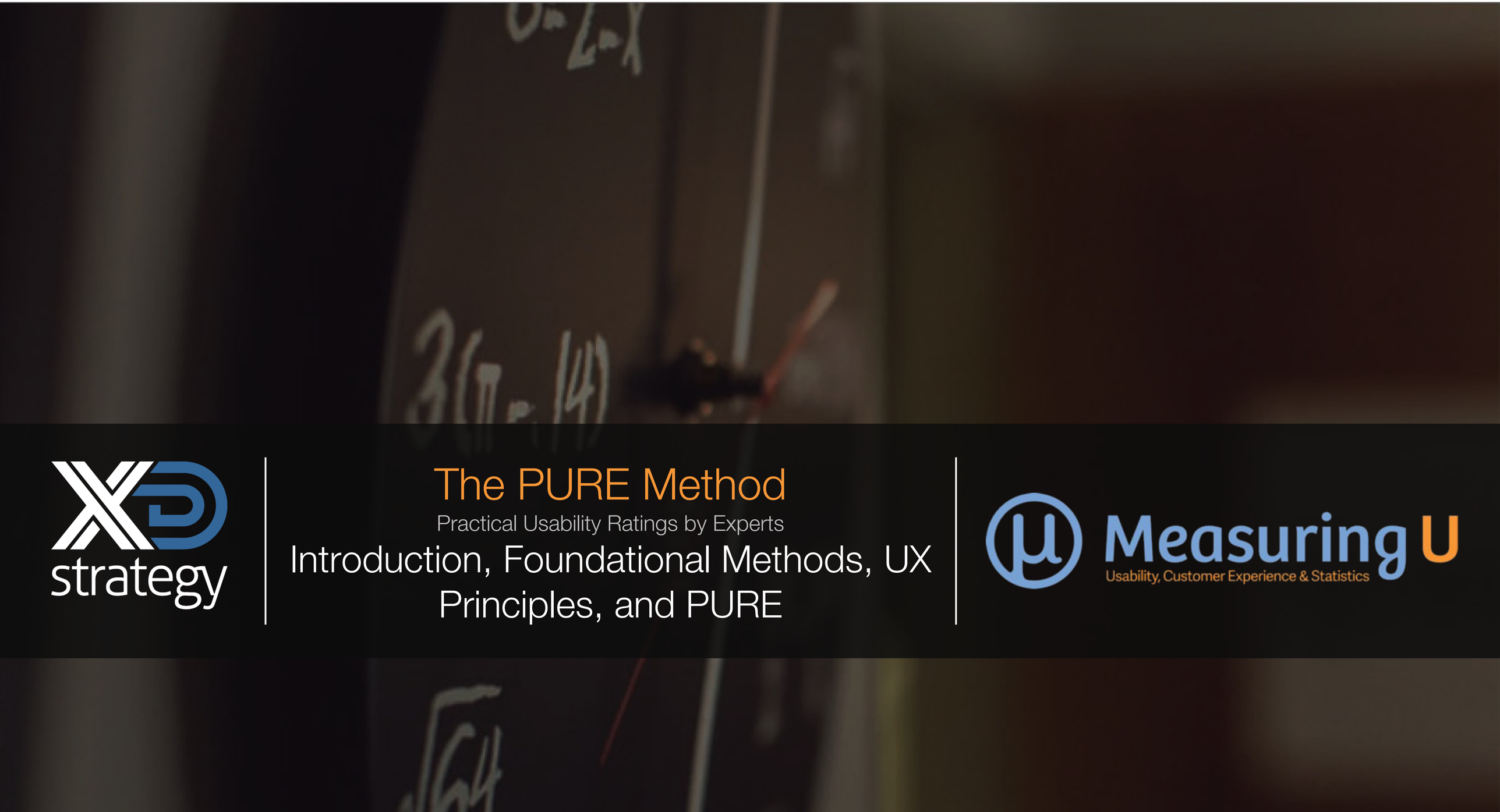


XD strategy

The PURE Method

Practical Usability Ratings by Experts

Introduction, Foundational Methods, UX Principles, and PURE

Measuring U

Usability, Customer Experience & Statistics

Introduction

QUESTIONS?

- Periodically, I will pause and take look at questions the Moderator(s) has/have been collecting. When you see this slide, that's what I'll be doing.
- Type your questions into the Zoom Chat and the Moderator will put them into a separate list for me to review. This is so similar questions can be combined.
- Questions that are out of scope we will be considered outside of the class.

Education and Experience

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.

1987 — UC Santa Cruz
Computer & Information Sciences (Honors BA, Phi Beta Kappa)

1992 — Stanford University
Cognitive Science, HCI, Education (PhD)

1998 — Usability Engineer
Director of User Experience Research

1999 — Senior Director of User Experience Design

2004 — VP of Enterprise Design & Research

2008 — SurveyMonkey
Interim VP, Product Design

2011 — UserZoom
Industry Advisor

2015 — Founder & Principal

2018 — Lecturer, School of Computing & Design

1987 — The Santa Cruz Operation (aka Caldera)
Unix/Xenix Tech Support Engineer & Trainer

1998 — Network Computer, Inc. (aka Liberate)
Ethnographic Field Researcher

2004 — Director of UER and Product Research

2011 — Vice President & Chief Design Officer

2018 — Invited Industry Instructor

Tech	Research	Design	Coach, Consult, Teach, Speak, Write

WHAT ARE WE ASSESSING TODAY?

- Usability?
- User Experience?

REGARDLESS, WE NEED TO DEFINE THEM FIRST

USABILITY DEFINED

"the quality or state of being usable : ease of use"
-Merriam Webster

"The **ease** of use and **learnability** of a human-made object such as a tool or device. In software, usability is the degree to which a software **can be used** by specified consumers to **achieve** quantified **objectives** with **effectiveness, efficiency, and satisfaction** in a quantified context of use"
-Wikipedia

According to Jakob Nielsen and NN/g: "Usability is defined by 5 quality components:

1. **Learnability**: How easy is it for users to accomplish basic tasks the first time they encounter the design?
2. **Efficiency**: Once users have learned the design, how quickly can they perform tasks?
3. **Memorability**: When users return to the design after a period of not using it, how easily can they reestablish proficiency?
4. **Errors**: How many errors do users make, how severe are these errors, and how easily can they recover from the errors?
5. **Satisfaction**: How pleasant is it to use the design?

THE ISO SAYS...

"The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use." -ISO 9241-11

Unpacking ISO 9241-11 a bit further yields:

effectiveness: accuracy and completeness with which users achieve specified goals

efficiency: resources used in relation to the results achieved

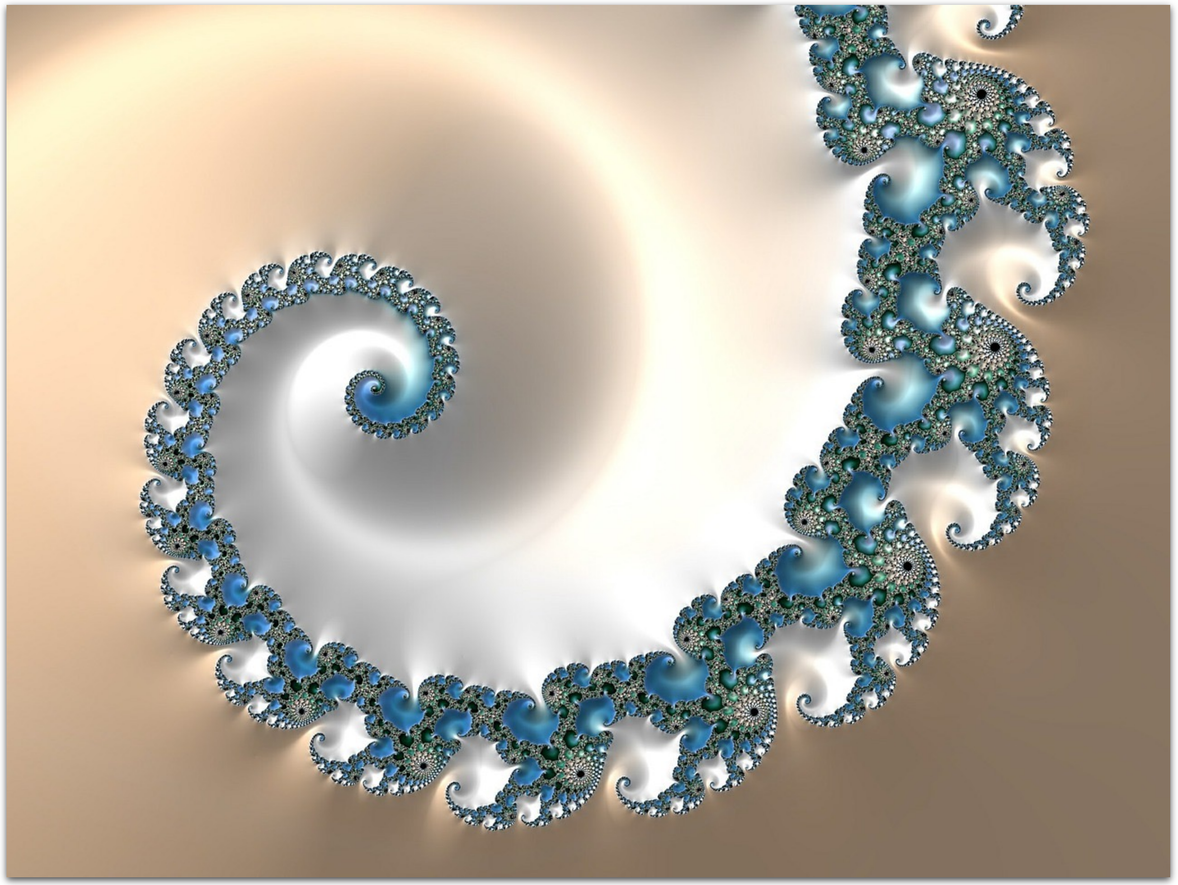
satisfaction: extent to which the user's physical, cognitive and emotional responses that result from the use of a system, product or service meet the user's needs and expectations

context of use: combination of users, goals and tasks, resources, and environment (including the technical, physical, social, cultural and organizational environments)

It is challenging to measure or assess all of these components.

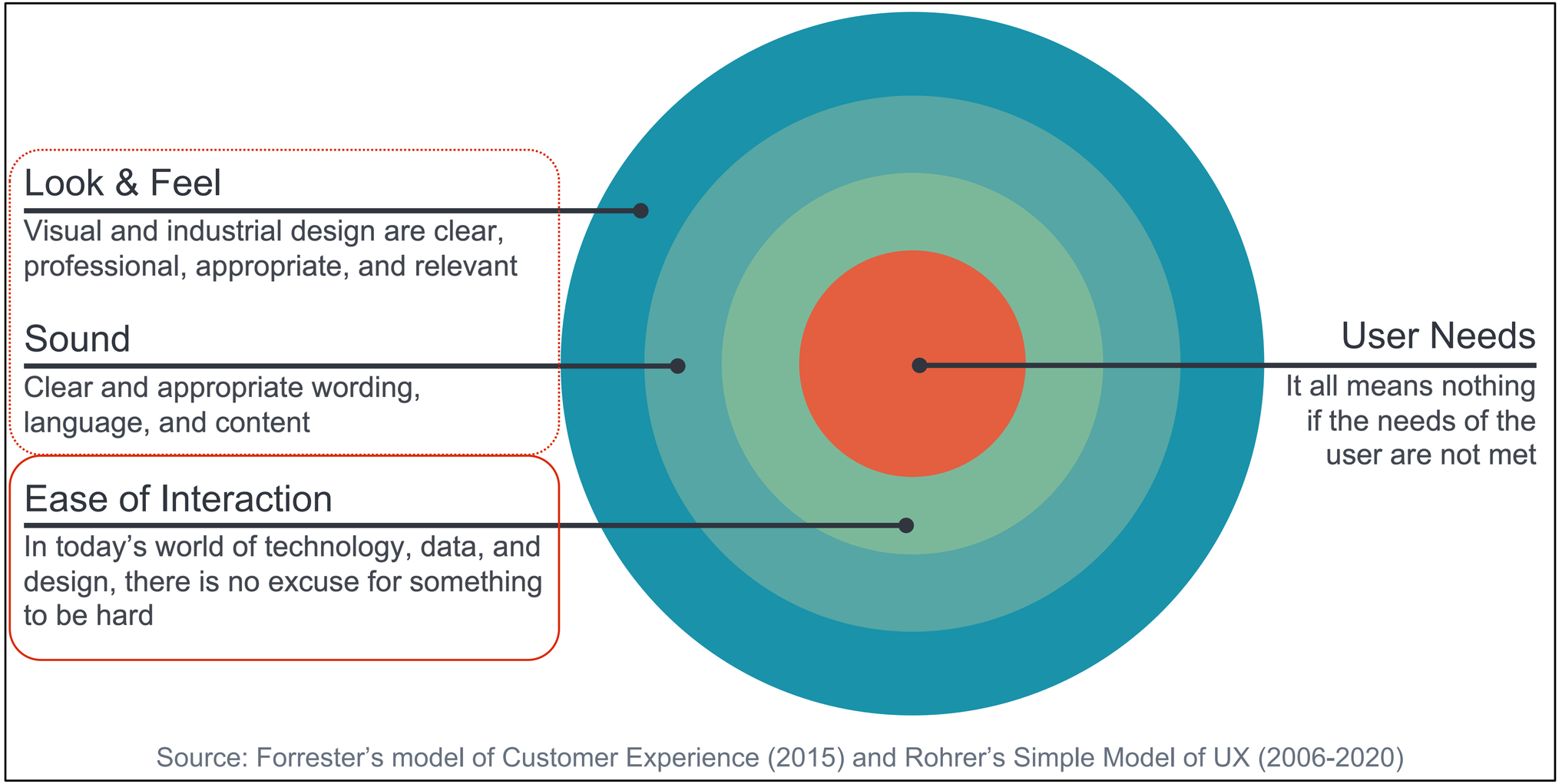
According to various authorities, User Experience is basically… “all aspects of every interaction”

NN Group: “UX encompasses **all aspects** of the end-user’s **interaction** with the company, its services, and its products.”
The ISO: “person's **perceptions and responses** resulting from the **use** and/or anticipated use of a **product, system or service**”
UXPA: “**Every aspect** of the user's **interaction** with a **product, service, or company** that make up the user's perceptions of the whole."
Wikipedia: “a person's **emotions** and **attitudes** about using a particular product, system or service"
IXDA: (They don’t offer a definition, but refer to the above. They reiterate the idea that user experience is “**everything**.”)



OK, let’s assess…***everything!*** Wait, can we do that easily?

A Simple Model of User Experience



Most UX Assessments tend to focus on Ease of Interaction, and some aspects of Look, Feel & Sound
User Needs are hard to assess this way, because they vary by user type, domain and value proposition

WHAT ARE "UX ASSESSMENTS" / "USABILITY INSPECTIONS" / "EXPERT REVIEWS"?

- Similar in spirit to code inspections
- Trained and qualified experts review a product and design from a user-centered perspective of some kind
- This "perspective" is driven by some of the following:
  - Familiarity with certain user-types (e.g., personas)
  - Knowledge of UX principles and heuristics
  - Asking critical user-centered questions about the experience at each step
  - The ability to review the design or product against known design and editorial guidelines
- Including preparation, takes 1-5 days to perform
- Typical output is a list of identified issues, details on why they are an issue, potential design solutions, and/or a UX scorecard
- Doesn't find all issues, but does find many that should be addressed (30-90% of what can be found in a usability study)
- Some issues found may be false positives
- Other issues are "in the long-tail" (i.e., only detectable in a large study, not in a typical usability study), which can be useful
- There are many types of UX Assessments - let's see how they compare

HOW DO UX ASSESSMENTS COMPARE?

	UX Teardown	Cognitive Walkthrough	Pluralistic Walkthrough	KLM-GOMS	Heuristic Evaluation	The PURE Method
Use this method when your main focus is to…	Provide an outside critique of an experience, often by reverse engineering a product's intent and approach	Assess learnability and ease-of-use for new users	Have a cross-functional team that wants to see the experience from the point of view of the end-user	Assess the estimated time (and difficulty) of tasks at a very low-level (motor skill and cognitive units)	Find as many potential usability problems across the experience/UI, based on heuristics (principles)	Score and diagnose the friction in a product for its most important customers across its Fundamental Tasks
Perspective	External critic	Internal	Internal	Internal or external	Internal or external	Internal or external
Evaluator Characteristics	UX professional, often with a specific area of expertise	Anyone who can learn the basics of user learning theory	Any role (designer, developer, UX professional)	Anyone with good attention to detail	UX expert well versed in heuristics and design principles	UX expert well versed in heuristics and design principles
Learning Curve	Moderate	Low-Moderate	Low-Moderate	Moderate	Moderate-High	Moderate-High
Scripted	No - experience depends on where the reviewer happens to go	Yes - only correct path is followed	Yes - only correct path is followed	Yes - can be applied to any usage sequence	No - reviewer passes through experience as they wish	Yes - start with the Happy Paths for the Fundamental Tasks; can expand later
Output	Sequential screens with callouts and answers to key questions about users and intent	List of potential learnability issues, along with rationale	List of potential user experience issues	Sequence of low-level actions; numeric score estimating time required for task	List of potential issues with associated heuristic violations; severity ratings	Scorecard on friction (ease-of-use) for Tasks and Entire product; also, details on score rationales
Metrics	None	None	None	Estimated task time	Severity ratings of issues found	Friction/Ease of Use Scorecard

QUESTIONS?

- You may now ask questions from this section
- You may type your question into Zoom Chat
- We may have to cover some questions as follow-ups after the seminar

UX Tear Downs

WHAT IS A UX TEAR DOWN?

- A review of an experience, shown screen by screen, that critiques the experience from a perspective outside the company that built it, often to understand their intended target users and business goals
- Key questions are asked, as the experience is reviewed (see example list)
- The author as critic provides his or her perspective of the site, based on how he or she answers to these questions
- My main critique: unless you have credibility, **it lacks objectivity**, but it can be very useful, if done well
- One of the more well-known (and recommended) set of tear downs comes from UserOnboard by Samuel Hulick, who focuses on the on-boarding experience for new users

Questions that are typically asked:

- Who do we think are the target users, needs and their goals?
- For new users, does the site or app "onboard" its users well and is it easy to learn to use?
- For experienced users, does the site or app work efficiently?
- What does the site or app do well in terms of good UX practices?
- Does the site or app have clear violations of good UX practices?
- Does it meet the needs of the target users or not? If not, why not?

UX Teardown Examples covering the user on-boarding from Samuel Hulick (pick one):

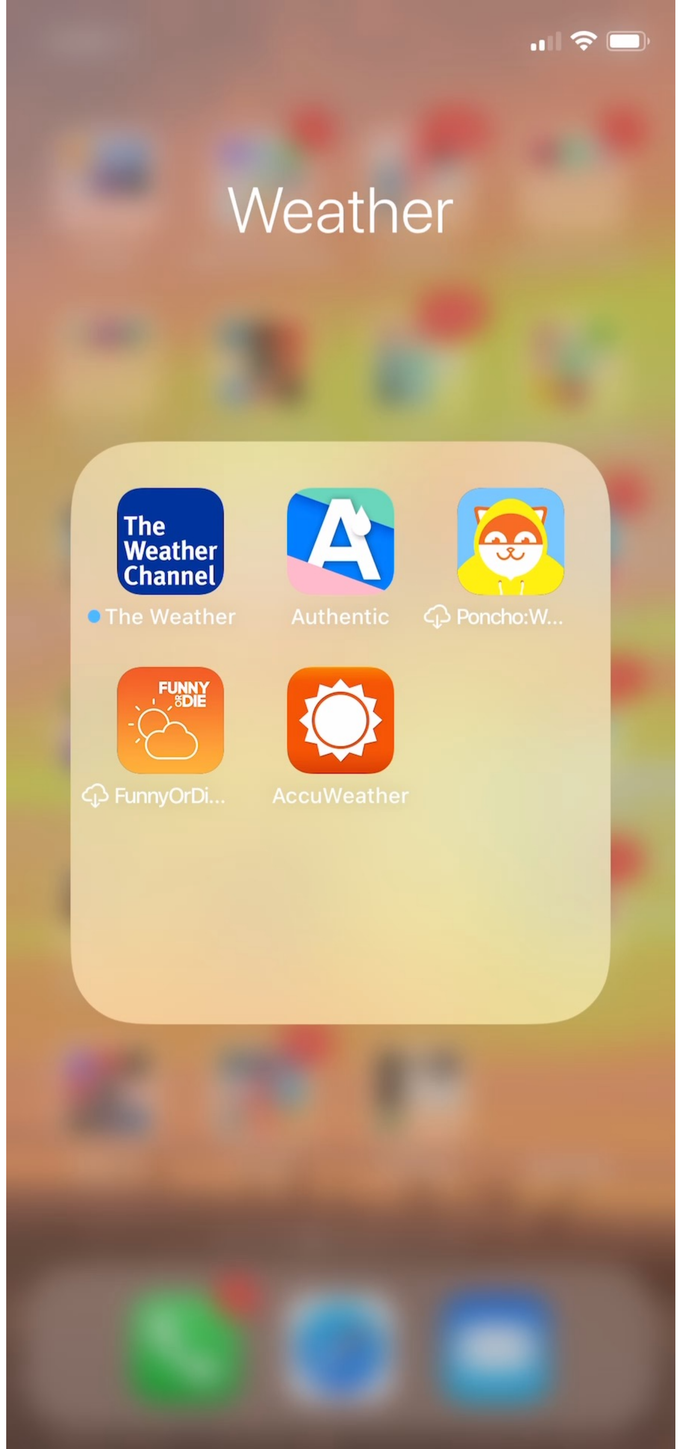
Dropbox | QuickBooks | Waze

Why Samuel Hulick's Teardowns? Two reasons: He's a UX expert on the onboarding process, so has a lot of domain knowledge there, and onboarding is something we can ALL relate to.

LET'S TRY TO DO A UX TEARDOWN

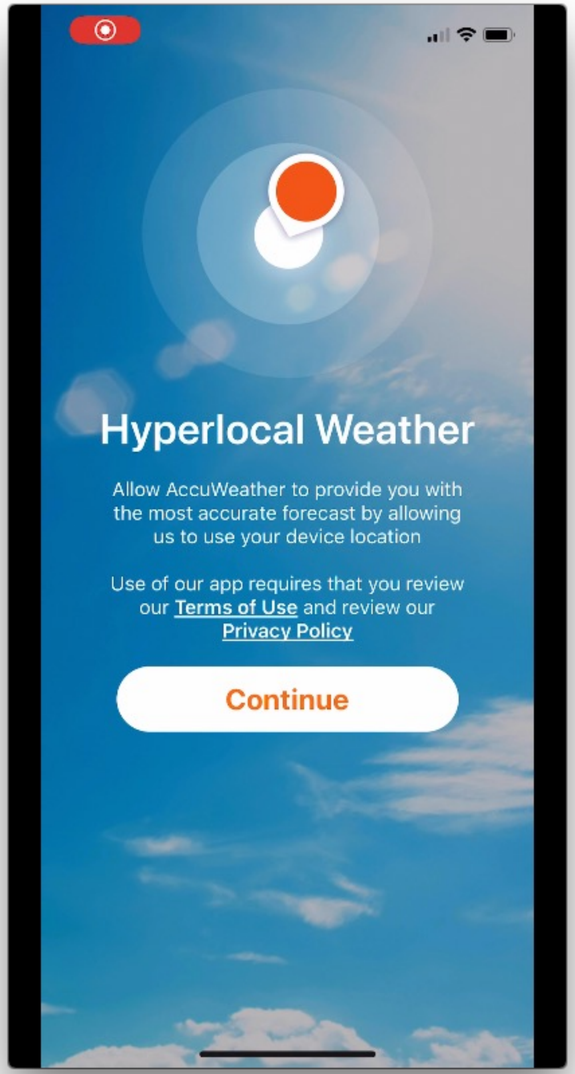
- AccuWeather on iPhone XS
- This is the first time experience after loading it from the App Store
- Adding a city that we are NOT in now (Los Angeles)
- Seeing today's (Thu Mar 12) weather
- Then checking the weather for that city for next weekend (Sat Mar 21)

Here's a 1:12 video of that experience:

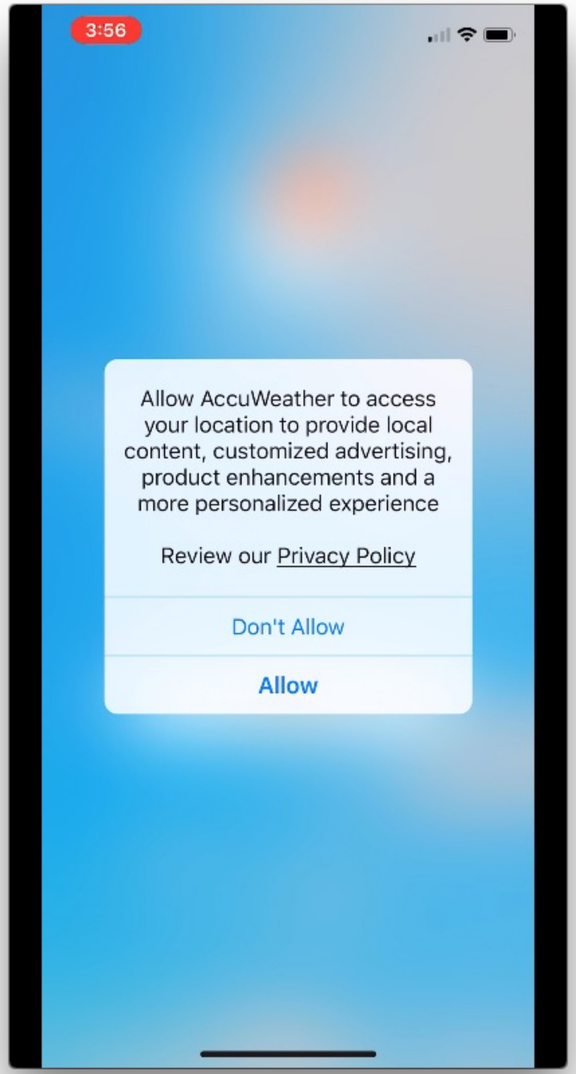


Exercise documents: http://bit.ly/nng-ux-assessments

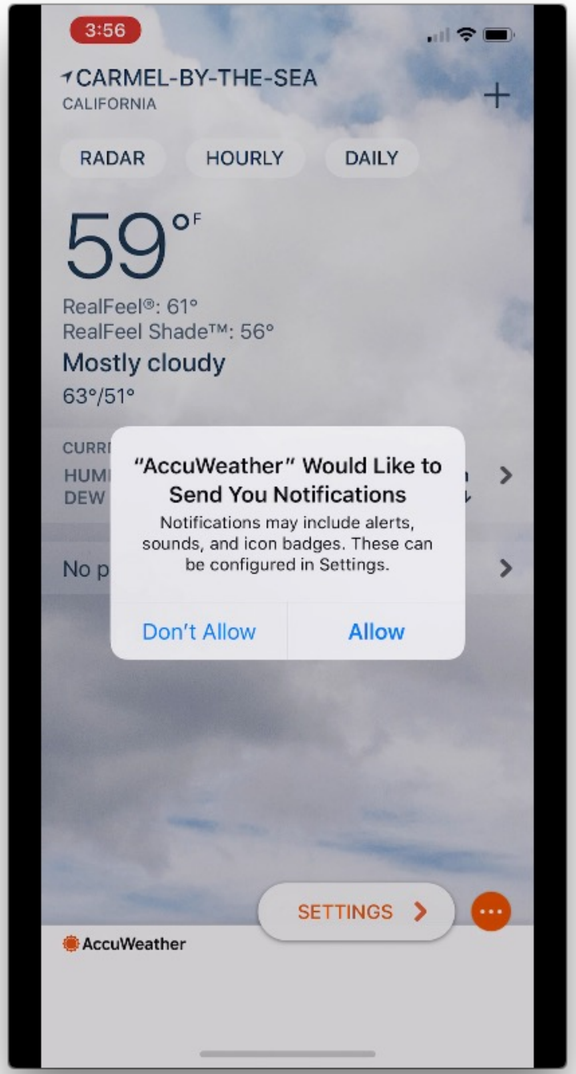
Tap on Continue



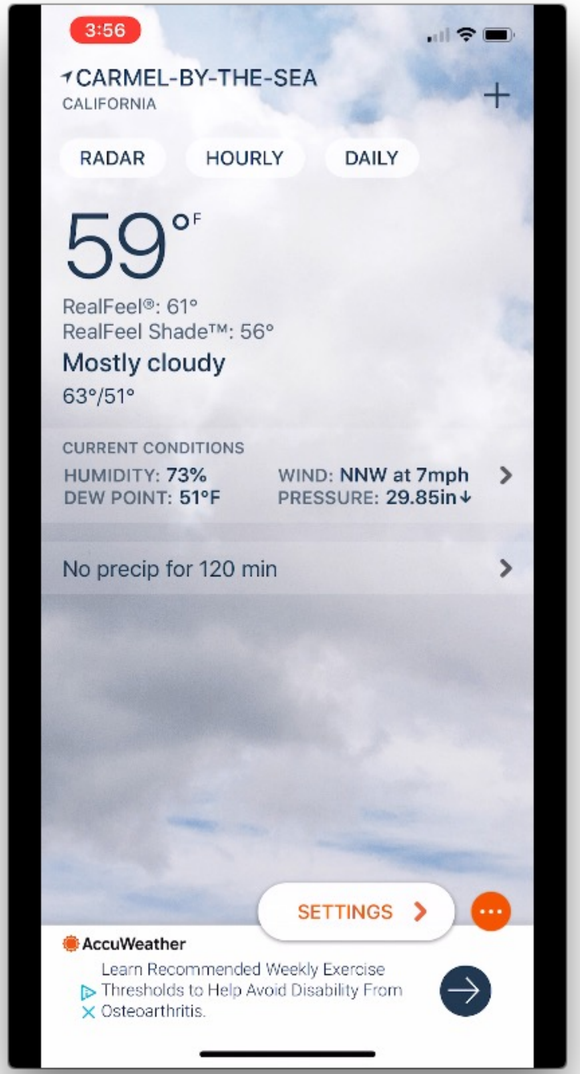
Tap on Don't Allow



Tap Don't Allow

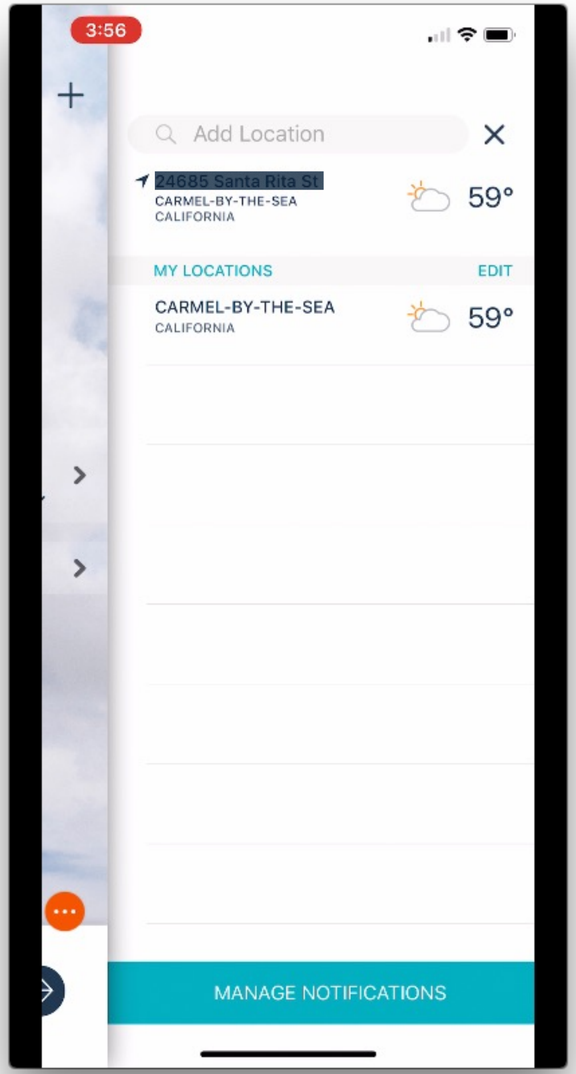


Tap '+' to add city

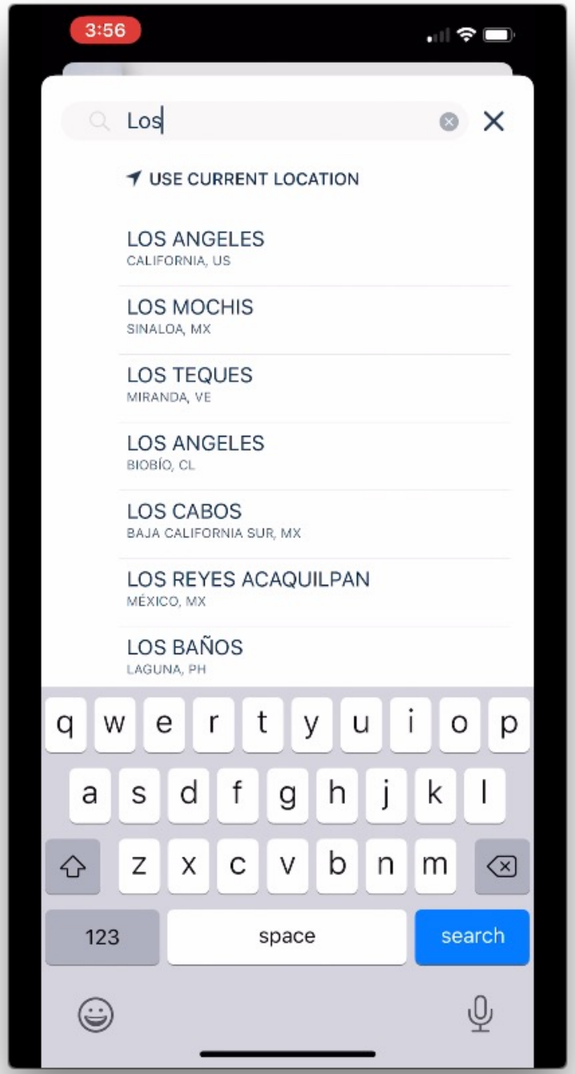


Exercise documents: http://bit.ly/nng-ux-assessments

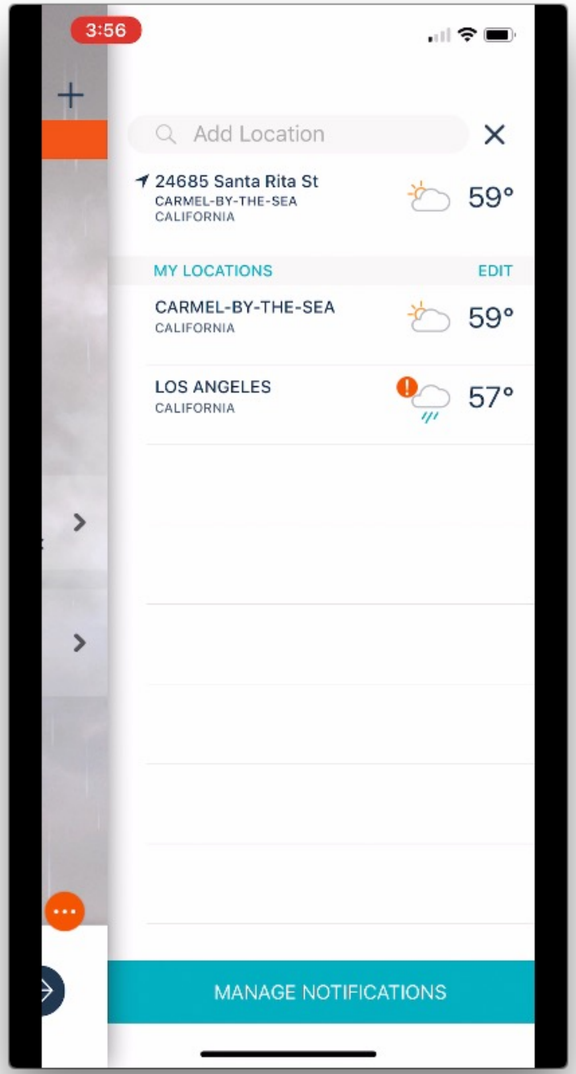
Add location screen



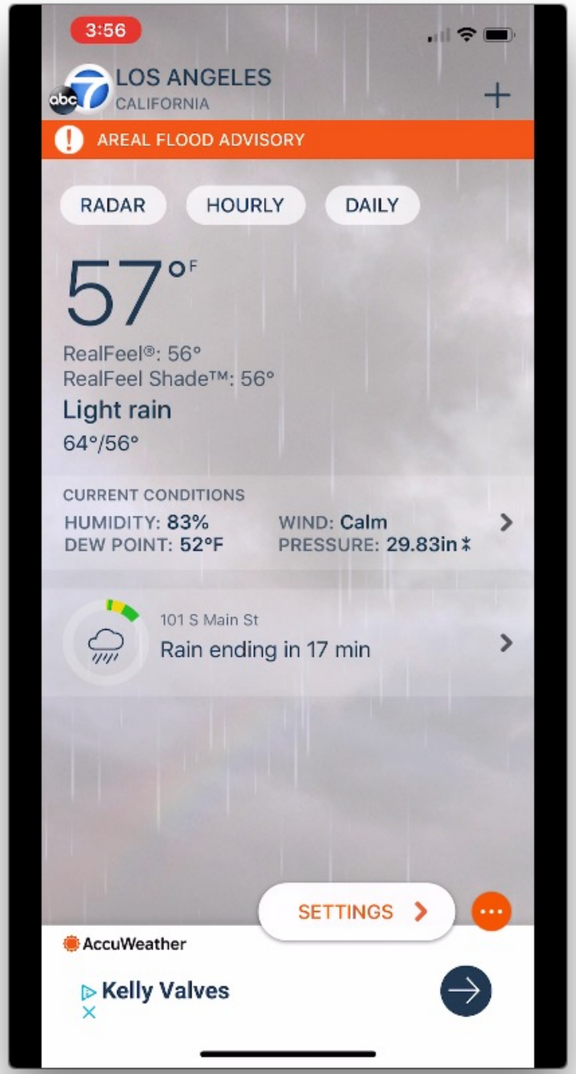
Type in “Los” then tap LOS ANGELES



Tap LOS ANGELES

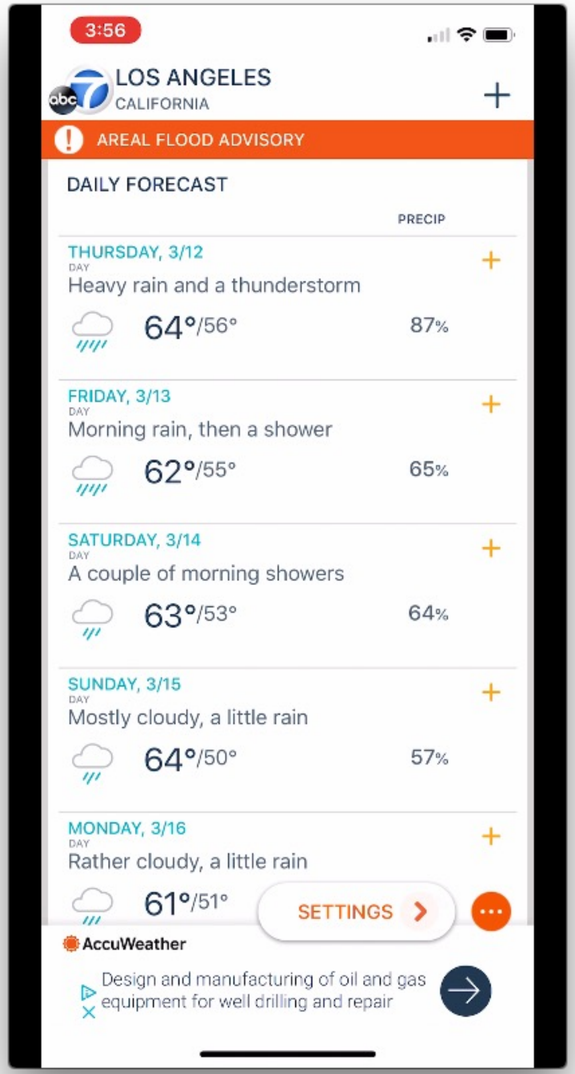


LA weather; tap DAILY

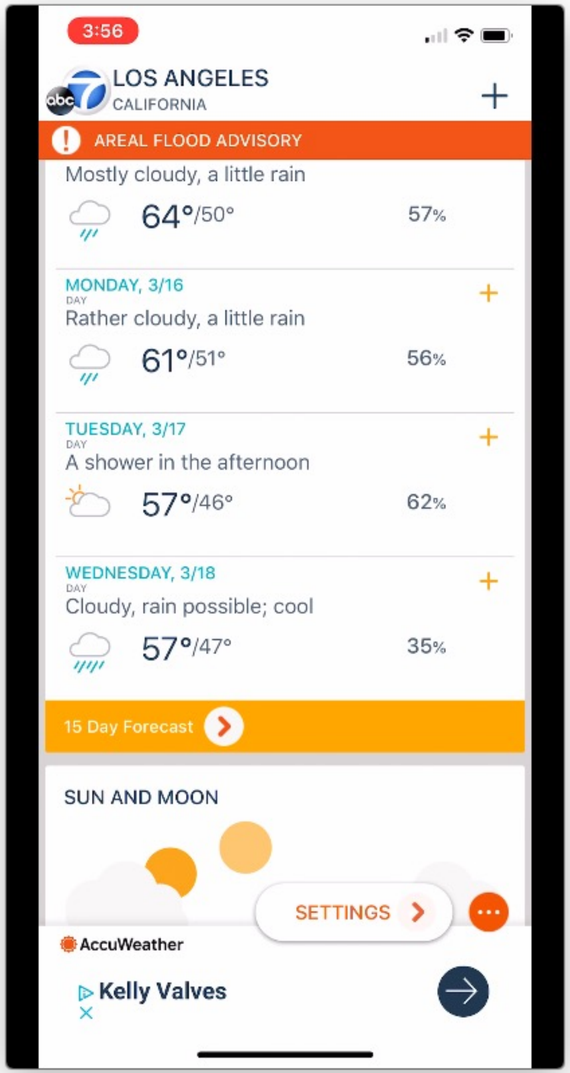


Exercise documents: http://bit.ly/nng-ux-assessments

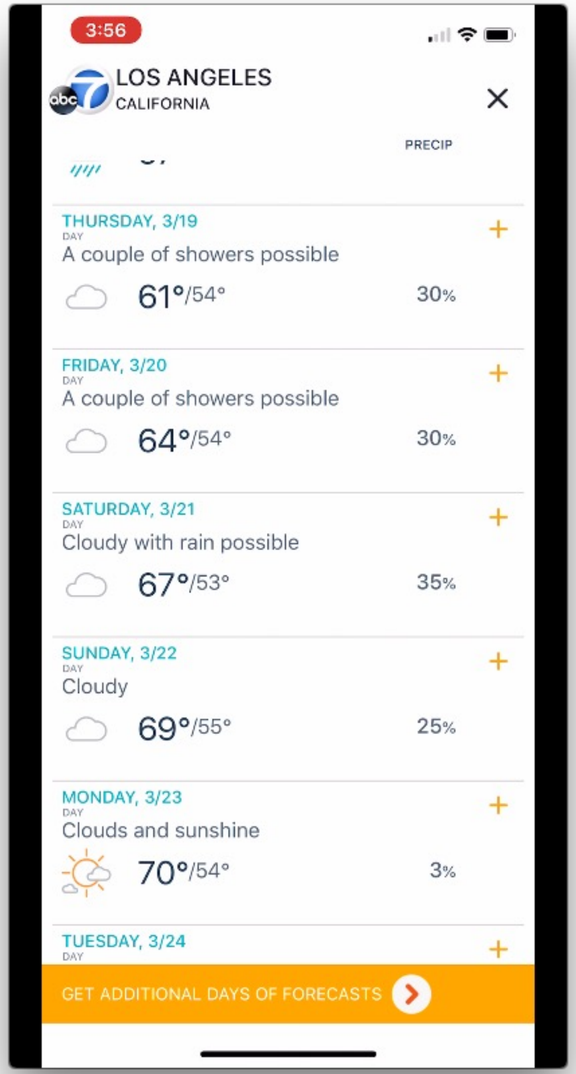
Scroll through 7 days



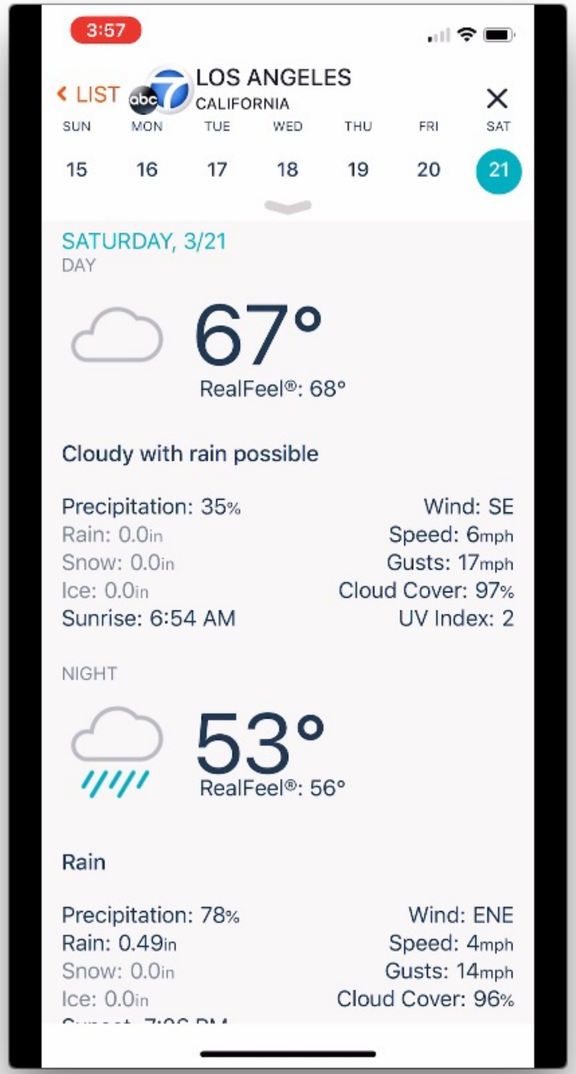
Tap 15 Day Forecast



Tap SATURDAY, 3/21



Review weather on Sat 3/21



Exercise documents: http://bit.ly/nng-ux-assessments

WHAT DID YOU THINK?

- Would this approach work for you in your organization?
- What would be needed for this to work for you?
- What knowledge is needed to conduct it at your organization?

QUESTIONS?

- You may now ask questions from this section
- You may type your question into Zoom Chat
- We may have to cover some questions as follow-ups after the seminar

Cognitive Walkthrough

WHAT IS A COGNITIVE WALKTHROUGH?

- A Cognitive Walkthrough is a usability inspection method that was specifically designed to assess **learnability** and **ease-of-use for new users**.
- It was inspired by and developed based on a theory of human learning and problem solving called CE+ (Polson & Lewis, 1990)
  - Main assumption: users learn a system by trying actions in a task they believe will help them reach their goal, and they monitor whether they are making progress or not based on system response.
- Focuses on "perfect performance" of the task, not mistakes they could make.

What to know

- **Level of difficulty:** Low-Moderate
- **Evaluators needed**: Single or group (better)
- **Evaluator characteristics:** Need to learn:
  - The basics of learning and problem solving methods implied by the theory
  - The questions that CW asks, and
  - How to walk through the system to get the answers
- **Time required**: Prep: 1-2 days; Analysis: < 1 day

WHEN TO USE COGNITIVE WALKTHROUGH

- When you want to focus on learnability
- If you have the system and the task flows available (hardcopy of screens are fine)
- A good complement to Heuristic Evaluation
- Note: It does not account for:
  - Expert user usability
  - All paths or parts of the system
  - Errors or difficulty encountered

Pros

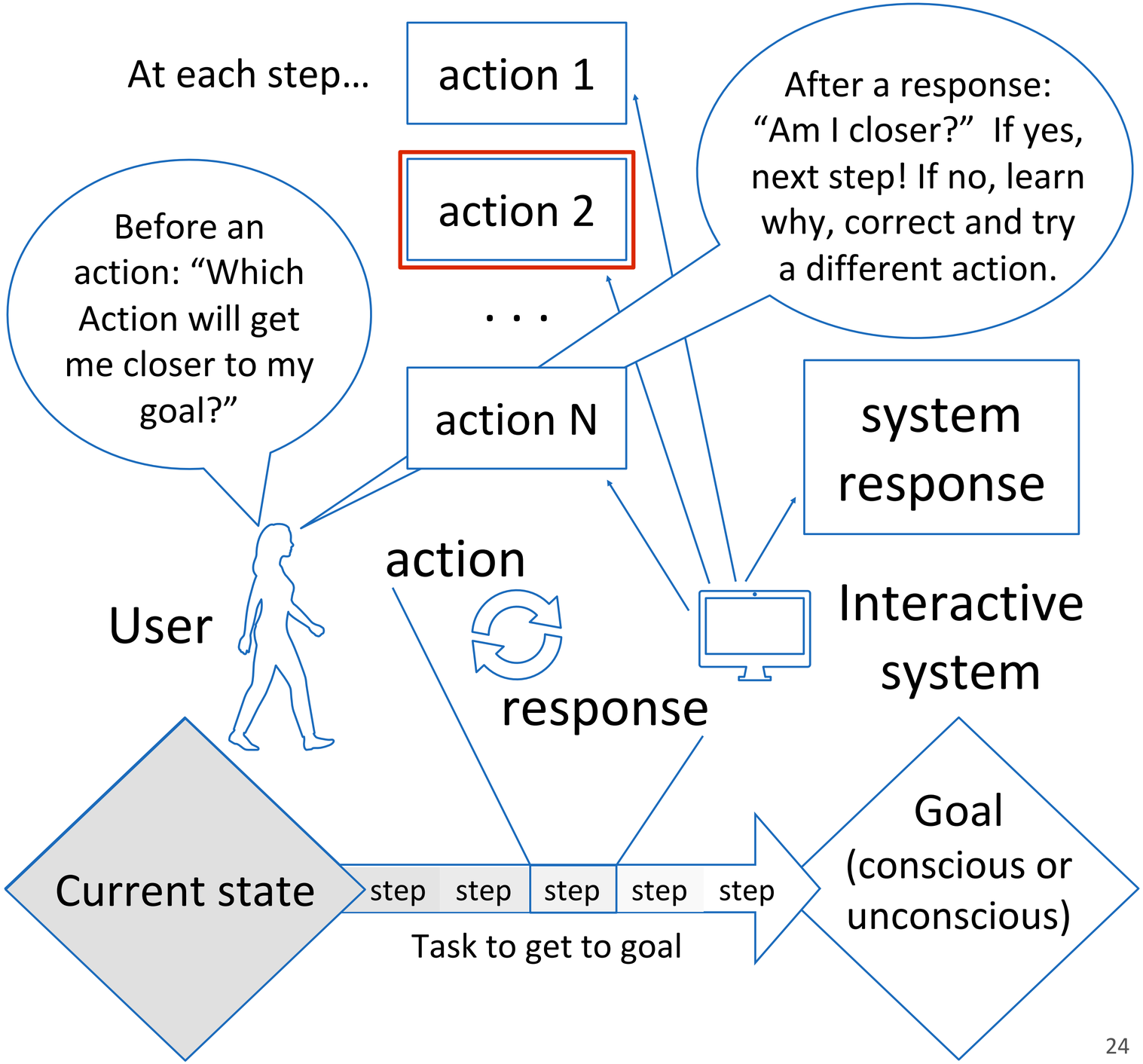
- Can be done quickly
- Finds many of the ease-of-learning problems that would be found testing with users
- Doesn’t have have multiple evaluators

Cons

- Isn’t enough alone - needs other methods to complement it

Theoretical Framework:
The CE+ Learning Model (Polson & Lewis, 1990)

- A user approaches all systems with a **goal**, which can be conscious or unconscious
- To get from the current state to the goal, the user takes a series of **steps**, which we call a "task"
- Each step in a task is composed of an **action** the user takes, followed by a **response** that the system has to that action
- When deciding what action to take, users examine the actions available, and selects one they believe will take them **closer to their goal**
- The system's response should clearly indicate **what happened** and the **new current state**
- If the user chooses the **correct action**, it should easy to **confirm they are closer** to their goal
- If the user makes a **mistake**, it should be **easy to realize**, **clear as to why**, and **easily correctable**

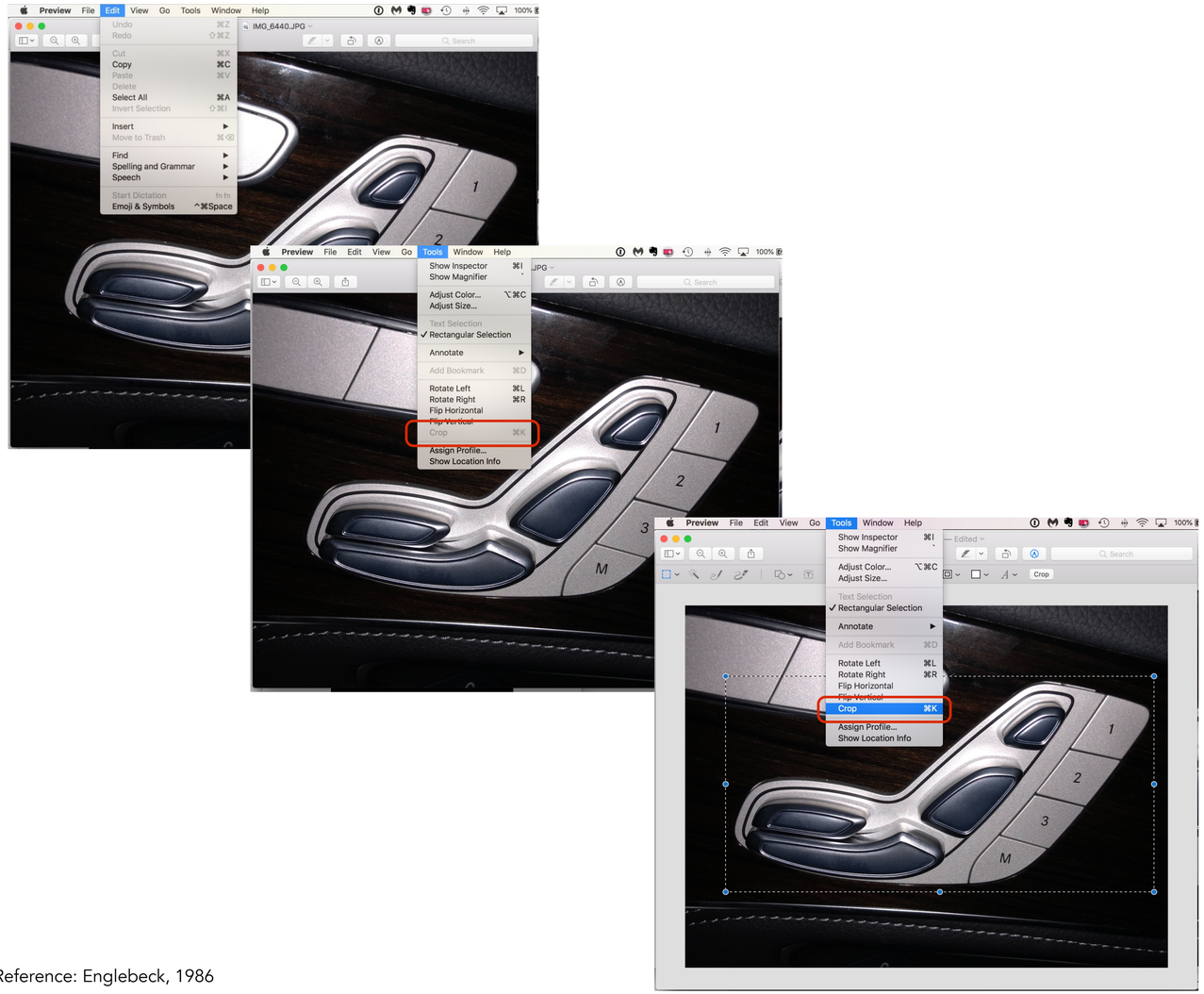


* From the CE+ Learning Model (Polson & Lewis, 1990)

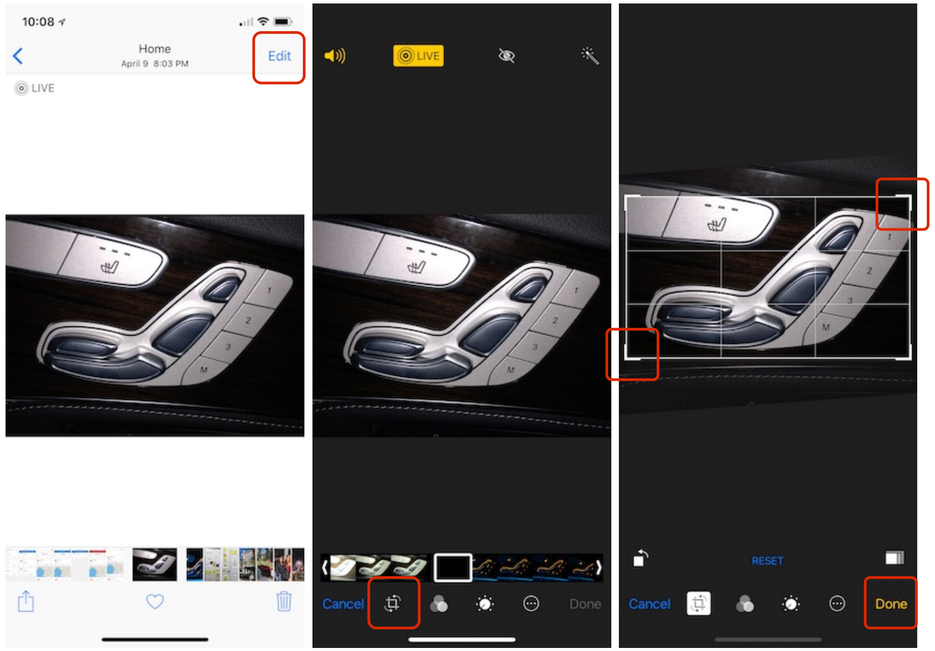
Principle: Label-Following

When users are not sure how to use a system for a specific task, they explore their options, looking for matches between the task or goal description (either given to them or in their head) and the labels of the actions available.

Example: Using the Preview Mac OS app to "crop" a photo.



Example: Cropping in iOS's Photos app



Reference: Englebeck, 1986

Polson & Lewis' Design Principles for Learning

Figure 7. **Design Principles for Successful Guessing**

1. Make the repertoire of available actions salient.
2. Use identity cues between actions and user goals as much as possible.
3. Use identity cues between system responses and user goals as much as possible.
4. Provide an obvious way to undo actions.
5. Make available actions easy to discriminate.
6. Offer few alternatives.
7. Tolerate at most one hard to understand action in a repertoire.
8. Require as few choices as possible.

PREPARING FOR A COGNITIVE WALKTHROUGH

Preparation Phase - Obtain and print out the following:

- **Product:** A description of the product/service
- **Users:** Relevant information about the target user, their goals and the context (i.e., typical conditions they would encounter the product with - for example, if it is a database they are searching, do we expect there to be many records to be realistic?)
- **Tasks:** A description of the task scenarios you will walk through. Choose tasks that are: first-time tasks, problematic, used by many users, and/or important to meet a user need.
- **Screens and Actions:** For each task scenario, specify the steps needed to succeed and the screens the user would see, including:
  - All the actions they could take
  - An indication of the correct action at each step
- **Roles:** Determine the evaluator team and assign the following roles:
  - UX specialist: guides the method; provides additional knowledge needed during walkthrough about user or market
  - Designer: presents each screen and actions
  - Peers (developers, product management): record or scribe (only one scribe needed)

Exercise documents: http://bit.ly/nng-ux-assessments

PERFORMING A COGNITIVE WALKTHROUGH

Analysis Phase - Gather evaluators together and then:

- **Walk through** each step of the **task**, reviewing the screen for that step and all the actions offered. **Ask and answer the Cognitive Walkthrough questions**:
  - Just before taking an action (e.g., "Does the user know what action to take to make progress toward the goal?")
  - Right after the system responds to an action (e.g., "Did the user understand if that step succeeded?")
- During the walkthrough, **capture or document** the following:
  - **Issues** identified from problematic answers to the Cognitive Walkthrough questions. Be sure to include a good visual or description of the problem area. Issues can be:
    - Any technical issues observed which might affect the experience reviewed
    - Any design issues observed which might affect the experience reviewed
  - New **assumptions or questions** about the user base that came up during the walkthrough which may require confirmation or investigation

The 4 Cognitive Walkthrough Questions*

Before an Action (for a step) is taken

1. "Will the users try to achieve the right effect?" (i.e., pick the correct next step)
2. "Will the user notice that the correct Action is available?" (related to Discoverability)
3. "Will the user associate the correct action with the effect trying to be achieved?"

Or better yet: "1. To make progress toward their goal/task, will the user know what Action to take next?"

After an Action (for a step) is taken

4. "If the correct action is performed, will the user see that progress is being made toward the solution of the task?" (related to Feedback, Visibility)

And then: "2. Based on the system response, will the user see that progress is being made toward their goal/task completion?"

*From Wharton, et al (1994).

Exercise documents: http://bit.ly/nng-ux-assessments

LET'S DO A SIMPLE COGNITIVE WALKTHROUGH

- **Product:** Poll Everywhere - an lightweight polling tool for obtaining real-time audience feedback via mobile, web or text messaging
- **Users:** Instructors or teachers who want real-time feedback during presentations; they are proficient in the use of computers, the web and smartphones. They have seen a poll used in a class before.
- **Tasks:** Users want **to set up a poll/survey easily and see how it responds to test data.**
- **Screens and Actions:** See the following video and screens, which should show you:
  - The actions the user could take
  - Some indication of the correct action at each step

To simplify, let's reduce the number of **Cognitive Walkthrough Questions** to just **2,** using a streamlined version of this method*:

Before an Action (for a step) is taken, ask:

1. **"To make progress toward their goal/task, will the user know what Action to take next?"**

After an Action (for a step) is taken, ask:

2. **"If the user does the right thing, will they see that progress is being made toward their goal (based on system response)?"**

**Adapted based on Rick Spencer. 2000. The streamlined cognitive walkthrough method, working around social constraints encountered in a software development company. In Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '00). ACM, New York, NY, USA, 353-359.*

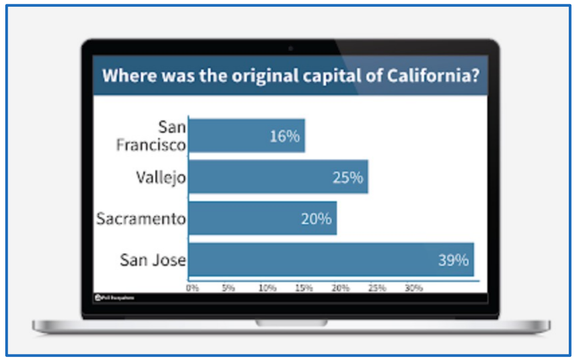
Exercise documents: http://bit.ly/nng-ux-assessments

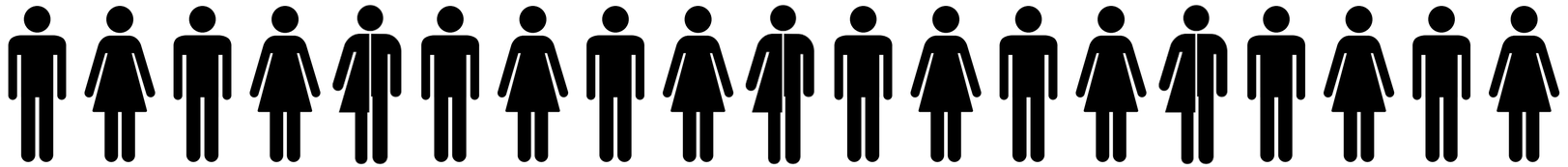
Poll Everywhere

User Goals/Task Scenarios:

- Set up a poll/survey easily and see how it responds to test data

Take a look at this video (< 5 mins), then we will look at it screen by screen for the Cognitive Walkthrough.





Safari File Edit View History Bookmarks Develop Window Help 100% Mon 3:39 PM Christian Rohrer

polleverywhere.com

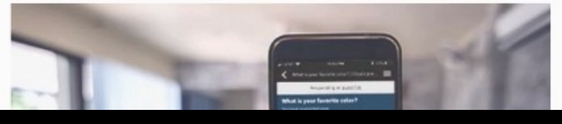
Poll Everywhere

How it works Pricing Support Enterprise Log in Sign up

Live interactive audience participation

Engage your audience or class in real time

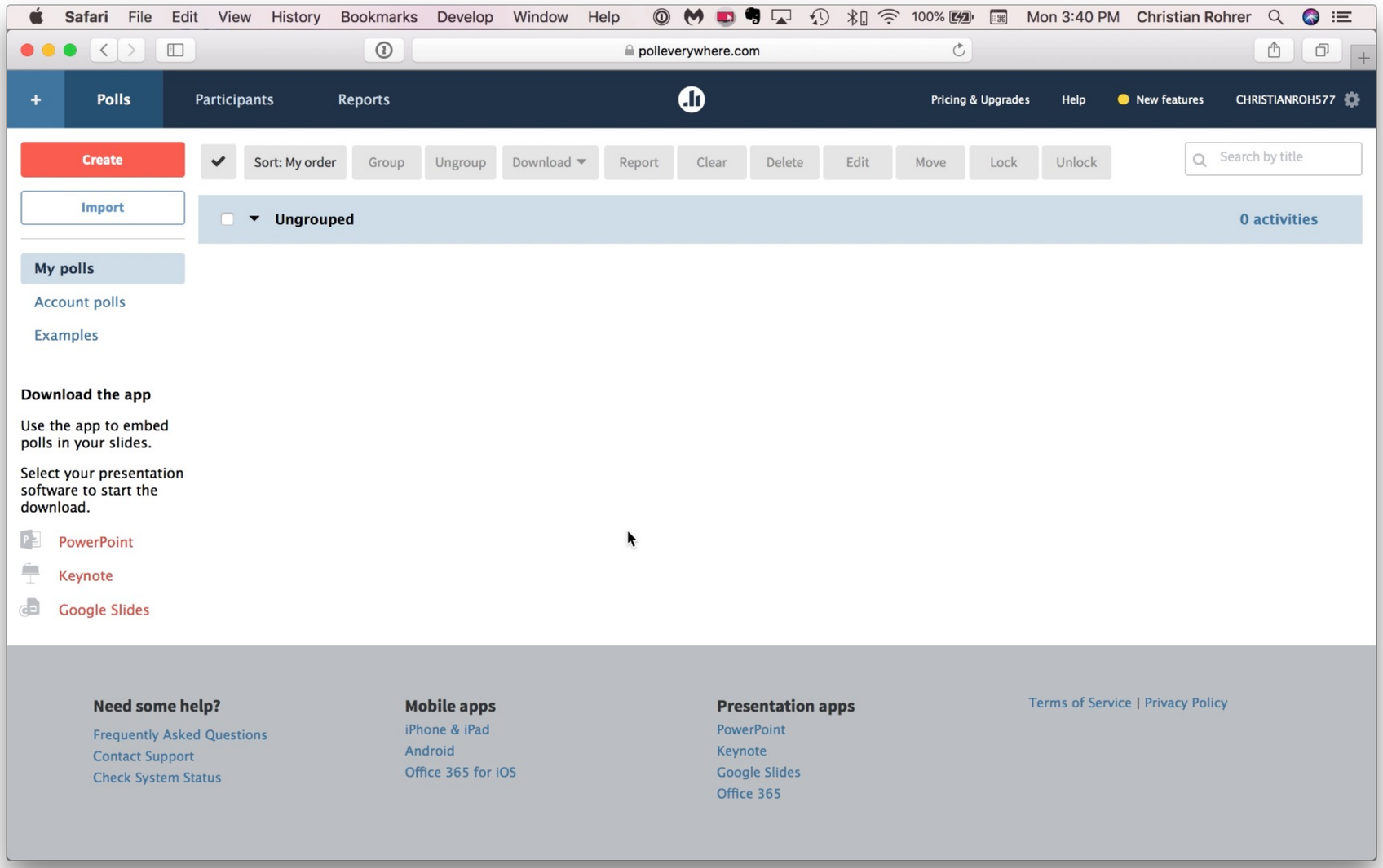
Get started



https://www.polleverywhere.com/login

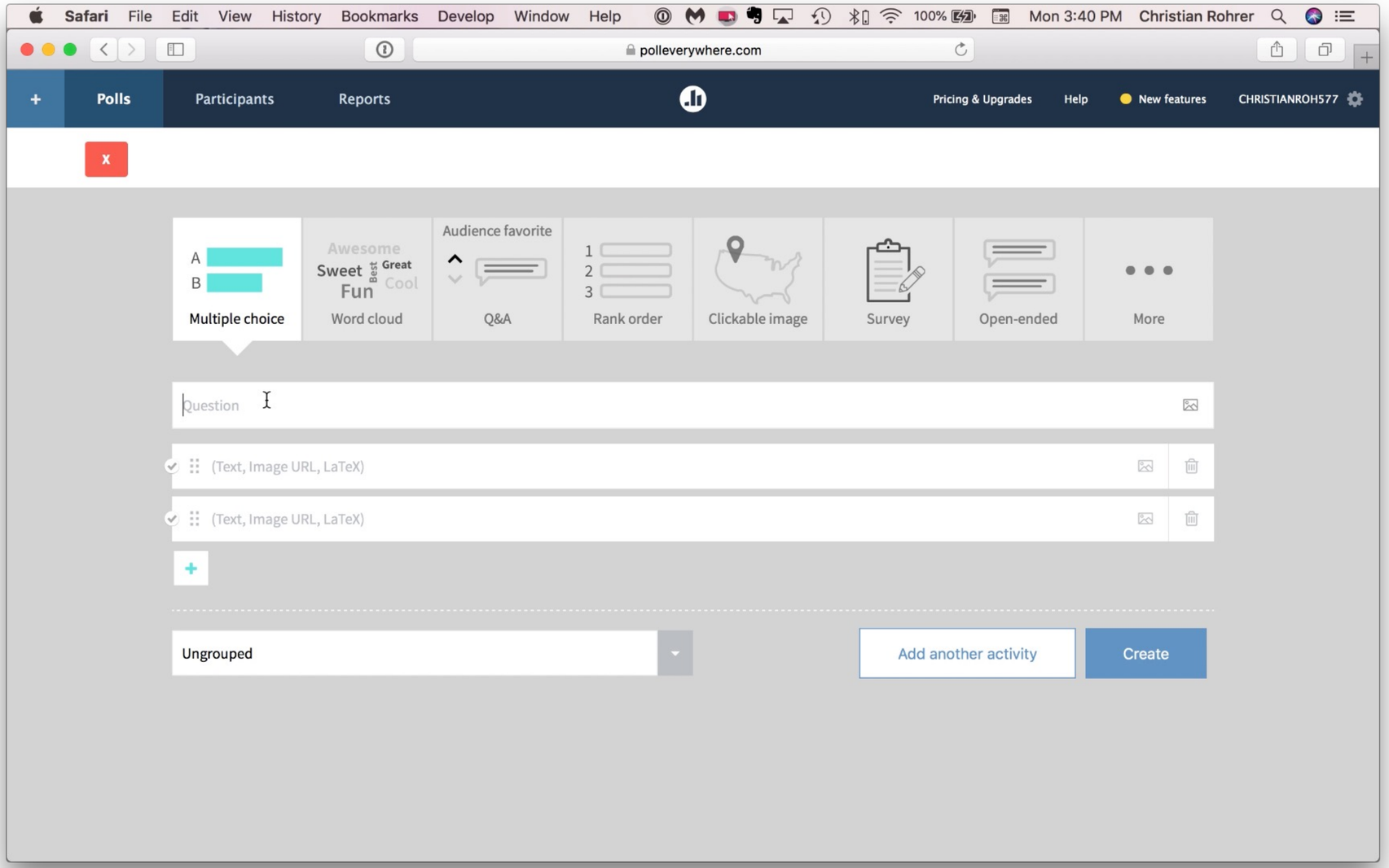
1. “To make progress toward their goal/task, will the user know what Action to take next?"

2. “If the user does the right thing, will they see that progress is being made toward their goal (based on system response)?



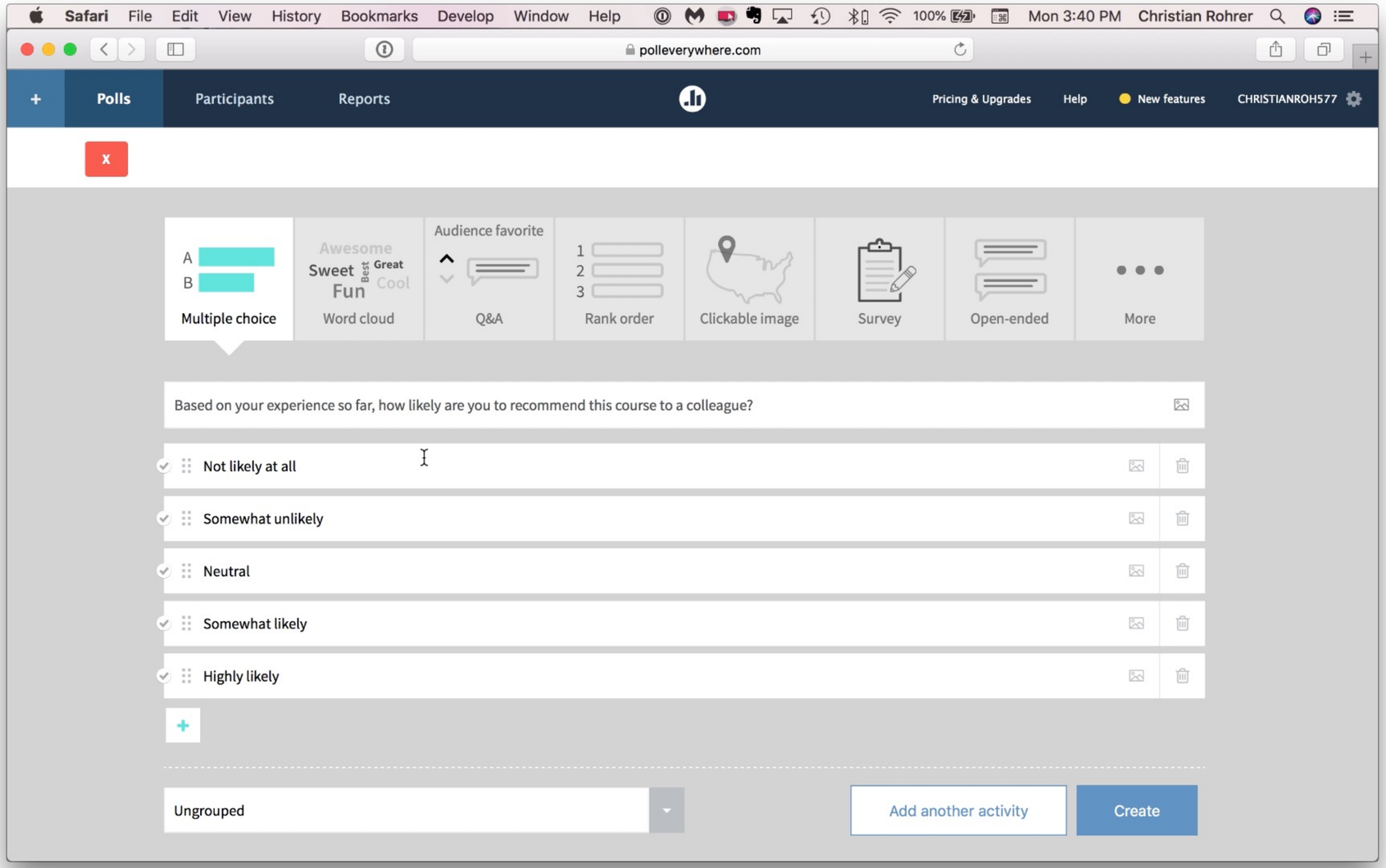
1. “To make progress toward their goal/task, will the user know what Action to take next?"

2. “If the user does the right thing, will they see that progress is being made toward their goal (based on system response)?



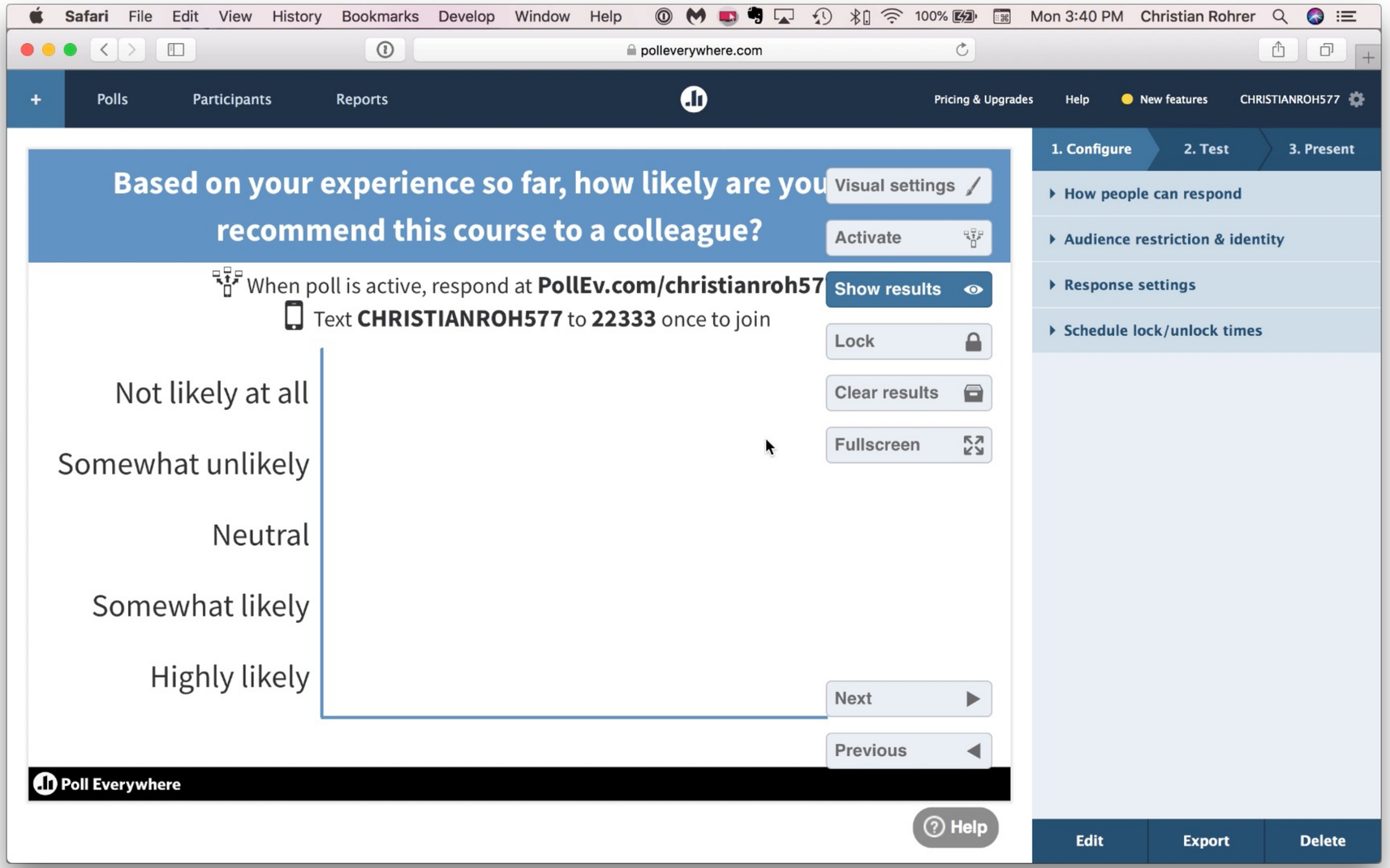
1. “To make progress toward their goal/task, will the user know what Action to take next?"

2. “If the user does the right thing, will they see that progress is being made toward their goal (based on system response)?



1. “To make progress toward their goal/task, will the user know what Action to take next?"

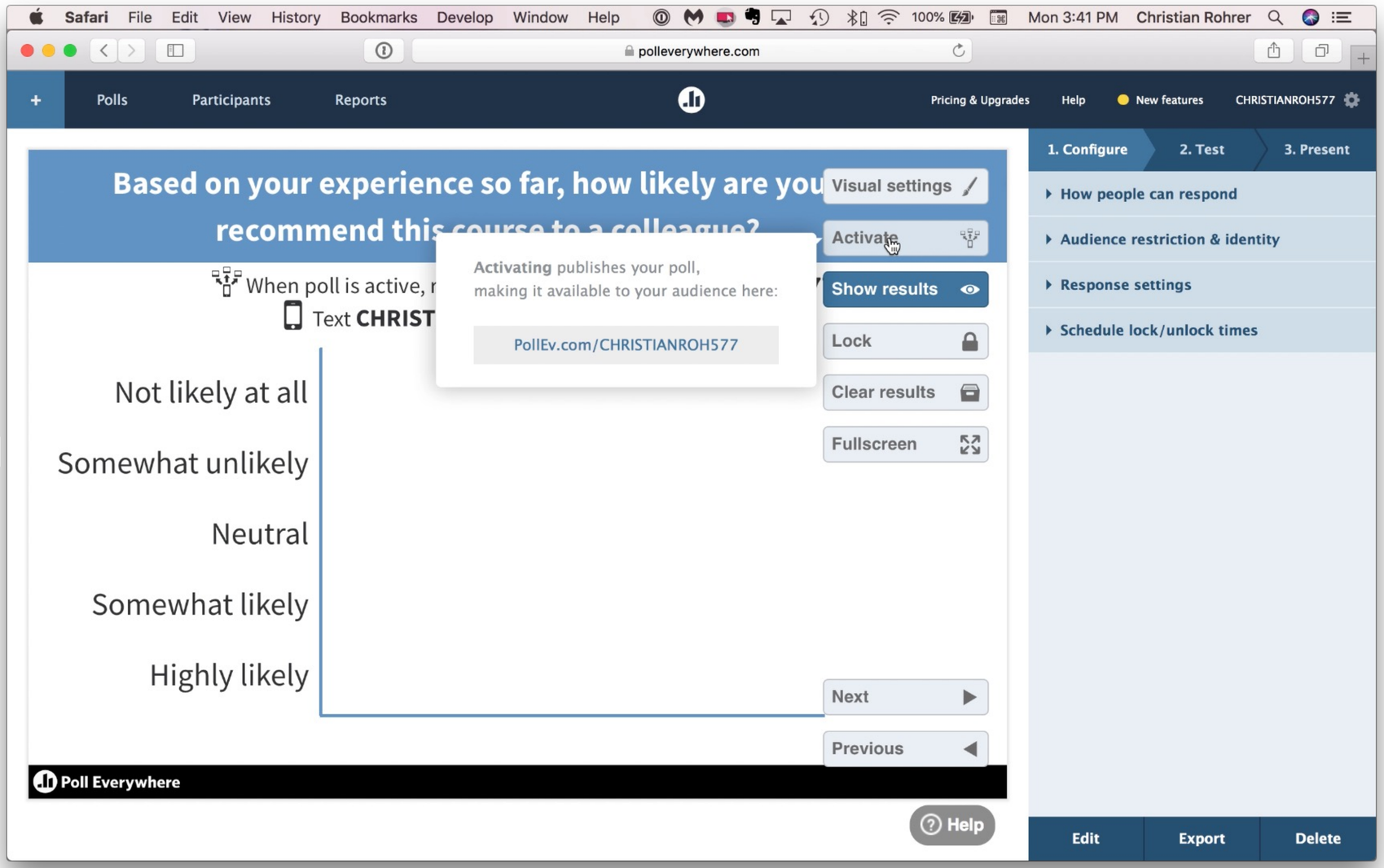
2. “If the user does the right thing, will they see that progress is being made toward their goal (based on system response)?



1. “To make progress toward their goal/task, will the user know what Action to take next?"

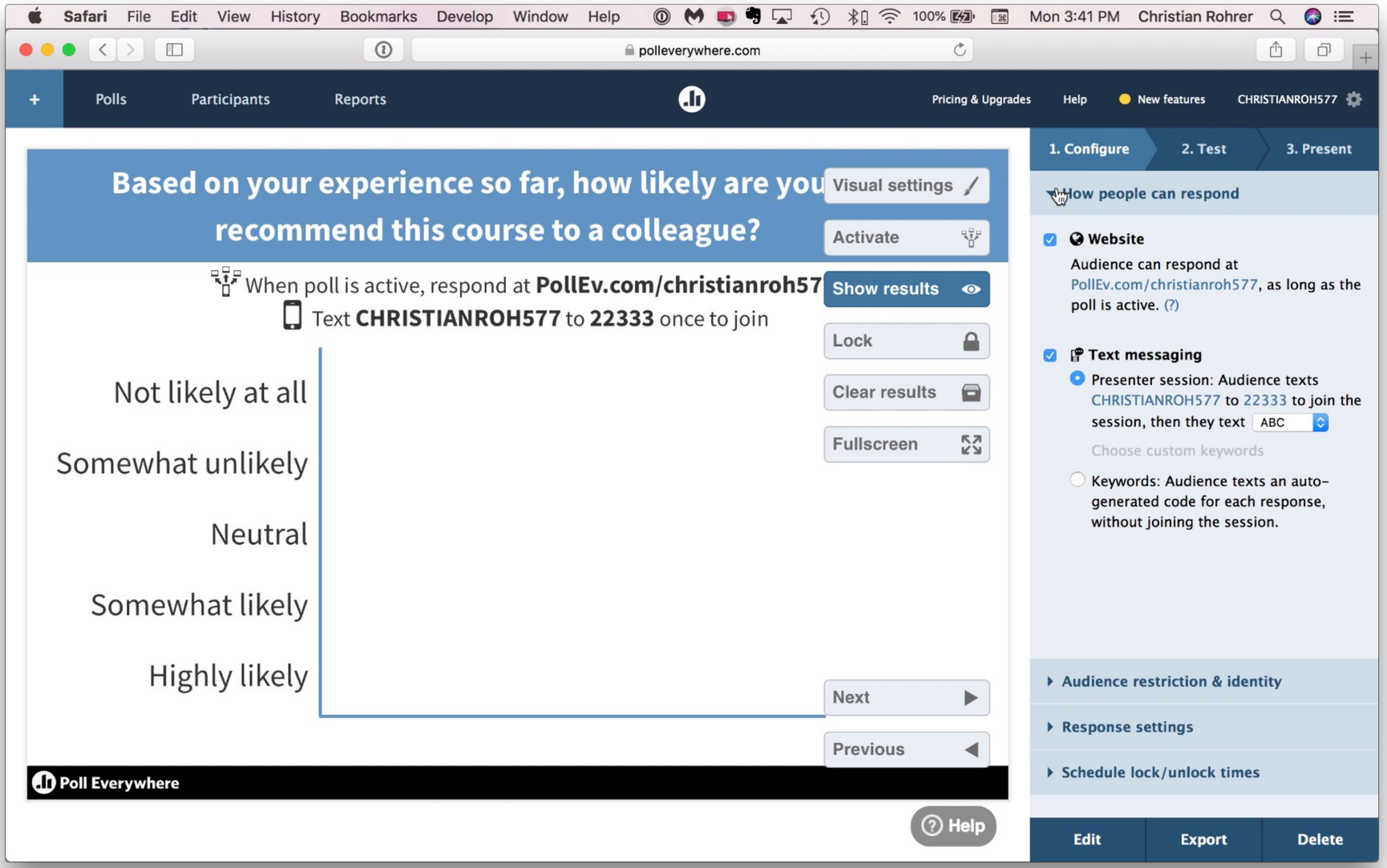
2. “If the user does the right thing, will they see that progress is being made toward their goal (based on system response)?

Hovering over “Activate”



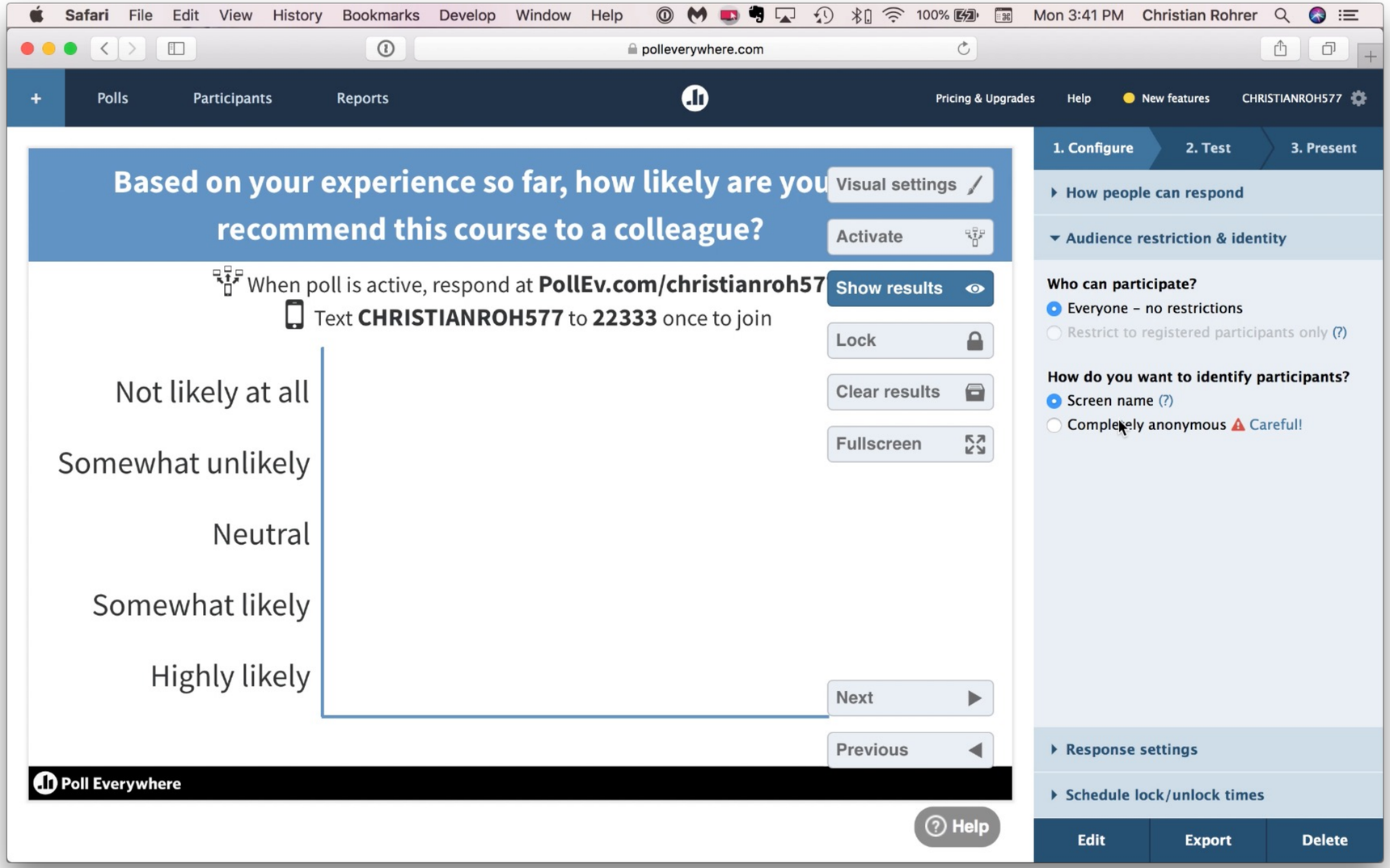
1. “To make progress toward their goal/task, will the user know what Action to take next?"

2. “If the user does the right thing, will they see that progress is being made toward their goal (based on system response)?



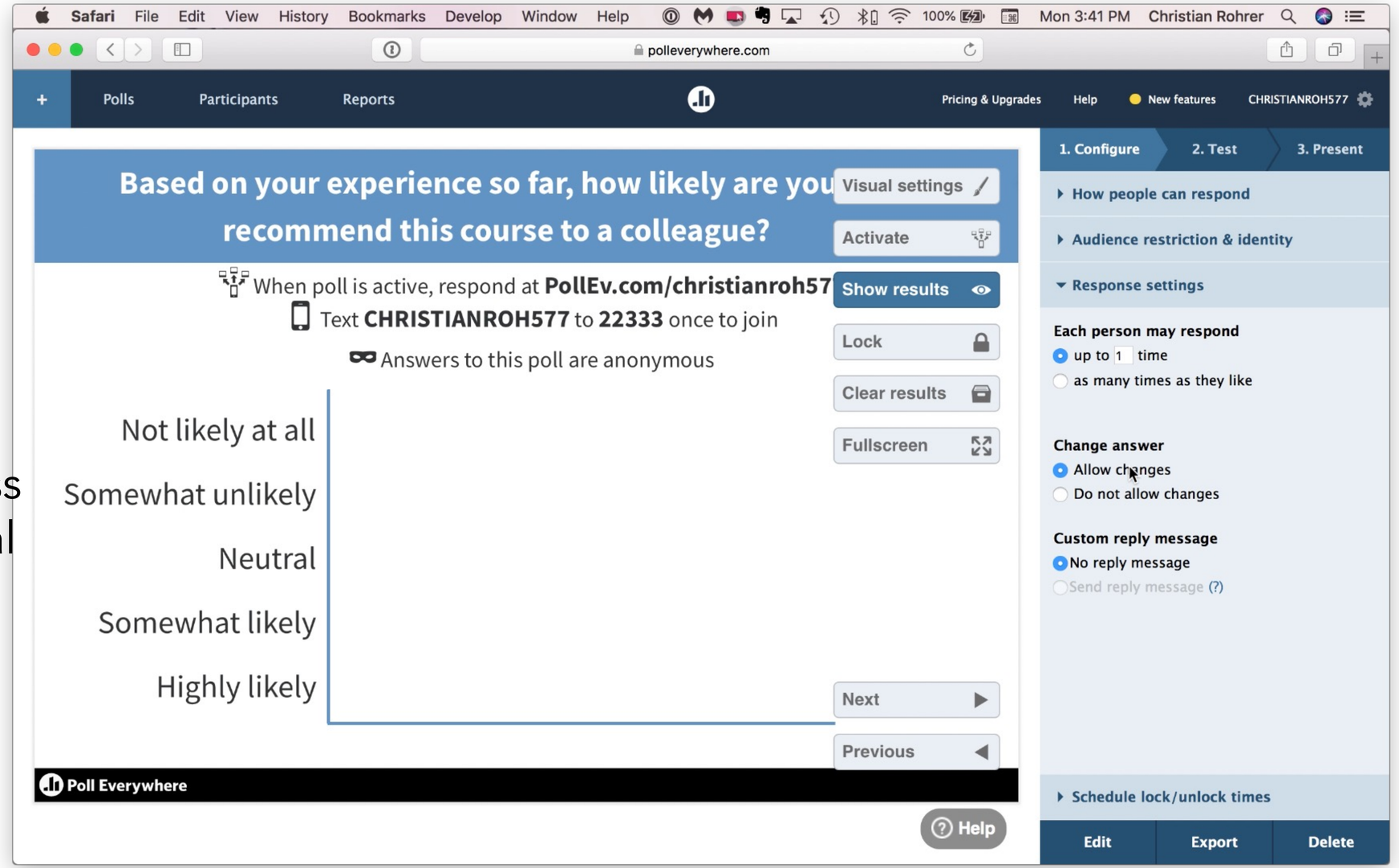
1. “To make progress toward their goal/task, will the user know what Action to take next?"

2. “If the user does the right thing, will they see that progress is being made toward their goal (based on system response)?



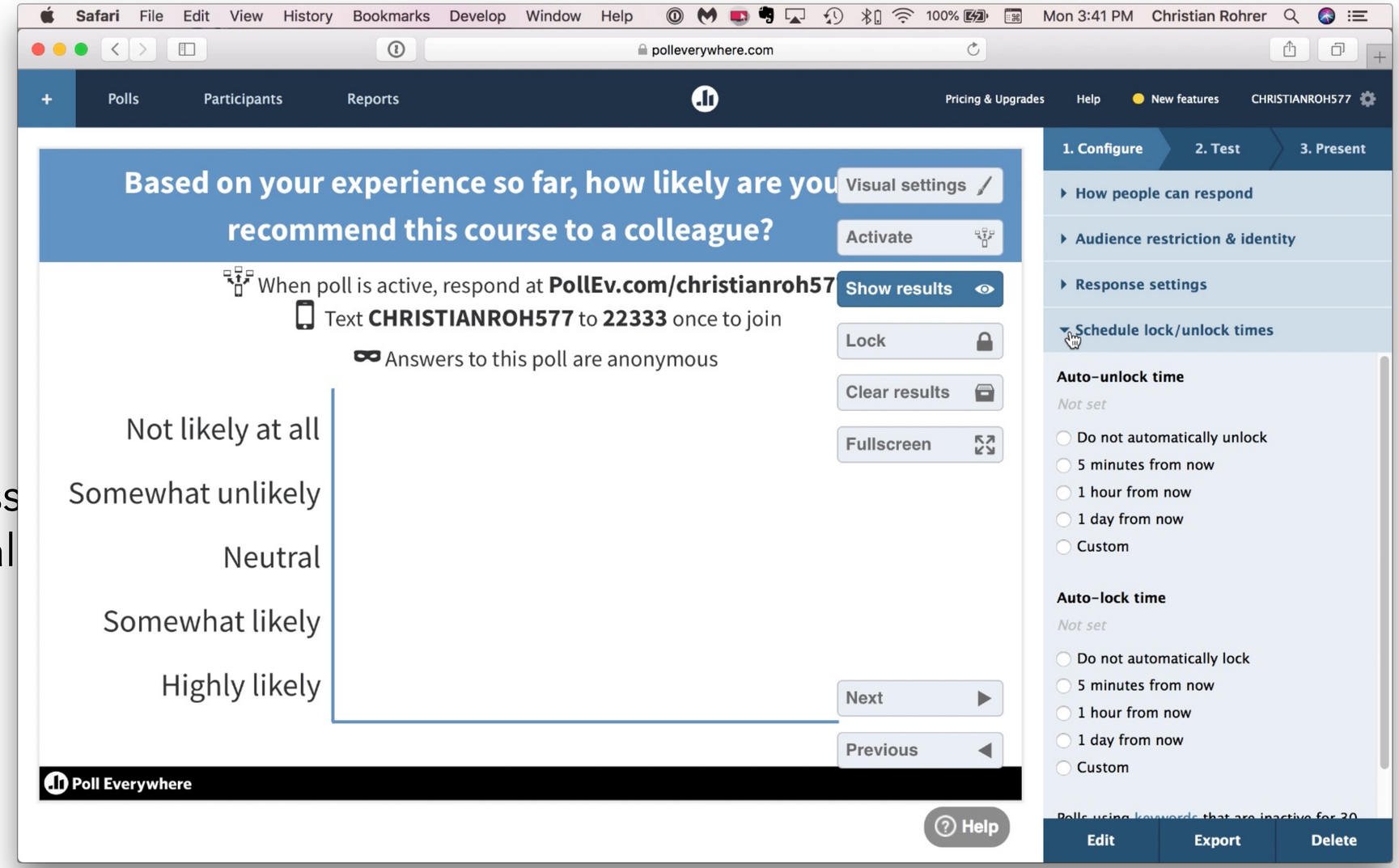
1. “To make progress toward their goal/task, will the user know what Action to take next?"

2. “If the user does the right thing, will they see that progress is being made toward their goal (based on system response)?



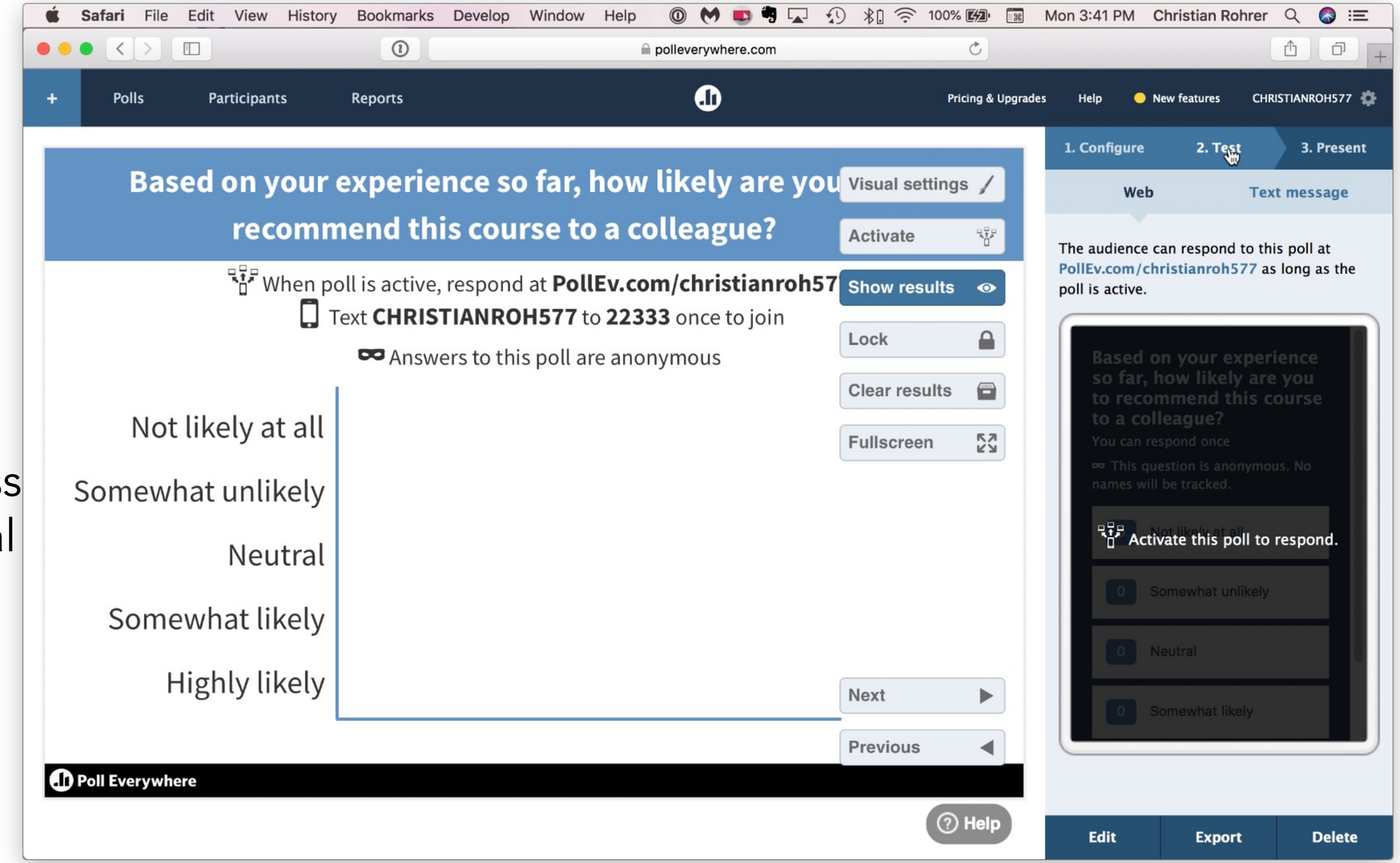
1. “To make progress toward their goal/task, will the user know what Action to take next?"

2. “If the user does the right thing, will they see that progress is being made toward their goal (based on system response)?



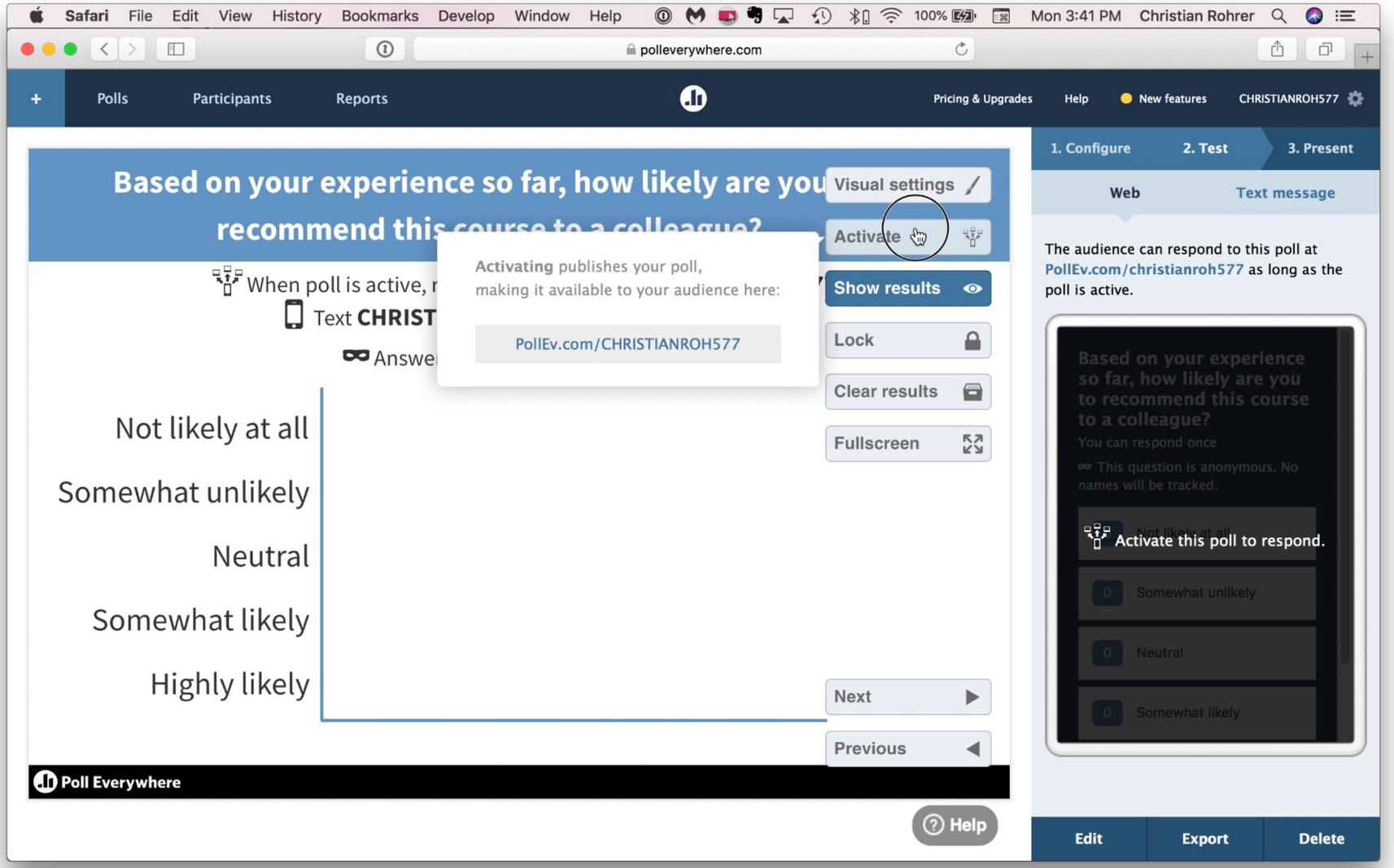
1. “To make progress toward their goal/task, will the user know what Action to take next?"

2. “If the user does the right thing, will they see that progress is being made toward their goal (based on system response)?



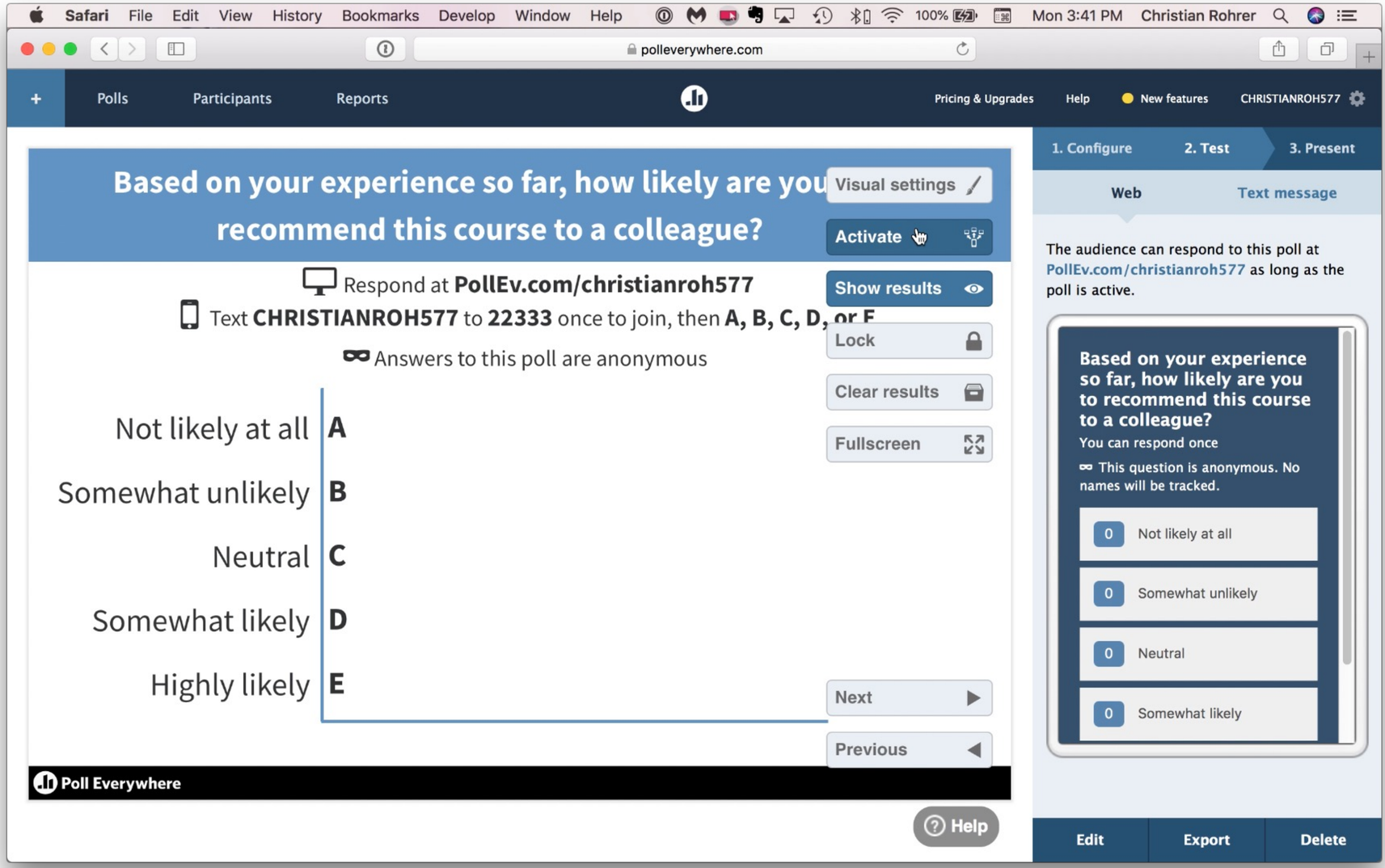
1. “To make progress toward their goal/task, will the user know what Action to take next?"

2. “If the user does the right thing, will they see that progress is being made toward their goal (based on system response)?

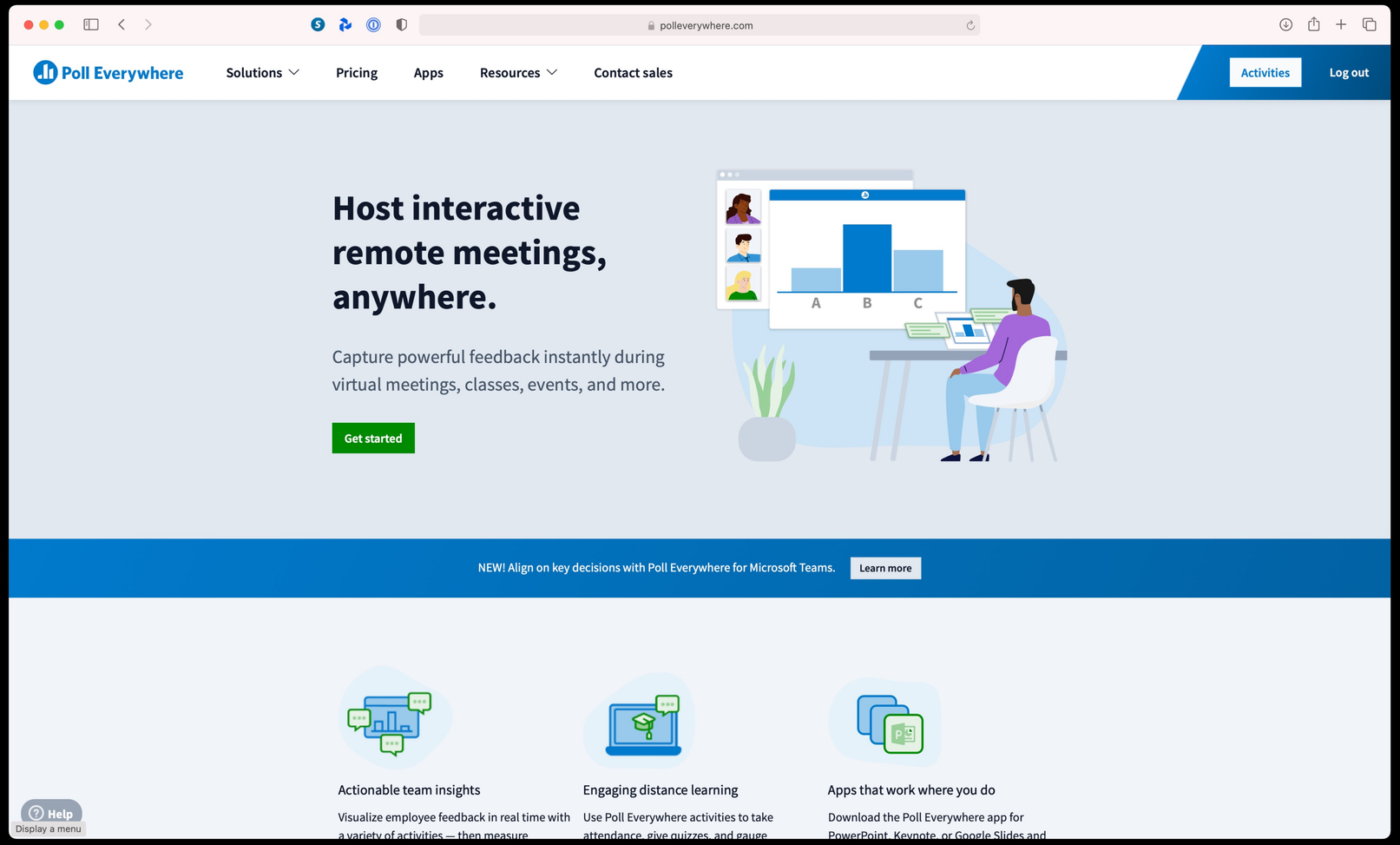


1. “To make progress toward their goal/task, will the user know what Action to take next?"

2. “If the user does the right thing, will they see that progress is being made toward their goal (based on system response)?



HEY! POLLEVERYWHERE WAS REDESIGNED! LET'S SEE IF IT'S BETTER IN TERMS OF LEARNABILITY...



Poll Everywhere

Solutions Pricing Apps Resources Contact sales

Sign up Log in

Host interactive remote meetings, anywhere.

Capture powerful feedback instantly during virtual meetings, classes, events, and more.

Get started

NEW! Align on key decisions with Poll Everywhere for Microsoft Teams. Learn more

COGNITIVE WALKTHROUGH EXERCISE DISCUSSION

- What were the most difficult parts of the product to learn to use? Why?
- What knowledge does the user base need to have in order to be successful at learning how to use this system?
- What Actions could be improved?
- What Feedback could be improved?
- What was done well?

OUR TOP THREE RECOMMENDATIONS FOR THESE TASKS

- Recommendation 1:
- Recommendation 2:
- Recommendation 3:

Exercise documents: http://bit.ly/nng-ux-assessments

QUESTIONS?

- You may now ask questions from this section
- You may type your question into Zoom Chat
- We may have to cover some questions as follow-ups after the seminar

Pluralistic Walkthrough

WHAT IS A PLURALISTIC WALKTHROUGH?

- A means of stepping through an experience with a **cross-functional team (product management, development, design) and a target user** to get a user-centered perspective on the experience, fostering discussion on potential improvements
- Similar to Cognitive Walkthroughs, but includes 3-4 roles, including end-users
- **When to do a Pluralistic Walkthough**: When you have access to an end-user type who is capable of talking about their experience to designers and developers **and** when said designers and developers are interested in looking at the experience together from an end-user perspective

Pros

- Cheap and quick (if end-user is easily available)
- Can be done on early-stage forms of the product (wireframes or mockups)
- Can quickly establish user-centric point of view for many cross-functional team members, resulting in buy-in and willingness to change
- Can get results that might not come from testing, due to discussion (e.g., "I got it right, but I was unsure about…")

Cons

- End-users are not usually so readily available, and many cannot give legitimate feedback to those who created an experience
- Screens reviewed slowly are not as useful or realistic as seeing the live experience directly, which impacts validity of findings
- Does not explore all parts of the experience, multiple correct paths, or go deep into problem areas

HOW TO DO AN PLURALISTIC WALKTHROUGH

Preparation:

- Find 3-5 reviewers for the group, including: an end user, developer, designer, and a user researcher/usability expert
- Create printouts of the screens to be reviewed or put screenshots into a spreadsheet
- Describe the tasks briefly
- Write down any important context or assumptions

Review:

- One person (the designer or developer) acts as the walkthrough administrator; they provide any info or assumptions about the target users
- Each task description is read and the group walks through the task, screen by screen. (No flipping ahead)
- Each reviewer takes on the view of an end user and writes down what they believe the next action would be and why, noting any areas of confusion
- If the end-user (or any reviewer) needs any info that would be normally available, the designers or developers can answer. The user researcher moderates and prevents "explaining away" too much
- The walkthrough administrator reveals what the "right" step is, and the group discusses their choices, with the end-user going first and others after
- The user researcher encourages candor and a focus on improvement of the UX
- Design ideas can be offered up

Report: Results from all expert evaluators can be combined into a single report, but much of the benefit of a Pluralistic Walkthrough occurs during the Review process

WHAT DO YOU THINK?

- Would this approach work for you in your organization?
- What would be needed for this to work for you?
- How would you modify this, if at all?

QUESTIONS?

- You may now ask questions from this section
- You may type your question into Zoom Chat
- We may have to cover some questions as follow-ups after the seminar

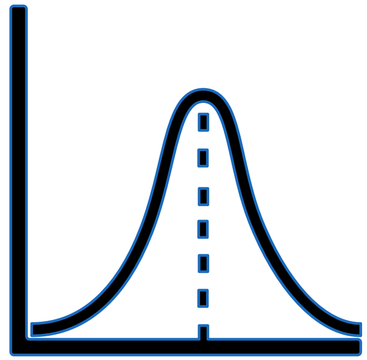
Metrics from UX Assessments

UX ASSESSMENTS CAN PRODUCE METRICS

	UX Teardown	Cognitive Walkthrough ✓	Pluralistic Walkthrough ✓	KLM-GOMS	Heuristic Evaluation	The PURE Method ✓
Use this method when your main focus is to…	Provide an outside critique of an experience, often by reverse engineering a product's intent and approach	Assess learnability and ease-of-use for new users	Have a cross-functional team that wants to see the experience from the point of view of the end-user	Assess the estimated time (and difficulty) of tasks at a very low-level (motor skill and cognitive units)	Find as many potential usability problems across the experience/UI, based on heuristics (principles)	Score and diagnose the friction in a product for its most important customers across its Fundamental Tasks
Perspective	External critic	Internal	Internal	Internal or external	Internal or external	Internal or external
Evaluator Characteristics	UX professional, often with a specific area of expertise	Anyone who can learn the basics of user learning theory	Any role (designer, developer, UX professional)	Anyone with good attention to detail	UX expert well versed in heuristics and design principles	UX expert well versed in heuristics and design principles
Learning Curve	Moderate	Low-Moderate	Low-Moderate	Moderate	Moderate-High	Moderate-High
Scripted	No - experience depends on where the reviewer happens to go	Yes - only correct path is followed	Yes - only correct path is followed	Yes - can be applied to any usage sequence	No - reviewer passes through experience as they wish	Yes - start with the Happy Paths for the Fundamental Tasks; can expand later
Output	Sequential screens with callouts and answers to key questions about users and intent	List of potential learnability issues, along with rationale	List of potential user experience issues	Sequence of low-level actions; numeric score estimating time required for task	List of potential issues with associated heuristic violations; severity ratings	Scorecard on friction (ease-of-use) for Tasks and Entire product; also, details on score rationales
Metrics	None	None	None	Estimated task time	Severity ratings of issues found	Friction/Ease of Use Scorecard

HOW DO THESE COMPARE WITH EMPIRICAL USER RESEARCH METRICS?

TRADITIONAL MEASURES OF USER EXPERIENCE & USABILITY



Empirical: Data from users and studies

USABILITY BENCHMARKING

CARD SORTING

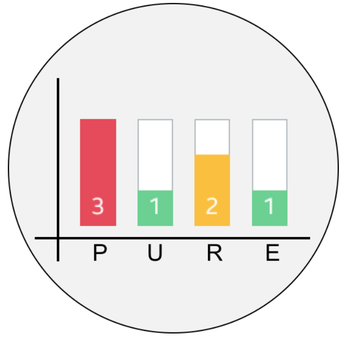
TREE TESTING

FIRST CLICK TESTING

Analytical: Data from experts

HEURISTIC EVALUATIONS

KEYSTROKE LEVEL MODELING



PURE

Empirical Data: Task-based attitudes and behaviors

Attitudinal Measures

Study Level Sat.
Task Level Sat.

Behavioral Measures

Completion Rates
Errors

Time on task

Example Attitudinal Measures: SUS and SEQ

System Usability Scale (SUS) - asked after study

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

Single Ease Question (SEQ) - asked after each task

Overall, how difficult or easy did you find this task?

Very Difficult 1	2	3	4	5	6	Very Easy 7
○	○	○	○	○	○	○

Example Behavioral Measures

Time on Task

- How long to complete task
- Total time to complete all tasks

Error Rate

- What % of users in the task had an error
- Number of errors per task

Help Needed

- What % of users in the task required help
- How many times did users need help to succeed

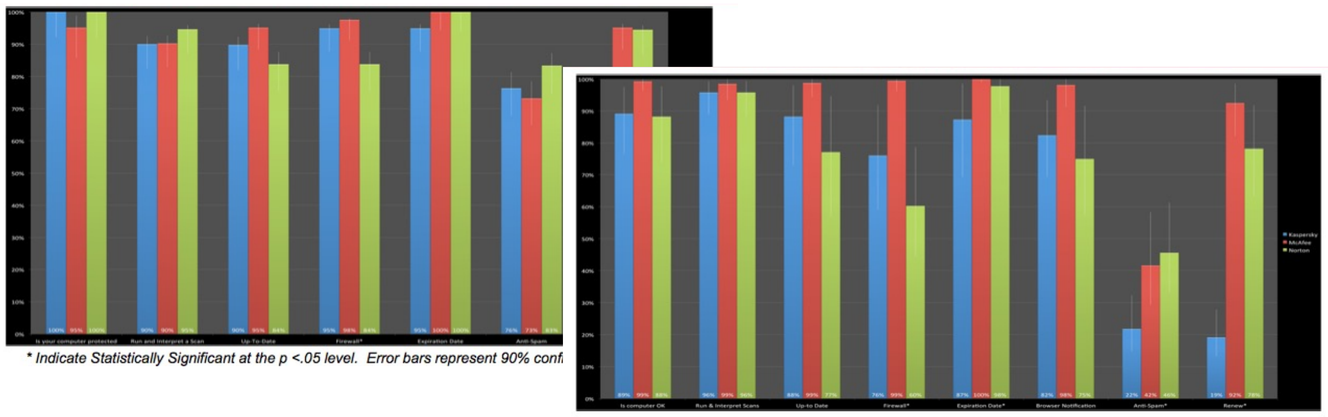
Success Rate

- What % of users succeeded in the task
- Numbers of tasks successfully accomplished

Quantitative Empirical Methods require time, money and many users

Usability Benchmark Example:

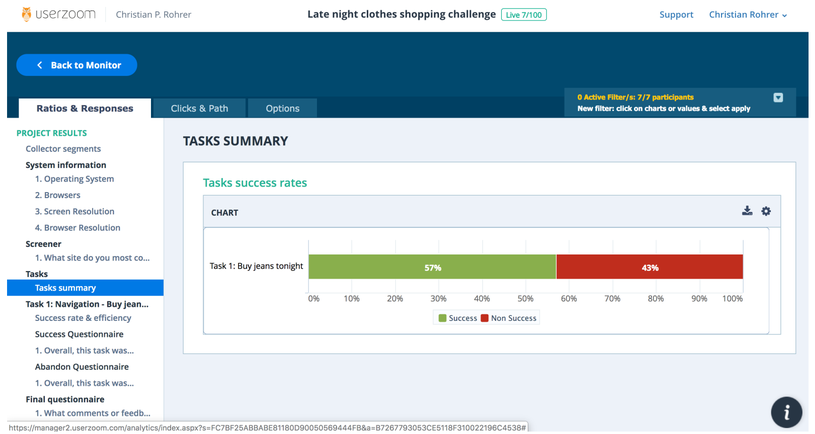
- Three product competitive benchmark (desktop SW)
- 7 tasks/scenarios
- 42 users per product
- 2+ months, $100K



** Indicate Statistically Significant at the p <.05 level. Error bars represent 90% confidence intervals*

Online Task-Based Testing Example (e.g., UserZoom):

- 1200 website users (3 sites)
- 8 tasks/scenarios
- 3 week setup and collection
- $90K license or $25-50K x 1



Metrics can also be produced by Analytical Methods (UX Assessments)

Analytical

Heuristic Evaluation

- Metrics are twofold:
  - Number of problems found by any rater
  - Severity rating (0-4) of problems

Keystroke Modeling (GOMS/KLM)

- Metrics are about time to complete tasks (based cognitive and motor)

PURE

- Scores of ease of use (friction or cognitive load) of key tasks and product

HEURISTIC EVALUATIONS

KEYSTROKE LEVEL MODELING

PURE

How the Analytical Methods produce their metrics differs

Heuristic Evaluation

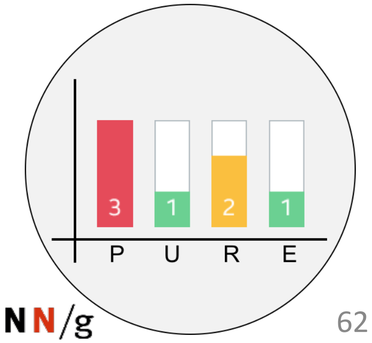
- 3-5 raters independently walk through an interface, to review whether good principles (heuristics) are present
- Goal is to find and document as many usability issues as possible and assign severity ratings of found problems

Keystroke Modeling (GOMS/KLM)

- 1 rater catalogs operations in an interface, using known cognitive/motor skill time limits to estimate efficiency

PURE

- 3 raters represent the perspective of a specific user type and reliably score core tasks of a product in terms of ease of use (aka friction or cognitive load)



Metrics from Heuristic Evaluation, KLM/GOMS, PURE

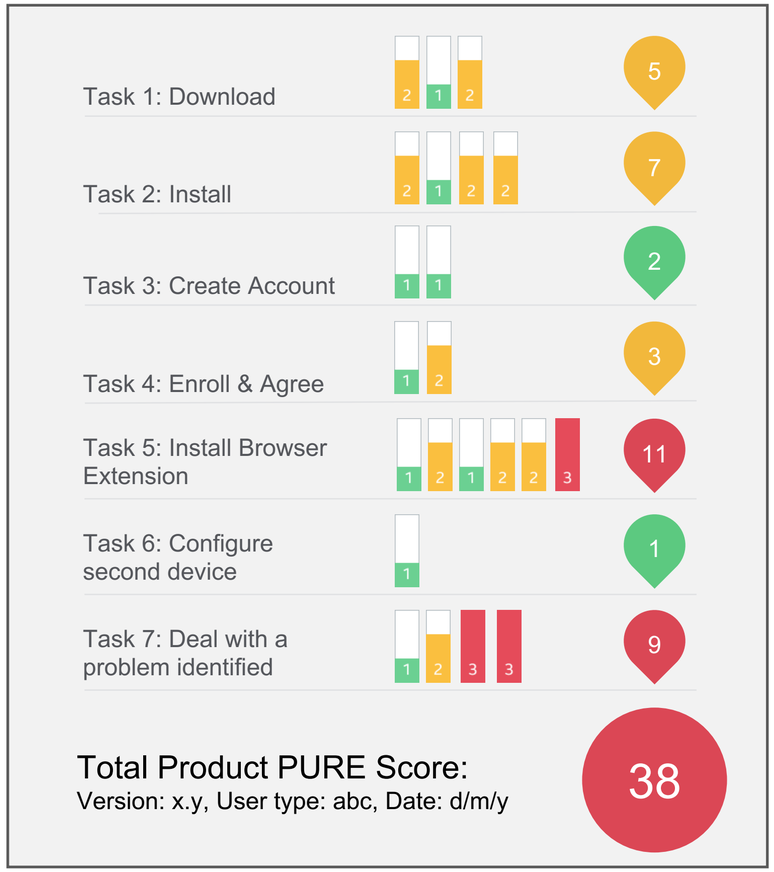
Heuristic Evaluation: A number of issues found, categorized by severity levels

	Heuristic Evaluator #1	Heuristi Evaluator #2	Heuristic Evaluator #3
Issues found - severity 1	16	13	10
Issues found - severity 2	4	3	6
Issues found - severity 3	3	2	2
Issues found - severity 4	2	1	3

KLM/GOMS: Estimated execution times a low level

Operator Description [Short name]	Time (ms)
Shift attention [Attend]	50
Initiate motor process [Initiate]	50
Swipe down from upper right corner [Swipe]	170
Wait for system to render Control Center UI [Response]	300
Find Wi-Fi icon [Look]	550
Initiate motor process [Initiate]	50
Tap Wi-Fi icon [Tap]	450
Verify that Wi-Fi is off [Look]	550
Initiate motor process [Initiate]	50
Swipe up from bottom to dismiss Control Center [Swipe]	170
Wait for system to close Control Center UI and return to Mail [Response]	300
TOTAL Time to Execute 2.69 secs	**2690**

PURE: An ease of use scorecard



ARE UX ASSESSMENTS RELIABLE AND VALID?

- This has been a long-standing topic in the field of Human-Computer Interaction
- **All** of the original **authors** of the various methods **claim some degree of validity**, meaning that their method finds true usability/UX issues that have been independently measured through some type of empirical study (e.g., usability study)
- However, this has **not been unequivocally proven** through rigorous experimental design and controlled study (Gray & Salzman, 1998)
- It has also been shown that **different evaluators find different issues** (Mahatody & Kolski, 2010)
- There's hope: **Some studies** for recent methods like PURE (Rohrer, et al., 2016) showed **reasonable convergent validity** (r=0.5) against empirical results (e.g., SEQ with 220 users across 3 products and 8 websites). PURE has also been shown to have **good** ($0.5 < r < 0.9$) inter-rater **reliability** after a handful of training sessions.

- Getting perspectives from **multiple evaluators** improves the likelihood of finding issues, if done carefully, and it can help evaluators sharpen their judgment
- When viewed as **a *supplementary* method**, rather than a complete replacement for empirical user research, UX Assessments find their purpose in a few key ways:

1. Issues identified against known principles are strong indicators of **potential problems**; this can be a good start for where to focus using another method to verify
2. The cost of a **false positive** for a given "issue" is **low** if:
    - A clearly superior solution is obvious
    - The cost of implementation of the solution is low
3. Simply having **buy-in** to address UX quality when with quick, low-cost UX Assessment methods is often valuable
    - If validity is questioned, it opens the door to budget for and conduct higher-quality empirical user research.
    - Either way, a **focus on UX quality** is a win

QUESTIONS?

- You may now ask questions from this section
- You may type your question into Zoom Chat
- We may have to cover some questions as follow-ups after the seminar

6

GOMS

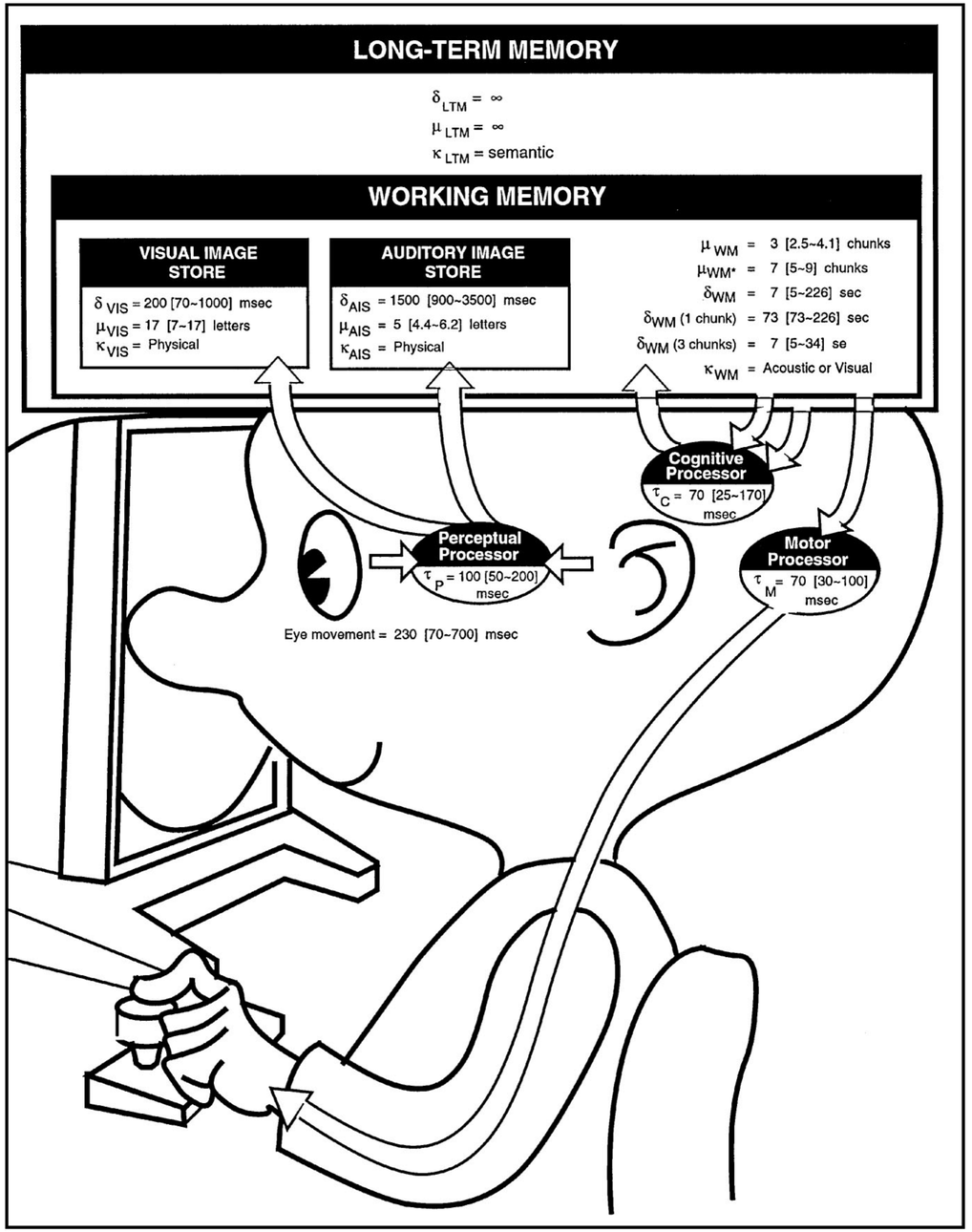
WHAT IS GOMS? GOALS, OPERATORS, METHODS & SELECTION RULES

- GOMS is a method of analysis that allows us to predict the performance of a skilled person using a specific system
- The most common form is KLM-GOMS, which stands for Keystroke Level Model, which provides specific values for many user actions (keys typed, buttons selected, mental choice, etc.)
- It is based on a model of cognition called Human Information Processing Theory
- It also relies on findings from research on motor processing and performance, such as Fitts's Law.

- **Goals**: what a person wants to accomplish, either at a high-level or low level. High level goals are met by several low-level goals (e.g., writing a book comes from typing/deleting characters).
- **Operators**: perceptual, cognitive or motor actions used to accomplish goals, such as pressing a key or tapping an icon on the screen or trackpad.
- **Methods**: procedures or algorithms of operators that accomplish a goal.
- **Selection Rules**: the decisions a given person makes to choose the method in a given context.

THEORETICAL FRAMEWORK: HUMAN INFORMATION PROCESSOR (PROCESSING) MODEL (1983)

- This model characterizes thinking, perceiving and acting much like how a computer processes information, including our specifications (how fast we can perform certain operations)
- In the 70s and early 80s, this was the popular view of human cognition*, especially on how memory, perception and attention work
- Actions like the perception and storage of a visual image, or movement of a cursor were empirically measured, derived and specified
- This view of human-computer interaction largely focused on how an individual human could effectively interact with a computer or other technology, based on our capabilities and whether the design of the system was well-matched for the user



Source: Card, Moran, Newell (1983): The Psychology of Human Computer Interaction

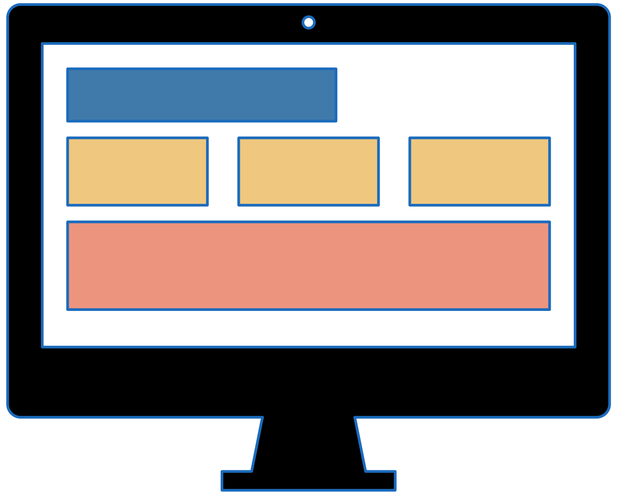
*For example, Newell, A & Simon, H. (1972): Human Problem Solving

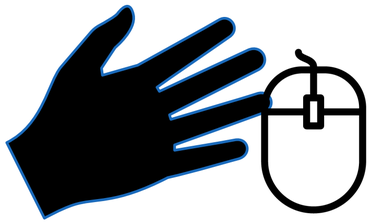
PUT MORE SIMPLY, GOMS IS A FRAMEWORK FOR ESTIMATING THE TIME IT TAKES FOR 3 MAIN TYPES OF INTERACTIONS

- **System Responses:** The time for the system to respond or display information to the user
- **Cognitive Processes:** The time for the user to perceive and think
- **Motor Processes:** The time for the user to take action physically on or with the system

E.g., Looking: 550ms

E.g., Screen render: 300 ms

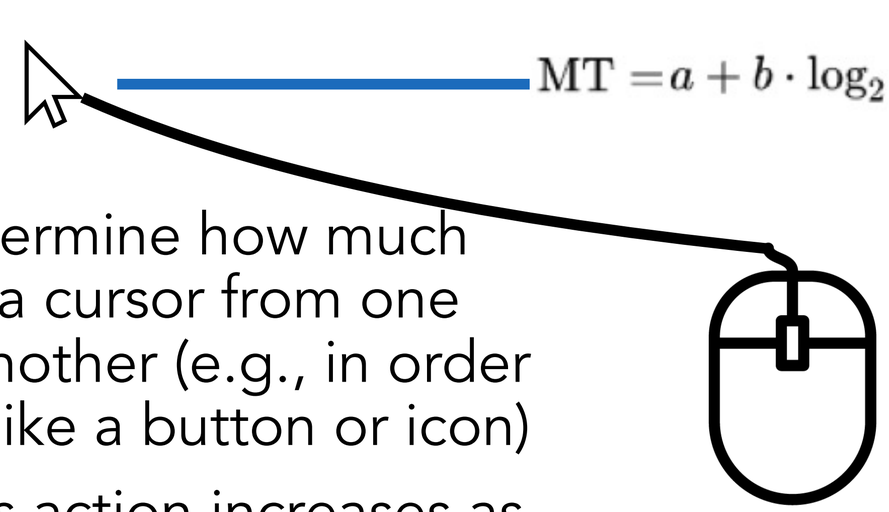




E.g., Move hand to mouse

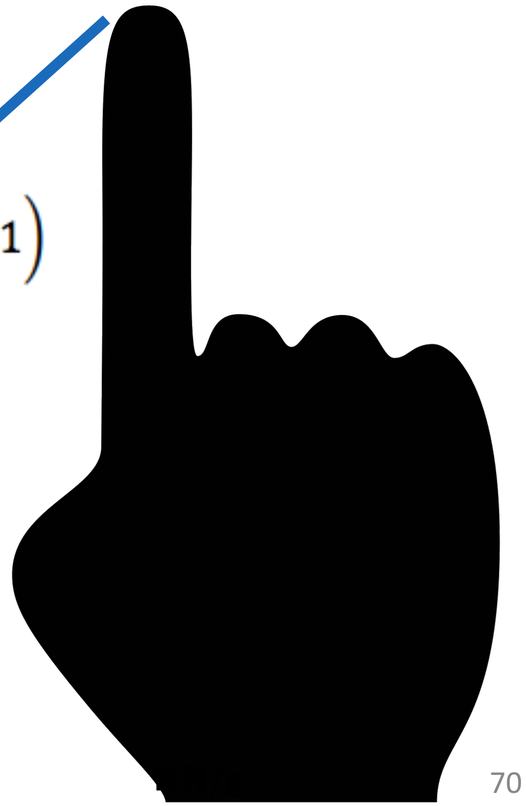
PRINCIPLE:
FITTS'S LAW

$$\mathrm{MT} = a + b \cdot \log_2\left(\frac{2D}{W}\right)$$



- A formula derived to determine how much time it will take to move a cursor from one place on the screen to another (e.g., in order to click/select an object like a button or icon)
- The **time to perform** this action increases as the **distance** needed to travel increases and/or the **size** of the target decreases.
- There are more precise, updated versions of Fitts's Law for touchscreens (e.g., "FFitts" shown here), but the basic idea is the same.

$$T = a + b\log_2\left(\frac{A}{\sqrt{2\pi e(\sigma^2 - \sigma_a^2)}} + 1\right)$$



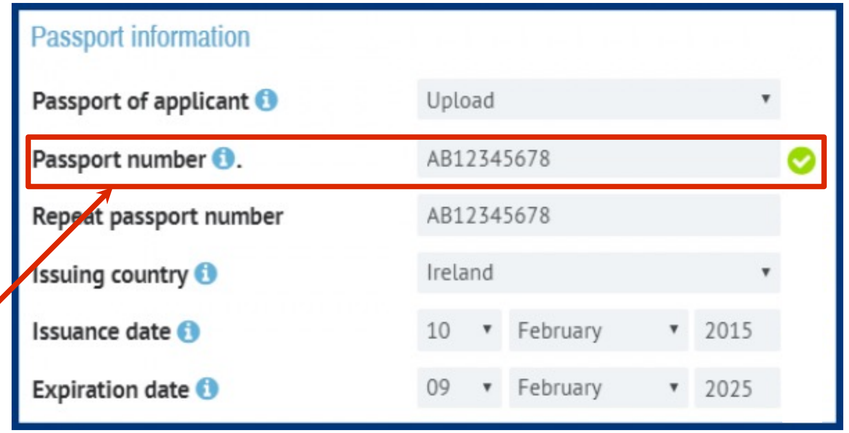
Fitts' Law: https://en.wikipedia.org/wiki/Fitts%27s_law

Touchscreen adaptation "FFitts": https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/41645.pdf

Mobile adaptation: https://webdesign.tutsplus.com/articles/applying-fitts-law-to-mobile-interface-design--webdesign-6919

A **GOMS ANALYSIS** SELECTS A TASK AND SUMS UP THE TIME FOR ALL PHYSICAL (E.G., KEYSTROKING, CLICKING/TAPPING, DRAWING) AND MENTAL OPERATIONS (E.G., FINDING AN ICON, DECIDING). IT CAN ALSO INCLUDE SYSTEM RESPONSE TIME AND DELAYS AS WELL.

A micro-example: Typing in a memorized string of characters (e.g, a passport number):



Using the sample performance times in the table to the right, here is the GOMS Analysis for entering the data into this single field:

- Attend to the screen: 50 ms
- Initiate the hand to move the cursor: 50 ms
- Move cursor into field and click: 1200 ms
- Put hands onto home row (hands): 450 ms
- Initiate the act of typing: 50 ms
- Type 'SHIFT', 'A', 'B': 200 ms x 3 = 600 ms
- Type 8 digits: 200 ms x 8 = 1600 ms
- Proofread the first 5 characters: 330 ms
- Proofread the last 5 characters : 330 ms

Prediction: 50+50+1200+450+50+600+1600+330+330 = 4660 ms or 4.66 secs

Example Performance Times

Attend to the "Passport number" field	50 ms
Initiate the body to perform a motor process	50 ms
Type a key on a computer keyboard (on home row, touch typing)	200 ms
Move the cursor 6" across the screen and click a 3/4" wide target	1200 ms
Tap a key on the keyboard of a smartphone (small target; finger already there)	80 ms
Proofread a word or a chunk of 3-5 characters	330 ms
Look at an item at a known position	550 ms
Place hands onto the home row ("hands")	450 ms

Values derived based on the Fitts' Law Calculator, Kieras (2001) and Batran & Dunlop (2014)

Q: How do you think this would this change if the number was **not** memorized?

WHEN TO USE KLM-GOMS

- When you want to model **low-level estimates** of how much time it will take for an **experienced user** to perform **specific tasks**
- Useful for evaluating and **comparing** different designs, especially for **efficiency**
- Note: It does not account for:
  - Time to learn the system
  - Errors or difficulty encountered

PERFORM A KLM-GOMS TASK ANALYSIS

1. Choose a task scenario (a Goal) on a working system or a prototype that supports this task.

2. Determine the best way to do the task, or the way that you assume users will do it (Method and Selection Rules). Walk through the Method of performing this task (i.e., all the steps), and assume no errors are made or wrong paths taken. It is helpful to number your steps.

3. List the keystroke-level actions and the corresponding **physical** Operators involved in doing the task (see KLM-GOMS table on the following slide for many of these actions).

4. If necessary, include Operators for when the user must wait for the **system** to respond.

5. Insert a **mental** Operator for when user has to monitor the system, such as to see how it responded to an action (aka feedback).

6. Look up the standard execution time to each Operator (see table).

7. Add up the execution times for the Operators. The total of the Operator times is the estimated time to complete the task.

Note: You could compare the time obtained with another way to complete this task by repeating steps 1-6 to see which requires more time (effort). You might do this after a redesign of this task flow, with a competitor's product, or even just with an alternate Method of accomplishing this task.

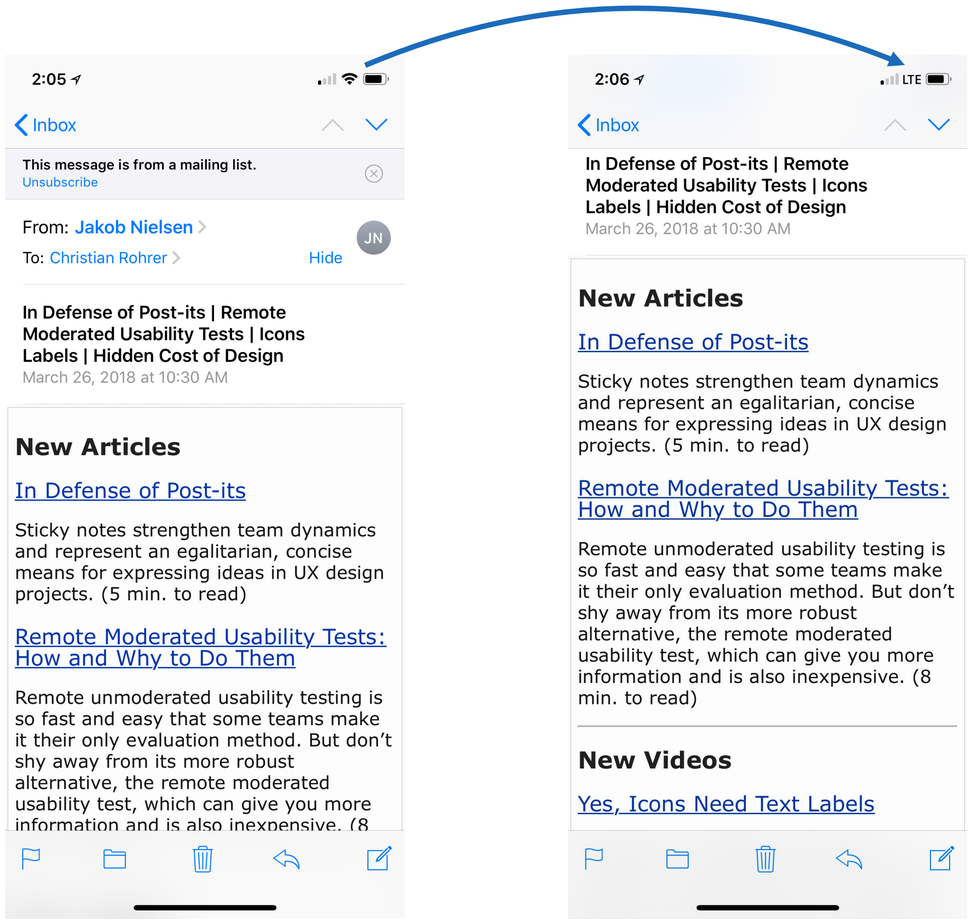
Exercise documents: http://bit.ly/nng-ux-assessments

GOMS OPERATORS

- The times for operators are usually empirically determined or based on known equations (e.g., Fitts's Law)
- If you don't have any values for your system and users, you can start with published tables or derive them manually from published formulae or calculators
- Each system and user type will vary to some degree, so make adjustments for your case.
- The table shown here is a combination of my own calculations (e.g., using Fitts' Law) and a summary from the makers of Cogulator, software that helps you build a GOMS model and analysis (source: http://cogulator.io/primer.html#OPERATORS)

	GOMS Operator	Time (in milliseconds)	Description
Visual	Look	550	Looking at an item at a known position
	Read	260	Read one word
	Proofread	330	Carefully read one word
	Search	1250	Visually search for an item at an unknown position
Audition & Speech	Hear	400	Attend to someone speak (begin listening)
	Say	400	Speak (preparing to)
Cognition	Attend	50	Shifting attention to stimuli
	Initiate	50	Initiate motor process
	Recall	550	Retrieve information from memory
	Store	50	Place item in working memory
	Think	1250	Generic operator for thinking
Motor	Click	320	Click of a mouse
	Drag	230	Drag item across touchscreen
	Grasp	750	Reach and grasp a nearby object with hand
	Hands	450	Move hands to mouse or keyboard
	Keystroke	280	Pressing a single key (not on home row, e.g., Enter or ESC)
	Point	950	Movement of cursor via mouse (or use Fitts' Law estimator to calculate value)
	Write	2000	Handwriting (2 seconds per word)
	Swipe	170	Swipe or flick touchscreen (gesture)
	Tap	450	Tap on a touch screen (finger already over or near the target)
	Turn	800	One turn of a knob of dial
	Type	200	Type a key on a keyboard (dependent on typist's speed; slower on touchscreens)
System	Render	300	Render the user interface (including animation) - iOS
	Load	500-10000	Load a web page (highly dependent on connection speed and system specs; users notice delays > 1000 ms, and will abandon more when Load > 10000 ms)

EXAMPLE TASK: TURN OFF WIFI THROUGH THE SETTINGS APP WHEN IN A THE EMAIL APP

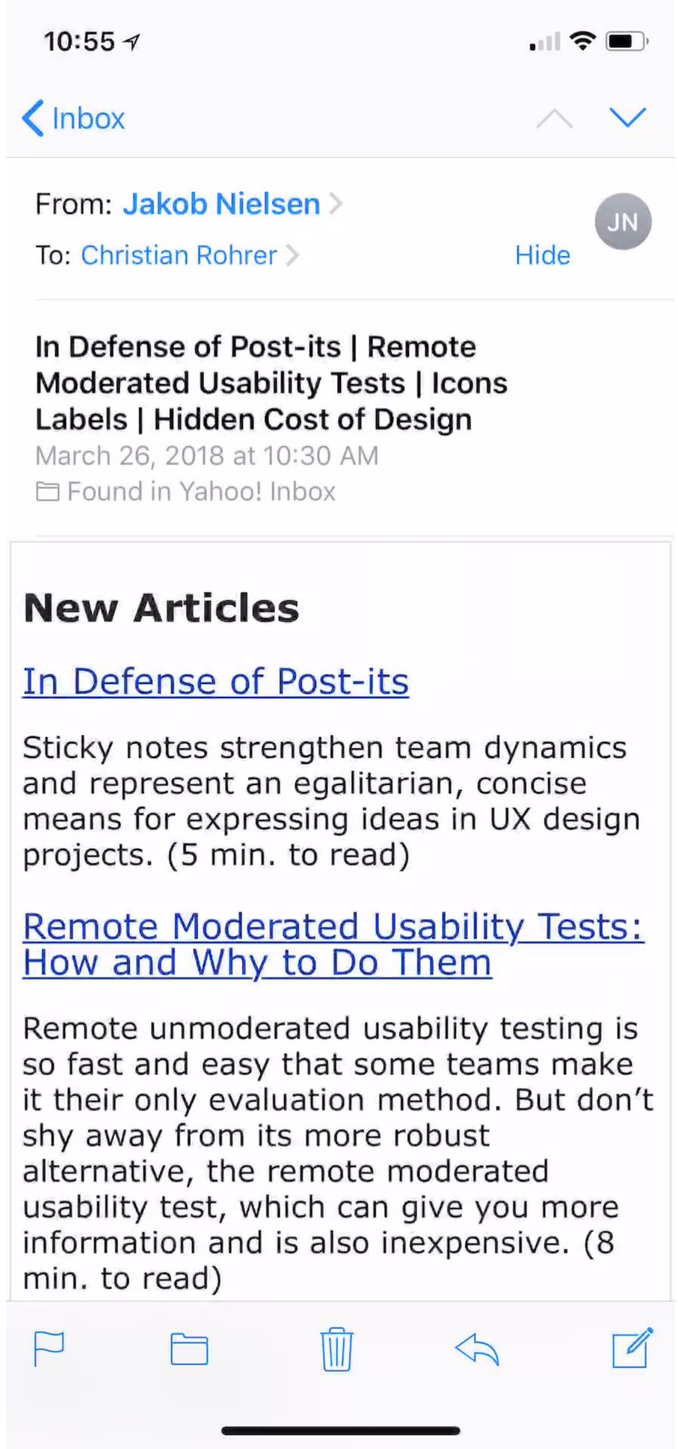


Start State

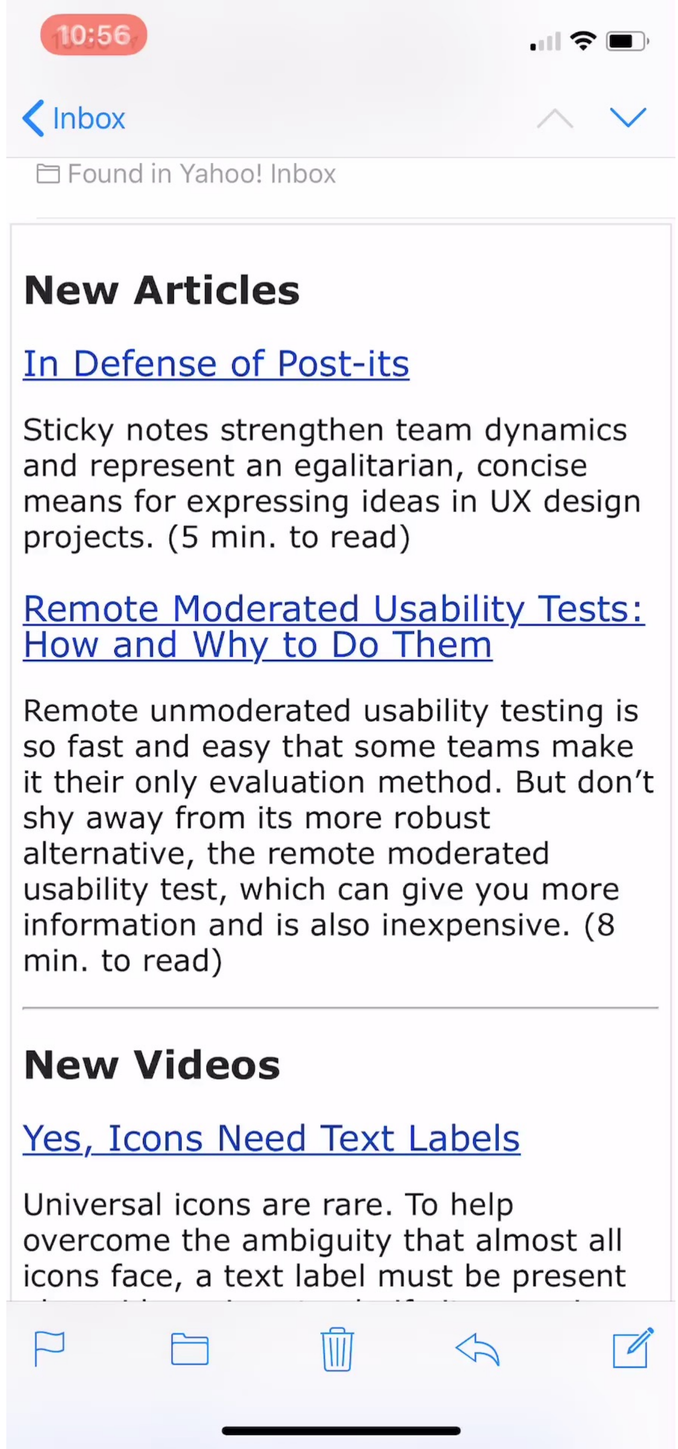
End State

Let's watch two ways to do this…

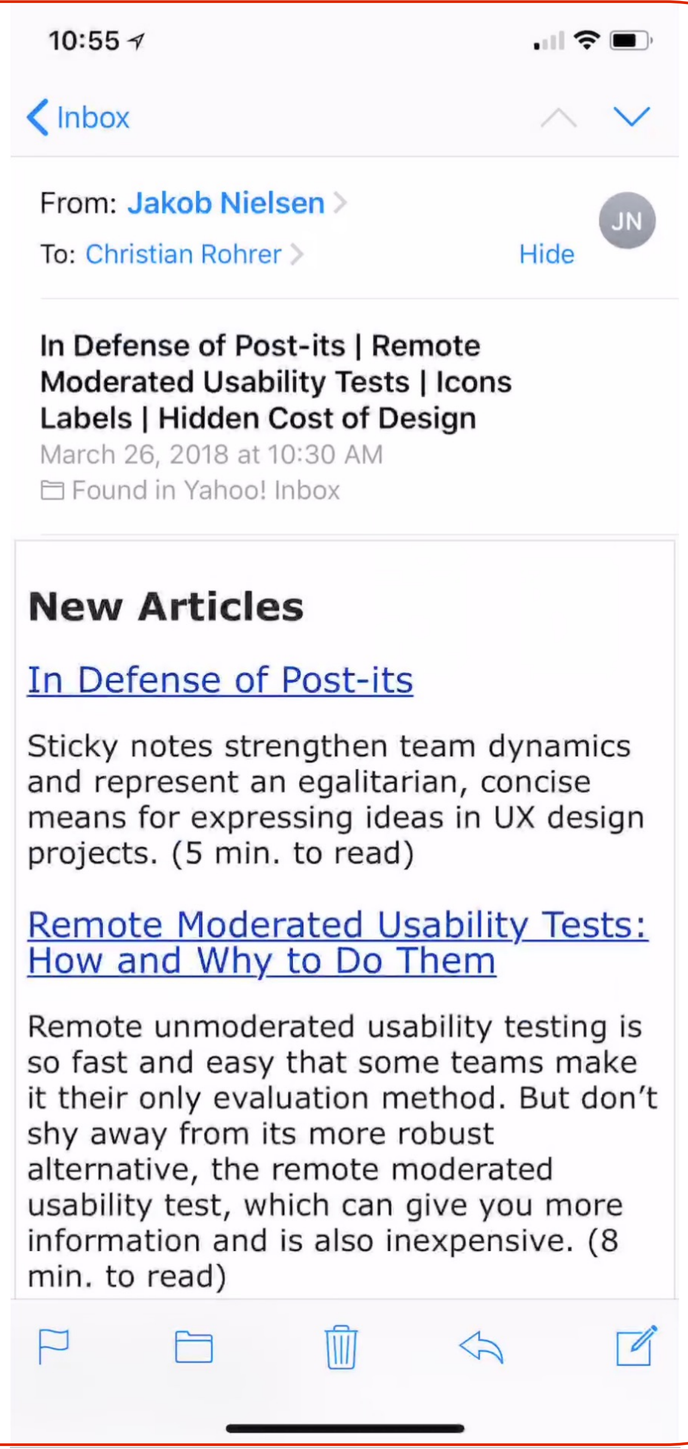
1. Through the Settings app



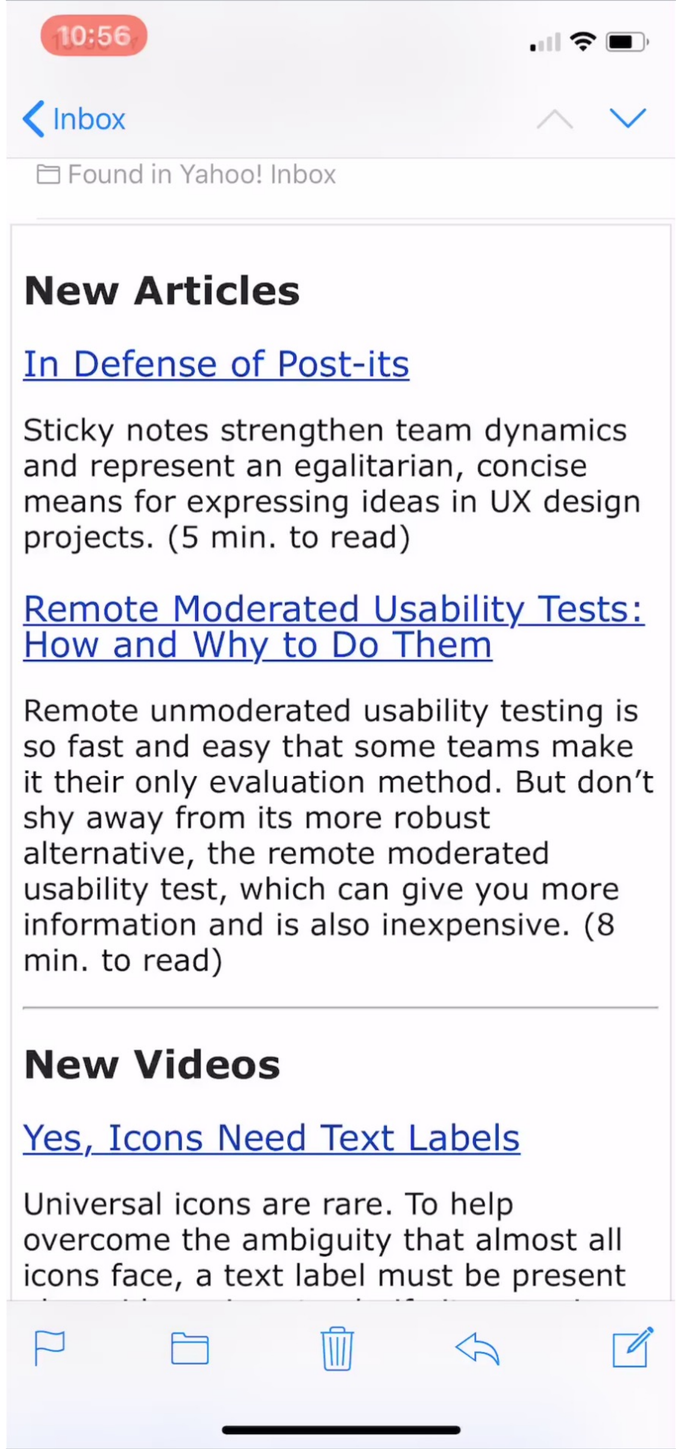
2. With Control Center

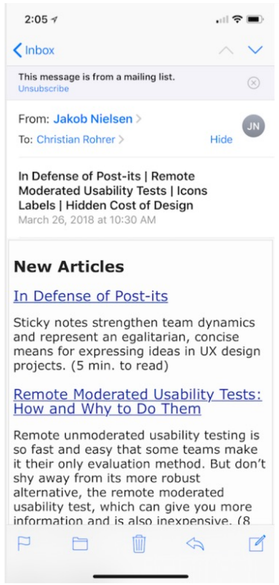


1. Through the Settings app

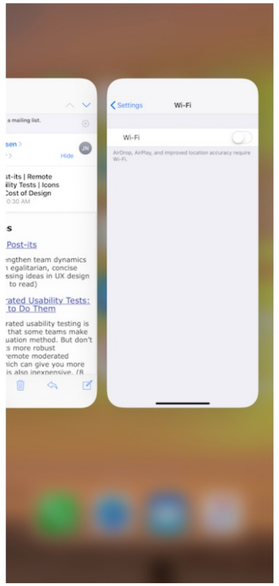


2. With Control Center

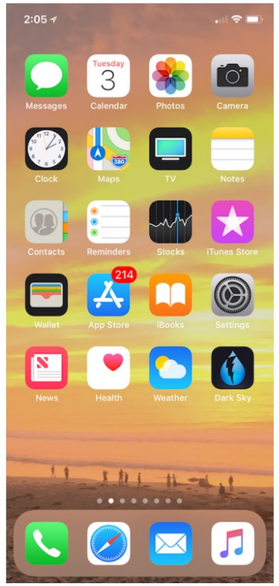




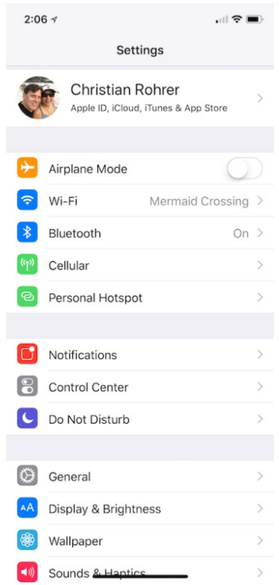
Screen: Email app



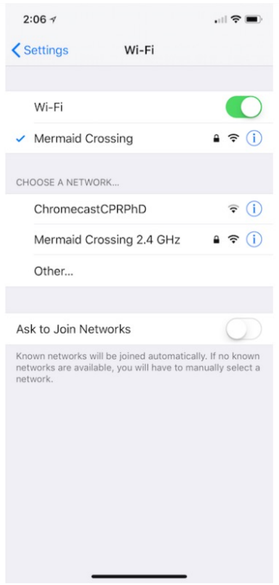
Screen: Swipe up to return to Home



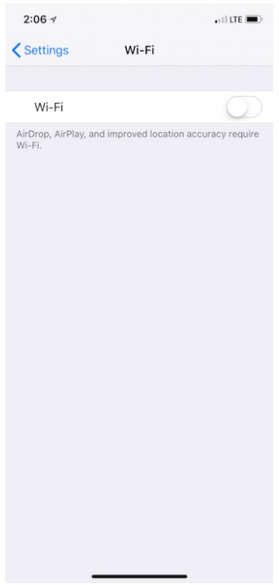
Screen: Home screen to select Settings app



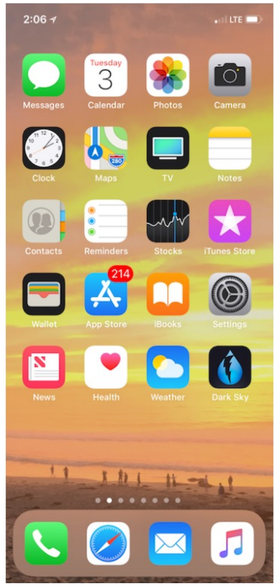
Screen: Settings app



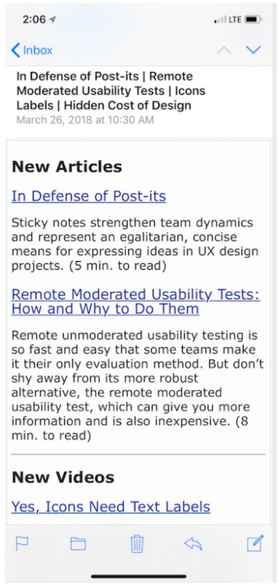
Screen: Wi-Fi settings



Screen: Wi-Fi turned off

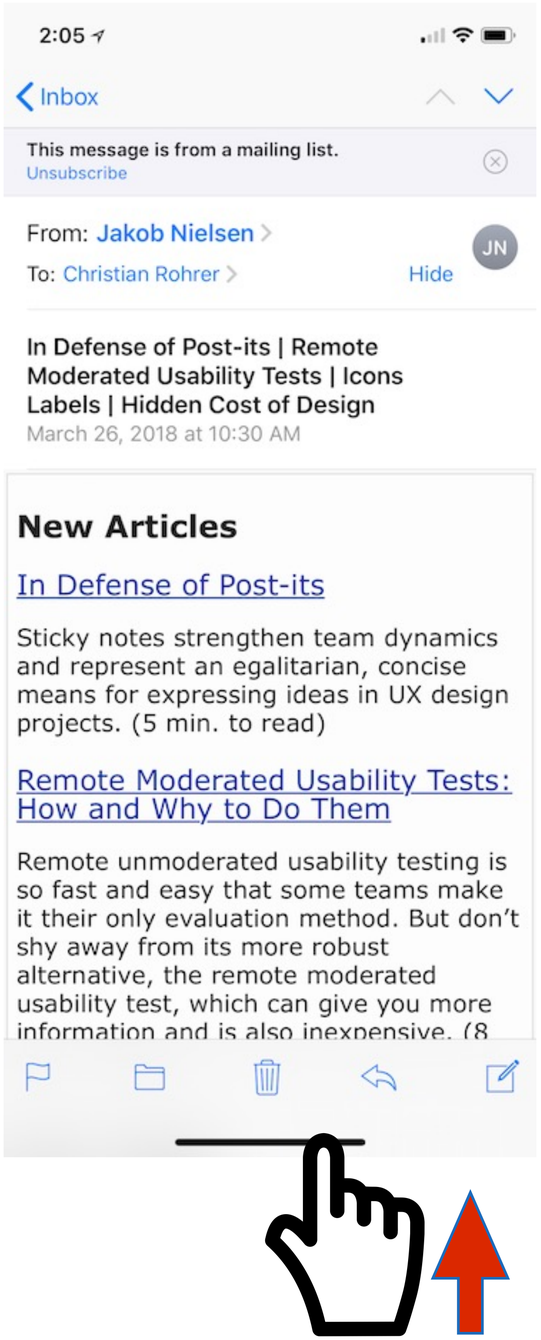


Screen: Return to home screen (with swipe up)

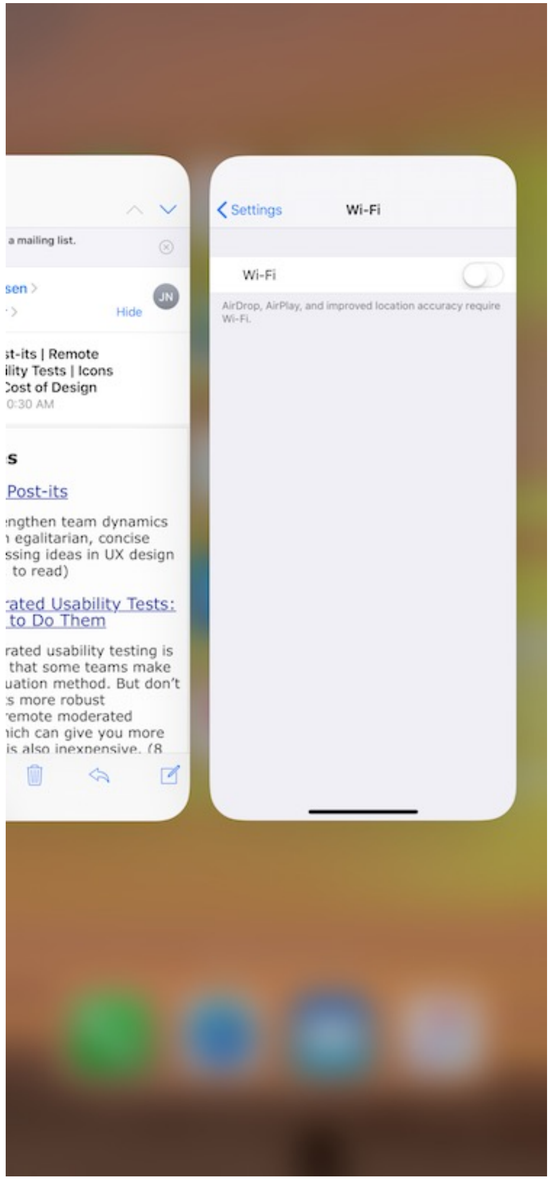


Screen: Tap on email to return to app

Exercise documents: http://bit.ly/nng-ux-assessments

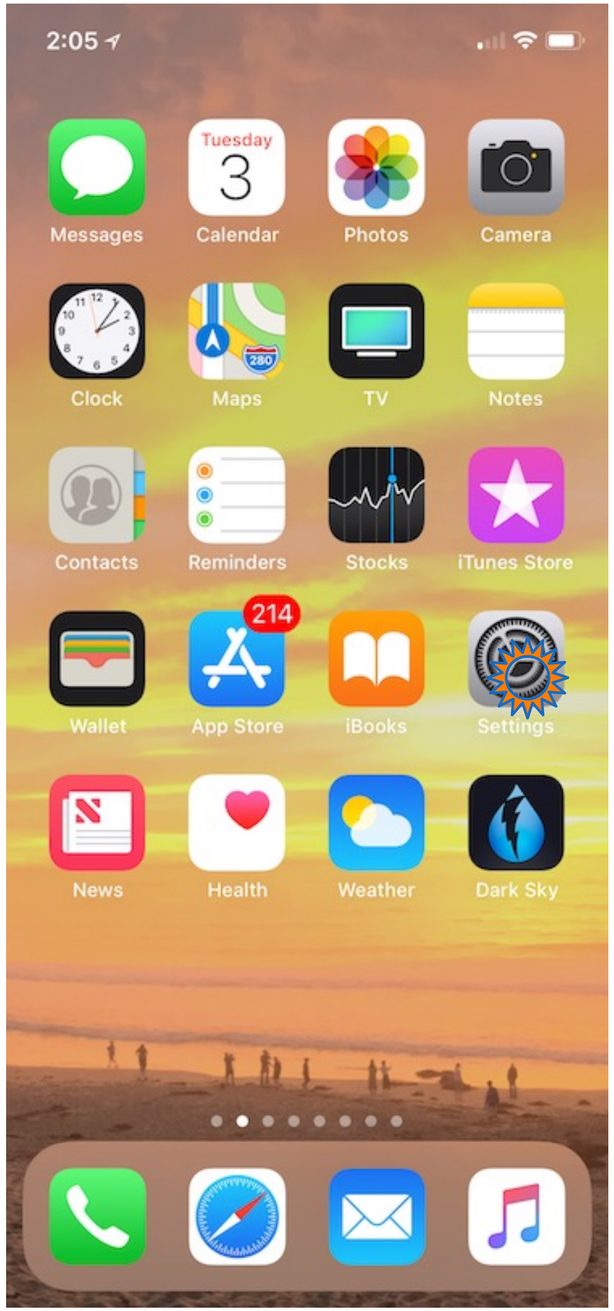


Screen: Email app; Swipe up from bottom



Screen: Return to Home (animation transition)

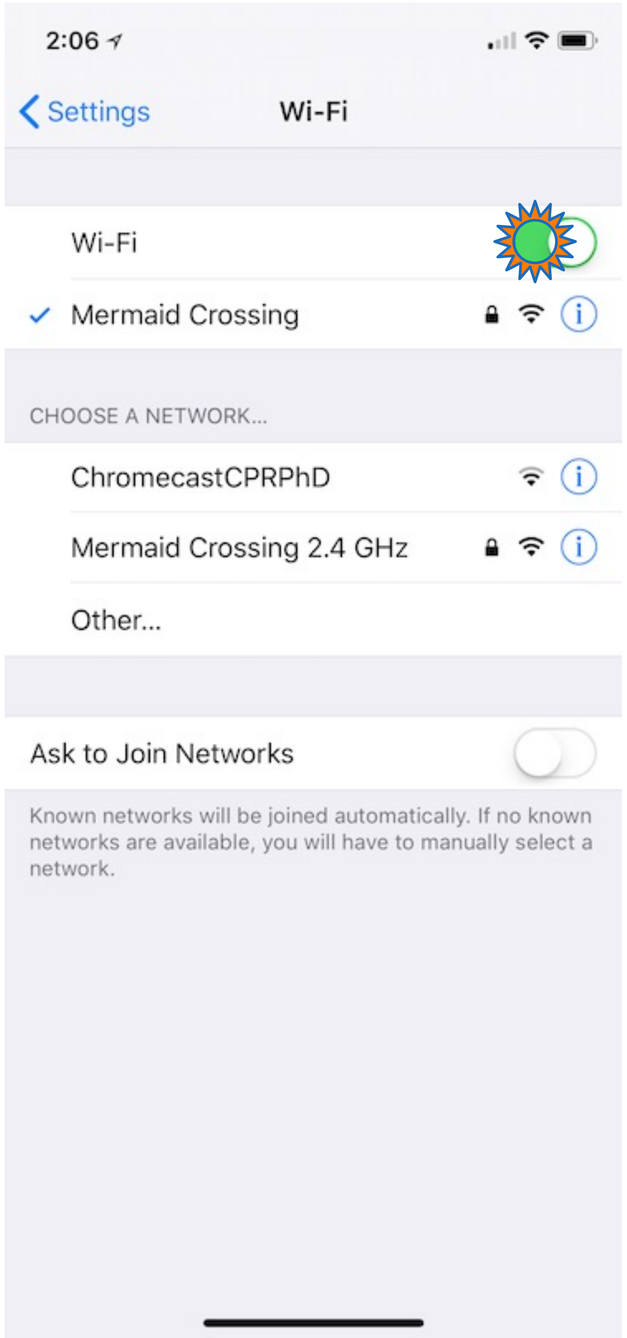
Exercise documents: http://bit.ly/nng-ux-assessments



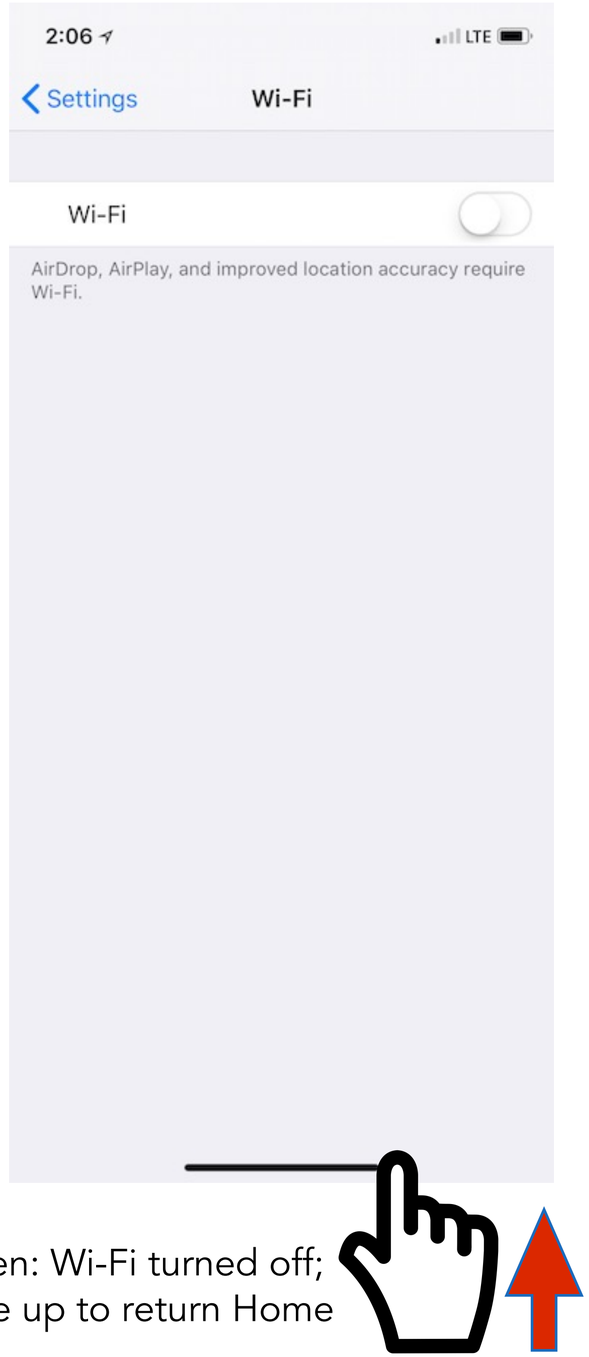
Screen: Home screen to select Settings app

Screen: Settings app

Exercise documents: http://bit.ly/nng-ux-assessments

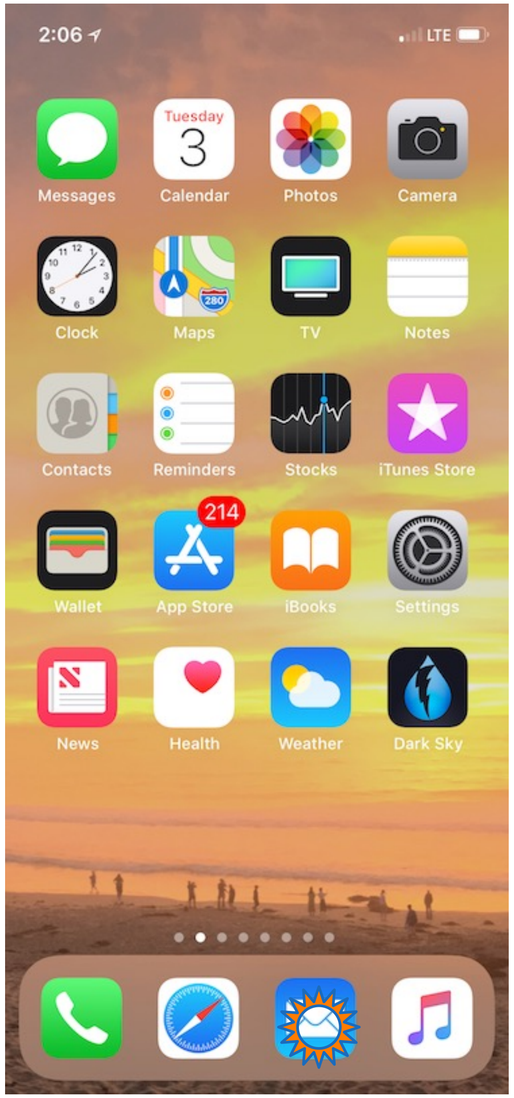


Screen: Wi-Fi settings

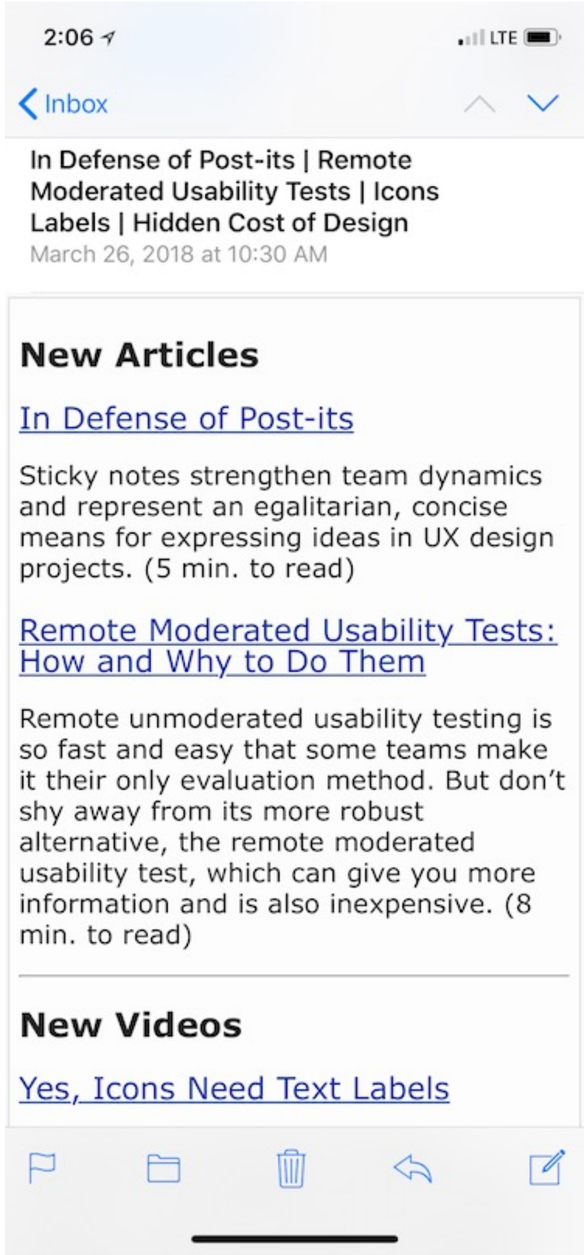


Screen: Wi-Fi turned off;
swipe up to return Home

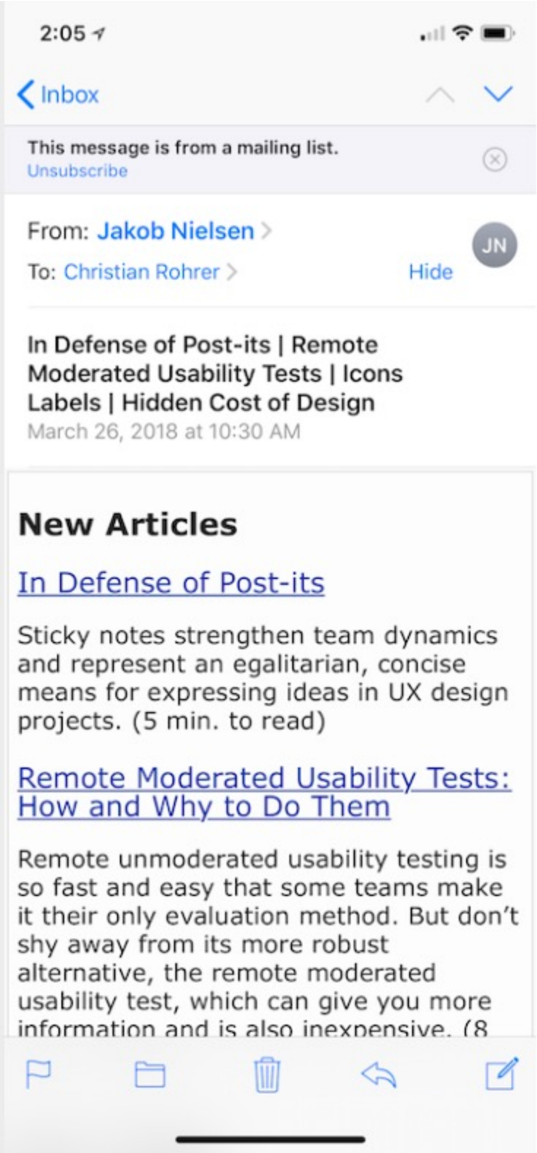
Exercise documents: http://bit.ly/nng-ux-assessments



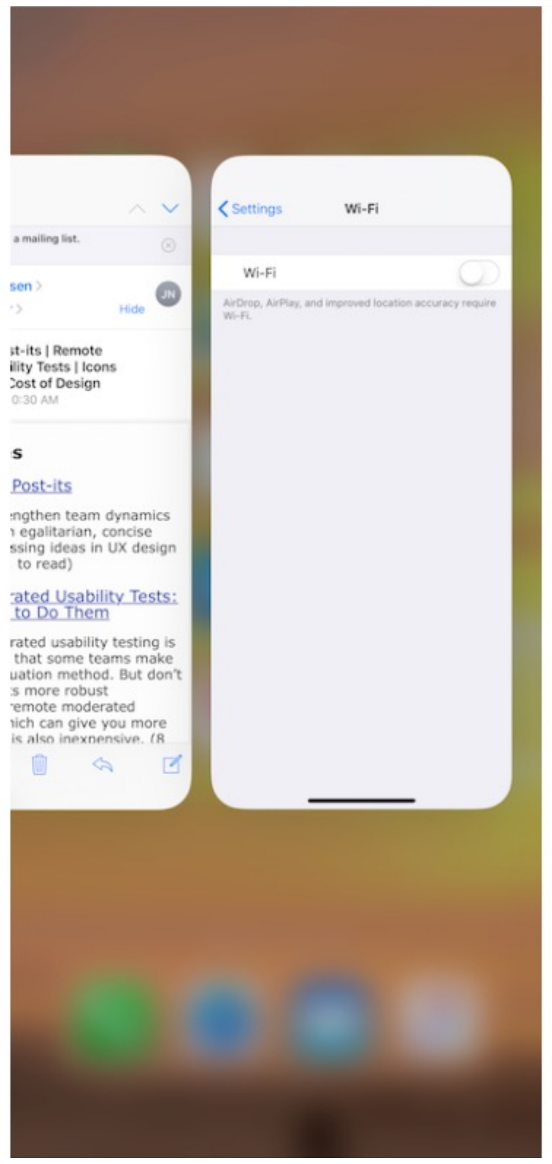
Screen: Return to home screen (after swipe up); Tap Email



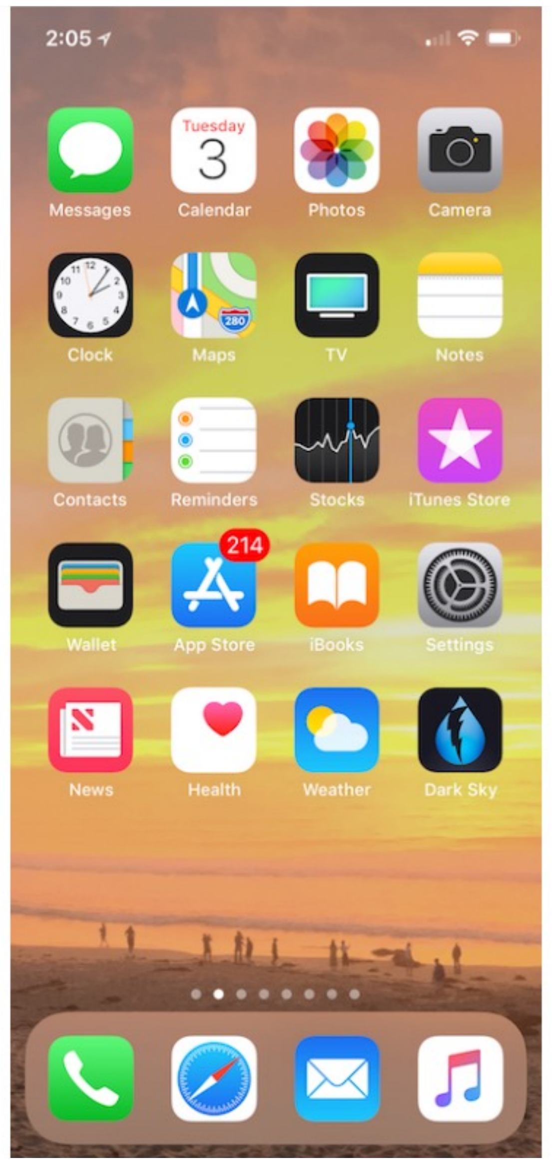
Screen: Return to app



Screen: Email app

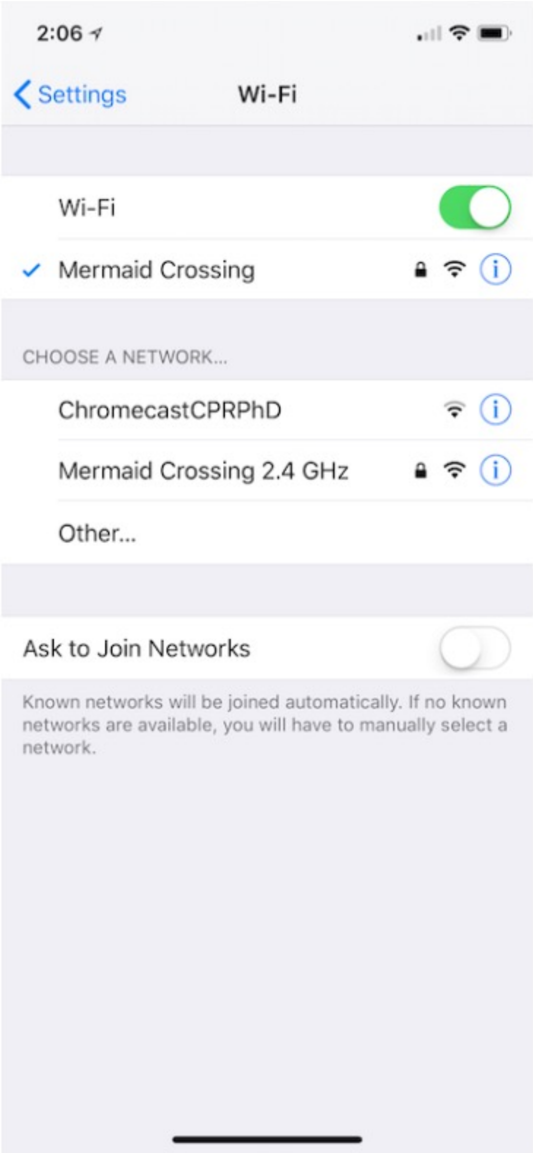


Screen: Swipe up to return to Home

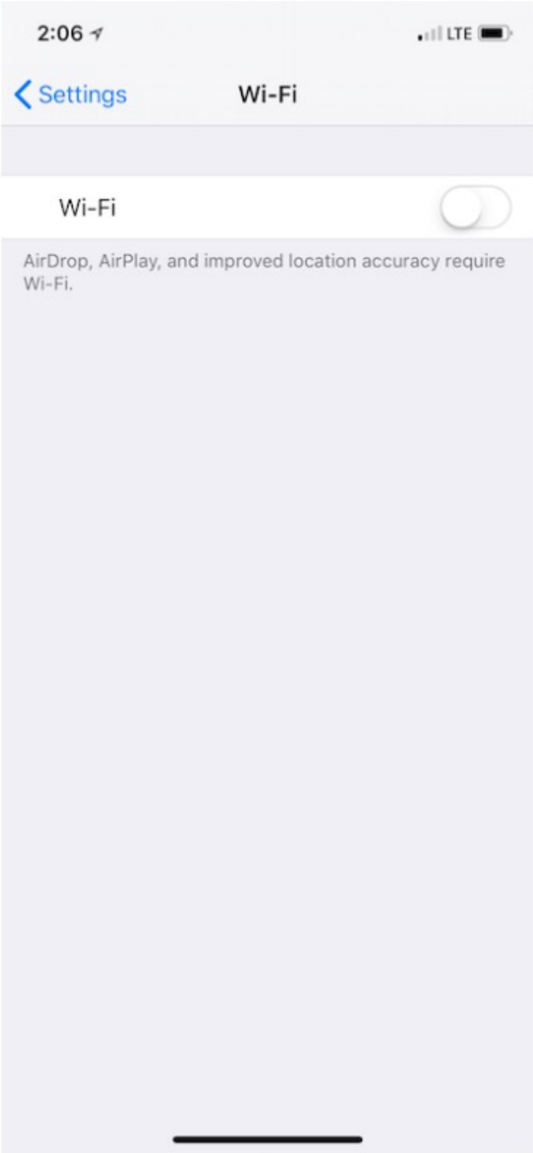


Screen: Home screen to select Settings app

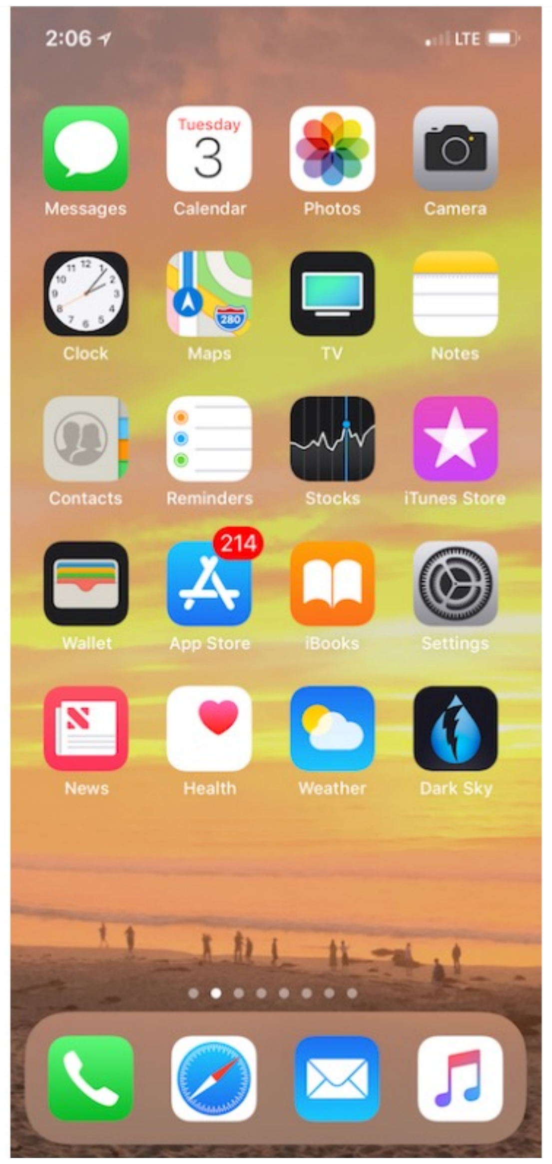
Screen: Settings app



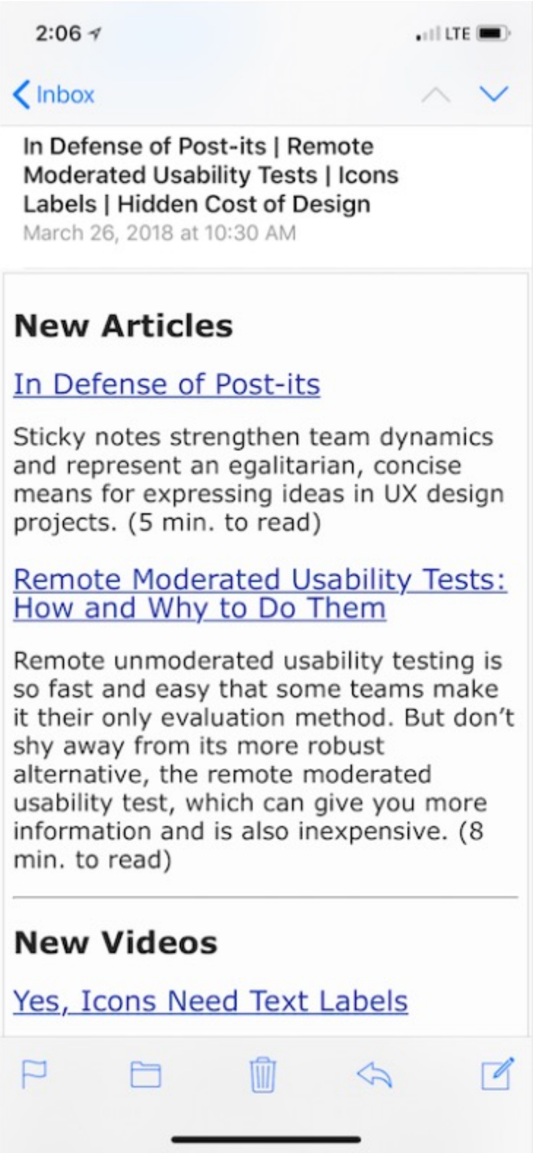
Screen: Wi-Fi settings



Screen: Wi-Fi turned off



Screen: Return to home screen (with swipe up)



Screen: Tap on email to return to app

Method 1: Turn off Wifi using the Settings app

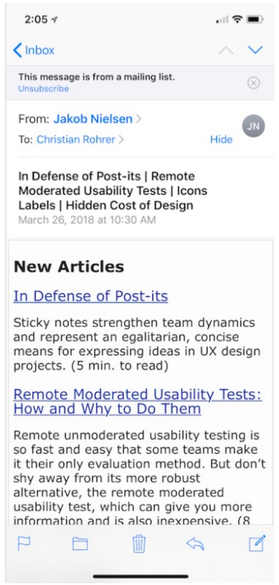
Instructions:

1. Using your own phone or using the screens above, go through the task from beginning to end.
2. Write down all the GOMS Operators (from the earlier table) that an efficient user would use to accomplish this task in this way. Some Operators, such as the Cognitive Operators or System Responses, occur between screens shown above.
3. Enter the Time needed for each Operator.
4. Sum up the time.

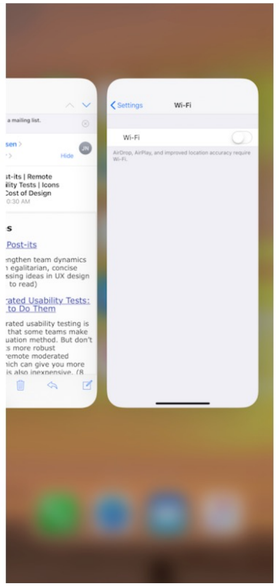
Table to capture Operators and their descriptions

Operator Description [Short name]	Time (ms)

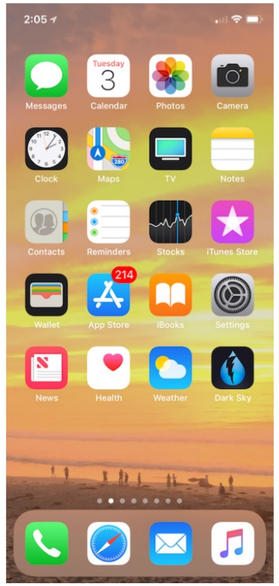
Exercise documents: http://bit.ly/nng-ux-assessments



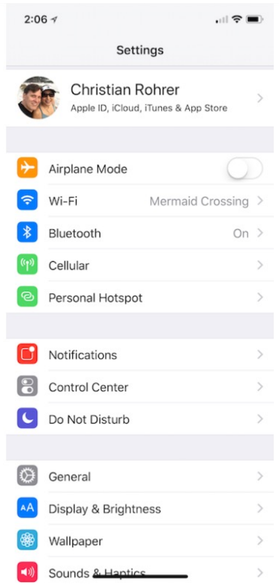
Screen: Email app



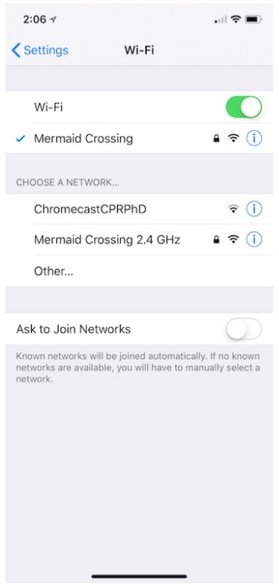
Screen: Swipe up to return to Home



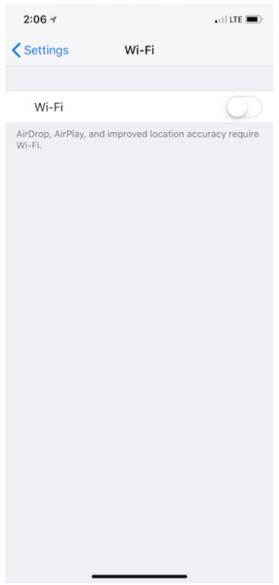
Screen: Home screen to select Settings app



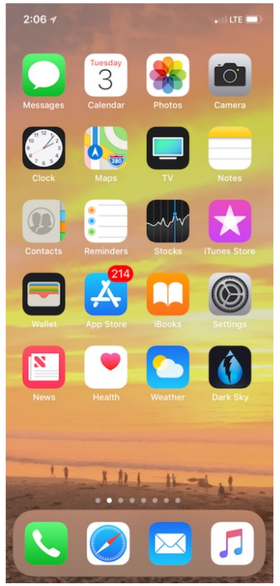
Screen: Settings app



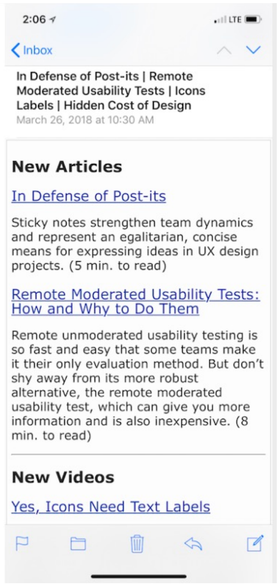
Screen: Wi-Fi settings



Screen: Wi-Fi turned off



Screen: Return to home screen (with swipe up)



Screen: Tap on email to return to app

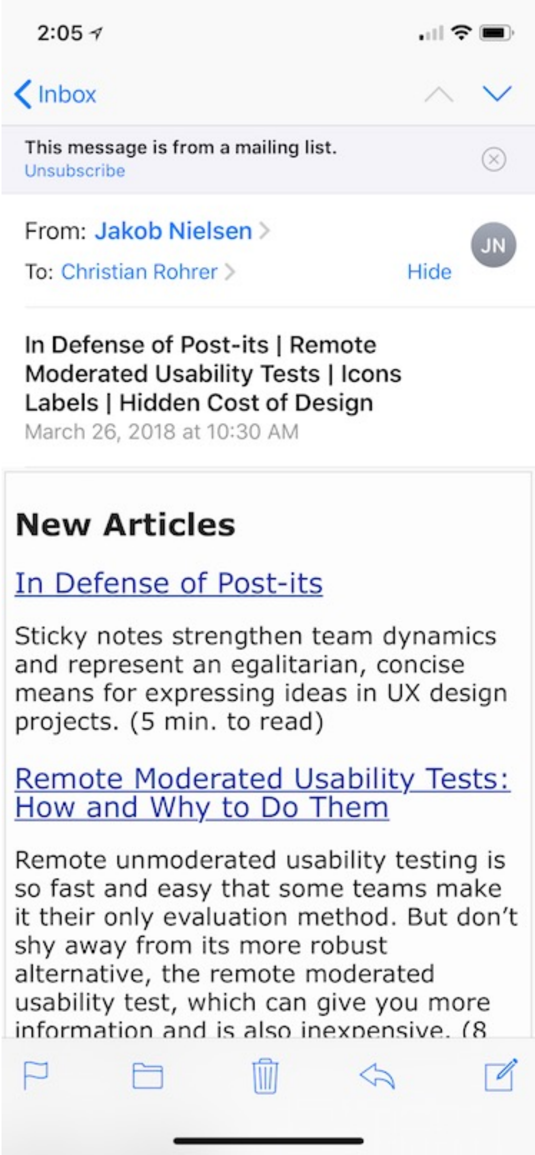
Method 1: Turn off Wifi using the Settings app

Instructions:

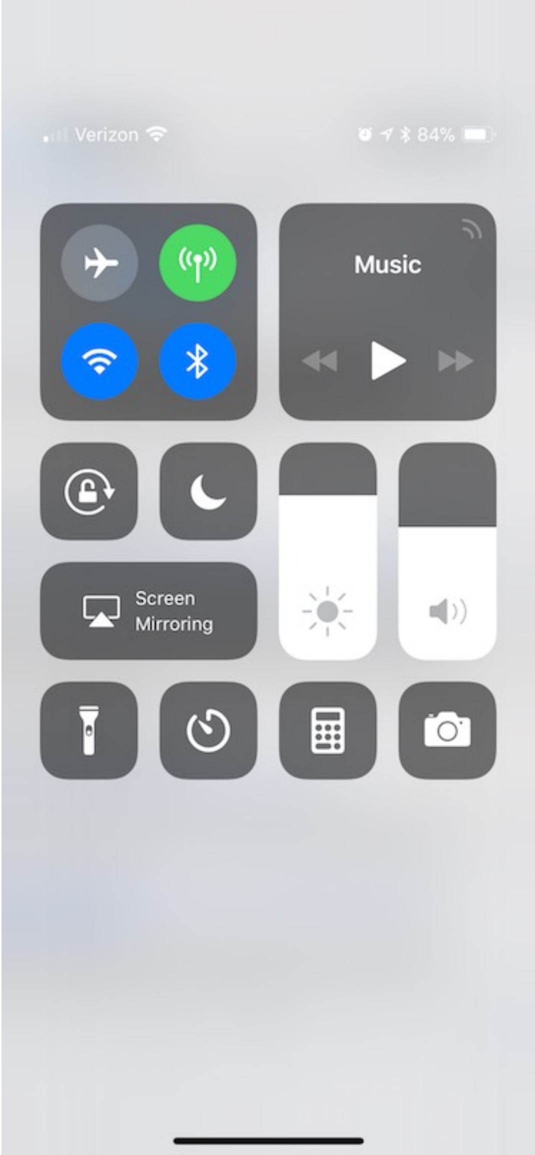
1. Using your own phone or using the screens above, go through the task from beginning to end.
2. Write down all the GOMS Operators (from the earlier table) that an efficient user would use to accomplish this task in this way. Some Operators, such as the Cognitive Operators or System Responses, occur between screens shown above.
3. Enter the Time needed for each Operator.
4. Sum up the time.

Table to capture Operators and their descriptions

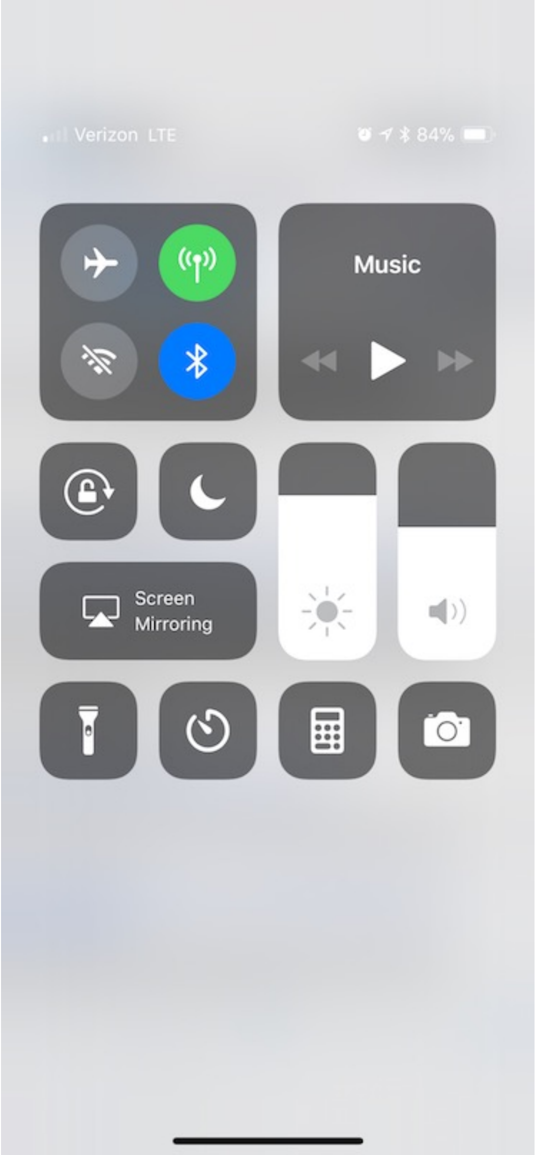
Operator Description [Short name]	Time (ms)
Shift attention [Attend]	50
Initiate motor process [Initiate]	50
Swipe up to go Home [Swipe]	170
System response: Put away Mail app [Render]	300
Visually find Settings App in Home [Look]	550
Initiate motor process [Initiate]	50
Tap on Settings App [Tap]	450
System response: Display Settings app UI [Render]	300
Search for Wi-Fi pane [Look]	550
Initiate motor process [Initiate]	50
Tap on Wi-Fi pane [Tap]	450
System response: Display Wi-Fi menu [Render]	300
Visually find Wi-Fi toggle [Look]	550
Initiate motor process [Initiate]	50
Tap toggle [Tap]	450
Confirm that toggle is off [Look]	550
Initiate motor process [Initiate]	50
Swipe up to go home [Swipe]	170
System response: Put away Settings app [Render]	300
Find mail app icon [Look]	550
Initiate motor process [Initiate]	50
Tap mail app icon [Tap]	450
System response: Display Mail app & message [Render]	300
TOTAL Time to Execute: 6.74 secs	**6740**



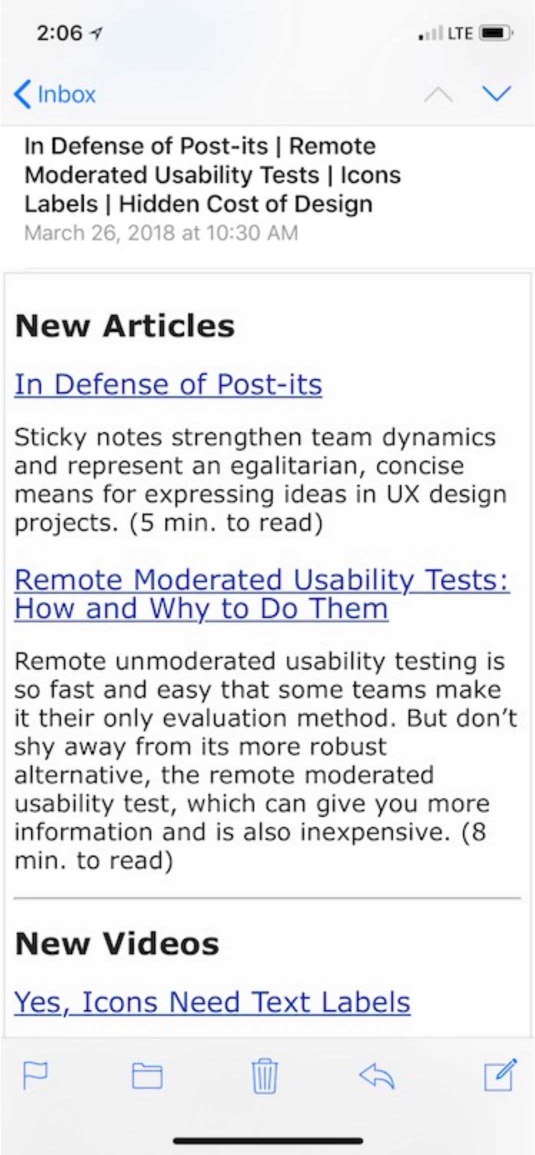
Screen: Email app



Screen: Control Center (rt. corner swipe down)



Screen: After tapping Wi-Fi icon, Wi-Fi is off



Screen: swipe up closes CC, returns to email

Method 2: Turn off Wifi using Control Center (if time permits)

Let's compare with an alternative Method to accomplish this same task by invoking the Control Center screen.

Instructions:

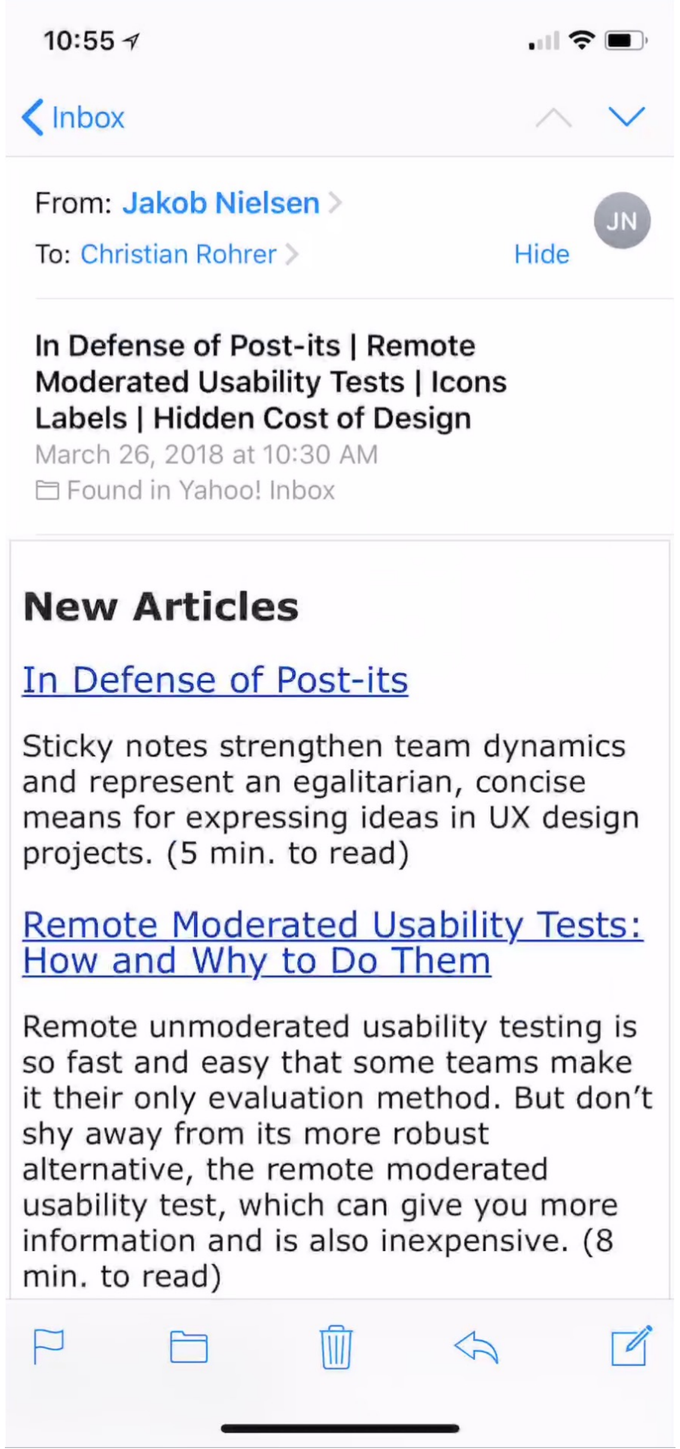
1. Using your own phone or using the screens above, go through the task from beginning to end.
2. Write down all the GOMS Operators (from the earlier table) that an efficient user would use to accomplish this task in this way. Some Operators, such as the Cognitive Operators or System Responses, occur between screens shown above.
3. Enter the Time needed for each Operator.
4. Sum up the time.

Table to capture Operators and their descriptions

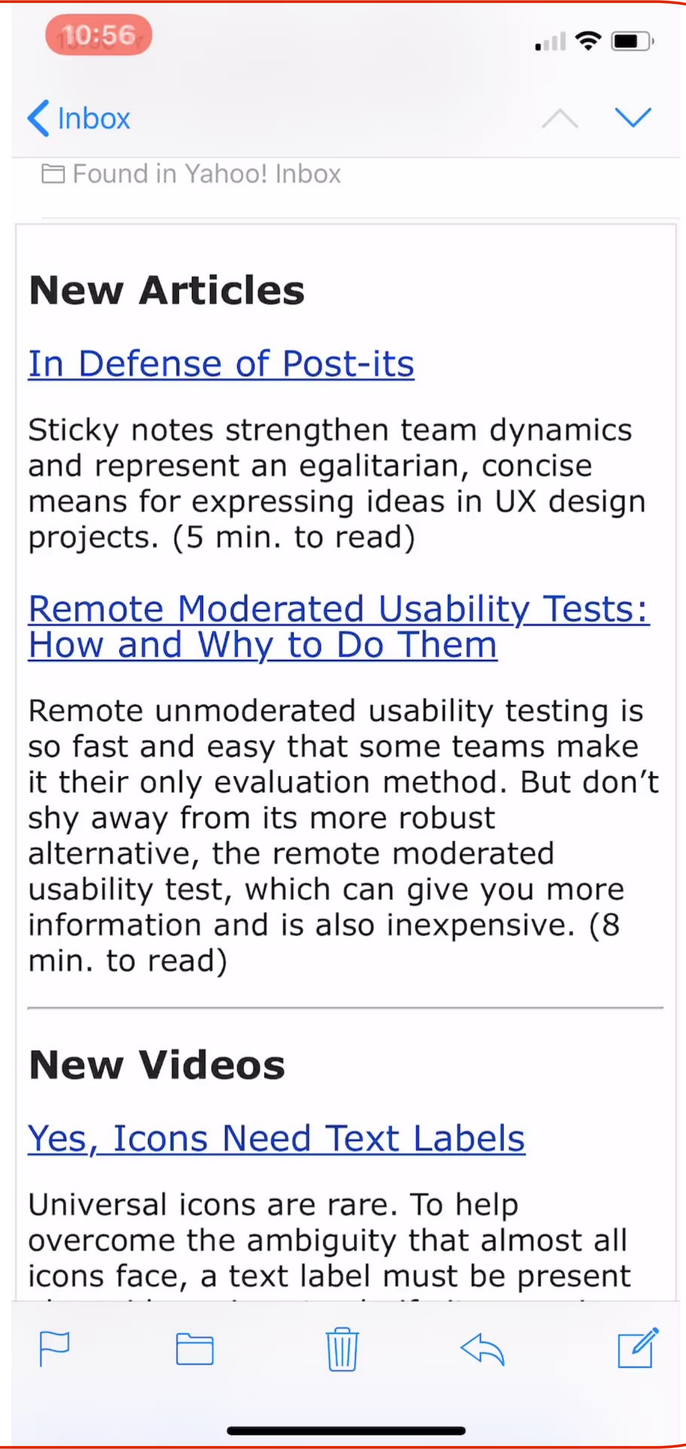
Operator Description [Short name]	Time (ms)
TOTAL Time to Execute	

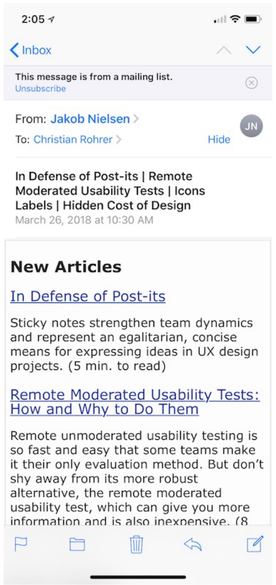
Exercise documents: http://bit.ly/nng-ux-assessments

1. Through the Settings app

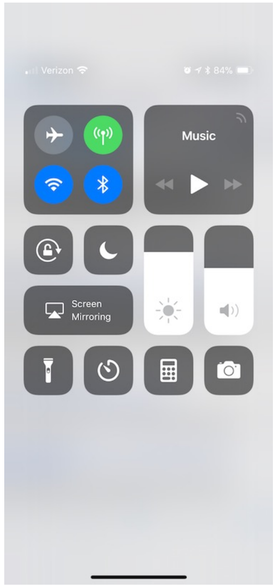


2. With Control Center

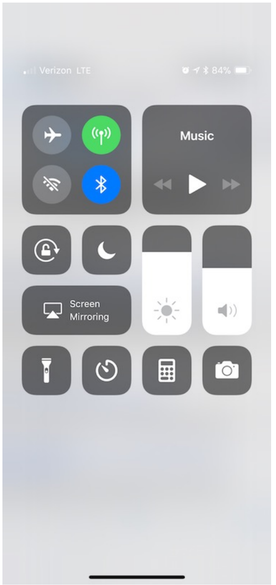




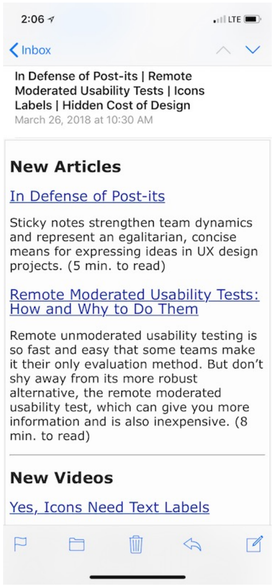
Screen: Email app



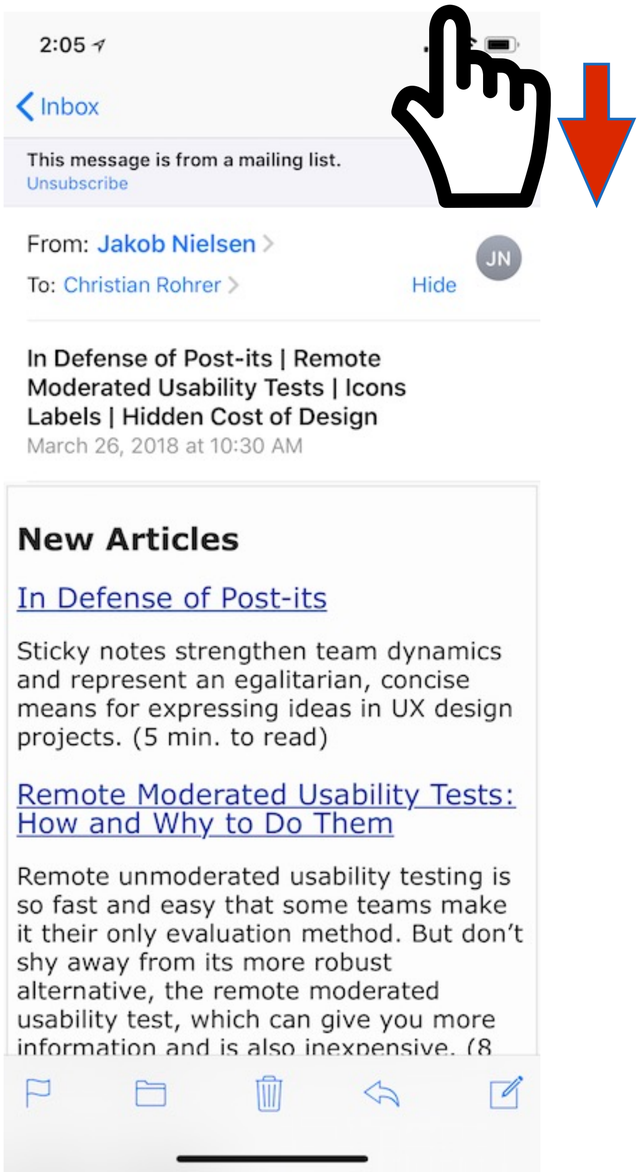
Screen: Control Center (rt. corner swipe down)



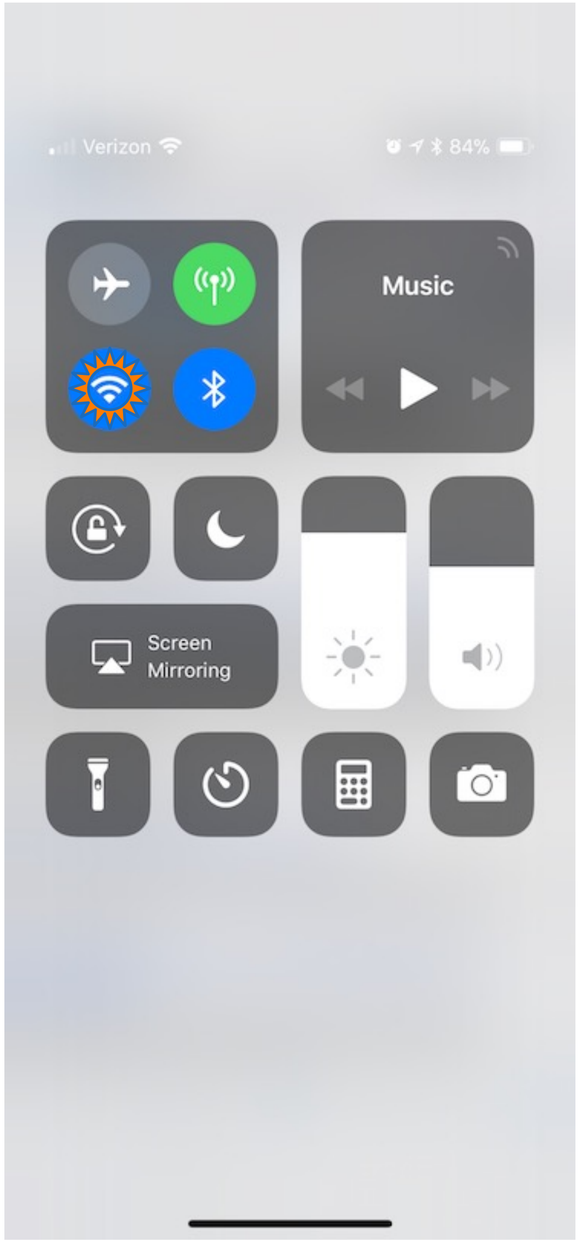
Screen: After tapping Wi-Fi icon, Wi-Fi is off



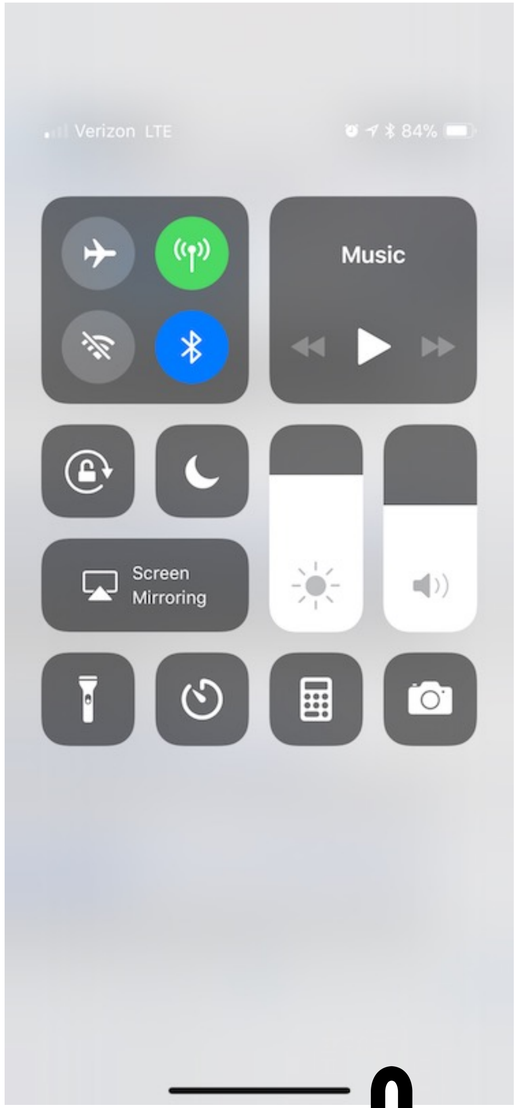
Screen: swipe up closes CC, returns to email



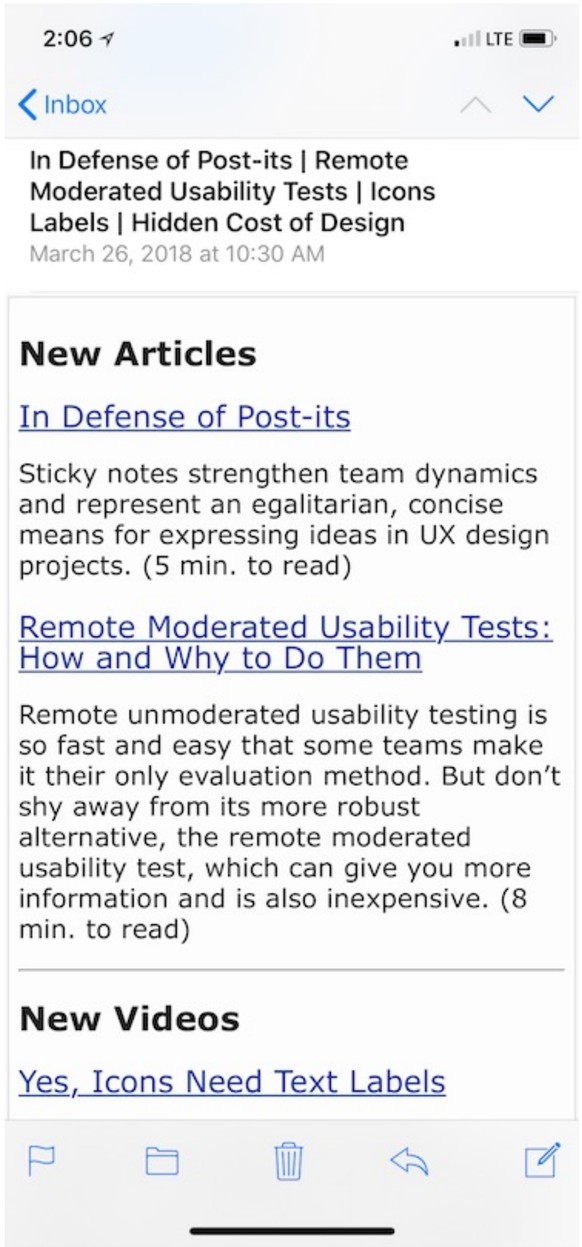
Screen: Email app; swipe down from corner



Screen: Control Center; tap WiFi icon

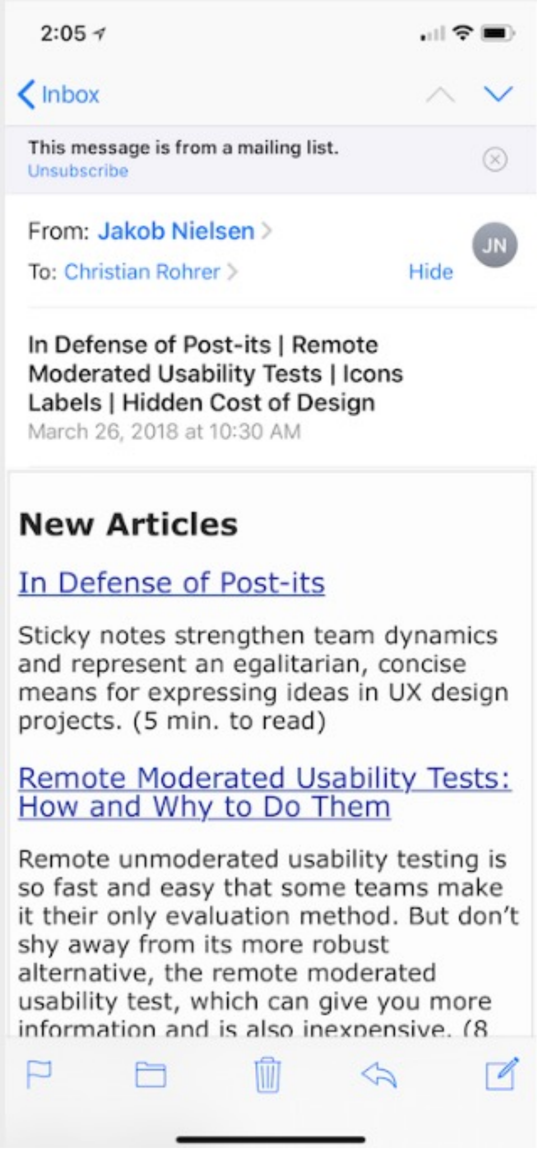


Screen: After tapping Wi-Fi icon, confirm; Wi-Fi is off; swipe up

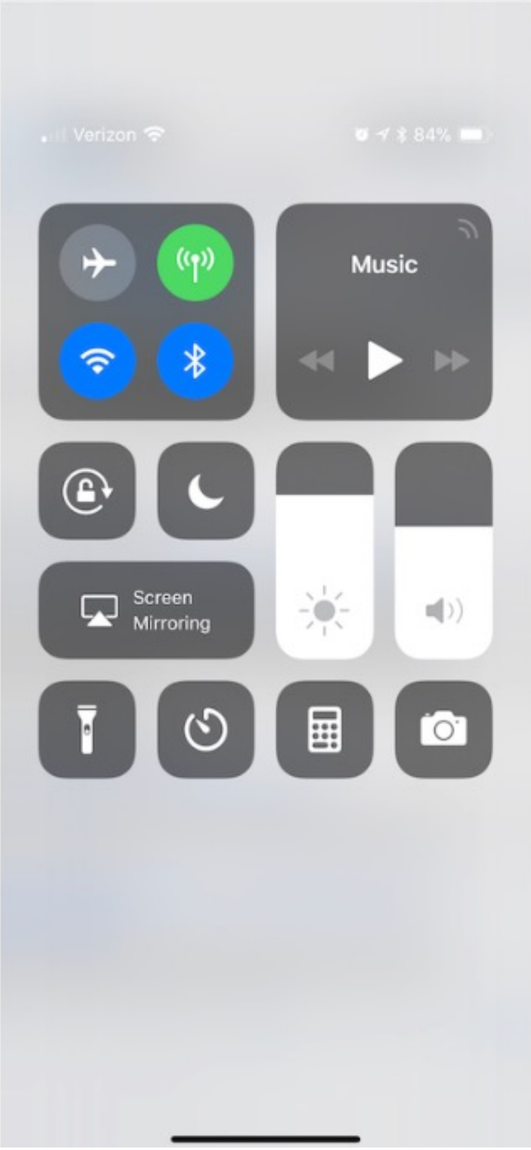


Screen: swipe up closes CC, returns to email

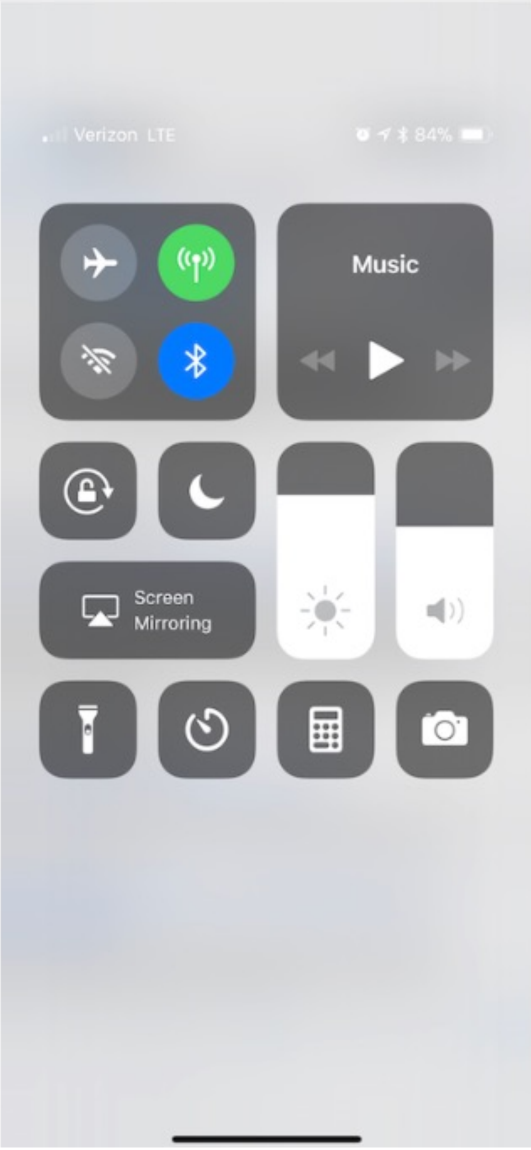
EXERCISE



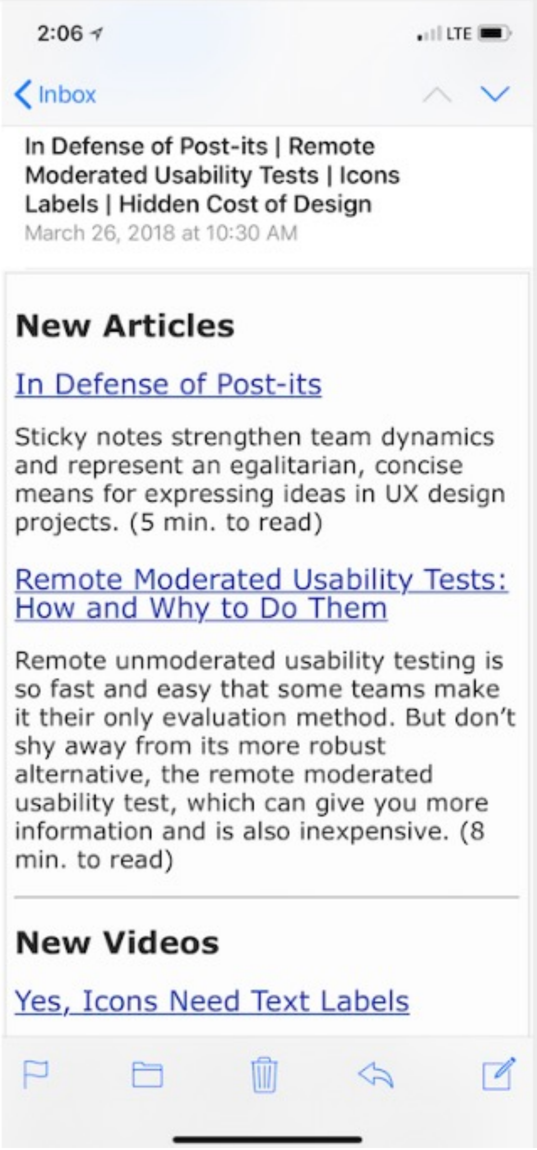
Screen: Email app



Screen: Control Center (rt. corner swipe down)



Screen: After tapping Wi-Fi icon, Wi-Fi is off



Screen: swipe up closes CC, returns to email

Method 2: Using Control Center

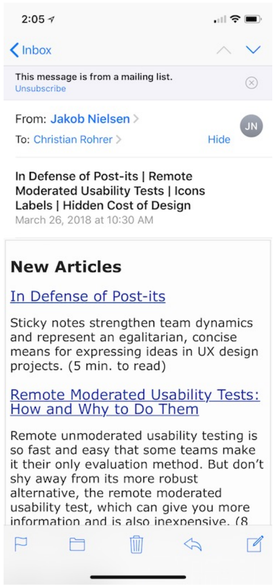
Let's compare with an alternative Method to accomplish this same task by invoking the Control Center screen.

Instructions:

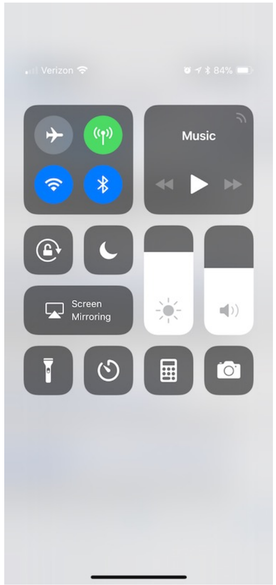
1. Using your own phone or using the screens above, go through the task from beginning to end.
2. Write down all the GOMS Operators (from the earlier table) that an efficient user would use to accomplish this task in this way. Some Operators, such as the Cognitive Operators or System Responses, occur between screens shown above.
3. Enter the Time needed for each Operator.
4. Sum up the time.

Table to capture Operators and their descriptions

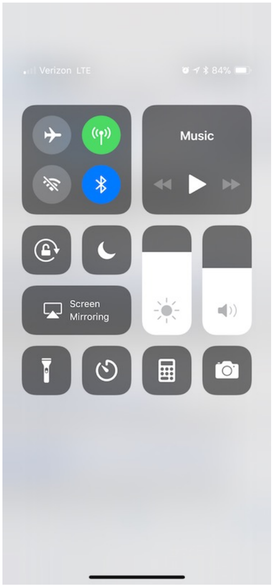
Operator Description [Short name]	Time (ms)
TOTAL Time to Execute	



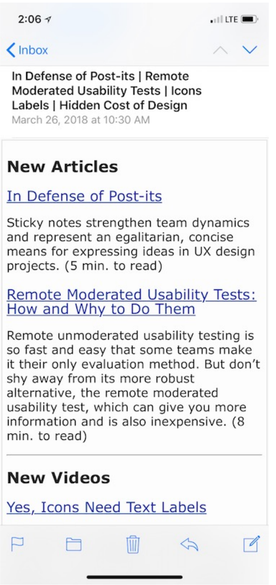
Screen: Email app



Screen: Control Center (rt. corner swipe down)



Screen: After tapping Wi-Fi icon, Wi-Fi is off



Screen: swipe up closes CC, returns to email

Method 2: Using Control Center

Let's compare with an alternative Method to accomplish this same task by invoking the Control Center screen.

Instructions:

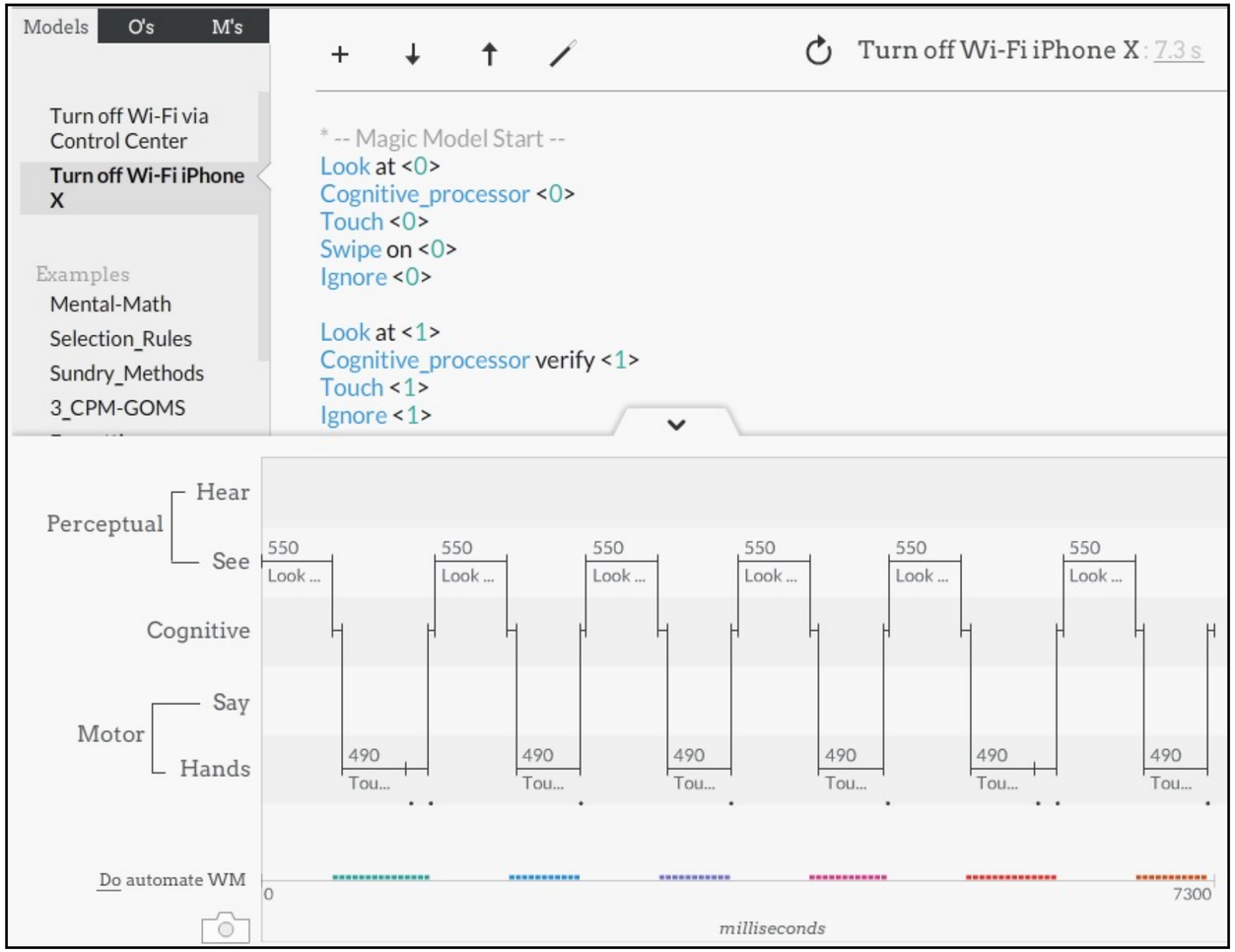
1. Using your own phone or using the screens above, go through the task from beginning to end.
2. Write down all the GOMS Operators (from the earlier table) that an efficient user would use to accomplish this task in this way. Some Operators, such as the Cognitive Operators or System Responses, occur between screens shown above.
3. Enter the Time needed for each Operator.
4. Sum up the time.

Table to capture Operators and their descriptions

Operator Description [Short name]	Time (ms)
Shift attention [Attend]	50
Initiate motor process [Initiate]	50
Swipe down from upper right corner [Swipe]	170
Wait for system to render Control Center UI [Response]	300
Find Wi-Fi icon [Look]	550
Initiate motor process [Initiate]	50
Tap Wi-Fi icon [Tap]	450
Verify that Wi-Fi is off [Look]	550
Initiate motor process [Initiate]	50
Swipe up from bottom to dismiss Control Center [Swipe]	170
Wait for system to close Control Center UI and return to Mail [Response]	300
TOTAL Time to Execute 2.69 secs	**2690**

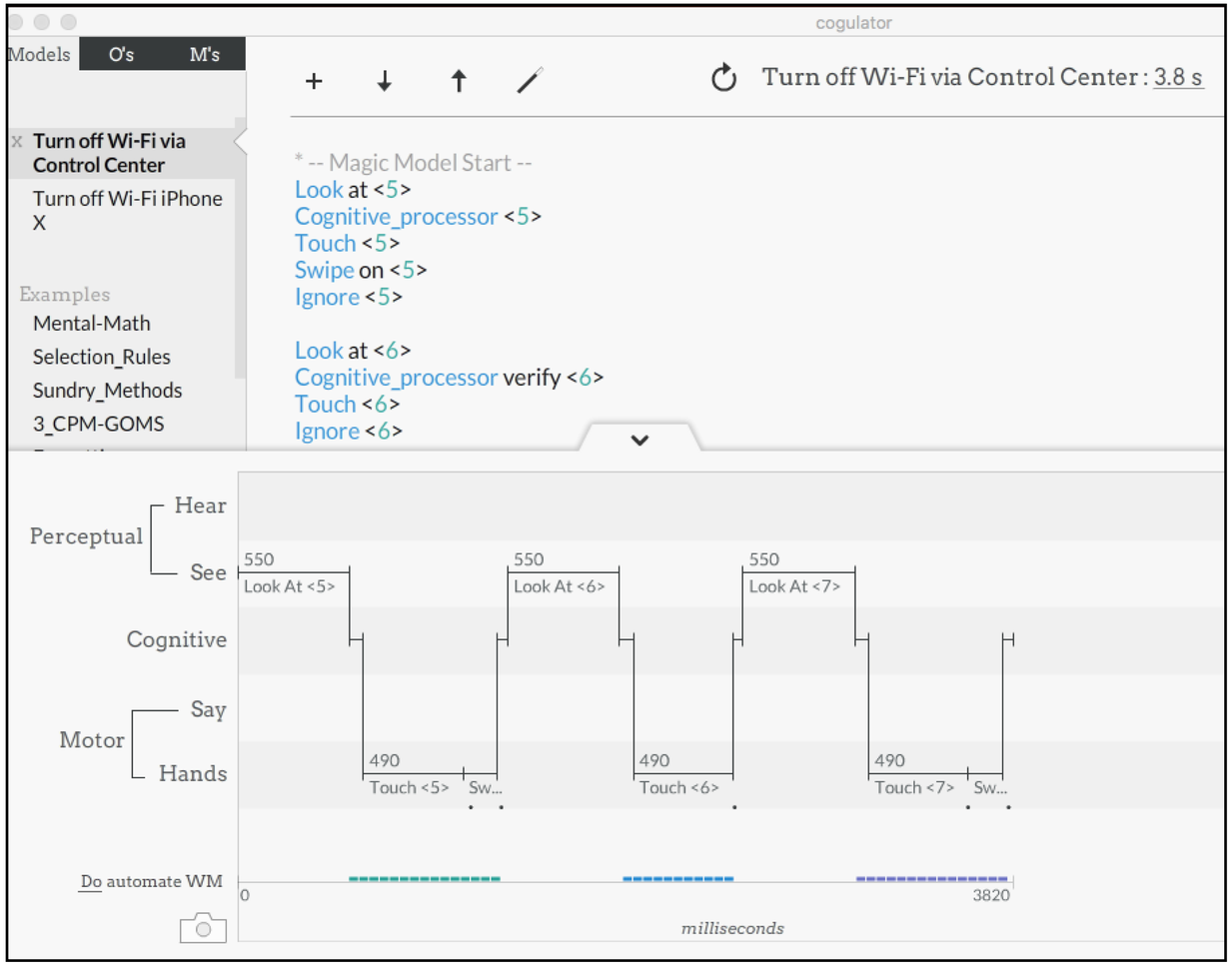
COGULATOR RESULTS [DEMO]

TURNING OFF WI-FI VIA THE SETTINGS APP



7.3 secs (vs. my 6.74 secs)

TURNING OFF WI-FI VIA CONTROL CENTER



3.8 secs (vs. my 2.69 secs)

NN/g

QUESTIONS?

- You may now ask questions from this section
- You may type your question into Zoom Chat
- We may have to cover some questions as follow-ups after the seminar

Heuristic Evaluation

WHAT IS HEURISTIC EVALUATION?

- "A method for finding the usability problems in a user interface design… [by]… having a small set of expert evaluators examine the interface and judge its compliance with recognized usability principles (the 'heuristics')."
  -Jakob Nielsen

- Similar to code inspection, in that experts examine the code for potential defects (instead of only relying on testing the code)

- Requires skills: knowledge of heuristics, familiarity with good UX design practices

Nielsen's Original 10 Heuristics

1. Visibility of system status
2. Match between system and the real world
3. User control and freedom
4. Consistency and standards
5. Error prevention
6. Recognition rather than recall
7. Flexibility and efficiency of use
8. Aesthetic and minimalist design
9. Help users recognize, diagnose, and recover from errors
10. Help and documentation

WHEN TO USE HEURISTIC EVALUATION

- When you have multiple experts available who are knowledgable about key UX principles (heuristics)
- If you want to quickly identify several potential UX issues throughout your product without necessarily testing with users
- When you want to explore any part of the experience
- Can provide severity rating for issues found, if desired:

*Severity rating definitions:
0 = I don't agree that this is a usability problem at all
1 = Cosmetic problem only: need not be fixed unless extra time is available on project
2 = Minor usability problem: fixing this should be given low priority
3 = Major usability problem: important to fix, so should be given high priority
4 = Usability catastrophe: imperative to fix this before product can be released

Pros

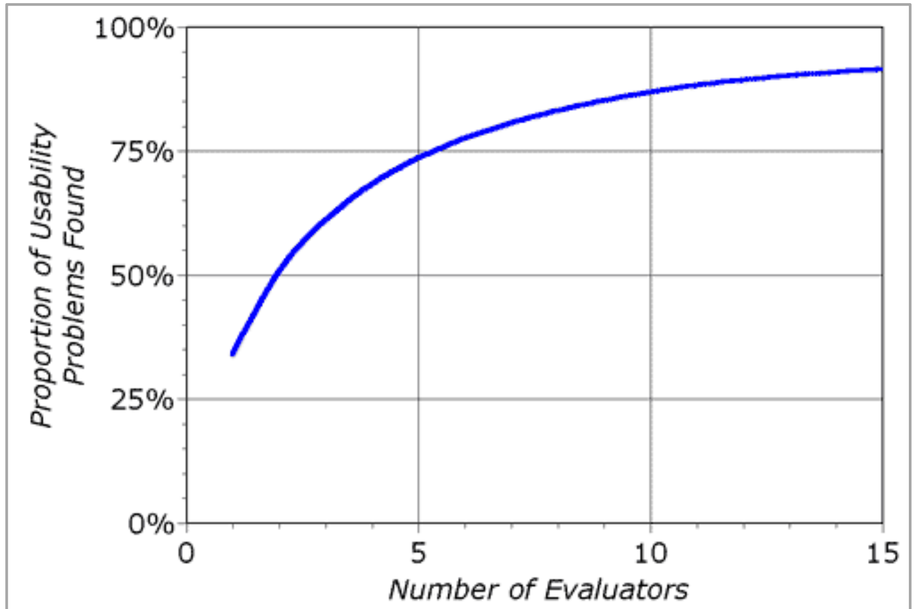
- Can be done quickly
- Allows you to examine any and all parts of the experience, and not limited to specific tasks
- Potentially finds 30-90% of the usability problems that could be found testing with users*

Cons

- Evaluators should be experts
- Need multiple evaluators to be effective

*This claim and number has been debated fiercely, but the numbers here were derived by combining results from Nielsen & Molich (1990) and Hollingsed, T and Novick, D (2007)

HOW TO DO AN HEURISTIC EVALUATION



Source: How to Conduct a Heuristic Evaluation

- Gather 3-5 usability experts knowledgeable about user experience design principles or heuristics to be used (see graph)
- Determine the scope of the evaluation in the interface and clarify the user types and capabilities
- Each expert reviews the interface alone (possibly with an observer); they should go through it quickly at first, and then more carefully next. They may go through multiple times, identifying usability issues as they do
- An observer can capture the experts' comments as they go through, if this resource is available. The observer should be familiar with the interface and the domain, so they can assist the expert in understanding its intent and users, if needed
- The output from each expert evaluator should be: (a) a list of usability issues, (b) the usability principle each issue violates, (c) details and rationale for why it is a potential issue, and (d) the severity of the issue, based on how frequently it will likely be encountered and the impact on the user experience
- Results from all expert evaluators are combined into a single report

PRINCIPLE

prin·ci·ple /ˈprinsəpəl/

noun:
1. a fundamental truth or proposition that serves as the foundation for a system of belief or behavior or for a chain of reasoning.

2. a fundamental source or basis of something.

HEURISTIC(S)

heu·ris·tic /hyo͝oˈristik/

adjective:
enabling a person to discover or learn something for themselves.

noun:
1. a heuristic process or method.
the study and use of heuristic technique

2. recognized usability principles (e.g., as in the “heuristics" in heuristic evaluation)

In today’s context, these essentially mean the same thing.

Principles are necessary to do a UX Assessment

- UX principles or heuristics (usability principles) should be based on known facts*
- Research and theory greatly inform principles
- Some are very domain-specific (e.g., in security, some users rely on a "trusted advisor" to make decisions or take action)
- Some are more very low-level or UI-specific (e.g., "drop-down choices should be mutually exclusive")
- Some are high-level or more abstract (e.g., "consistency")

*A possible exception: Some Design Guidelines include "Principles," which may simply convey a point of view on what the experience should be like "in principle." Ideally these are based on research, but in practice, this is not always the case. We will learn about Design Guidelines later in the course.

NIELSEN'S 10 USABILITY HEURISTICS*

*Nielsen, J (1994): *Usability Engineering*

1. Visibility of system status
The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.

2. Match between system and the real world
The system should speak the users' language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.

3. User control and freedom
Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support undo and redo.

4. Consistency and standards
Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.

5. Error prevention
Even better than good error messages is a careful design which prevents a problem from occurring in the first place.

6. Recognition rather than recall
Make objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.

7. Flexibility and efficiency of use
Accelerators -- unseen by the novice user -- may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.

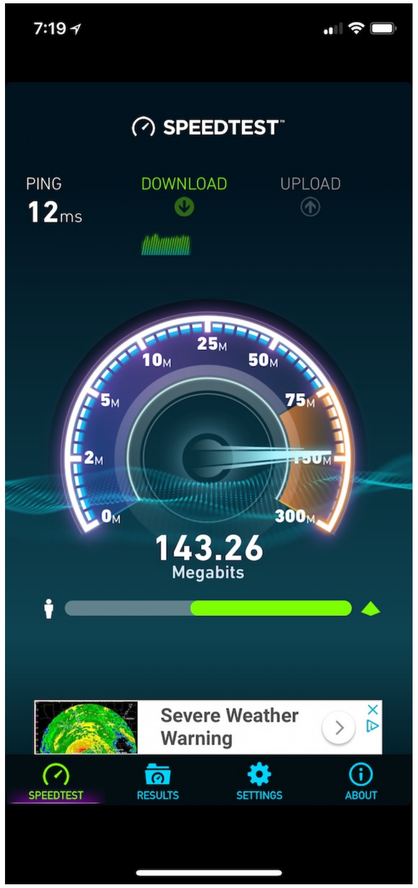
8. Aesthetic and minimalist design
Dialogues should not contain information which is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.

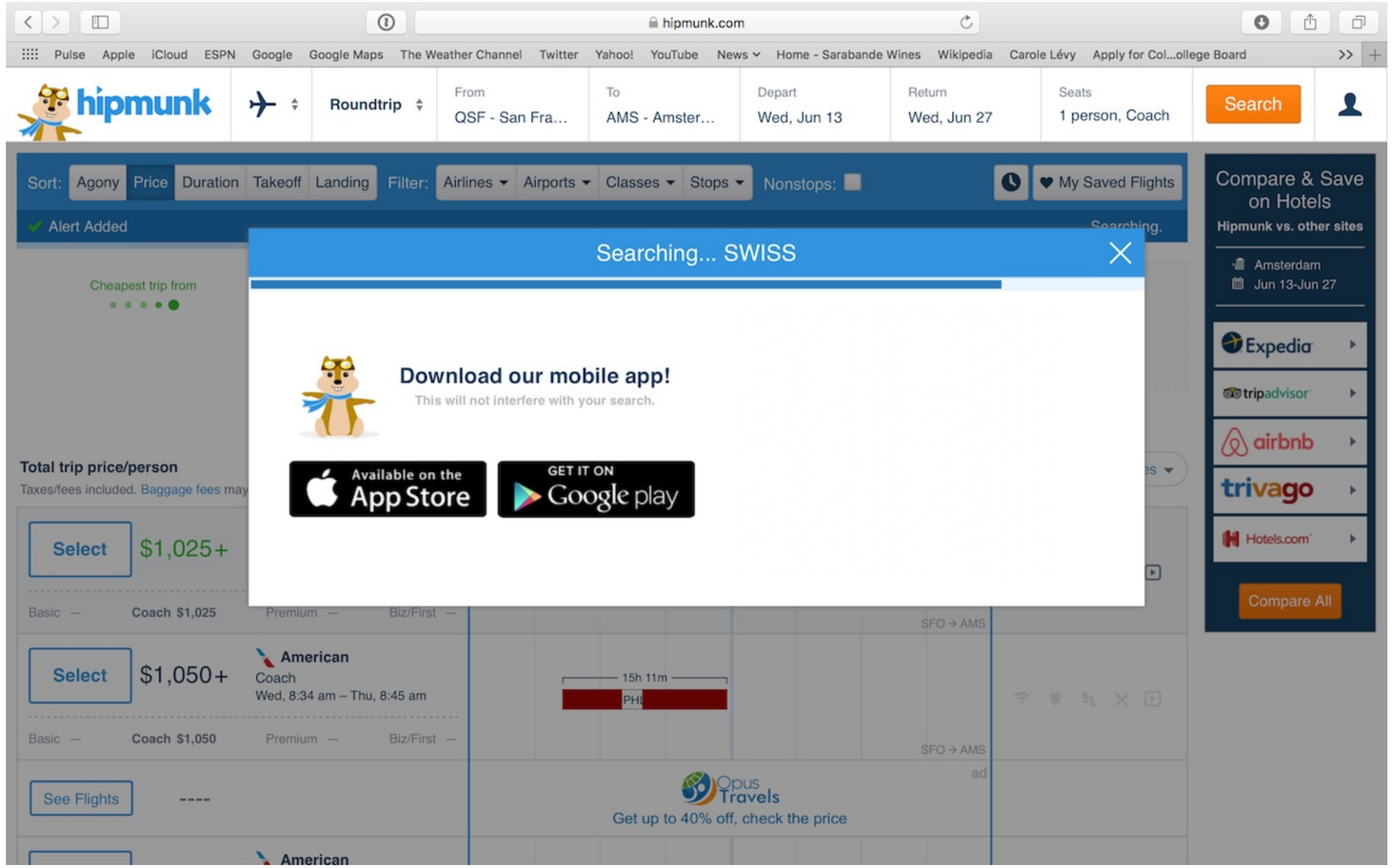
9. Help users recognize, diagnose, and recover from errors
Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.

10. Help and documentation
Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large.

NIELSEN'S 10 HEURISTICS

1. **Visibility of system status:** The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.



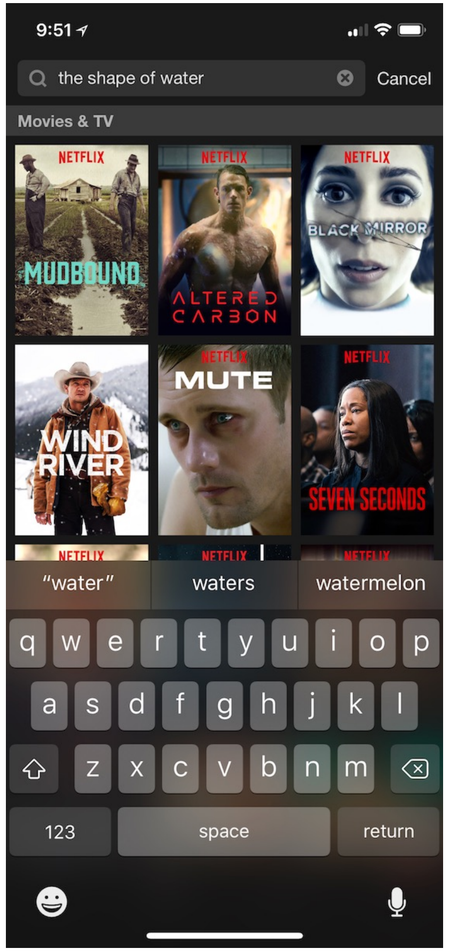


Good visibility (Speedtest and Freshbooks)

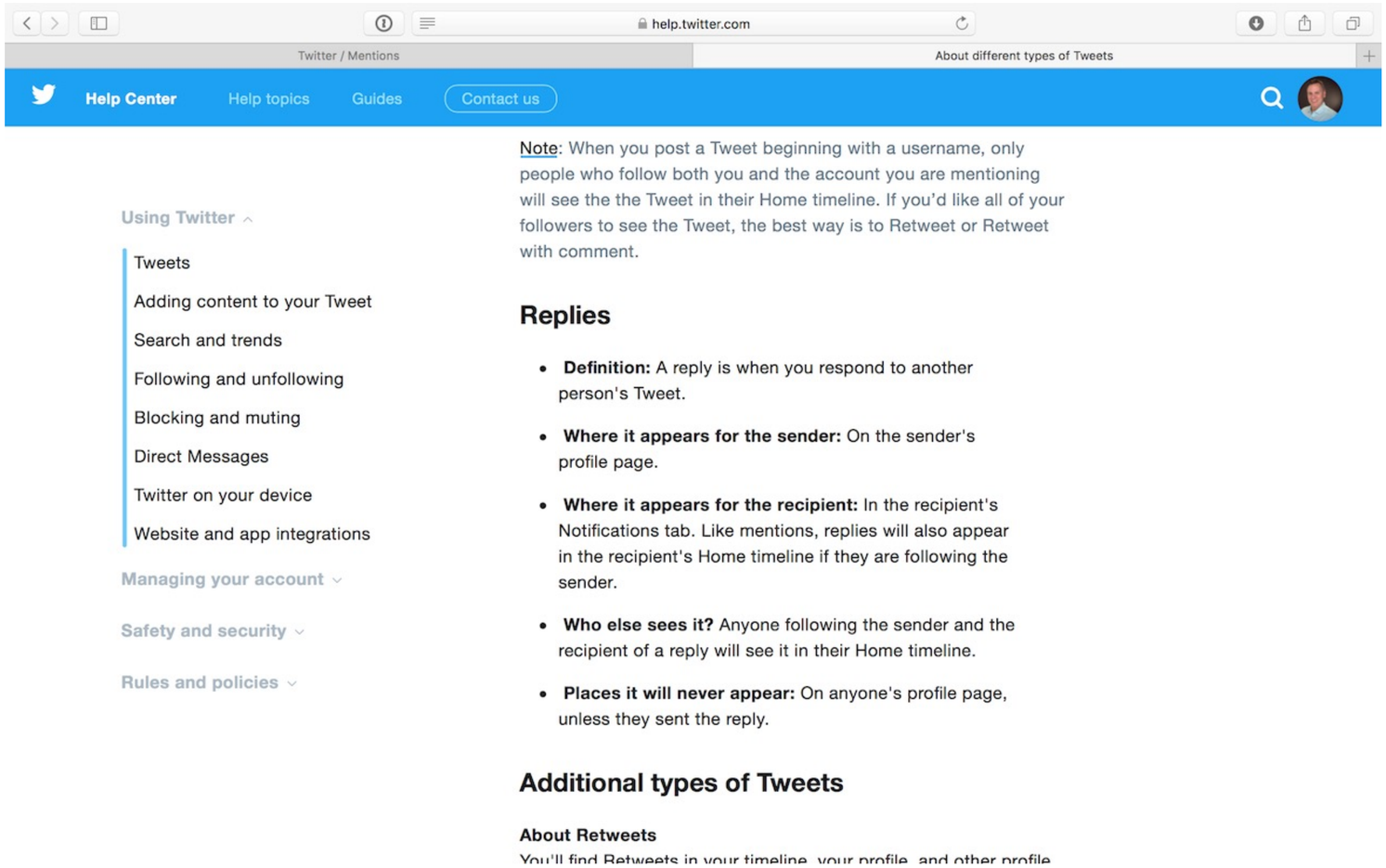
Reasonably good visibility in Hipmunk

NIELSEN'S 10 HEURISTICS

1. **Visibility of system status:** The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.



Poor: Netflix search results when there are 0 title matches



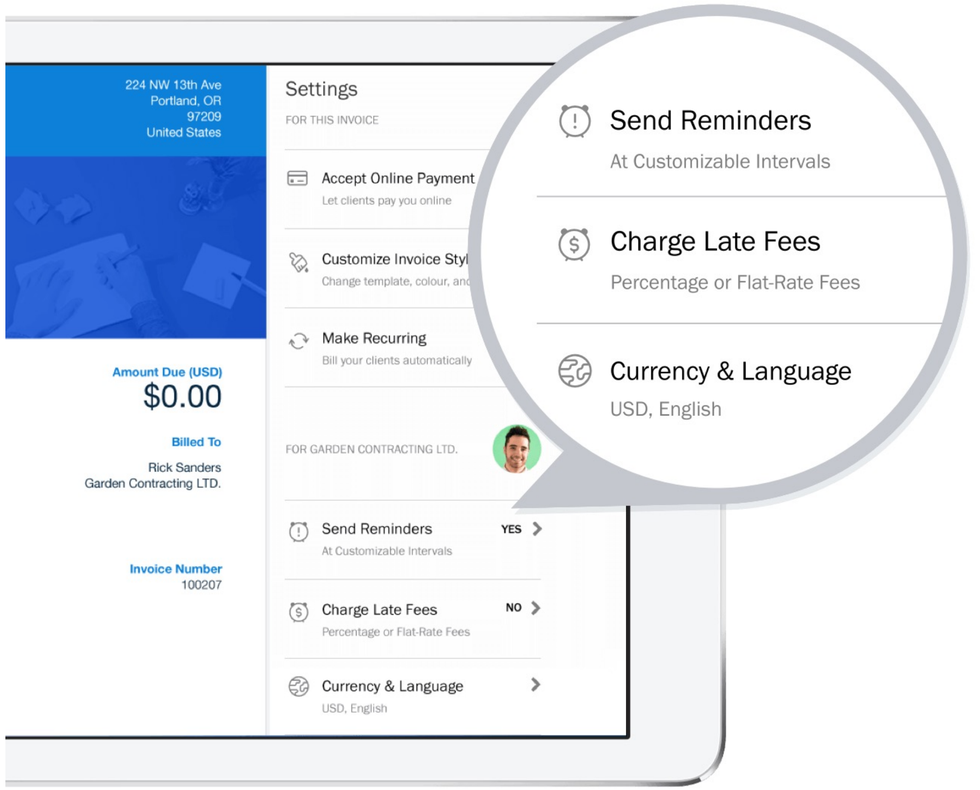
Poor visibility: How Twitter handles what is public

NIELSEN'S 10 HEURISTICS

2. Match between system and the real world: The system should speak the users' language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.

Good match: Shake Shack - this is what you're getting!

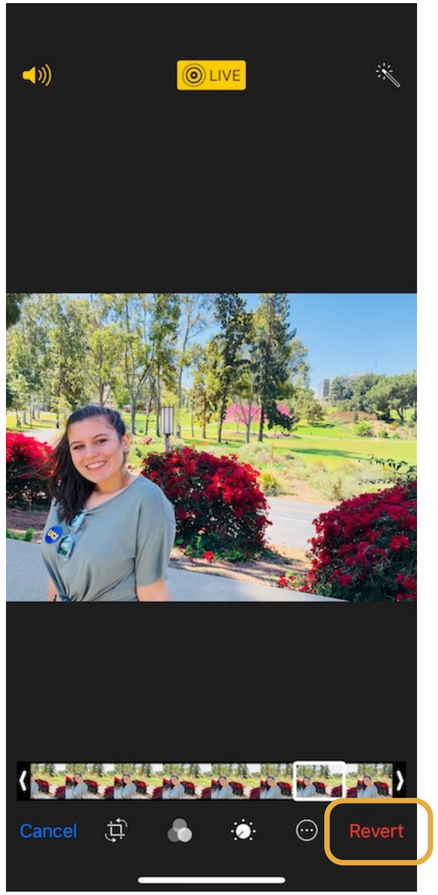
Good match: a Level app



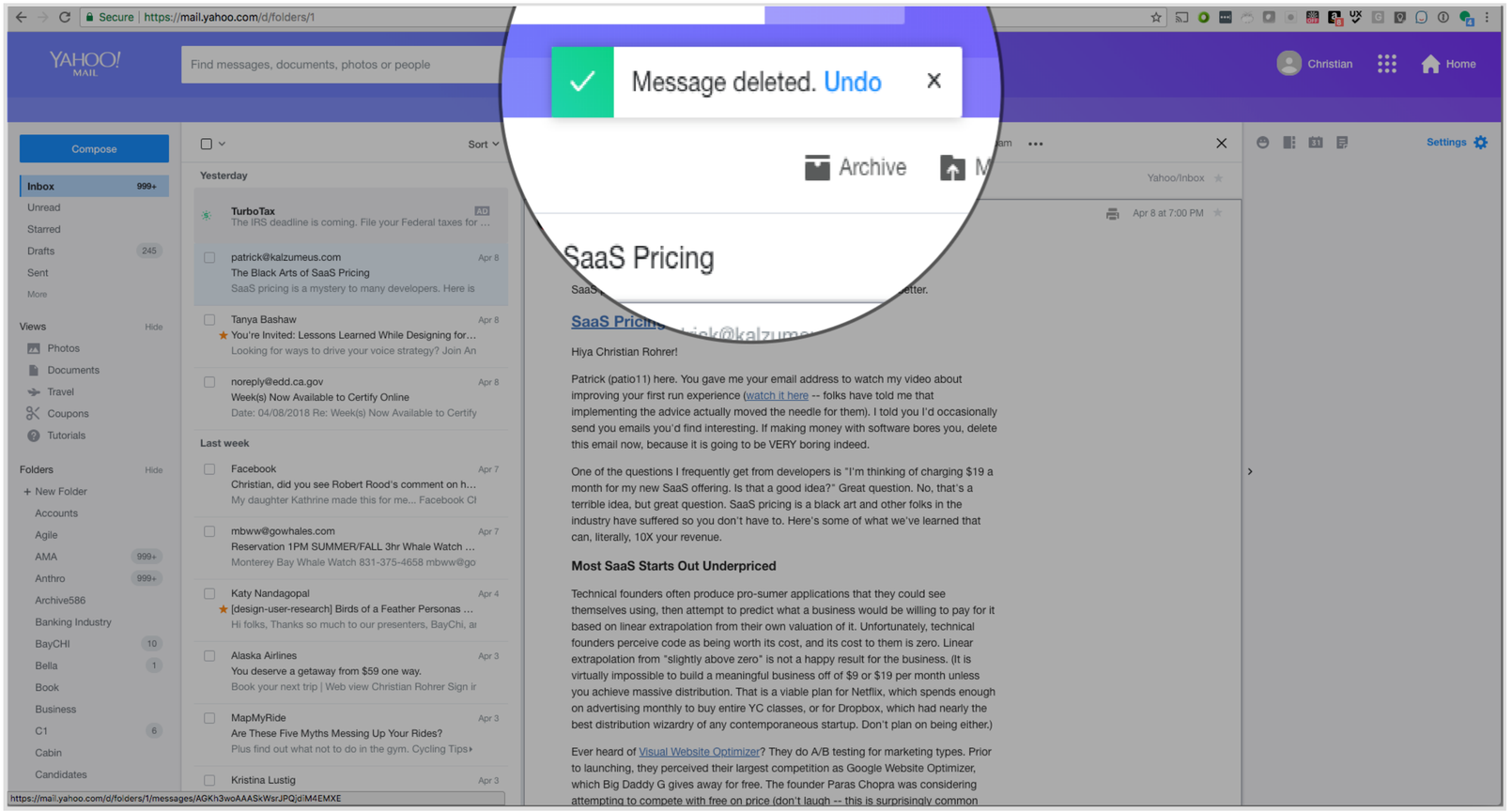
Good match: FreshBooks makes accounting and invoicing easy by using everyday language familiar to small business users, rather than accounting terms.

NIELSEN'S 10 HEURISTICS

3. User control and freedom: Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support undo and redo.



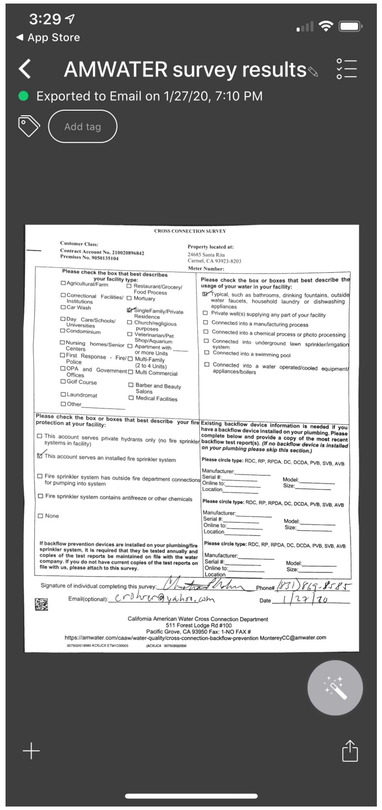
Good Control: Ability to revert to original photo

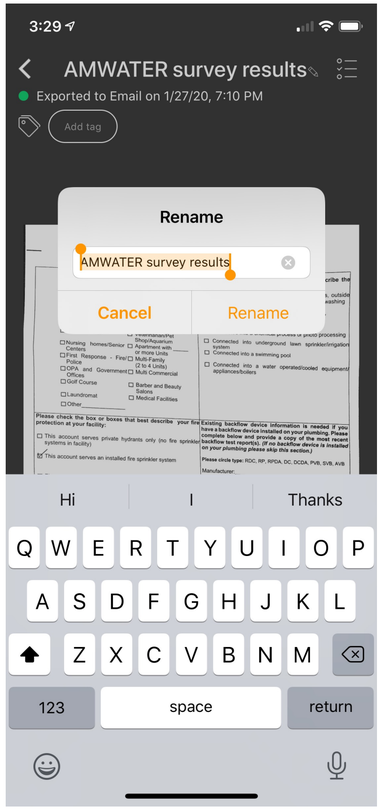


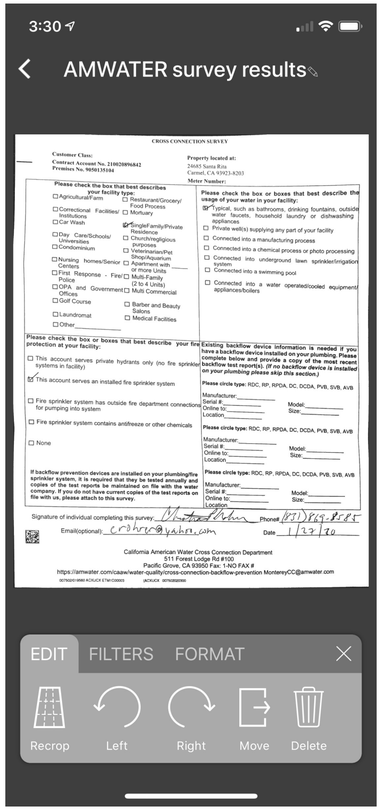
Good Control: Yahoo! Mail's undo

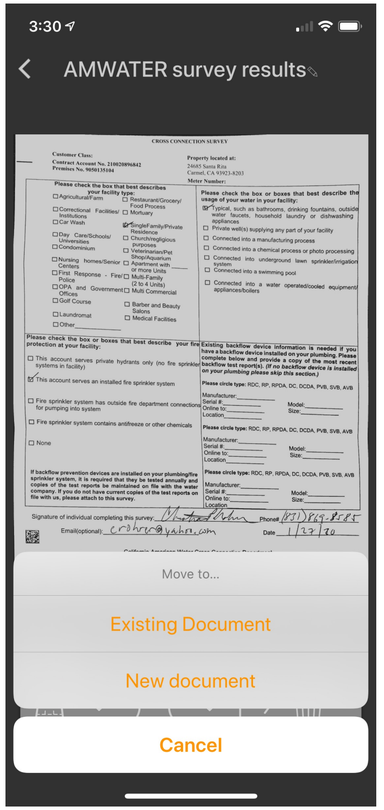
NIELSEN'S 10 HEURISTICS

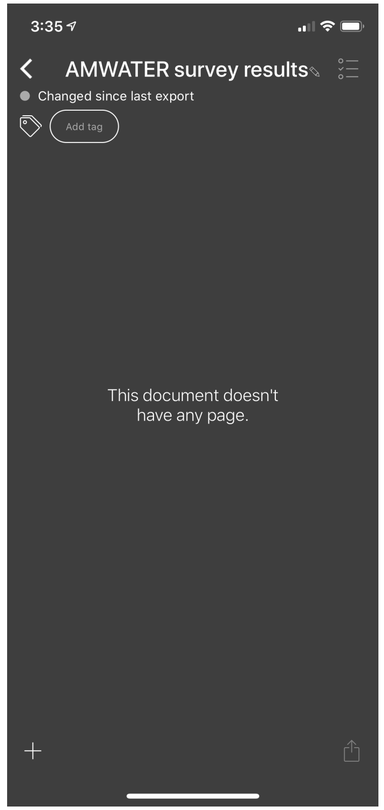
3. User control and freedom: Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support undo and redo.







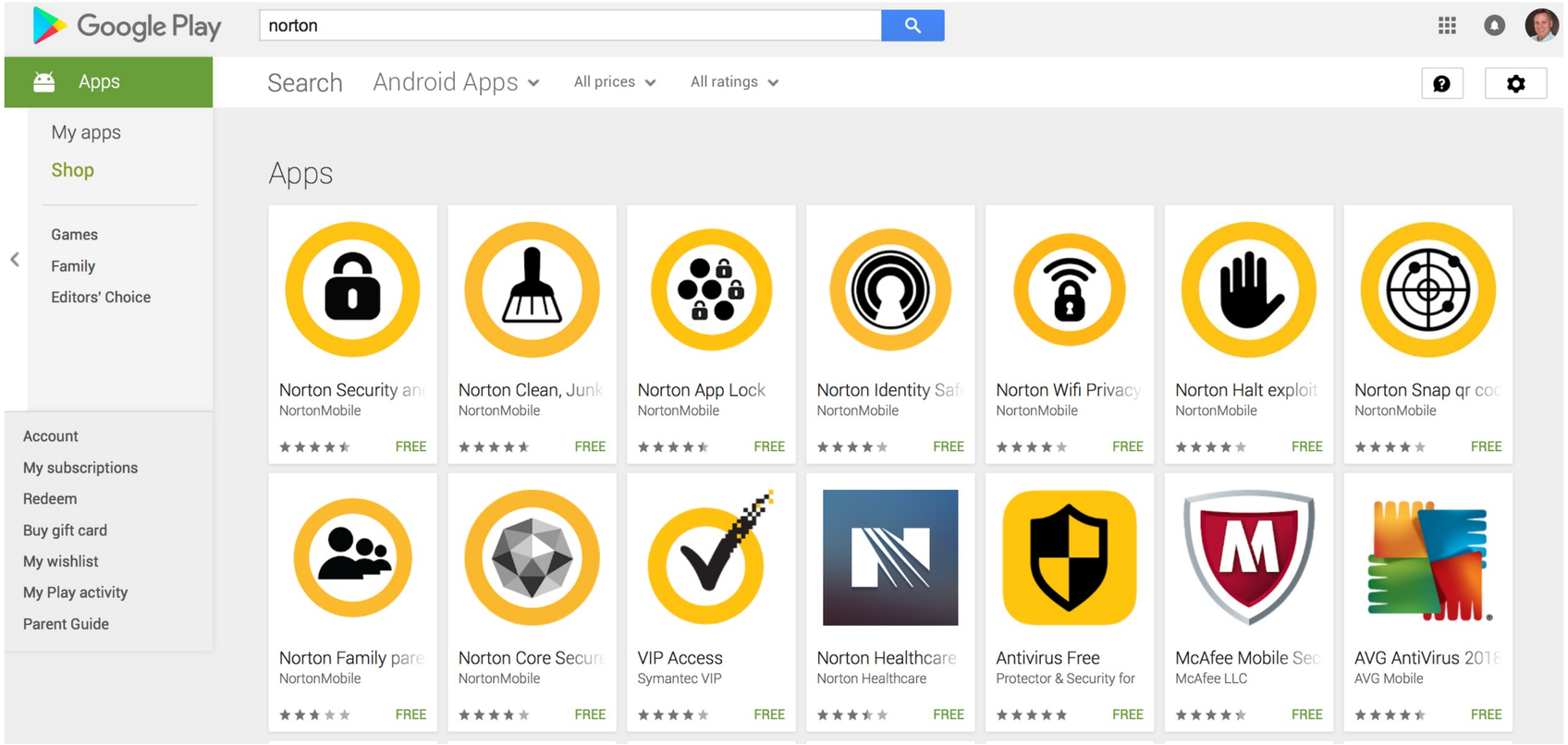




OK Control: Genius Scan's has the freedom to rename and move documents, but you can't undo deleting

NIELSEN'S 10 HEURISTICS

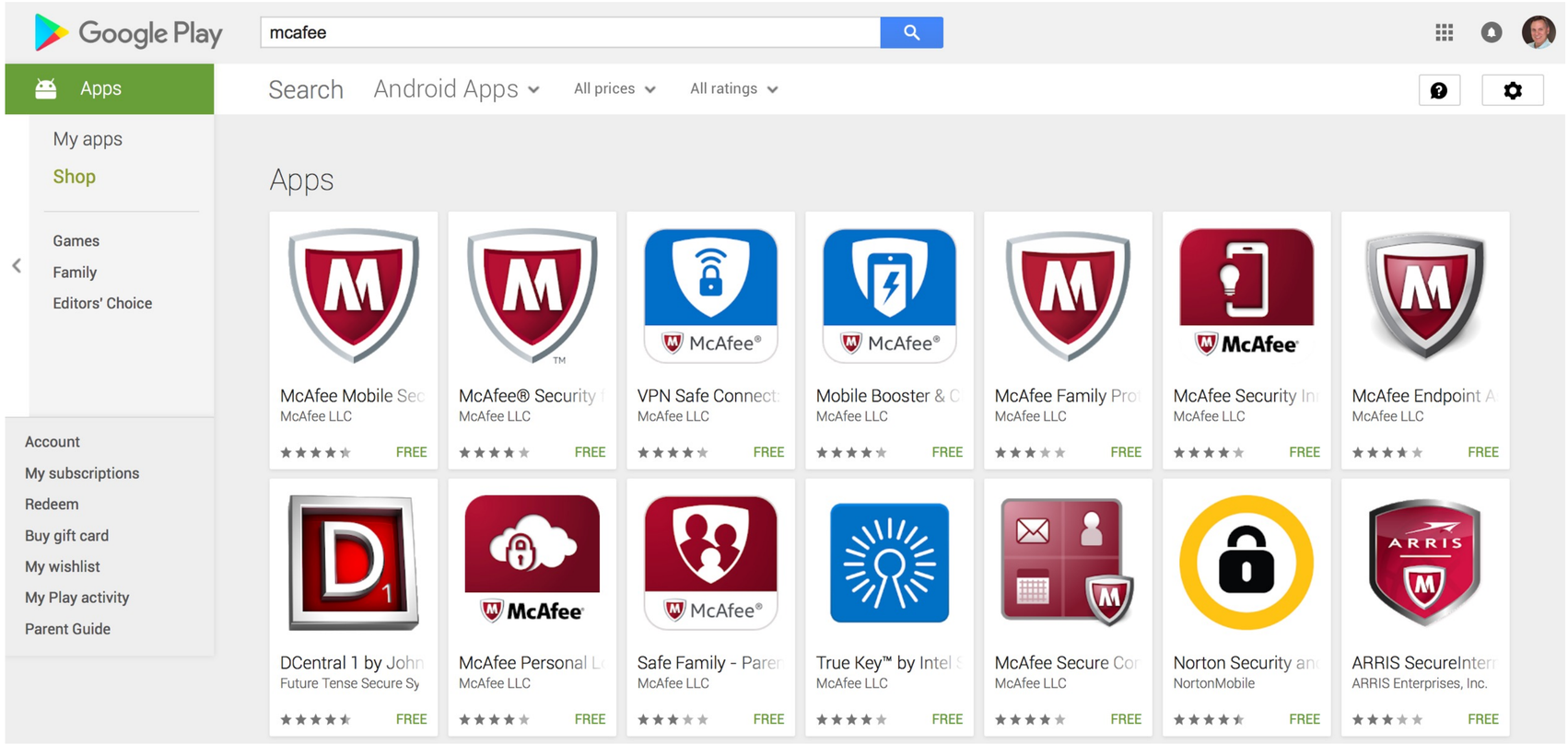
4. Consistency and standards: Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.



Good consistency and standards: Norton branding

NIELSEN'S 10 HEURISTICS

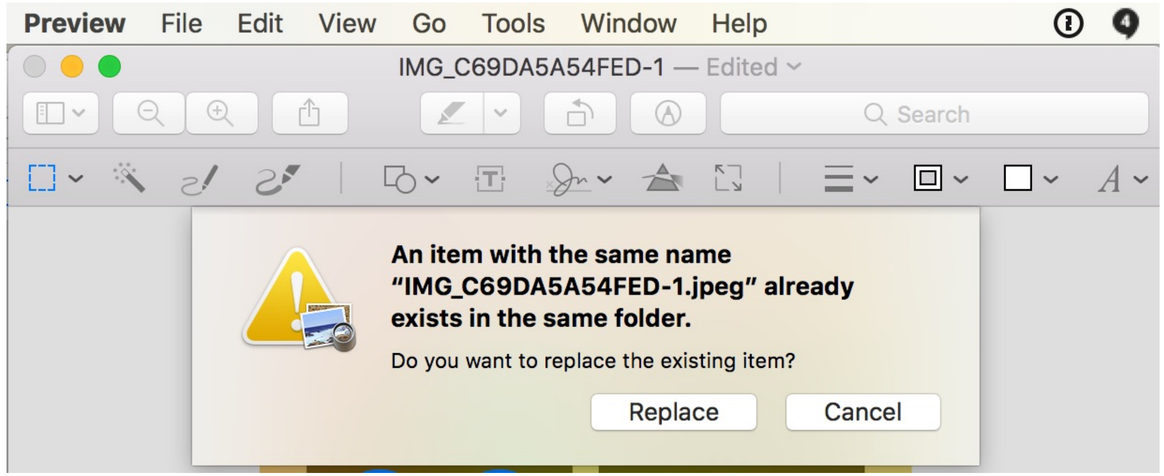
4. Consistency and standards: Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.



Poor consistency and standards: McAfee branding and poor product naming

NIELSEN'S 10 HEURISTICS

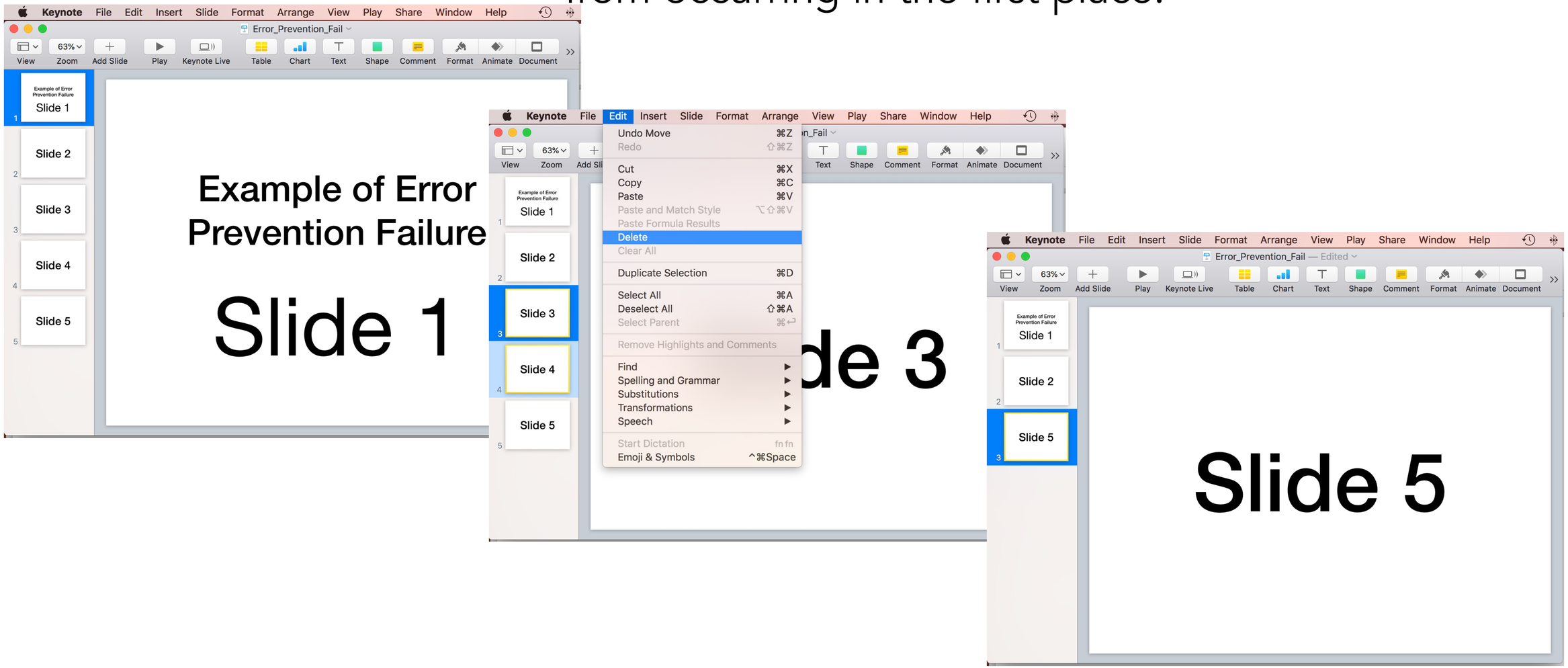
5. Error prevention: Even better than good error messages is a careful design which prevents a problem from occurring in the first place.



Good error prevention: Mac OS X: Prevents you from overwriting existing files (Even better if it gave the user the option to rename or move the file in this dialog.)

NIELSEN'S 10 HEURISTICS

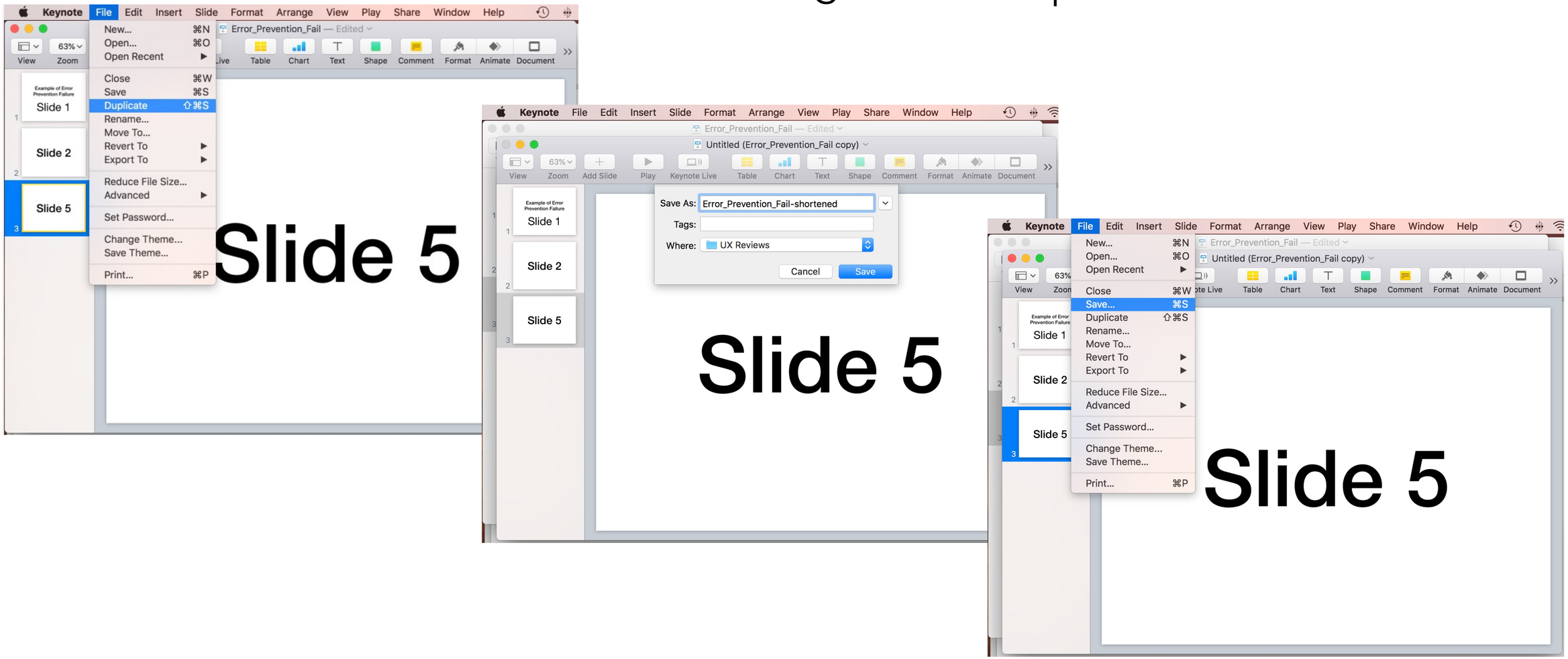
5. Error prevention: Even better than good error messages is a careful design which prevents a problem from occurring in the first place.



Poor error prevention: Mac OS X AutoSave feature makes it too easy to lose application data

NIELSEN'S 10 HEURISTICS

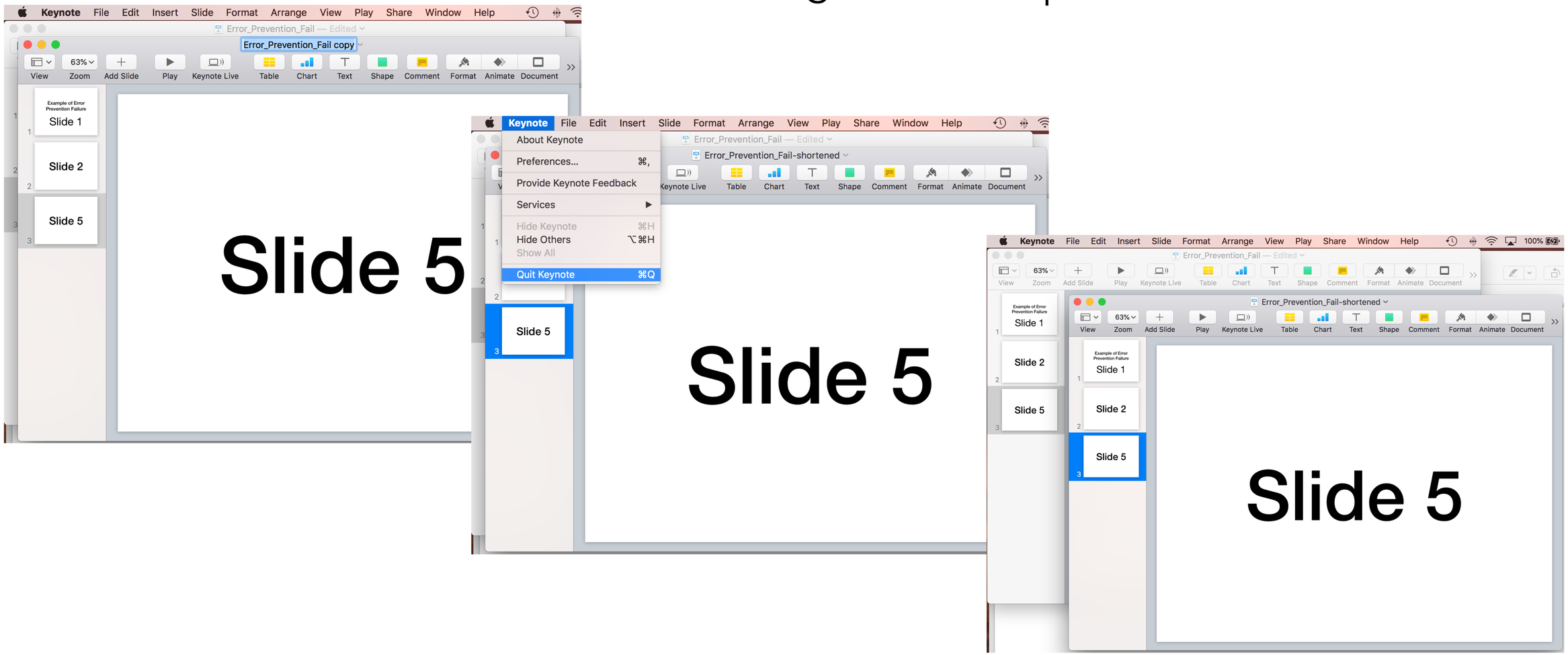
5. Error prevention: Even better than good error messages is a careful design which prevents a problem from occurring in the first place.



Poor error prevention: Mac OS X AutoSave feature makes it too easy to lose application data

NIELSEN'S 10 HEURISTICS

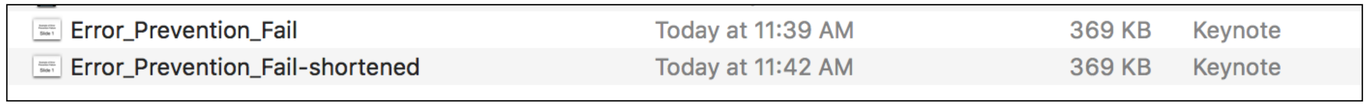
5. Error prevention: Even better than good error messages is a careful design which prevents a problem from occurring in the first place.

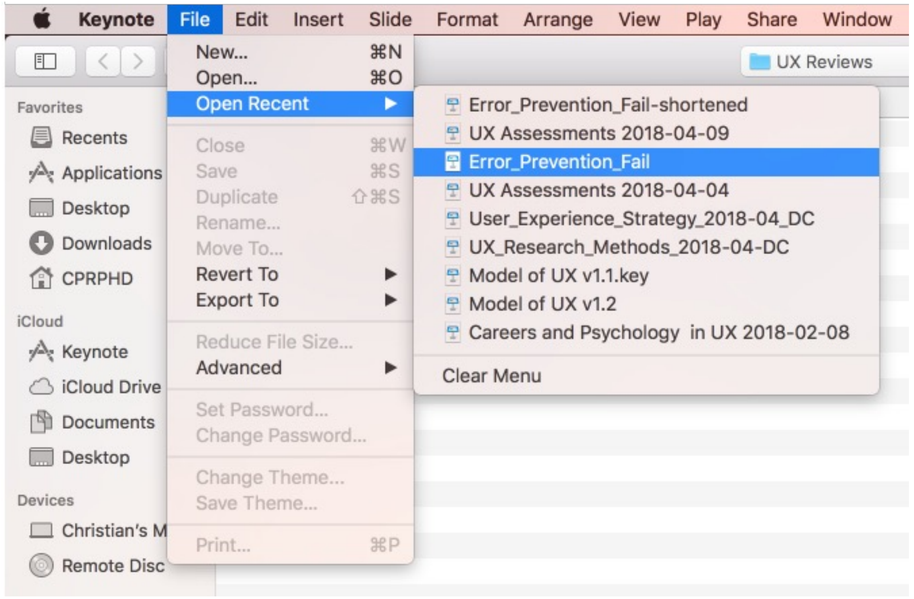


Poor error prevention: Mac OS X AutoSave feature makes it too easy to lose application data

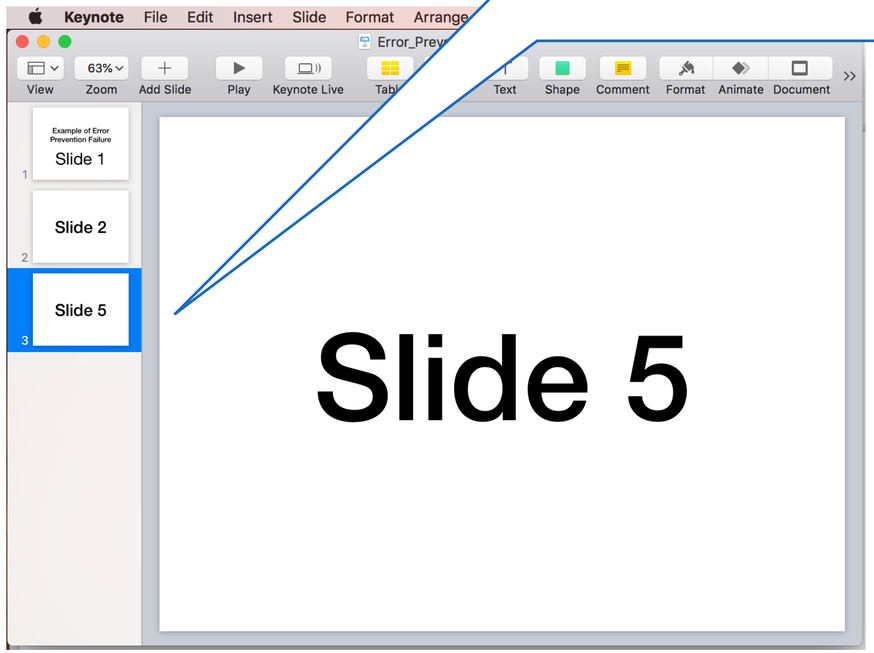
NIELSEN'S 10 HEURISTICS

5. Error prevention: Even better than good error messages is a careful design which prevents a problem from occurring in the first place.





"Where are original slides 3 and 4???"

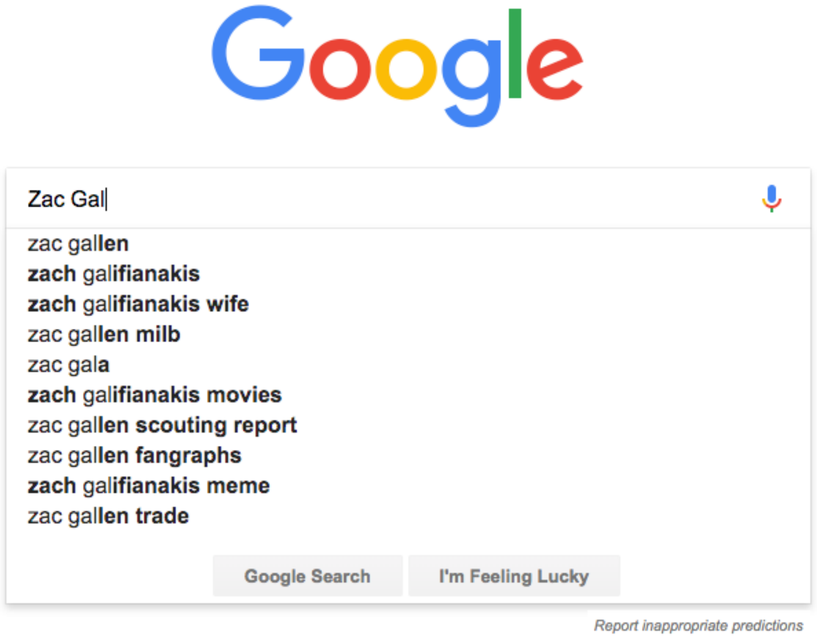


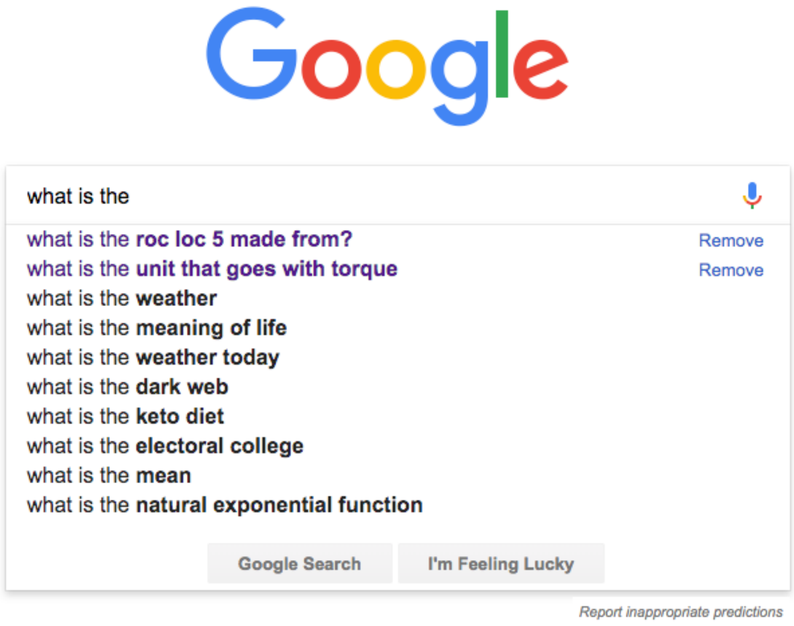
The problem is that AutoSave silently assumed changes made before the Duplication were meant to be kept.

Poor error prevention: Mac OS X AutoSave feature makes it too easy to lose application data

NIELSEN'S 10 HEURISTICS

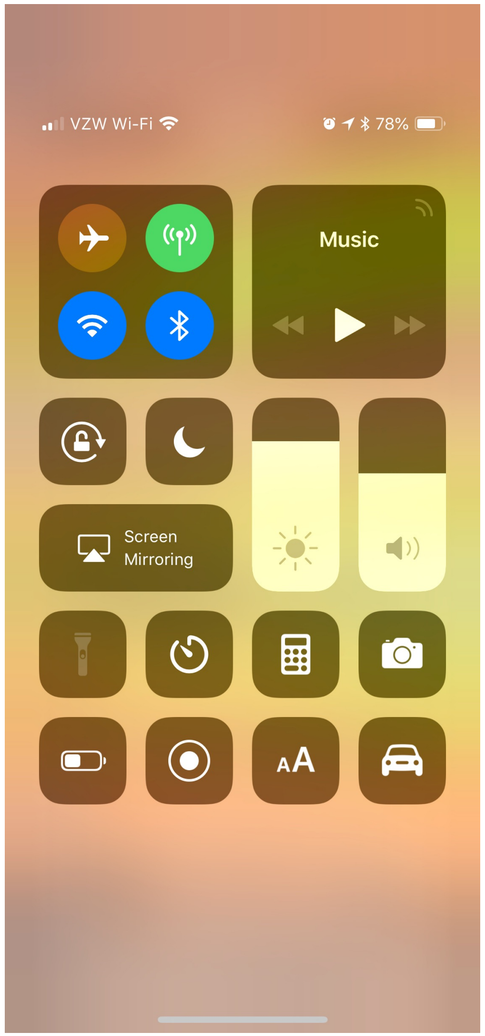
6. Recognition rather than recall: Make objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.





Good Recognition from Google Suggest: Helps avoid spelling AND search term recall

NIELSEN'S 10 HEURISTICS



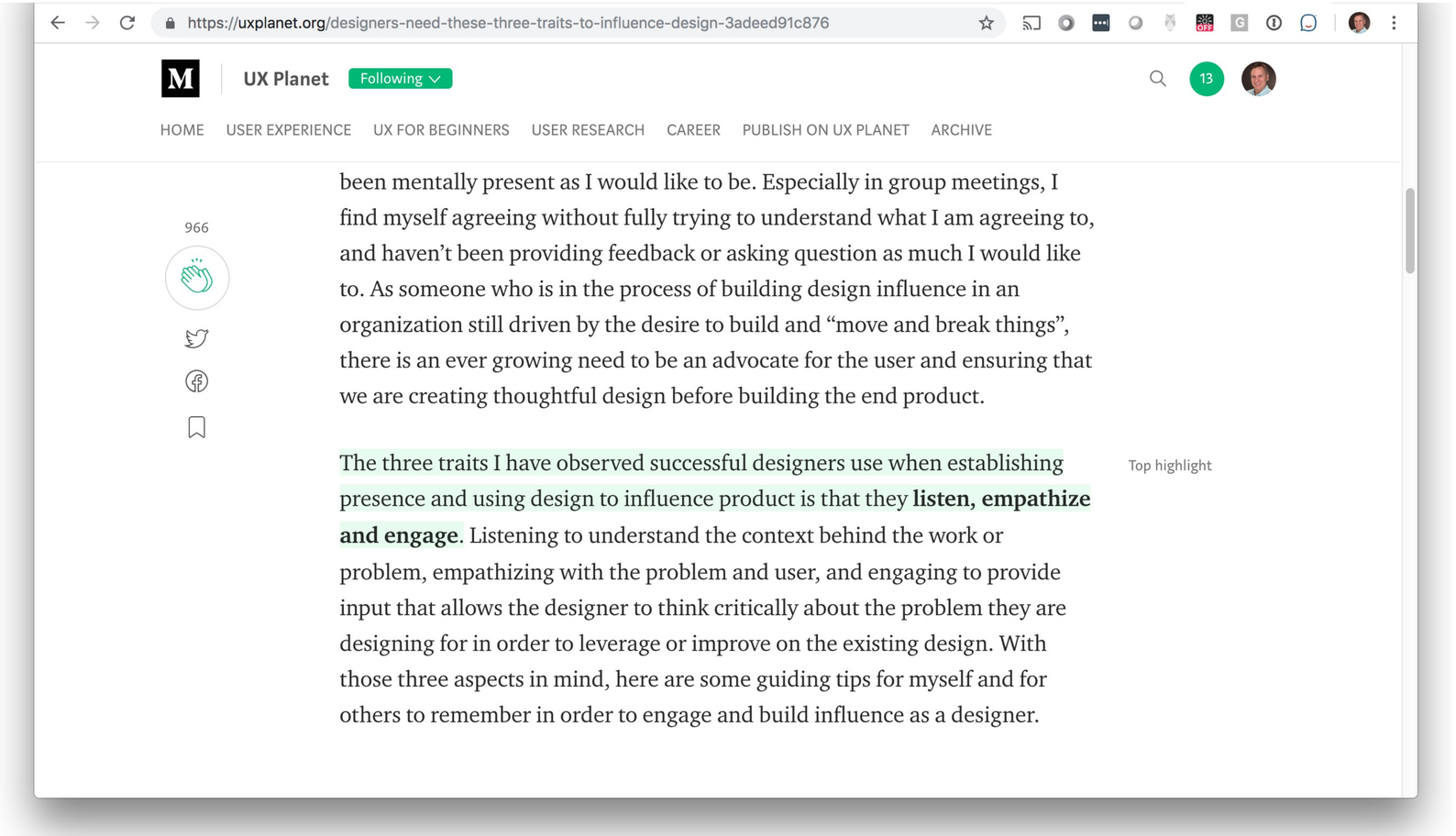
7. Flexibility and efficiency of use: Accelerators -- unseen by the novice user -- may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.

Good efficiency: Control Center speeds up interaction of common tasks

NIELSEN'S 10 HEURISTICS

8. Aesthetic and minimalist design: Dialogues should not contain information which is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.

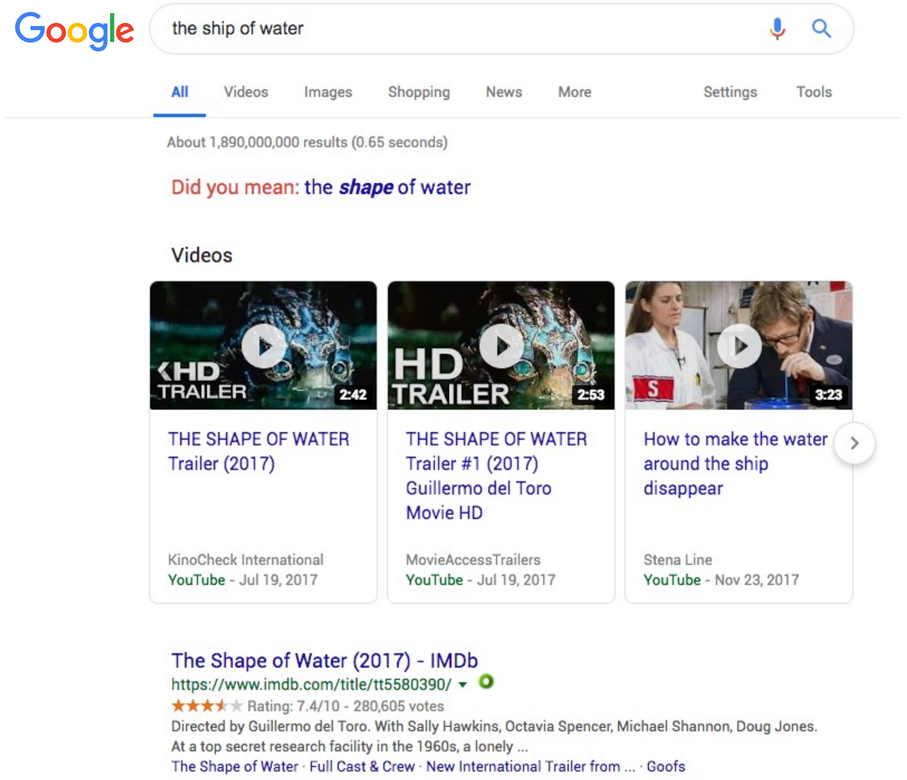
Good Aesthetic: Medium.com puts almost nothing in the way of your reading or writing



NIELSEN'S 10 HEURISTICS

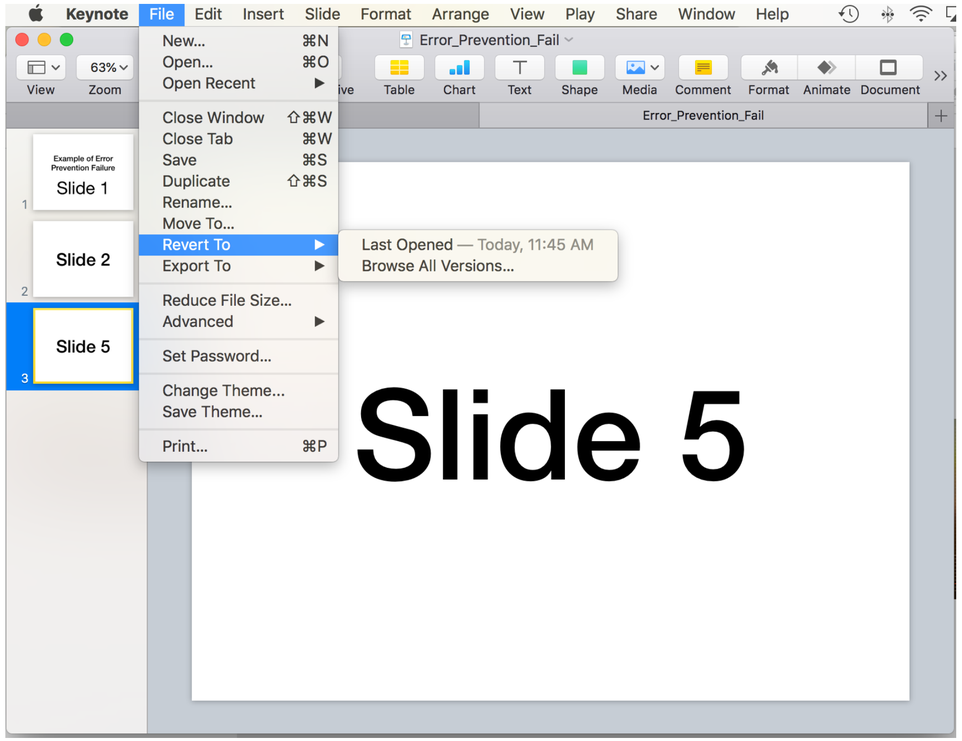
9. Help users recognize, diagnose, and recover from errors: Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.

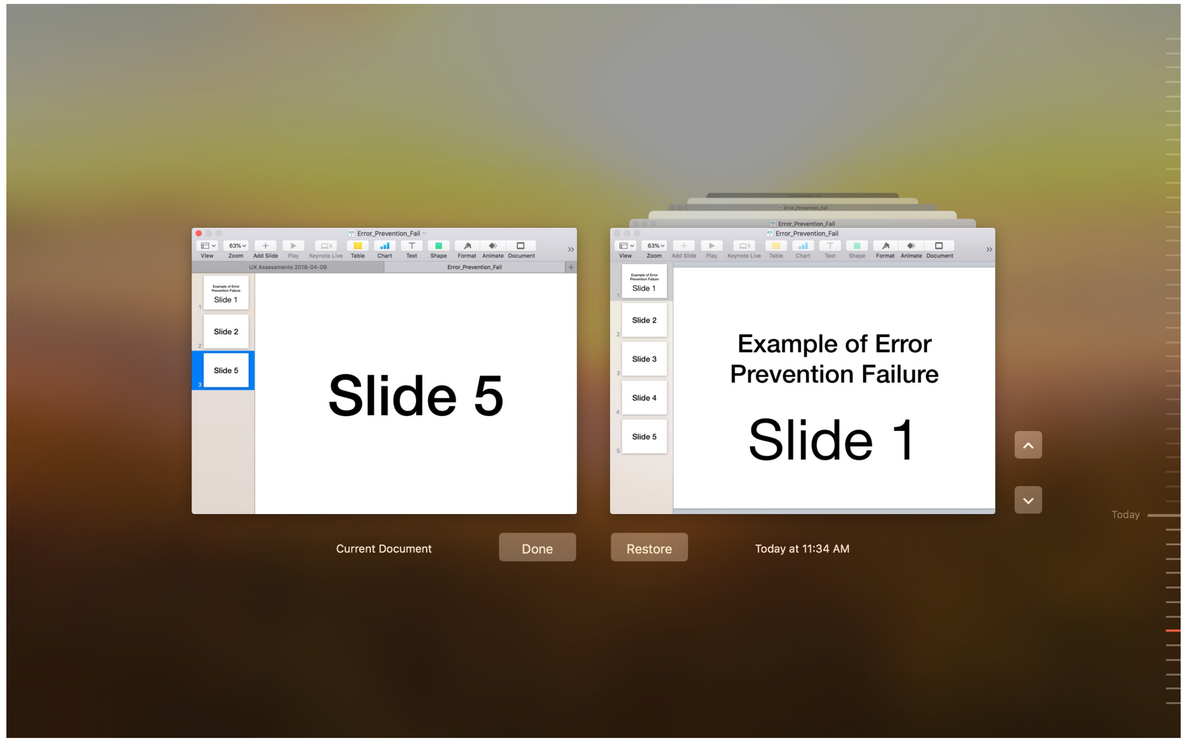
Good error recognition & recovery: Google realizes I probably mis-typed the word "shape" as "ship" and offers me a way to recover (and actually shows me results as if I typed it as "shape")



NIELSEN'S 10 HEURISTICS

9. Help users recognize, diagnose, and recover from errors: Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.



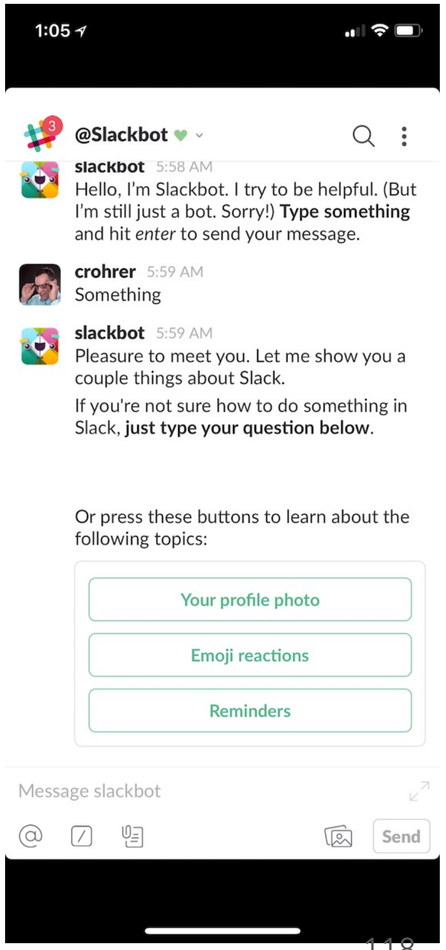


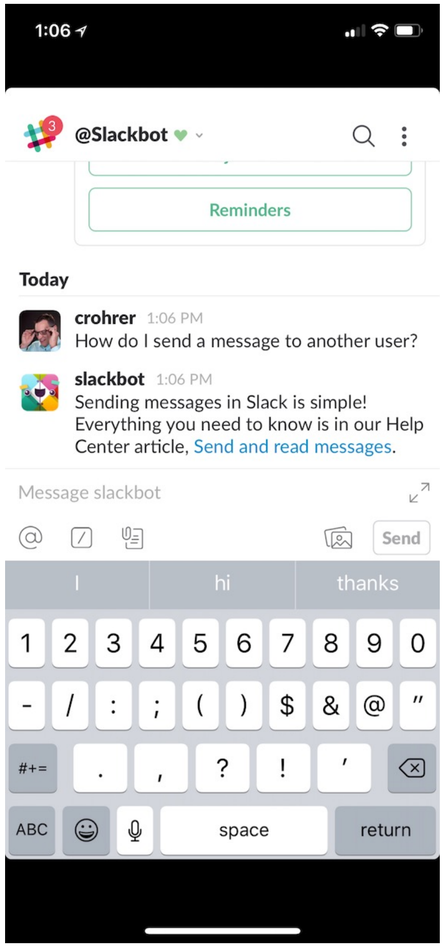
Good error recovery: Keynote lets you revert back to an older version of your presentation (File -> Revert To), should you screw up, as shown in Error Prevention.

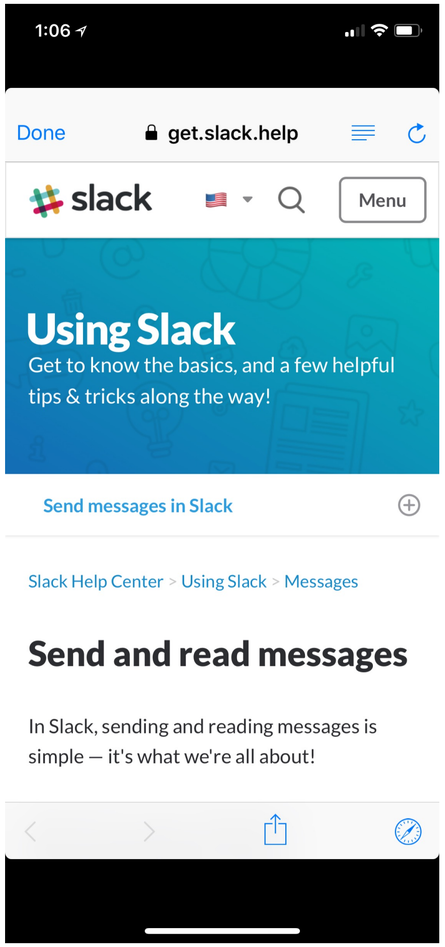
NIELSEN'S 10 HEURISTICS

10. Help and documentation: Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large.

Good Help and Documentation: Slack does a great job of helping new users, using a chatbot interface, FAQs and a lot of contextual help.







NN/g

QUESTIONS?

- You may now ask questions from this section
- You may type your question into Zoom Chat
- We may have to cover some questions as follow-ups after the seminar

NORMAN'S 7 FUNDAMENTAL PRINCIPLES*

1. **Discoverability**. It is possible to determine what actions are possible and the current state of the device.

2. **Feedback**. There is full and continuous information about the results of actions and the current state of the product or service. After an action has been executed, it is easy to determine the new state.

3. **Conceptual model**. The design projects all the information needed to create a good conceptual model of the system, leading to understanding and a feeling of control. The conceptual model enhances both discoverability and evaluation of results.

4. **Affordances**. The proper affordances (capabilities)exist to make the desired actions possible.

5. **Signifiers**. Effective use of signifiers ensures discoverability and that the feedback is well communicated and intelligible.

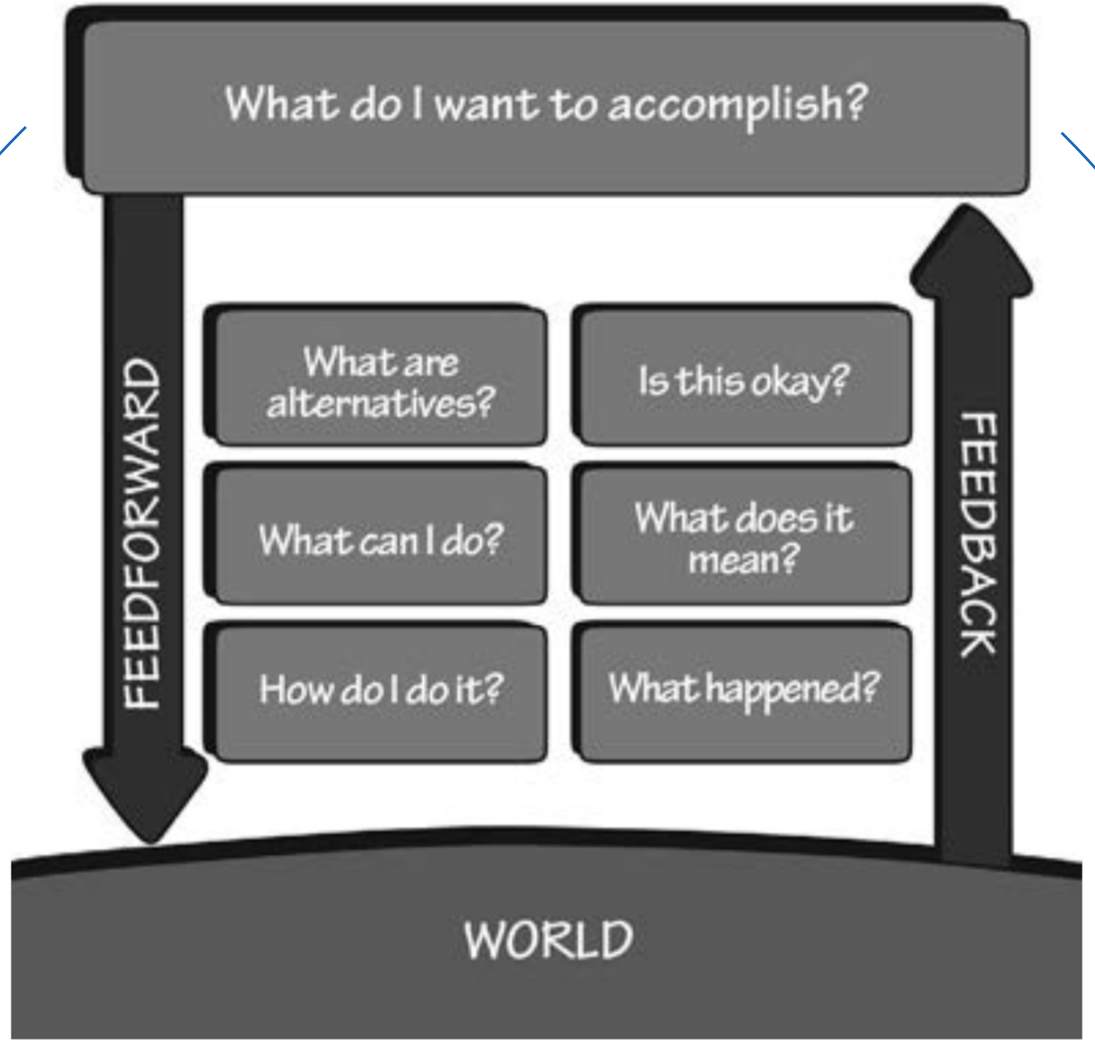
6. **Mappings**. The relationship between controls and their actions follows the principles of good mapping, enhanced as much as possible through spatial layout and temporal contiguity.

7. **Constraints**. Providing physical, logical, semantic, and cultural constraints guides actions and eases interpretation.

*Norman, D. (1988 & 2013): *The Design of Everyday Things*

NORMAN'S PRINCIPLES AND THE SEVEN STAGES OF ACTION

Don Norman's principles are influenced by how he views the basic human action with the world:



The Gulf of Execution*

*Aided by:
- Signifiers
- Constraints
- Mappings
- Conceptual Models

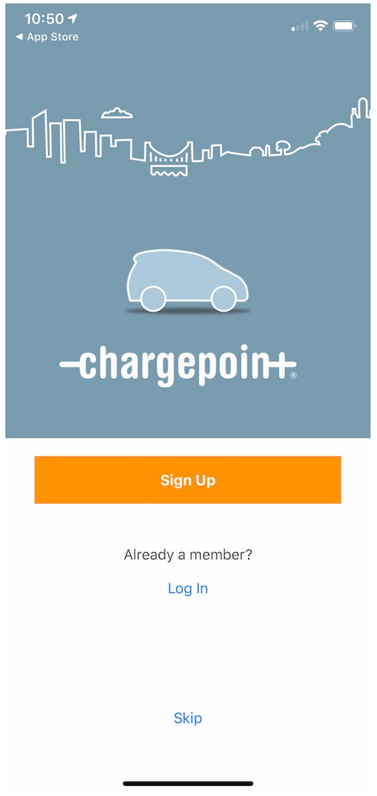
The Gulf of Evaluation**

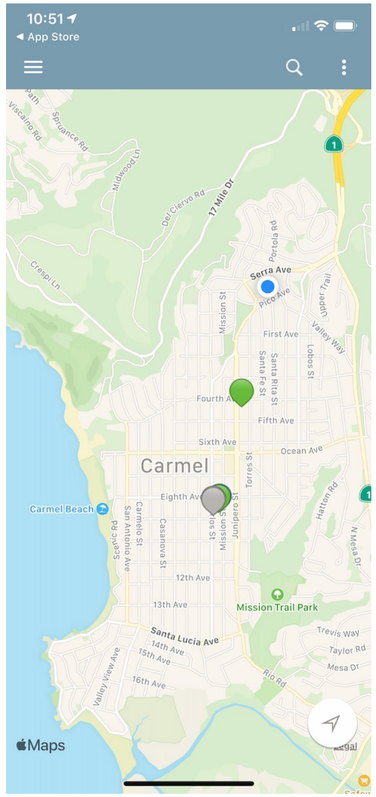
**Aided by:
- Feedback
- Conceptual Models

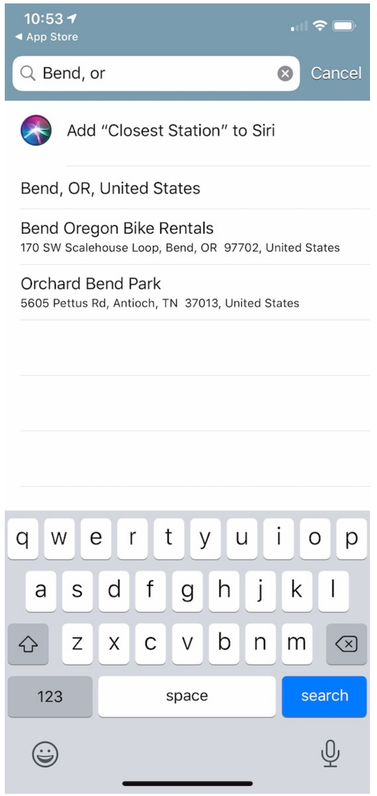
From Norman, D.A. (2013): *The Design of Everyday Things*

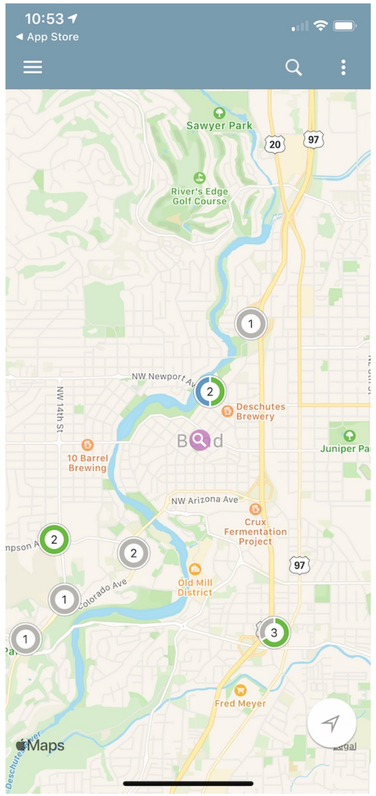
NORMAN'S PRINCIPLES

1. Discoverability. It is possible to determine what actions are possible and the current state of the device.





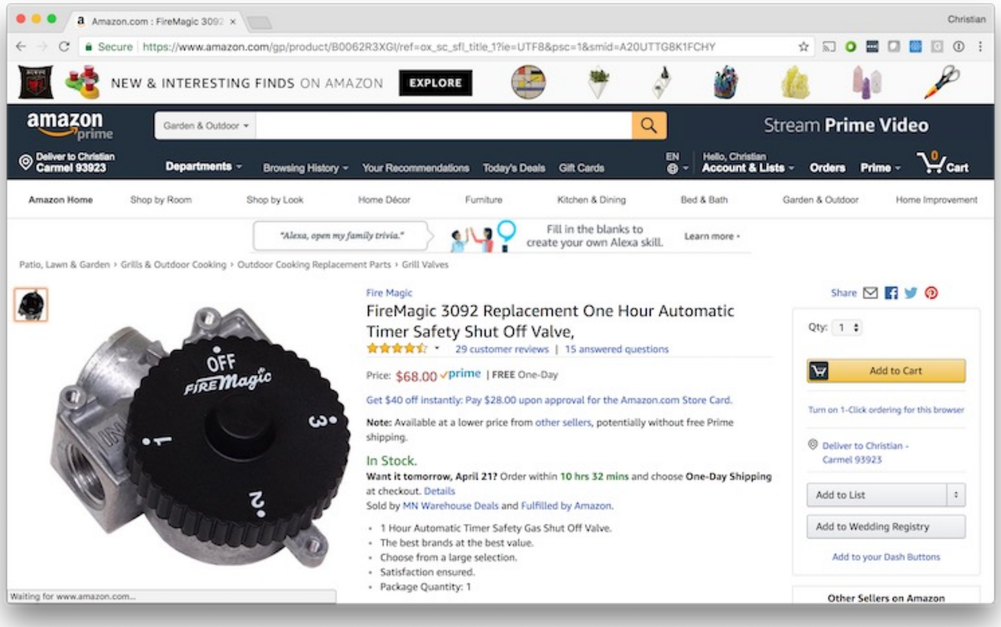


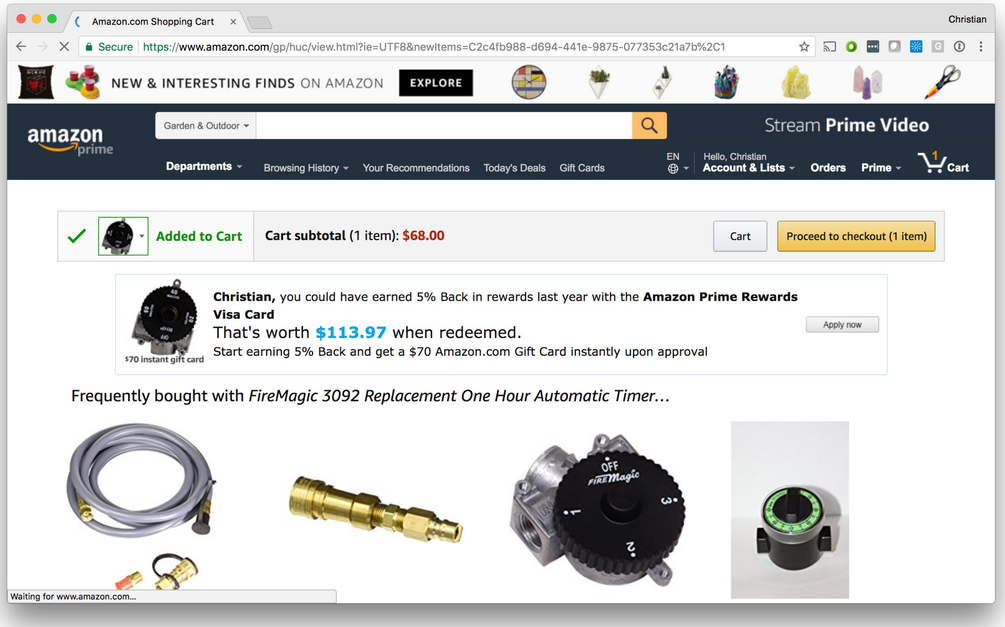


Good Discoverability in Chargepoint (an electric car charging station location app): Defaults to current location, but can easily search others

NORMAN'S PRINCIPLES

2. Feedback. There is full and continuous information about the results of actions and the current state of the product or service. After an action has been executed, it is easy to determine the new state.



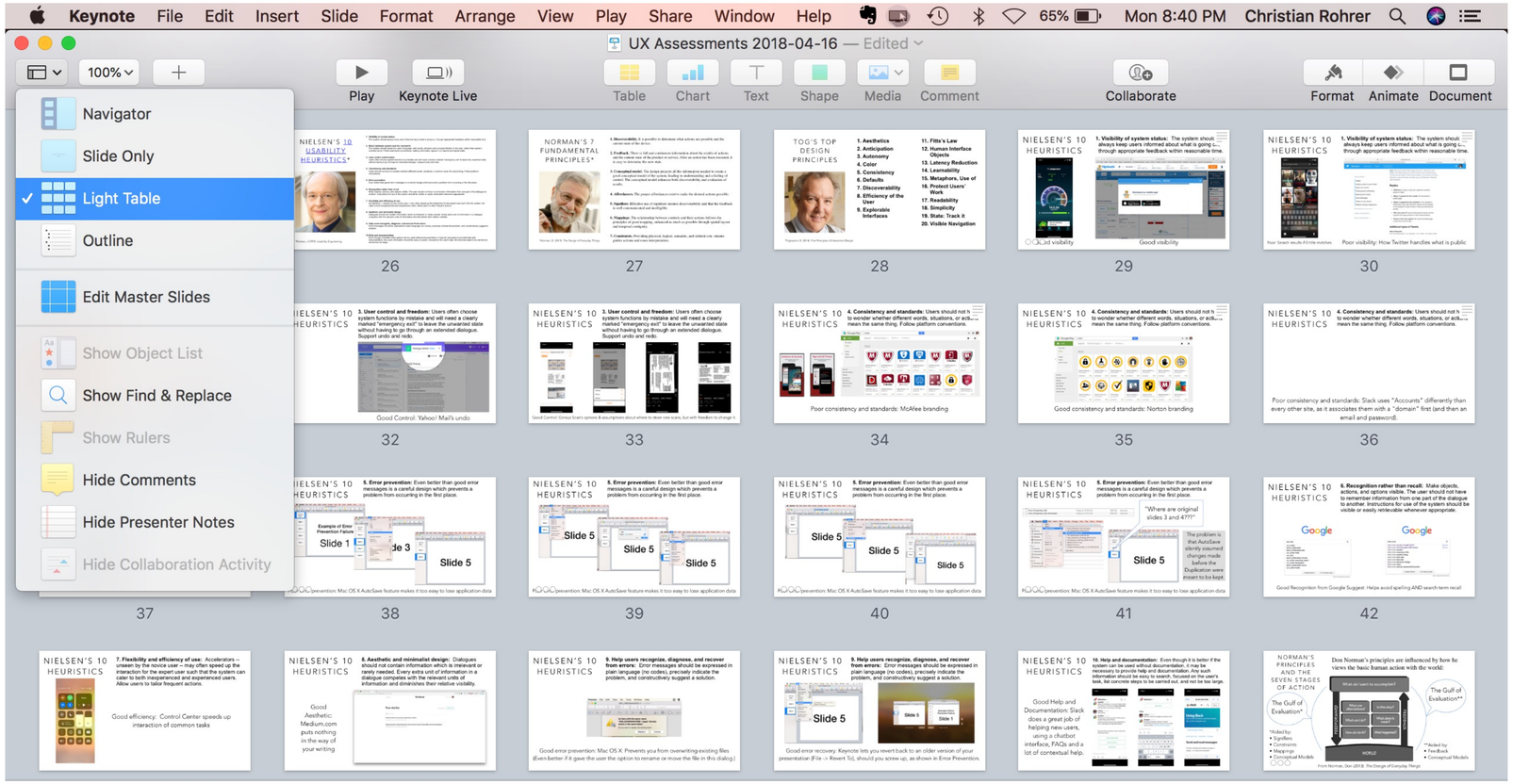


Good feedback: Amazon confirms with strong typography and a graphic (green check) that the item was added to the cart

NORMAN'S PRINCIPLES

3. Conceptual model. The design projects all the information needed to create a good conceptual model of the system, leading to understanding and a feeling of control. The conceptual model enhances both discoverability and evaluation of results.

Good Conceptual Model: Keynote can display slides as if they were slides on a “light table,” able to be seen at a high level, recognized, and moved around easily.



NN/g

NORMAN'S PRINCIPLES

4. Affordances. The proper affordances (capabilities) exist to make the desired actions possible.

Good Affordance: What can you do with this?

And this?

NORMAN'S PRINCIPLES

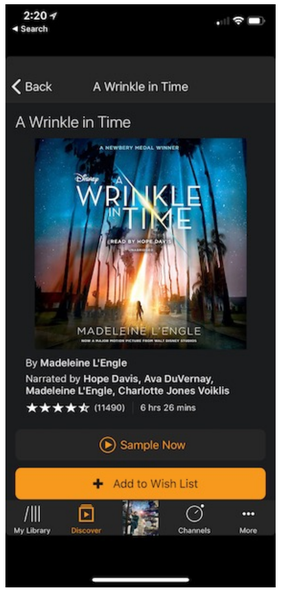
4. Affordances.

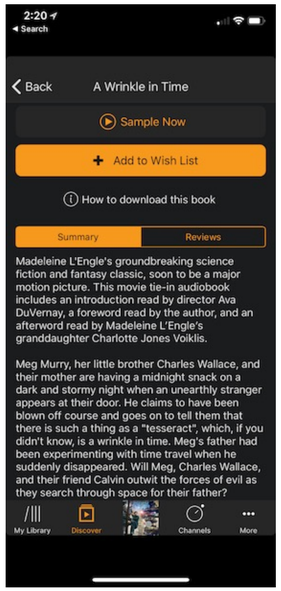
An Affordance is a **relationship** between an object (physical or digital) and a user of that object. The object "affords" certain actions, if the relationship is right. For example, size matters in terms of whether the affordance is truly there.

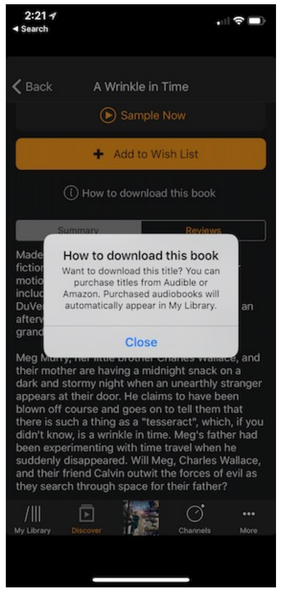
NORMAN'S PRINCIPLES

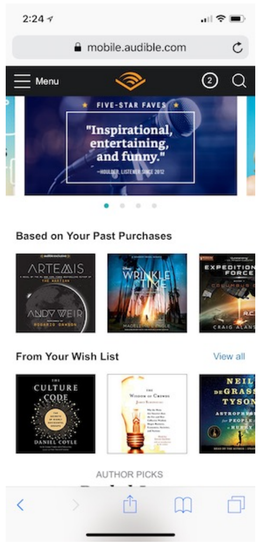
4. Affordances. The proper affordances exist to make the desired actions possible.

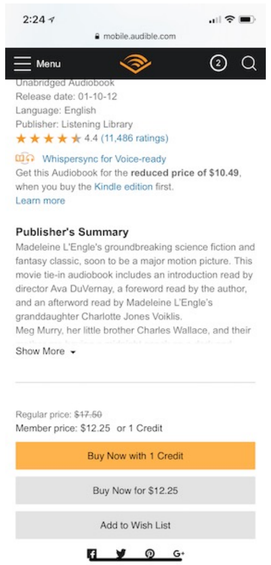
No Affordance: Audible (App) formerly did not afford the user the ability to purchase an audiobook from their iPhone app (due to rev share with Apple). The user had to use their mobile or desktop website first.

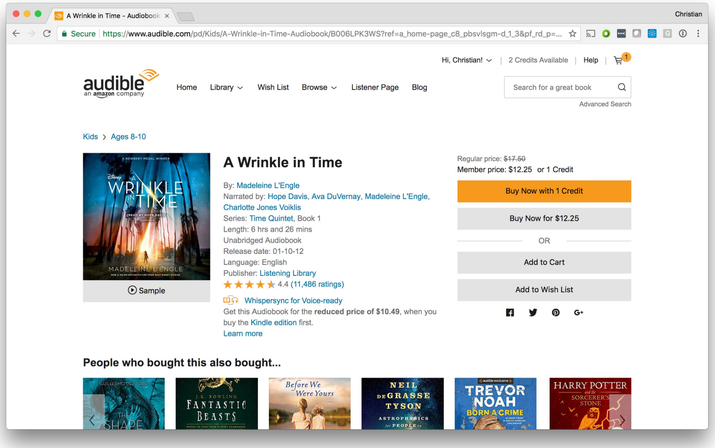










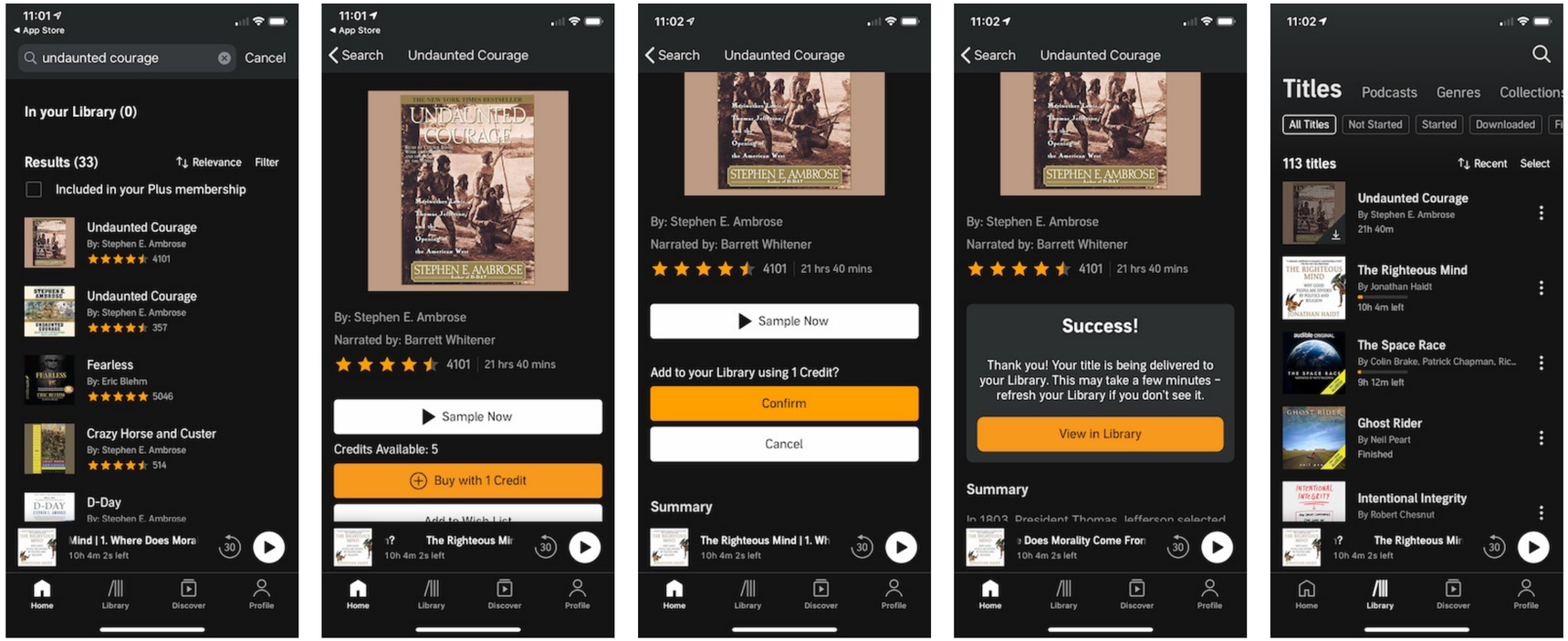


NN/g

NORMAN'S PRINCIPLES

4. Affordances. The proper affordances exist to make the desired actions possible.

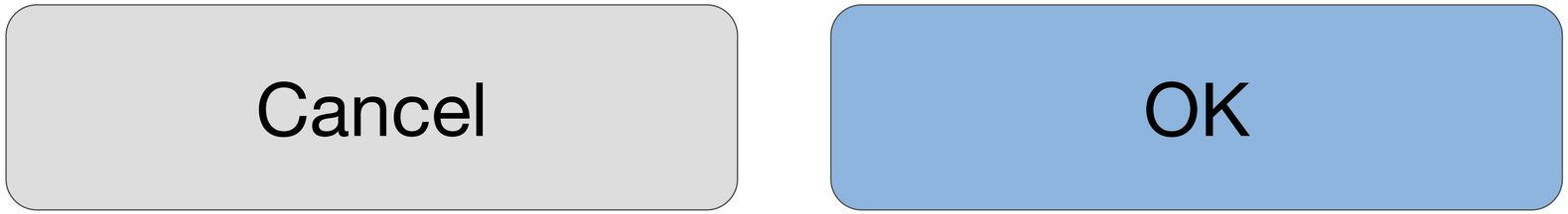
Improved Affordances with Audible today: The Affordance to buy a book is now there in the app itself (somebody finally woke up)



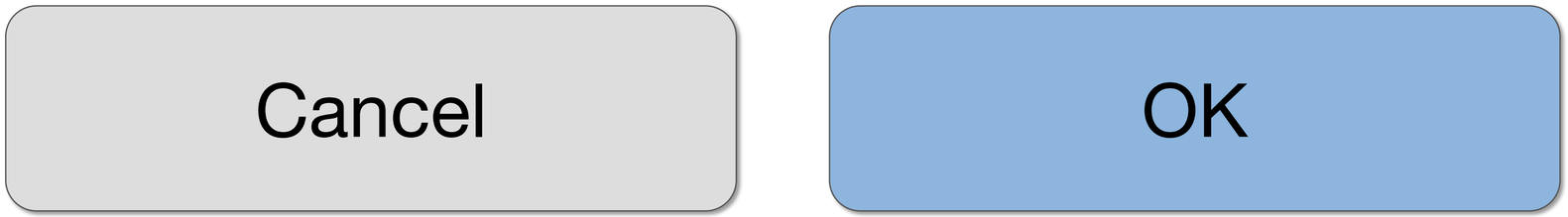
NORMAN'S PRINCIPLES

4. Affordances. The proper affordances (capabilities) exist to make the desired actions possible.

Digital Affordances: What can you do with these?



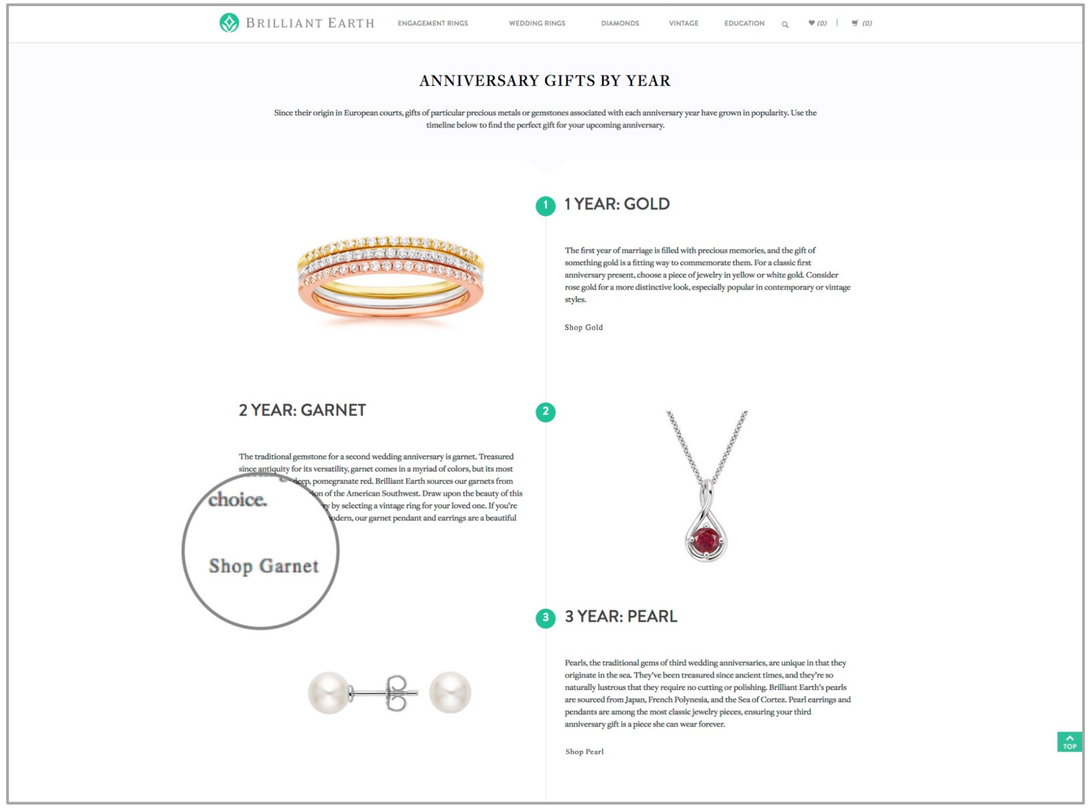
What about these?



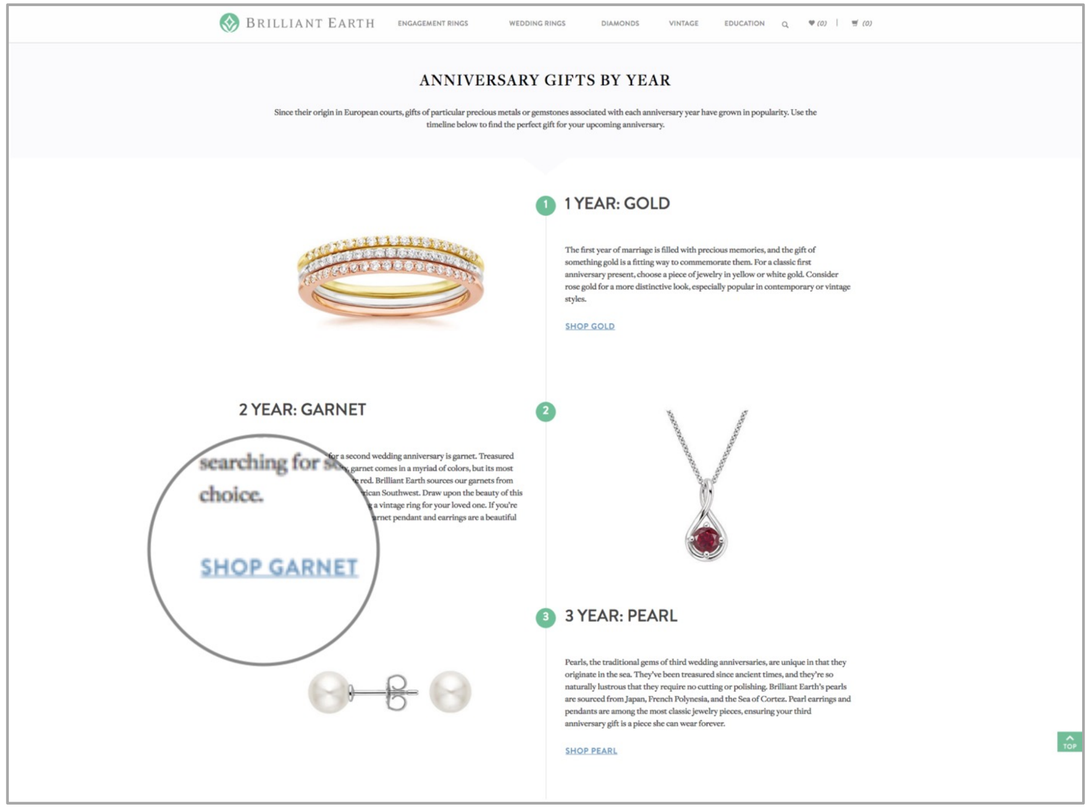
The Affordance is the same. It's the "Signifier" that differs.

NORMAN'S PRINCIPLES

5. Signifiers. Effective use of signifiers ensures discoverability and that the feedback is well communicated and intelligible.



Flat UI with weak signifiers



Traditional UI with strong signifiers

Good and Bad Signifiers in Traditional vs. Flat UI Design (by Kate Meyer: https://www.nngroup.com/articles/flat-ui-less-attention-cause-uncertainty/)

NN/g

NORMAN'S PRINCIPLES

A word about **Affordances** and **Signifiers**…

The term “Affordance” in practice, came to mean two different but related ideas:

1. Real Affordance: When a capability is actually available for a user whether perceived or not.
2. Perceived Affordance: When the user is able to see that the capability was there, as in strong visual cues (e.g., buttons that look clickable).

This led to additional modifications of the term as in a “Weak Affordance”: when the perception of the affordance is difficult, as in Flat UI design styles, in which the user might have to hover over an item to tap it to know if it’s clickable.

In 2013, Norman clarified this by better distinguishing these two concepts, and he now uses the term “Signifier” to mean Perceived Affordance and encourages designers to use strong signifiers to aid discovery of affordances.

NORMAN'S PRINCIPLES

6. Mappings. The relationship between controls and their actions follows the principles of good mapping, enhanced as much as possible through spatial layout and temporal contiguity.

Good Mapping: Vivino guides a user taking a photo of a wine bottle label with the way it masks the real time camera feed.

Notice how the curve of the label matches the curved mask of the camera feed?

NN/g

NORMAN'S PRINCIPLES

6. Mappings. The relationship between controls and their actions follows the principles of good mapping, enhanced as much as possible through spatial layout and temporal contiguity.

Good Mapping: Mercedes Benz chose to show how seats can be controlled with a natural mapping.

NORMAN'S PRINCIPLES

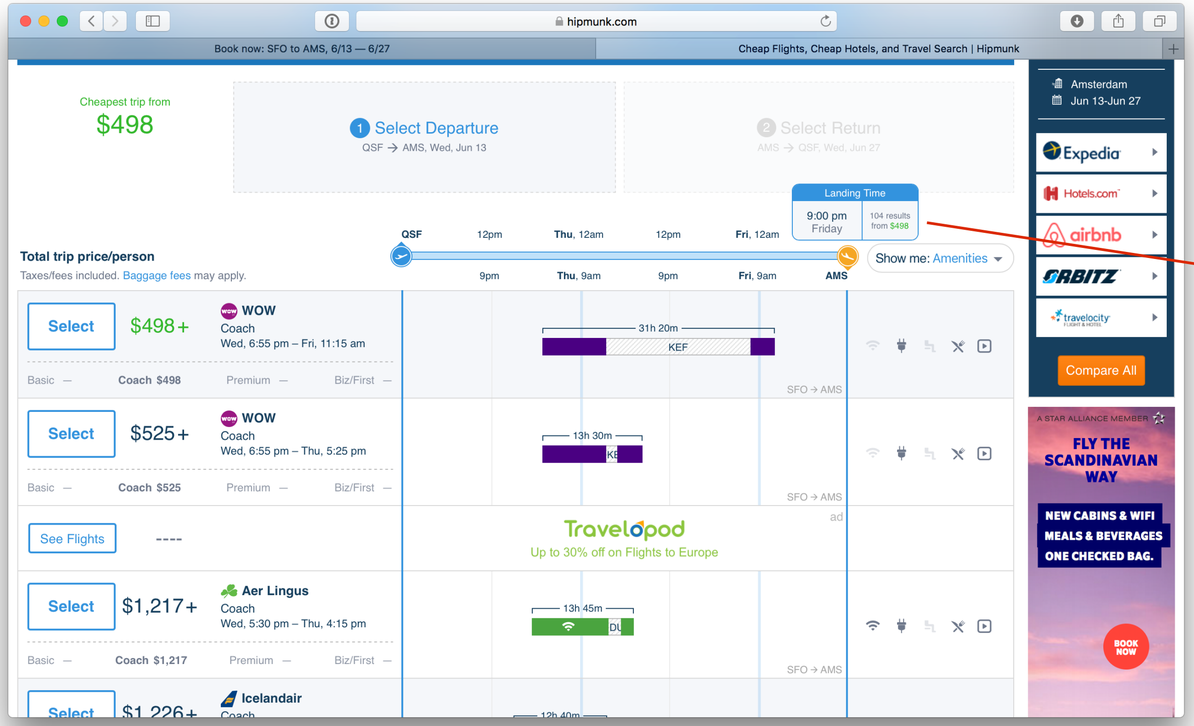
7. Constraints. Providing physical, logical, semantic, and cultural constraints guides actions and eases interpretation.

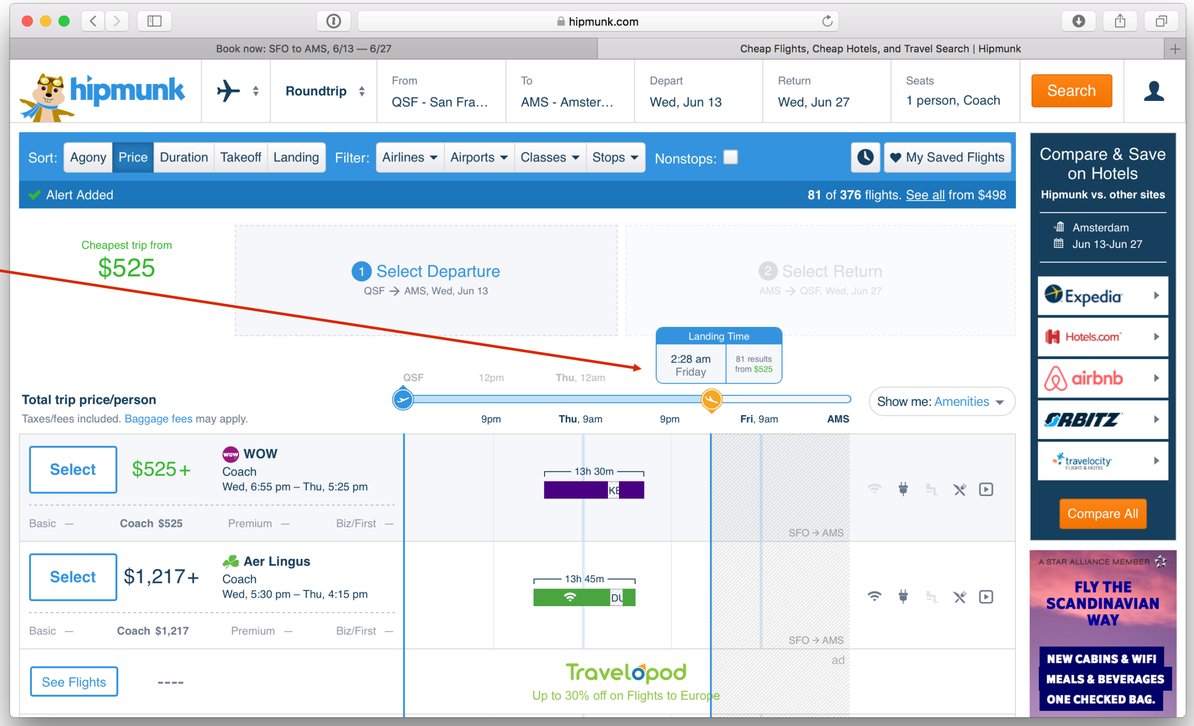
These doors have very clear constraints about which way they open (by the jamb and visible catch plate).

NORMAN'S PRINCIPLES

7. Constraints. Providing physical, logical, semantic, and cultural constraints guides actions and eases interpretation.

Good Constraints: Hipmunk's sliders allow the user to constrain time, but **also** provide a good mapping to duration (time)





NN/g

NORMAN'S PRINCIPLES

7. Constraints. Providing physical, logical, semantic, and cultural constraints guides actions and eases interpretation.

Good Constraints: Constraints show you what is possible given the current state, and what is not. Example: greyed out menu items

TOG'S TOP DESIGN PRINCIPLES

1. Accessibility
2. Æsthetics
3. Animation
4. Anticipation
5. Autonomy
6. Consistency
7. Control
8. Customization
9. Defaults
10. Discoverability
11. Efficiency of the user
12. Envision Information
13. Explorable Interfaces
14. Feedback
15. Flow
16. Freedom
17. Help
18. Human-Interface Objects
19. Illusions Good & Bad
20. Learnability
21. Leverage
22. Memorability
23. Metaphors
24. Modes
25. Protect Users' Work
26. Readability
27. Response Time & Latency
28. Simplicity
29. Situational Awareness
30. Stability
31. Staged Revelation
32. State, Track It
33.Undo
34. Visibility

*Tognazzini, Bruce (2014): First Principles of Interaction Design

Tog teaches an entire course on these principles - worth taking!

QUESTIONS? (EXERCISE FOLLOWS)

- You may now ask questions from this section
- You may type your question into Zoom Chat
- We may have to cover some questions as follow-ups after the seminar

Recap: Nielsen and Norman Heuristics & Principles (and others)

Jakob Nielsen's 10 Heuristics

1. Visibility of system status
The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.

2. Match between system and the real world
The system should speak the users' language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.

3. User control and freedom
Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support undo and redo.

4. Consistency and standards
Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.

5. Error prevention
Even better than good error messages is a careful design which prevents a problem from occurring in the first place.

6. Recognition rather than recall
Make objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.

7. Flexibility and efficiency of use
Accelerators -- unseen by the novice user -- may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.

8. Aesthetic and minimalist design
Dialogues should not contain information which is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.

9. Help users recognize, diagnose, and recover from errors
Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.

10. Help and documentation
Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large.

Don Norman's 7 Principles

1. **Discoverability**. It is possible to determine what actions are possible and the current state of the device.

2. **Feedback**. There is full and continuous information about the results of actions and the current state of the product or service. After an action has been executed, it is easy to determine the new state.

3. **Conceptual model**. The design projects all the information needed to create a good conceptual model of the system, leading to understanding and a feeling of control. The conceptual model enhances both discoverability and evaluation of results.

4. **Affordances**. The proper affordances (capabilities)exist to make the desired actions possible.

5. **Signifiers**. Effective use of signifiers ensures discoverability and that the feedback is well communicated and intelligible.

6. **Mappings**. The relationship between controls and their actions follows the principles of good mapping, enhanced as much as possible through spatial layout and temporal contiguity.

7. **Constraints**. Providing physical, logical, semantic, and cultural constraints guides actions and eases interpretation.

Other Principles

Label-following: When users are not sure how to use a system for a specific task, they explore their options, looking for matches between the task or goal description (either given to them or in their head) and the labels of the actions available.

Fitts' Law: The time to perform this action increases as the distance needed to travel increases and/or the size of the target decreases.

TRY A HEURISTIC EVALUATION [20 MINS]

Instructions

- Select one of the product choices, based on your interest
- On your own, go through the product/application, noting any problems or potential usability issues
- If possible, reference the heuristic or principle violated for any issues noted
- Document assumptions or questions about the user type
- Capture screenshots or perform a screen recording so you can refer back to it later
- Spend 15 minutes exploring on your own
- Then document an example in a one-pager, possibly to share on Slack or Zoom chat (optional) - 5 mins
- Tip: If you are using a Smartphone, take screenshots or screen recordings, put into a Google Slides or PowerPoint document with your annotations and callouts (see Exercise Slides for a template slide)

Product options (choose one)

If you want to work on a smartphone app, choose one of these:

1. Vivino wine rating smartphone app (see next slide for images of wine bottles to use with this app)

2. Accuweather weather app

If you want to work on a website, choose one of these (useful if you are working with another on Zoom and want to share screens)

3. Accuweather.com website (can be done alone or with others)

Exercise documents: http://bit.ly/nng-ux-assessments

USE THIS IMAGE IF EVALUATING THE "VIVINO" APP

Tip: Get your camera close enough to the screen to make out the text clearly, but not so close that the images get blurry.

What did you discover in your Heuristic Evaluation?

Heuristic/Principle:

Supported or violated?

Severity Rating Level*

Recommendations?

Screenshot(s) here:

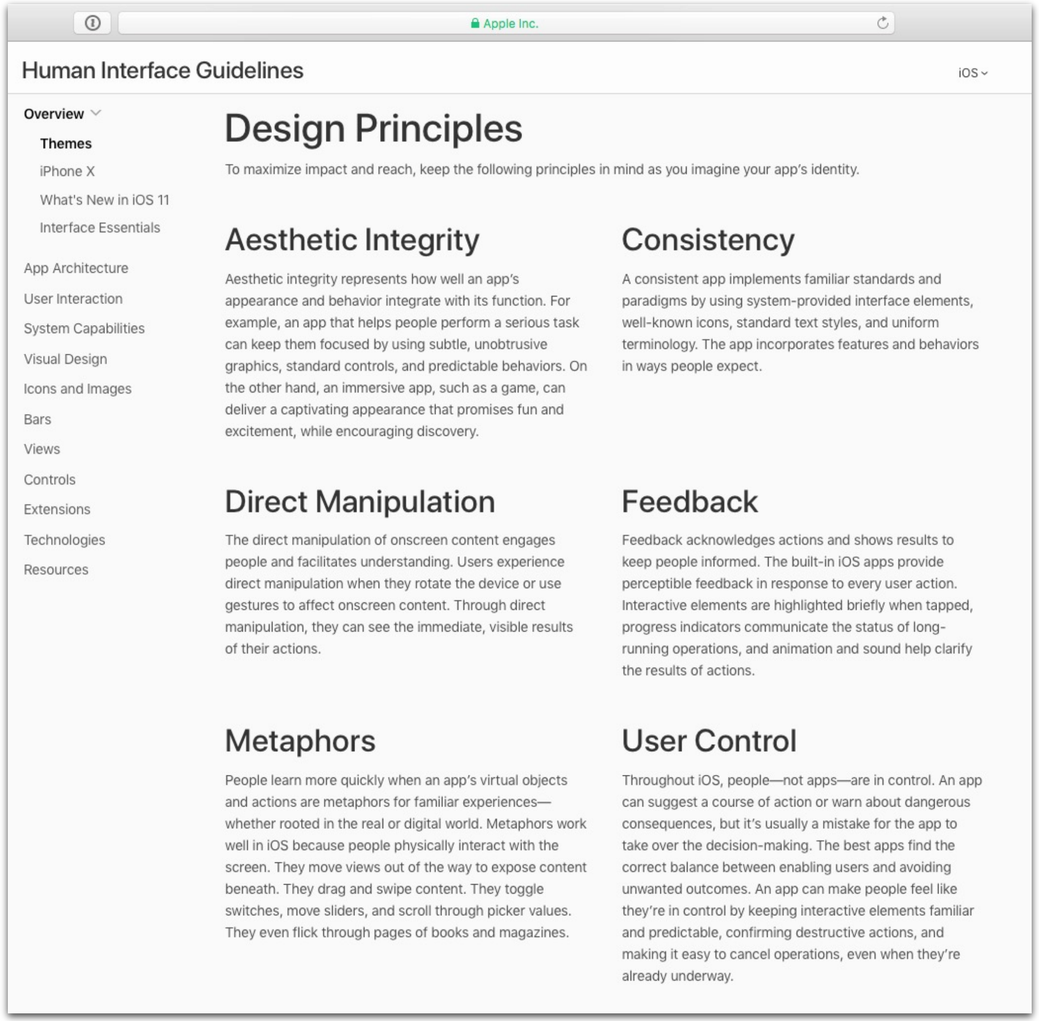
*Severity rating definitions:
0 = I don't agree that this is a usability problem at all
1 = Cosmetic problem only: need not be fixed unless extra time is available on project
2 = Minor usability problem: fixing this should be given low priority
3 = Major usability problem: important to fix, so should be given high priority
4 = Usability catastrophe: imperative to fix this before product can be released

Design Principles & Guidelines

DESIGN PRINCIPLES AND GUIDELINES

- Earlier, we reviewed design principles and heuristics, but the term “design principles” has another meaning
- **Design Principles** are sometimes used to describe the underlying themes that drive decisions about a certain style and associated set of visual and editorial guidelines.
- These “principles” can be very similar to the “principles" or “heuristics” discussed earlier, and are often research based (note resemblances in this example):

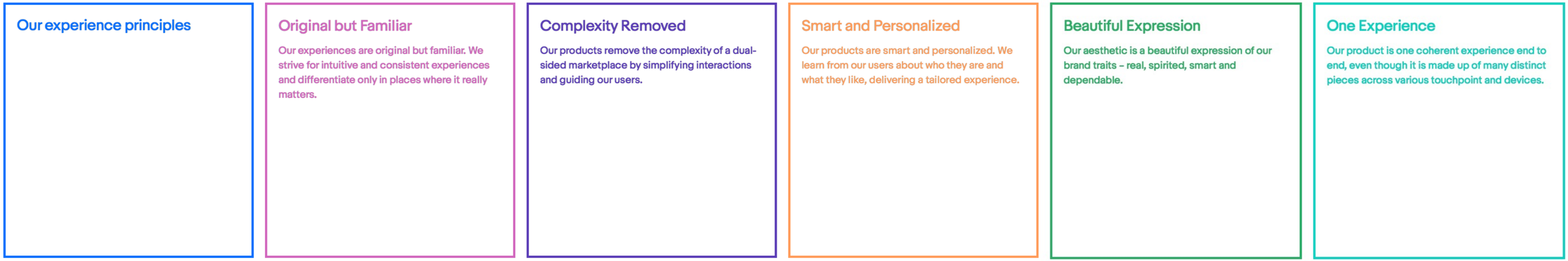
Apple’s iOS Design Principles



DESIGN PRINCIPLES AND GUIDELINES (CONT.)

- However, some “design principles” are mainly used to articulate a top-down direction the company and brand want to see expressed
- This done so that various internal and external design teams follow the same spirit intended for the brand
- Typically, these principles take into account the current brand attributes, the desired brand attributes, and aspects of the domain that matter, as with the principles for eBay, shown here:

eBay’s Experience Principles



DESIGN PRINCIPLES AND GUIDELINES (CONT.)

- These principles usually set the stage for specific examples of how at least two aspects of design should function:
  - The visual design (how it should look)
  - The content or writing (how it should sound)
- Taken together, these principles, the guidelines, examples and rules (do's and don'ts) comprise the guidelines, which are sometimes called the "design language"

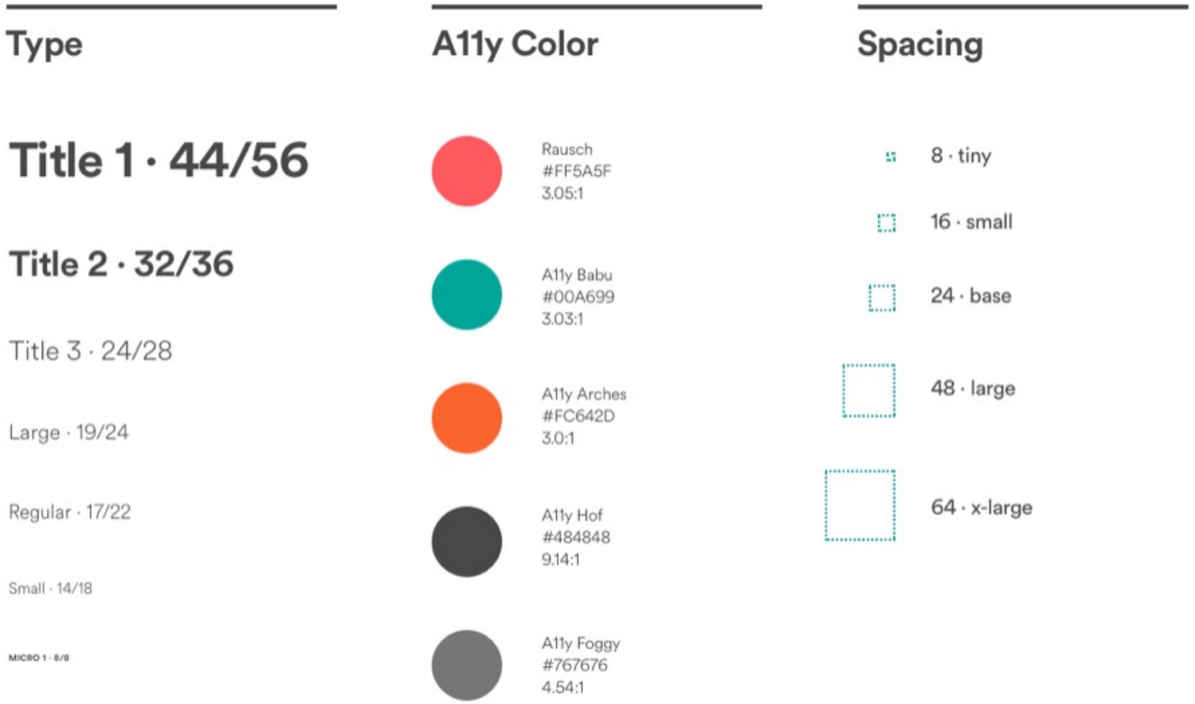
Airbnb's Design Principles:

- **Unified**: Each piece is part of a greater whole and should contribute positively to the system at scale. There should be no isolated features or outliers.
- **Universal**: Airbnb is used around the world by a wide global community. Our products and visual language should be welcoming and accessible.
- **Iconic**: We're focused when it comes to both design and functionality. Our work should speak boldly and clearly to this focus.
- **Conversational**: Our use of motion breathes life into our products, and allows us to communicate with users in easily understood ways.

HOW PRINCIPLES AND GUIDELINES ARE USED

- At a minimum design principles and guidelines are inspirational and directional for designers and writers
- They can be used to do as important documentation on whether future designs are created in a way that complies with these guidelines
- Content (writing) is an important part of the design and experience and needs an expression for how it is used, just as much as visual design (perhaps even more so)

Airbnb



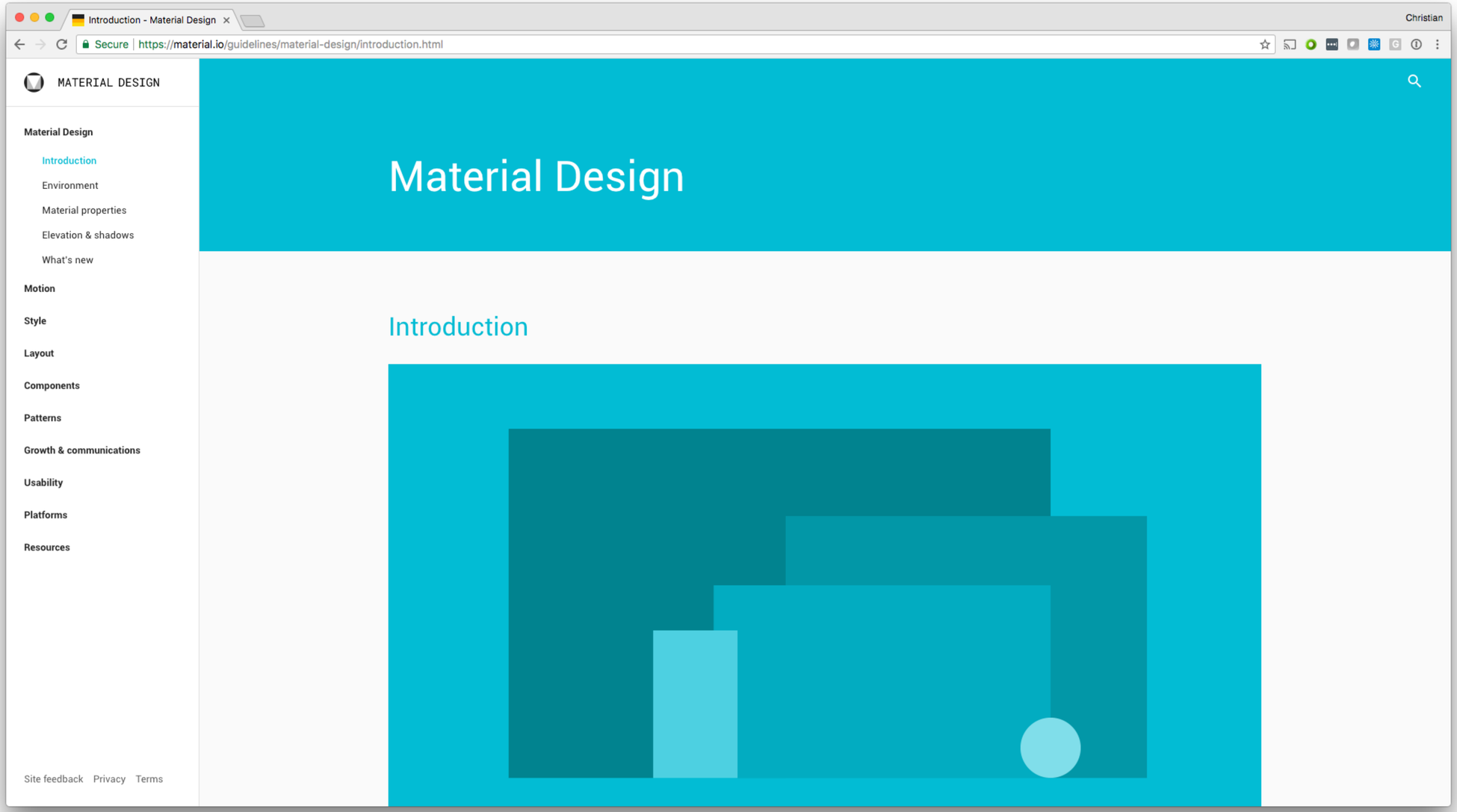
"Our voice is the extension of our brand and personality within the product, and it's the foundation of everything we write."

…

"Our voice is **straightforward**, **inclusive**, **thoughtful**, and **spirited**."

-Marissa Phillips, Head of Content Strategy, Airbnb

ANDROID'S MATERIAL DESIGN GUIDELINES: ONE OF THE BEST



https://material.io/guidelines/material-design/introduction.html

NN/g

DESIGN CONSISTENCY REVIEWS

- Design Reviews can take many forms, but for today's purpose, we will focus specific on whether a given design is consistent with the accepted design guidelines for a brand or company
- There are several aspects to consider when conducting a design consistency review, starting with the target user and going all the way down to the visuals and the voice
- Here are some key questions that are asked when reviewing a design for consistency:

Target Users: Who are they, what are their goals, what matters most?

Design Patterns: Are there recurring interactions or components that we re-use frequently, and if so, have we researched and defined how those patterns should work? If so, are we using them?

Design Components: These are the smaller building blocks of Patterns, like buttons and interactive elements. Have we defined these and are we using them correctly?

Visual Design Hierarchy: Have we established a way of creating a strong visual hierarchy, and are we using that when appropriate?

Typography: Have we defined what typeface to use, sizes, spacing and colors, and are we following these rules?

Logos and Colors: Have we used the right logo in the correct place, and are we using the color palette defined for the brand and this touchpoint?

Icons and Imagery: What styles of icons are we using? For imagery, are we using photographic styles, illustrations or something else? What is the role of animation?

Voice and Tone: Are we writing in a manner that is consistent with our brand voice, and is the tone used appropriate for the context the user is in?

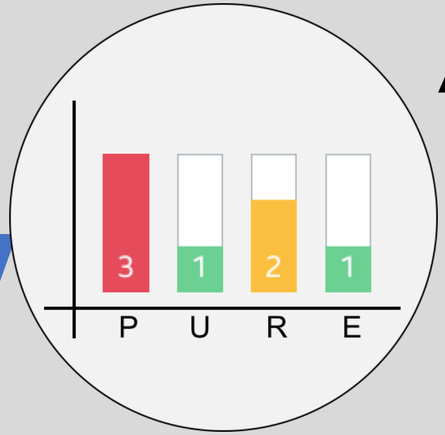
QUESTIONS?

- You may now ask questions from this section
- You may type your question into Zoom Chat
- We may have to cover some questions as follow-ups after the seminar

The PURE Method

The PURE Method

PRACTICAL
USABILITY
RATINGS by
EXPERTS



A Pragmatic Approach for Scoring Product Usability

A little background on where this method came from

WHAT IS THE PURE METHOD?

- The PURE Method is a way to estimate an important aspect of User Experience: Ease-of-use
- It represents how hard a given product is to use for its most important user type
- It is usually limited to the most fundamental or critical tasks, but it can be applied to any set of tasks, time permitting
- It results in a scorecard for each of these tasks and the entire product as a whole
- It also includes the detailed analysis of how those scores were derived

How PURE came to be: Business leaders, product professionals, and engineers are obsessed with dashboards, metrics, quantitative test & scores, as evidenced by statements like these:

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

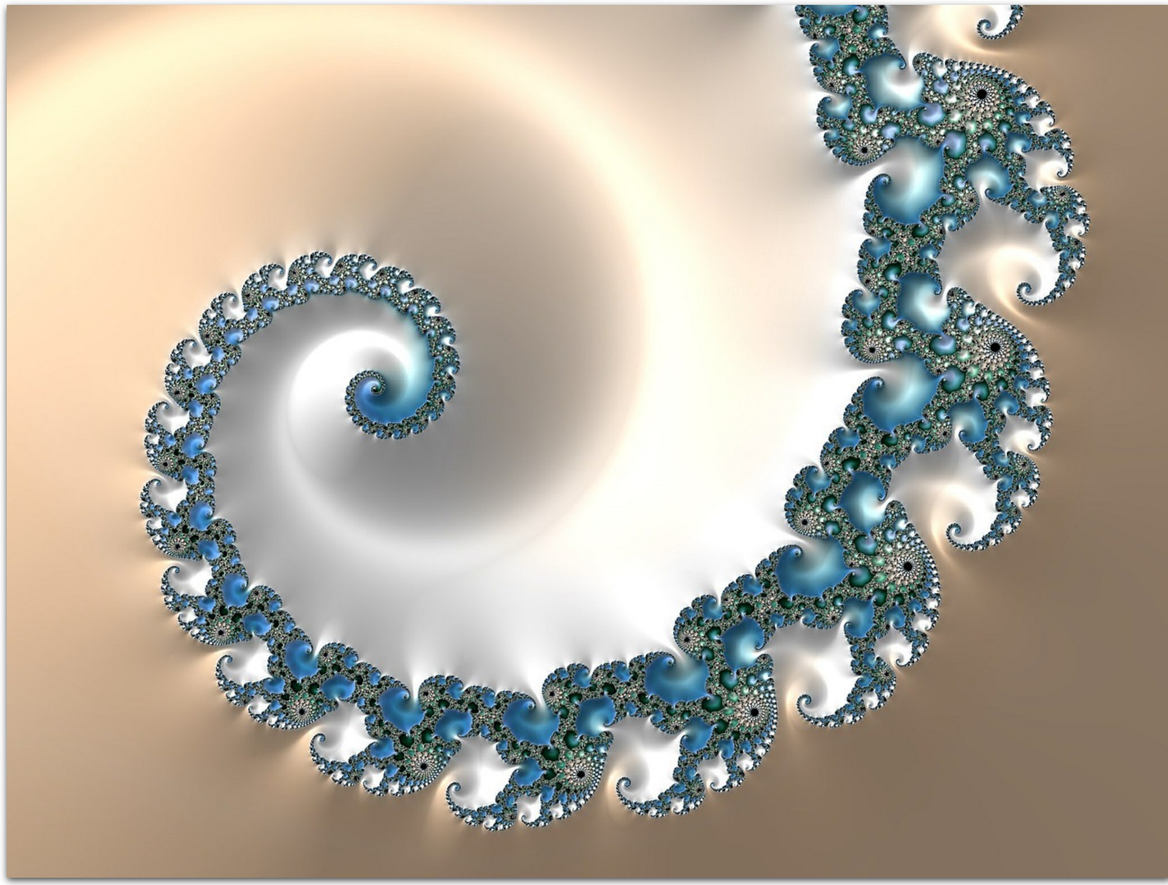
Common measures for UX include: Conversion rates, App downloads, customer reviews, A/B testing performance, uptime, abandon rate, Google Analytics data, path flow patterns

Q?

But are these really measures of User Experience?

Remember: According to various authorities, User Experience is basically… “all aspects of every interaction”

NN Group: “UX encompasses **all aspects** of the end-user’s **interaction** with the company, its services, and its products.”
The ISO: “person's **perceptions and responses** resulting from the **use** and/or anticipated use of a **product, system or service**”
UXPA: “**Every aspect** of the user's **interaction** with a **product, service, or company** that make up the user's perceptions of the whole."
Wikipedia: “a person's **emotions** and **attitudes** about using a particular product, system or service"
IXDA: (They don’t offer a definition, but refer to the above. They reiterate the idea that user experience is “**everything**.”)



Yep! Everything.

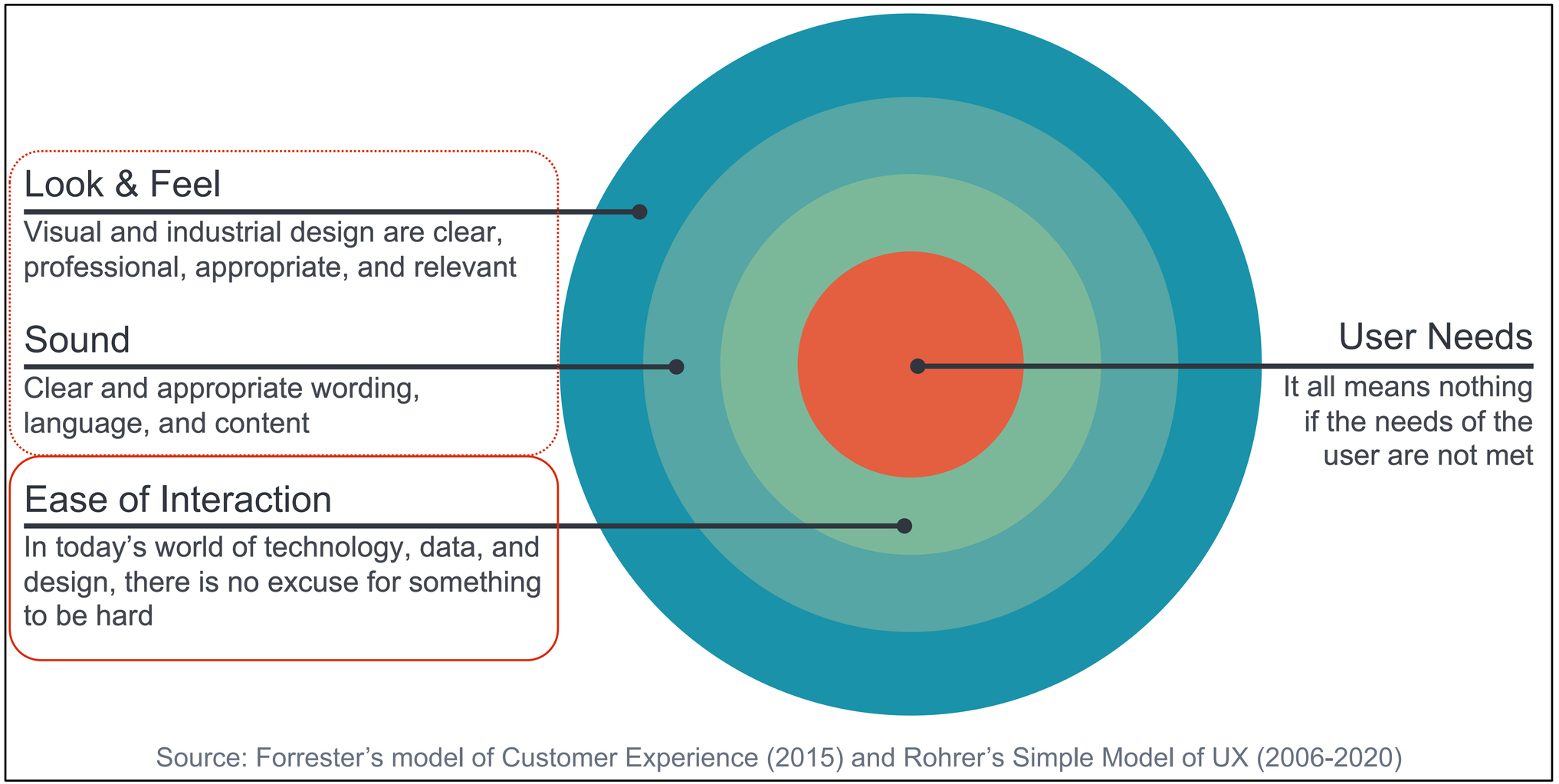
Forrester's Model of Customer Experience

The experience delivers value to customers

It's not difficult for customers to get value from the experience

Customers feel good about their experience

A Simple Model of User Experience



PURE focuses on Ease of Interaction (and Look, Feel & Sound, when they support Ease)

ANALOGY: IN PURE, WE JUST 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"

Photo credit: Agência Brasil Fotografias (source: Flickr Creative Commons)

Simone Biles' floor exercise score by panel of judges at the Rio Olympics in 2016

Difficulty: How difficult the moves performed in her routine, based on a panel of 2 judges. Calculated prior to the routine

+

Execution: How well she performed those move. Starting from 10 possible points, judges deduct points for each mistake

=

SCORE

The PURE Rating Rubric is much simpler than in gymnastics

A rating of 1, 2 or 3 is given for each step in the task, based on the following PURE Rating Rubric:

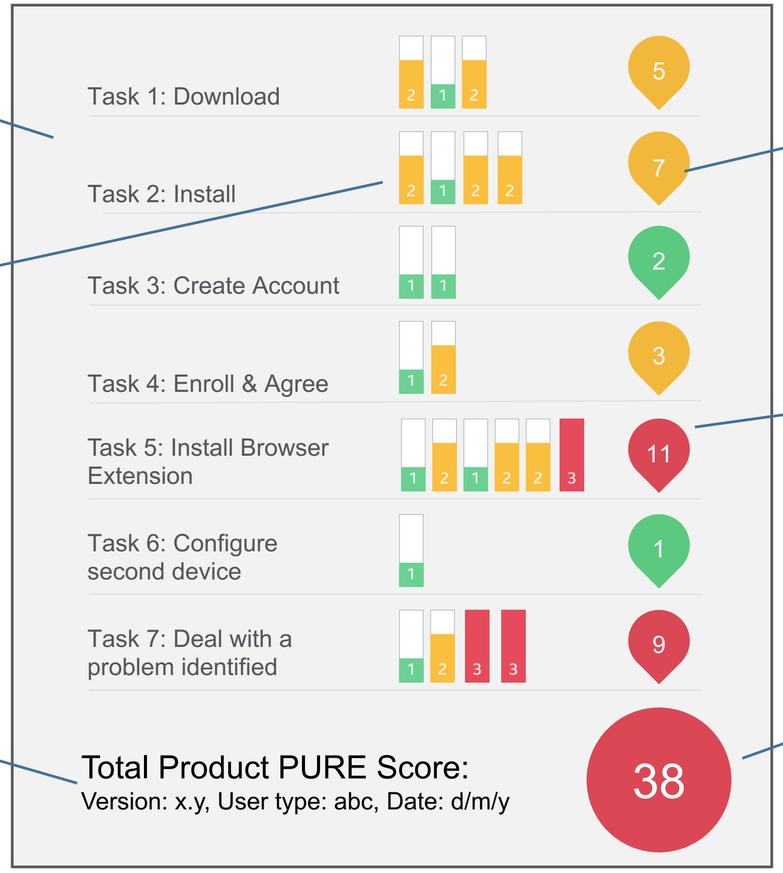
- **1** = The step can be accomplished easily and quickly by the Target User, because there is very little cognitive load. Examples:
  - Easy to understand language/user interface
  - A single, recognizable and clear call to action
  - A familiar interaction pattern, such as the acceptance of a EULA (end user license agreement)
- **2** = The step requires some degree of thought by the Target User, but can generally be accomplished with such effort
- **3** = The step is difficult for the Target User, due to significant cognitive load; some of the Target Users would likely fail at this step of the task

The output is an Ease of Use Scorecard for the most desirable paths of a product for a given point in time

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

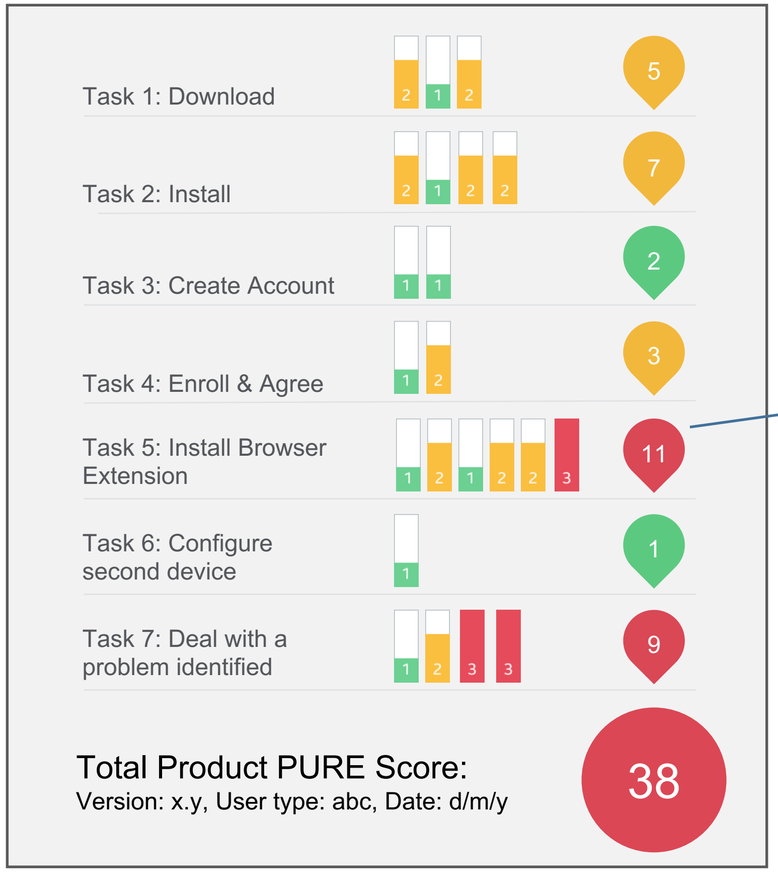


Each task gets a score and color. As in golf, green is good, smaller numbers are better

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

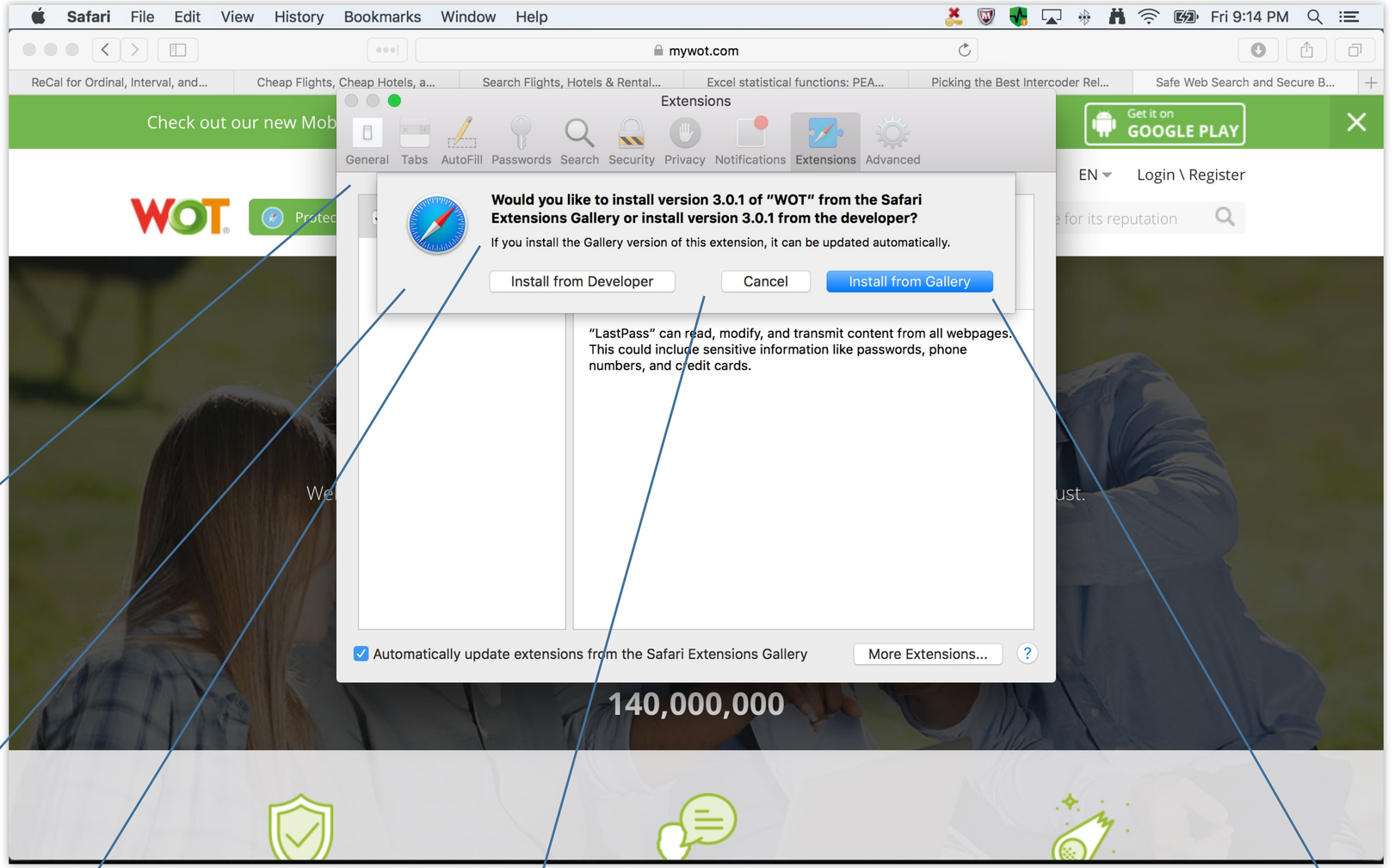
11

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)

QUESTIONS?

- You may now ask questions from this section
- You may type your question into Zoom Chat
- We may have to cover some questions as follow-ups after the seminar

HOW TO CONDUCT THE “PURE” METHOD

Preparing to conduct the PURE Method

What you'll need:

- Collaboration from Product, Design and Technology for the Kickoff/Setup meeting and supporting materials
- At least 3 expert evaluators (user researchers, typically)
- Displays, Devices, Recording Equipment, Spreadsheets
- 1-3 days to conduct the method (the first few times, it requires a bit more time for evaluators to calibrate on the rating scale)

The PURE Method Overview

Kick Off Meeting

1. Define the Target User Type(s)
2. Determine the Fundamental Tasks
3. Specify the “Happy Path” or Typical Path of each Fundamental Task

(Ensure that Product Management, Design and Engineering are present)

Typical 1-2 hours

Preparation

1. Get access to flows in paths to be evaluated
2. Record paths (e.g., w/ screen recorder)
3. Identify the “step boundaries”
4. Name tasks and step boundaries
5. Document

(Can be done by single, lead evaluator on own)

Typical 1-2 days

Team Evaluation

Team of 3+ evaluators:

1. Walk though each path and silently rate each step on a 1-3 scale (the PURE rubric)
2. Discuss ratings and rationale to determine the singular, “Decided On” team ratings
3. Calculate inter-rater reliability of silent scores
4. Sum and color the Task Ratings; sum up the Task scores to get the Total PURE
5. Report the descriptive insights behind each Step Score

Typical ½ day

Kickoff Step 1: Define User Type

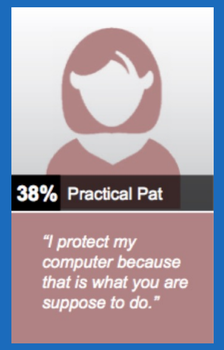
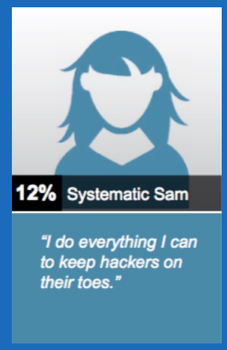
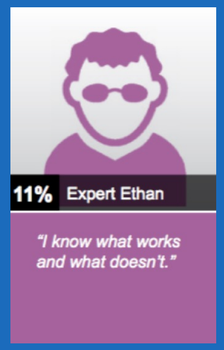
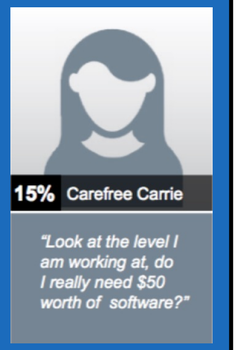
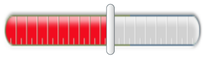
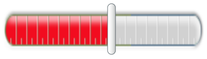
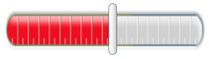
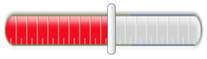
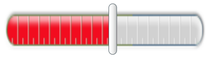
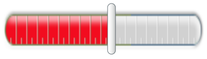
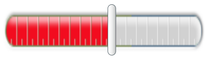
- The Target User type is clearly defined and described by the Design Lead and Product Management (e.g., via well-known personas):

- Consider key behavioral and attitudinal attributes that are relevant to how the product will be used.

Defining User Type in Detail

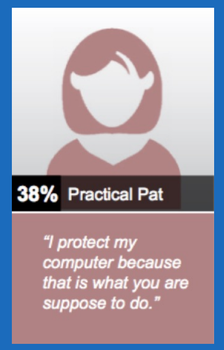
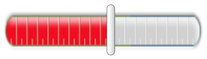
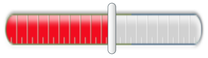
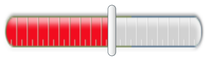
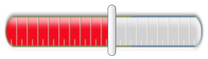
- To conduct a PURE Evaluation, you must describe and document the user type in **at least these three main ways**:
  1. **Technical skill**. Specify how much the user type knows about the technology being assumed in the PURE evaluation, because someone who is technically proficient will have far less difficulty than a novice.
  2. **Domain familiarity**. Indicate how much a user type knows about a given domain (e.g., banking, security or commerce), as this also greatly affects the "cognitive load" or difficulty faced when going through a flow.
  3. **Contextual assumptions**. Describe the user type's usage scenario and what constraints and pressures they are likely to have. This helps improve the accuracy and reliability of the PURE evaluation.
- The above can be used in lieu of a Persona. Personas that don't have this type of description usually need to have it specified to work well with PURE.

Example: Users in Consumer Security

Consumer Security Personas					
Awareness of consumer security issues					
Concerns about all devices (not just PCs)					
Level of concern about identity protection					
Level of interaction with security solutions					
Technical skill and early adoption level					

- In this description of users, both **technical skill** and **domain familiarity** are specified.
- What remains is to document their assumptions about their context (goals, current status, type of system being used, etc.)

Example of User Type with Contextual Assumptions described in detail (**bold**)

Consumer Security Personas	
Awareness of consumer security issues	
Concerns about all devices (not just PCs)	
Level of concern about identity protection	
Level of interaction with security solutions	
Technical skill and early adoption level	

- Practical Pat has been working for a mid-sized company for over 15 years, and **uses a PC and an Android phone many hours a day**. She is familiar with how it operates, including the MS Office suite, Internet Explorer and Chrome. She has fairly recent versions of software on her work computer, but due to her company's IT department having concerns aoout supporting new releases, it is rarely the very latest version. At home, she has a laptop **PC and an iPad**.
- Although she is not intimately familiar with the technology, Pat **knows how to use the default browsers on her PC, her iPad and Android effectively**, even though the experiences are different. She's **able to navigate** between **desktop web browsing**, **mobile web browsing** (both iPad and Android), **desktop software** (e.g., Office), **iPad apps**, and **Android apps**. She **updates** these systems and apps primarily **when she is prompted** to.
- Her feelings about **consumer security** are not strongly negative or positive - it's **just a maintenance task** she needs to perform on her PCs. She doesn't think about this much for her iPad or Android, as they seem to be automatically taken care of in the update process.
- Pat has received **several notices** that her **security software subscription has expired**. For the purposes of this evaluation, we will assume **Pat is looking to update her Security software on her personal PC laptop, and renew the subscription starting from the security window that keeps popping up, and using the default web browser (Microsoft Edge), if needed, while connected at home through high-speed WiFi.**

Kickoff Step 2: Identify Fundamental Tasks

The Fundamental Tasks are identified by PM and the Design Lead. These are defined as the **5-10 tasks that the Target User(s) MUST be able to do for both the user and the business to be successful**.

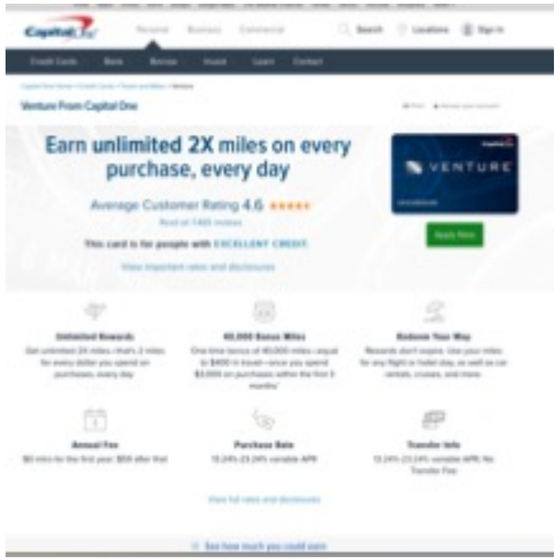
Example Fundamental Tasks for software that monitors security of multiple devices:

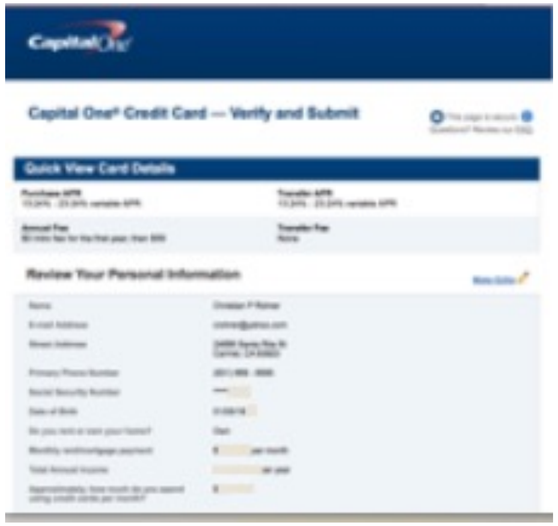
1. Discover & Learn
2. Download & Install
3. Create Account
4. Add a device
5. Access Dashboard
6. Send Message

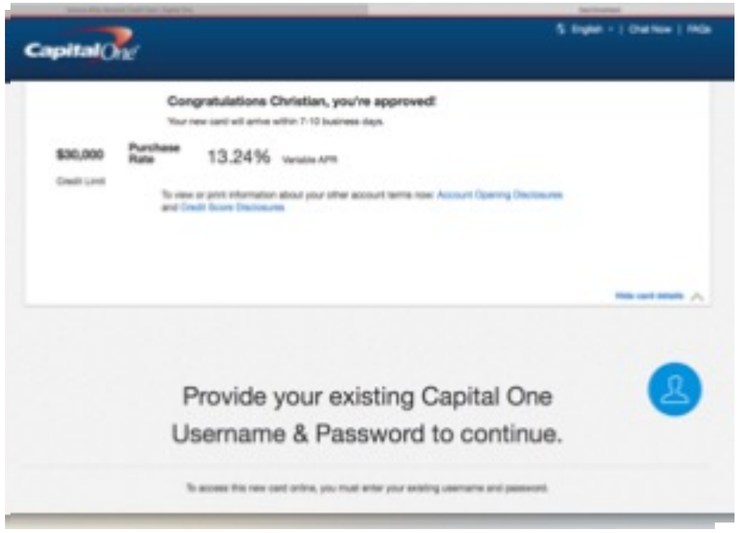
Kickoff Step 3: Specify the “Happy Path” of each Task (or typical path)

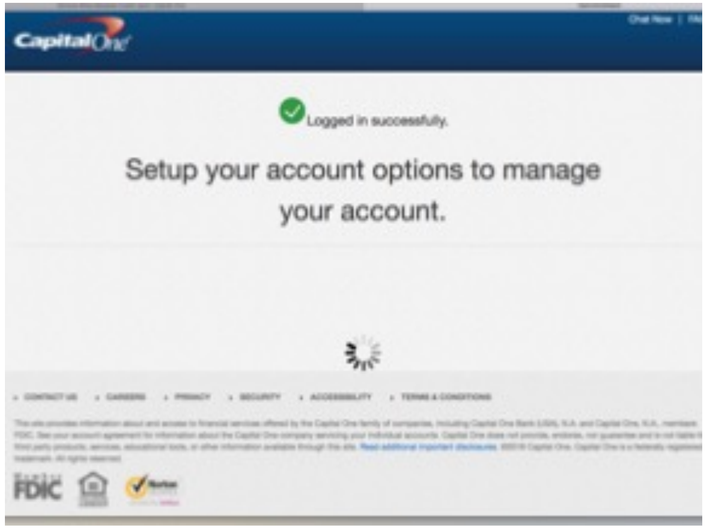
- The “**Happy Path**” of each Fundamental Task is defined as **the most desired path to accomplish the task, as specified by the Design Lead.**
- Alternatively, you can choose the “**typical path**,” if you have analytic data that shows this. Either way, pick one path and stick with it for the analysis.
- Pro Tip: You may need help from Technology to invoke the Happy Path realistically (e.g., usernames/passwords or QA server access)
- We do not evaluate divergent paths, because it creates too much variation and we want to score the “best current performance.”

Example Happy Path with 6 steps (Capital One credit card application):









Preparation Phase (1-2 days)

- After the kick-off meeting, you should have key information locked down: Target User Type, Fundamental Tasks, Happy Paths for those tasks
- Now you need to prepare for the Evaluation Phase of PURE
- Take a few days to prepare for the PURE evaluation, checking back in with stakeholders on your decisions, assumptions and progress

What you will do:

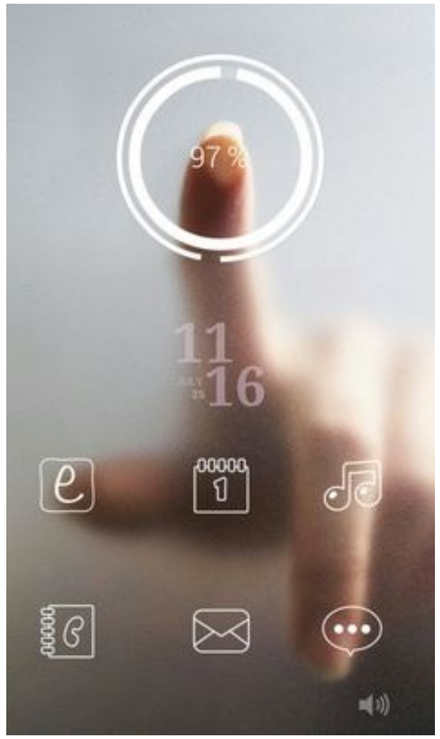
- Give each Fundamental Task a name
- Ensure you have access to the Happy Paths of each task
- Record the Happy Paths of the Fundamental Tasks (e.g., use a screen recorder)
- Determine where the “step-boundaries” of each step are (see forthcoming definition of a “step”)
- Mark where each step-boundary is on the recording, either by calling it out in the audio channel or with a text overlay in the video
- Give each step a name
- Create a scoring spreadsheet with a tab for each rater, providing a place to score each step of each task (templates are available)
- Insert the task names and the step names
- Document user type(s) and flesh out descriptions, as needed

What is the **Default Definition of a "Step?"**

A step **begins** when the system presents the user with a set of options (in the UI) and is waiting for user input to proceed.

STEP BEGINS:
UI is rendered

Micro-interactions may occur within the UI until the user reaches a point where they take an action to proceed to the next step

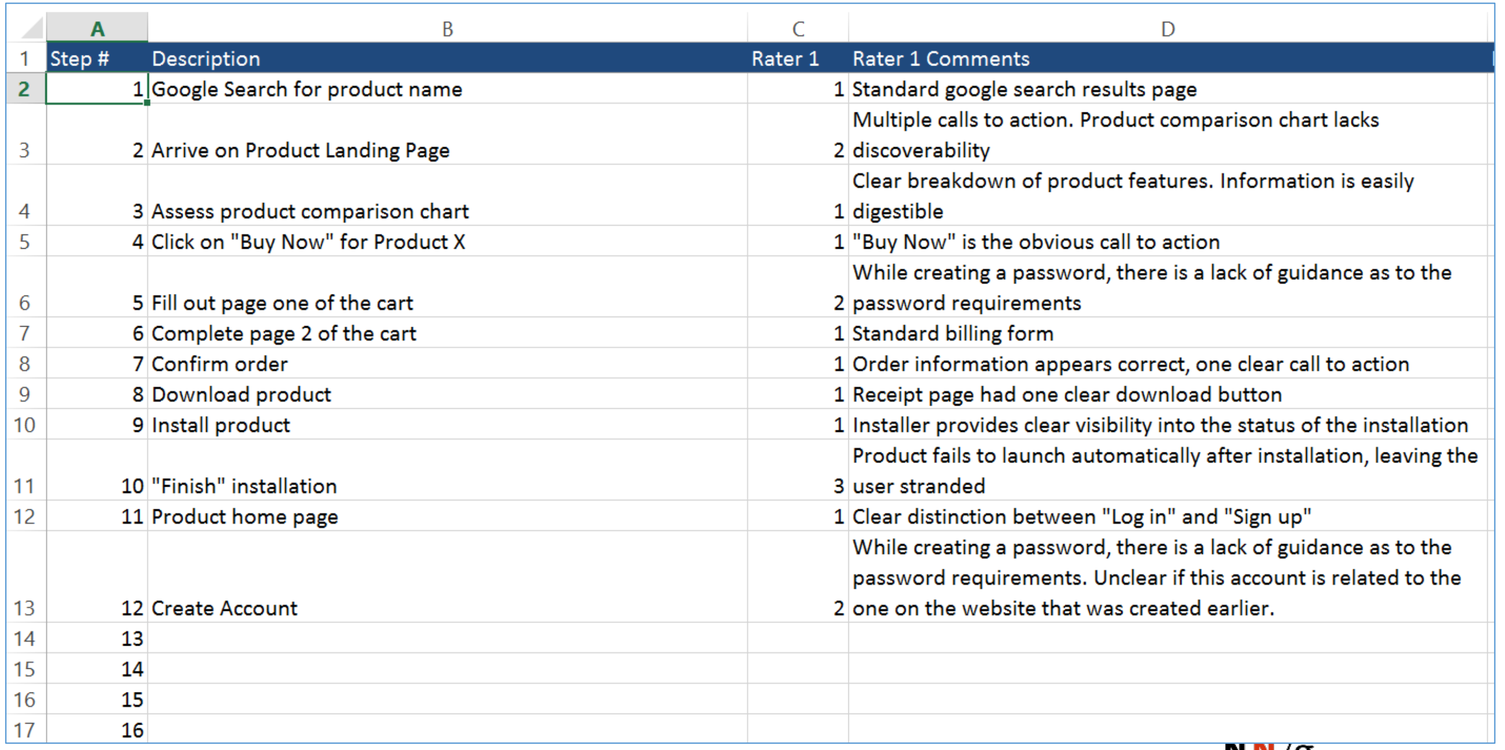


STEP ENDS:
When user makes selection and expects a significant system response

Note: You are free to change the definition of where step boundary occurs. But if you do so, you need to document your decision and rationale so you can apply it in future PURE scoring in a consistently way.

Team Evaluation Step 1: Walk through each Step of each Task and Rate

- The three raters walk through the Happy Path of each Task together (ie., they watch the **same thing at the same time**).
- They silently provide a 1-3 difficulty rating for each step of each Fundamental Task, based on the **PURE Rating Rubric** (next slide).
- Each rater stores his or her PURE Scores in a spreadsheet, including score in each step.
- Pro tip: Use an online spreadsheet with separate tabs per rater.



	A	B	C	D
1	Step #	Description	Rater 1	Rater 1 Comments
2	1	Google Search for product name	1	Standard google search results page
3	2	Arrive on Product Landing Page	2	Multiple calls to action. Product comparison chart lacks discoverability
4	3	Assess product comparison chart	1	Clear breakdown of product features. Information is easily digestible
5	4	Click on "Buy Now" for Product X	1	"Buy Now" is the obvious call to action
6	5	Fill out page one of the cart	2	While creating a password, there is a lack of guidance as to the password requirements
7	6	Complete page 2 of the cart	1	Standard billing form
8	7	Confirm order	1	Order information appears correct, one clear call to action
9	8	Download product	1	Receipt page had one clear download button
10	9	Install product	1	Installer provides clear visibility into the status of the installation
11	10	"Finish" installation	3	Product fails to launch automatically after installation, leaving the user stranded
12	11	Product home page	1	Clear distinction between "Log in" and "Sign up"
13	12	Create Account	2	While creating a password, there is a lack of guidance as to the password requirements. Unclear if this account is related to the one on the website that was created earlier.
14	13			
15	14			
16	15			
17	16			

Reminder: The PURE Rating Rubric Definition

A rating of 1, 2 or 3 is given for each step in the task, based on the following PURE Rating Rubric (a guide for scoring or grading):

1 = The step can be accomplished easily and quickly by the Target User, because there is very little cognitive load. Examples:

- Easy to understand language/user interface
- A single, recognizable and clear call to action
- A familiar interaction pattern, such as the acceptance of a EULA (end user license agreement)

2 = The step requires some degree of thought by the Target User, but can generally be accomplished with such effort

3 = The step is difficult for for the Target User, due to significant cognitive load, so they would likely fail this task or quit (“fail or bail”)

Team Evaluation Step 2: Discuss ratings & rationale to determine the "decided on" score

- Goal: Determine a single PURE rating for each step (the "decided on score")
- For each step and ask raters for their rating rationale; seek to understand what they observed, the UX principles supported/violated, and assumptions
- Do not use the average; consensus is more valuable than false precision here
- First consider the "mode" (most common score) as a starting point for the "decided on rating"; as rationales and assumptions are shared, this may change

Task # & Step #	Rater 1	Rater 2	Rater 3	Avg	Mode	Decided Rating
T1 S1	1	1	1	1	1	1
T1 S2	1	1	1	1	1	1
T1 S3	3	2	3	2.66666667	3	3
T1 S4	1	1	1	1	1	1
T1 S5	2	2	1	1.66666667	2	2
T1 S6	1	1	1	1	1	1

NN/g

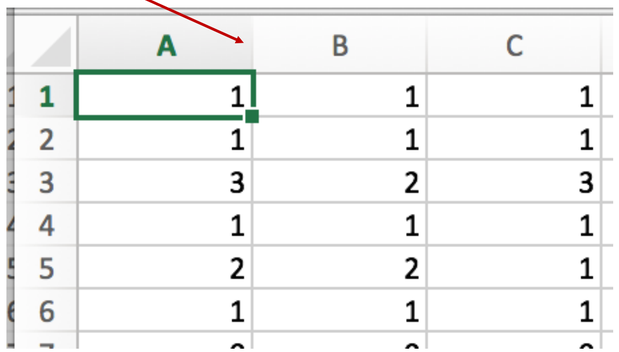
Team Evaluation Step 2 (cont.): Discuss ratings & rationale and determine the "decided on" score

- Take turns discussing your rationale, starting with those in the minority. You may decide on a rating that was not the mode; this is OK if rationale is good and the team agrees on the assumptions that lead to this rating.
- Discuss why you have different initial scores to establish common criteria for interpreting the Rubric; in the future, your ratings will most likely converge.

Task # & Step #	Rater 1	Rater 2	Rater 3	Avg	Mode	Decided Rating
T1 S1	1	1	1	1	1	1
T1 S2	1	1	1	1	1	1
T1 S3	3	2	3	2.66666667	3	2 ~~3~~
T1 S4	1	1	1	1	1	1
T1 S5	2	2	1	1.66666667	2	2
T1 S6	1	1	1	1	1	1

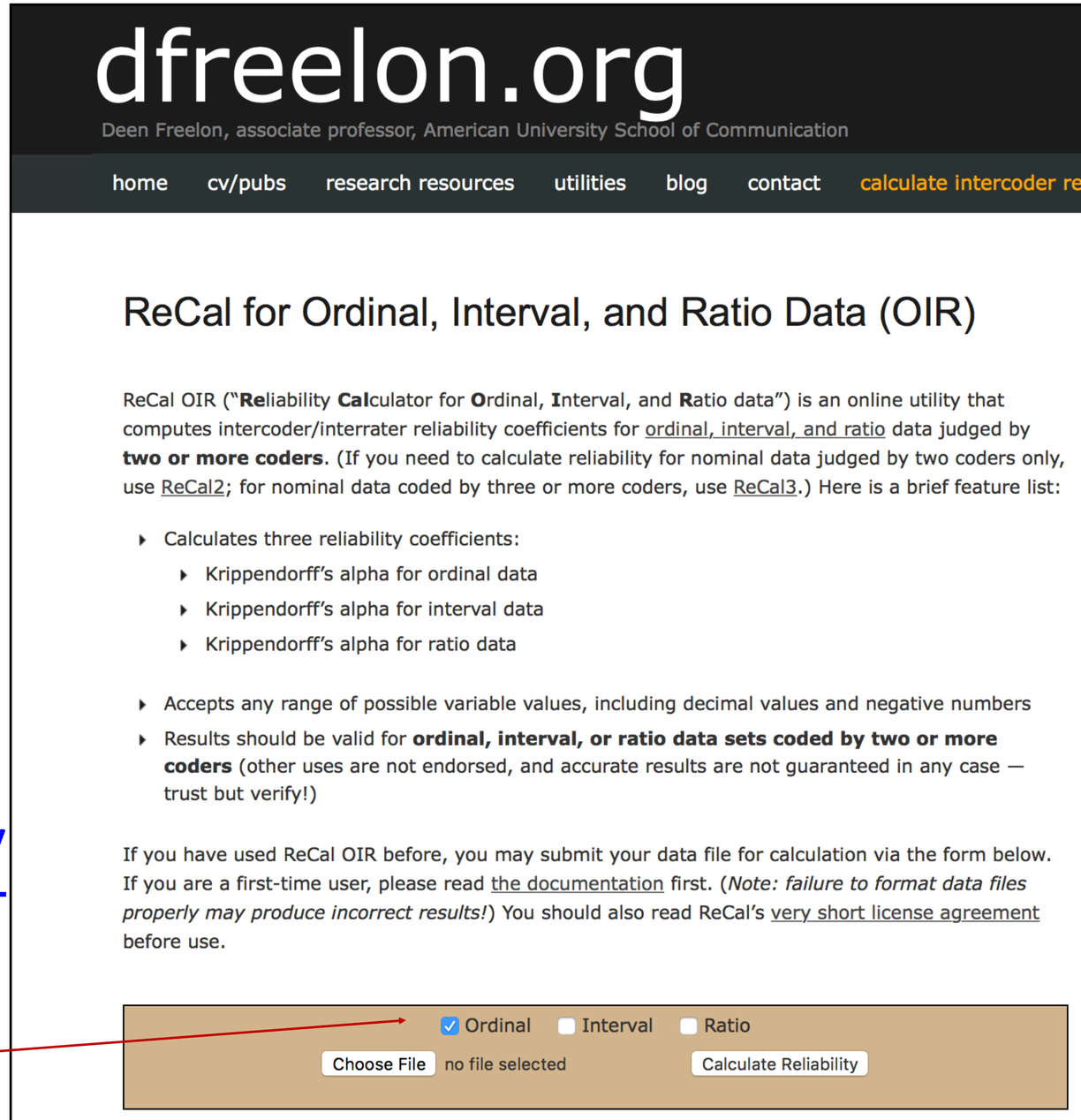
Team Evaluation Step 3: Calculate Inter-rater Reliability (IRR)

- Put all ratings of all tasks for all raters into a single sheet w/ no headers

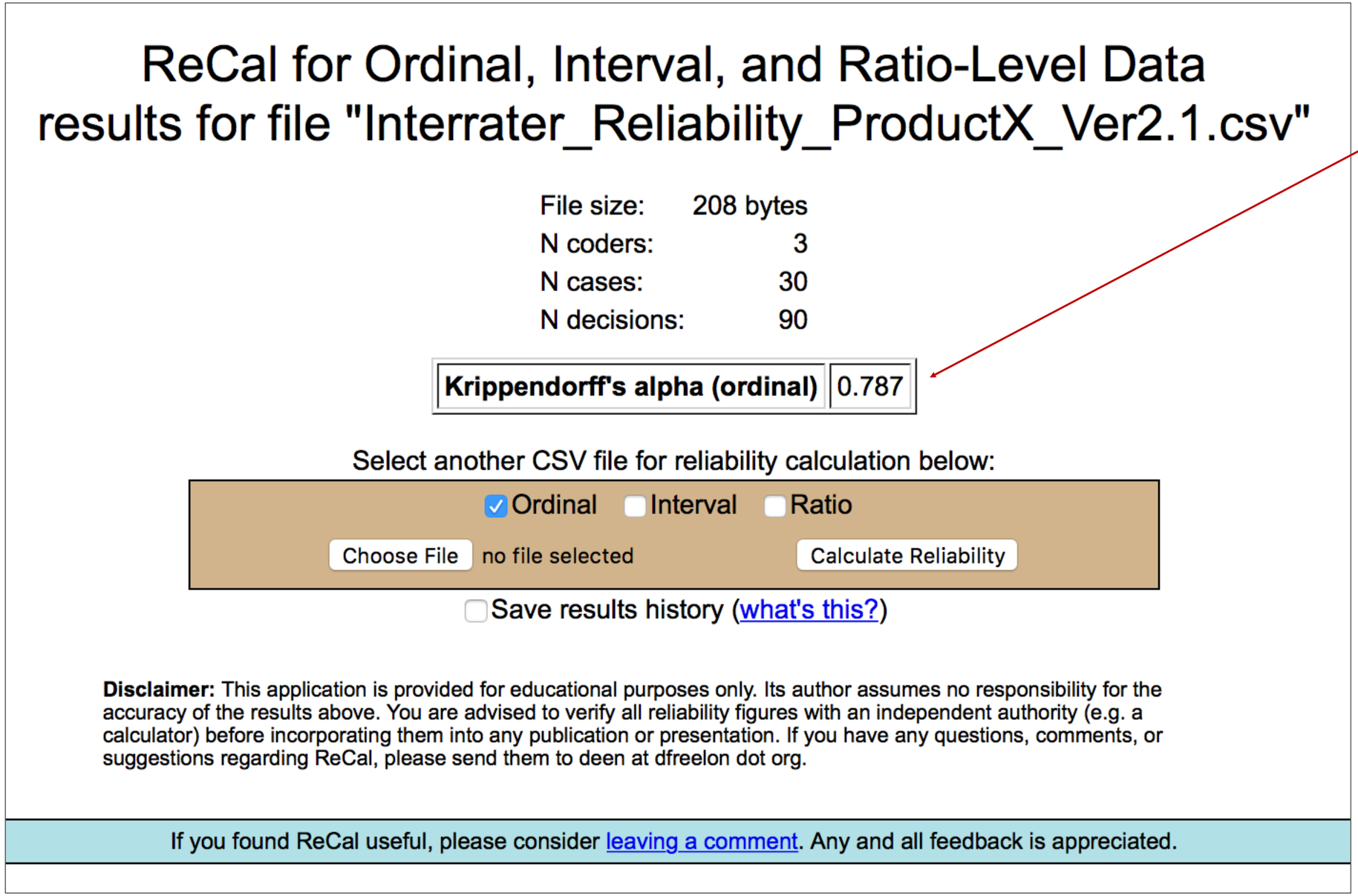


	A	B	C
1	1	1	1
2	1	1	1
3	3	2	3
4	1	1	1
5	2	2	1
6	1	1	1

- Save as a CSV
- Upload CSV file to ReCal OIR calculator: http://dfreelon.org/utils/recalfront/recal-oir/
  - Select “Ordinal” before uploading



Example output of ReCal for IRR



ReCal for Ordinal, Interval, and Ratio-Level Data
results for file "Interrater_Reliability_ProductX_Ver2.1.csv"

File size:	208 bytes
N coders:	3
N cases:	30
N decisions:	90

Krippendorff's alpha (ordinal)	0.787

Select another CSV file for reliability calculation below:

☑ Ordinal ☐ Interval ☐ Ratio

Choose File | no file selected | Calculate Reliability

☐ Save results history (what's this?)

Disclaimer: This application is provided for educational purposes only. Its author assumes no responsibility for the accuracy of the results above. You are advised to verify all reliability figures with an independent authority (e.g. a calculator) before incorporating them into any publication or presentation. If you have any questions, comments, or suggestions regarding ReCal, please send them to deen at dfreelon dot org.

If you found ReCal useful, please consider leaving a comment. Any and all feedback is appreciated.

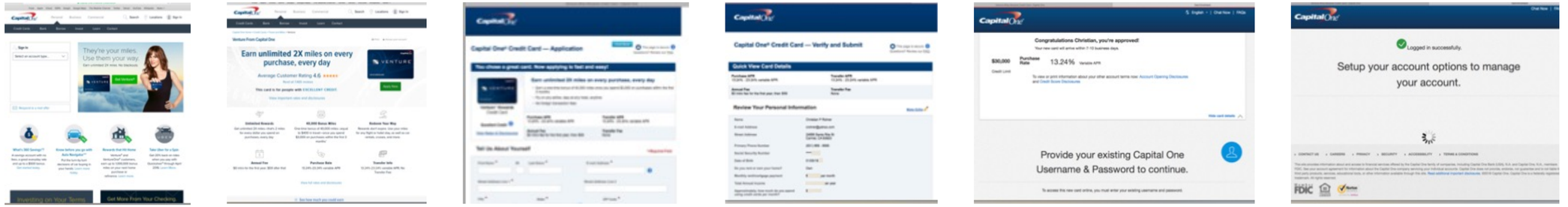
Output from ReCal OIR with inter-rater reliability at 0.787 (well above 0.5!)

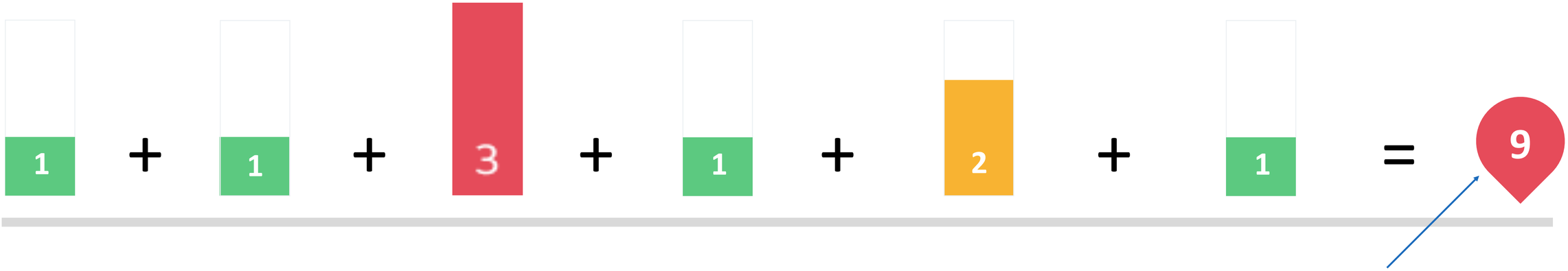
If we're reporting the "decided on" score, why do we need inter-rater reliability of the original scores?

- Calculating IRR on the original (silent) scores does two things:
  1. Helps us learn how to rate more consistently with each other
  2. Provides more methodological soundness to PURE
- The premise of any "rubric" is that it is objective enough to be applied consistently by **trained** raters
  - The rationale discussions helps your team establish a consistent understanding of the rubric, and IRR is your grade
- Once the rating team is trained (usually takes 2-3 times), they will be able to produce reliable scores time after time
- If our IRR is < 0.5 (using Krippendorff's alpha), the team is estimating the difficulty of the task too differently for this to be a reported.
- NOTE: We need a sufficient N (~25+), so include at least 2-3 tasks and all their steps before calculating it

Team Evaluation Step 4: Sum and color the PURE Scores - Tasks

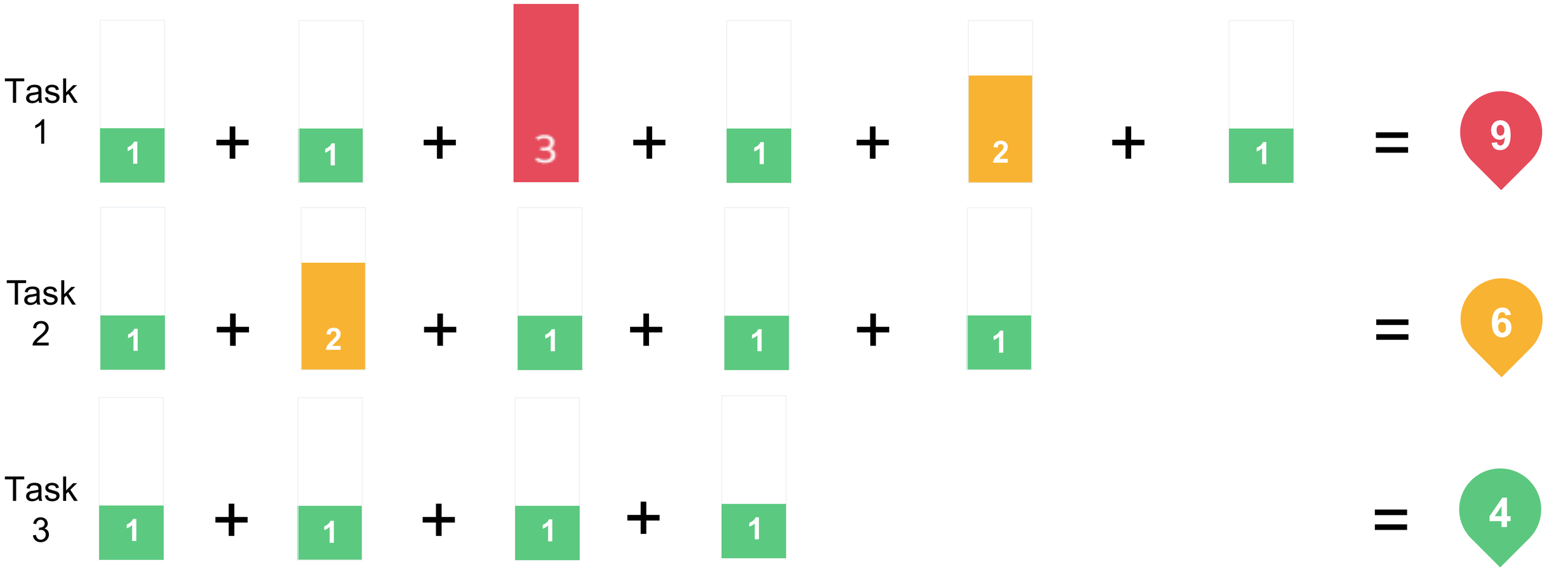
The sum of all ratings for a given Fundamental Task is the PURE Score for that task. Here is a PURE Task Score, along with associated colors & bars:





NOTE: PURE Task Scores are also assigned a color, based on the worst color of the components. One red or yellow score makes the whole task red or yellow. Like golf, smaller numbers are better and green is good.

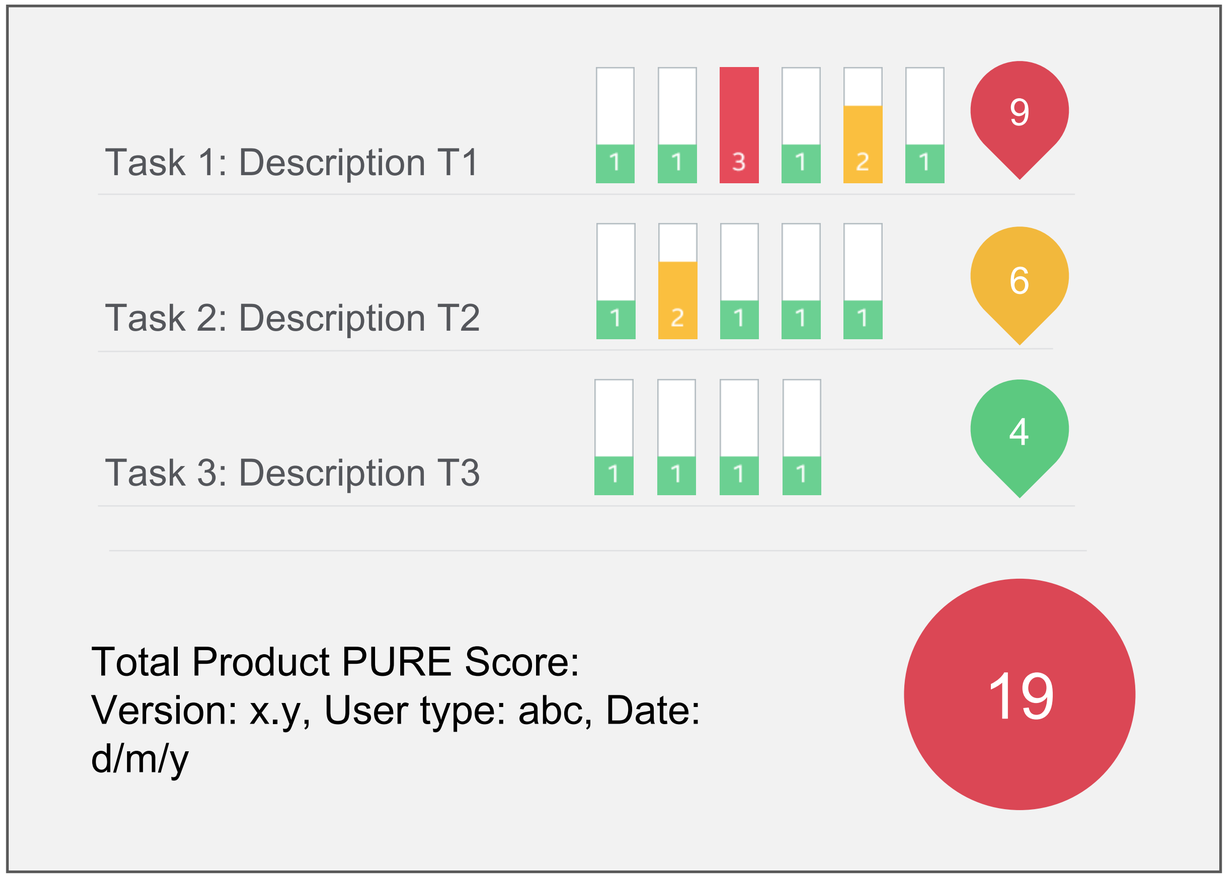
Team Evaluation Step 4: Sum Tasks Score for PURE Product Score



Total PURE Product Score = 19

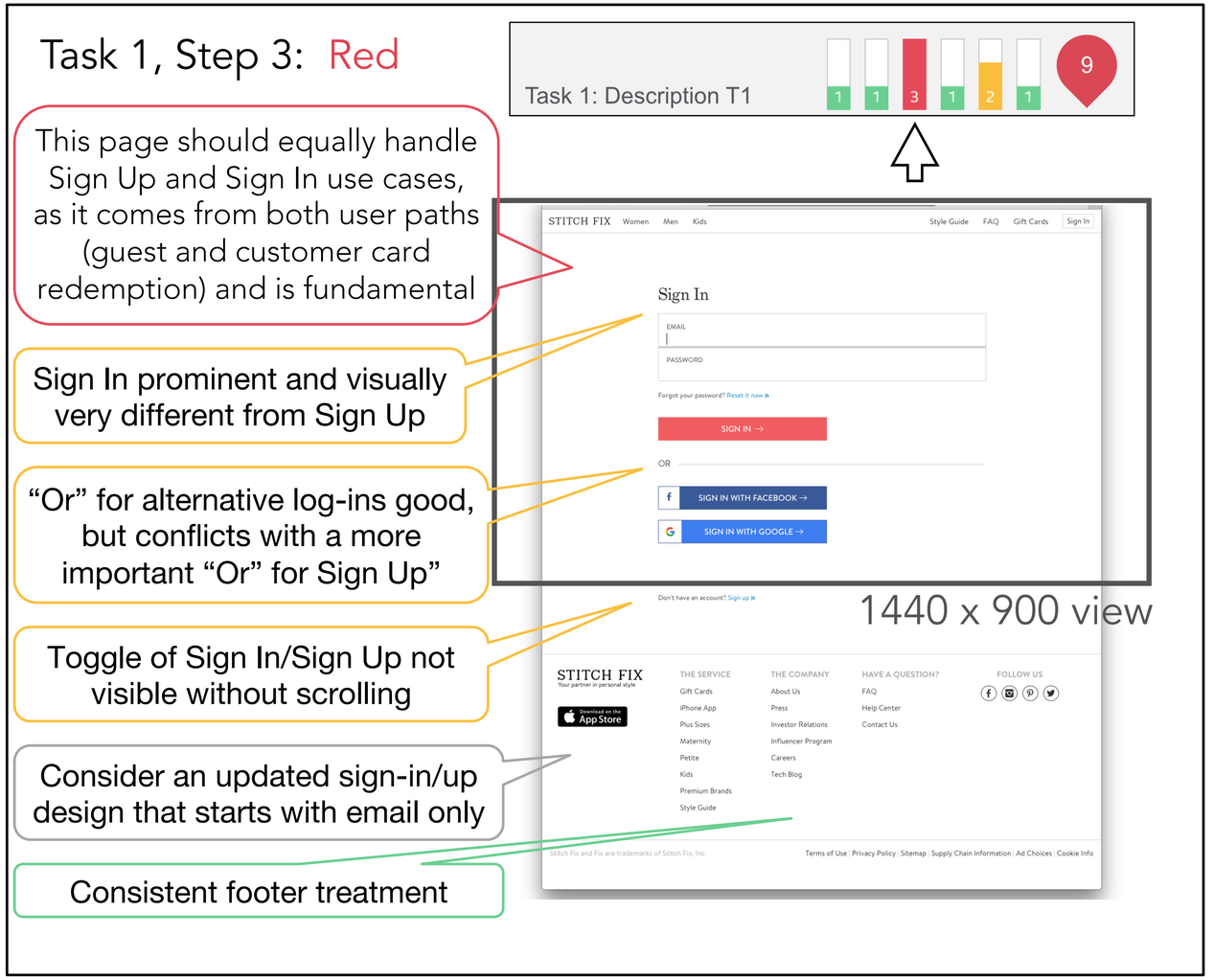
NN/g

Team Evaluation Step 4: PURE Product Score Card



TEAM EVALUATION STEP 5: CREATE THE DESCRIPTIVE INSIGHTS REPORT BEHIND EACH OF THE STEP SCORES

- Use 1+ slide per step w/callouts
- This is perhaps one of the most under-rated benefits of PURE
- Once the scorecard is shown, the impetus is to fix the problems
- The qualitative insights reports gives stakeholders details about why the scores are what they are, and possibly what can be done to address them



Suggested call out coloring

3: Red for severe issues (fail/bail)

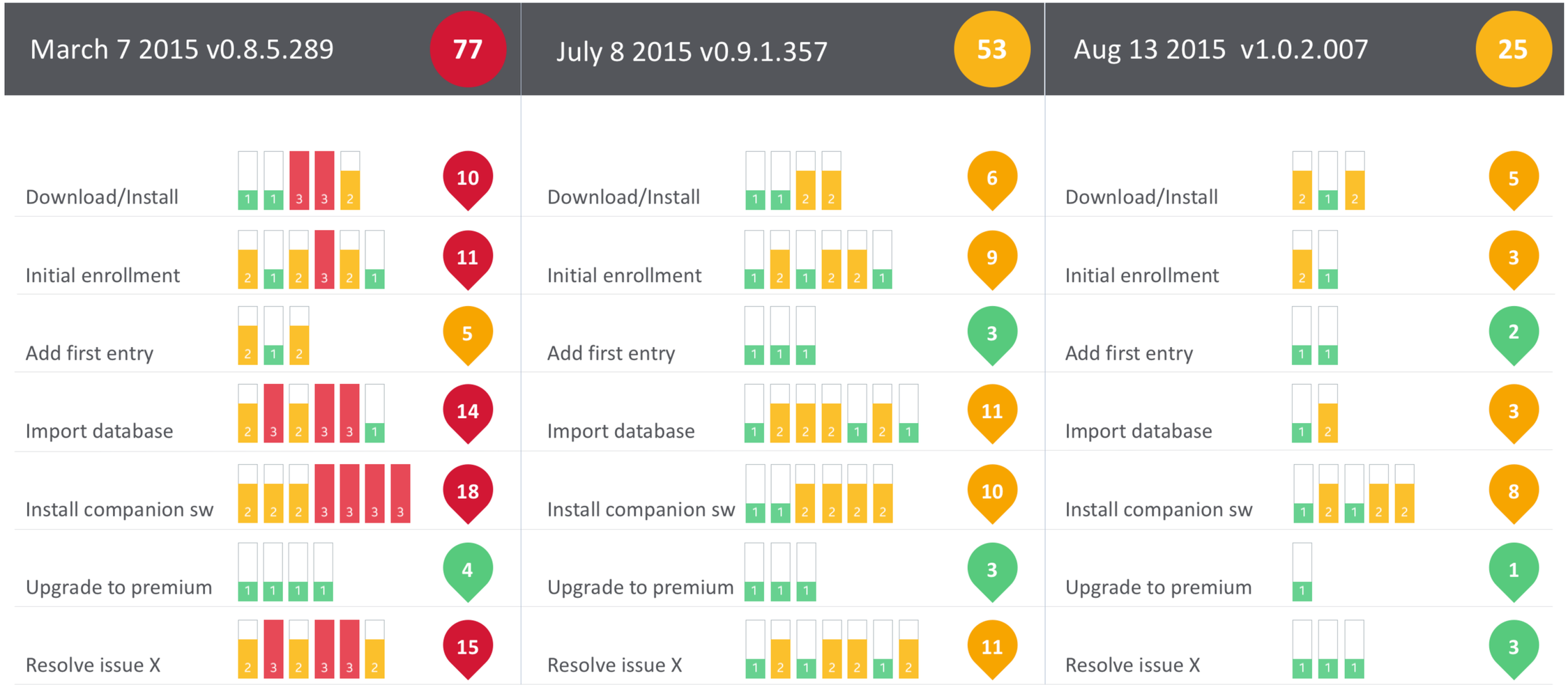
2: Yellow for "some" friction

1: Green for good UX practices

Grey for neutral comments/suggestios

NN/g

Leaders and Teams **want** to get better over time. An actual example:



Competitive reviews of Ratings also foster improvement

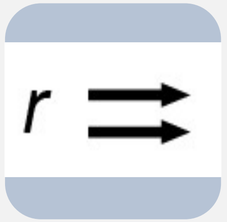
Windows vX.X.XXXX.X iOS vX..X | Android vX.X	Our Product	92	54	Competitor	Windows vX.X.X.XX | iOS vX.X.X Android vX.X.X-XX
Example Task 1*		NA		Example Task 1	2
Example Task 2	1 1 2 2 2 3 2	13		Example Task 2	2 3 3 1 1 1 3	14
Example Task 3 that goes on two lines	2 2 2 3 1 3 1 3 2	19		Example Task 3 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 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

PURE Method Reliability & Validity (courtesy of MeasuringU)

Validity: is same issue found in other studies?



Convergent Validity
(comparing results from PURE to results from a traditional benchmarking study)

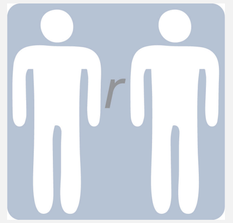
220 Users 3 Products + 8 Websites

SEQ: r = .5

SUS/SUPRQ r = .4

Completion/Time: r < .2

Reliability: can the issue be consistently found?



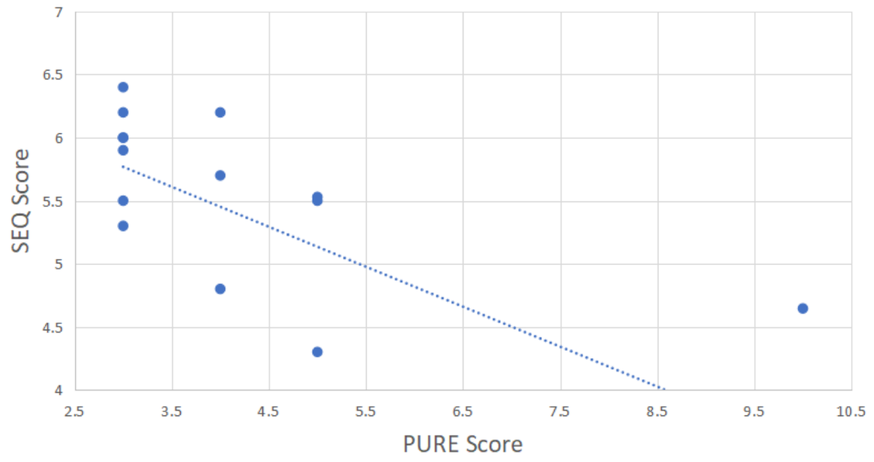
Inter-rater Reliability
(calculated both among raters inside the same company and across raters at the company and at the agency MeasuringU)

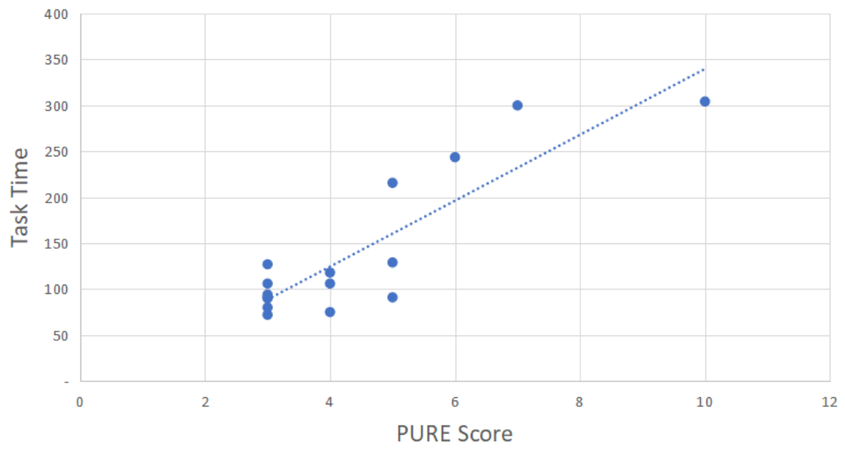
r = .5 to .9

Validity: A classic usability benchmarking study was conducted by MeasuringU and values for attitudinal measures (SEQ, SUS, SUPRQ) and behavioral measures (completion time) were gathered. Then, a PURE score was conducted by MeasuringU researchers, and the results were compared. **Reliability:** MeasuringU researcher PURE scores and Intel researcher PURE scores were compared.

Results from more recent study

Replication Validation Study





313 Users, 6 Banking + 4 Health Insurance Websites

Predicting UX Metrics with the PURE Method. https://measuringu.com/ux-metrics-pure/

Results from more recent study (cont.)

Replication Validation Study: Stronger Evidence for Predictive Validity

Metric	*r*
Completion Rate	-0.21
Time	0.88
Confidence	-0.72
SEQ	-0.67
SUPR-Q	0.55

PURE Scores can explain (or predict) about 45% of the variation in SEQ scores

77% of the variation in Time.

30% of overall UX quality perception.

Predicting UX Metrics with the PURE Method. https://measuringu.com/ux-metrics-pure/

A note about Empirical data vs. PURE

- It is **always** better to use solid empirical data, if you can do so with relatively the time and effort available to have a real impact.
- Having qualitative data is always better than nothing (and is sometimes better than some kinds of quantitative data)
- But examining an experience systematically with PURE isn't "nothing". We use:
  - Collective experience (past research, knowledge of UX principles) from raters
  - Considering multiple viewpoints on the rating team, as objectively as possible
  - Team learning and continuous improvement
- The Insights behind the scores can be an excellent starting point, either for obvious fixes or deeper, empirical studies. Better yet, just go fix the product!

QUESTIONS?

- You may now ask questions from this section
- You may type your question into Zoom Chat
- We may have to cover some questions as follow-ups after the seminar

EXERCISE: Let's try rating a product experience together with PURE

- **Target User**: Weather Prepared Pat
  - Pat is very interested in the weather, and wants to be prepared for what may happen
  - Pat has friends in other cities and visits often; likes to know what they are experiencing
  - Pat is reasonably web savvy, and has had a smartphone for 9 years and is proficient with it
- There are **2 Fundamental Tasks** (I've recorded these from a Smartphone for you):
  1. Check the **weather at current location** (assume first time open after download of app)
  2. Check the **weather in Paris 9-10 days from now**
- Evaluate both task 1 and 2 of the following products (if time is short, focus on task 2):
  - Product A: Weather Bug (as a class)
  - Product B: Weather Puppy (as a class)
- I will share the video recordings of these experiences through Zoom, but you may also examine the pre-recorded screen recordings in the documents folder: http://bit.ly/nng-ux-assessments
- We will review these flows together as a class on Zoom, and then go through and rate each step
- You can record your ratings and rationale on paper, in a document, or using a PURE scoresheet

EXERCISE: Small Group PURE Evaluation (if time permits or as HW)

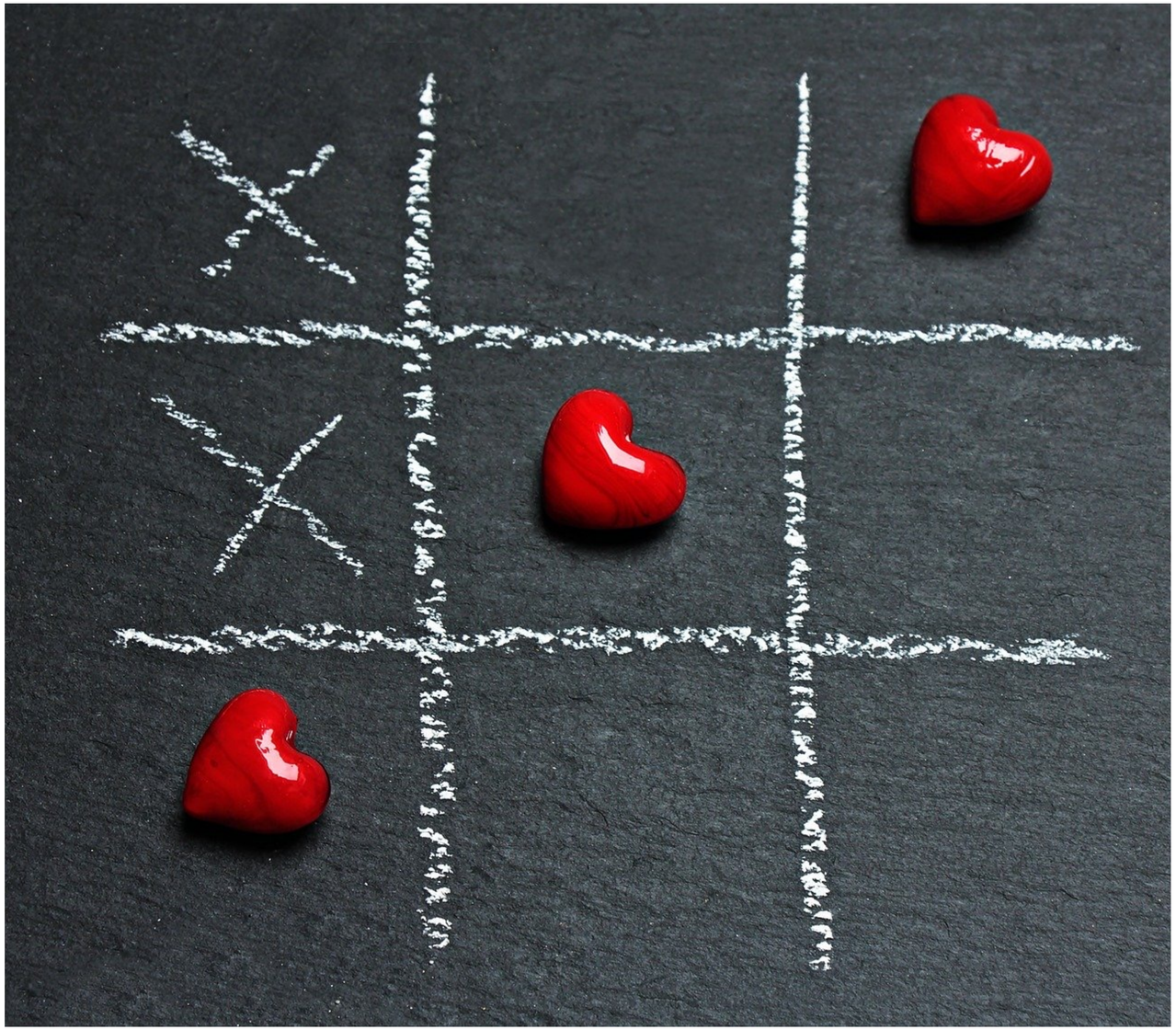
- **Target User**: Weather Prepared Pat
  - Pat is very interested in the weather, and wants to be prepared for what may happen
  - Pat has friends in other cities and visits often; likes to know what they are experiencing
  - Pat is reasonably web savvy, and has had a smartphone for 9 years and is proficient with it
- There are **2 Fundamental Tasks** (I've recorded these from a Smartphone for you):
  1. Check the **weather at current location** (assume first time open after download of app)
  2. Check the **weather in Paris next week**
- You will evaluate in small groups of 4 (or 3). We will look at websites, since these are easier to screen share in Zoom.
- Evaluate both task 1 and 2 of the following products (if time is short, focus on task 2):
  - Product A: Weather Puppy (smartphone app)
  - Product B: Weather Bug (smartphone app)
- You can examine and play with these sites on your own, but for the purposes of evaluation, use the pre-recorded videos of tasks 1 and 2 in document folder: http://bit.ly/nng-ux-assessments
- You can record your ratings and rationale on paper, in a document, or using a PURE scoresheet

QUESTIONS?

- You may now ask questions from this section
- You may type your question into Zoom Chat
- We may have to cover some questions as follow-ups after the seminar

PURE Summary

- PURE is not a substitute for quantitative measures of usability/UX
- Conducting PURE and measuring usability can greatly supplement other studies
- Qualitative studies strongly inform PURE Ratings
- Business leaders are obsessed with metrics, so **give them to them**
- End result: more improvements to UX and usability



Even if PURE is not accepted as a valid measure, the conversation should **now** shift toward what better (empirical) measures of Ease of Use (and other aspects of User Experience) you should now fund and deploy.

Any focus on UX is a win.

Thank you!

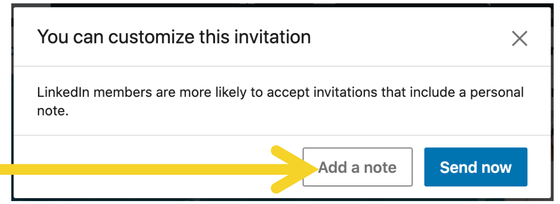
Christian Rohrer

Connect with me!

Twitter: @christianrohrer

LInkedIn:
https://www.linkedin.com/in/crohrer/

Suggestion: Please include a note in your LinkedIn invitation and indicate that you were in this course, so we have a record of how we met



References

- Bi, Xiaojun et. al (2013): "FFitts Law: Modeling Finger Touch with Fitts' Law" [PDF]
- Card, S.K., Newell A., Moran, T.P. (1983): *The Psychology of Human Computer Interaction*
- Card, S. K., Moran, T. P., & Newell, A. (1980): "The keystroke-level model for user performance time with interactive systems" in Communications of the ACM, **23**(7), 396-410
- El Batron, K & Dunlop, M.D., "Enhancing KLM (Keystroke-Level Model) to Fit Touch Screen Mobile Devices" MobileHCI'14, Toronto, ON, Canada, September 23-26, 2014
- Engelbeck, G. E. (1986): *Exceptions to generalizations: Implications for formal models of human–computer interaction*. Master's thesis, Department of Psychology, University of Colorado, Boulder, CO.
- Fitts, Paul (1954): "The Information Capacity of the Human Motor System in Controlling the Amplitude of Movement" in Journal of Experimental Psychology, **47**, 381-391 [PDF]
- Fitts' Law Calculator: https://codepen.io/gsus/details/QyKBGy
- Freelon, Dean. "ReCal: reliability calculation for the masses" http://dfreelon.org/utils/recalfront/recal-oir/
- Gray, W.D., Salzman, M.C. (1998). "Damaged Merchandise? A review of experiments that compare usability evaluation methods." Human-Computer Interaction, 11, pp. 203-261.
- Greeno J.G. & Simon, H.A. (1988): "Problem Solving and Reasoning" in Steven's Handbook of Experimental Psychology
- Hertzum, Morten & Jacobsen, Niels. (2003). "The Evaluator Effect: A Chilling Fact About Usability Evaluation Methods." Int. J. Hum. Comput. Interaction. 15. 183-204. 10.1207/S15327590IJHC1304_05.
- Hollingsed, T and Novick, D (2007). "Usability inspection methods after 15 years of research and practice" in Proceedings of the 25th annual ACM international conference on Design of communication (SIGDOC '07). ACM, New York, NY, USA, 249-2 5 [PDF]
- Kieras, David (2001): "Using the Keystroke-Level Model to Estimate Execution Times" [PDF]
- Mahatody, T., Sagar, M., & Kolski, C. (2010). "State of the art on the Cognitive Walkthrough method, its variants and evolutions. International Journal of Human-Computer Interaction, 26(8), 741-785.
- Newell, A & Simon, H. (1972): *Human Problem Solving*
- Nielsen, J and Mack, R.L. (1994): *Usability Inspection Methods*
- Nielsen, J., and Molich, R. (1990). "Heuristic evaluation of user interfaces" in Proceedings of the Conference on Human Factors in Computing Systems (CHI 90), Seattle, WA, April 1-5, 1990, 249-256
- Polson, X and Lewis, Y (1990): "Theory Based Design for Easily Learned Interfaces" in Human-Computer Interaction **5**: 191-220
- Rohrer, Christian P. (2000): *Joint Exploratory Learning: A Case Study of How People Learn to Use Interactive Technology Together.* Stanford University PhD Dissertation
- Spencer, Rick (2000): "The Streamlined Cognitive Walkthrough Method, Working Around Social Constraints Encountered in a Software Development Company." CHI 2000

Appendix:
TYPICAL PURE METHOD SETUP CHALLENGES & REMEDIES

Frequently Asked Questions

Q: Why do we specify the Target User?

A: In order to reduce the variance around how raters would interpret the PURE rating for a given step. This can vary greatly by user.

Q: Why do we only look at the Fundamental Tasks?

A: In order to have a consistent baseline measure **and** to force the team to prioritize what matters most.

- You can choose to apply PURE to other tasks, even all tasks, if you plan to re-score and compare later.

Q: Why don't we deviate from the Happy Path?

A: In order to have a consistent baseline measure and to show that we're rating "our best shot" at solving a user problem.

- Once this is done, you can choose to apply a PURE rating/score to any flow, but it probably shouldn't be part of an official PURE score for the product

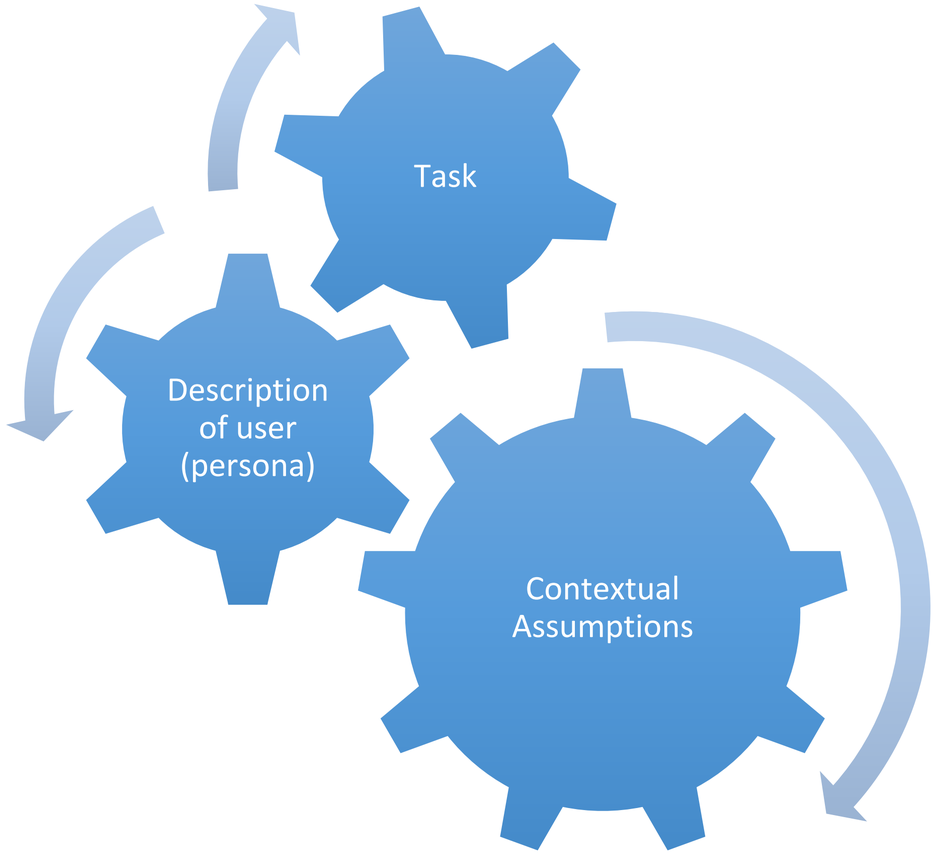
Q: Why do all raters watch the same thing?

A: Unlike Heuristic Evaluation, where you are looking for a wide number of problems, here we are trying to get a score we can reliably count on.

- If raters don't see the same thing, they will undoubtedly vary, not based on how hard it was, but based on what they happened to see

Proper set up and framing is important for an efficient evaluation

Teasing out the assumptions prior to the evaluation ensures stakeholders are aligned around the most common tasks/flows and persona, leading to a more efficient evaluation process down the line. Being more concrete and explicit about the context, the scenario, path, and steps, eliminates the guess-work during the evaluation to keep things moving properly.



Task: Make a mobile check deposit using the Capital One Mobile Banking App

Persona: User who primarily manages their account online. Occasionally uses their smartphone app to check transactions on the phone and pay bills, but has has never deposited a check using mobile deposit

Contextual Assumptions: This is very important for setting up the context and narrowing down the scenario

Typical kickoff discussion themes and how a skilled Evaluator can tease out assumptions about our task and persona

Evaluator: I want to make sure we clearly call out the experience and the happy path we believe our user takes so we can properly set up the session

Product Manager: Keep in mind that not everybody sees the same experience, it depends on the account type and the eligibility requirements of the user. Different account types, such as Small Business, may have a different flow.

Evaluator: Great, let's concentrate on a 'consumer persona'.

Product Manager: We also need to think about users eligibility to make a deposit. That is driven by compliance, legal, and fraud, so it depends on the user and what they are trying to do at a given point. So what bucket should we look at?

Evaluator: let's assume that the user doesn't fall in any risk-buckets that prevent him from making a deposit. We'll concentrate on users who wish to deposit and have no specific restrictions

Product Manager: What about users who try to deposit via a tablet, or Android phone, the UX for Android is slightly different.

Designer: Also, we should focus on the first time experience. Those folks will see the coaching and help screens, and have to agree to the terms and conditions first time and we're aggressively marketing to get new folks in the door.

Evaluator: Great, if we can all agree, we'll narrow down the folks to reign it in

Evaluator: We narrowed it down to Mobile app users with an iPhone who are depositing a check for the first time into their personal checking account.

Designer: Do we want them to read through the all the onboarding screens and tool tips since that's part of the experience?

IT Lead: Let's narrow it to a user who skips the onboarding screens, data tells us that only 20% spend time looking at those screens. We'll factor in the Terms and conditions screens since that's a requirement for first time users.

Designer: There are 3 different ways you can get to Mobile deposit, from the mobile landing page, from the overflow menu and from the accounts page.

Typical kickoff discussion themes and how a skilled Evaluator can tease out assumptions about our task and persona (cont.)

Evaluator: Based on analytics, which in the most commonly used? We'll pick the one most used. Ok, so we've narrowed it to entry from within the Accounts Page, since that's the most common entry point.

Product Manager: Since you have to be logged in, are you also going to start from the 'Sign In' experience? There are several ways users can sign in in and depending on that experience the number of steps will vary

Evaluator: Great question, we're going to assume the user is an active Capital One customer, and knows their UN/PW.

IT Lead: In order to replicate the experience, you'll need the credentials to the test account:

IT Lead: You won't be able to actually submit the check and get final confirmation, because the back-end system for the UAT site, is currently on lock-down, so you can't actually submit, will that create a problem?

Evaluator: Got it. Can you provide a clickable prototype or a set of screenshots that simulates what happens after a user actually submits the check? We will be able to do it, but note that during the evaluation

Evaluator: Going through the experience, I notice that when I click the 'Mobile Deposit' button, it asks for access to my camera and GPS. When I go to allow it, it takes me through the phone's setting experience. Do we want to capture this experience as well, or assume the user has access to GPS and camera activated in settings?

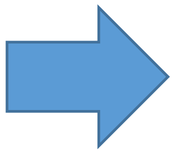
Product Manager: Let's assume user previously allowed access to GPS and has camera turned on. Data tells us that most of our folks have that setting turned on when they download our app

As you can see, you have to have detailed discussions ahead of time and DOCUMENT YOUR ASSUMPTIONS! This is so that you can follow the same assumptions in future PURE evaluations, if you are comparing scores.

Proper set up and framing is important for an efficient evaluation. Here's a real-life example of a task and persona specified, and how one team was able to tease out assumptions and define context

Task: Make a mobile check deposit using the Capital One Mobile Banking App

Persona: User who primarily manages their account online. Occasionally uses their smartphone app to check transactions on the phone and pay bills, but has has never deposited a check using mobile deposit



Assumptions we teased out in the conversation:

- Evaluating first time user experience for a personal account user who's never deposited a check with any financial institution
- User will skip initial "coaching" screens
- Using iPhone 6S w/GPS and camera app access already 'turned on'
- Persona is using their username and password to log in, and their username is stored on their device
- Persona is not bound by account restrictions
- Persona is entering experience via 'Account' page, but we'll start at the Home Page
- We are not able to 'submit' at this point; we will evaluate the 'Verification' and 'Confirmation' pages using screenshots and will not be able to really interact

Typical PURE LEAD Duties

- Document assumptions and decisions the team has made on:
  - Target User type
  - Fundamental Task Choices
  - Happy Path specification and choices
  - Why the Target User type is following this happy path
- Run the session so all raters can see the experience reasonably well
- Let raters know when a task is beginning and ending
- Let raters know when a step is beginning and ending
- Facilitate discussion afterwards
  - Counterbalance who shares their ratings first, last, etc.
  - Record and later report inter-rater reliability score