HCI:
LAB STUDIES

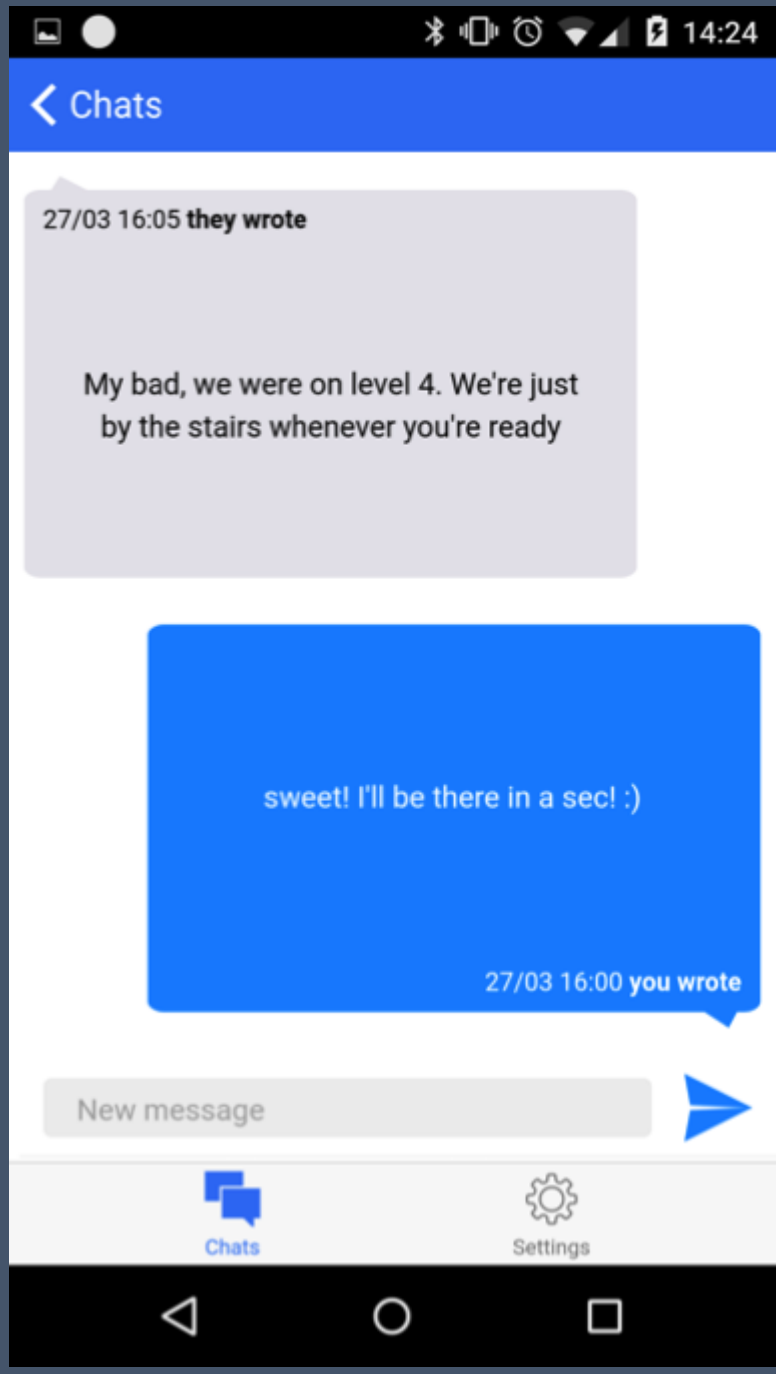
Dr Kami Vaniea

First, the news…

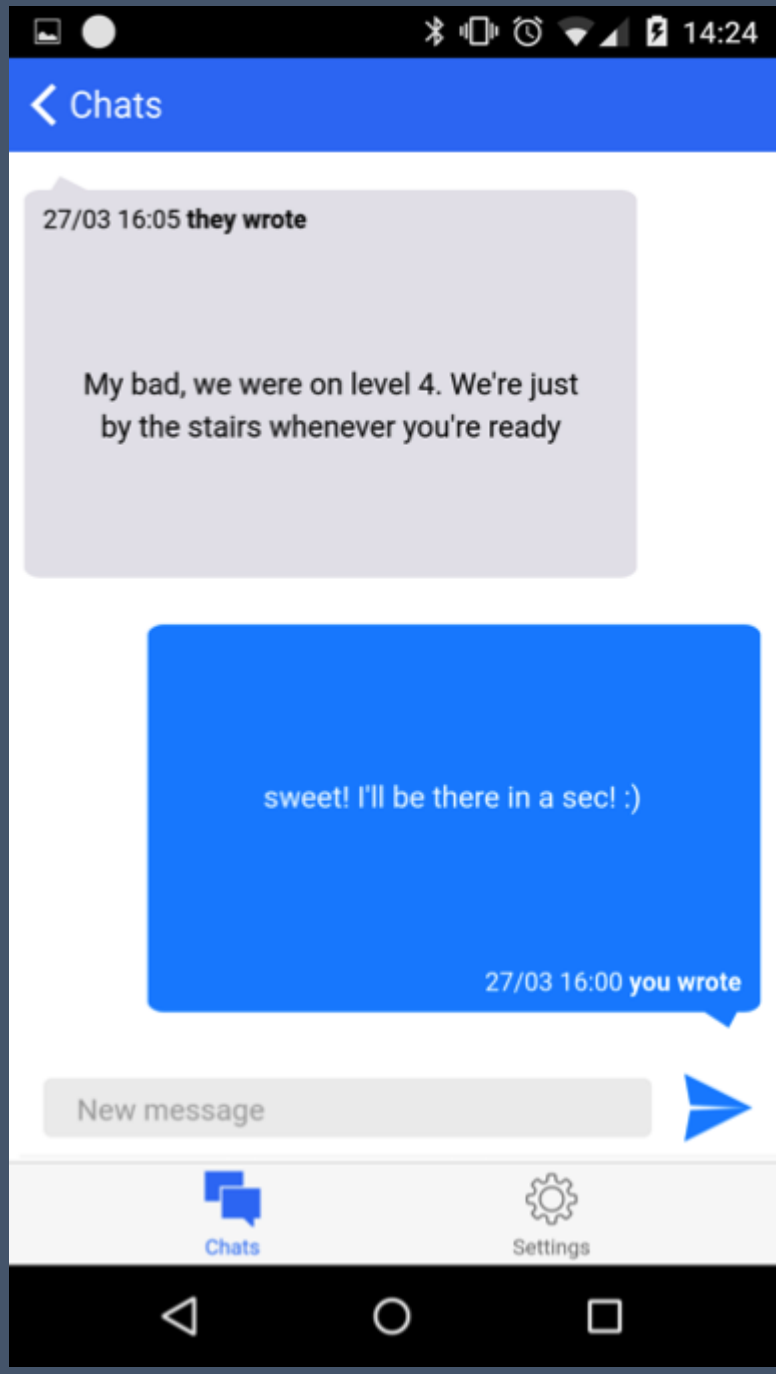
- Forget-me-not: History-less mobile messaging
- Mattias Rost, Christos Kitsos, Alexander Morgan, Martin Podlubny, Pietro Romeo, Edoardo Russo, and Matthew Chalmers. 2016. Forget-me-not: History-less Mobile Messaging. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI '16). ACM, New York, NY, USA, 1904-1908. DOI: http://dx.doi.org/10.1145/2858036.2858347

What if a mobile conversation was more like a verbal one?

The Forget-me-not app only shows you the last thing you and they said, no other history.



“We ran a trial with 10 participants for 2 weeks. The participants were recruited from two groups with existing social ties.”



Lab Studies

Usability testing & research

Usability testing

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

Experiments for research

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

Many ways to test usability

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

Lab studies are a simple idea. You ask a user to come into a physical space and ask them to interact with the interface there.

Lab Study

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

Mixed-methods study

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

Pre/post questionair

1. Participant fills out a questionair immediately after they complete informed cosent.
2. They engage with the content of the study (do stuff)
3. Participant fills out a very similar or identical questionair
4. The researcher comparres the answers in the pre and post questionairs to determine if the content of the study had an impact on the participant

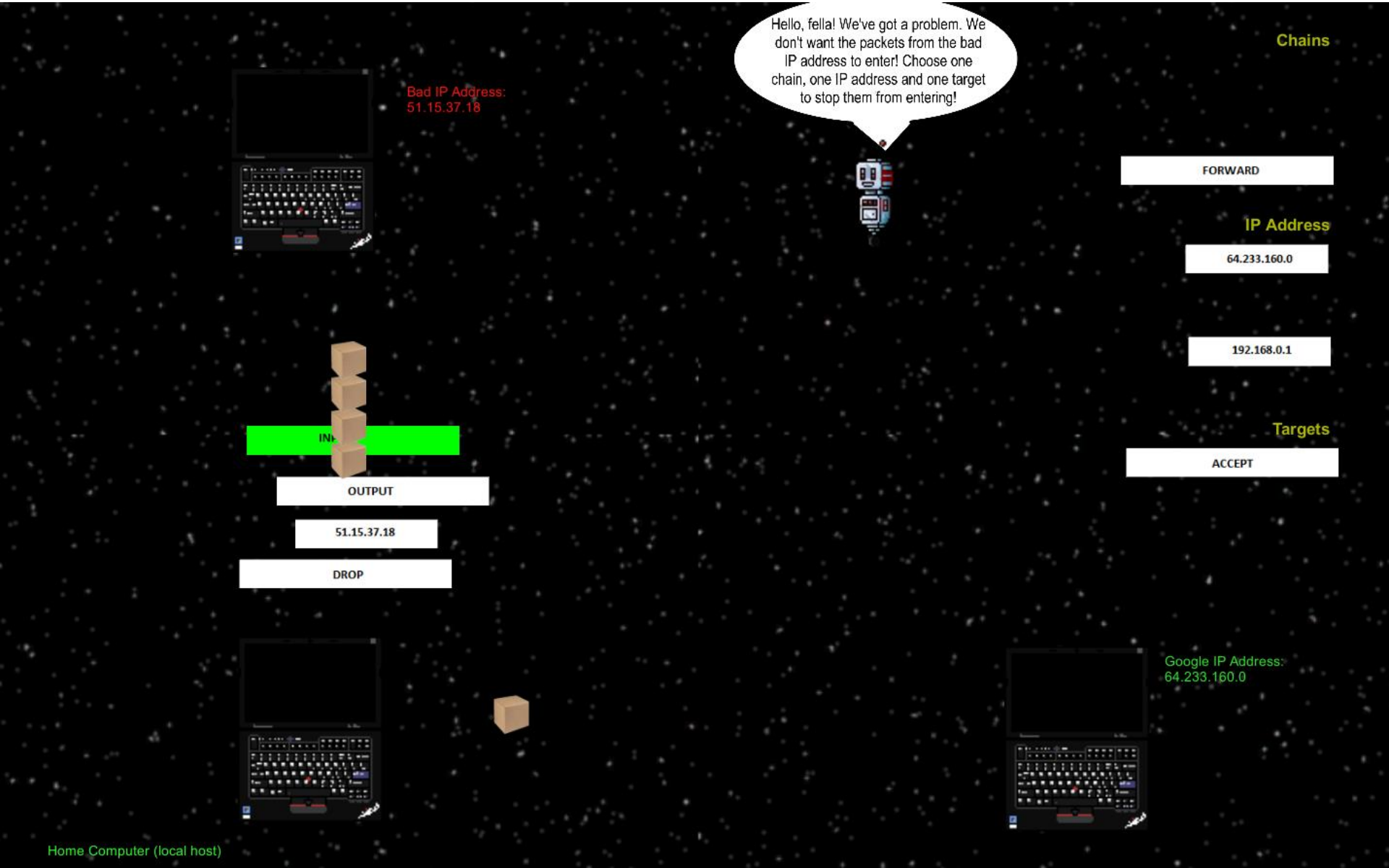
Pre/post questionair

- Give the participant a questionair before and after the main study
- Pros
  - Easy way to determine if the study had an impact on the participant
  - Very useful in education or attitude chaning studies
  - Easy to compare
- Cons
  - Learning effect – if you give the same questionair pre/post the participant may have just thought more about the answers
  - Often need to compare to prove that the results are not just a learning effect

Permission Impossible
by Sibylle Sehl

Early lab test done as part of an MSc project

Is this screen usable?



Lab test of Security Game

1. Informed Consent
2. Pre-questionair
3. Play the game
4. Post-questionair
5. Post discussion with participants (mini focus group)

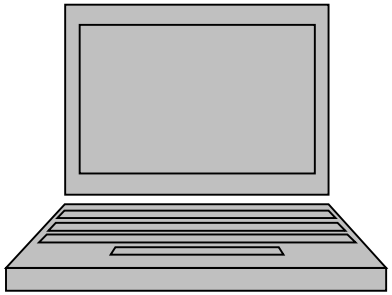
Betrayed by Updates: How Negative Experiences Affect Future Security

by Kami Vaniea, Emilee Rader, and Rick Wash in CHI 2014

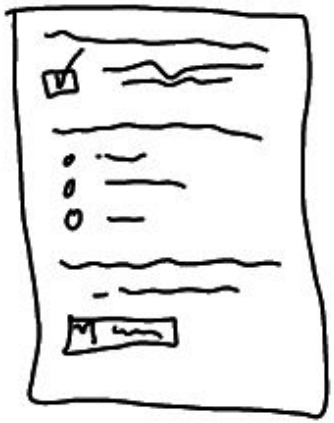
Why are people not updating software?

Multi-methods approach

Settings and logs



Survey



Interview

- Kami Vaniea, Emilee Rader, and Rick Wash; Betrayed by Updates: How Negative Experiences Affect Future Security, CHI 2014
- Rick Wash, Emilee Rader, Kami Vaniea, and Michelle Rizor; Out of the Loop: How Automated Software Updates Cause Unintended Consequences, SOUPS 2014

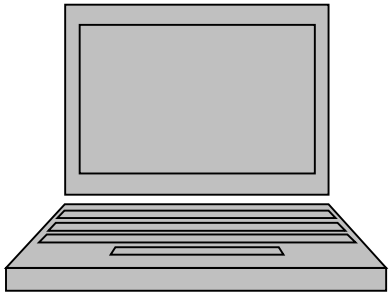
Participants

- 37 non-technical graduate students
  - Aged 21-57 (Mean: 31)
  - 17 were male
- Advertised via emailing a random sample of PhD students

Protocol

1. Informed consent
2. Start data collecting from participant's computer (Powershell script)
3. Participant fills out survey on computer
4. Interviewer conducts semi-structured survey with participant
5. Technician analyzes data and provides Interviewer with report about computer
6. Interviewer shows participant report and discusses

Settings and logs



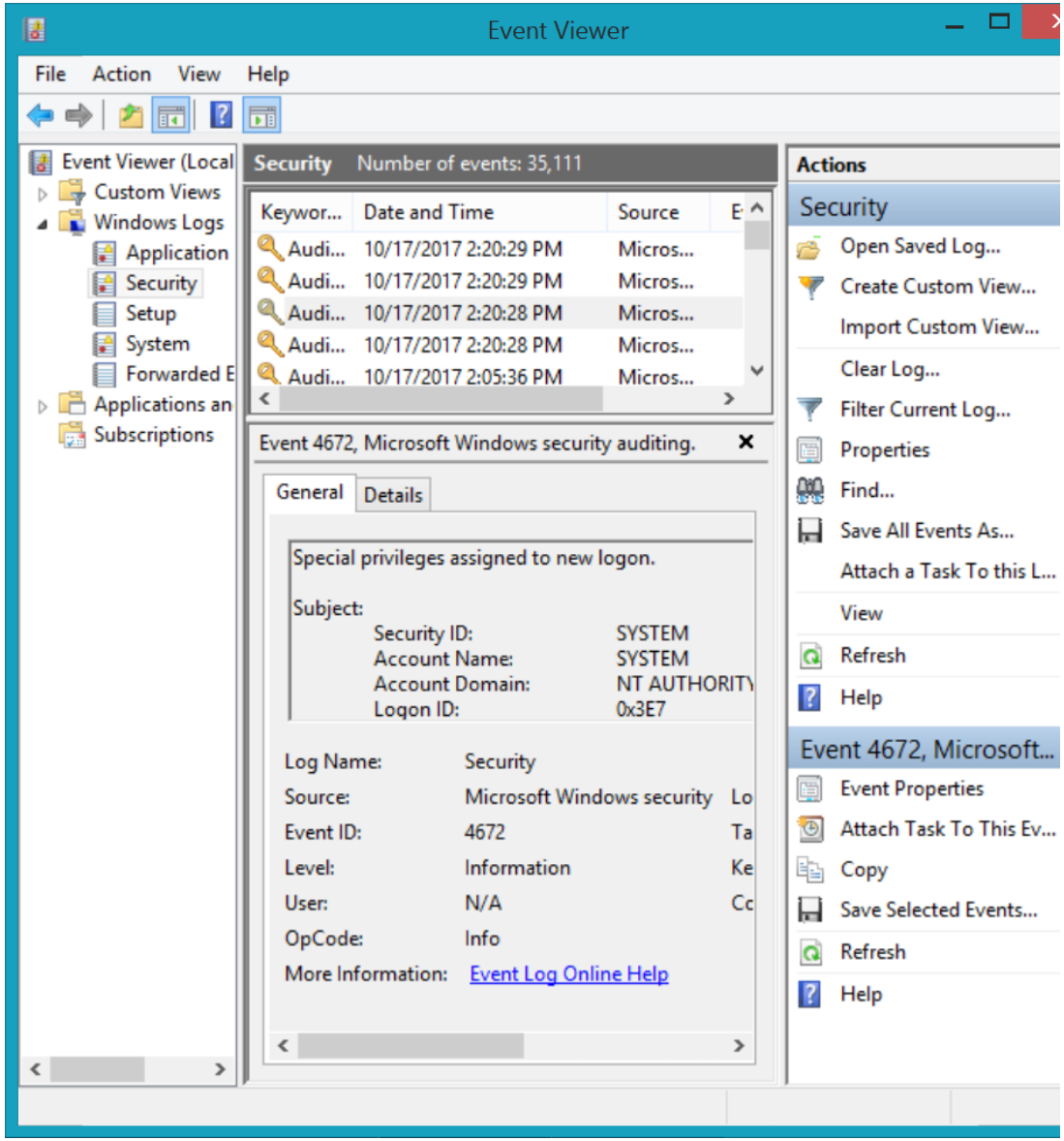
Survey

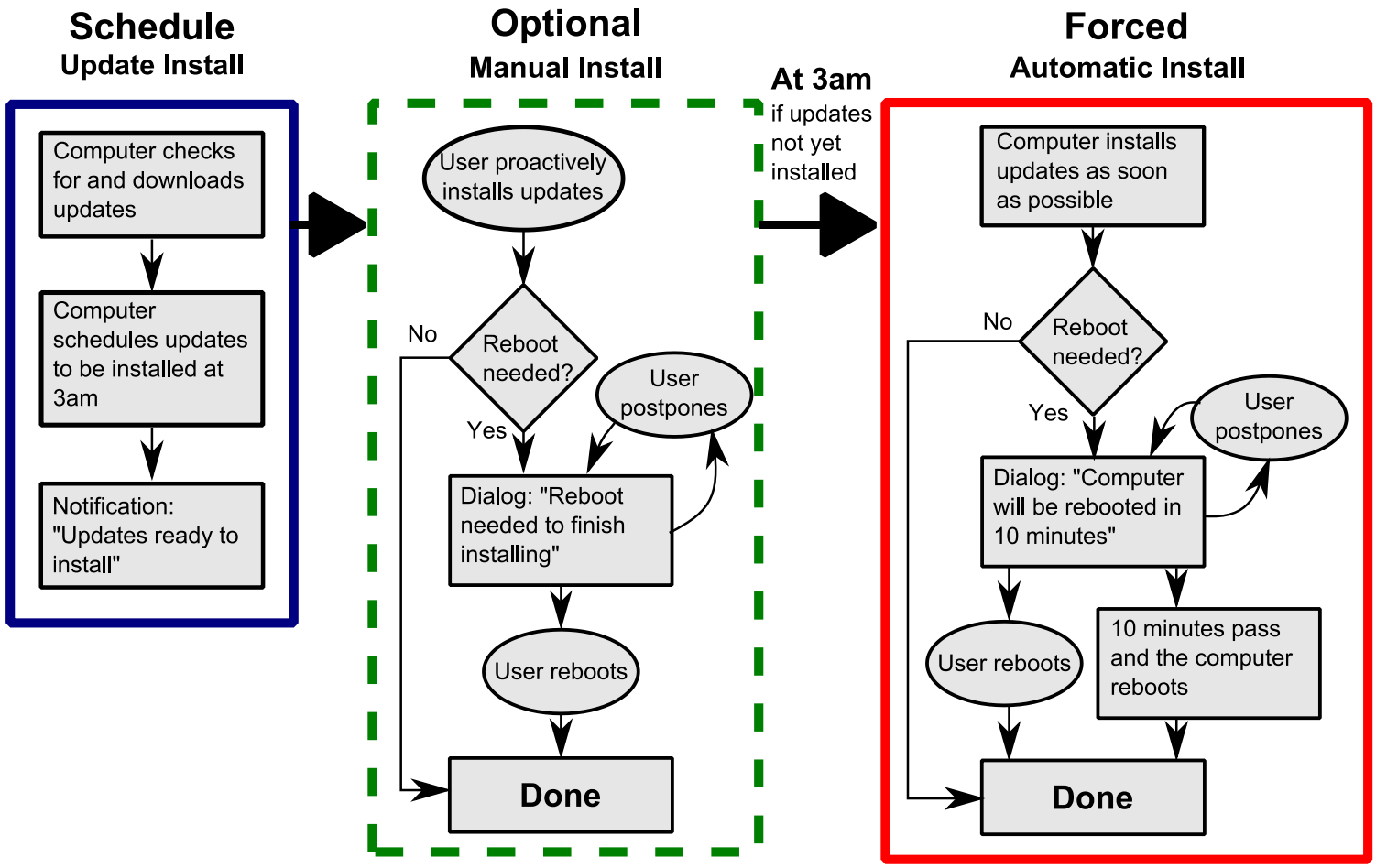
Interview

- Rick Wash, Emilee Rader, Kami Vaniea, and Michelle Rizor; Out of the Loop: How Automated Software Updates Cause Unintended Consequences, SOUPS 2014
- Kami Vaniea, Emilee Rader, and Rick Wash; Mental models of software updates, International Communication Association

Settings and Logs

- Powershell script to gather settings and log files
- Ruby script to parse logs into human-recognizable events
- Latex to create user-facing report




Schedule
Update Install
Computer checks for and downloads updates
Computer schedules updates to be installed at 3am
Notification: "Updates ready to install"
Optional
Manual Install
User proactively installs updates
Reboot needed?
No
Yes
User postpones
Dialog: "Reboot needed to finish installing"
User reboots
Done
At 3am
if updates not yet installed
Forced
Automatic Install
Computer installs updates as soon as possible
Reboot needed?
No
Yes
User postpones
Dialog: "Computer will be rebooted in 10 minutes"
User reboots
10 minutes pass and the computer reboots
Done

Participants

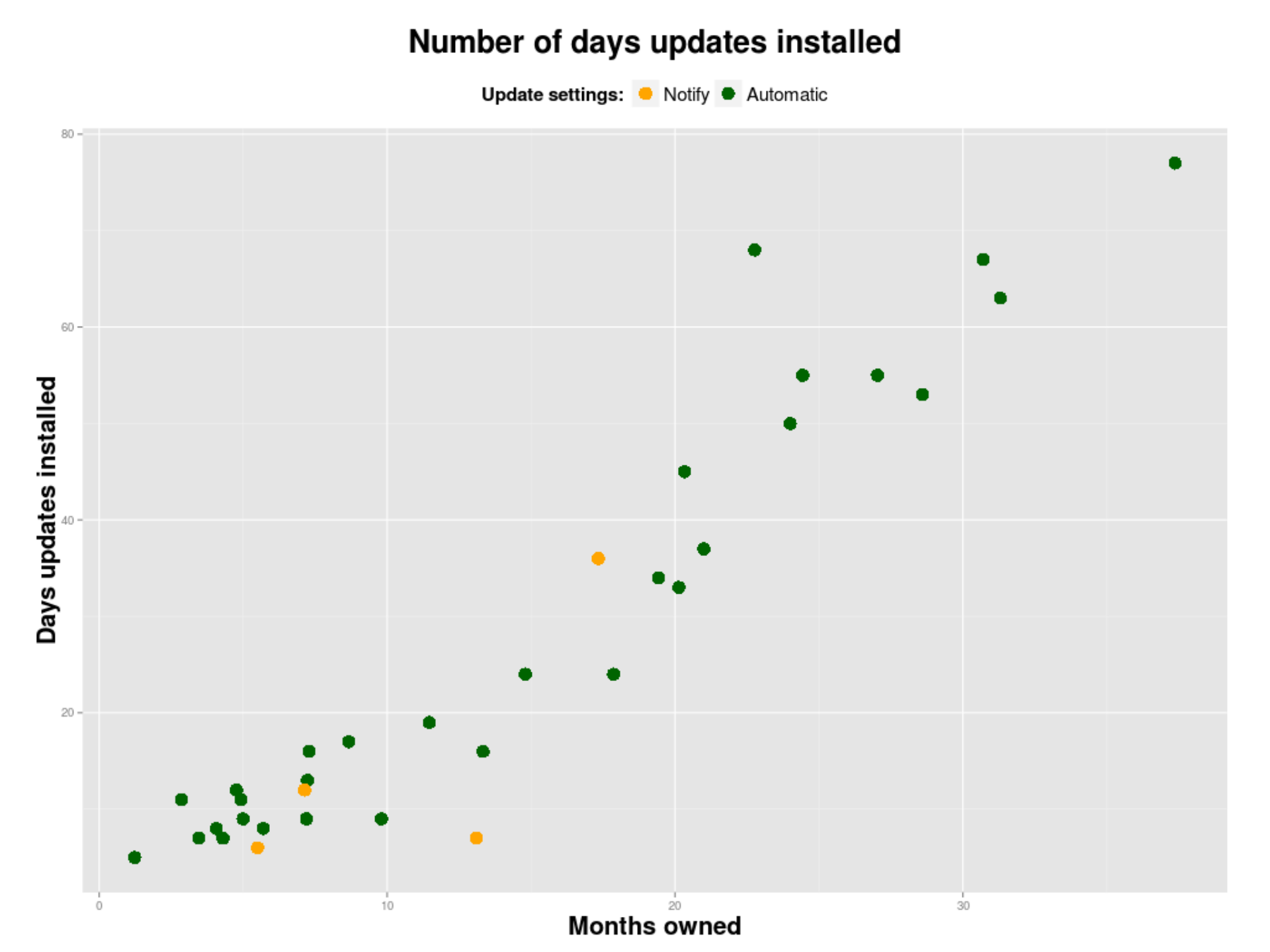
- 15 participants had at least one security issue
  - 6 - Anti-virus not installed or disabled
  - 2 - Firewall off
  - 7 - User account control off or limited
  - 6 - Windows updates off or limited

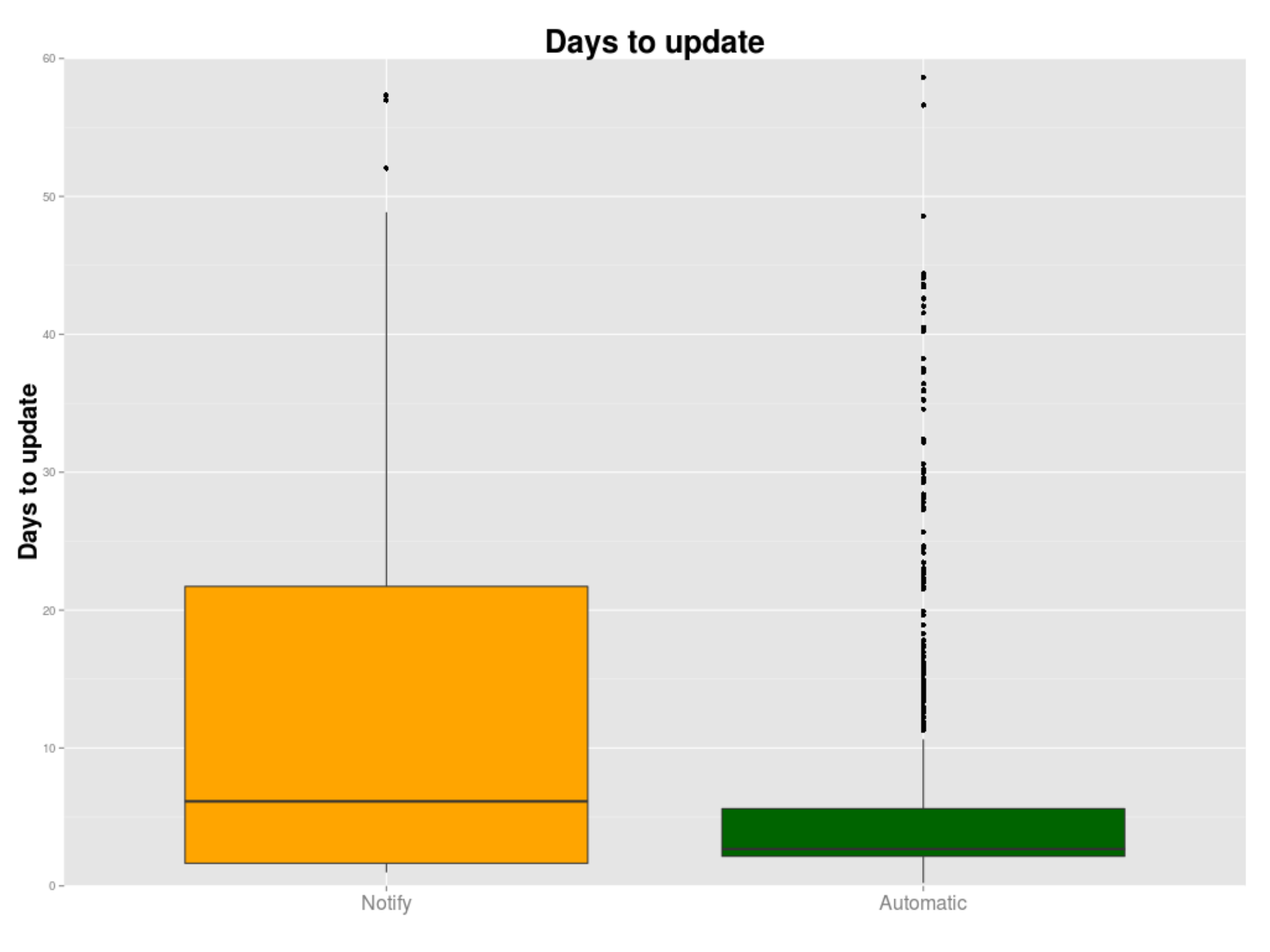
Update Settings

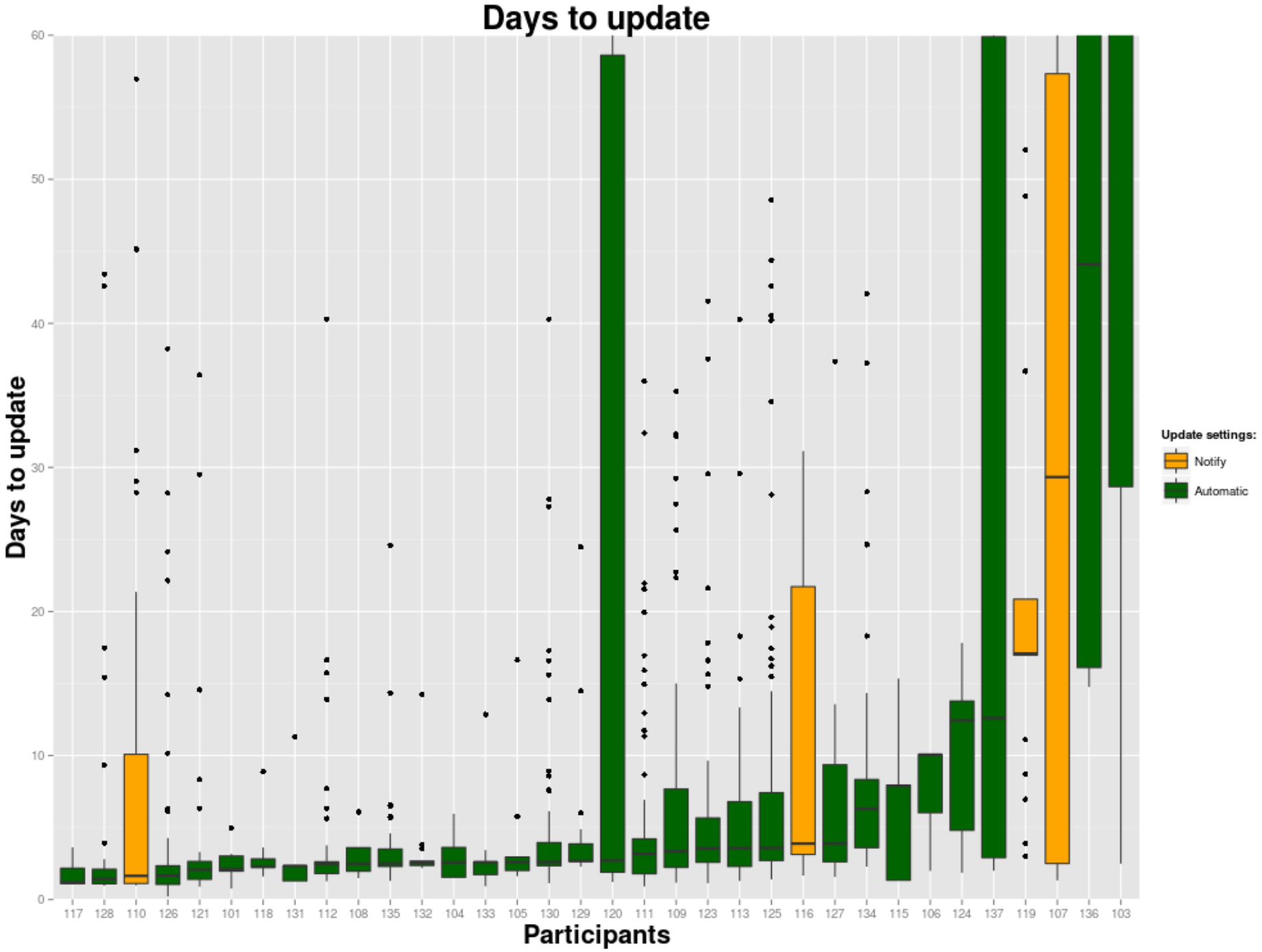
- Automatic install (31 participants)
- Notify (4 participants)
- Disabled (2 participants)

Research Questions

- Are users installing updates?
- Are users installing updates quickly?

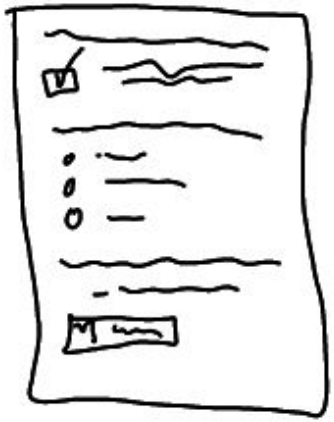

Number of days updates installed
Update settings:
Notify
Automatic
Days updates installed
80
60
40
20
0
10
20
30
Months owned


Days to update
Days to update
60
50
40
30
20
10
0
Notify
Automatic


Days to update
Days to update
0
10
20
30
40
50
60
117 128 110 126 121 101 118 131 112 108 135 132 104 133 105 130 129 120 111 109 123 113 125 116 127 134 115 106 124 137 119 107 136 103
Participants
Update settings:
Notify
Automatic

Settings and logs

Survey



Interview

Survey

- Demographics
- Computer maintence experience
- Installed software
- Auto-update settings
- Updating behavior

Settings

Survey	Checks for updates (Notify)	Checks and installs (Automatic)	Disabled
Notify before install	4	20	-
Notify after install	-	8	1
Both	-	3	-

Settings

Survey	Checks for updates (Notify)	Checks and installs (Automatic)	Disabled
Notify before install	4	20	-
Notify after install	-	8	1
Both	-	3	-

Settings and logs

Survey

Interview

Kami Vaniea, Emilee Rader, and Rick Wash, Betrayed by Updates: How Negative Experiences Affect Future Security, CHI 2014

Semi-structured interview

- Free-listing activity
  - "Things that can happen if the software on your computer is too old or out of date."
- Discuss each listed item
- Hypothetical scenarios
  - Promted to restart computer mid-task
  - Seeing a large number of urgent Windows updates are available
  - Reading a news article about a virus
  - A software program that costs money to update
  - A slow computer that is generating lots of warnings
- Discussion of final report contents

Thematic analysis

We found people avoid updates because of:

- Surprise user interface changes
- Unused and unwanted software
- Currently functional software

Conclusions

- Users are installing Windows updates
- Users with Notify settings are updating slower
- Users are not aware of what their Windows Update settings are
- Users are avoiding updating software due to:
  - Surprise user interface changes
  - Unused and unrecognized software
  - Currently functional software

Think Aloud

Think aloud

- Basic idea: Have a participant use the interface and speak aloud while they do so
- Think aloud is a very versatile, can be long or short, detailed or minimal, planned or ad-hoc
- Pros
  - Get a sense of what the user is trying to do and why they click on some things
  - Very detailed information
  - Testing with 5 users will find the majority of major issues
- Cons
  - Small sample sizes
  - Talking aloud changes how long a user spends on tasks so this method cannot be combined with timing

Think aloud

- Think aloud sessions are typically scripted, that is, you write down everything you will say in advance
- Everything you say to the participant will change their behavior so you have to be very careful
- Typical session
  1. Tell the participant what the session will involve including things like how long it will be and what kind of data recording you will be doing (informed consent)
  2. Train them in thinking aloud
  3. Ask them to accomplish several tasks which have been previously written down, reading aloud each task before starting it
  4. End by thanking them and offering to answer any questions they may have

Usability testing the iPad

- 7 participants with 3+ months experience with iPhones
- Signed an informed consent form explaining:
  - what the participant would be asked to do;
  - the length of time needed for the study;
  - the compensation that would be offered for participating;
  - participants' right to withdraw from the study at any time;
  - a promise that the person's identity would not be disclosed; and
  - an agreement that the data collected would be confidential and would be available to only the evaluators
- Then they were asked to explore the iPad
- Next they were asked to perform randomly assigned specified tasks

Examples of the tasks

App or website	Task
iBook	Download a free copy of *Alice's Adventures in Wonderland* and read through the first few pages.
Craigslist	Find some free mulch for your garden.
eBay	You want to buy a new iPad on eBay. Find one that you could buy from a reputable seller.
Time Magazine	Browse through the magazine and find the best pictures of the week.
Epicurious	You want to make an apple pie for tonight. Find a recipe and see what you need to buy in order to prepare it.
Kayak	You are planning a trip to Death Valley in May this year. Find a hotel located in the park or close to the park.

Table 14.1 Examples of some of the tests used in the iPad evaluation (adapted from Budiu and Nielsen, 2010).

Source: Copyright Nielsen Norman Group, from report available at http://www.nngroup.com/reports/.

Think aloud training

“In this observation, we are interested in what you think about as you perform the tasks we are asking you to do. In order to do this, I am going to ask you to talk aloud as you work on the task. What I mean by “talk aloud” is that I want you to tell me everything you are thinking from the first time you see the statement of the task until you finish the task. I would like you to talk aloud constantly from the time I give you the task until you have completed it. I don’t want you to try to plan out what you say or try to explain to me what you are saying. Just act as if you were alone, speaking to yourself. It is most important that you keep talking. If you are silent for any long period of time, I will ask you to talk. Do you understand what I want you to do?”

Observe-pair-share

- I am going to do a live think-aloud with myself as the subject
- You need to record:
  - Tasks and subtasks I engage in
  - Any critical issues I have
  - Any unexpected behaviors I engage in

Observe-pair-share

- I am going to do a live think-aloud with myself as the subject
- You need to record:
  - Tasks and subtasks I engage in
  - Any critical issues I have
  - Any unexpected behaviors I engage in
- I need:
  - A website
  - A task

Pair-share

- Share your observations with your neighbor
- If you could change how this website is designed, what would you change?

Think aloud Analysis

- Task analysis
- Critical incident analysis

Second Lecture

Think aloud sounds easy but…

- Wording is important. Everything you say to the participant will have an impact and change their behavior
  - Read from a script, or better yet, memorize the script
  - Some researchers hire actors to run think aloud sessions because they can memorize scripts
- It is VERY important that you not talk to the participant during the tasks (this is surprisingly hard to do)
- The only things you can say off-script are:
  - “Please keep talking”
  - Provide a hint if the participant is clearly frustrated
- Train the participant in thinking aloud, this takes time, but it is very important to do

Think aloud sounds easy but…

- Never call the protocol “think aloud” in front of a participant. It is always “talk aloud” or “speak aloud”.
  - The word “think” implies “explain” which will cause participants to start explaining what they are doing to you. This is bad, because when they start explaining they stop behaving normally, because they are thinking about it more.
- Immediately stop the protocol, or step in if the participant becomes distressed
- Help the participant if they ever spend more than 3 minutes on a sub-task

Task Analysis

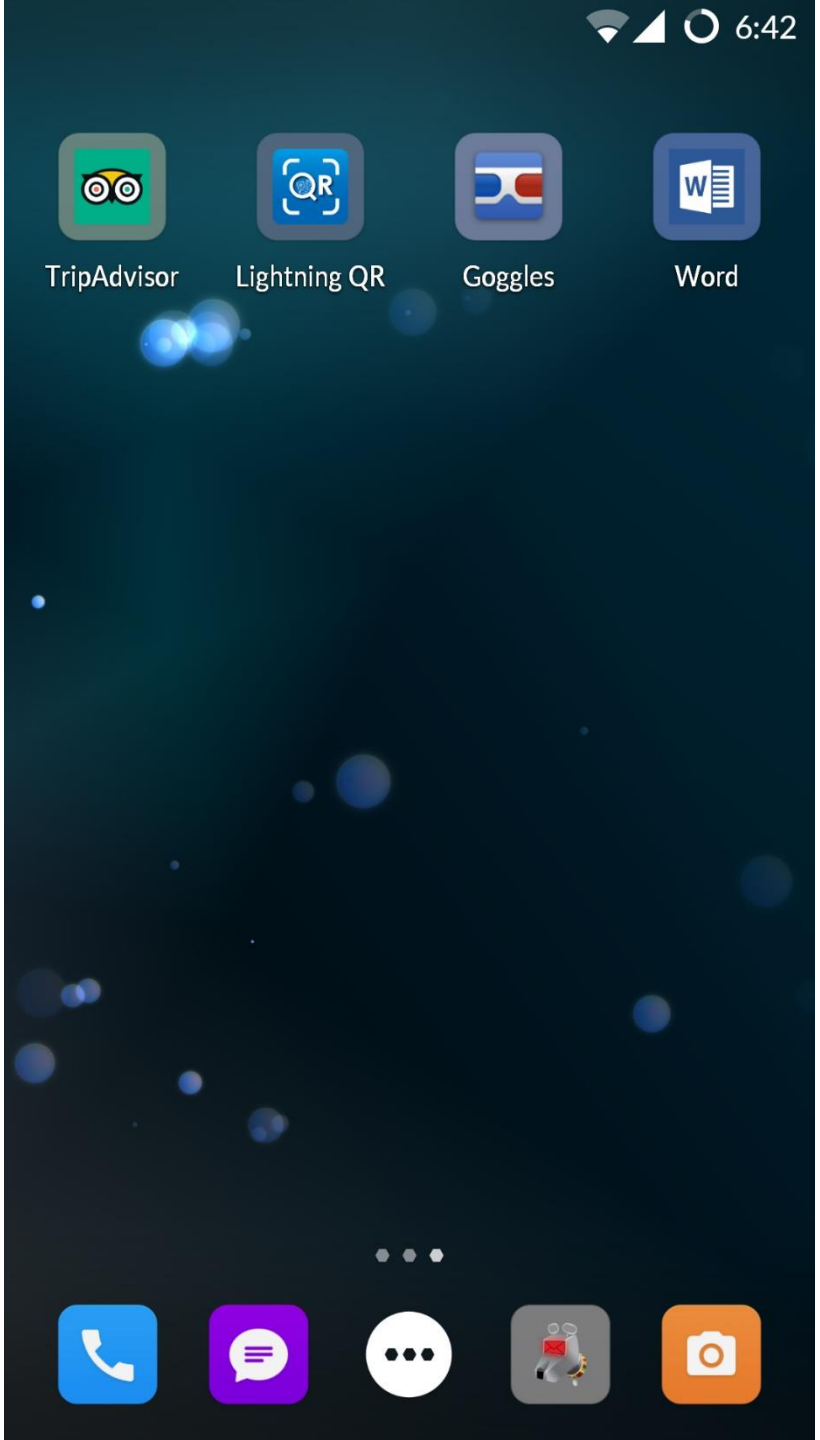
Break participant activity into tasks

- As you observe the participant you should break their activity into tasks and subtasks.
- Primary tasks: The primary goal the user has when they started this interaction.
- Subtasks: smaller task goals that are necessary to complete the larger task

Simple example:

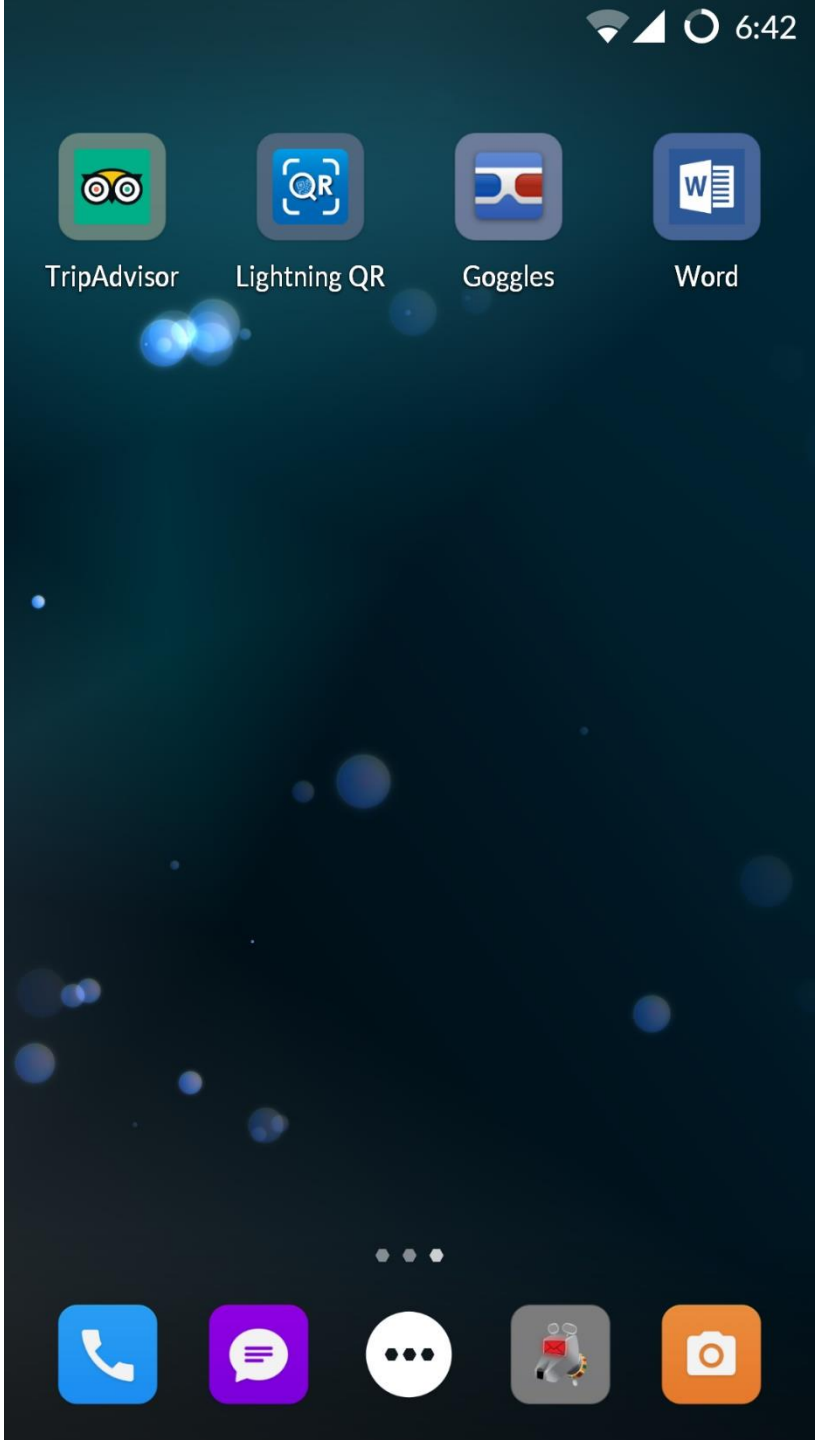
Task: Set an alarm for 7:00am

Task: Set an alarm for 7:00am



Task: Set an alarm for 7:00am

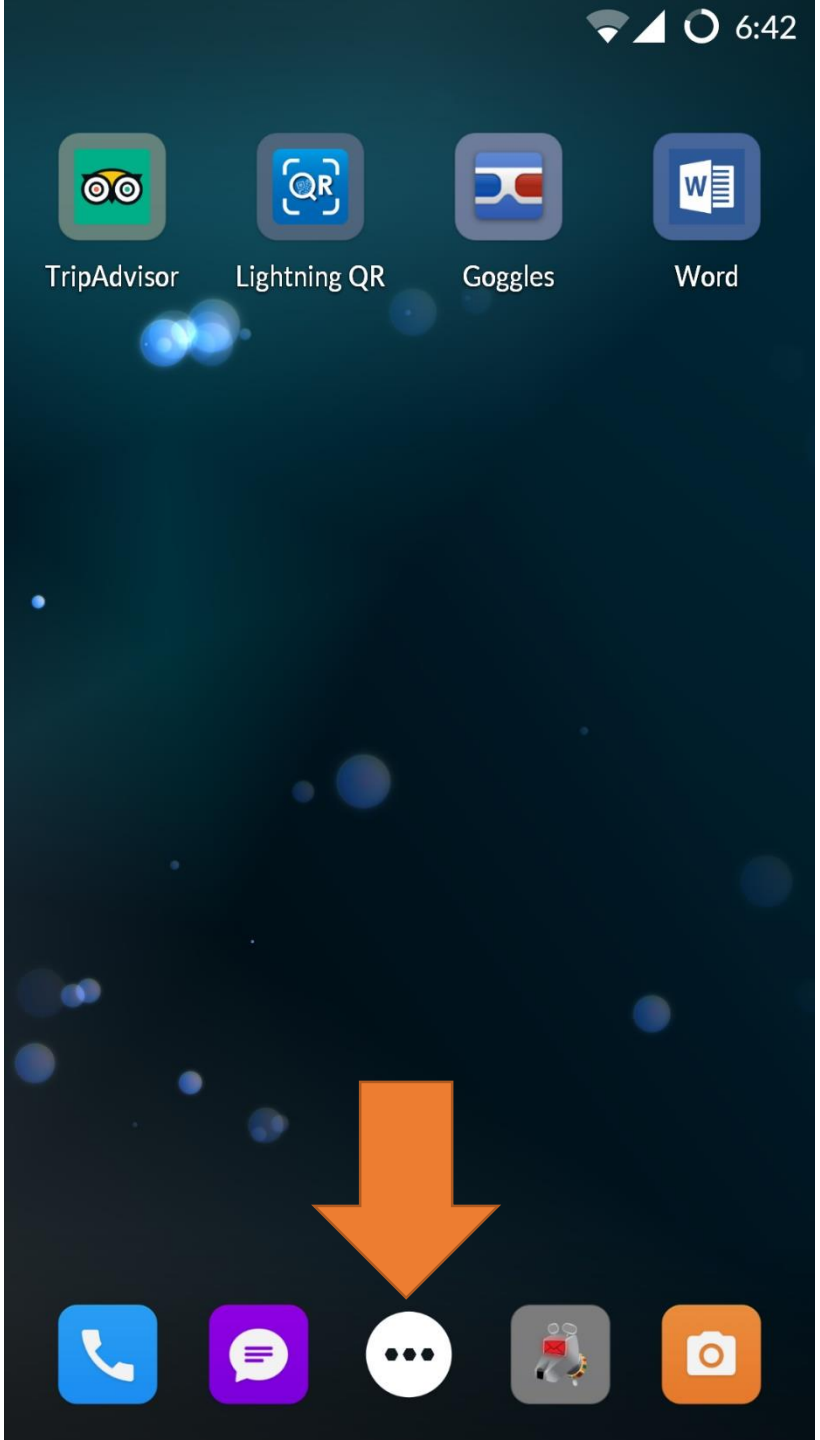
Subtask 1:
Find an app that supports “alarm clock” type functionality.



Task: Set an alarm for 7:00am

Subtask 1:
Find an app that supports “alarm clock” type functionality.

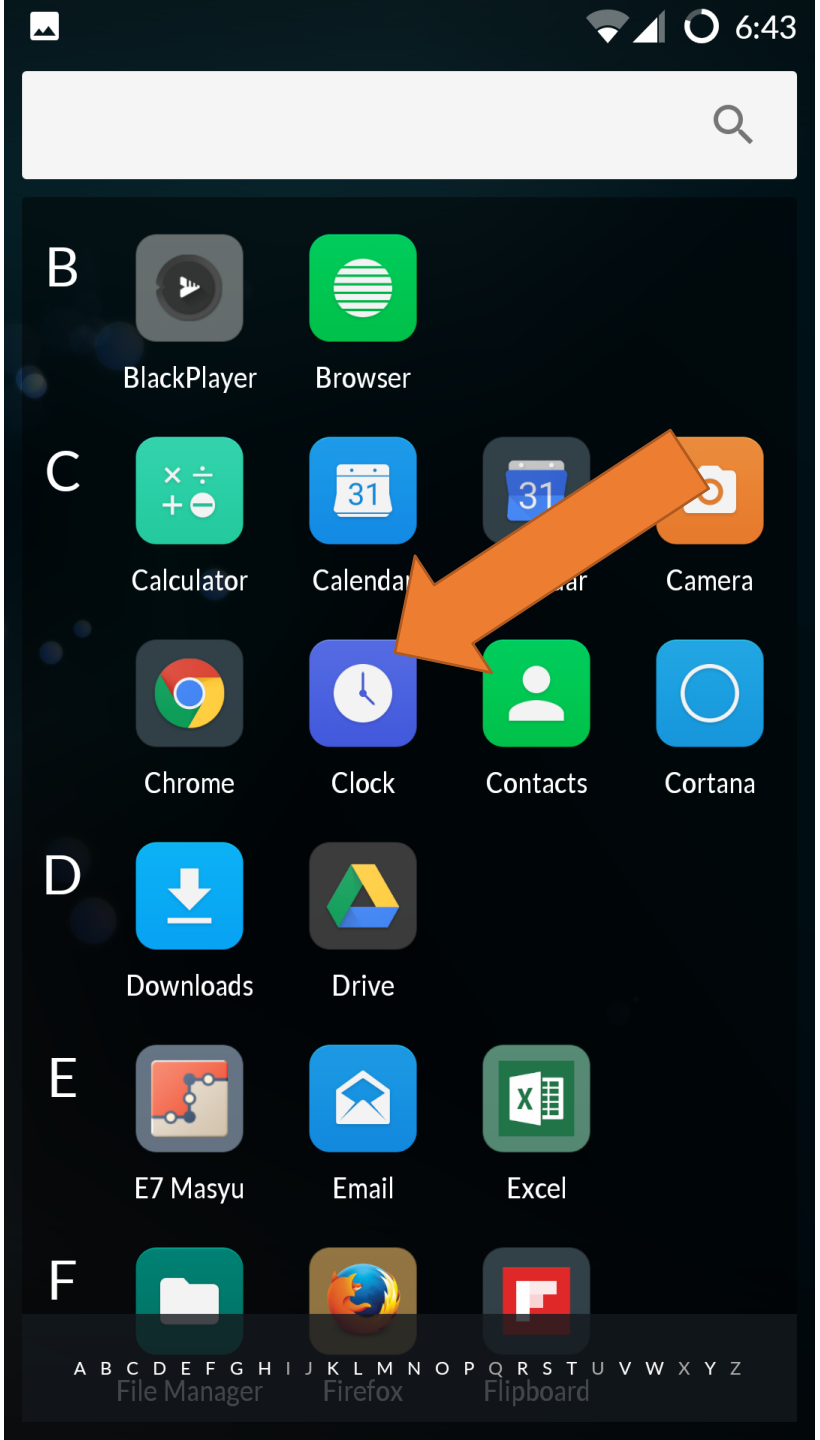
Subtask 2:
Find a list of all apps



Task: Set an alarm for 7:00am

Subtask 1:

Find an app that supports “alarm clock” type functionality.

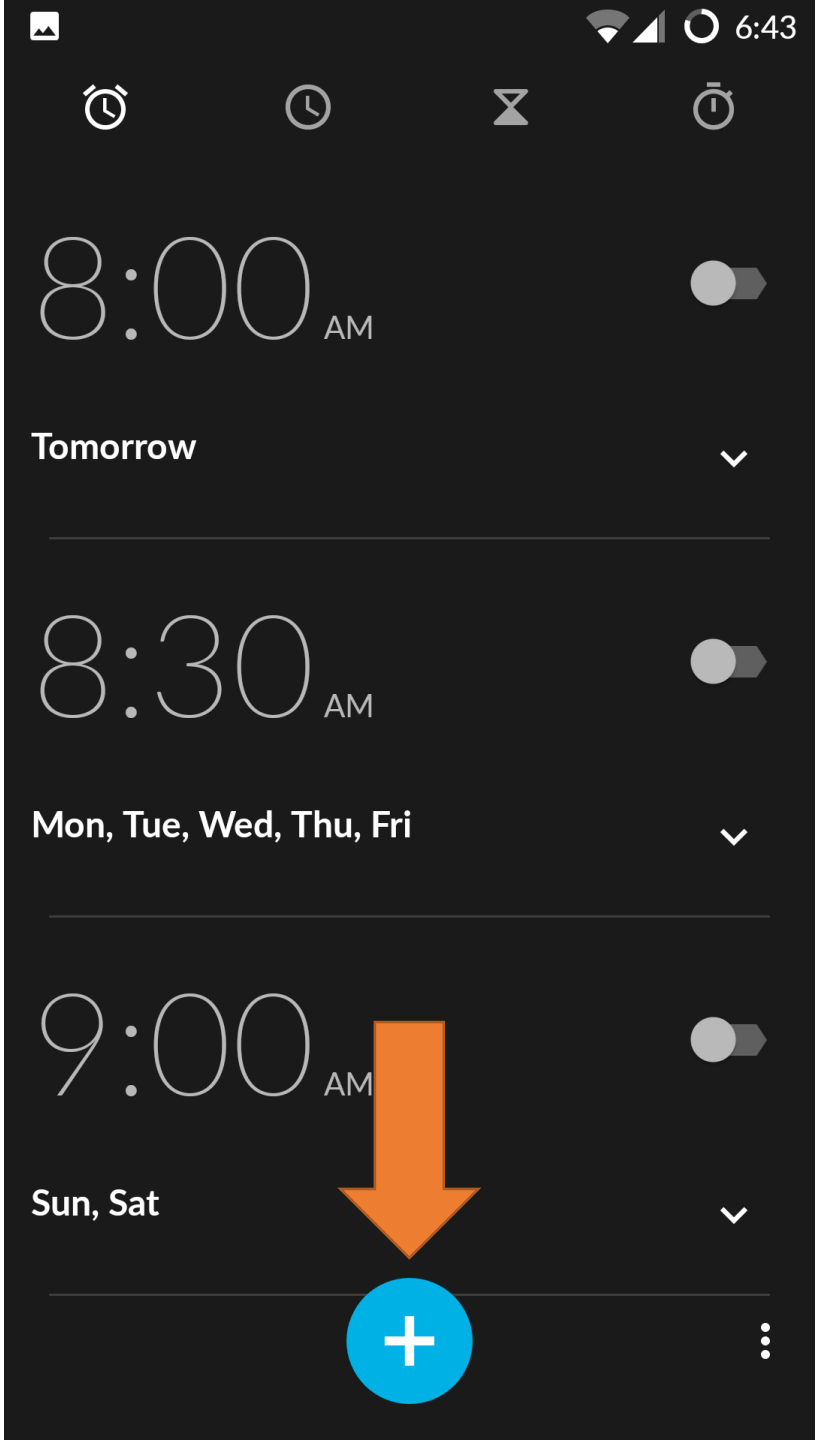


Task: Set an alarm for 7:00am

Subtask 1:
Find an app that supports “alarm clock” type functionality.

Task: Set an alarm for 7:00am

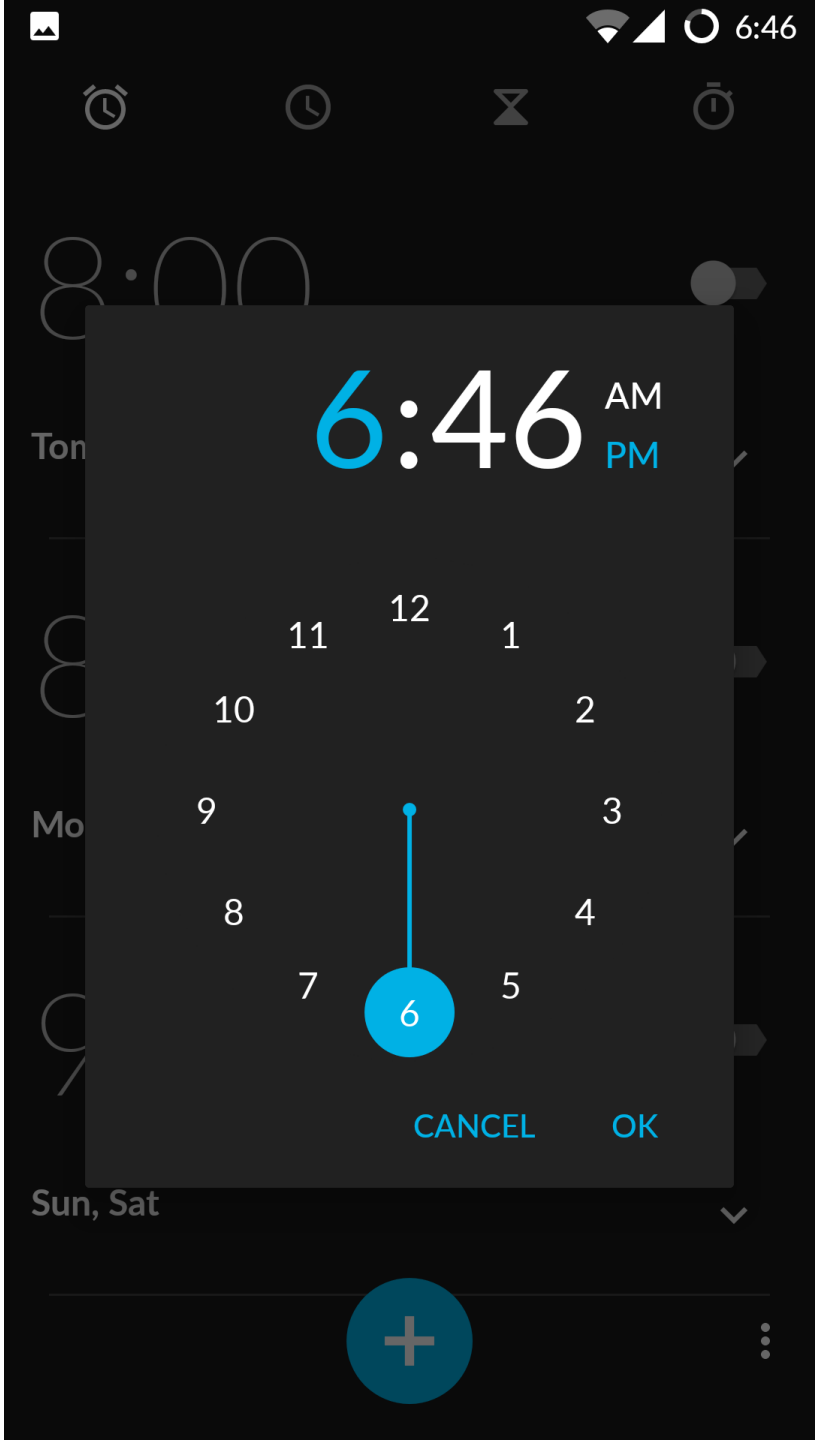
Subtask 3:
Create a new scheduled alarm.



Task: Set an alarm for 7:00am

Subtask 3:
Create a new scheduled alarm.

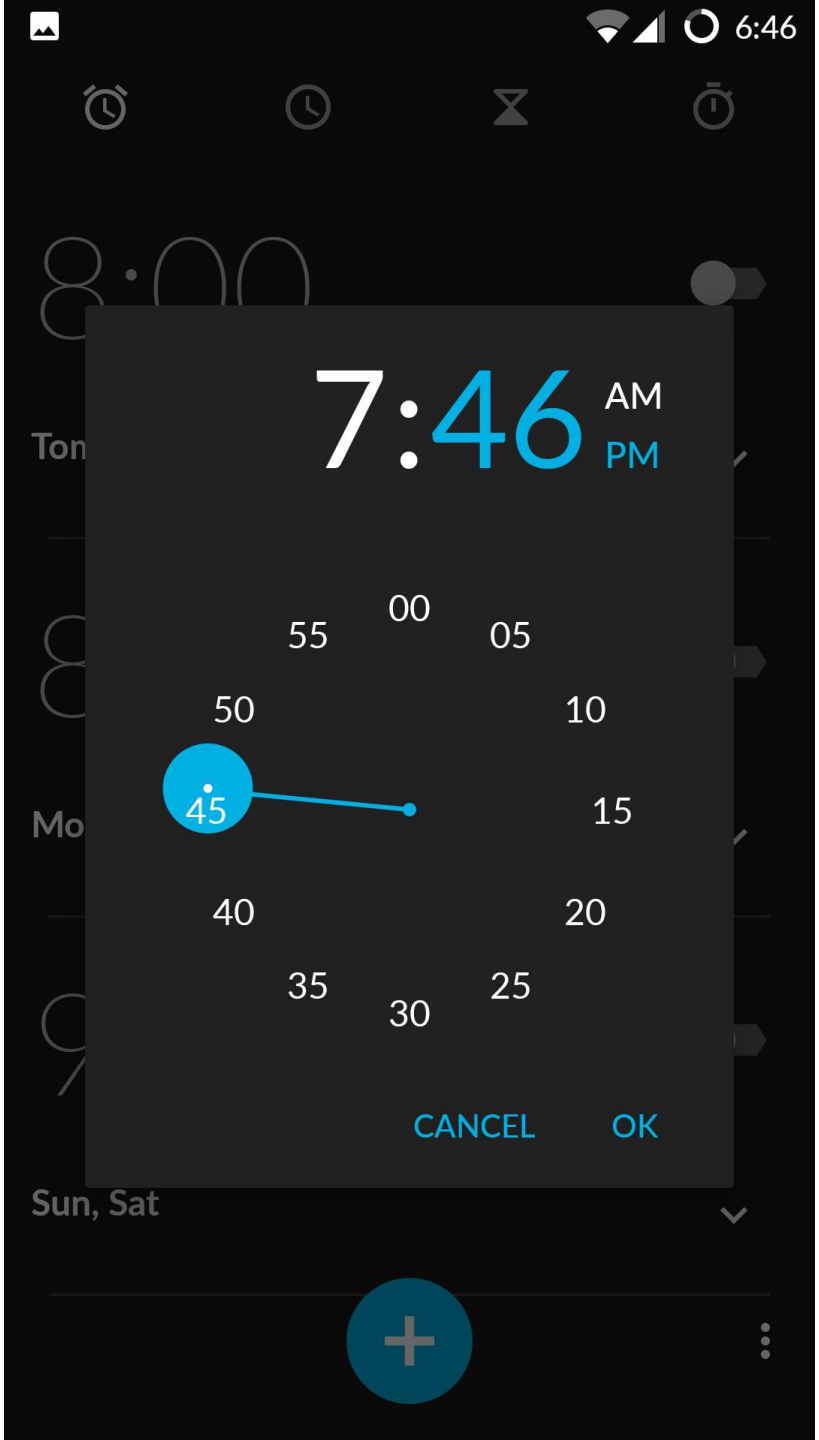
Subtask 4:
Set the hour to 7



Task: Set an alarm for 7:00am

Subtask 3:
Create a new scheduled alarm.

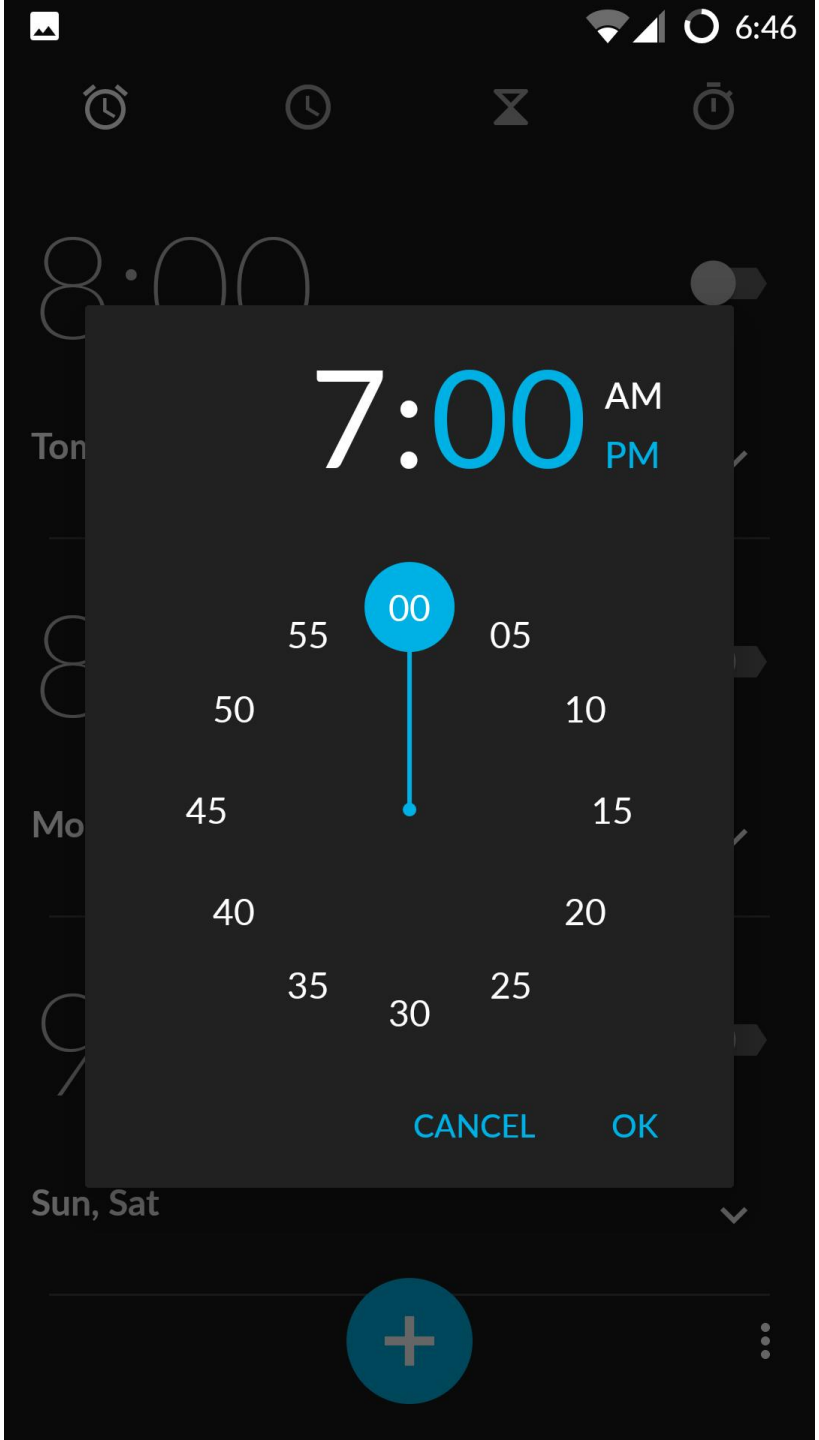
Subtask 5:
Set minutes to 00



Task: Set an alarm for 7:00am

Subtask 3:
Create a new scheduled alarm.

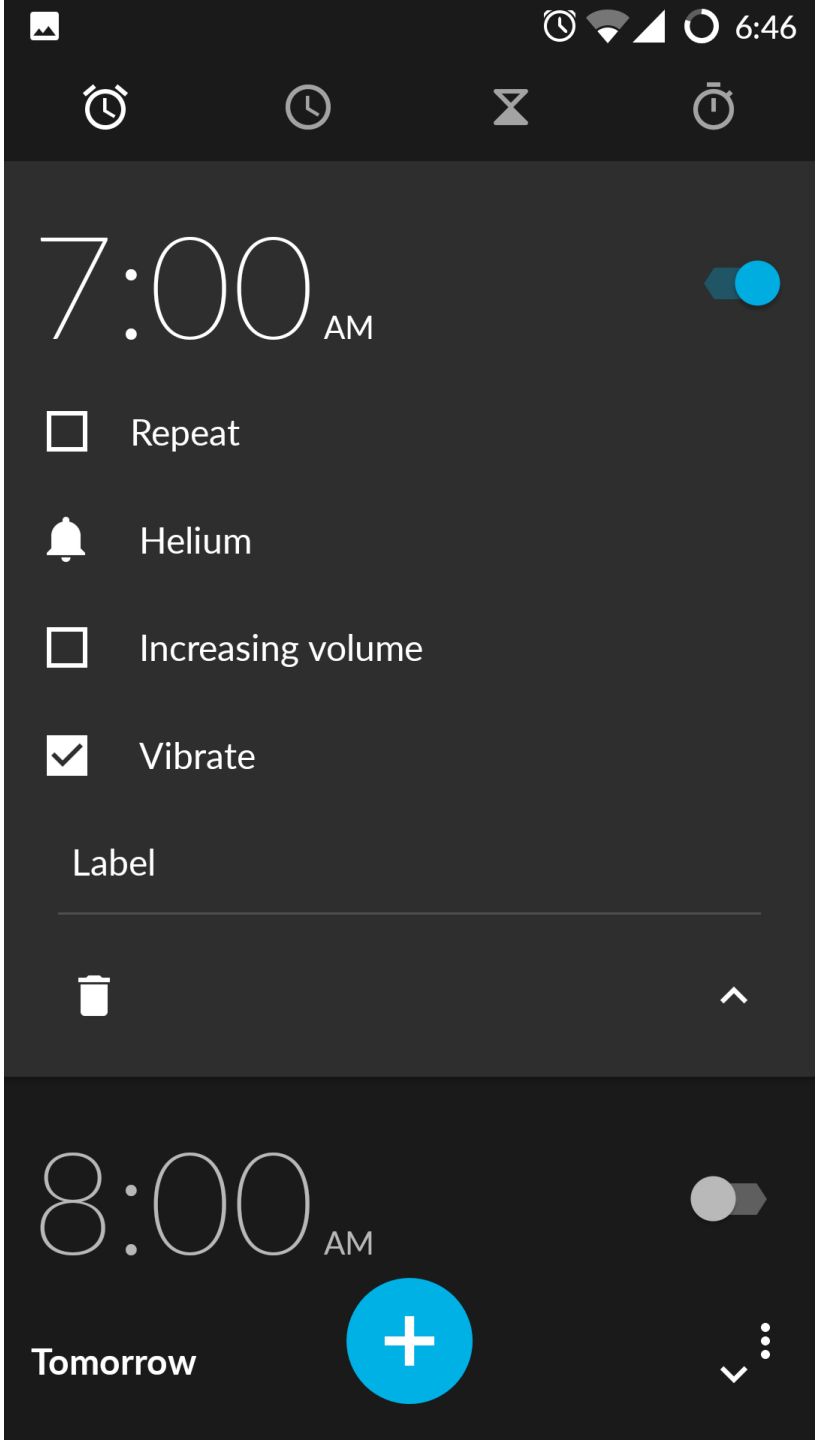
Subtask 6:
Set to “AM”



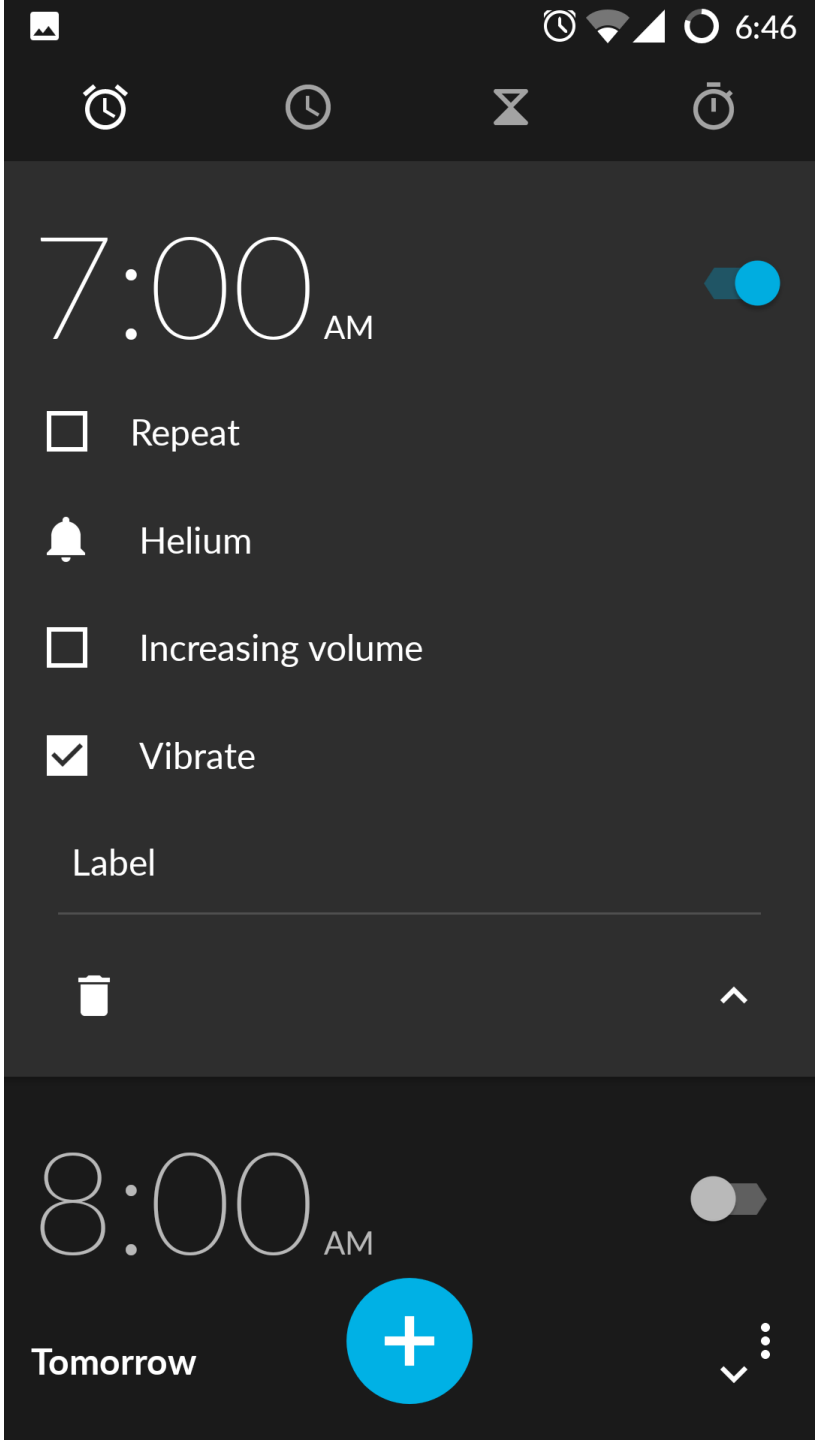
Task: Set an alarm for 7:00am

Subtask 7:

Check that the time has been correctly set and the alarm is now “on”



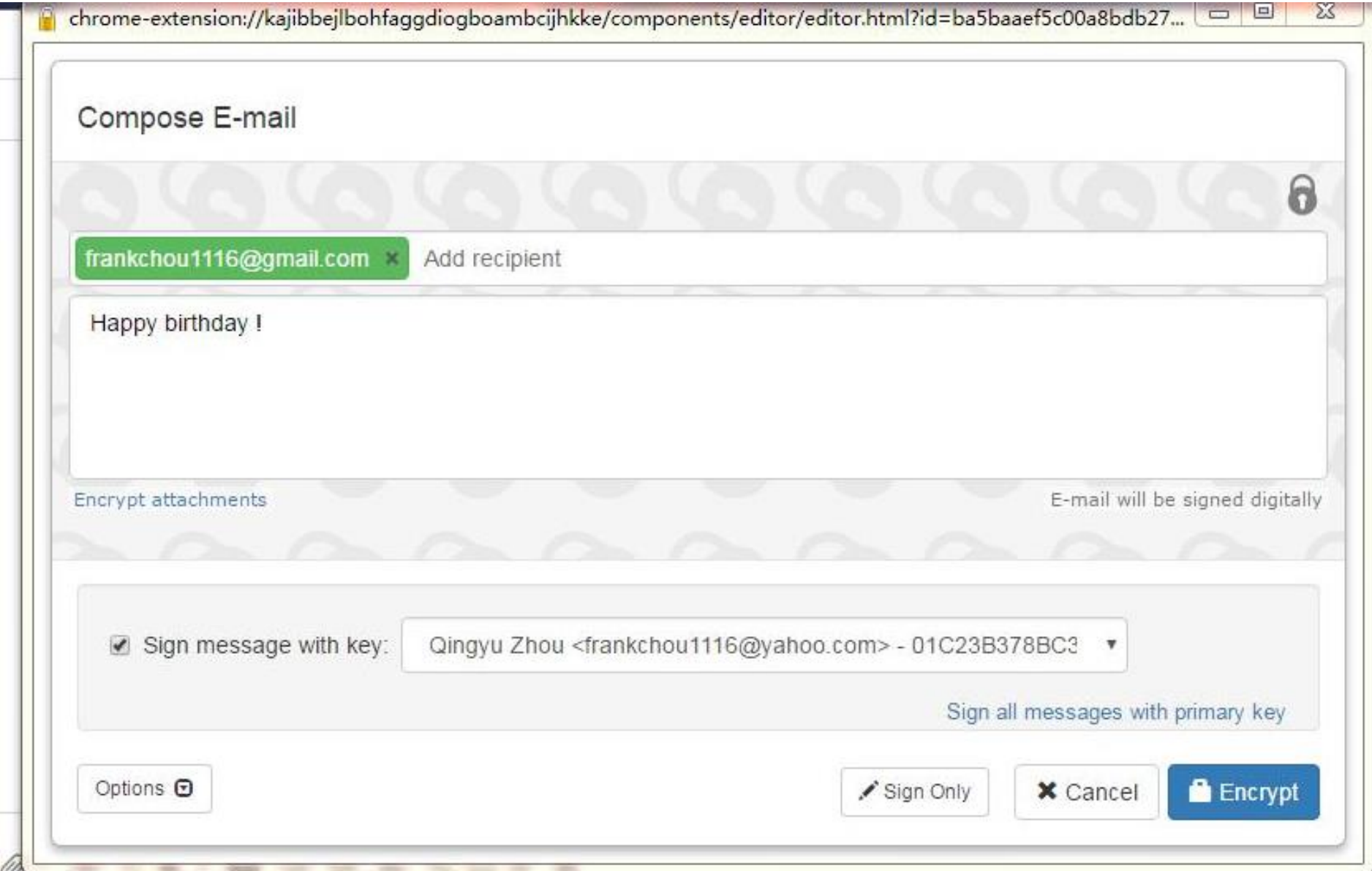
Task Completed!



One analysis is to list tasks/subtasks

- What kinds of subtasks did participants try?
- Did they spend all their time on the “correct” path?
- Are there subtasks that need to be integrated into your product that aren’t?
- Which subtasks are users successful or unsuccessful at?

Mailevelop Plugin



	Task 2: Write an encrypted email			
	Webmail login	Composing email on	Opening Mailvelope popup	Sending encrypted email
T1	Success(hint)	Webmail editor	Failure	Failure
T2	Success(hint)	Webmail editor	Failure	Failure
T3	Success(hint)	Webmail editor	Failure	Failure
T4	Success(hint)	Webmail editor	Failure	Failure
T5	Success(hint)	Webmail editor	Failure	Failure
T6	Success	Webmail editor	Failure	Failure
T7	Success	Webmail editor	Failure	Failure
T8	Success	Webmail editor	Failure	Failure
T9	Success(hint)	Mailvelope popup	Success	Success
T10	Success	Webmail editor	Failure	Failure

Table 4.3: Completion details of Task 2 for each participant.

Time on Task

Time on Task

- Measure “usability” as how quickly a task can be accomplished by the participant.
- Analysis is fairly simple: average, min, max, median, and mode.
- Pros
  - Easy to understand and measure
  - Easy to combine with subtask analysis
- Cons
  - Cannot be combined with a Think Aloud
  - Time is not always the best measure of usability
  - Time changes with experience with the interface

Empirical Evaluation

- Is design A “better” than design B?
- Also known as an A/B test.
- Evaluate two (or more) designs and compare the results to see which design is better.
- Pros
  - Good way to compare two designs
  - Easy to combine with other methods like Think Aloud, Survey, or even Interviews
- Cons
  - Need at least two comparable designs
  - Care is needed to make sure the setups are comparable

Questions?