysis, Eyetracking

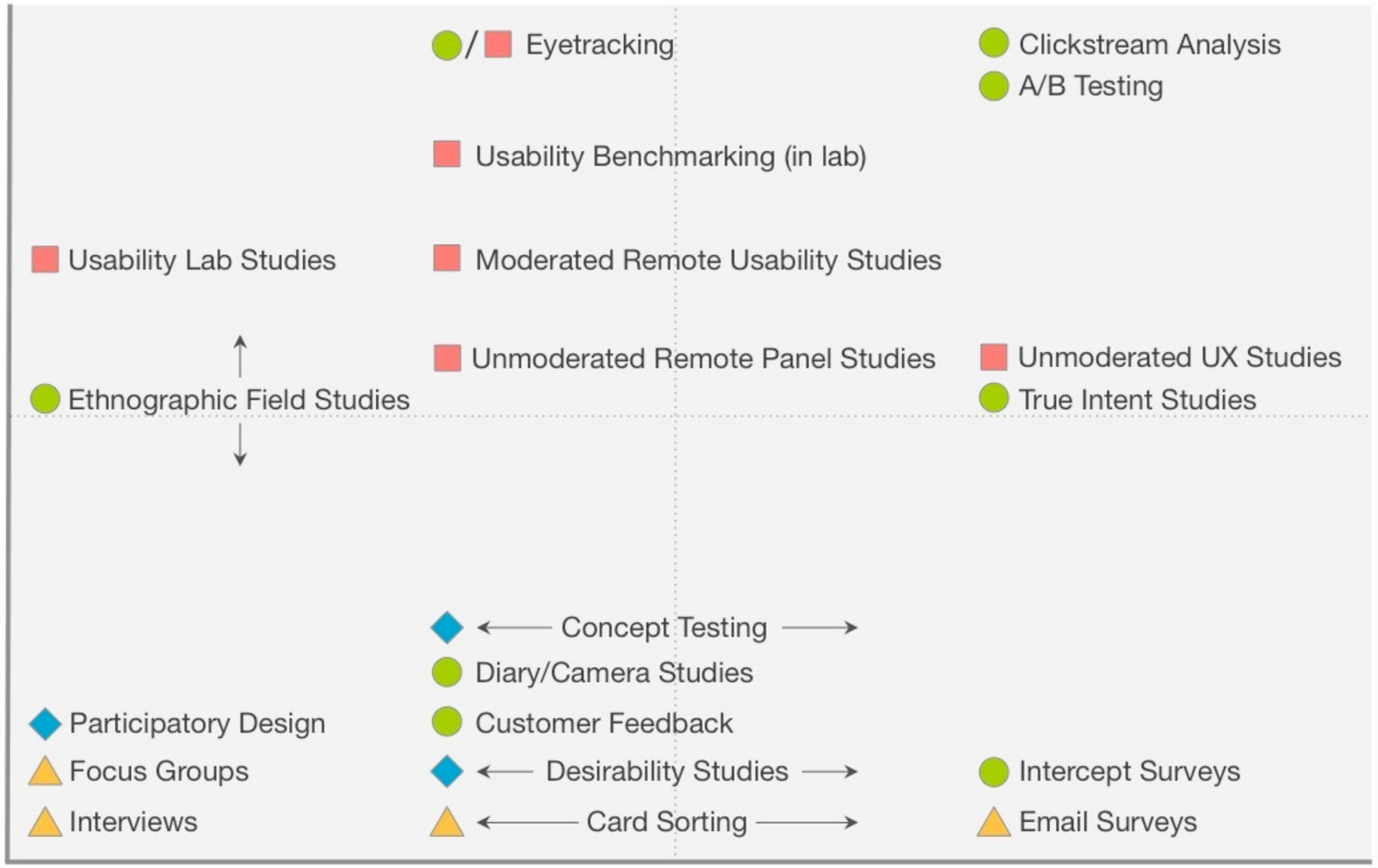
Behavioral vs. Attitudinal



QUESTIONS ANSWERED BY RESEARCH METHODS ACROSS THE LANDSCAPE

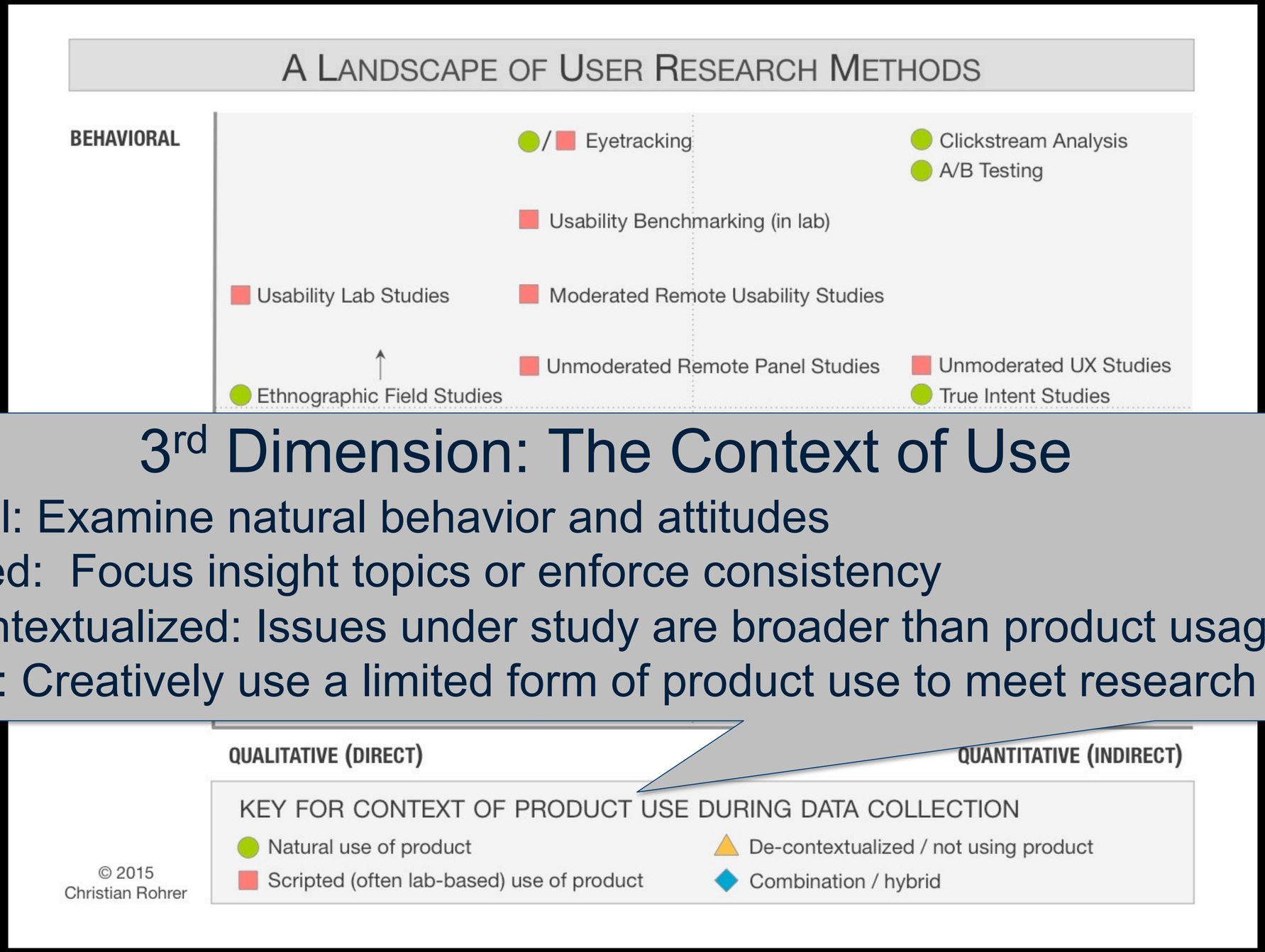


A Landscape of User Research Methods



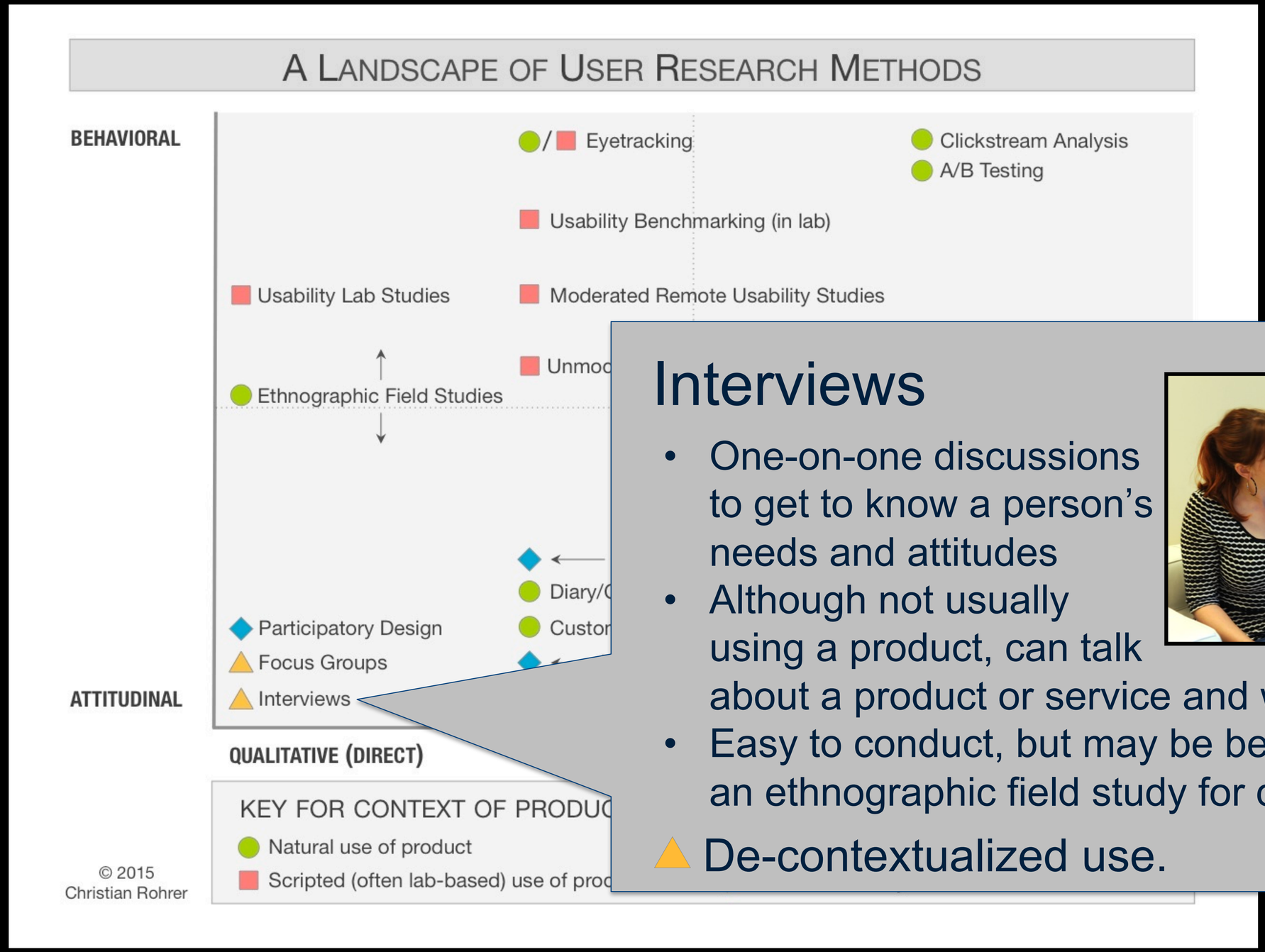
KEY FOR CONTEXT OF PRODUCT USE DURING DATA COLLECTION

- Natural use of product
- Scripted (often lab-based) use of product
- De-contextualized / not using product
- Combination / hybrid



3rd Dimension: The Context of Use

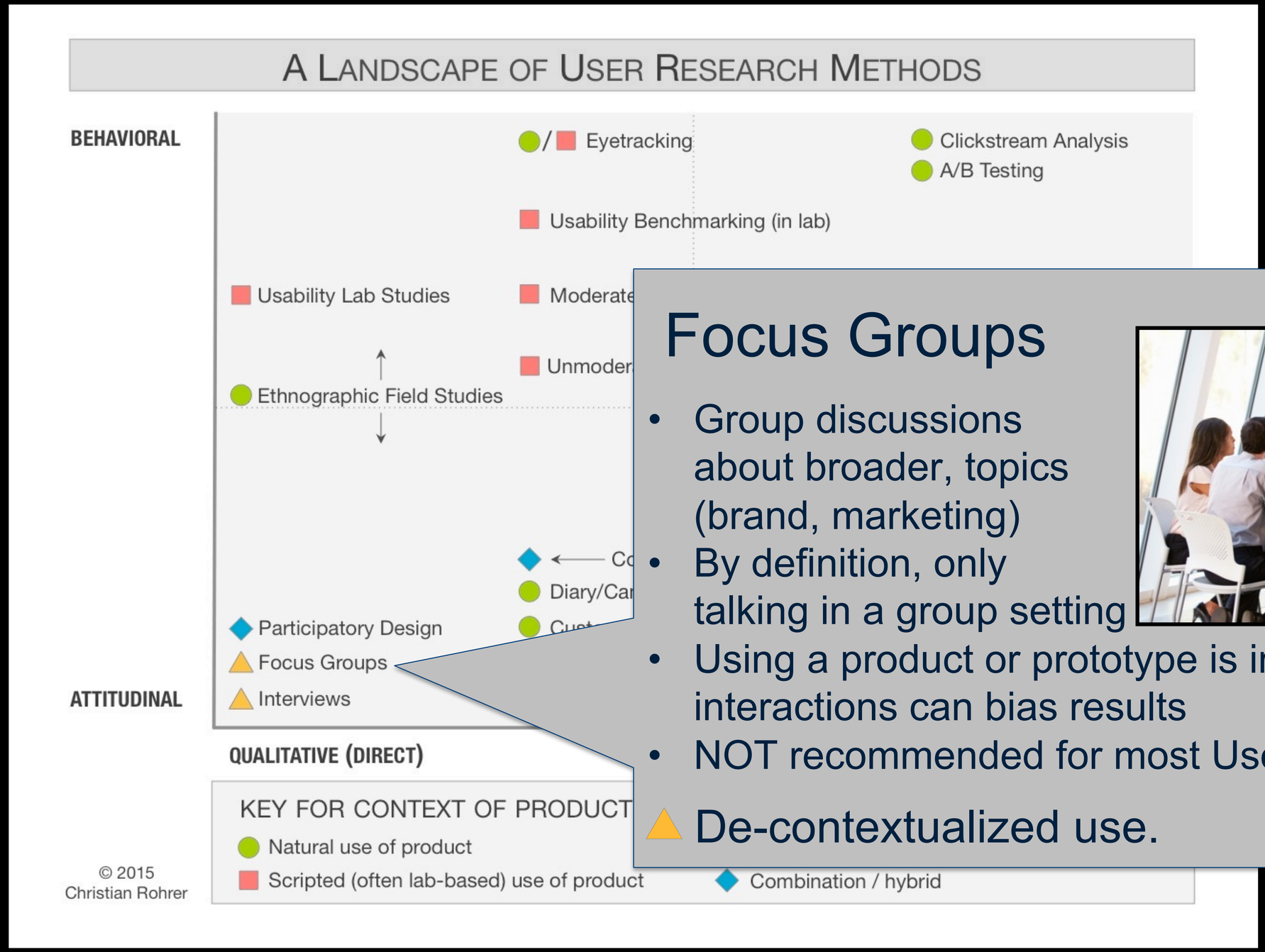
- Natural: Examine natural behavior and attitudes
- Scripted: Focus insight topics or enforce consistency
- De-contextualized: Issues under study are broader than product usage
- Hybrid: Creatively use a limited form of product use to meet research goals



Interviews

- One-on-one discussions to get to know a person's needs and attitudes
- Although not usually using a product, can talk about a product or service and what it should do
- Easy to conduct, but may be better to do a field study or an ethnographic field study for deeper insights

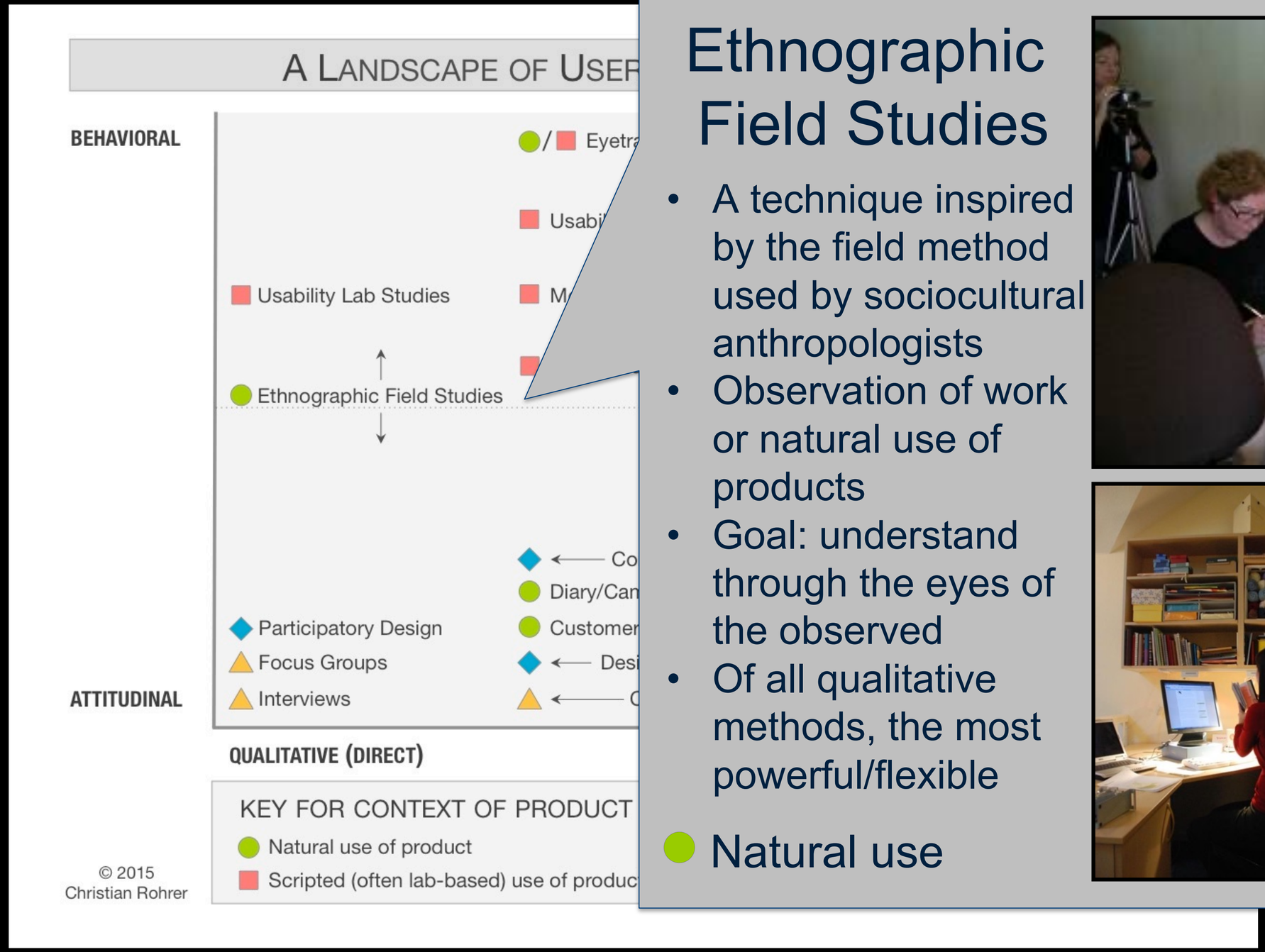
▲ De-contextualized use.



Focus Groups

- Group discussions about broader, topics (brand, marketing)
- By definition, only talking in a group setting
- Using a product or prototype is impractical, and group interactions can bias results
- NOT recommended for most User Research purposes

△ De-contextualized use.



Ethnographic Field Studies

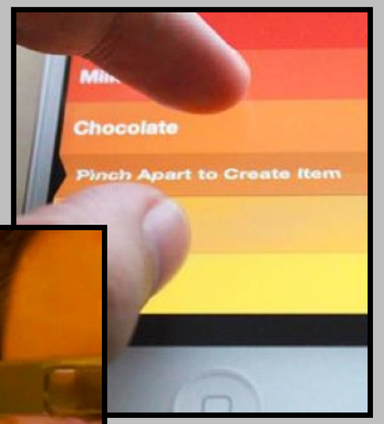
- A technique inspired by the field method used by sociocultural anthropologists
- Observation of work or natural use of products
- Goal: understand through the eyes of the observed
- Of all qualitative methods, the most powerful/flexible

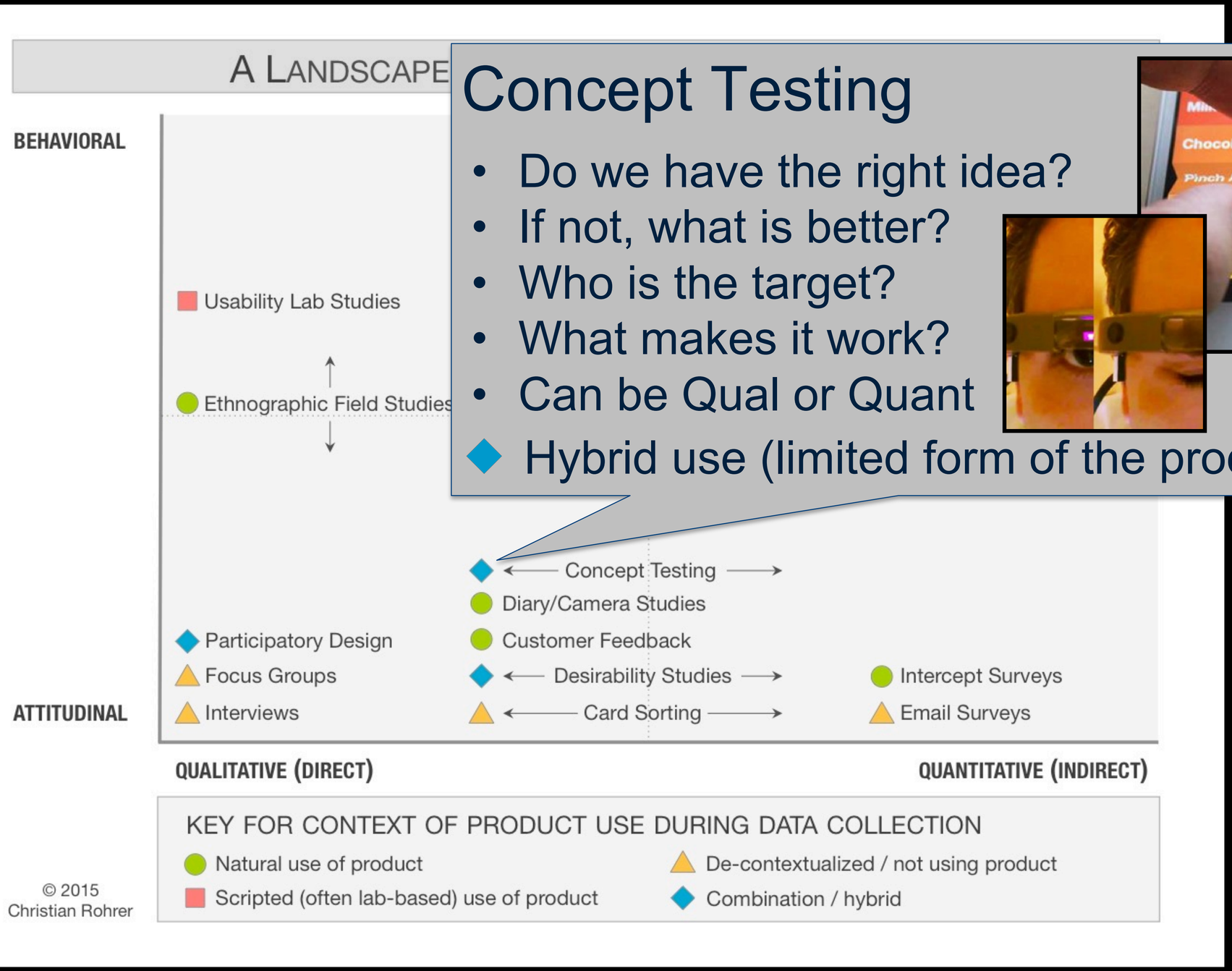
● Natural use

Concept Testing

- Do we have the right idea?
- If not, what is better?
- Who is the target?
- What makes it work?
- Can be Qual or Quant

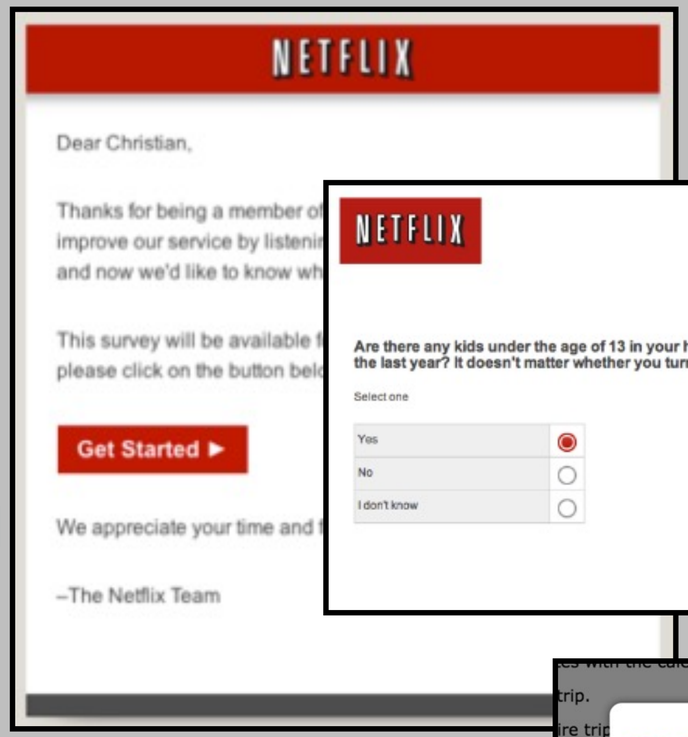
◆ Hybrid use (limited form of the product)

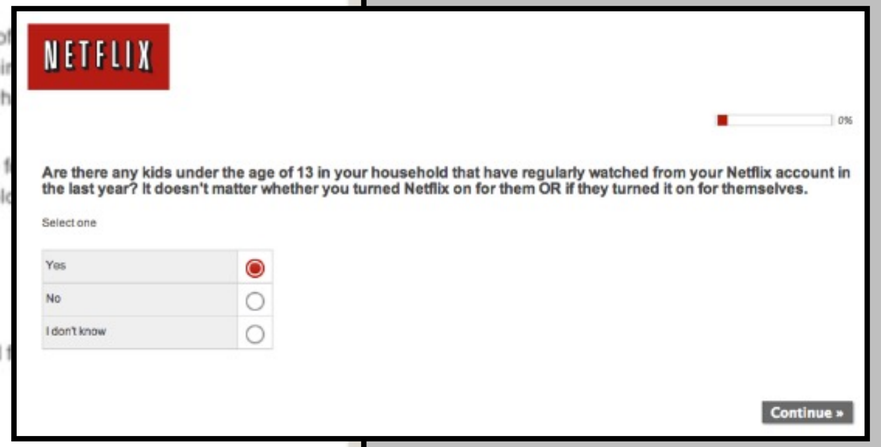


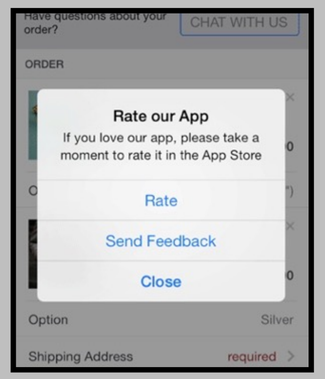


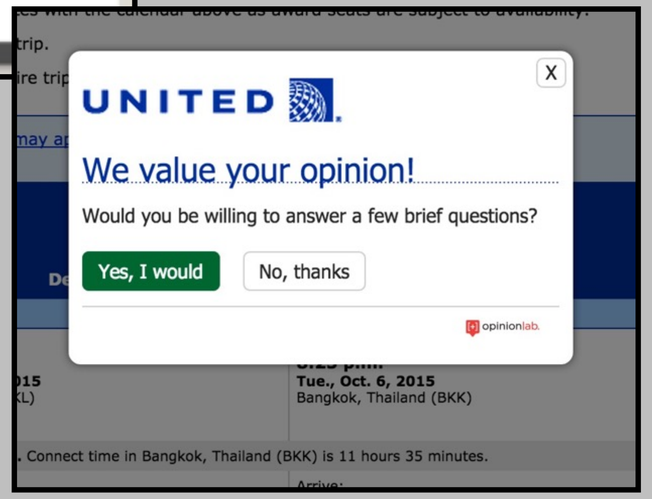
Surveys

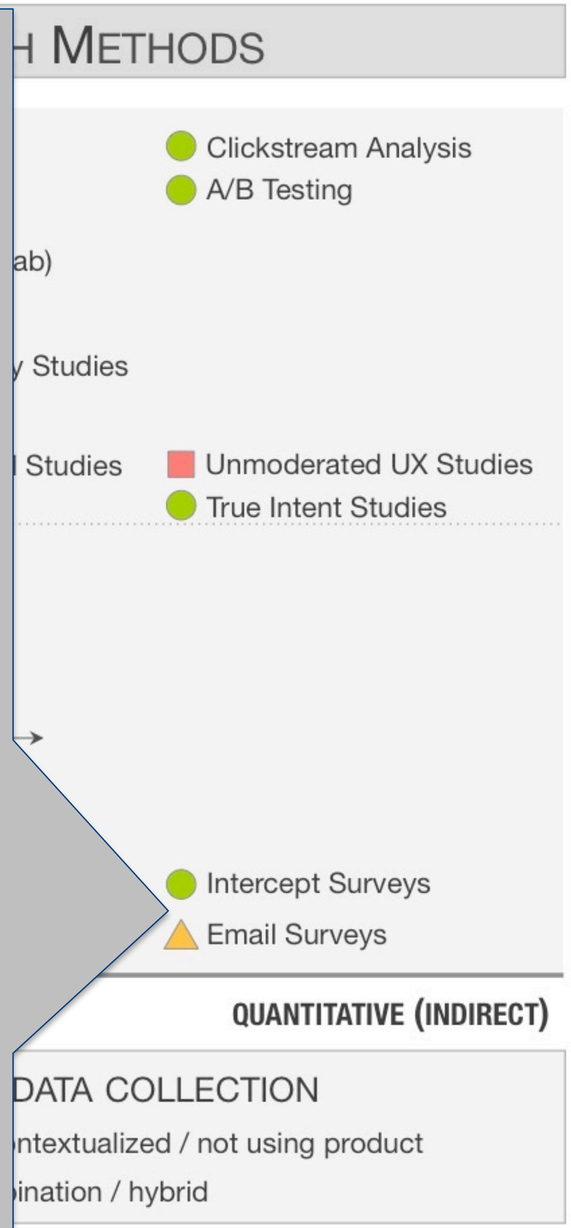
- Asking large numbers of users what they think in a structured way
- Email surveys: invite participants via email
- Intercept surveys: randomly invite a percentage of users on a site or using an app for their opinion about the site/app











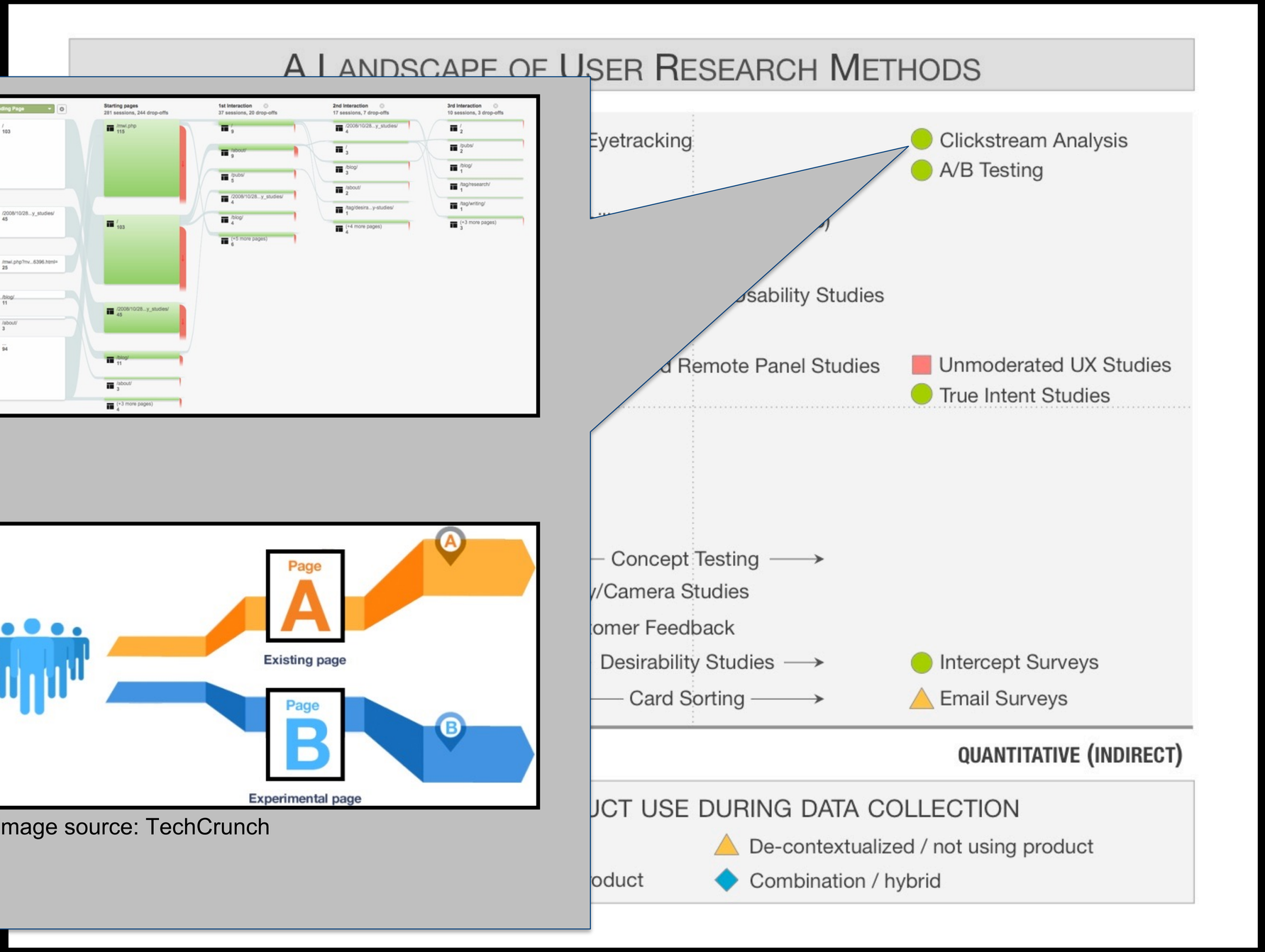
▲ De-contextualized use or ● Natural use (intercept)

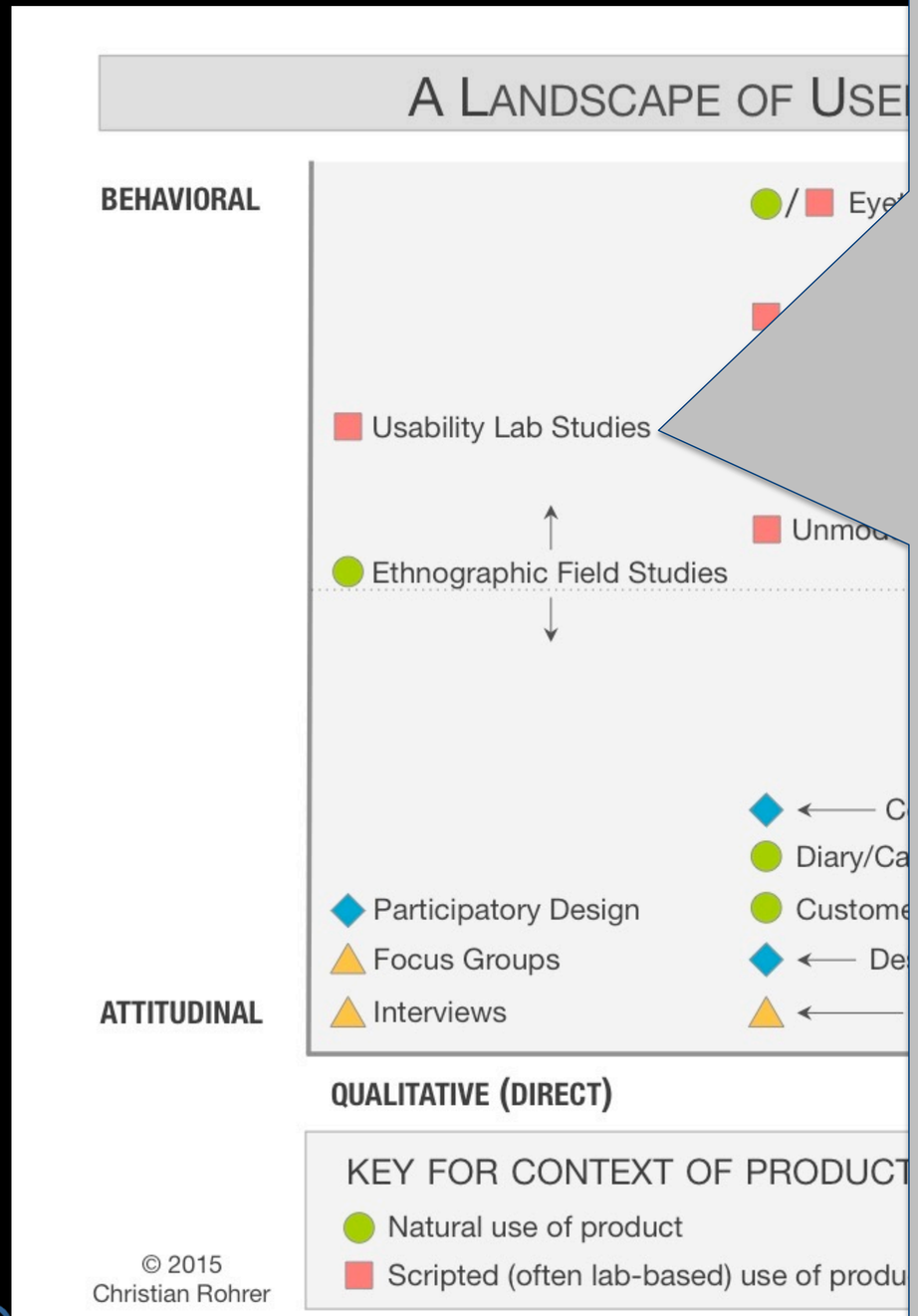
Clickstream Analysis & A/B Testing

- Analysis of data stored in logs (web or SW telemetry) on what users click
- A/B Tests: give a random sample of users alternative version of website design; compare logs of behavior to current site design

● Natural use

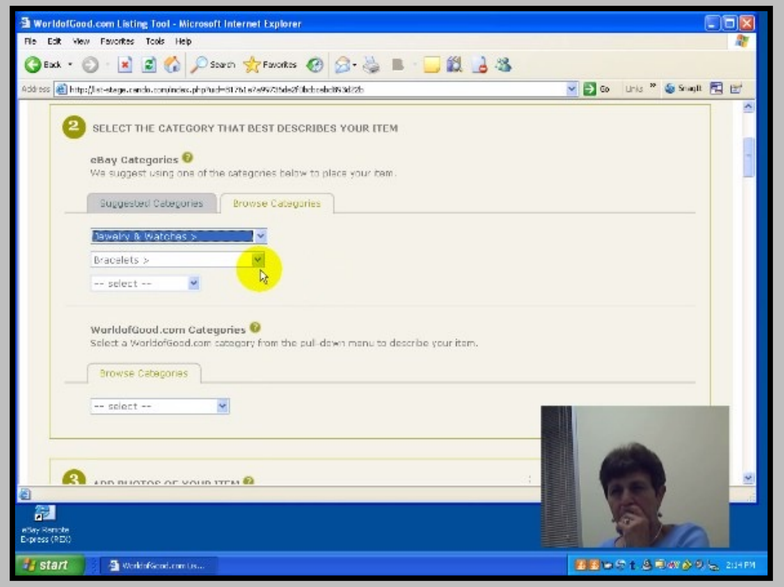
Image source: TechCrunch





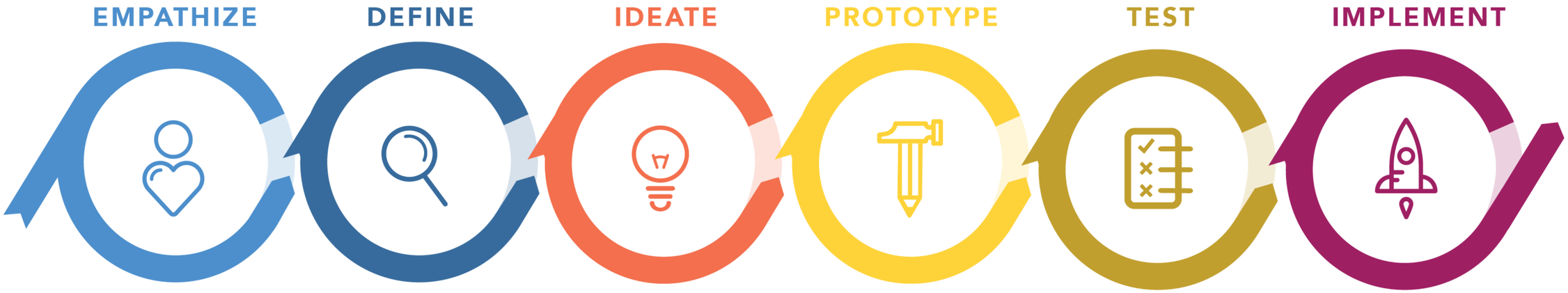
Usability Lab Studies

- Single-user using product/prototype
- Lab-based
- Usually scripted
- Behavior > Attitude



Scripted use (and in lab typically)

Time: The 4th Dimension



Generative (strategic)

Goal: Inspire new ideas; discover opportunities; inform strategy

What shall we do?

- Ethnographic field studies
- Competitive analysis
- Market segmentation
- Design Thinking
- Develop & test concepts
- Feature/task analysis

Formative (optimizing)

Goal: Understand user goals and tasks; Improve/refine the design; reduce execution risk

How shall we do it?

- Card Sorting
- Participatory design
- Usability inspections
- Iterative design & testing
- Desirability studies
- Usability (lab) studies

Summative (assessing)

Goal: Measure or compare against self or competition; feed back into strategy

How well did we do?

- Surveys (CSAT, Loyalty, NPS)
- Online UX Assessments
- Usability benchmarks
- Web Analytics
- Live (A/B) Testing
- Qualitative insights

XD strategy

BENEFITS OF TEACHING

- Design professionals benefit from this (sometimes a lot)
- Three classes of research generally match the Design Process (double-diamond, stage-gate, even agile or lean UX)
- Good for other research colleagues and data analysts to "nerd out " with

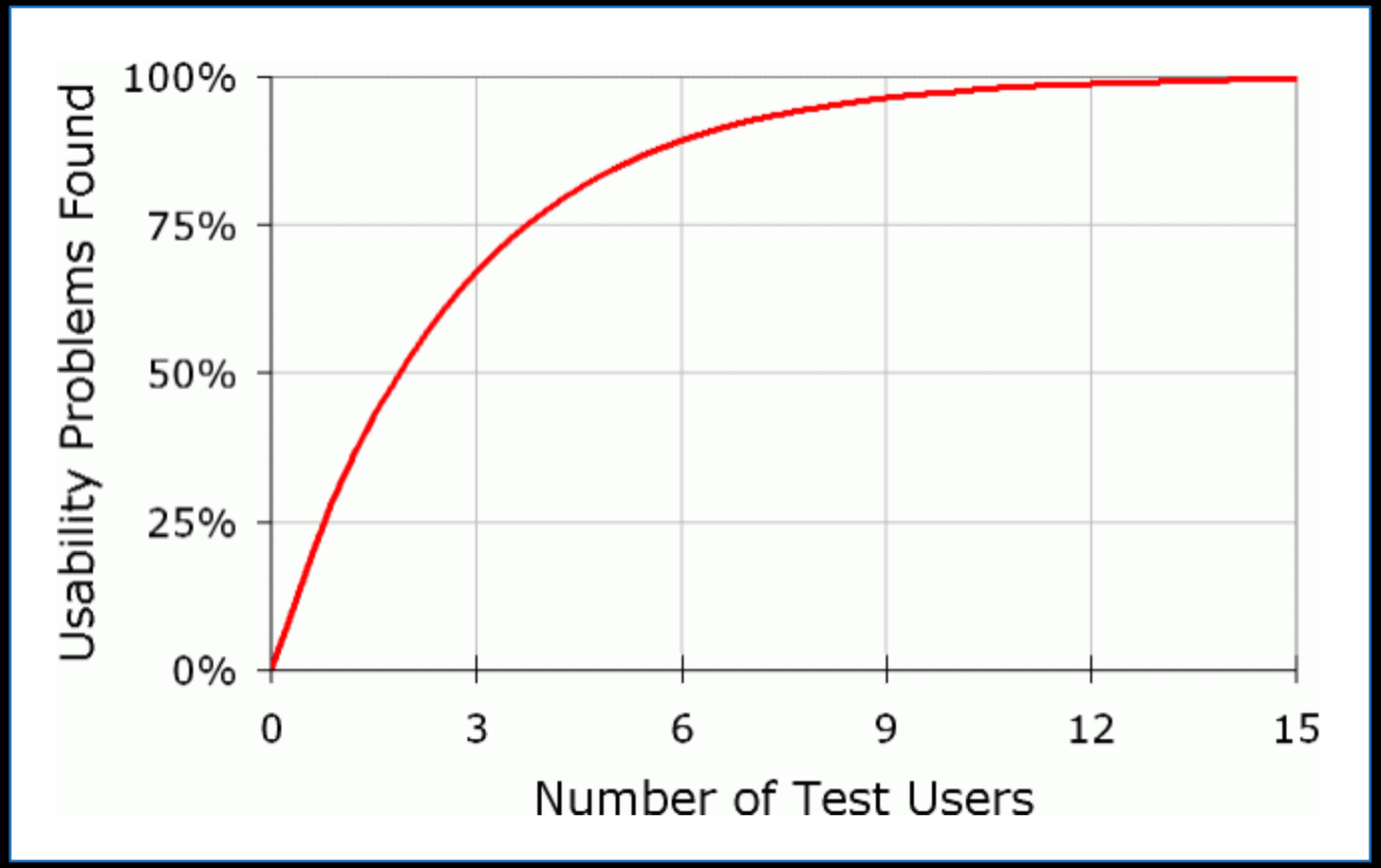
PITFALLS OF TEACHING

- Most decision-makers are not interested in being taught
- "Are you here to teach or get stuff done?"
- It's usually better for stakeholders to learn by doing; may be true for some in design as well

STRATEGIES I'VE USED IN THE PAST (IN THIS ORDER, FOR BETTER OR WORSE)

1. Teach Everyone About Research Methods
2. Sell the Benefits of Qualitative Research
3. Conduct both Qualitative and Quantitative Research
4. Develop Your Own Metrics
5. Lead With the Design Process

HAVE YOU EVER USED THIS GRAPHIC BEFORE?



Source: Nielsen Norman Group

"You only need 5 users …"

MANY THINK SIMPLISTICALLY:

"BIG NUMBERS GOOD.
SMALL NUMBERS BAD."

"Sorry, but I'll never make a decision based on research with just 5 people."
–Product Management

LET'S RETURN TO THE EARLY 2000s

PEOPLE WERE FIGHTING ABOUT "HOW MANY USERS WERE ENOUGH…"

"You only need to test with 5 users… [which finds] 85% of the usability problems."
-Jakob Nielsen

"This is the 'parabola of optimism.'" –Jared Spool

"[In CUE,] 70% of the usability findings … were unique."-Rolf Molich

"Identify and fix as many issues as possible and verify the effectiveness of these fixes." –M. Medlock

THE "RITE METHOD" PAPER (2002) WAS GROUNDBREAKING

- Although the paper was originally rejected by CHI, it went on to be one of the most influential and impactful methodological contributions of the early 2000s. Some of my takeaways were:
  - Finding and fixing problems is the top purpose of usability studies
  - Traditional validity, reliability and statistical power are less important than trying a new solution and testing it in the same study
  - Some classes of findings don't require multiple observations to be acted upon
- Major benefit of this kind of qualitative: The Impact Ratio (Found:Fixed problems)

Using the RITE method to improve products; a definition and a case study

Michael C. Medlock, User Testing Lead, Microsoft Games Studios (mmedlock@microsoft.com)
Dennis Wixon, Usability Manager, Microsoft (denniswi@microsoft.com)
Mark Terrano, Game Designer, Ensemble Studios (mterrano@EnsembleStudios.com)
Ramon L. Romero, User Testing Lead, Microsoft Games Studios (ramonr@microsoft.com)
Bill Fulton, User Testing Lead, Microsoft Games Studios (billfu@microsoft.com)

ABSTRACT

This paper defines and evaluates a method that some practitioners are using but has not been formally discussed or defined. The method leads to a high ratio of problems found to fixes made and then empirically verifies the efficacy of the fixes. We call it the Rapid Iterative Testing and Evaluation method – or RITE method. Application to the tutorial of a popular game, Age of Empires II, shows this method to be highly effective in terms of finding and fixing problems and generating positive industry reviews for the tutorial.

INTRODUCTION

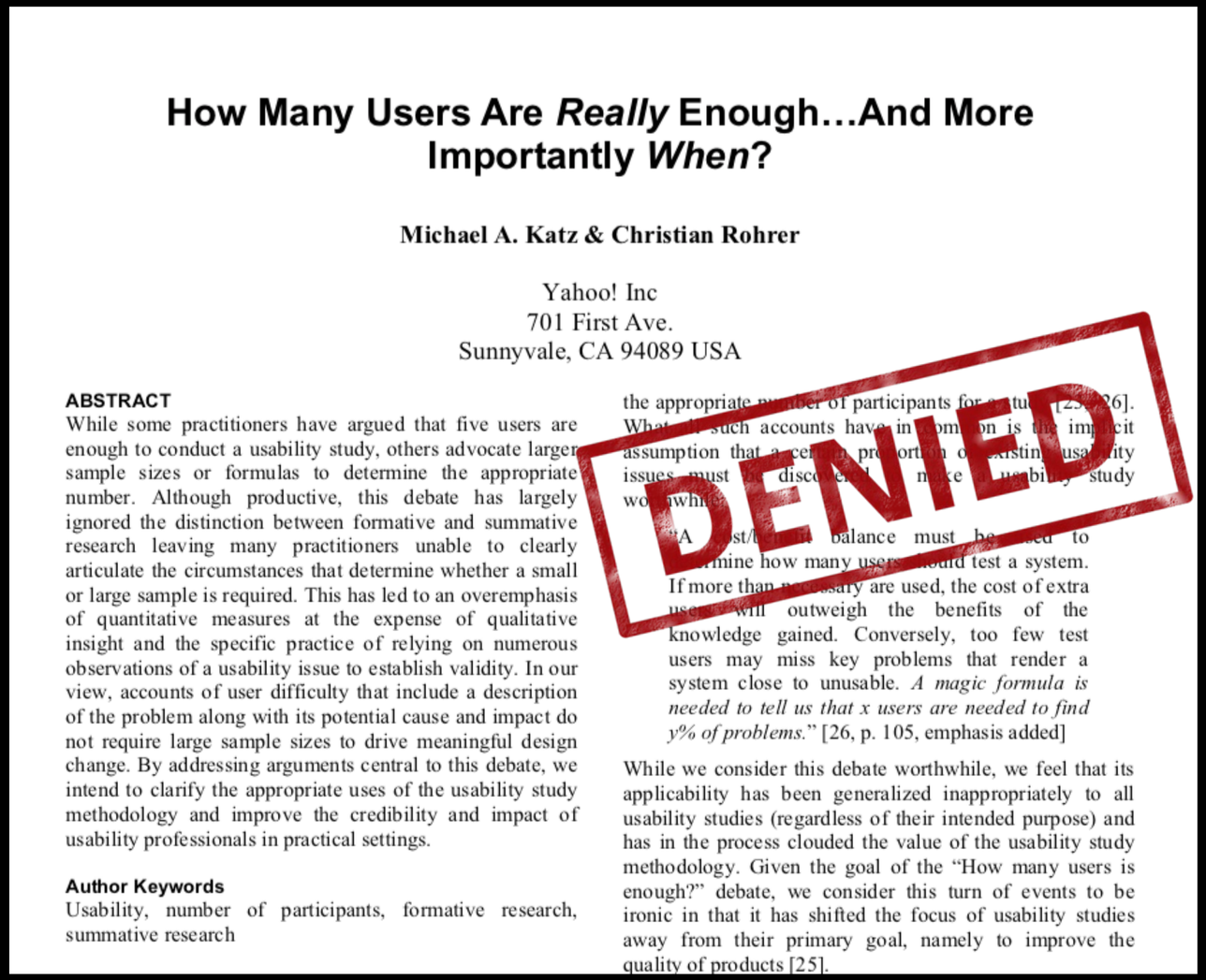
Traditionally the literature on sample sizes in usability studies has focused on the likelihood that a problem will be found [11, 12, 13, 16, 17, 18, 20]. This literature suggests:

- Running zero participants identifies zero problems.
- The more participants used, the fewer new problems are discovered.
- That calculating the number of participants needed to uncover "enough" problems can be done via a formula based on the binomial probability distribution –but that this number will vary depending on what the experimenter sets as the likelihood of problem detection. It is important to note that this calculation is based on the assumption that the experimenter might see the problem at least once. For example,
  - Observing 4-5 participants will uncover approximately 80% of the problems in a user interface that have a high likelihood of detection (0.31 and higher) [10, 11, 18].
  - Problems in a user interface that do not have a high likelihood of detection (for whatever reason) will require more participants to detect [16, 20].

When the researcher is interested in problems that have a high likelihood of detection, the suggestion has been made that it is more efficient to test with 4-5 users and test more often compared to running fewer, large-sample studies [11, 13]. Depending on the goals and context of the test, there are situations in which running even fewer than 4-5 participants is appropriate and more efficient. Lewis [11] noted that as long as the likelihood of problem detection was very high (0.50 and higher) that 87.5 % of these problems will be uncovered by at least 1 of 3 participants.

However, the usability literature on sample size has often not focused on what we as practitioners view as the primary goal of usability testing in an applied commercial setting: shipping an improved user interface as rapidly and cheaply as possible. We stipulate that when the determination has been made that a discount usability method is appropriate it is more important to get the team to fix problems and to determine the likelihood that a "fix" has solved a problem than to agonize over if every problem has been uncovered. The same likelihood of detection calculation based on the binomial probability distribution can be used for this purpose (and all the same caveats apply). It is noteworthy that relatively few studies have focused on the likelihood that change recommendations will be implemented [7, 15]. A small number of studies have focused on the magnitude of improvement in the user interface of a shipped product or tool, or the relative effectiveness of these improvements in affecting commercial sales or user efficiency [1, 5, 9, 19].

RITE INSPIRED US TO WRITE A PAPER AND SUBMIT TO CHI 2004…



How Many Users Are *Really* Enough…And More Importantly *When*?

Michael A. Katz & Christian Rohrer

Yahoo! Inc
701 First Ave.
Sunnyvale, CA 94089 USA

DENIED

ABSTRACT
While some practitioners have argued that five users are enough to conduct a usability study, others advocate larger sample sizes or formulas to determine the appropriate number. Although productive, this debate has largely ignored the distinction between formative and summative research leaving many practitioners unable to clearly articulate the circumstances that determine whether a small or large sample is required. This has led to an overemphasis of quantitative measures at the expense of qualitative insight and the specific practice of relying on numerous observations of a usability issue to establish validity. In our view, accounts of user difficulty that include a description of the problem along with its potential cause and impact do not require large sample sizes to drive meaningful design change. By addressing arguments central to this debate, we intend to clarify the appropriate uses of the usability study methodology and improve the credibility and impact of usability professionals in practical settings.

Author Keywords
Usability, number of participants, formative research, summative research

the appropriate number of participants for a study [25, 26]. What all such accounts have in common is the implicit assumption that a certain proportion of existing usability issues must be discovered to make a usability study worthwhile.

"A cost/benefit balance must be used to determine how many users should test a system. If more than necessary are used, the cost of extra users will outweigh the benefits of the knowledge gained. Conversely, too few test users may miss key problems that render a system close to unusable. *A magic formula is needed to tell us that x users are needed to find y% of problems.*" [26, p. 105, emphasis added]

While we consider this debate worthwhile, we feel that its applicability has been generalized inappropriately to all usability studies (regardless of their intended purpose) and has in the process clouded the value of the usability study methodology. Given the goal of the "How many users is enough?" debate, we consider this turn of events to be ironic in that it has shifted the focus of usability studies away from their primary goal, namely to improve the quality of products [25].

The CHI 2004 Papers committee didn't like it, because we (correctly) pointed out that researchers created this mess themselves! You can have a copy here: **http://bit.ly/2FUj682**

So, we simplified the paper to an article for *User Experience*, Volume 4, Issue 4, 2005:

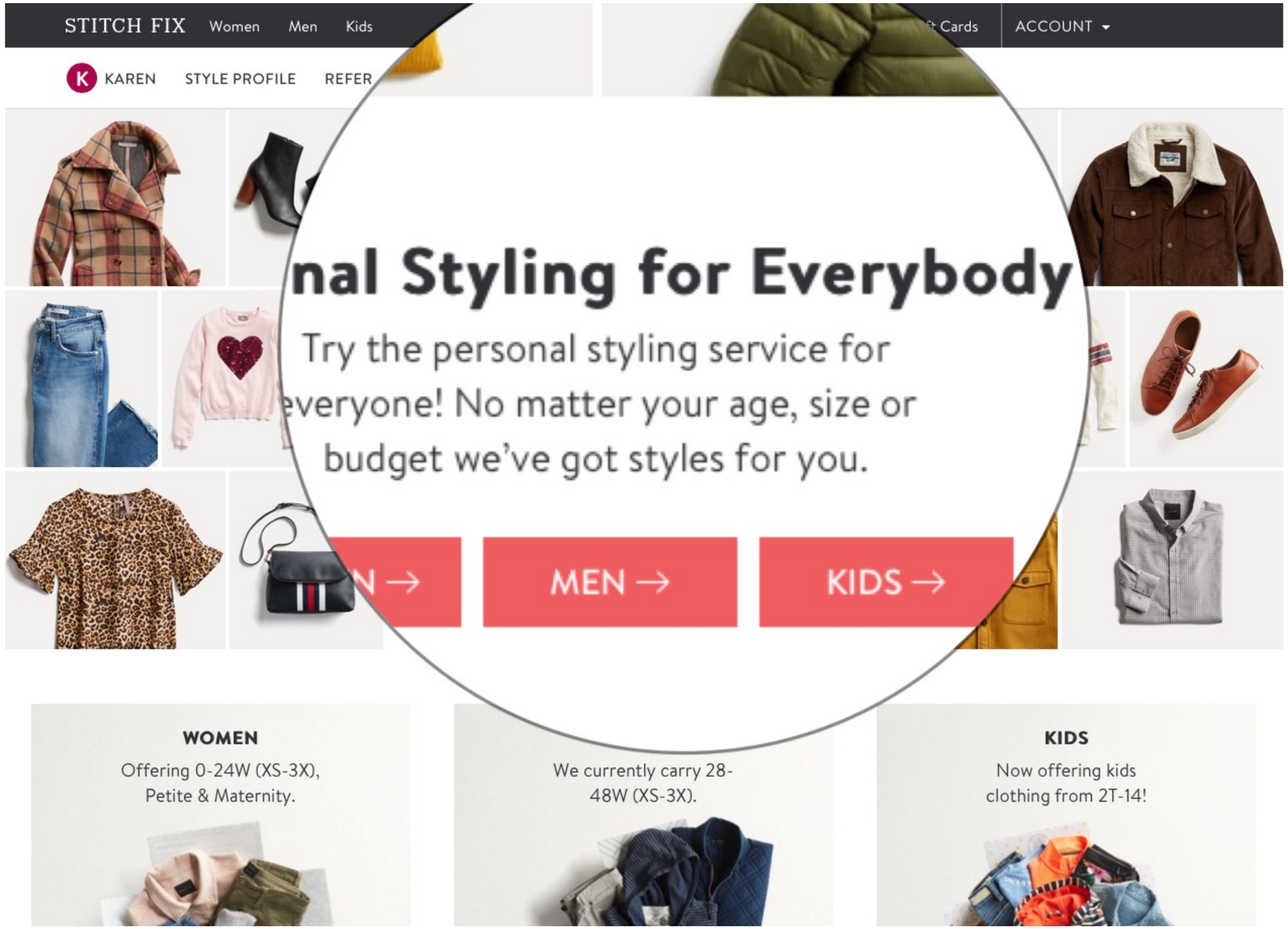
Criteria for a Valid Usability Issue

- The participant is representative of the target users for the product.
- The difficulty stemmed from a behavior that was reasonable, given the product domain.
- You can clearly describe the problem or difficulty.
- You can clearly describe the impact of the difficulty.
- You can provide a rational account of the cause of the problem.

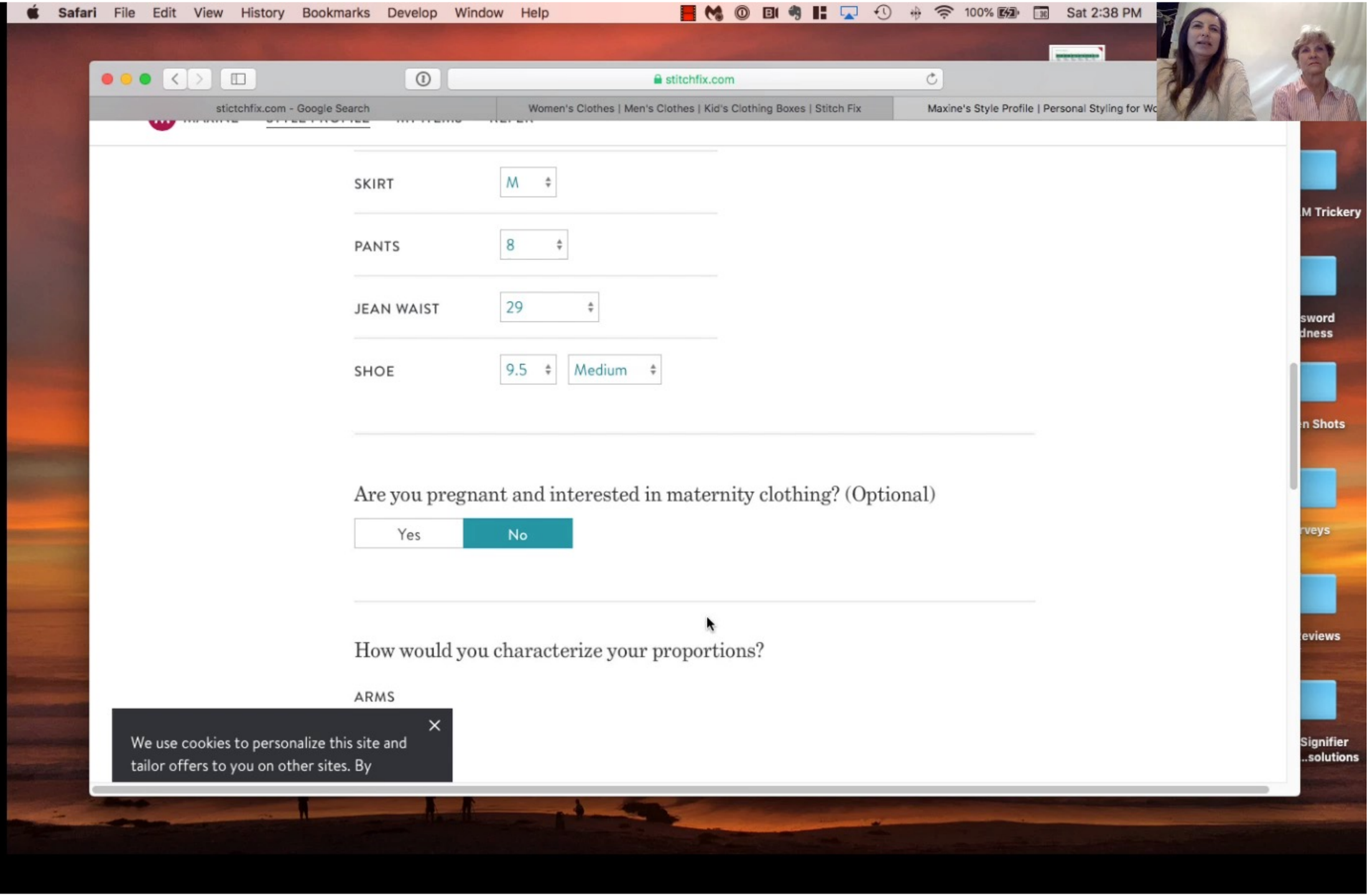
LET'S SEE AN EXAMPLE

STITCHFIX.COM

StitchFix.com: a styling service for … "everybody"?

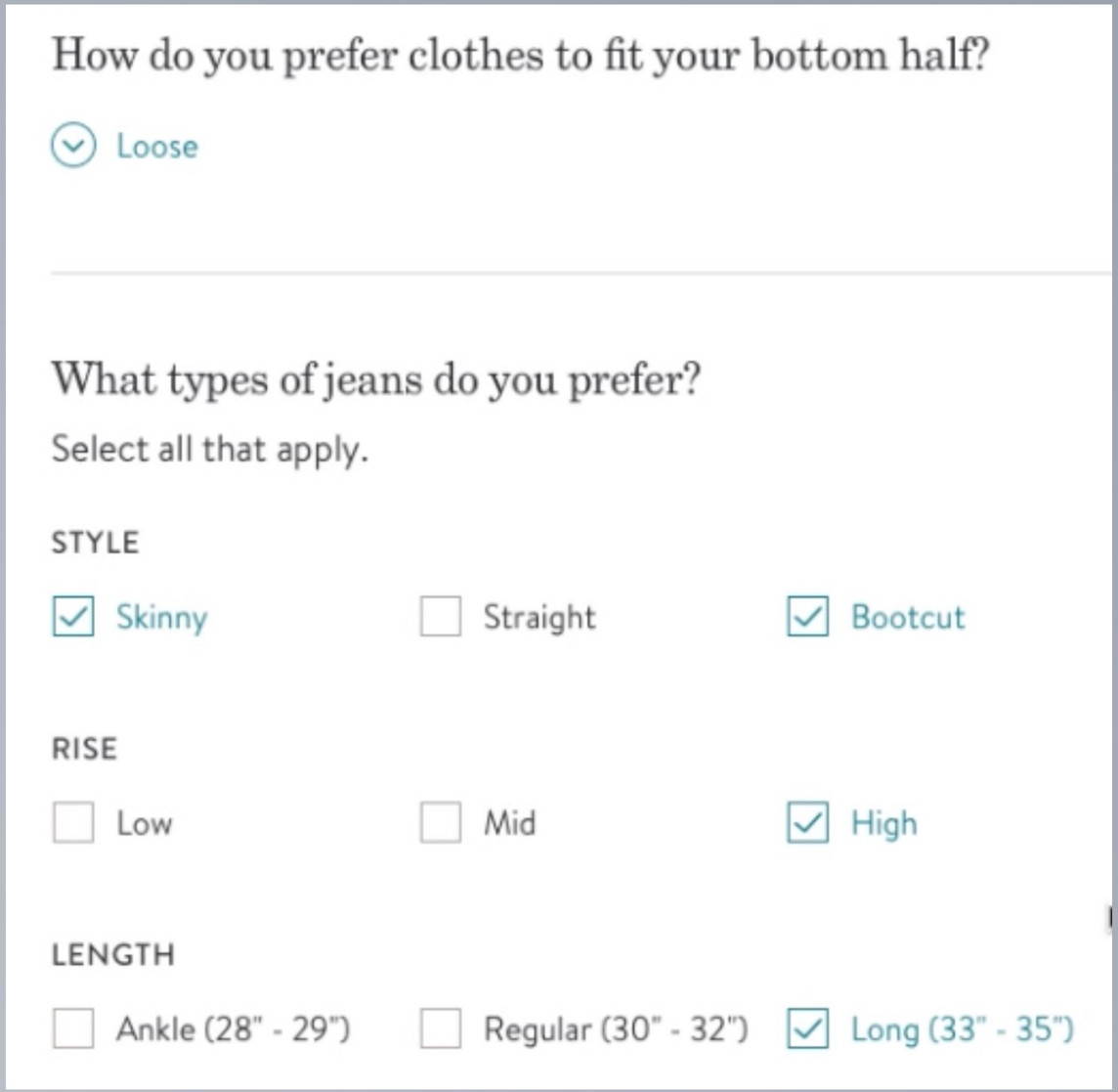


STITCHFIX.COM: MIDDLE-AGED FEMALE (IN CO-DISCOVERY STUDY) INTERESTED IN NEW FASHIONS, FILLING OUT THE STITCHFIX STYLE PROFILE…


Safari File Edit View History Bookmarks Develop Window Help
100% Sat 2:38 PM
stitchfix.com
stictchfix.com - Google Search
Women's Clothes | Men's Clothes | Kid's Clothing Boxes | Stitch Fix
Maxine's Style Profile | Personal Styling for Wo
SKIRT M
PANTS 8
JEAN WAIST 29
SHOE 9.5 Medium
Are you pregnant and interested in maternity clothing? (Optional)
Yes No
How would you characterize your proportions?
ARMS
We use cookies to personalize this site and tailor offers to you on other sites. By
M Trickery
sword
dness
n Shots
rveys
eviews
Signifier
..solutions

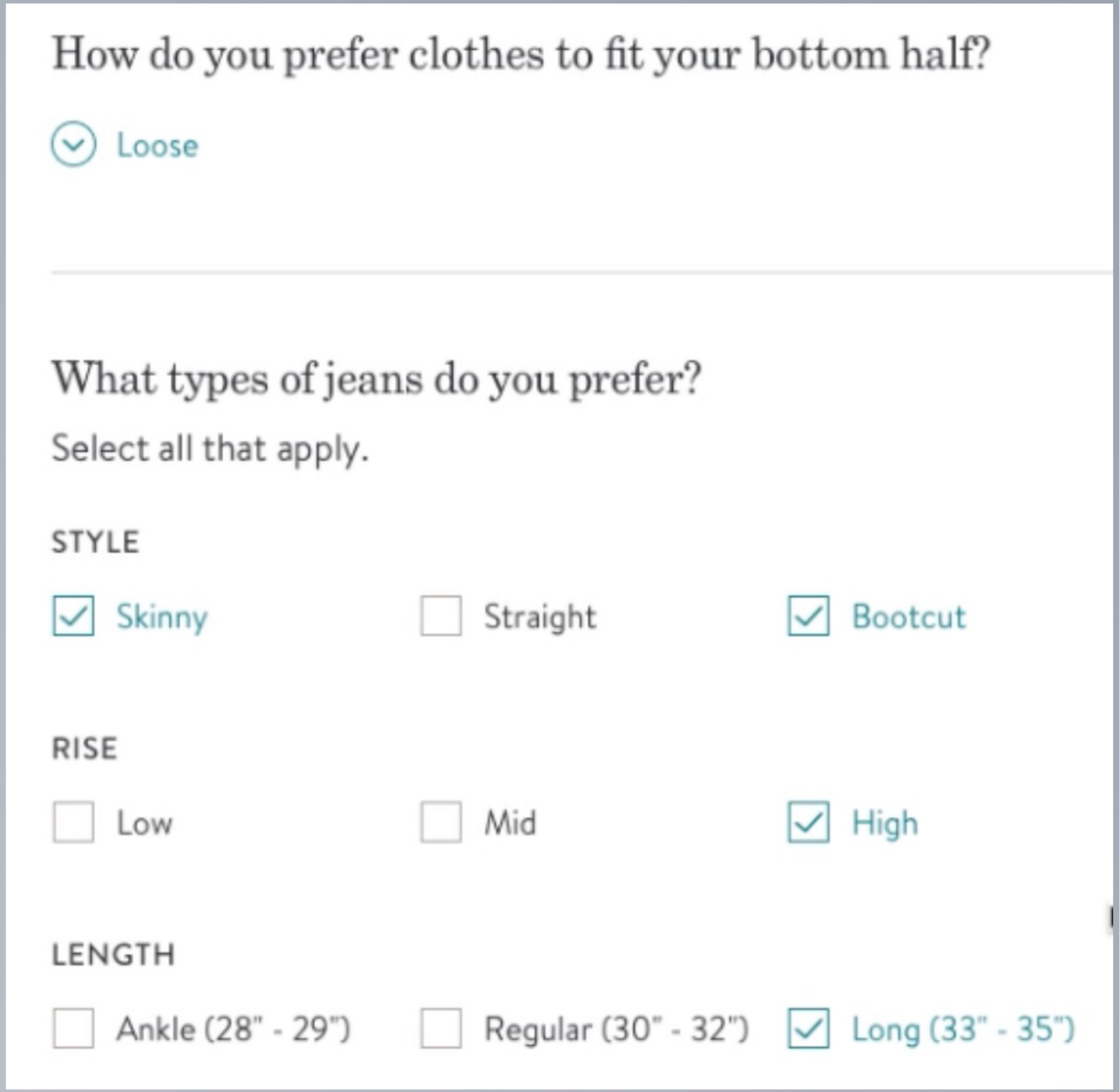
IS THIS A VALID ISSUE?

- Representative user(s)?
- Difficulty stemming from reasonable behavior?
- Clear description of the problem?
- Clear description of the impact?
- Rational account of the cause(s) of the problem?



HERE'S WHY IT'S A VALID FINDING:

- Target users include everyone interested in getting clothes this way
- **Priming:** Users saw single-select repeatedly before the multi-select
- **Form-filling behavior:** Typically, users don't read helper text unless they think they need it. They just focus on the titles and form elements.
- Visual design of single-select and multiselect are non-standard and even look visually similar.



HAVE YOU EVER REPORTED A QUALITATIVE RESEARCH FINDING IN THIS KIND OF WAY?

- "Four out of 10 participants said they liked this feature."

HOW ABOUT THIS?

- "Three of our 9 participants had difficulty with this feature in our study."

WHY?

EMPHASIZING NUMBERS IN A QUALITATIVE STUDY DOES MORE HARM THAN GOOD

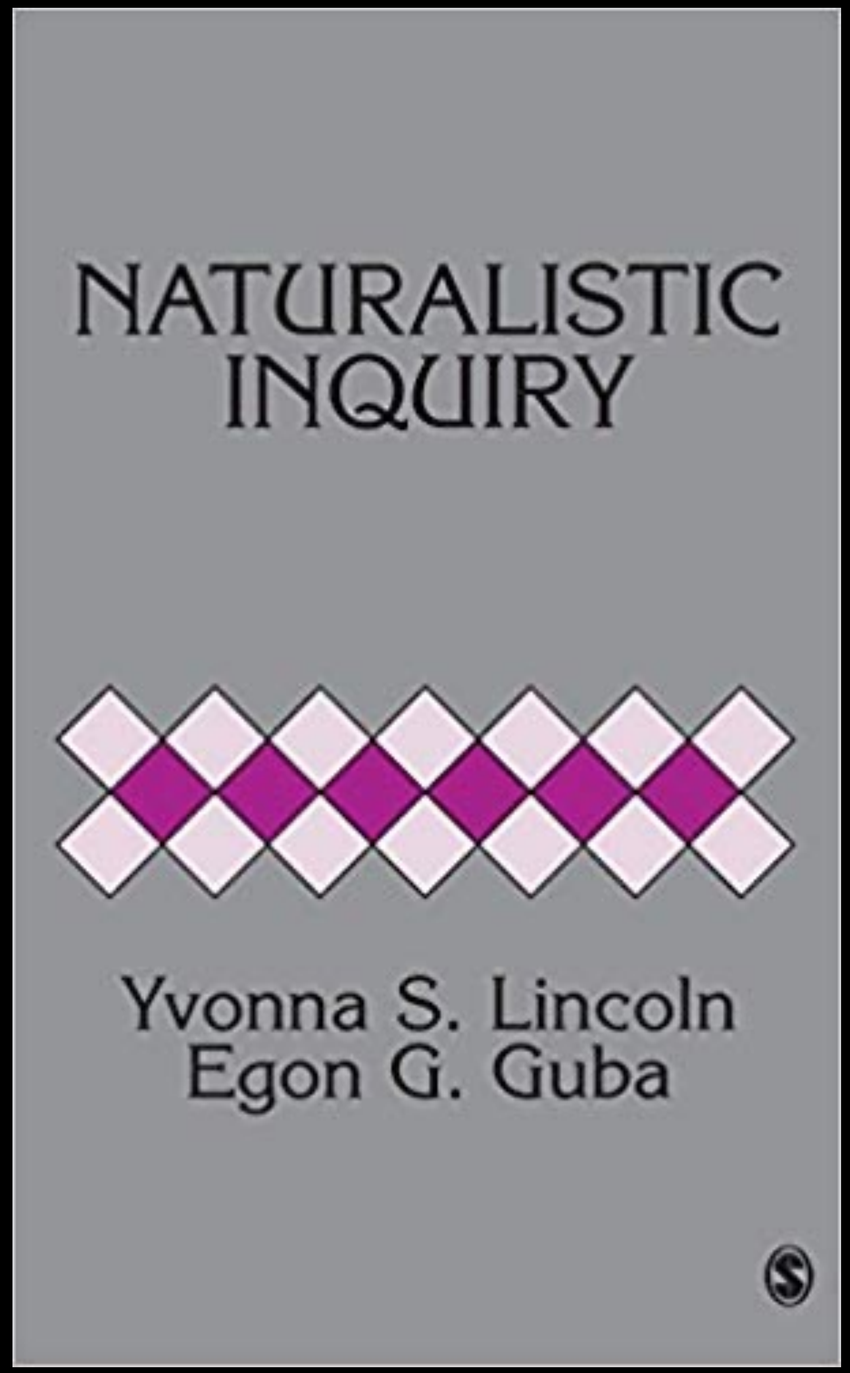
- Puts the focus on the number, not the quality of the experience
- It wrongly signals stakeholders to judge the validity based on the rules of quantitative research
- Qualitative research has different rules to determine its “validity”

WHAT DO WE MEAN BY (EXTERNAL) "VALIDITY?"

- Merriam-Webster: the quality of being well-grounded, sound, or correct the validity of an argument/theory. Example: "Other researchers have questioned the validity of the test results."
- Validity of a (statistical) measure (Wikipedia): the degree to which the tool measures what it claims to measure
- Qualitative validity (Cornell professor M.K. Trochim): when "soundness" of qualitative research is achieved
- **I suggest: "The extent to which the findings of a study can be generalized" (and therefore believed and acted upon)**

LINCOLN & GUBA (1985)

- Seminal work on qualitative research traditions (in contrast to quantitative)
- Qualitative research focuses on natural settings, is interpretive and is context-specific
- Field studies fit into this category



LINCOLN & GUBA'S 4 CRITERIA FOR QUALITATIVE RESEARCH, COMPARED TO QUANTITATIVE CRITERIA

Traditional Criteria for Judging Quantitative Research	Alternative Criteria for Judging Qualitative Research
internal validity	credibility
external validity	transferability
reliability	dependability
objectivity	confirmability

Put simply, you need:

- Breadth
- Depth
- Triangulation (method, source, analyst, theory/POV)
- Peer Debriefing
- Thick Description

TRANSFERABILITY

Transferability refers to the degree to which the results of qualitative research can be generalized or transferred to other contexts or settings. From a qualitative perspective **transferability is primarily the responsibility of the one doing the generalizing**. The qualitative researcher can enhance transferability by doing a thorough job of describing the research context and the assumptions that were central to the research. **The person who wishes to "transfer" the results to a different context is then responsible for making the judgment of how sensible the transfer is**.

Source: https://socialresearchmethods.net/kb/qualval.php

It is the job of the skilled design researcher to know when a finding is "transferable" to a domain outside of a usability study. It is NOT (usually just) about numbers and statistics.

ARE USABILITY STUDIES NATURALISTIC?

- Yes and no.
- We usually are interested in a "realistic" usage of the product
- But we script usability studies in a lab so that:
  1. We have consistency among participants (for "control"); or
  2. We can focus on the areas we are improving in the product at this time
- Being in a moderated usability study (lab or online) is convenient, but less natural

SO WHAT WORKS BETTER THAN SELLING OR EXPLAINING?

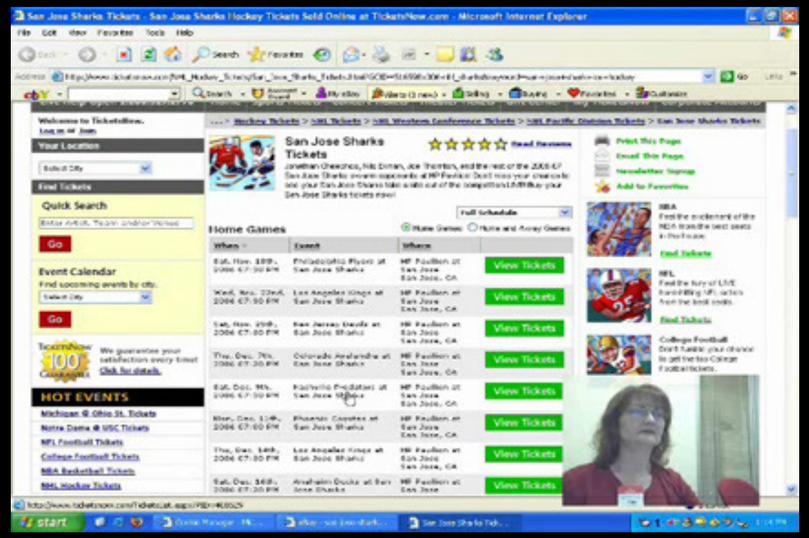
- Involvement.
- Study planning, session participation, even analysis.
- Seeing is believing.
- Even clips do better than explaining.

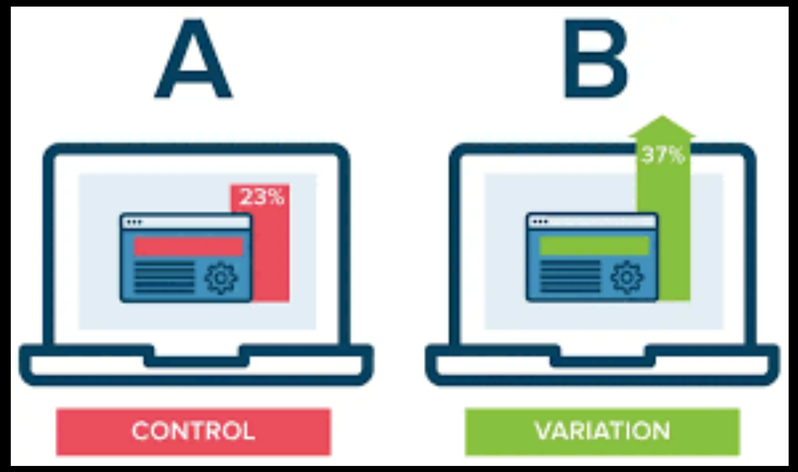
STRATEGIES I'VE USED IN THE PAST (IN THIS ORDER, FOR BETTER OR WORSE)

1. Teach Everyone About Research Methods
2. Sell the Benefits of Qualitative Research
3. Conduct both Qualitative and Quantitative Research
4. Develop Your Own Metrics
5. Lead With the Design Process

EBAY SEARCH WAS HORRIBLE IN THE EARLY 2000s, AND EXECS DIDN'T KNOW WHY

- Redesigned Search (aka "Finding")
- Coordinated a Usability Study with a simultaneous A/B Test
- Results from Qualitative and Quantitative result were combined
- We finally understood why we got the results we did in A/B testing
- The first time Execs had NO follow-up questions





IT'S 2011-2015, AND *THIS* IS MY COMPANY'S FLAGSHIP PRODUCT...

Consume... | Web Gate... | Web Gate... | IXDA 201... | MMS5.0_... | Certificat... | Professio... | My HR | true key -... | Yahoo | New T

https://www.yahoo.com

mcafee popup

Most Visited | Getting Started

Home | Mail | Answers | Groups | Flickr | Tumblr | Games | Live | Screen | Mobile | More

YAHOO!

Search Web

My Yahoo | Hi, Dmitri P | Mail

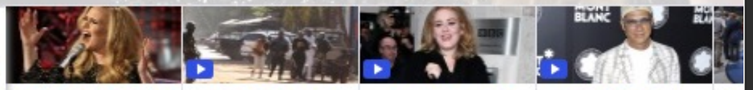
Show ad

- Mail
- News
- Sports
- Finance
- Weather
- Autos
- Fantasy
- Screen
- Dating
- Shopping
- Makers
- Parenting
- Health
- Style
- Beauty
- Politics
- Movies
- Travel
- Tech
- TV

More Yahoo Sites ›

Song-by-song review of Adele's new album, '25'

Adele's postsurgery voice sounds better than ever on what many are saying is the highly anticipated album. Best track »



Adele's new album reviewed | Mali hotel siege | Adele releases new album | Apple exec in hot water

All Stories | News | Local | Entertainment | Sports | More

Sylvester Stallone Shows Off His Stunning Daughters 'Creed' Premiere

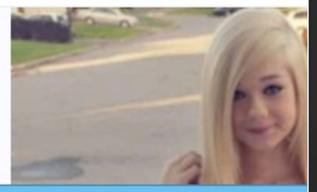
Sylvester Stallone's daughter Sistine, an aspiring model, looks a Minka Kelly.

ETonline

All Jaws Dropped When She Showed Up To The Prom

You Will Be Shocked To See What They Wore To The Prom!

ViralBoom Sponsored



Trending Now | Holiday Searches

1 Olivia Munn
2 Ronda Rousey
6 Phone plans
7 Valerie Letourneau

McAfee An Intel Company | LiveSafe – Internet Security

Your computer is at risk

Buy a subscription

Your trial subscription has expired. Don't leave your PC defenseless against the latest threats—buy a subscription now to stay protected.

Buy | Check for updates

Real-Time Scanning: On
Schedule and run scans

Firewall: On
View firewall and anti-spam settings

PC and Home Network Tools

McAfee Updates
Updates: Current
Check for updates and view settings

Your Subscription
Subscription: Expired
View details

Parental Controls

Home | Subscription | Help

Files checked during your last scan:

80,163

Security Report

Copyright © 2015 McAfee, Inc.

Threat Map | About

Mozilla Firefox | IXDA 2015 Keynote... | Photoshop - Shortcut | McAfee Securi... | Vis_Design_... | IXDA 2015 Keynote ... | Flash | Adobe Illustrat... | InDesign - Shortcut | CreativeCl... | Adobe Acrobat 8 ... | Google Chrome | iTunes | LWAPlugin... | Dropbox | CitrixOnline... | McAfee Security Sc... | Citrix Receiver | Acrobat Reader DC

L... | Visual_Design - Shortcut | Quick projects | D... | Yap | 2013_myA... | fb2 | My | Desktop Oct 2013

PROBLEM #1: THE UX WASN'T GREAT

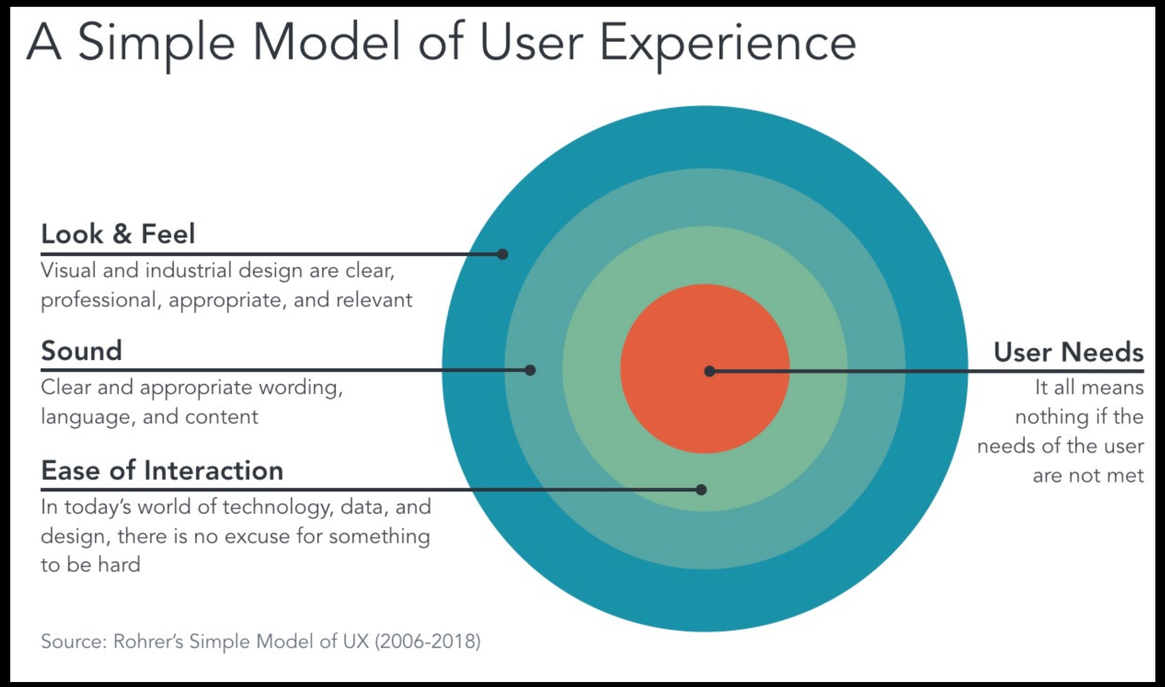
- "Hey, you know, this product makes a lot of money."
- "The **UI** isn't that bad, is it?"
- Few people actually like it – it was considered a necessity for the times we lived in

PROBLEM #2: PROVING THIS WAS HARD

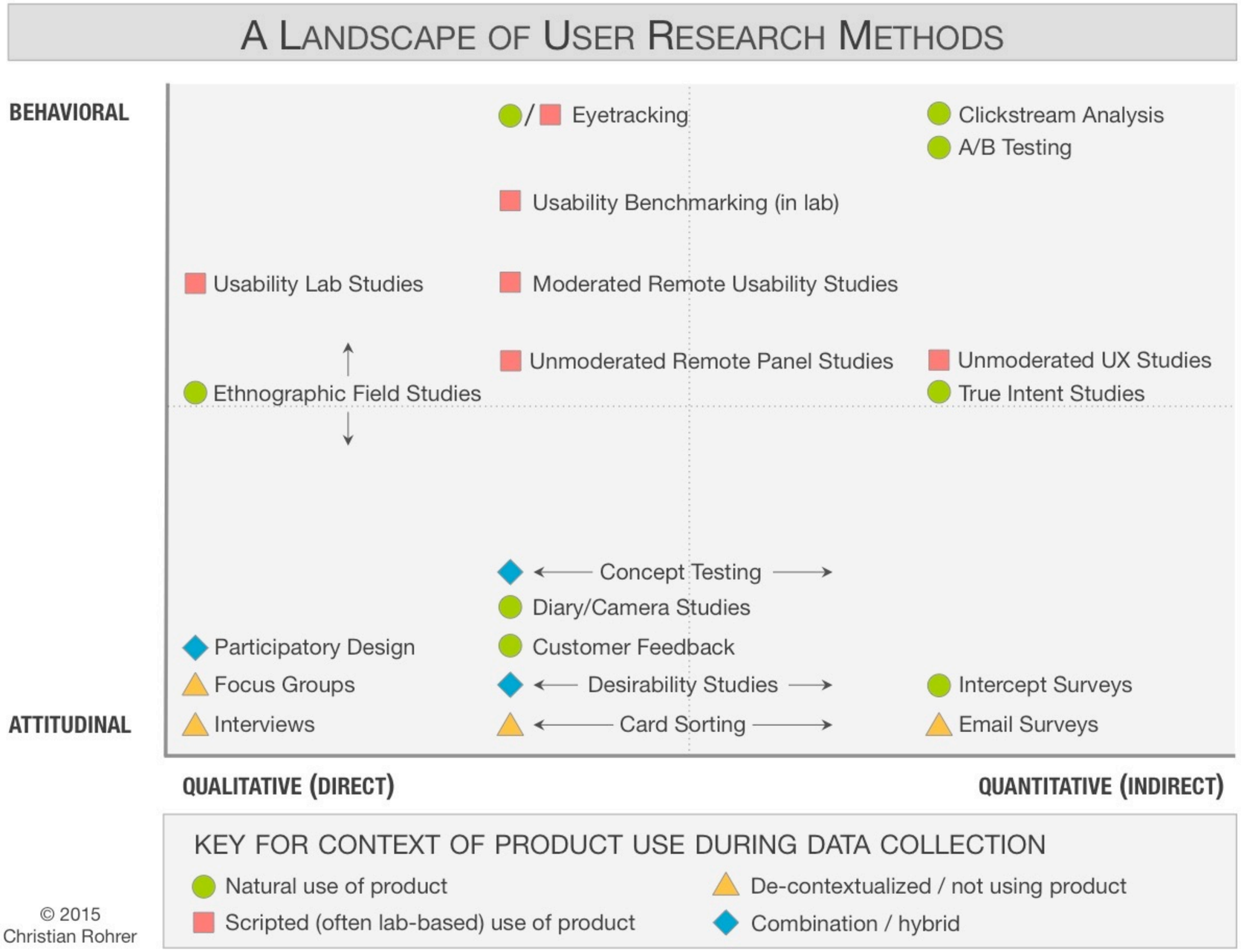
- Measuring UX on desktop software isn't easy
- New devices (mobile security) emerging
- Data and metrics were not really UX-related

SO, I DID SOME EDUCATING...

I EXPLAINED WHAT USER EXPERIENCE WAS…



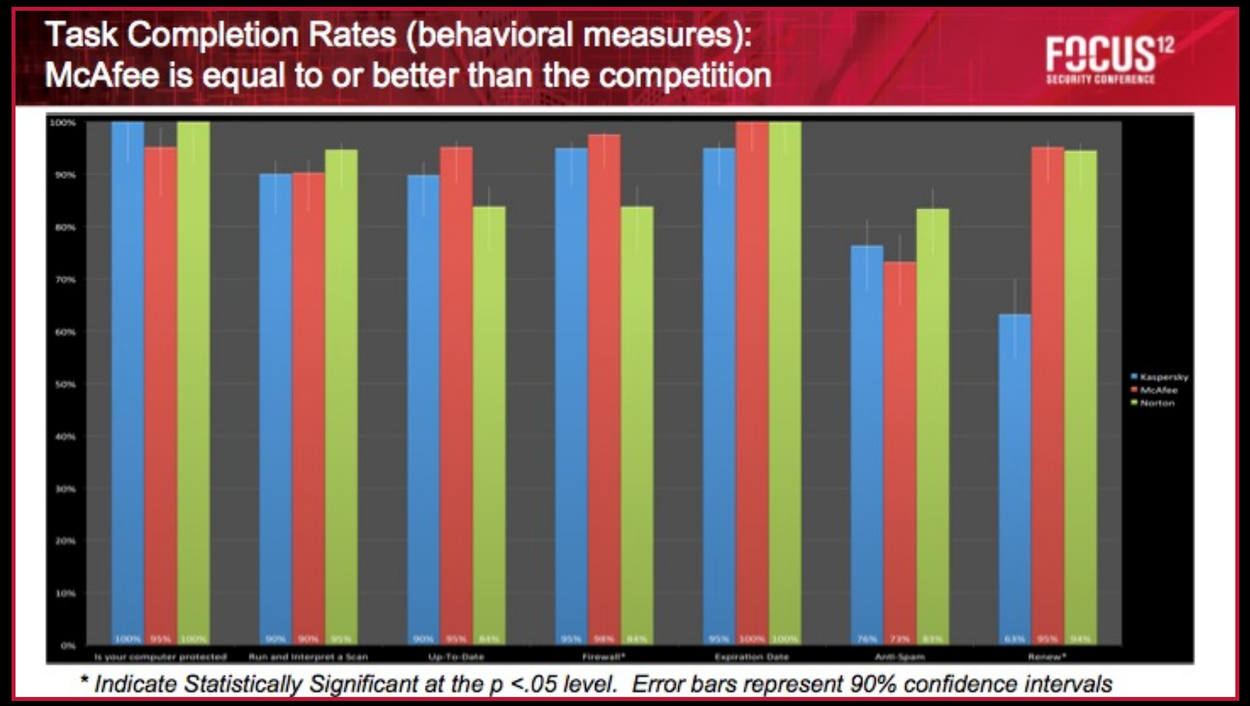
I TALKED ABOUT RESEARCH & DATA...

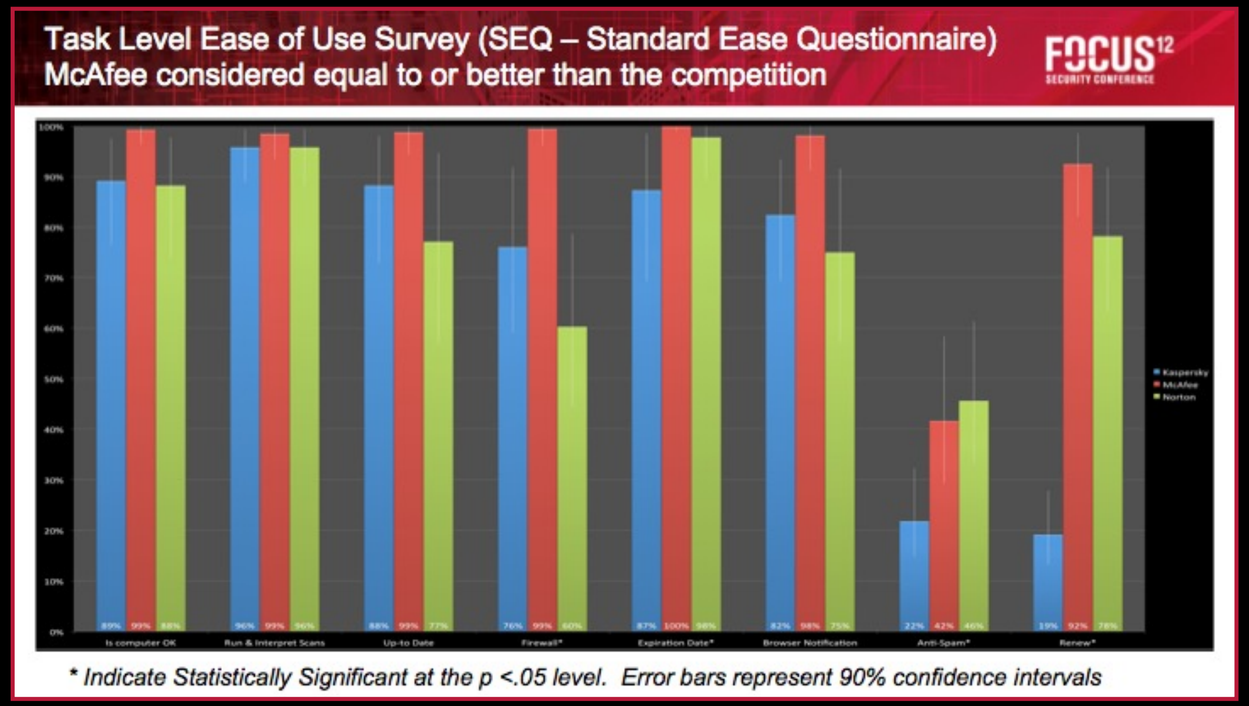


AND I GOT THE FUNDS TO MEASURE OUR UX

(TWO MONTHS AND $100K LATER...)

I HAD EMPIRICAL UX BENCHMARK RESULTS!





- Execs LOVED it, even though with N = 42, there were very few statistically significant differences
- Design liked having it, but would have been just fine with a 3 week qualitative study
- Takeaway: UX Benchmarks are not practical to conduct frequently or regularly

"QUANTIFICATION BIAS"

"The unconscious valuing of the measurable over the immeasurable."
- Tricia Wang, Technology Ethnographer

- "Quantifying is addictive."
- "It's very easy to throw out data that is not quantified."
- "But the future we need to predict is very often not quantified."

https://www.ted.com/talks/tricia_wang_the_human_insights_missing_from_big_data?language=en

"THICK DATA"

- "Thick and meaty data helps us understand the narrative of the human condition."
- Big Data + Thick Data = Complete Picture

SHOULD YOU DO BOTH QUALITATIVE AND QUANTITATIVE?

- If you have the time and money, then do it.
- Ensure that Qualitative and Quantitative mutually inform each other.
- Don’t make one better than the other. They have different jobs to do.
- You can cycle between qual-quant-qual-quant and get a lot.
- Don’t forget about Behavior vs. Attitudinal and Context Of Use dimensions.

STRATEGIES I'VE USED IN THE PAST (IN THIS ORDER, FOR BETTER OR WORSE)

1. Teach Everyone About Research Methods
2. Sell the Benefits of Qualitative Research
3. Conduct both Qualitative and Quantitative Research
4. Develop Your Own Metrics
5. Lead With the Design Process

HERE'S WHAT YOU HEAR A LOT OF WHEN WORKING WITH EXECUTIVES...

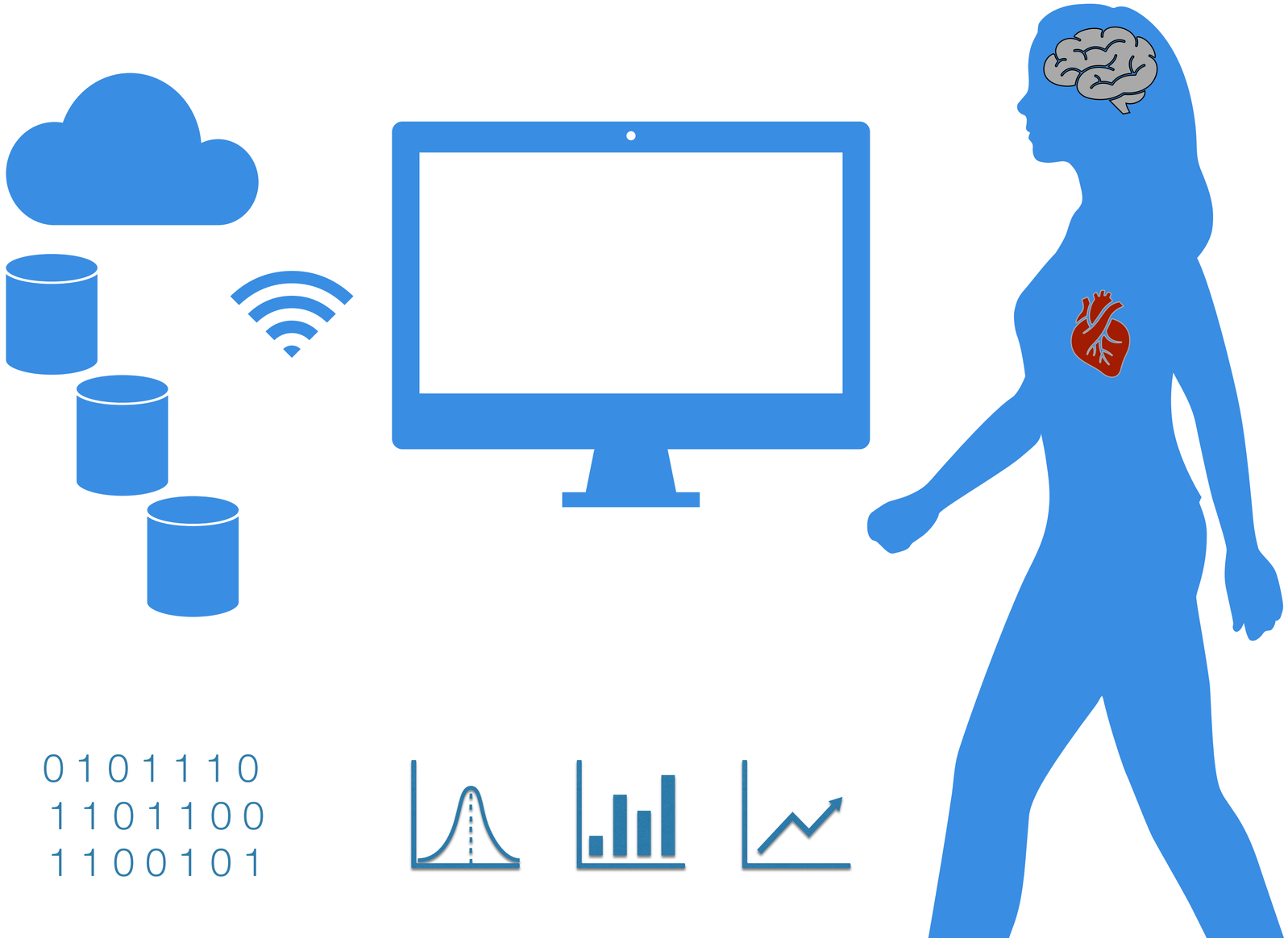
The insatiable demand for numbers…

- "You can't **manage** what you can't **measure**"
- "I need a **dashboard** to control my business"
- "How does our **NPS** compare?"
- "Invest in **data scientists** and **big data**"
- "I want to be more **scientific**"
- "Build, **measure**, learn"
- "The design needs to be **validated**"

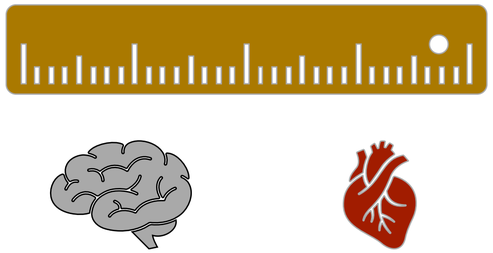
Conversion rates, downloads, NPS, customer reviews, A/B testing performance, uptime, abandon rate, analytics data, path/click flows

Numbers… Any numbers?

Q: Where is the "User Experience?"



Challenge: How can you measure THESE?



Numbers and metrics can come from…

Empirical studies

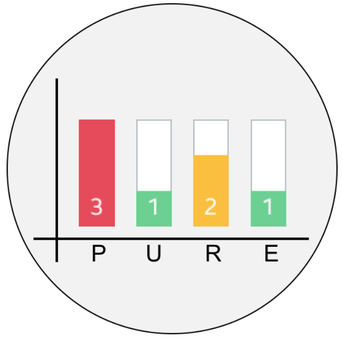
USABILITY BENCHMARKING

ONLINE TESTING

…and…

Analytical methods

HEURISTIC EVALUATIONS*



PURE

*(e.g,.severity ratings)

Measures of Ease through Analytical Methods using The PURE Method (Practical Usability Rating by Experts)

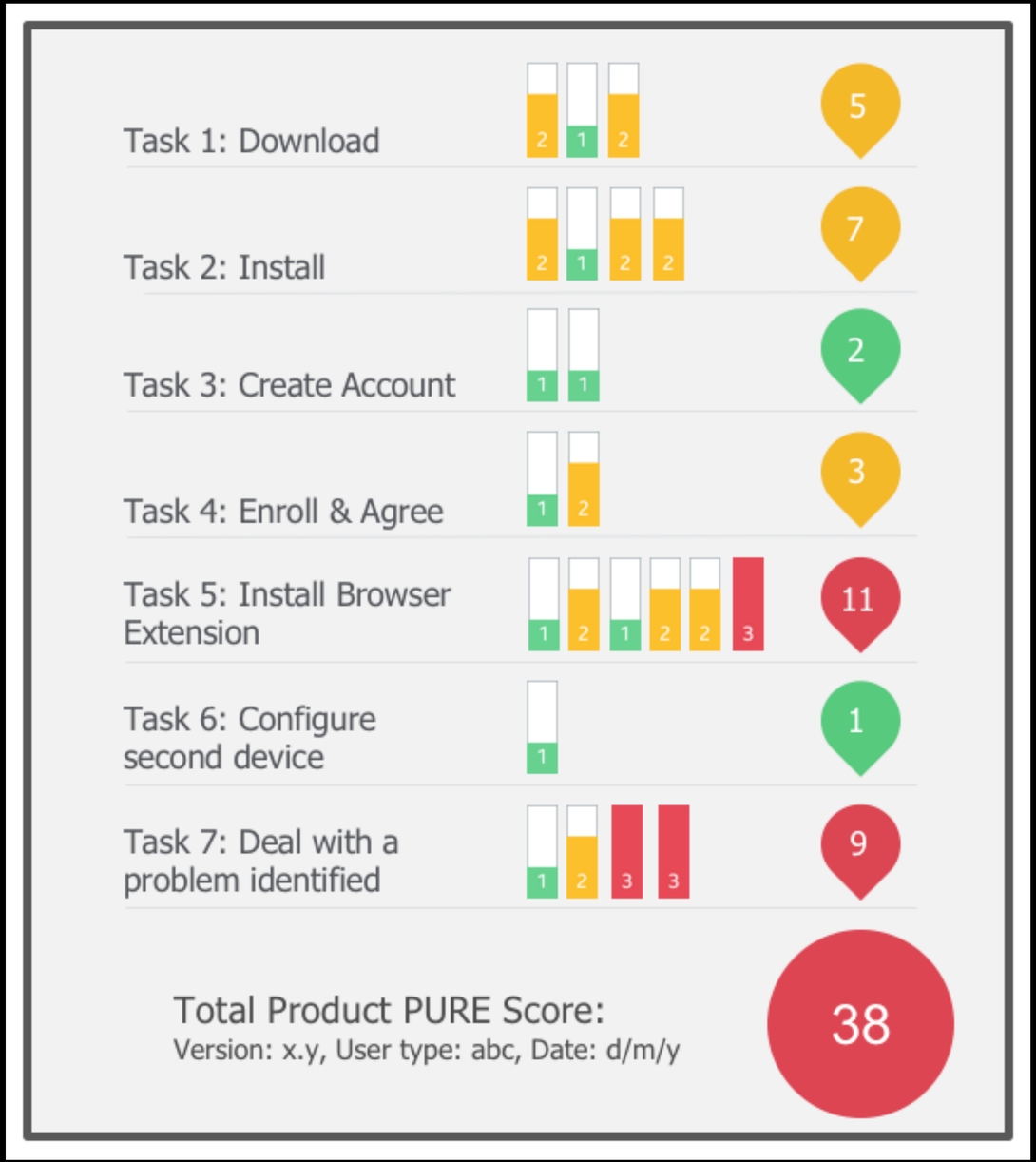
Published at CHI 2016, San Jose, CA USA

PURE: A WAY TO A "FRICTION METRIC"

PURE provides a cognitive load "score" based on :

1. A well-defined user type
2. A small set of "fundamental tasks"
3. The "performance" of those tasks (e.g., the Happy Path) for that user type, based on usability heuristics & UX design principles

Like in golf, smaller numbers are better, and green is good.



ANALOGY: IN PURE, WE JUDGE A SPECIFIC PERFORMANCE, AS IN SKATING & GYMNASTICS

- A panel of judges each silently rates a specific performance they are all witnessing
- A known rubric defines how much of a deduction results from a given mistake
- PURE rates every step, as if it were a "move"

The PURE step rubric is a simple 3 point scale:

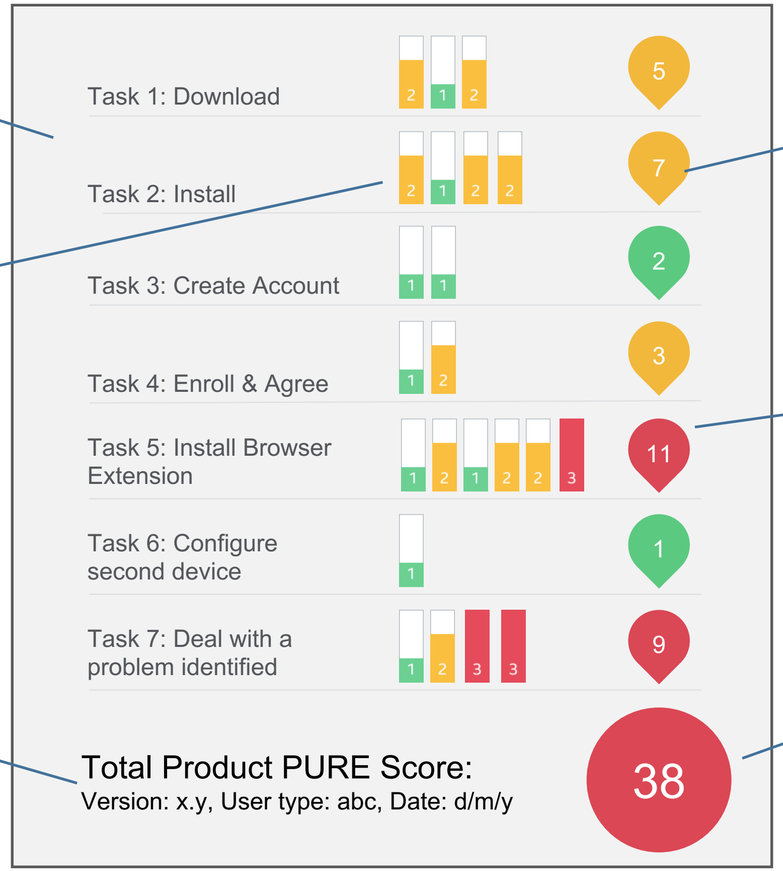
"Don't Make Me Think" ⟷ "The User will likely fail/bail"

PURE: an Ease of Use Scorecard of the "friction" for the **Target User** going through the **Happy Paths** of a product's handful of **Fundamental Tasks**:

Only the most fundamental tasks are scored

Both the length (number of bars) and ease of use (numbers/colors)of the tasks are quickly discerned

Version, target user type and date are documented for comparison later

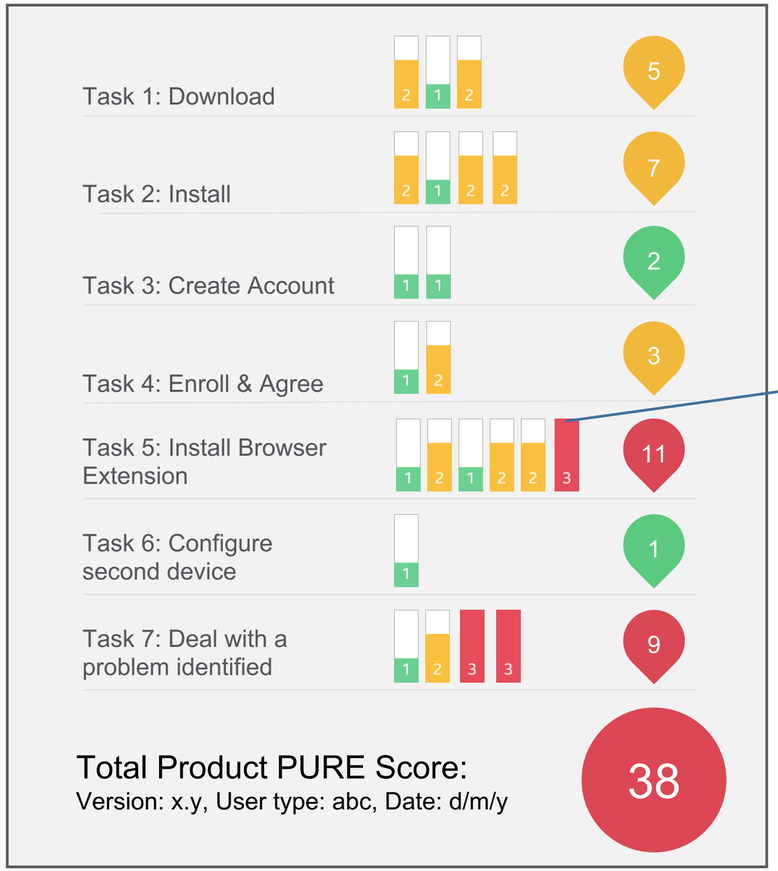


Each task gets a score and color. As in golf, green is good, smaller numbers are better

Behind every rating and score, there are helpful reasons for the scores, which drive fixes

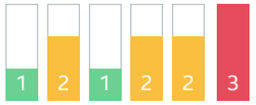
The whole product PURE score is driven by the task PURE scores

We can dig deeper on any given score



Behind every rating and score, there are helpful reasons for the scores, which drive fixes

Task 5: Install Browser Extension, Step #6



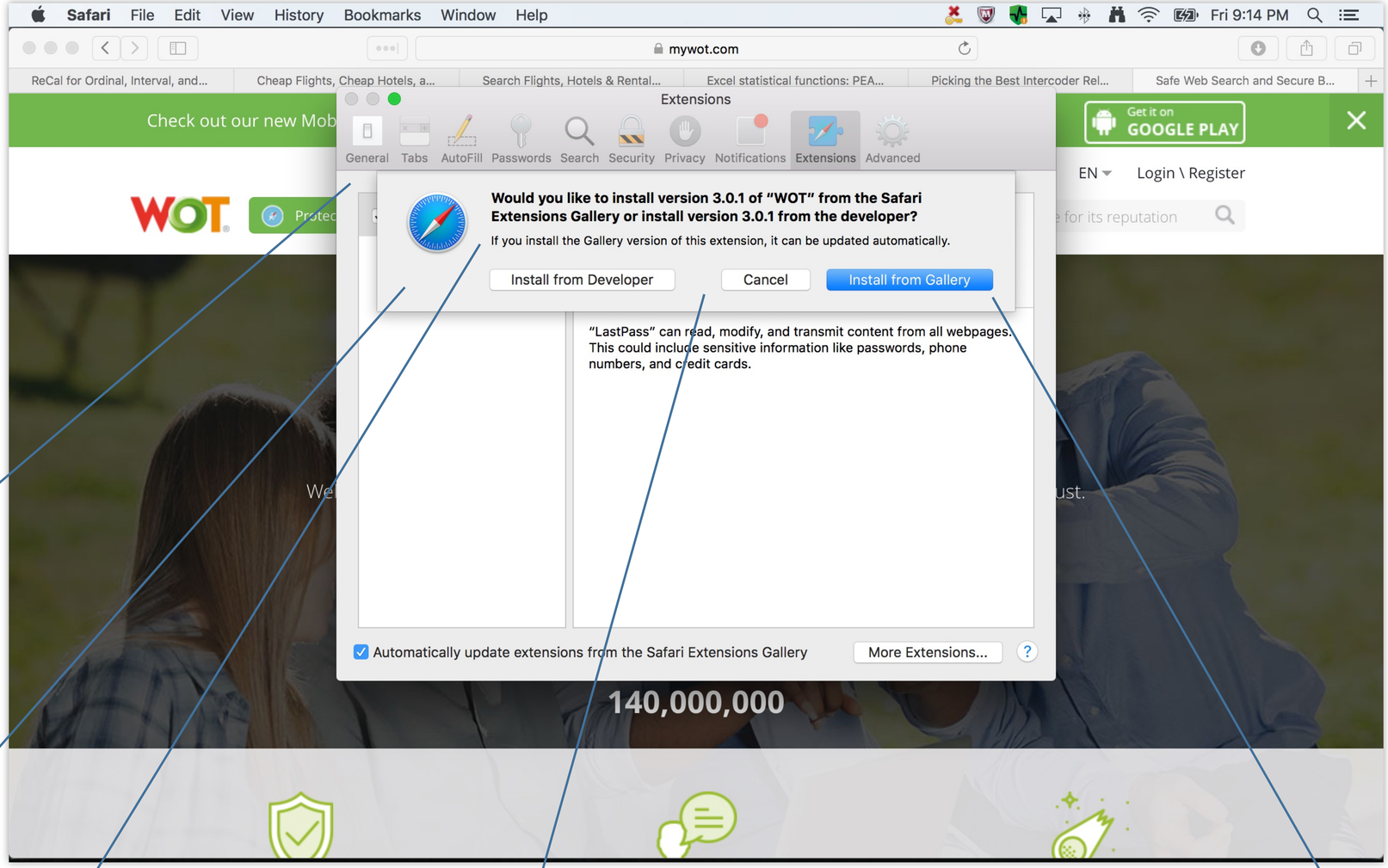
11

Why was this step a red 3?

This Extensions tab from Safari settings comes out of nowhere after the previous step (unexpected)

The dialogue box appears at the same time as the Safari settings tab, partially obscuring its content and masking the context the dialog is related to.

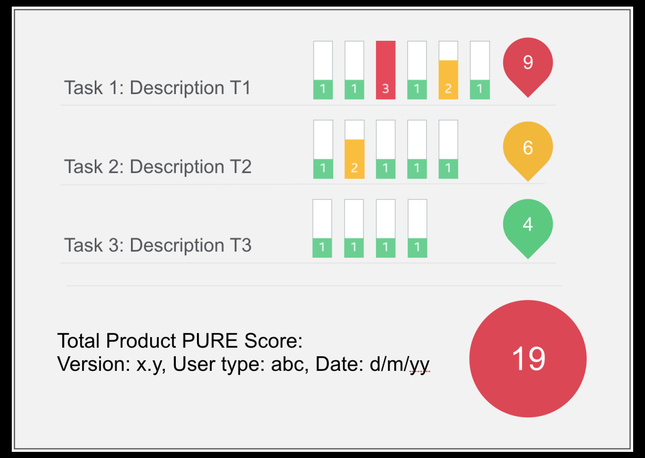
The language used here is difficult for the target user to fully understand without significant effort. (At least a benefit is explained, however.)

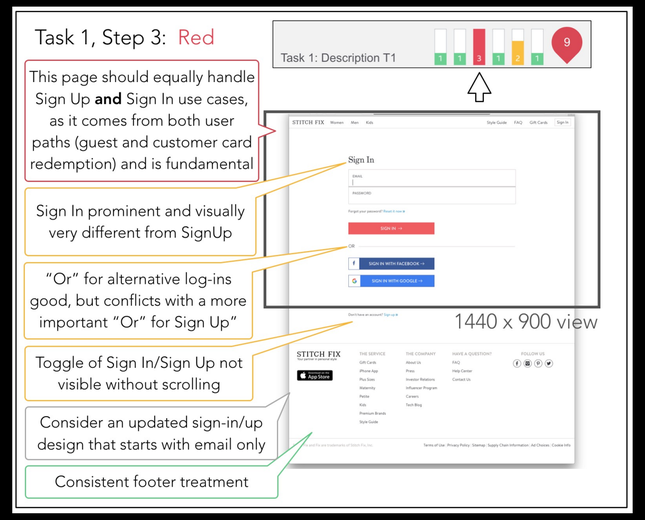


There are three choices, not uniformly spaced (so looks sloppy). Most problematic, it's likely not clear to this user type what "Cancel" does at this point without some cognitive effort.

After selecting the default button "Install from Gallery" both the dialogue and the Safari setting disappear and it appears not to have done anything (this issue is technically part of the next step)

A GOOD COMBO OF QUANT AND QUAL: THE PURE SCORECARD + STEP INSIGHTS DECK





- Quantification Bias provides face validity to PURE overall
  - Metrics are derived in a legitimate way (like gymnastics judgements or grading a paper based on a rubric)
- Raters must be experienced in qualitative research, knowledgeable in UX principles, skilled in being objective and open, and willing to score themselves to improve
- Scorecard + Insights can become a roadmap of what to fix or study further (highly applicable output)

WHERE DO THESE DESCRIPTIVE INSIGHTS COME FROM?

1 FROM THE RATERS' PREVIOUS QUALITATIVE RESEARCH EXPERIENCE

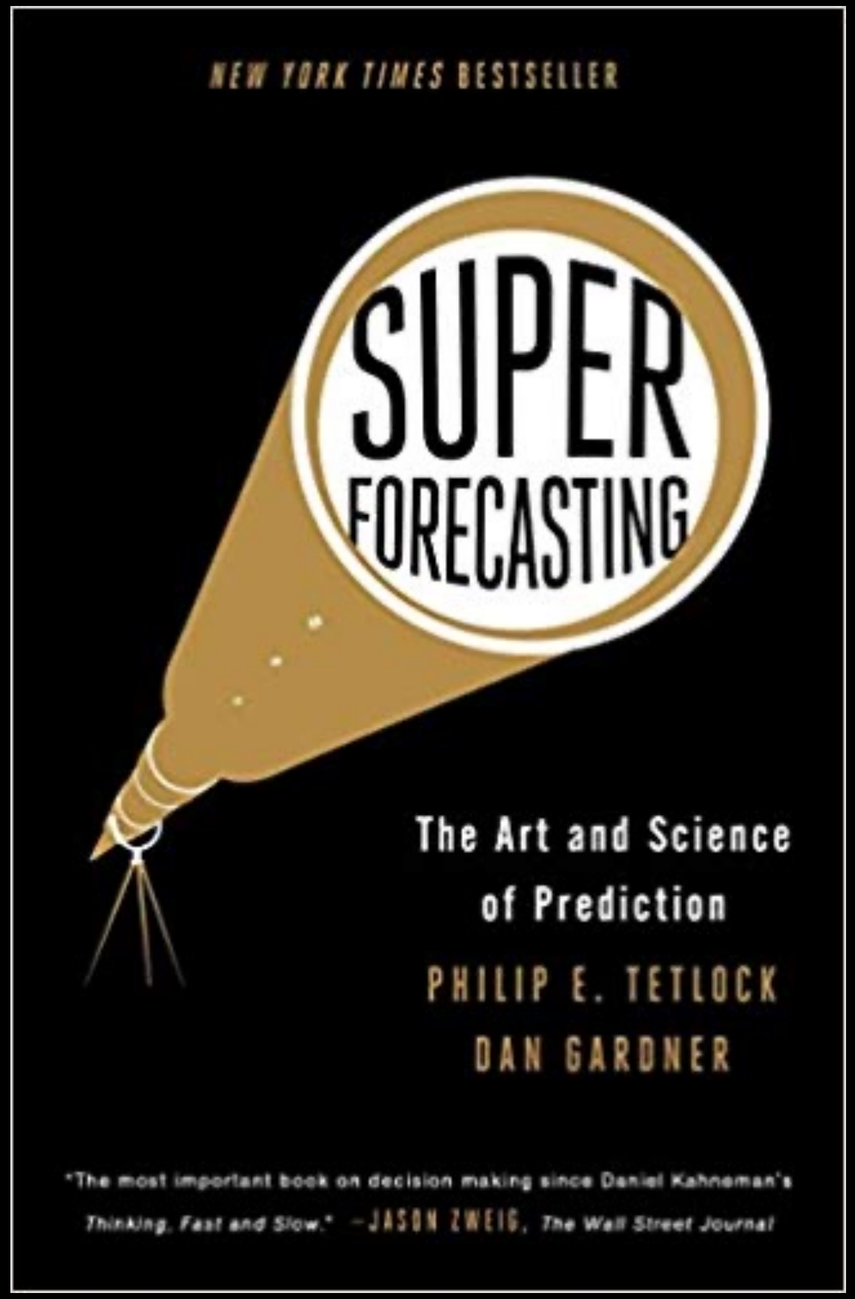
- Domain-related expertise and knowledge accumulated
- Similar experiences observed
- Users similar to Target User Type studied

2 FROM THE RATERS' UNDERSTANDING OF UNDERLYING UX PRINCIPLES

- Pattern recognition
- Application of "heuristic evaluation"
- Basic knowledge of human-computer interaction theory, interaction design patterns, visual design principles, content strategy, etc.

3 TIED TOGETHER WITH PEER DISCUSSION & INTER-RATER RELIABILITY CALCULATIONS DURING PURE SCORING

HOW DOES THIS RELATE TO VALIDITY?



SUPERFORECASTING: THE ART & SCIENCE OF PREDICTION, BY TETLOCK & GARDNER (2016)

- Superforecasting is a learnable skill, but the best tend to have these attributes:
  - Philosophy: Cautious, humble, nondeterministic
  - Thinking style: Open-minded, intelligent, curious, reflective, numerate
  - Work ethic: Growth mindset and grit
- Keeping score (the Brier score) helps Forecasters improve

HOW TO USE THESE NUMBERS TO MOTIVATE ACTION AND FOCUS ON THE UX

PURE SCORECARDS OVER TIME (VERSION OVER VERSION)

March 7 2015 v0.8.5.289		77	July 8 2015 v0.9.1.357		53	Aug 13 2015 v1.0.2.007		25
Download/Install	1 1 3 3 2	10	Download/Install	1 1 2 2	6	Download/Install	2 1 2	5
Initial enrollment	2 1 2 3 2 1	11	Initial enrollment	1 2 1 2 2 1	9	Initial enrollment	2 1	3
Add first entry	2 1 2	5	Add first entry	1 1 1	3	Add first entry	1 1	2
Import database	2 3 2 3 3 1	14	Import database	1 2 2 2 1 2 1	11	Import database	1 2	3
Install companion sw	2 2 2 3 3 3 3	18	Install companion sw	1 1 2 2 2 2	10	Install companion sw	1 2 1 2 2	8
Upgrade to premium	1 1 1 1	4	Upgrade to premium	1 1 1	3	Upgrade to premium	1	1
Resolve issue X	2 3 2 3 3 2	15	Resolve issue X	1 2 1 2 2 1 2	11	Resolve issue X	1 1 1	3

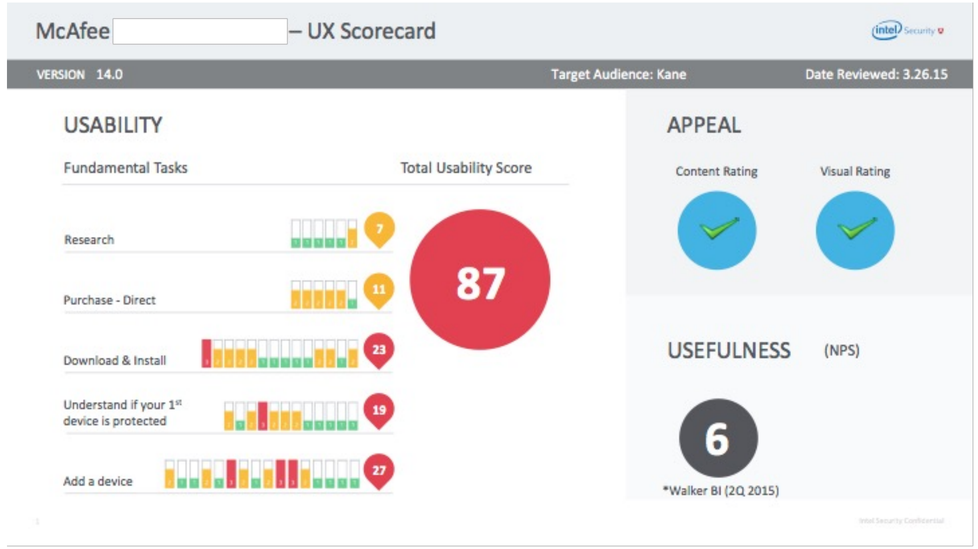
COMPETITIVE PURE SCORECARDS

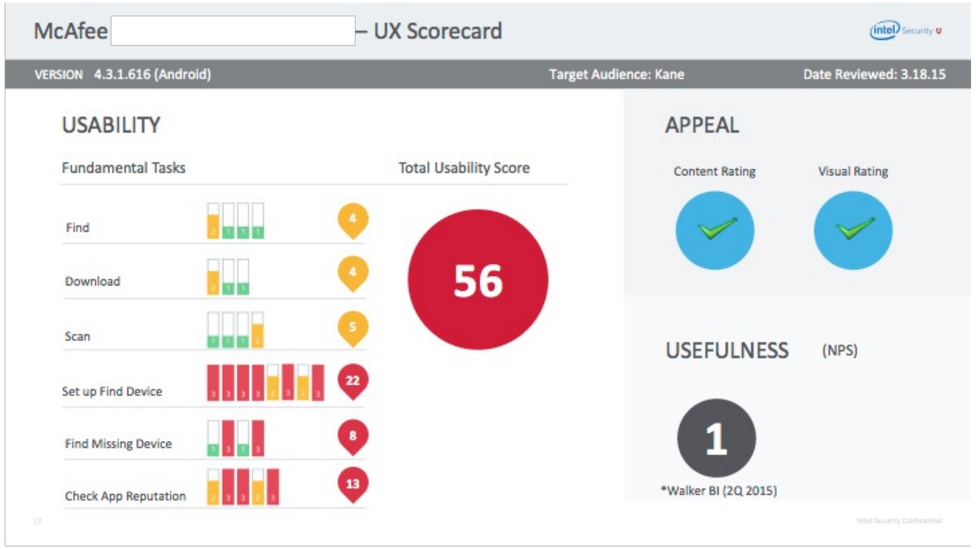
Windows vX.X.XXXX.X iOS vX..X | Android vX.X	Our Product	92	54	Competitor	Windows vX.X.X.XX | iOS vX.X.X Android vX.X.X-XX
Example Task 1*	NA			Example Task 1	2
Example Task 2	1 1 2 2 2 3 2	13		Example Task 2	2 3 3 1 1 1 3	14
Example Task 3 that goes on two lines	2 2 2 3 1 3 1 3 2	19		Example Task 3 that goes on two lines	2 2 2 1 2 3	12
Example Task 4**	3 2 3 3 2 3	16		Example Task 4**	1 2 1 1 2 2 1	10
Example task 5 possibly with way tinier font	1 3 3 3 1 2 2 3 3 2 3 3 2	32		Example task 5 possibly with way tinier font	1 2 2 1 1 2 3 3 3 3 3 3 3 1	31
Task 6 with two lines	2 3 3	8		Task 6 with two lines	2 2	4
Example Task 7 on two lines	1	1		Example Task 7 on two lines	1	1
Example Task 8 on two lines	3	3		Example Task 8 on two lines	2 3 2 2 3	12

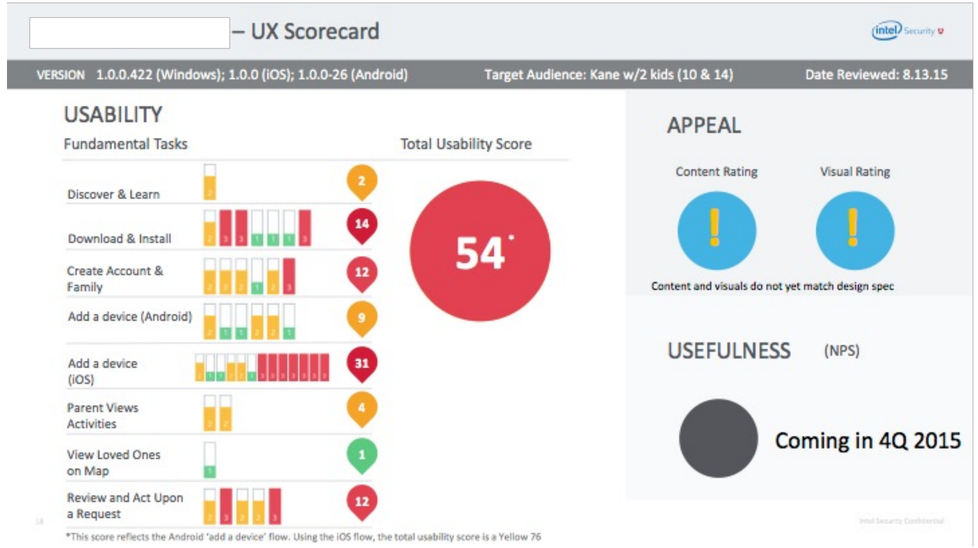
*Task 1 was absorbed into Task 2, so is not rated separately

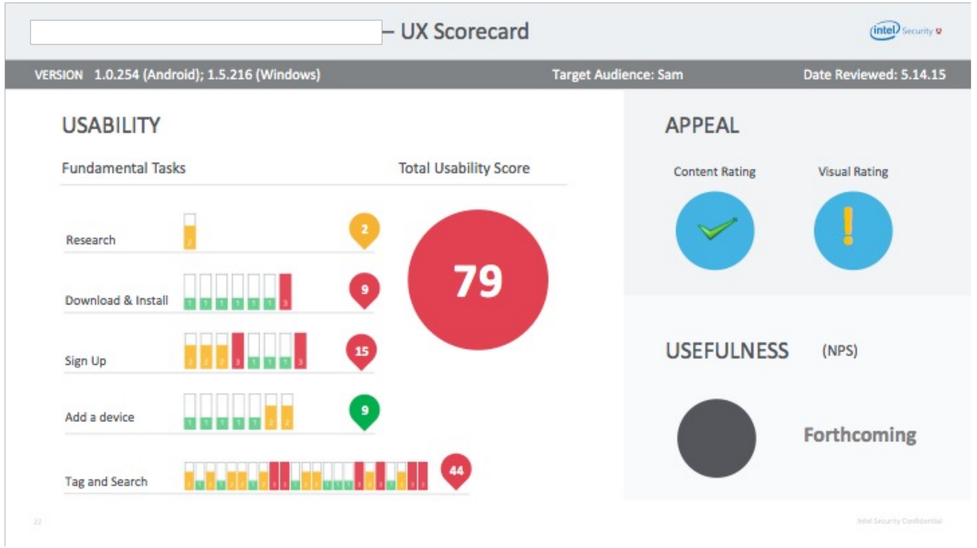
**Note: the number of clicks/steps in Competitor is higher, but level of effort for Task 4 is actually lower

EXAMPLE UX/PURE SCORECARDS









Note: Some of this information is proprietary and confidential so some information has been redacted.

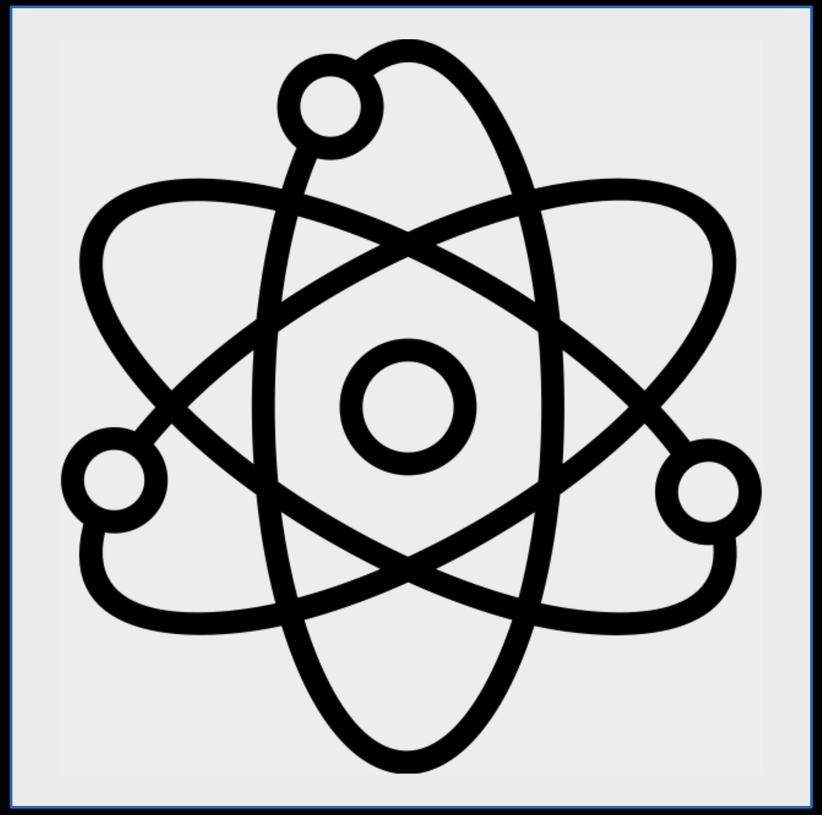
WHAT SHOULD YOU KNOW ABOUT CREATING MY OWN METRICS?

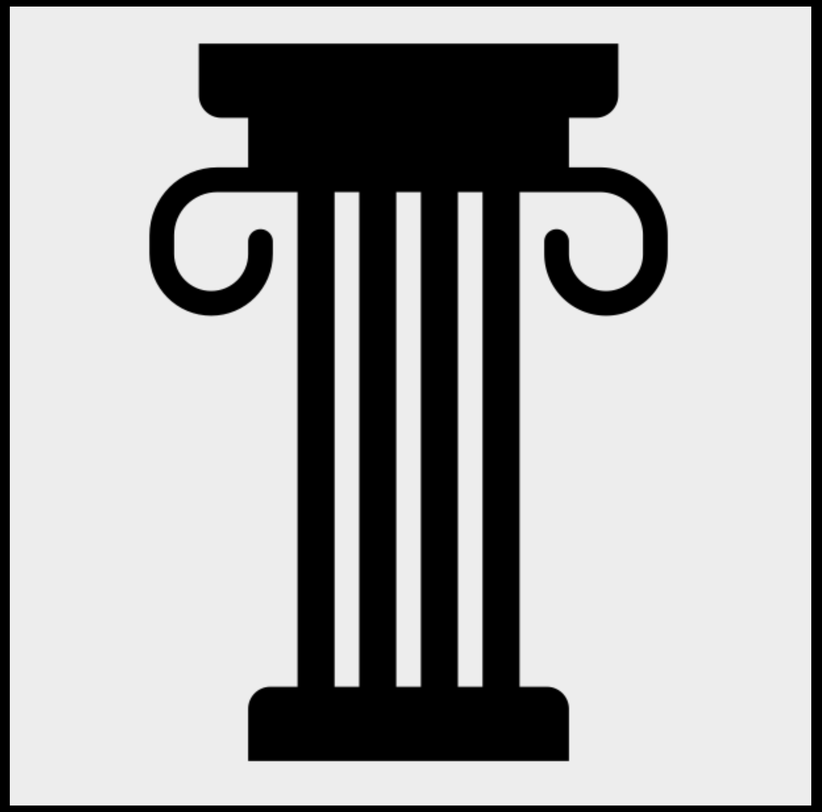
- It’s not simple, so take your time and iterate.
- But you can create something highly relevant to your organization.
- PURE is meant to be universal (for Ease of Use estimates).
- But we don’t yet have a universal way to assess User Needs.

STRATEGIES I'VE USED IN THE PAST (IN THIS ORDER, FOR BETTER OR WORSE)

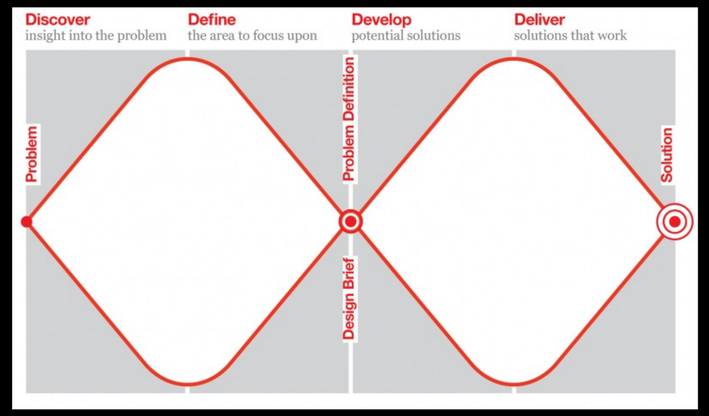
1. Teach Everyone About Research Methods
2. Sell the Benefits of Qualitative Research
3. Conduct both Qualitative and Quantitative Research
4. Develop Your Own Metrics
5. Lead With the Design Process

Q: IS DESIGN A SCIENCE OR HUMANITIES?

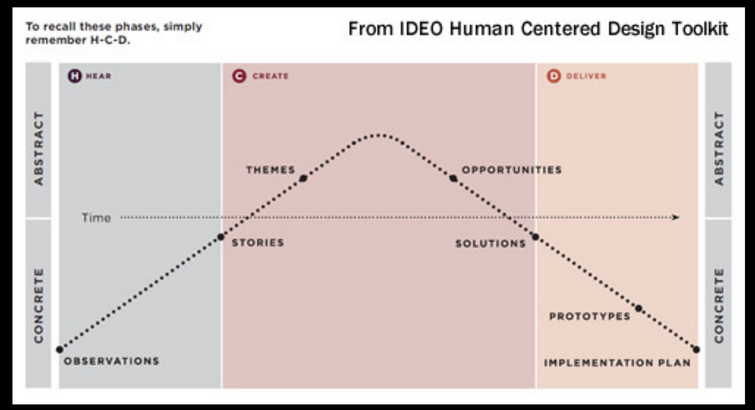




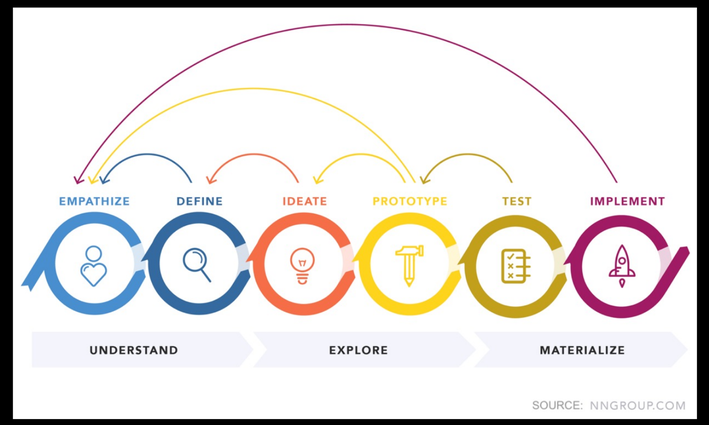
ALL DESIGN PROCESSES USE INSIGHTS, BUT HOW EXACTLY?



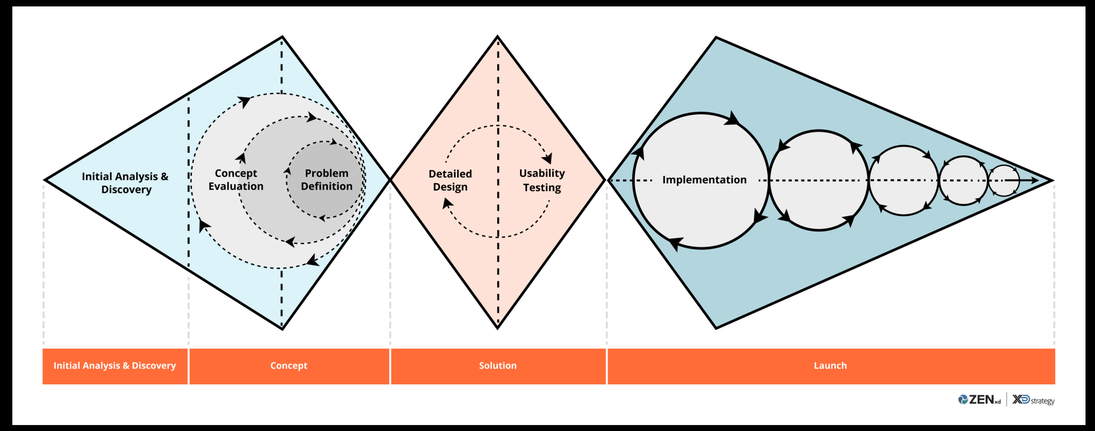
Source:
UK Design Council



Source: IDEO

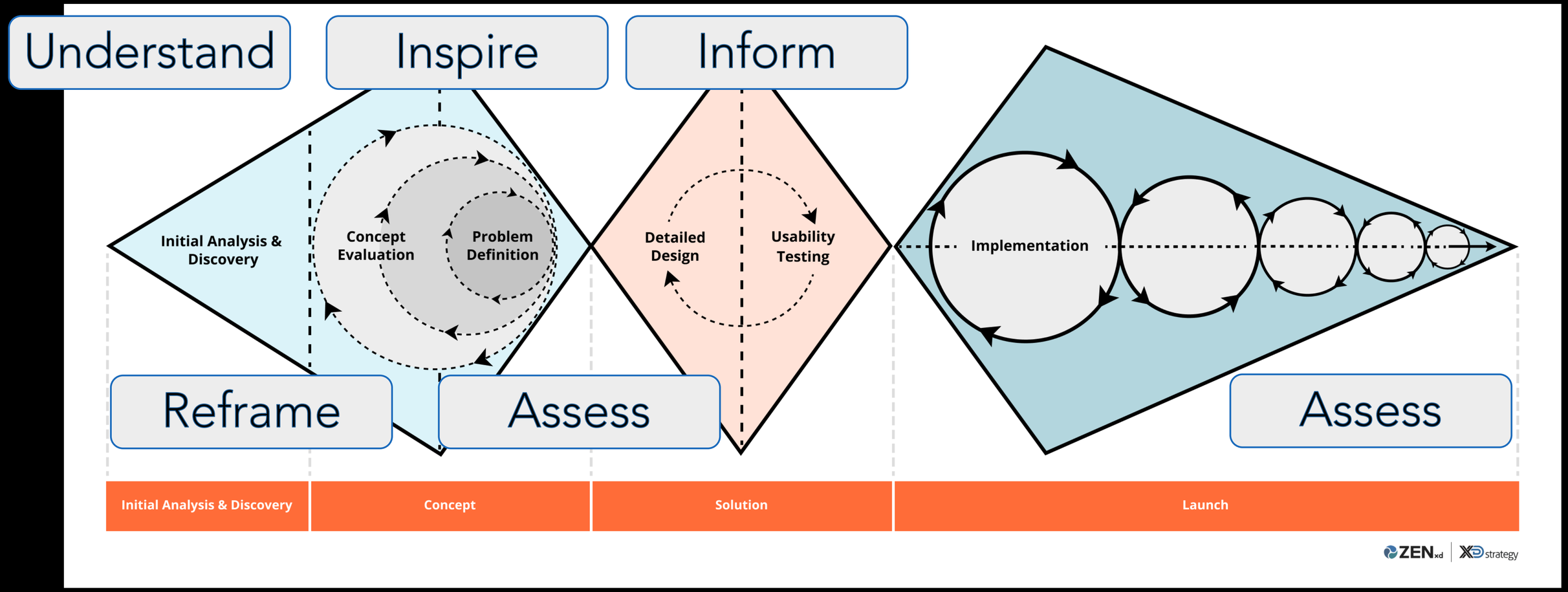


Source: NNGroup

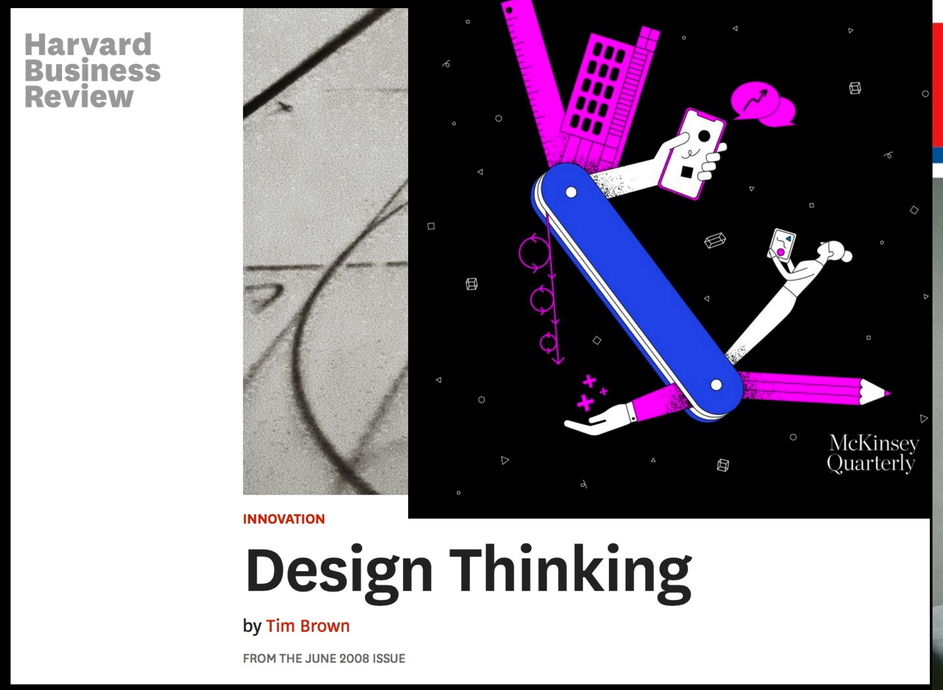


Source: zenXD and XD Strategy

HOW INSIGHTS IMPACT THE DESIGN PROCESS



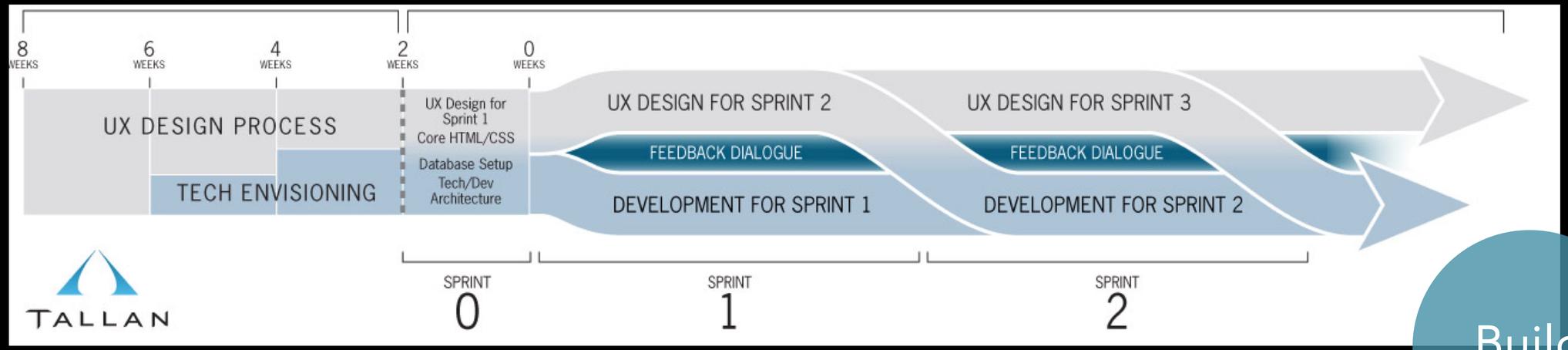
SOUNDS LIKE DESIGN THINKING, WHICH IS A GOOD THING



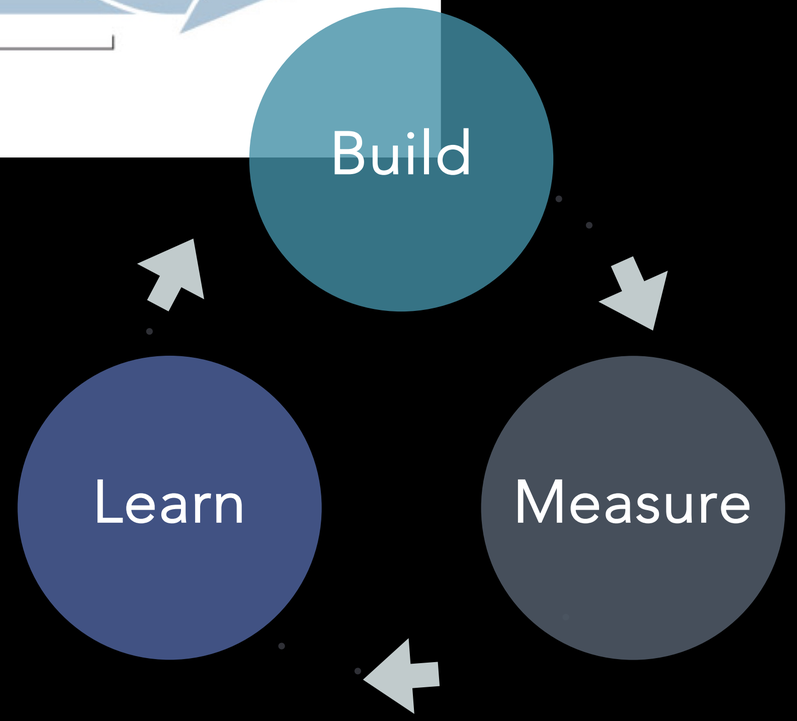
- Promoted by IDEO and Stanford University, it has created an avenue for user-centered design to be more relevant at the highest levels of business
- Essentially, Design Thinking focuses on solving problems in a human-centric way, rather than a tech-centric or product-centric manner

AND YOU CAN POTENTIALLY TIE THIS TO YOUR PRODUCT DEVELOPMENT METHOD

Discover + Define + Design



Agile/UX



Lean UX

SUMMARY AND TAKEAWAYS

NONE ARE PERFECT, BUT ALL HAVE MERIT

1. Teach Everyone About Research Methods
   → Read the room. Not everyone wants to know it all.

2. Sell the Benefits of Qualitative Research
   → Focus on the why, involve stakeholders, show the work.

3. Conduct both Qualitative and Quantitative Research
   → Do both when you have the time and money.

4. Develop Your Own Metrics
   → The Quantification Bias works with many types of metrics.

5. Lead With the Design Process
   → Ultimately, it's not about insights, but what can be designed and built.

PDF of these slides:

https://bit.ly/torchi-2023-09-27

THANK YOU!

CR@XDSTRATEGY.COM