HCI: STUDY DESIGN

Dr Kami Vaniea

Today: designing studies to test usability

Step 1: Define what “usable” means
Step 2: Identify your variables
Step 3: Setup your study
Step 4: Evaluate the outcome

The problem:
You just built a new widget and now you need to evaluate it

Step 1: Define what “usable” means

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

Started by designing requirements

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

Define your usability goal

- This step is very similar to specifying tasks
- 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
- Bad 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
- Interface makes the client happy

Step 2: Identify your variables

What kind of data do you want?

- Attitudinal – User attitudes and opinions

vs.

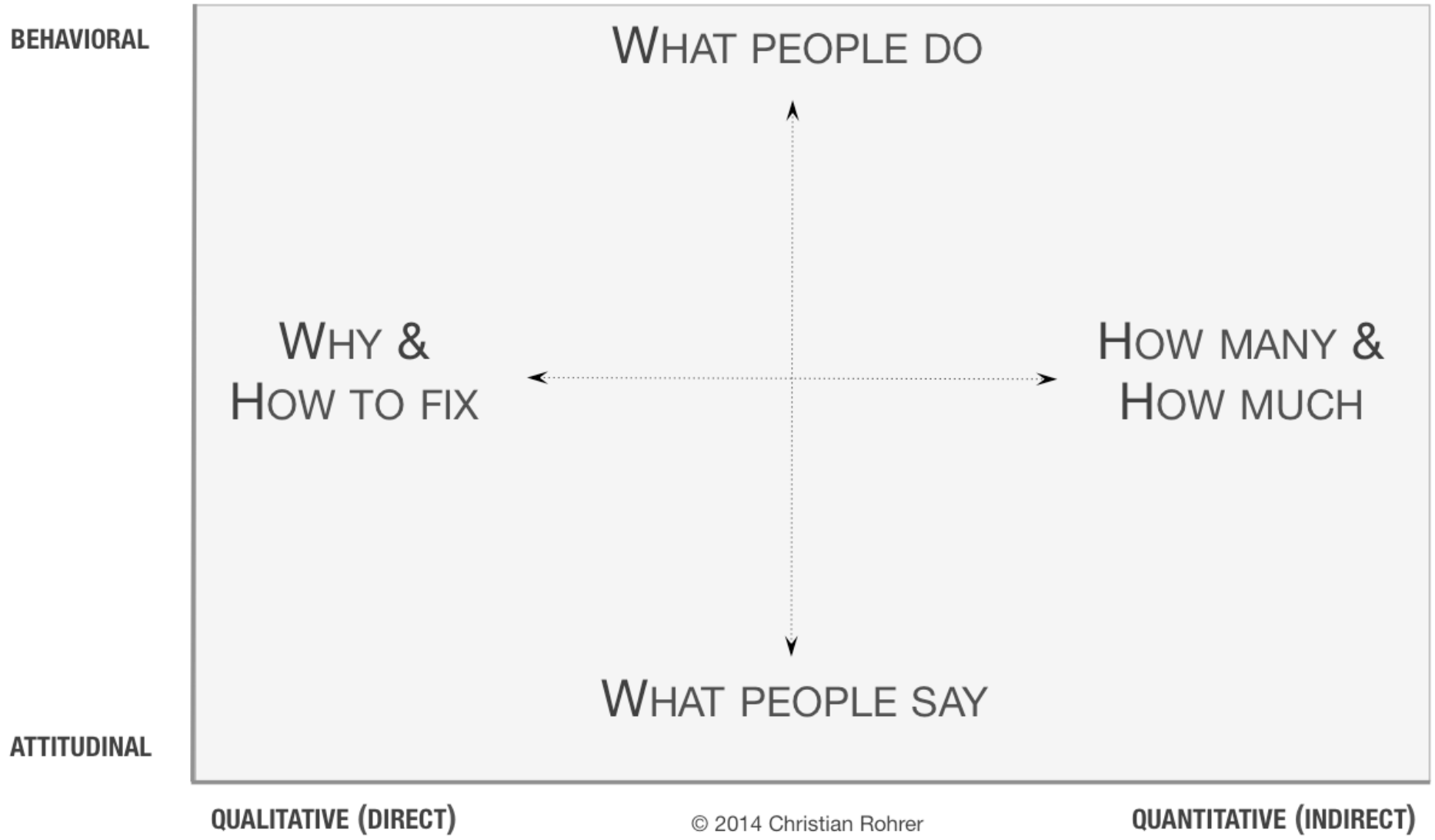
- Behavioral – What the user actually does or is capable of doing

- Qualitative – Unstructured data. Typically unstructured language data

vs.

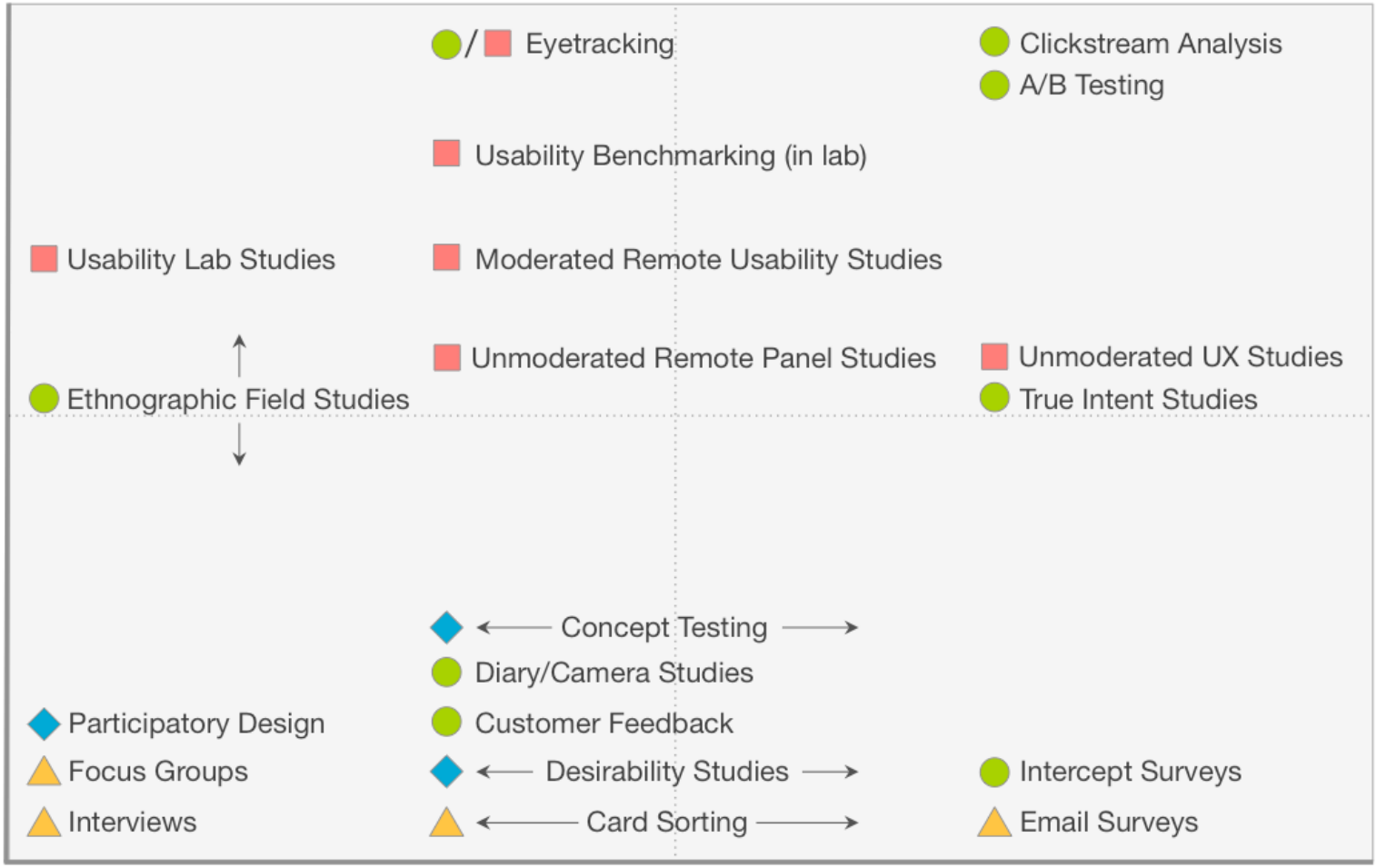
- Quantitative – Structured data. Typically numerical data that can be summed or counted

QUESTIONS ANSWERED BY RESEARCH METHODS ACROSS THE LANDSCAPE



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

A Landscape of User Research Methods



KEY FOR CONTEXT OF PRODUCT USE DURING DATA COLLECTION

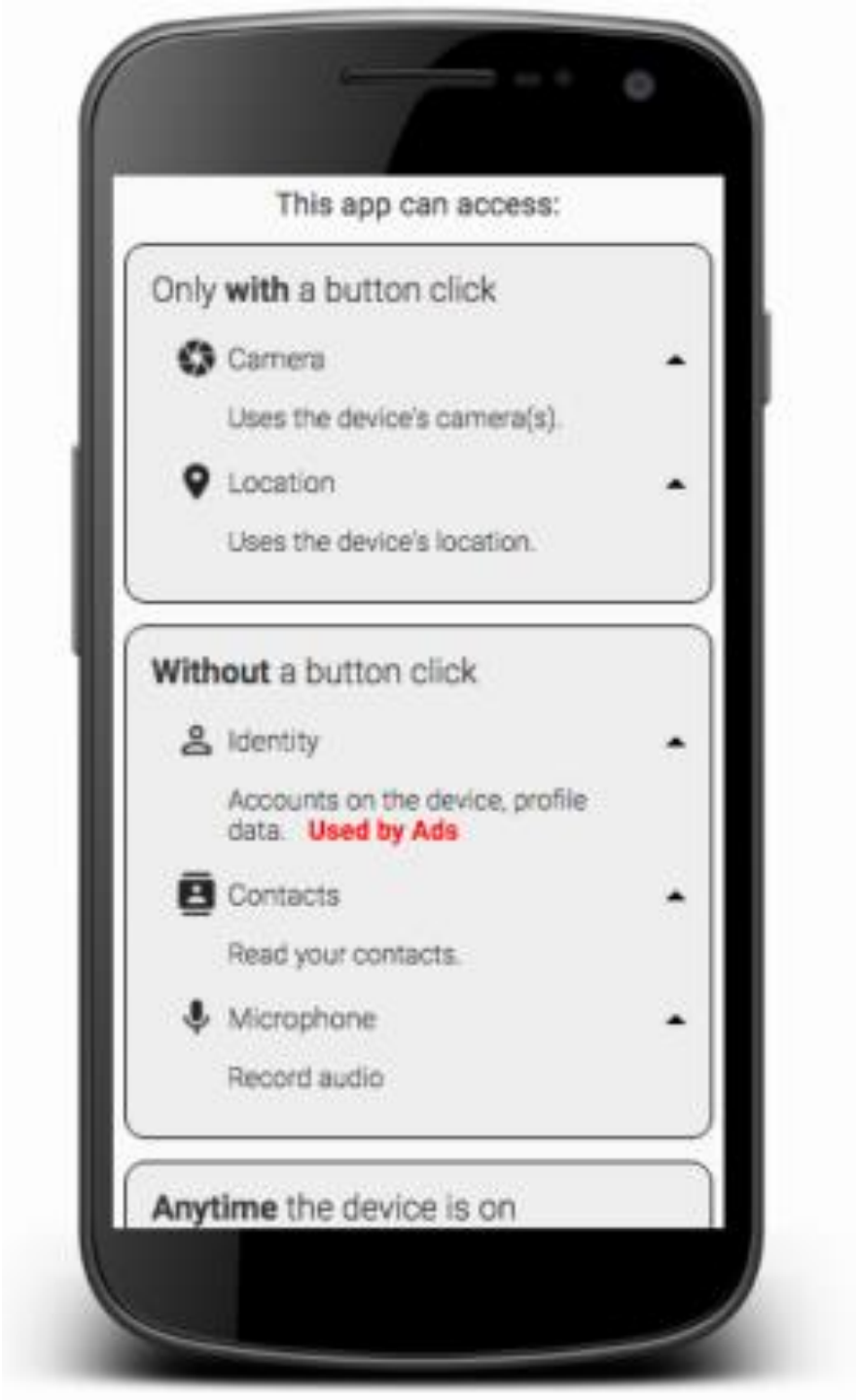
- Natural use of product
- Scripted (often lab-based) use of product
- De-contextualized / not using product
- Combination / hybrid

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

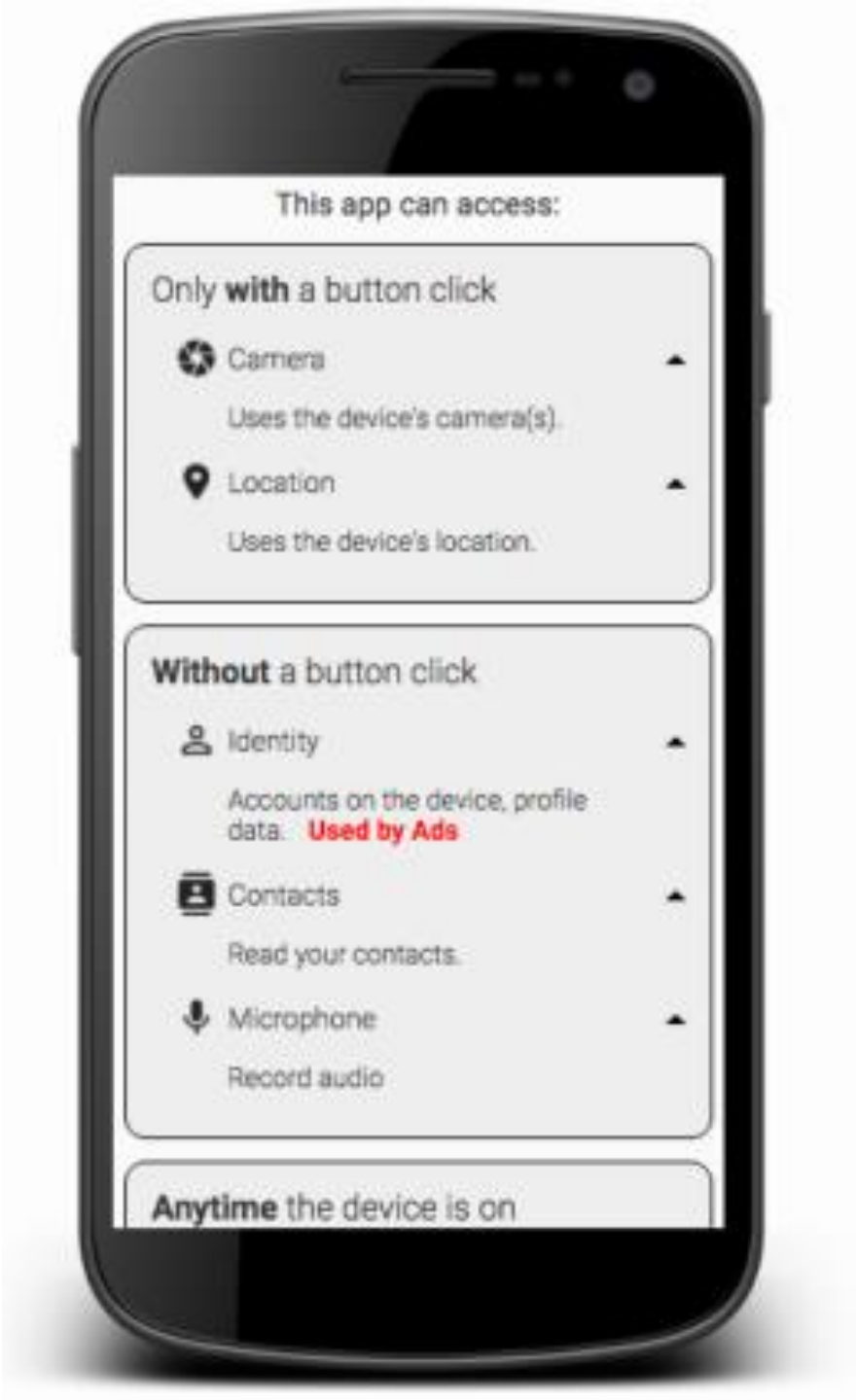
What are you going to measure?

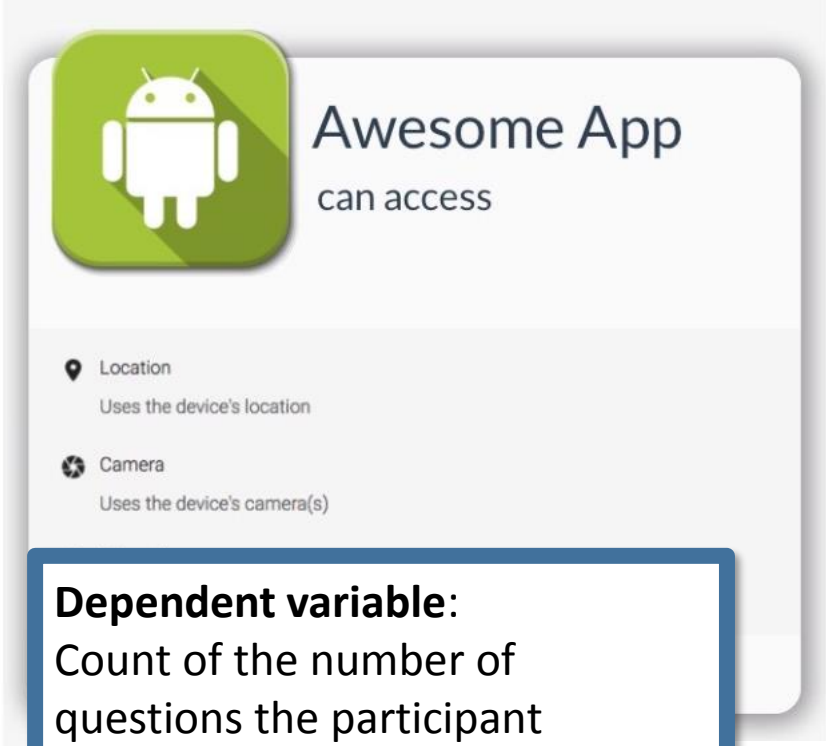
- In statistics there are classically two types of measurements (variables): dependent and independent
- Dependent
  - Also known as the outcome variable
  - Measures the usability goal
- Independent
  - Anything you are directly manipulating
  - An element of the study which is under your control
 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

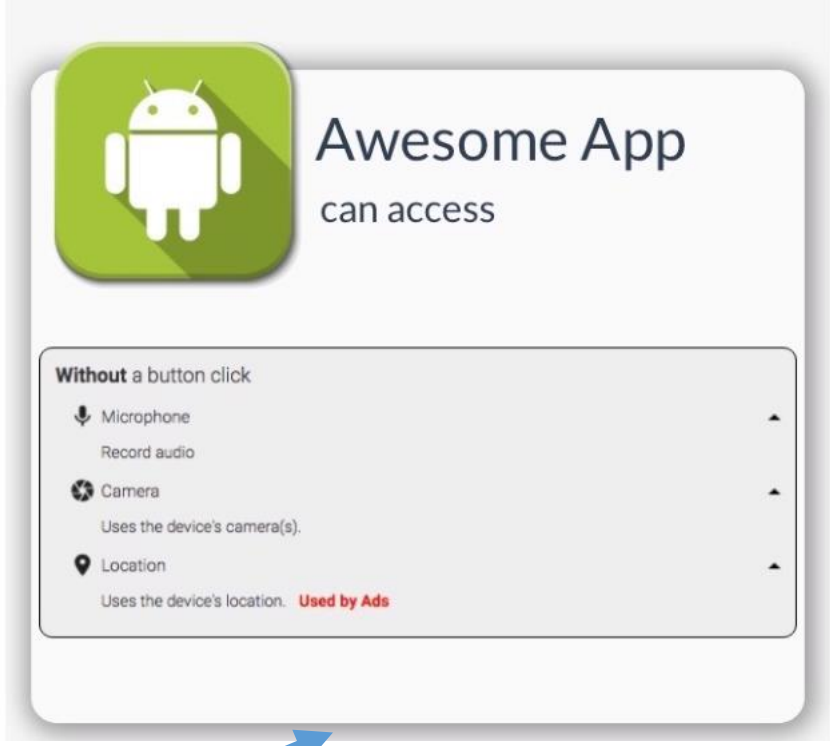
Lets use this study as an example



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







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

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

...an this app do?

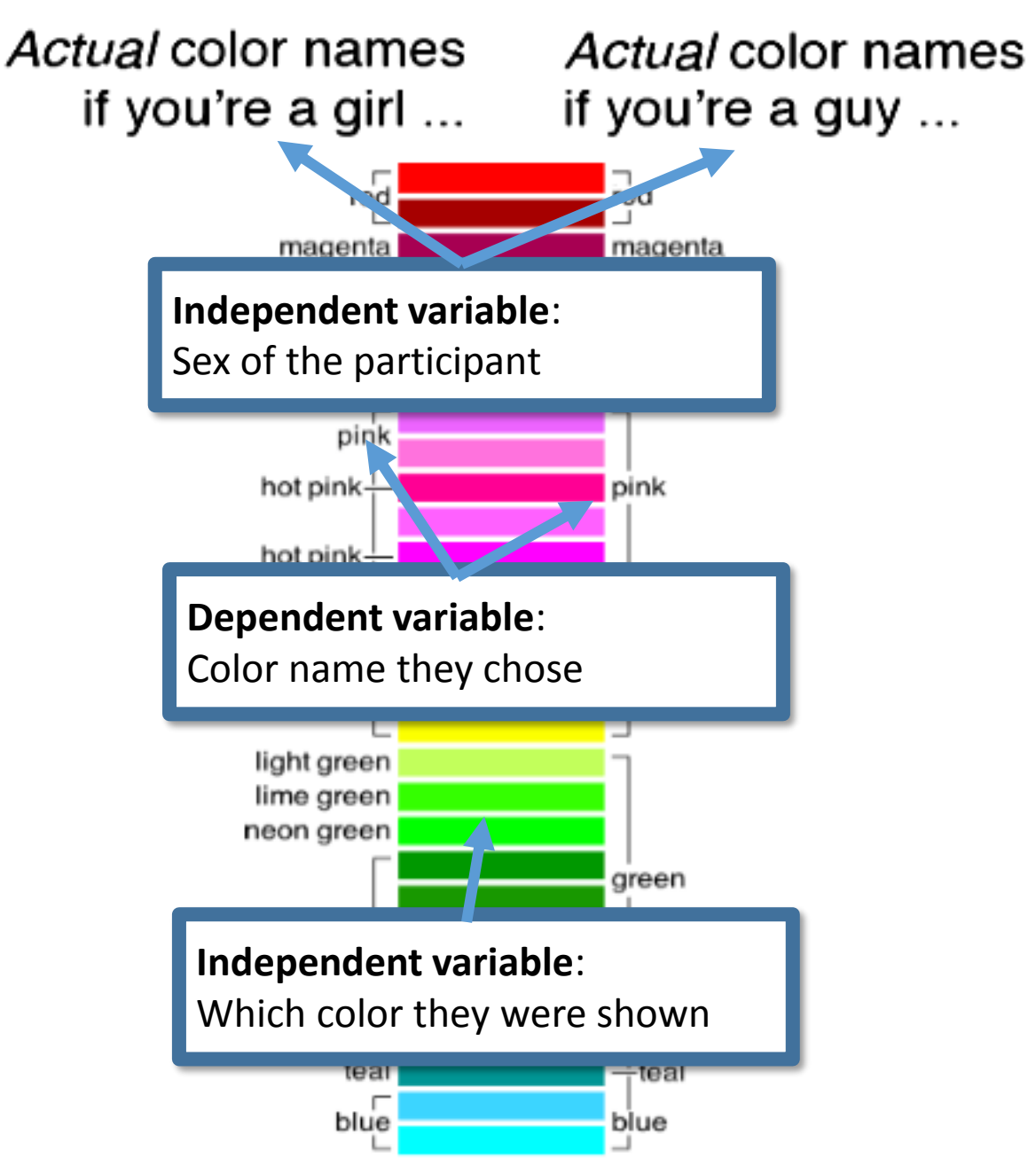
					Absolutely Possible
Charge purchases to your credit card at any time.					○
Get your location.	○	○	○	○	○
Allow ads to know your location.	○	○	○	○	○
Load ads.	○	○	○	○	○
Write on the SD card					

Variables that would make sense

- Goal: User can identify if an app can or cannot perform an action directly tied to a permission.
- Dependent
  - Number of permissions correctly/incorrectly read
  - Time spent reading each permission screen
- Independent
  - Study group
  - Order of the tasks
  - Time of day
  - Type of device (laptop, mobile, PC)
  - Demographics of the participants

XKCD ran a study to see what men and women call different colors

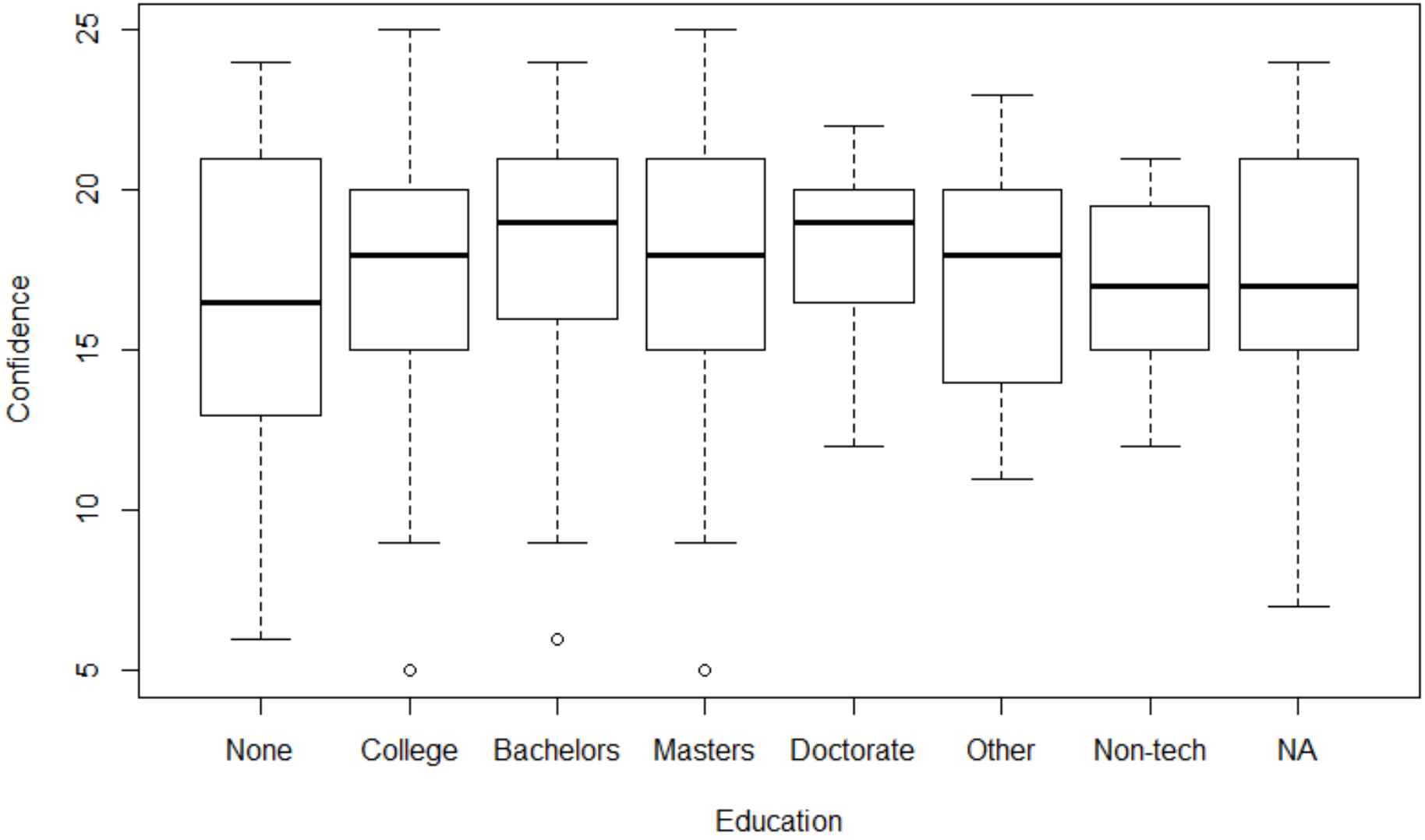
- Dependent
  - The color name they typed in
- Independent
  - Sex (man or woman)
  - Color they were shown (RGB)



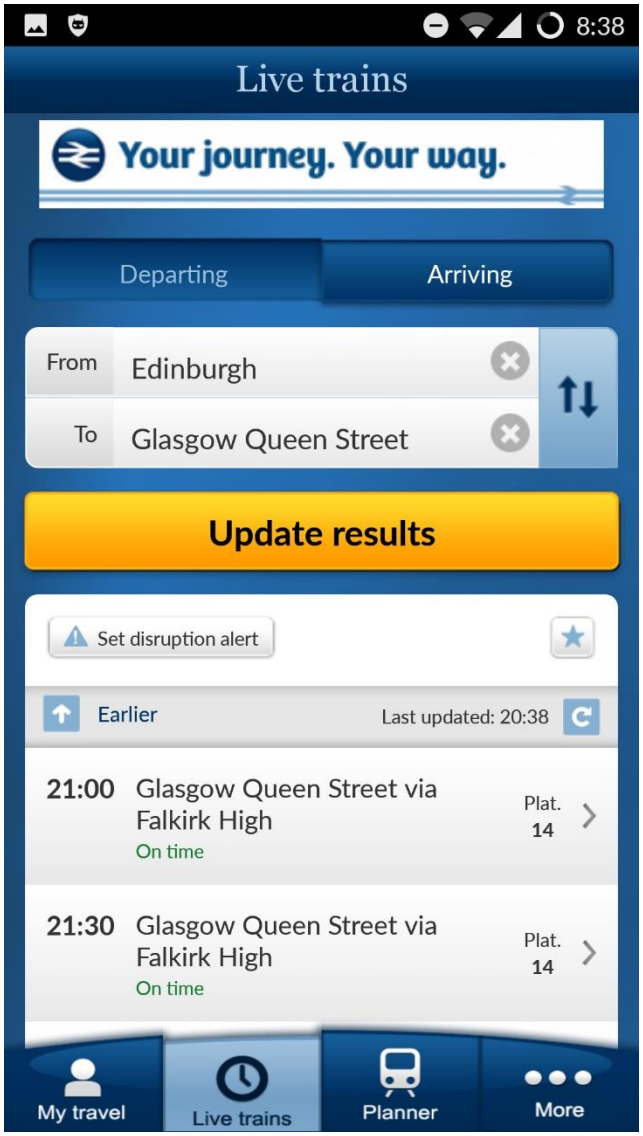
https://blog.xkcd.com/2010/05/03/color-survey-results/

MSc Project on reading config files

- Goal: Does the order of lines in a configuration file impact the way people interpret the file?
- Dependent
  - True/False – did the participant consider order
  - 1-7 – How confident were they in their answer
- Independent
  - Education level for technical professions only
  - Self-efficacy statements around programming and configuration file modification
  - Prior experience with configuration files
  - Other demographics


Confidence
5
10
15
20
25
None
College
Bachelors
Masters
Doctorate
Other
Non-tech
NA
Education

Attitudinal example



1. I think that I would like to use this system frequently.

   Strongly Agree ---------- Strongly Disagree

1. I found the system unnecessarily complex.
2. I thought the system was easy to use.
3. I think that I would need the support of a technical person to be able to use this system.
4. I found the various functions in this system were well integrated.
5. I thought there was too much inconsistency in this system.
6. I would imagine that most people would learn to use this system very quickly.
7. I found the system very cumbersome to use.
8. I felt very confident using the system.
9. I needed to learn a lot of things before I could get going with this system.

https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html

System Usability Scale

- Have the participants interact with the system
- Have them answer the questions on the right
- Follow the scale instructions
- Use the resulting number as a dependent variable

1. I think that I would like to use this system frequently.

   Strongly Agree ---------- Strongly Disagree

1. I found the system unnecessarily complex.
2. I thought the system was easy to use.
3. I think that I would need the support of a technical person to be able to use this system.
4. I found the various functions in this system were well integrated.
5. I thought there was too much inconsistency in this system.
6. I would imagine that most people would learn to use this system very quickly.
7. I found the system very cumbersome to use.
8. I felt very confident using the system.
9. I needed to learn a lot of things before I could get going with this system.

https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html

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

Common dependent things to measure

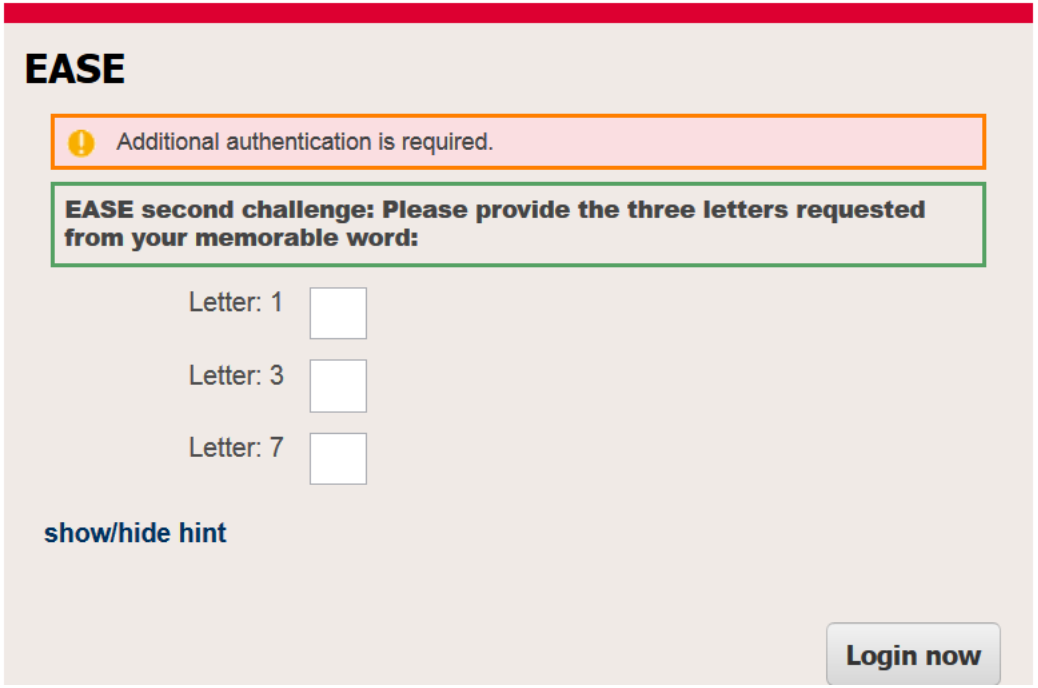
- Time to complete task
- Percent of task completed
- Percent of task completed per unit of time
- Ratio of successes to failures
- Time spent in errors
- Percent or number of errors
- Percent or number of competitors better than it
- Number of commands used
- Frequency of help and documentation use
- Percent of favorable/unfavorable user commands

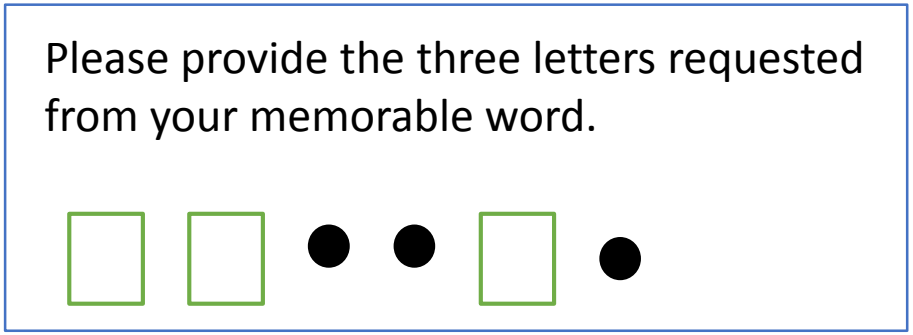
Common dependent things to measure

- 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

Think-pair-share

- Which partial passcode entry is more usable?
- How might you define “usable”
- What are the dependent variables?
- What are the independent variables?





Step 3: Setup your study

What do you want to be able to say after the evaluation is done?

- X interface is better than Y interface
  - Run an A/B study
  - Randomly assign users to groups
  - Have all users complete the same tasks
- My new interface is better than my old interface
  - Same as above
  - Or use rapid usability approach
- Users can use interface X to accomplish Y
  - Have users accomplish a set of tasks using X
  - Measure the usability (see step 2)
- Using my interface makes people better/smarter
  - Pre/post test – give them the same test before and after using your system

Between vs. Within subjects

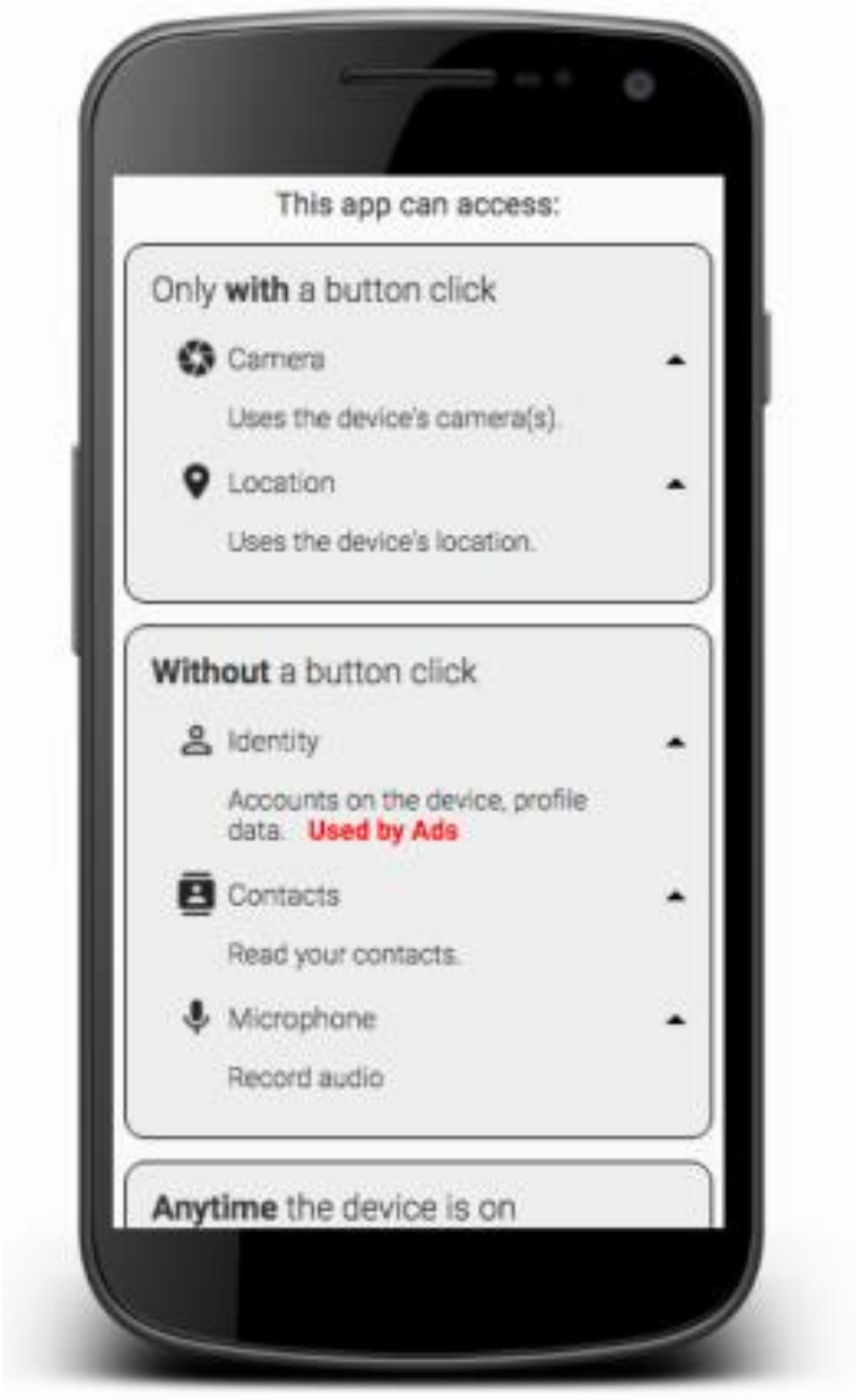
- Between subjects
  - Your study only shows one interface to one person
  - You are measuring how well the people randomly assigned to the A interface did compared to the people randomly assigned to the B interface
  - Lots of variability with this method
- Within subjects
  - Your study shows all interfaces 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

Scripted vs observational

- Scripted studies are planned 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

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



Study design

- Between subjects
- Multiple tasks
- Dependent
  - The color name they typed in
- Independent
  - Sex (man or woman)
  - Color they were shown

https://blog.xkcd.com/2010/05/03/color-survey-results/

Step 4: Evaluate the outcome

Evaluation options

- Basic
  - Counts of effectiveness on tasks
- Academically sound
  - Statistics

Basic version

- 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

We are about to learn about some of the basic statistics used in HCI

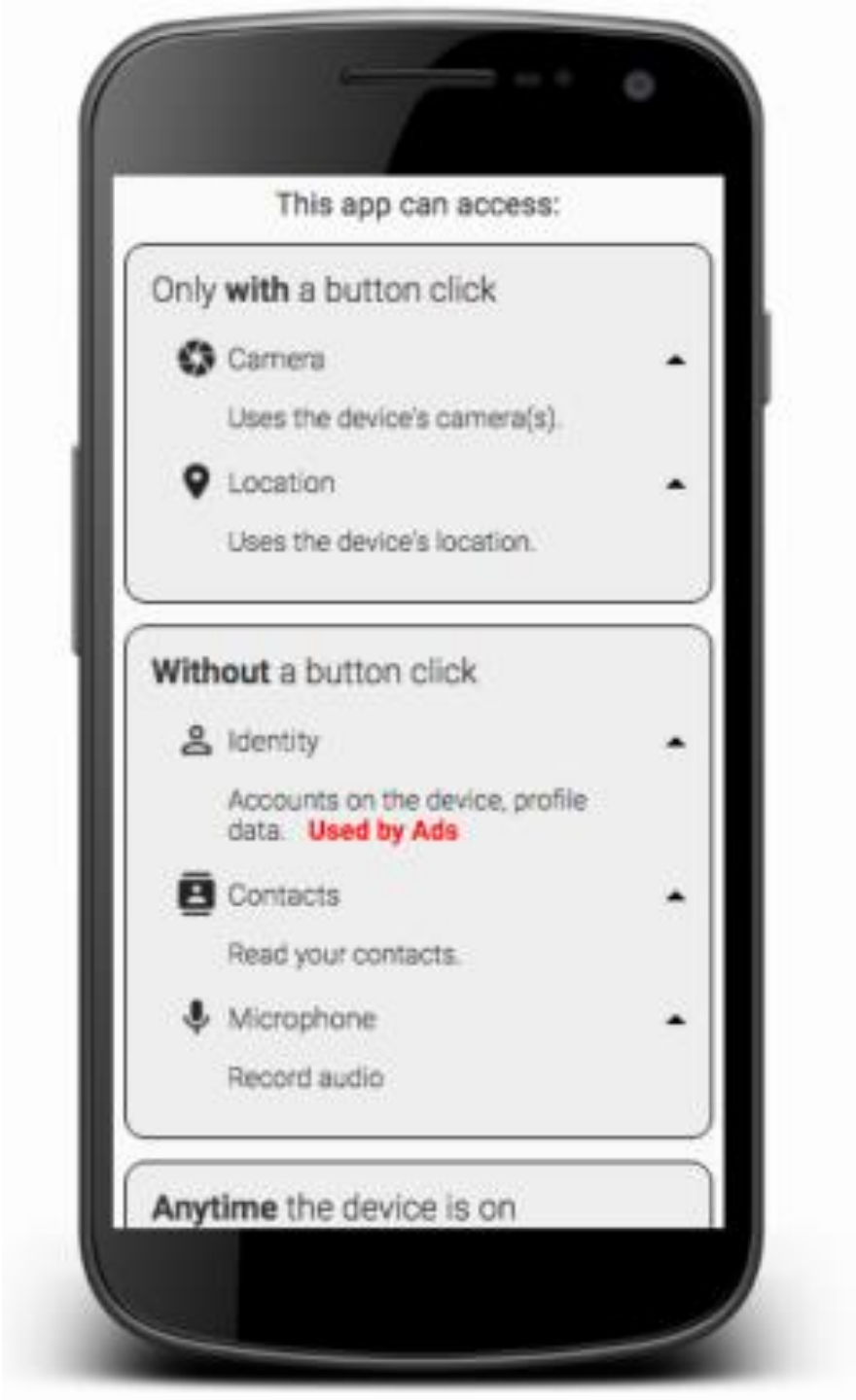
These are only needed if you want to scientifically prove that a statement is true

Types of data

- Numeric
  - **Continuous** – Any value on the range is possible including decimal
  - **Discrete** – Only certain values on the range are possible
  - **Interval** – Only certain values on the rage are possible and each has equal distance from its neighboring values
- Categorical
  - **Binary** – Only two possibilities
  - **Ordinal** – The values have an ordering (slow, medium, fast)
  - **Nominal** – The values have no ordering (apple, pear, kiwi, bannana)

Study design

- Accuracy on task
  - Categorical ordinal
- Which interface
  - Categorical binary



	Absolutely Impossible	Impossible	Neutral	Possible	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					

Study design

- Dependent
  - The color name they typed in
  - Categorical - nominal
- Independent
  - Sex (man or woman)
  - Color they were shown
  - Categorical – ordinal
  - Could be argued that it is integral based on RGB

https://blog.xkcd.com/2010/05/03/color-survey-results/

The next slide will not appear on any tests, it is here for your future reference

Comparing	Dependent	Independent	Parametric (Dependent variable is mostly normally distributed)	Non-parametric
The means of two independent groups	Continuous / scale	Categorical / nominal	Independent t-test	Mann-Whitney test
The means of 2 paired (matched) samples	Continuous / scale	Time variable (before/after)	Paired t-test	Wilcoxon signed rank test
The means of 3+ independent groups	Continuous / scale	Categorical / nominal	One-way ANOVA	Kruskal-Wallis test
3+ measurements on the same subject	Continuous / scale	Time variable	Repeated measures ANOVA	Friedman test
Relationship between 2 continuous variables	Continuous / scale	Continuous / scale	Pearson's Correlation Coefficient	Spearman's Correlation Co-efficient
Predicting the falue of one variable from the value of a predictor variable	Continuous / scale	Any	Simple Linear Regression	
Assessing the relationship between two categorical variables	Categorical / nominal	Categorical / nominal		Chi-squared test

Chi Squared

- $\chi^2 = \sum \frac{(Observed\ Value - Expected\ Value)^2}{(Expected\ Value)}$
- Answers the question:
  - Does the observed data have the same ratio as expected

OR

  - Do two counts come from the same distribution

Questions?