Studying User Expectations about Data Collection and Use by In-Home Smart Devices

Julia Bernd, ICSI
Serge Egelman, UC Berkeley / ICSI
Maritza Johnson, ICSI
Nathan Malkin, UC Berkeley
Franziska Roesner, UW
Madiha Tabassum, UNCC
Primal Wijesekera, UC Berkeley / ICSI
Thomasz Kosinski, Chalmers University

Approach

How are we applying CI?

1. Observe a data flow
2. Measure expectations
3. Deconstruct the data flow into CI parameters
4. Infer how expectations change based on the parameters

Applying CI to Mobile

Inferring context is hard

- We cannot know exactly how data will be used

Proxies can help

- Knowledge of recipient
- Descriptions of the data (e.g., source, permissions, etc.)
- What the app does
- What else was happening on the device

Smart TV study

Large survey on data collection and sharing – and protections

- Exploring differences by data type/format and by recipient

Wide variation in assumptions about data collection and flows

- Most people are against data being repurposed
- …but assume it will happen regardless!

People believe legal protections exist to prevent egregious violations

- (They don't)

N. Malkin, J. Bernd, M. Johnson, and S. Egelman. *"What Can't Data Be Used For?" Privacy Expectations about Smart TVs in the U.S.* In Proceedings of the European Workshop on Usable Security (EuroUSEC), 2018.

Examining reactions to changing CI parameters

Factorial vignette surveys:

"How would you feel if <X> shared <Y> with <Z>?"

Goal: uncover *relative* levels of concern

A. P. Felt, S. Egelman, and D. Wagner. *I've got 99 problems, but vibration ain't one: a survey of smartphone users' concerns*. In Proceedings of the second ACM workshop on Security and privacy in smartphones and mobile devices (SPSM '12), 2012, Raleigh, North Carolina, USA.

L. N. Lee, J. H. Lee, S. Egelman, and D. Wagner. *Information Disclosure Concerns in The Age of Wearable Computing*. In Proceedings of the NDSS Workshop on Usable Security (USEC '16), 2016.

What do people care about?

Sharing with computers is more acceptable than humans

People really care about multimedia

People care less about observable traits

How does this work for continuous sensing?

Sender: a particular IoT device
Recipient: entity *initially* receiving data
Subject: ?
Data attributes: ?
Transmission principle: ?

Fewer proxies to determine context

- Data is not strongly typed
- Processing often occurs remotely
- Device is shared

STUDY OF CURRENT USERS

Applying CI

What are contextual societal norms?

- What are people's expectations?
- How do these align with what devices already do?
- How do these change across contexts?

How do we define information sharing contexts?

- What other factors influence these expectations?

Study of current users

When do users expect to be recorded?

What do they expect will happen to that data?

How often do current devices record inappropriately?

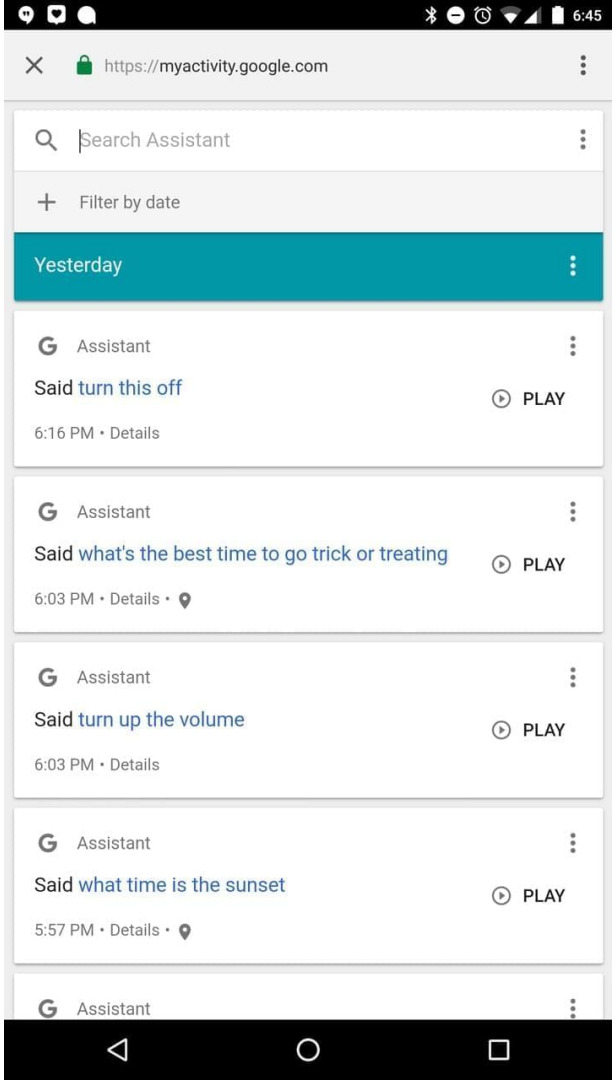
Home
Now Playing
Music & Books
Shopping & To-do Lists
Timers & Alarms
Skills
Smart Home
Things to Try
Settings
Help & Feedback
Not Thomas? Sign out

Settings

History

History shows your voice interactions with Alexa. Tap a line to see details, hear recordings, provide feedback, or delete recordings. Learn more.

- alexa what time is it
  Today at 12:45 PM on Bedroom
- alexa set volume six
  Today at 12:45 PM on Bedroom
- what time is it
  Today at 12:45 PM on Bedroom
- alexa
  Today at 12:45 PM on Bedroom
- stop
  Yesterday at 10:26 PM on Bedroom
- alexa
  Yesterday at 10:26 PM on Bedroom
- alexa set volume four
  Yesterday at 9:59 PM on Bedroom
- next
  Yesterday at 9:56 PM on Bedroom
- alexa
  Yesterday at 9:56 PM on Bedroom
- alexa next
  Yesterday at 9:46 PM on Bedroom
- next
  Yesterday at 9:41 PM on Bedroom


6:45
https://myactivity.google.com
Search Assistant
Filter by date
Yesterday
Assistant
Said turn this off
PLAY
6:16 PM • Details
Assistant
Said what's the best time to go trick or treating
PLAY
6:03 PM • Details •
Assistant
Said turn up the volume
PLAY
6:03 PM • Details
Assistant
Said what time is the sunset
PLAY
5:57 PM • Details •
Assistant

Methodology

Chrome and Firefox extensions

- Recruit existing users
- Screen scrape device activity pages
- Sample recordings to survey:
  - Did the user know it was recording?
  - How do they feel about sharing the given recording?
  - How long should Amazon/Google keep it for?
  - Would they like to delete it now?

Initial question

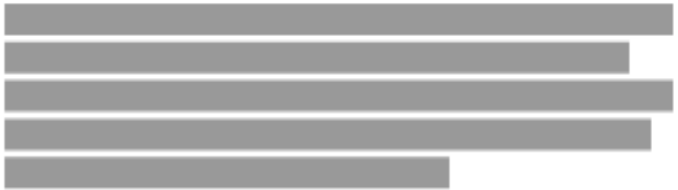
After you ask your device a question or make a command, what do you believe happens to the audio?

- It gets deleted immediately
- It gets saved temporarily
- It gets saved indefinitely
- I don't know

Fun fact: Amazon/Google store your recordings until you delete them

In-Home Assistant Study

http://blues.cs.berkeley.edu/iot-study/



1. Who is this a recording of?

- (•) This is a recording of me.
- ( ) This is a recording of someone else in my household.
- ( ) This is a recording of a guest.
- ( ) This is a recording of the TV, music, or other pre-recorded audio.
- ( ) This is a recording of noise/gibberish.

In-Home Assistant Study

http://blues.cs.berkeley.edu/iot-study/

2 Did [person speaking] address the device, or was this recording an accident?

- (•) [person] was speaking to Alexa.
- ( ) It was an accident.

3. Do you remember making this request?

- (•) Yes
- ( ) No
- ( ) I'm not sure

Retention questions

Would you like to delete this recording?

How would you feel if similar audio recordings were stored for…

- …just long enough to complete the request
- …one week
- …one month
- …one year
- …forever
- …as long as you own/use the device

Human vs. computer recipients

How acceptable would it be for this audio recording to be processed and analyzed by…

…a computer program performing quality control?

…a human, working for the device manufacturer, performing quality control?

Audio vs. transcript

How would you feel if Amazon used this audio recording for…

How would you feel if Amazon used a transcript of this interaction for…

- …Improving Alexa's functions/services
- …Providing you with additional services from Amazon
- …Providing you with offers from Amazon
- …Providing you with additional services and offers from other companies

STUDYING SENSITIVE CONTENT

When should a device not record?

Study idea: examine sensitive conversations

What makes a conversation sensitive?

- Keywords?
- Tone?
- Body language?

Proposed methodology

Crowdsource annotations of existing conversations

“If you were the speaker, would you expect this conversation to be shared with…?”

Perform feature analysis to examine what makes something inappropriate to share

Proposed methodology

How to get existing conversations?

- Existing corpora

Proposed methodology

Results can be used to train a classifier

Future validation studies

STUDYING BEHAVIOR

Self-reports are insufficient

Behavior doesn’t usually follow stated preferences

- e.g., the “privacy paradox”

Solution: give people new devices to study behavior

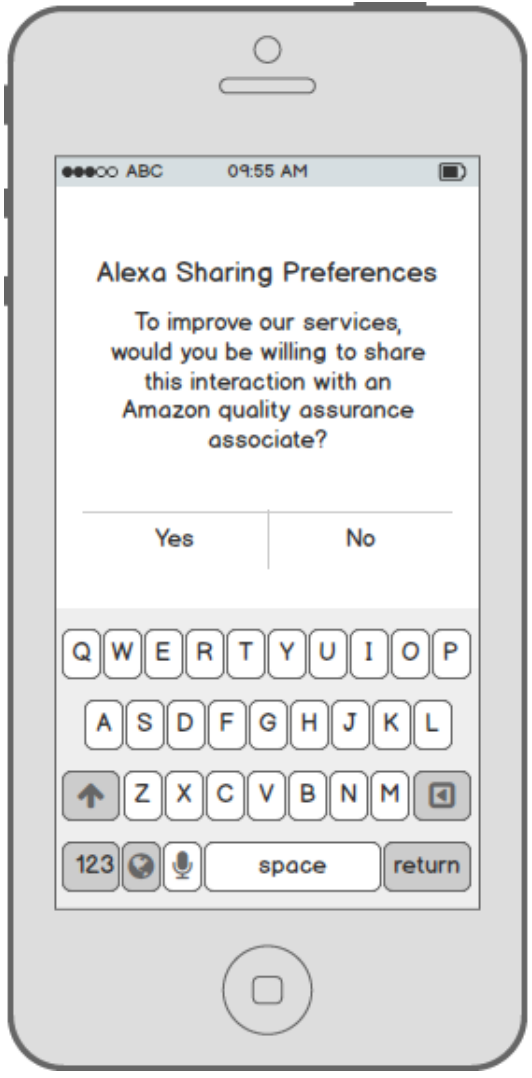
Metrics

Question:

At what point are norms violated?

Possible metrics:

- willingness to share data based on experience sampling



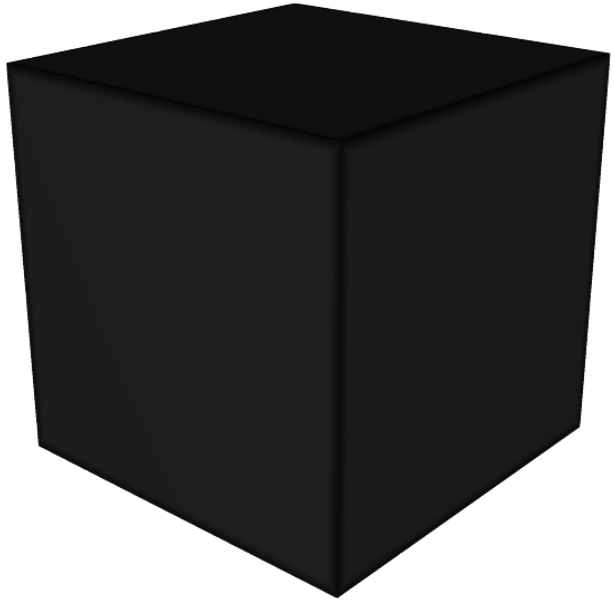
Metrics

Question:

At what point are norms violated?

Possible metrics:

- circumstances surrounding choice to disengage with device



Questions?

Serge Egelman
egelman@cs.berkeley.edu
@v0max