Human-Computer Interaction:
Study Design

Dr Kami Vaniea

Outline

- Lab studies
- Online studies
- Study design
  - Define what “usable” means
  - Select a study methodology
  - Identify the data to collect
  - Design the study structure
  - Plan the evaluation

Usability testing & research

Usability testing

- **Improve** products
- Few participants
- Results inform design
- Usually not completely replicable
- Conditions controlled as much as possible
- Procedure planned
- Results reported to developers

Experiments for research

- **Discover** knowledge
- Many participants
- Results validated statistically
- Must be replicable
- Strongly controlled conditions
- Experimental design
- Scientific report to scientific community

Lab Study

- Basic idea: Have a participant come to a physical place (lab) and interact with the interface there under observation.
- You setup the lab so it mimics the situation you want to test.
- Pros
  - Full control over the environment so limited confounds.
  - Detailed data from each subject .
  - Ability to ask them why they did something.
- Cons
  - Small sample sizes.
  - Being in the lab changes user behavior. They feel safer and their normal distractions are gone. They may also be more stressed.

Mixed-methods study

- Mix several HCI methods together in one study to get a better understanding of the topic.
- Most lab studies are a form of mixed-methods research. One of the more common is to do an experiment followed by a post-interview.
- Pros
  - You get more data.
  - One method will likely catch what another method missed.
- Cons
  - More methods take longer to plan and longer to run.
  - Data from different sources sometimes contradict each other and you must resolve the conflict.

Online Experiment

- Similar to a lab study, but remote.
- Basic idea: Have a participant interact with something via the internet and record aspects of their interactions.
- Pros
  - Larger and more diverse sample size.
  - More realistic (if the tested interface is meant to be used online).
  - Consistent structured data.
- Cons
  - Limited control over the environment so confounds likely.
  - No easy way to ask the user about observed behavior.
  - Careful design vital, users may not properly understand or ignore instructions.

To create a study you need to:

- Define what “usable” means
- Select a study methodology
- Identify the data to collect
- Design the study structure
- Plan the evaluation

To create a study you need to:

- **Define what “usable” means**
- Select a study methodology
- Identify the data to collect
- Design the study structure
- Plan the evaluation

A system which is undefined can never be wrong, it can only ever be surprising.

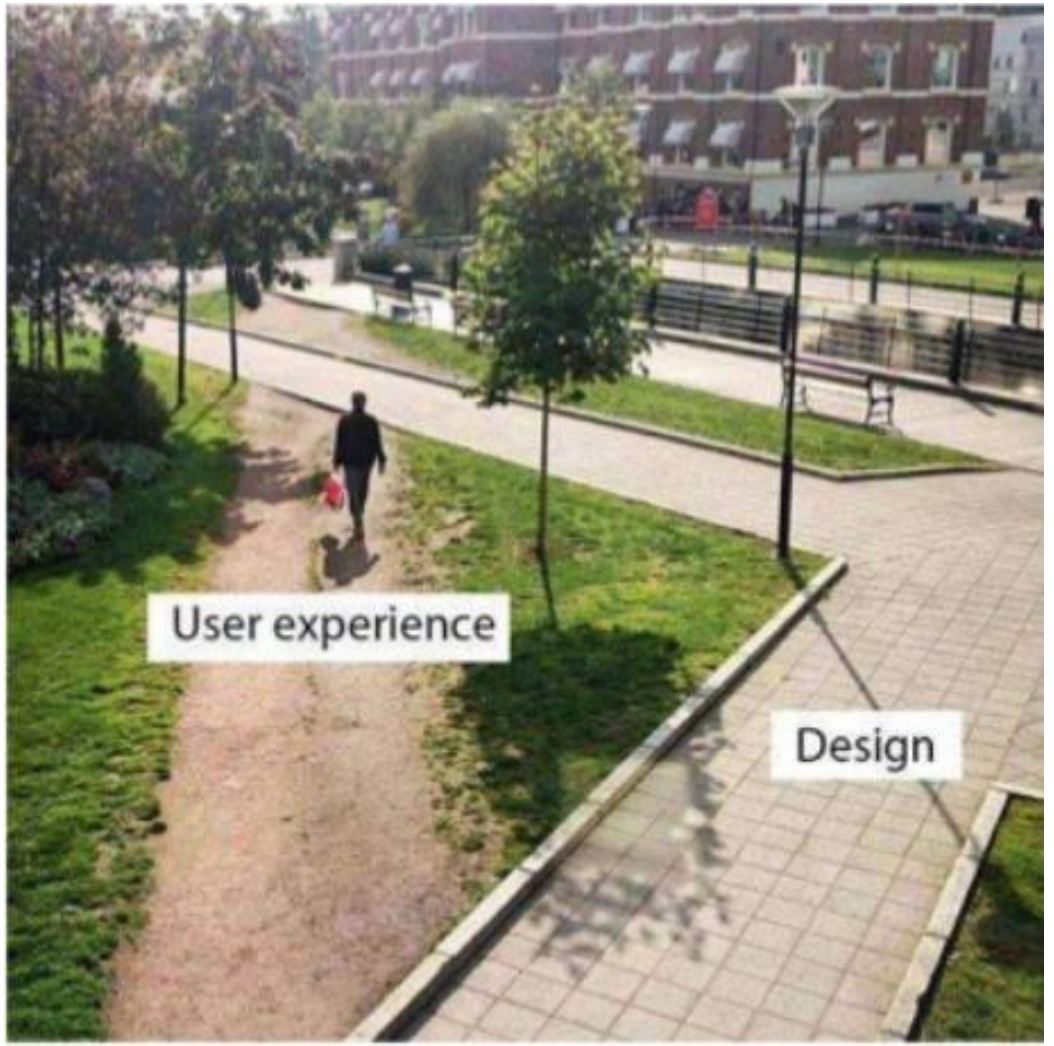
Is [my tool] usable?

Needs to be more specific to be testable.

Usability

A design is not usable or unusable *per se*

- its **features**, together with the **users**, **what the users want** to do with it, and the **users' environment** in performing tasks, determine its level of usability



Usability

A design is not usable or unusable *per se*

- its **features**, together with the **users**, **what the users want** to do with it, and the **users' environment** in performing tasks, determine its level of usability

Start by identifying requirements

Approaches:

- Interviews with users
- Interviews with experts
- Contextual inquiries
- Surveys
- Focus groups
- Reading background literature
- Diary studies
- Artifact analysis

Define your usability goals

- Identify what you think your users need to be able to do using your system or what kind of attitude you want them to have.
- The goals need to be specific and easy to identify if they have or have not been completed.
- Examples:
  - Find a stool on a shopping page and purchase it.
  - Be willing to give the app 5 stars after interacting with it for the first time.
- Poor examples:
  - Have fun using the site.
  - Find a bus to go somewhere.

“Usable” could mean:

- User can accomplish a task in Y minutes.
- User can accomplish task with no major errors.
- User can learn to use the interface the first time they see it so that they can accomplish a task later.
- More users buy products on the site.
- Users buy more expensive products on the site.
- Expert users can navigate from A to B in less than X seconds.
- Users rate the app highly.
- Interface breaks no major HCI heuristics.
- Users can confidently send an encrypted and signed email to someone else who is able to open it.

Some of my recent research questions:

- Can people differentiate between a subdomain and a domain when reading a URL?
- Does [my new system] help people differentiate between malicious URLs and safe ones?
- Can users use [my new password manager] faster and with less errors than [the old password manager]?
- Does knowing how an app will use its permissions impact app installation decisions?
- What factors impact end-users' willingness to update software?
- Is the guidance given by some static analysis tools better at helping developers identify and fix security errors in their code?

To create a study you need to:

- Define what “usable” means
- **Select a study methodology**
- Identify the data to collect
- Design the study structure
- Plan the evaluation

Many ways to test usability

- A/B Testing
- Affinity Diagraming
- Card Sorting
- Case Studies
- Cognitive Walkthrough
- Competitive Testing
- Critical Incident Technique
- Customer Experience Audit
- Desirability Testing
- Diary Studies
- Ergonomic Analysis
- Experience Sampling
- Experiments
- Eye tracking
- Fly-on-the-wall Observation
- Focus Groups
- Graffiti Walls
- Heuristic Evaluation
- Interviews
- KJ Technique
- Observation
- Participatory Action Research

What kind question are you asking?

- Attitudinal – User attitudes and opinions

vs.

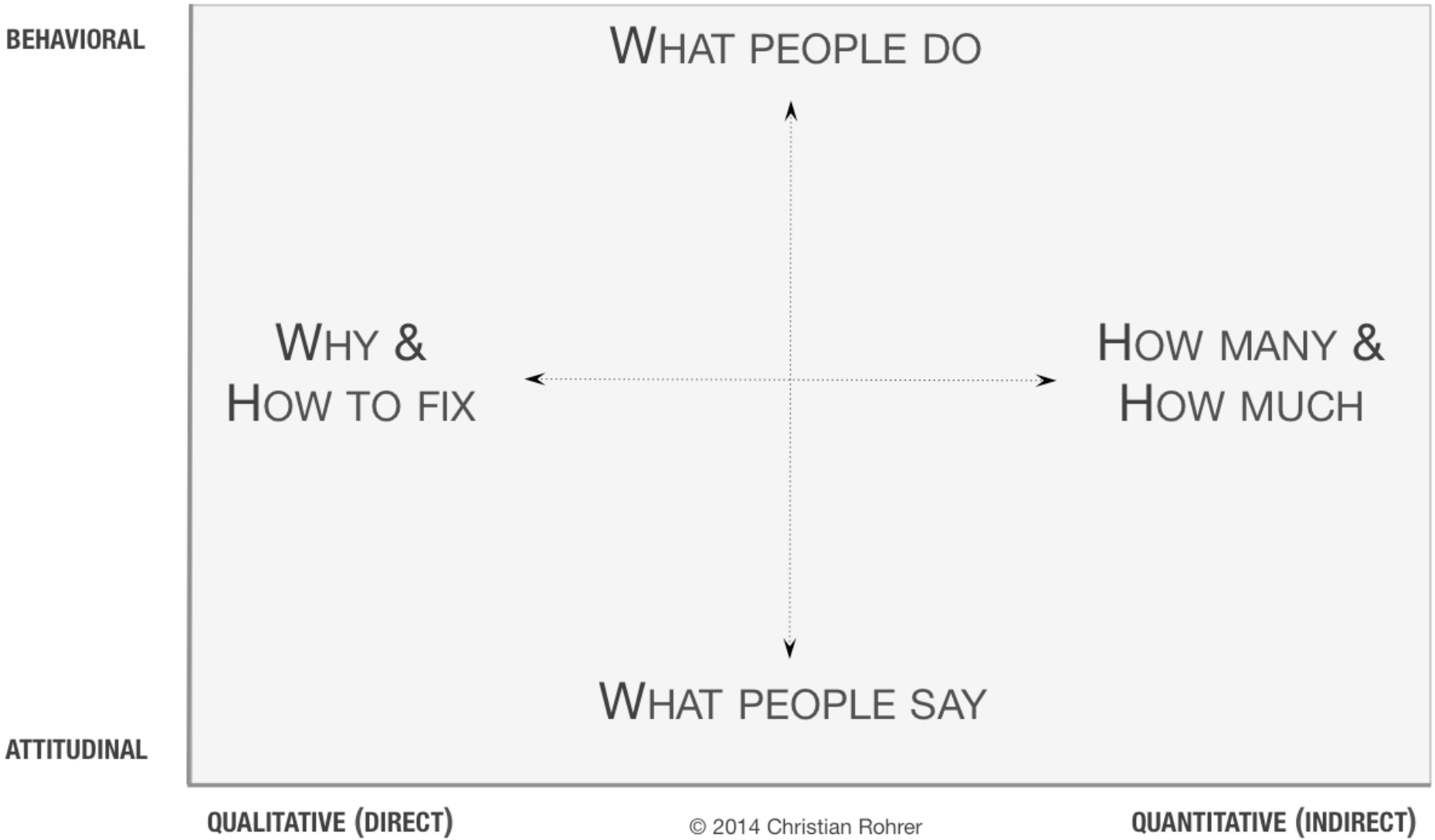
- Behavioral – What the user actually does or is capable of doing

- Qualitative – Unstructured data. Typically unstructured language data

vs.

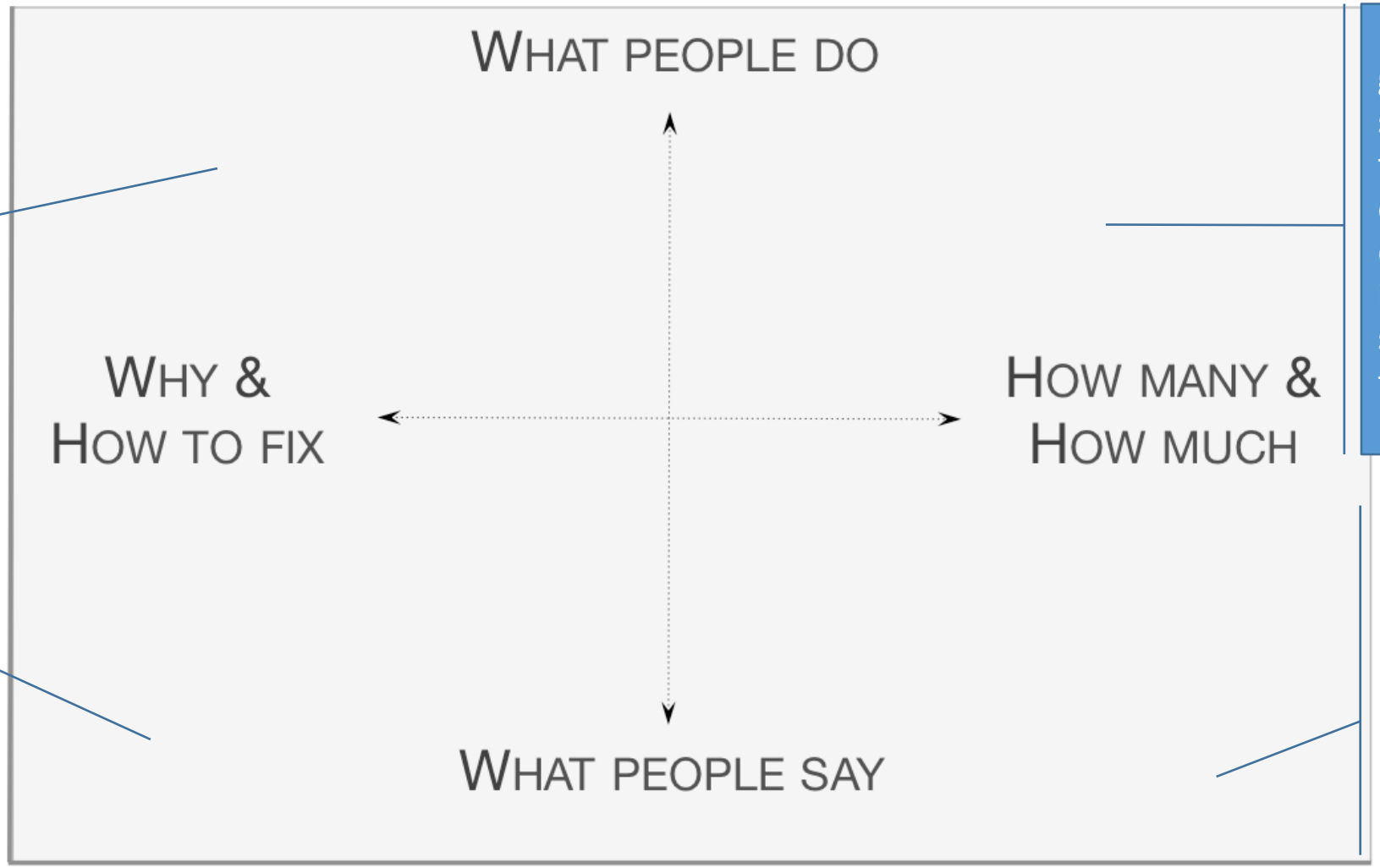
- Quantitative – Structured data. Typically numerical data that can be summed or counted

QUESTIONS ANSWERED BY RESEARCH METHODS ACROSS THE LANDSCAPE



https://www.nngroup.com/articles/which-ux-research-methods/

QUESTIONS ANSWERED BY RESEARCH METHODS ACROSS THE LANDSCAPE



What software update topics do system administrators discuss in online communities of practice?

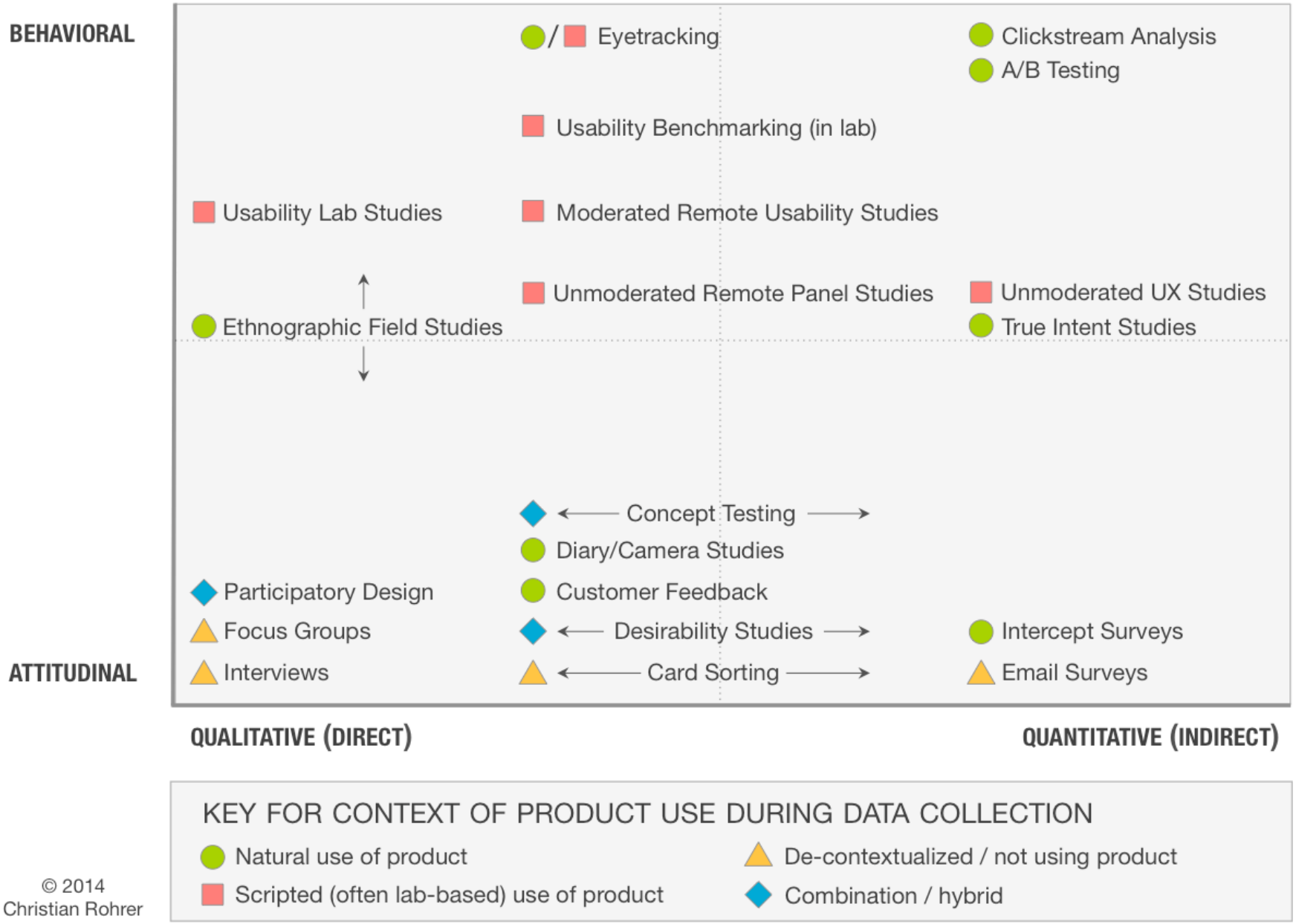
Is the guidance given by some static analysis tools better than others at helping developers identify and fix security errors in their code?

What factors impact end-users' willingness to update software?

What percentage of people think that cookie warning dialogs are linked to online tracking?

https://www.nngroup.com/articles/which-ux-research-methods/

A Landscape of User Research Methods



https://www.nngroup.com/articles/which-ux-research-methods/

To create a study you need to:

- Define what “usable” means
- Select a study methodology
- **Identify the data to collect**
- Design the study structure
- Plan the evaluation

What are you going to measure?

- In statistics there are classically two types of measurements (variables): dependent and independent.
- Dependent
  - Also known as the outcome variable
  - Measures the usability goal
- Independent
  - Anything you are directly manipulating
  - An element of the study which is under your control
  - A pre-existing feature of your participant or study environment

What are you going to measure?

- In statistics there are classically two types of measurements (variables): dependent and independent.
- Dependent
  - Also known as the outcome variable
  - Measures the usability goal
- Independent
  - Anything you are directly manipulating
  - An element of the study which is under your control
  - A pre-existing feature of your participant or study environment
- Controlled
  - An element of the study that you ensure is the same for all participants

Common dependent things to measure

- Time to complete task
- Percent of task completed
- Ratio of successes to failures
- Time spent in errors
- Percent or number of errors
- Percent or number of competitors better than it
- Number of commands used
- Frequency of help and documentation use
- Percent of favorable/unfavorable user commands
- Qualitative data, like answers to interview questions

Common dependent things to measure

- Number of:
  - Repetitions of failed commands
  - Runs of successes and failures
  - Times interface misleads the user
  - Good and bad features recalled by users
  - Available commands not invoked
  - Regressive behaviors
  - Users preferring your system
  - Times users need to work around a problem
  - Times the user is disrupted from a work task
  - Times the user loses control of the system
  - Times user expresses frustration or satisfaction

Common independent things to measure

- User demographics (age, gender, nationality, …)
- Prior experience with the tool or similar tools
- Prior experience with technology
- Existing attitudes
- Answers to scale questions like privacy attitudes
- Aspects of the study
  - Which interface design they were assigned
  - What computer they were asked to use
  - Which recruitment platform they came from
  - Which condition they were assigned to
  - How hot or cold the room was
  - Order tasks were given to the user

Common things to control for

- Computing equipment – laptops can have different types of mice and screens.
- Researcher script – users can be impacted by what the researcher says, saying the same thing to all participants helps.
- Require specific browser – online studies can look different on different browsers, so one option is to require users to take the study on a desktop using a specific browser.
- Task order – users will behave differently on the first task than they will on the last. Also they will be influenced by earlier tasks.
- Pre-define hints – if a user gets stuck on a task you don't want them too frustrated, but giving unscripted answers can result in variations. So pre-creating hint cards can control for variation.

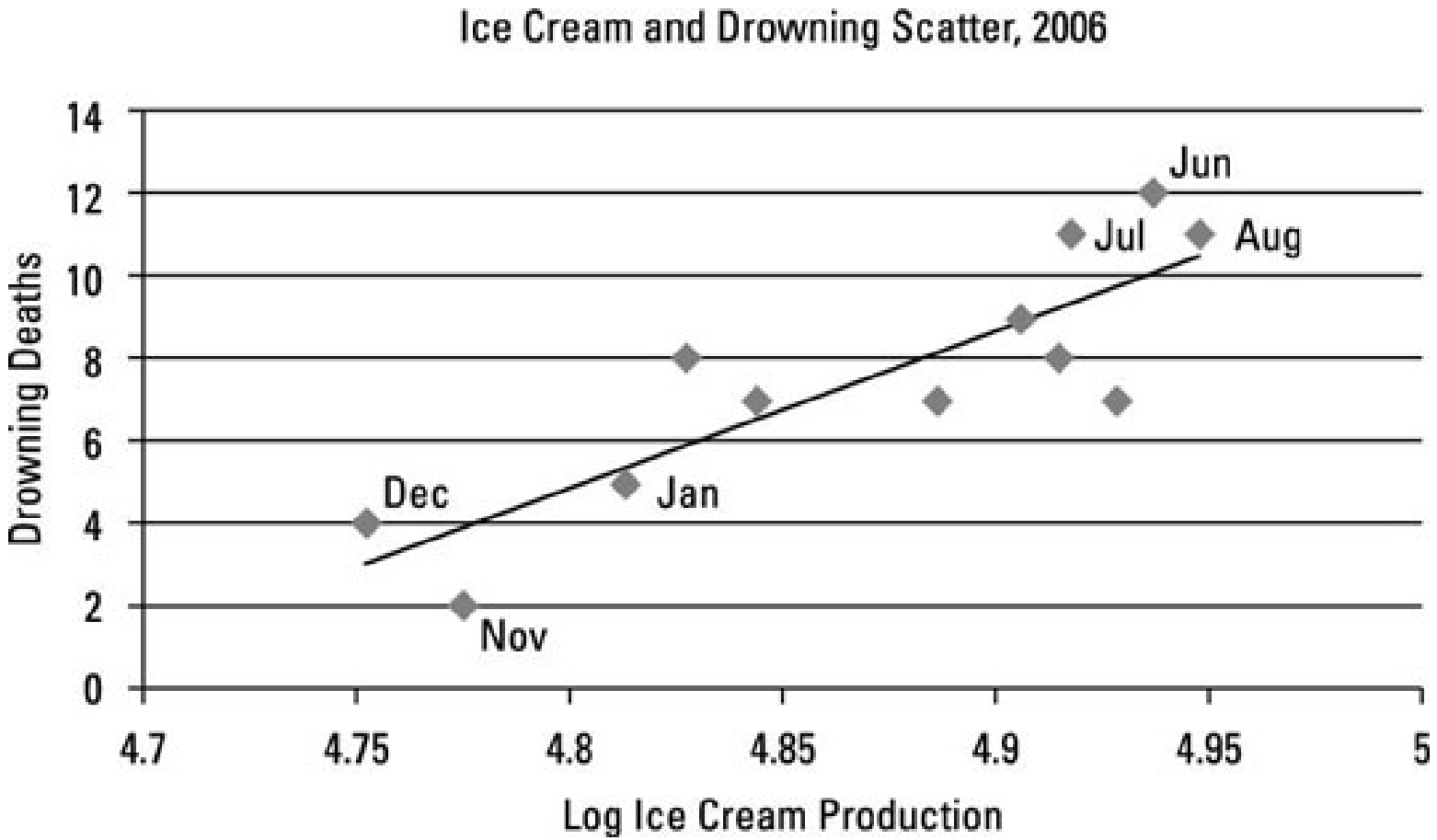
What are you going to measure?

- In statistics there are classically two types of measurements (variables): dependent and independent.
- Dependent
  - Also known as the outcome variable
  - Measures the usability goal
- Independent
  - Anything you are directly manipulating
  - An element of the study which is under your control
  - A pre-existing feature of your participant or study environment
- Controlled
  - An element of the study that you ensure is the same for all participants

Confounding Variable

An unmeasured influence that may be the source of the observed effect.

Think about what might impact your dependent variable in advance and measure it.



Confounding Variable

An unmeasured influence that may be the source of the observed effect.

Think about what might impact your dependent variable in advance and measure it.

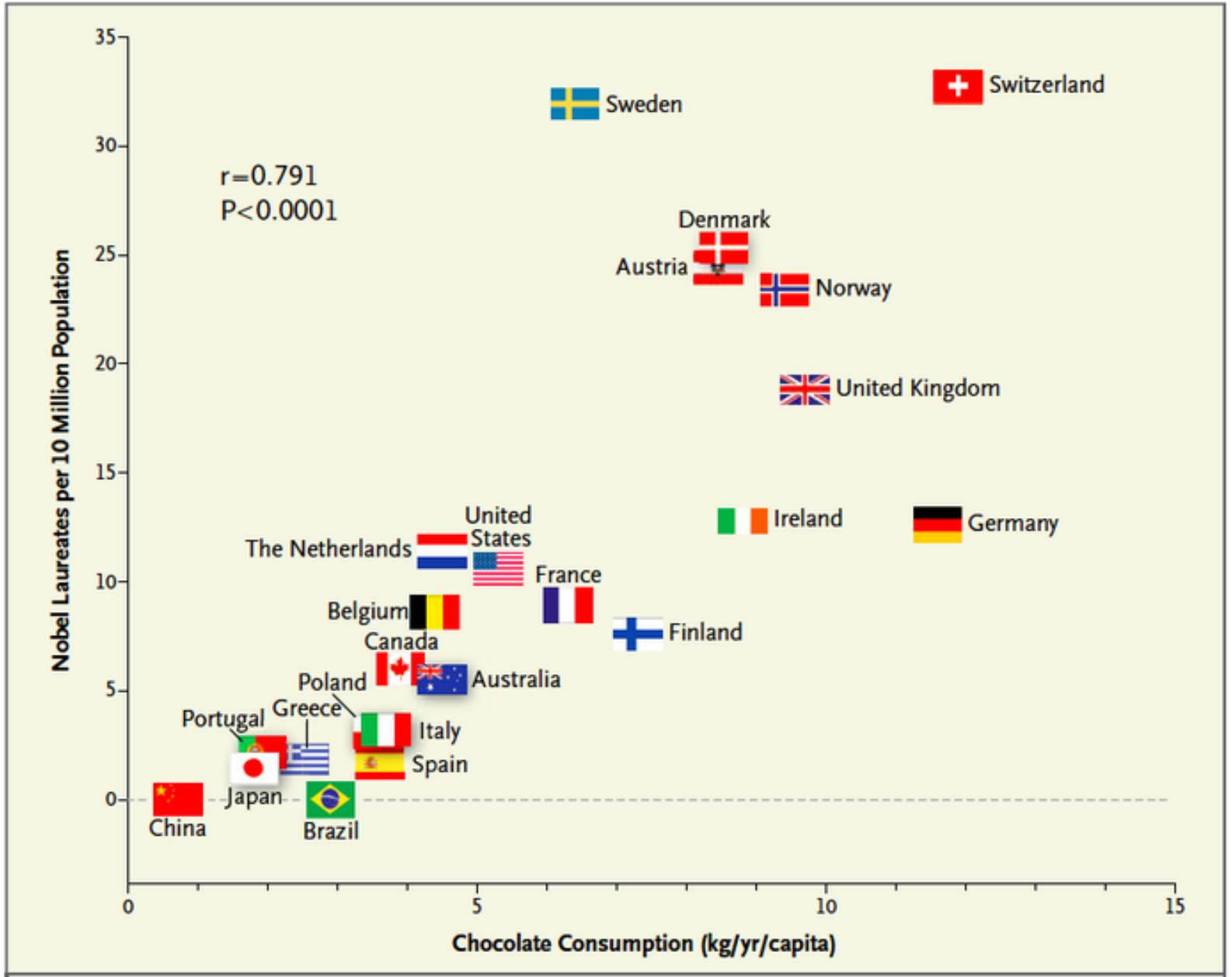
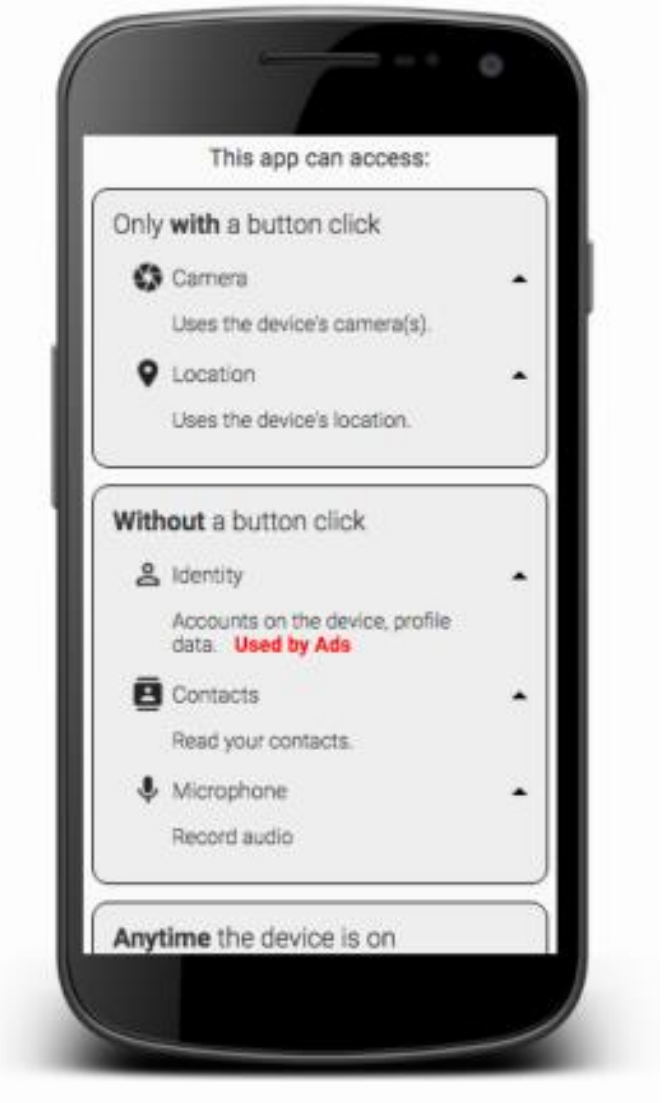


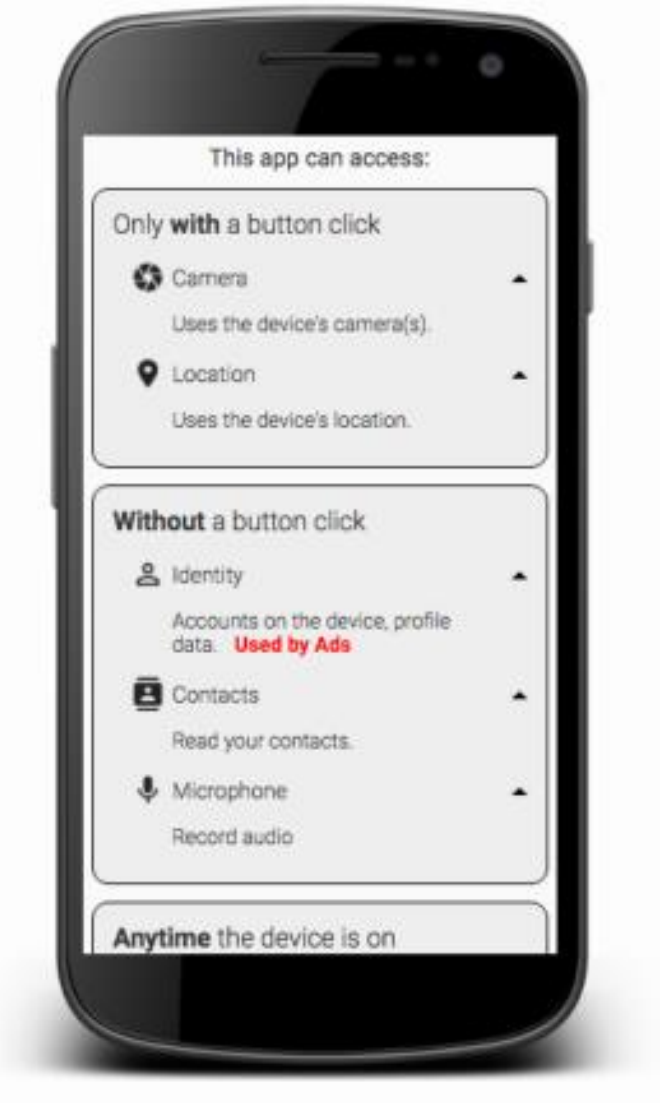
Figure 1. Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.

Example:

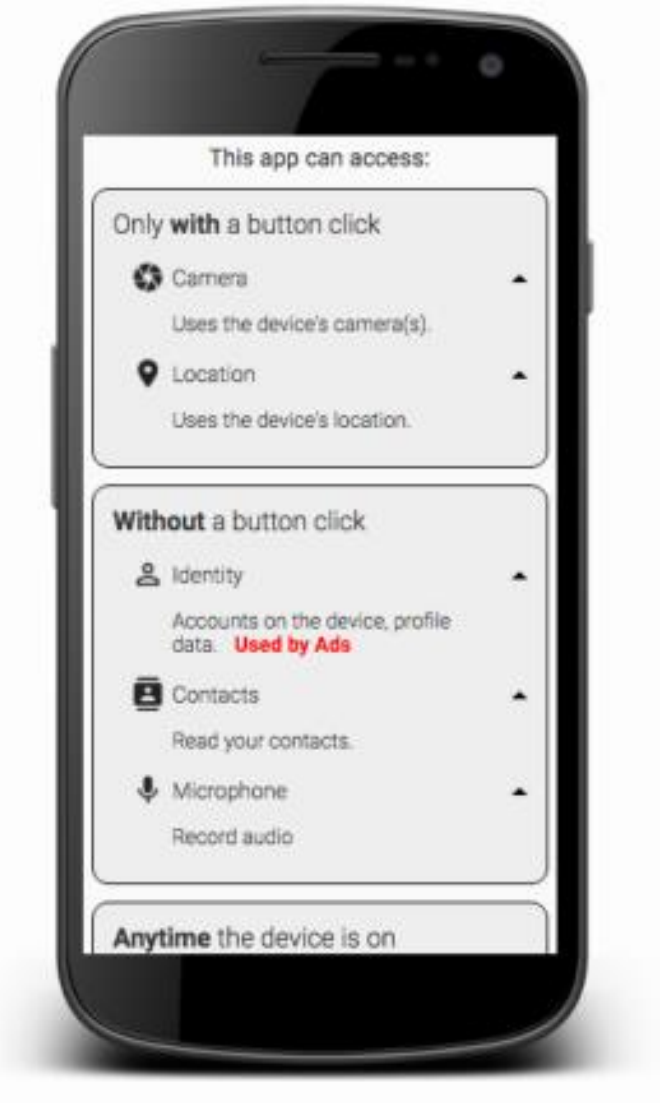
Project to make app permissions easier for people to understand.

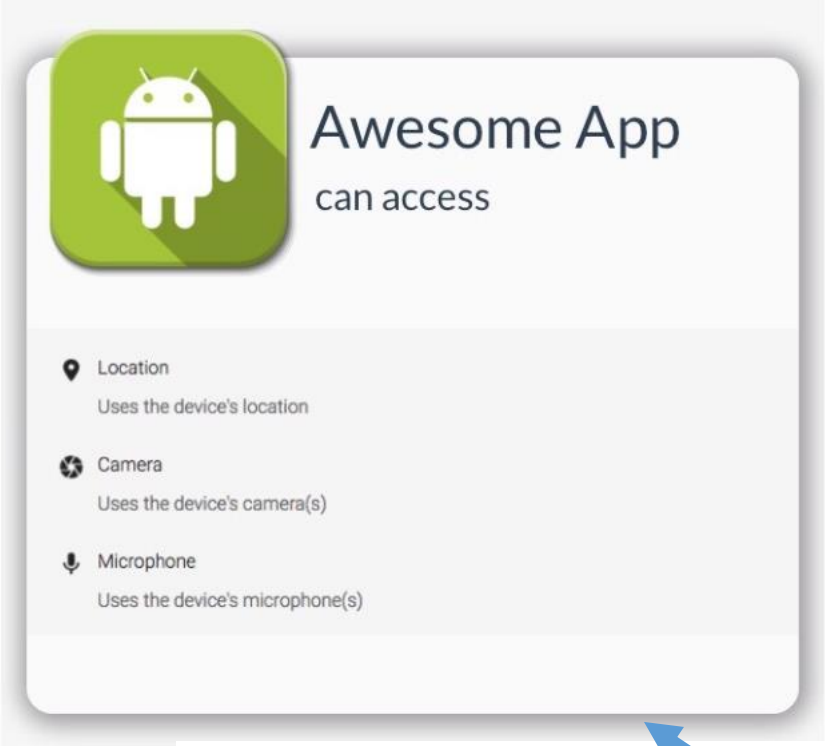


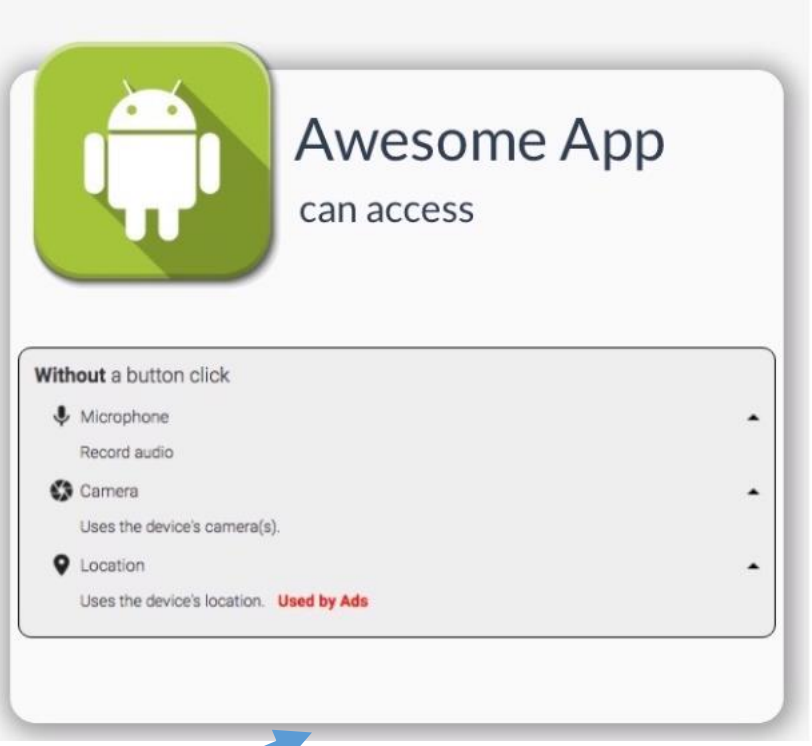
Usability: User can identify if an app can or cannot perform an action directly tied to a permission.



Method: Online study where users are asked to look at the interface and answer questions based on it.







Dependent variable:
Count of the number of questions the participant answered correctly

Independent variable:
Which of the two interfaces the participant was shown

Which of the following can this app do?

					Absolutely Possible
Charge purchases to your credit card at any time.					○
Get your location.	○	○	○	○	○
Allow ads to know your location.	○	○	○	○	○
Load ads.	○	○	○	○	○
Write on the SD card					

Example 2

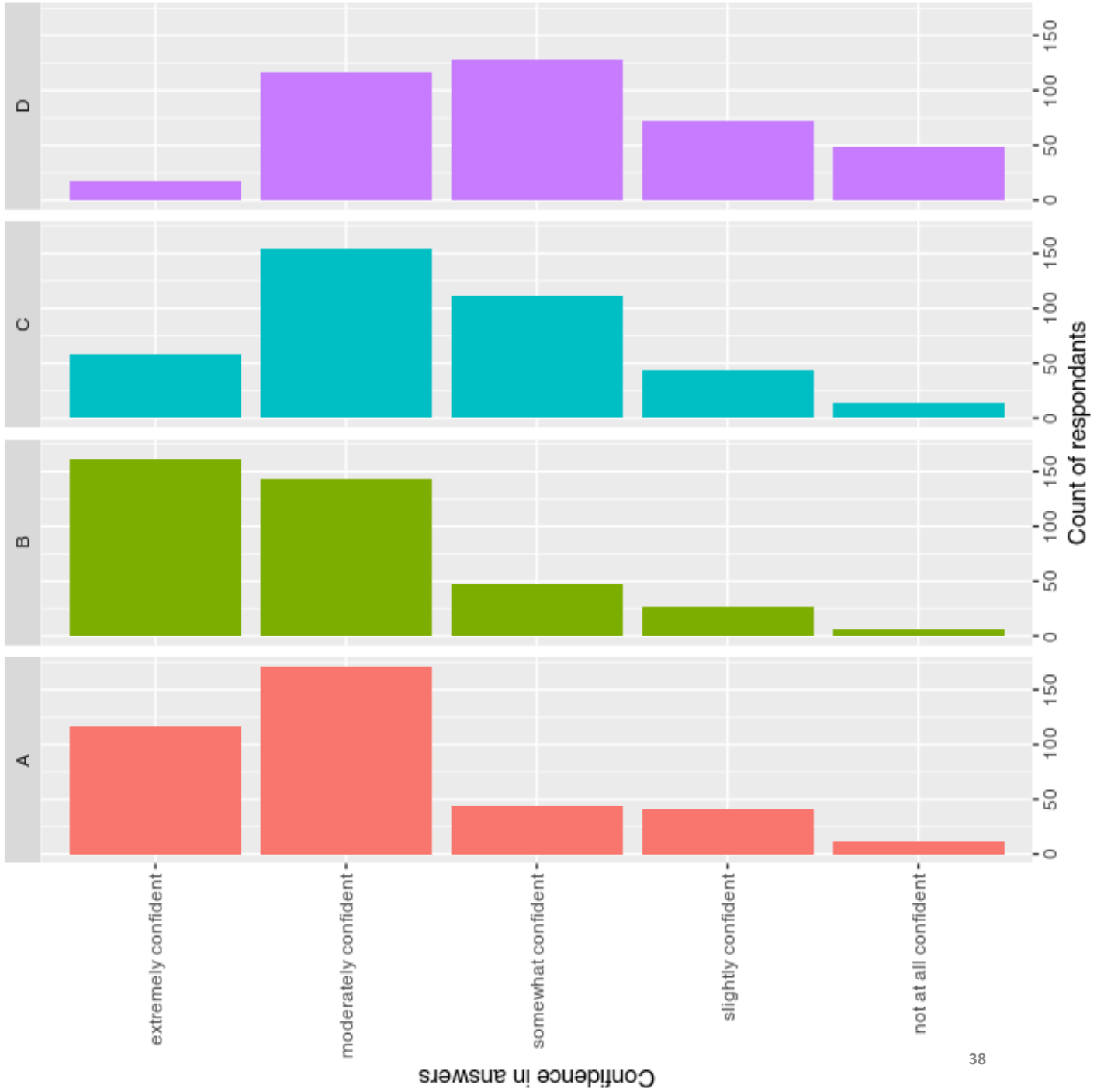
I showed participants 4 code samples and asked them what the code would do. I then asked them how confident they were in their answer.

Research Question: Does the code sample shown impact confidence in their answer?

Research Question:
Does the code sample shown impact confidence in their answer?

Independent:
Which code sample shown

Dependent:
Confidence



What I really want you to learn:

Think about what variables you are interested in and what graph / plot / table you want **before** you conduct the study.

To create a study you need to:

- Define what “usable” means
- Select a study methodology
- Identify the data to collect
- **Design the study structure**
- Plan the evaluation

Scripted vs observational

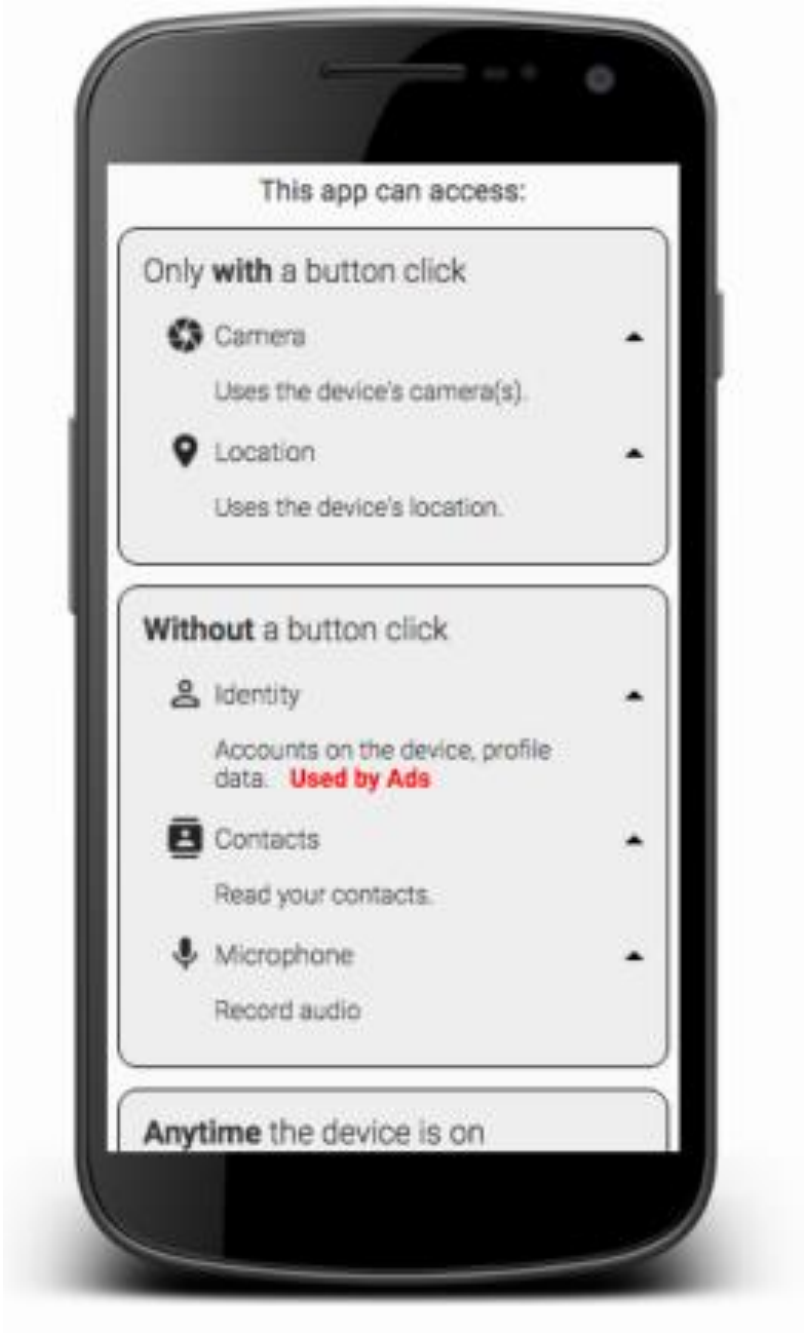
- Scripted studies are planned in detail in advance.
  - Tasks are prepared in advance.
  - Participants are in a controlled environment such as a lab.
  - Nearly all lab based studies are scripted.
  - Think-aloud is scripted.
- Observational studies are not planned and simply observe users doing their own tasks.
  - Participants may not even be notified that they are part of a study.
  - Participants are in their natural environment doing what they would normally do.
  - Hard/impossible to prove what task the user was trying to accomplish.
  - Some planning still needed around data collection.

Between vs. Within subjects

- Between subjects
  - Your study only shows one interface to one person
  - You are measuring how well the people randomly assigned to the A interface did compared to the people randomly assigned to the B interface
  - Lots of variability with this method
- Within subjects
  - Your study shows all interfaces to all people
  - You are measuring the difference in how they do on the two interfaces
  - Less variability (same person) but more learning effects and priming

Study design

- A/B test between the existing and new interface
- Between subjects
- 10 Tasks shown in the same order to all participants
- Dependent variables
  - Accuracy on task
- Independent variables
  - Which interface
  - User demographics

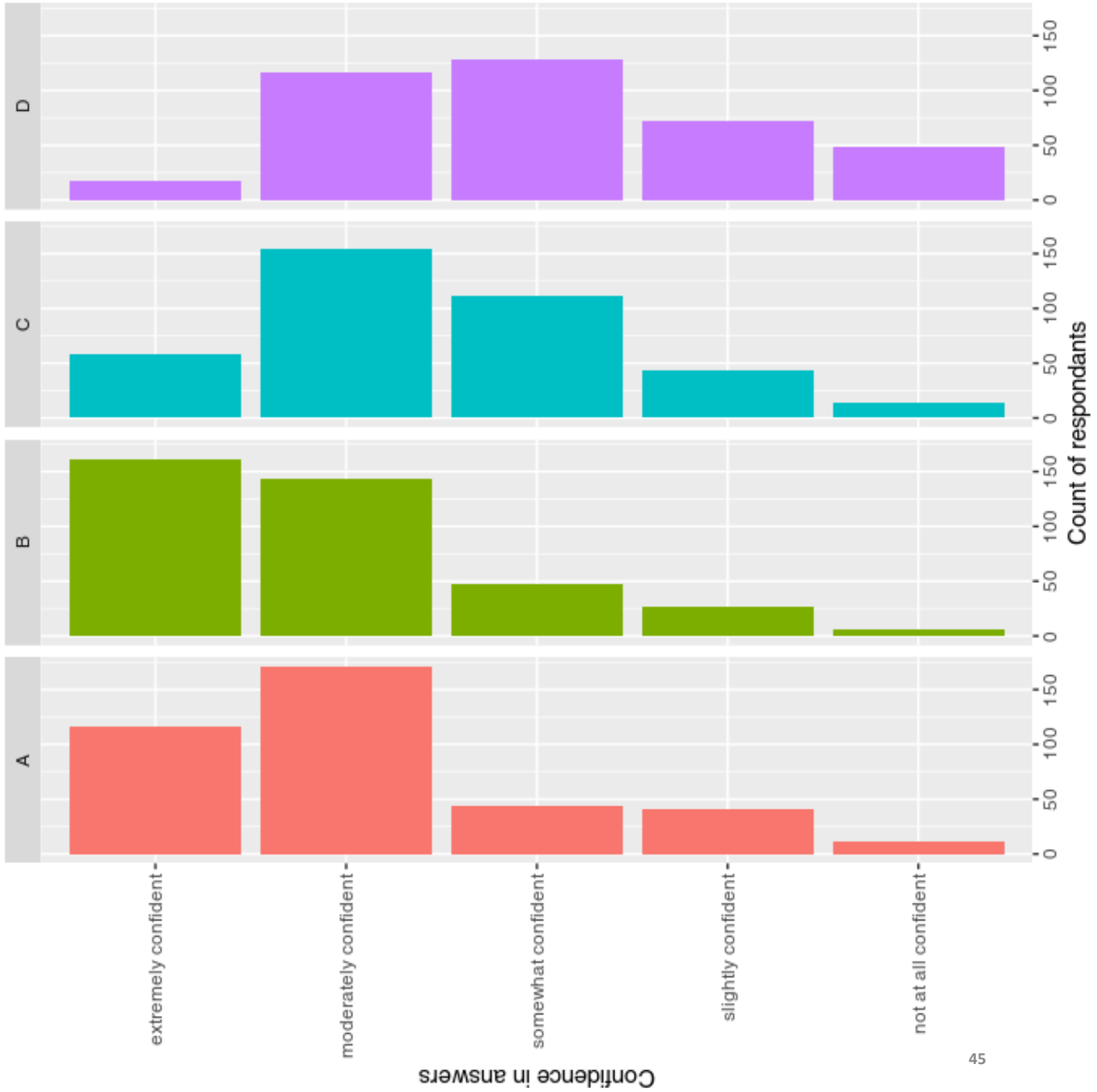


Online experiment

Within-subjects

Independent:
Which code sample shown

Dependent:
Confidence



To create a study you need to:

- Define what “usable” means
- Select a study methodology
- Identify the data to collect
- Design the study structure
- **Plan the evaluation**

Evaluation depends heavily on questions and data

- Goals
  - Usability testing
  - Evaluation for research
- Data type
  - Qualitative data
  - Quantitative data
- Research question
- Researcher skills
  - Statistics?

Usability testing & research

Usability testing

- **Improve** products
- Few participants
- Results inform design
- Usually not completely replicable
- Conditions controlled as much as possible
- Procedure planned
- Results reported to developers

Experiments for research

- **Discover** knowledge
- Many participants
- Results validated statistically
- Must be replicable
- Strongly controlled conditions
- Experimental design
- Scientific report to scientific community

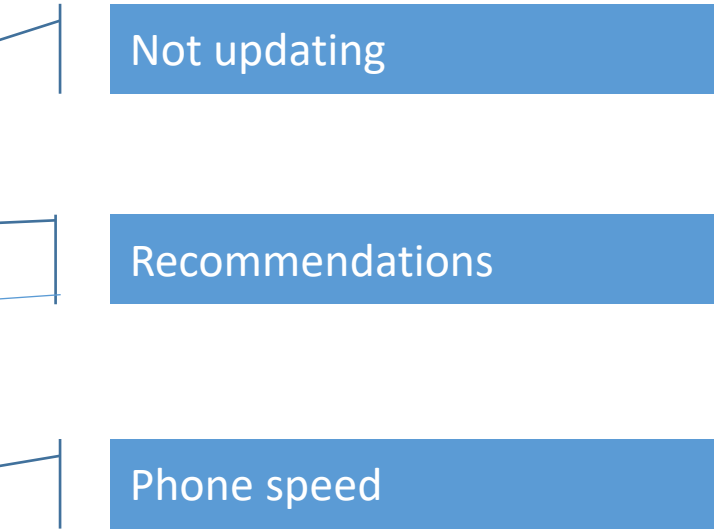
Task accuracy

- Count the number of tasks where the participant was able to accomplish your goal.
- If most participants were able to accomplish the goal then Yay! The interface is usable.

	Current Interface	New Interface
Task 1	15	12
Task 2	12	14
Task 3	11	10
Task 4	7	4

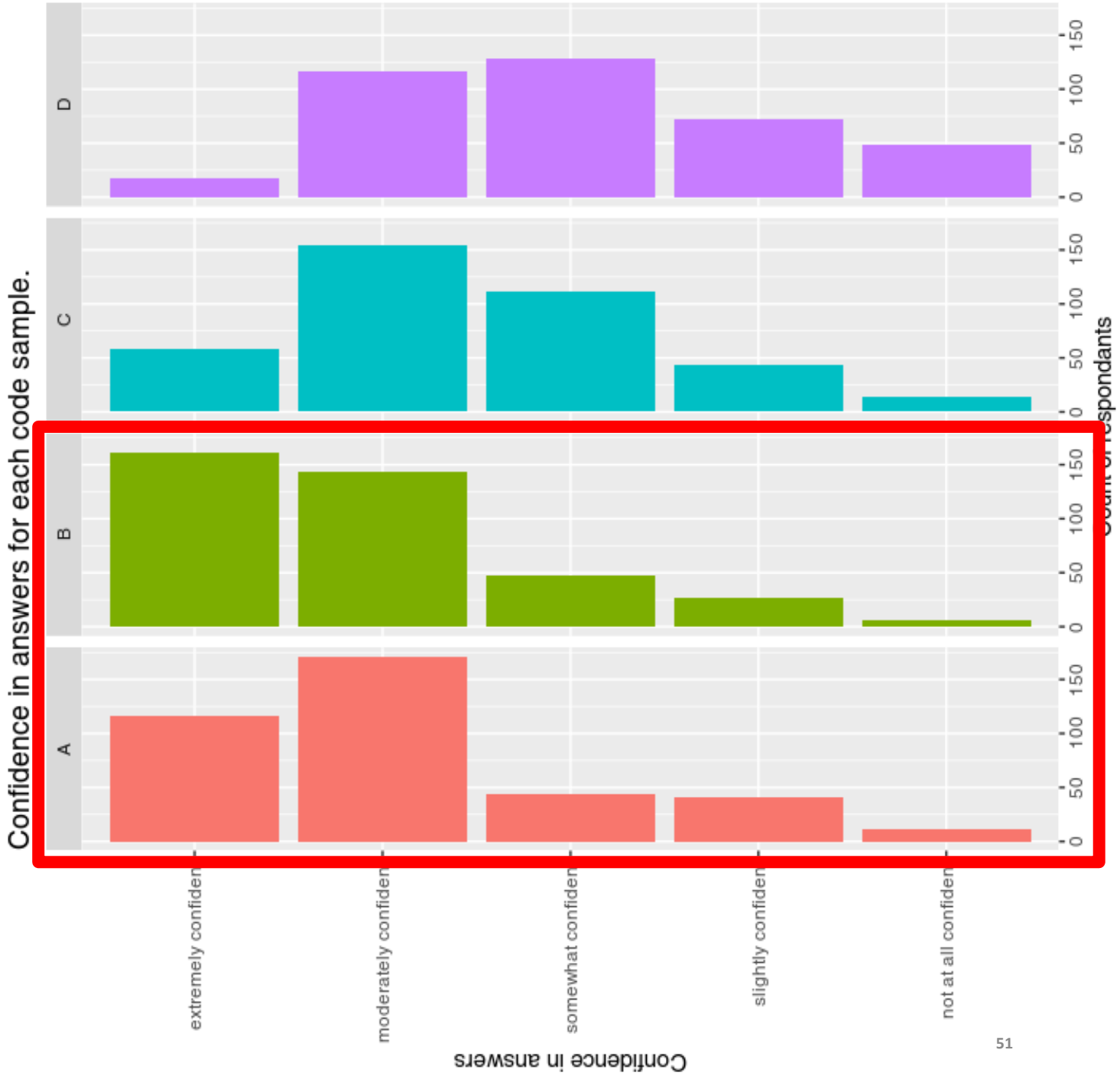
Content coding: Open coding

"I decided that I wasn't going to install the update because I have heard all the reviews online about how it generally makes your phone slower in every respect."

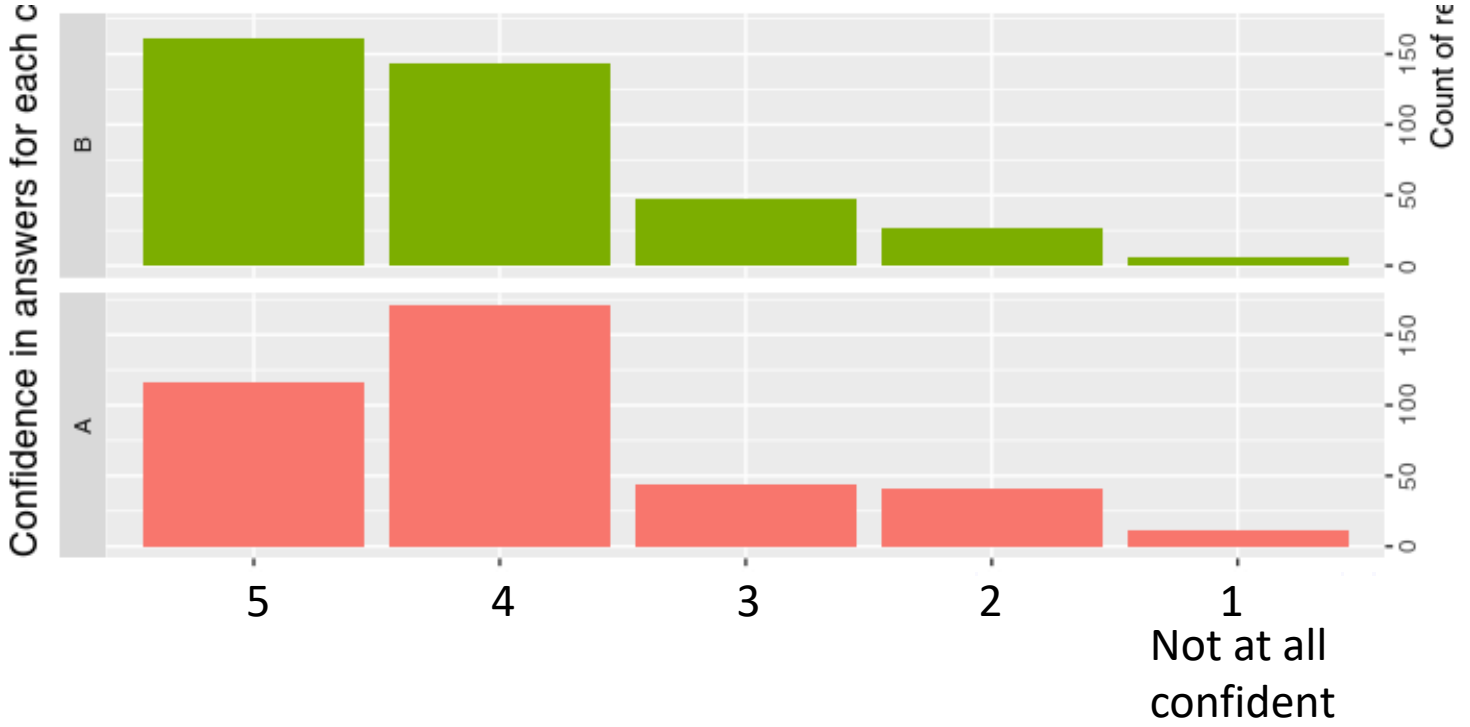


Statistics

Question: Do people have a different level of confidence when reading code samples A vs B?



Yes, they have different confidence levels between A and B.




```
> t.test(a.confidence,b.confider

        Paired t-test

data:  a.confidence and b.confidence
t = -5.2699, df = 383, p-value = 2.285e-07
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.3218198 -0.1469302
sample estimates:
mean of the differences
                -0.234375
```

To create a study you need to:

- Define what “usable” means
- Select a study methodology
- Identify the data to collect
- Design the study structure
- Plan the evaluation

Human-Computer Interaction:
Study Design

Dr Kami Vaniea