TAKING DATA OUT OF CONTEXT TO HYPER-PERSONALIZE ADS:

Crowdworkers' Privacy Perceptions And Decisions To Disclose Private Information

Julia Hanson*, Miranda Wei*, Sophie Veys, Matthew Kugler, Lior Strahilevitz, Blase Ur

* co-lead authors

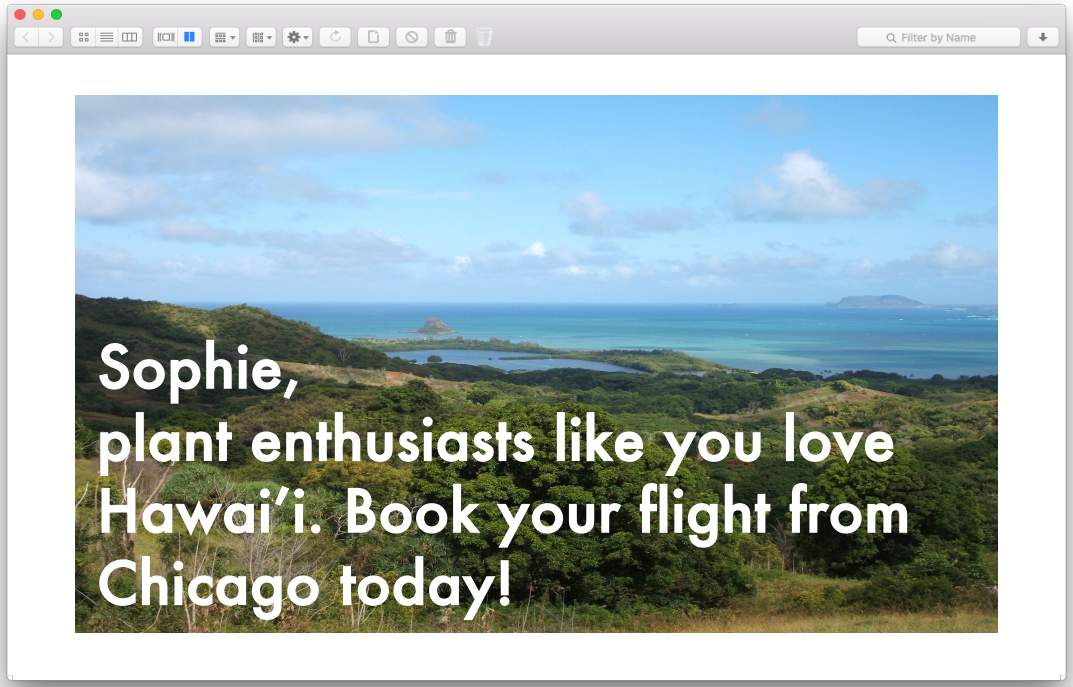
CHI 2020 May 2020

THE UNIVERSITY OF CHICAGO Northwestern

UNIVERSITY of WASHINGTON

UCHICAGO SUPER GROUP


Book your next vacation
to Hawai'i today!


Filter by Name
Sophie,
plant enthusiasts like you love
Hawai'i. Book your flight from
Chicago today!

People can find personalized ads creepy

Personalization reflects potential misuse of data:

data collected in one context, but re-used in another

1. How would you respond to seeing your data re-contextualized in a hyper-targeted ad?

2. How could you prevent unwanted reuse of your data in the future?

One strategy: limit information disclosure

Would seeing a creepy, hyper-personalized ad…

→ …cause feelings of privacy invasion?

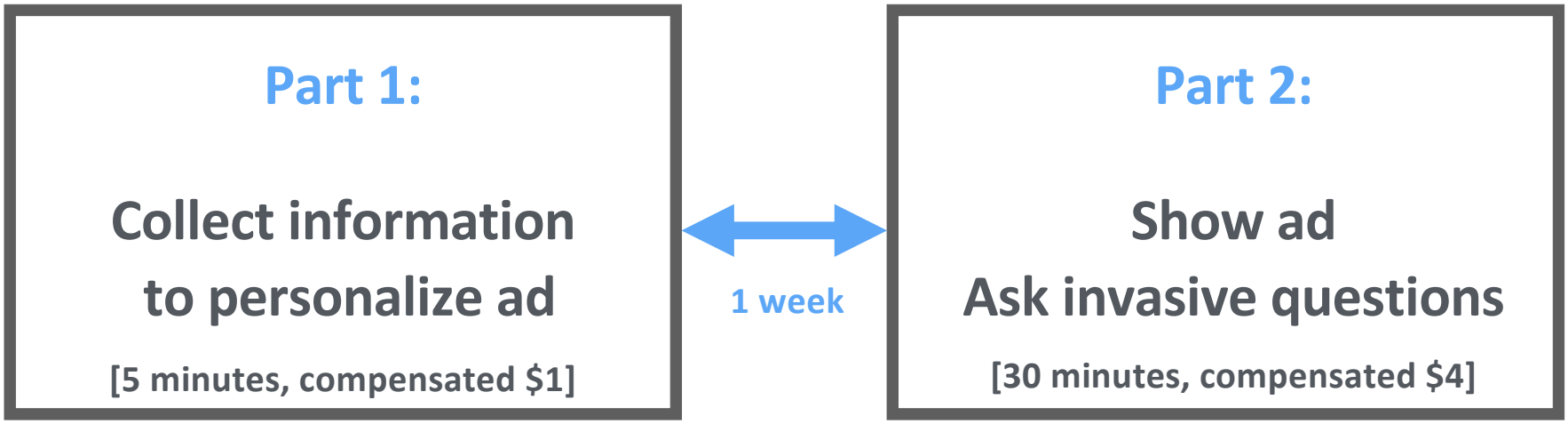
→ …alter information disclosure behavior?

Our experiment:

1. Show either a generic ad or hyper-personalized ad
2. Ask invasive questions (with an option not to answer)

METHODS

Deception protocol

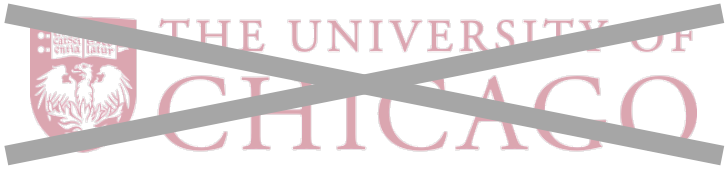


Deception protocol

Study Title:

~~The Impact of Hyper-Personalized Marketing on Information Disclosure~~

Affiliation:



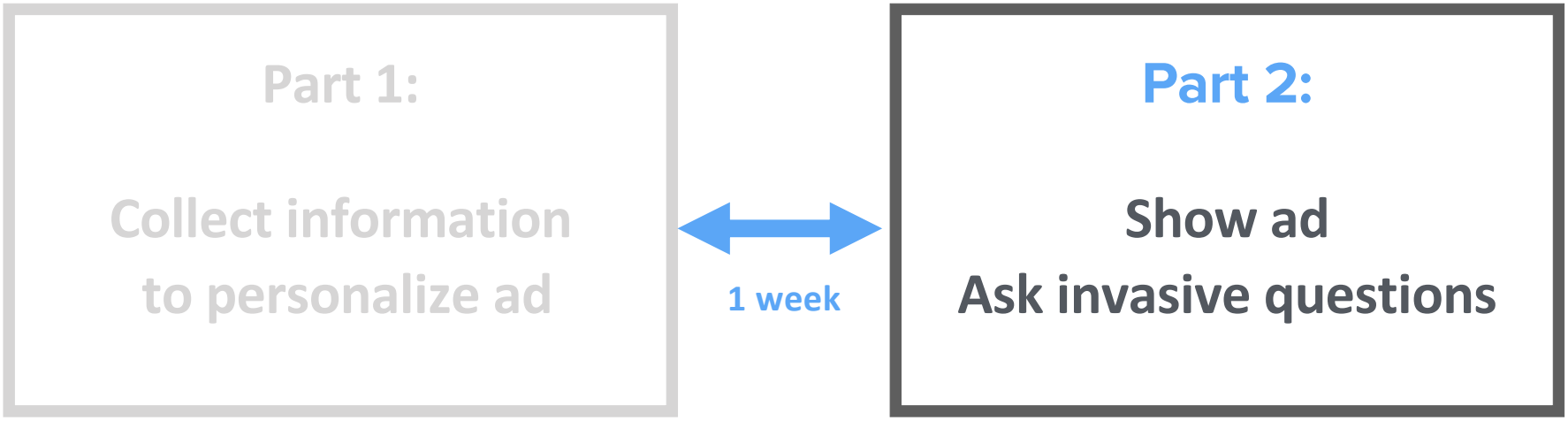
Part 1

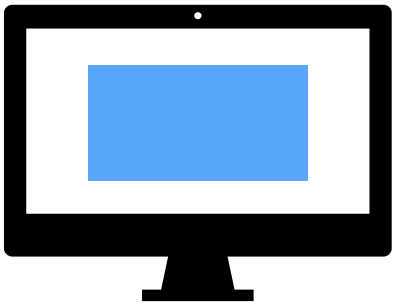
1. Obtain consent
2. Gather information without arousing suspicion

distractor distractor distractor distractor distractor distractor distractor distractor distractor first name distractor distractor distractor distractor distractor distractor distractor distractor phone number distractor distractor distractor distractor distractor distractor distractor distractor preferred cuisine distractor relationship status distractor distractor distractor distractor distractor distractor distractor distractor distractor partner's name

+ location

Deception protocol





or

Banner

Robotext

first name

location

partner's name

preferred cuisine

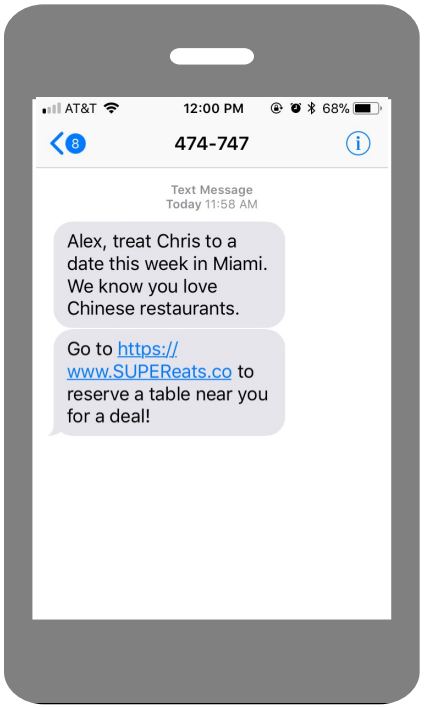
Banner ads

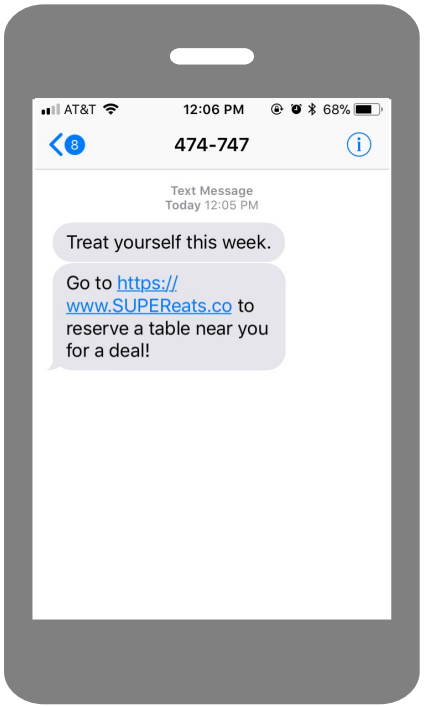
Personalized

Generic

Robotext ads

Personalized

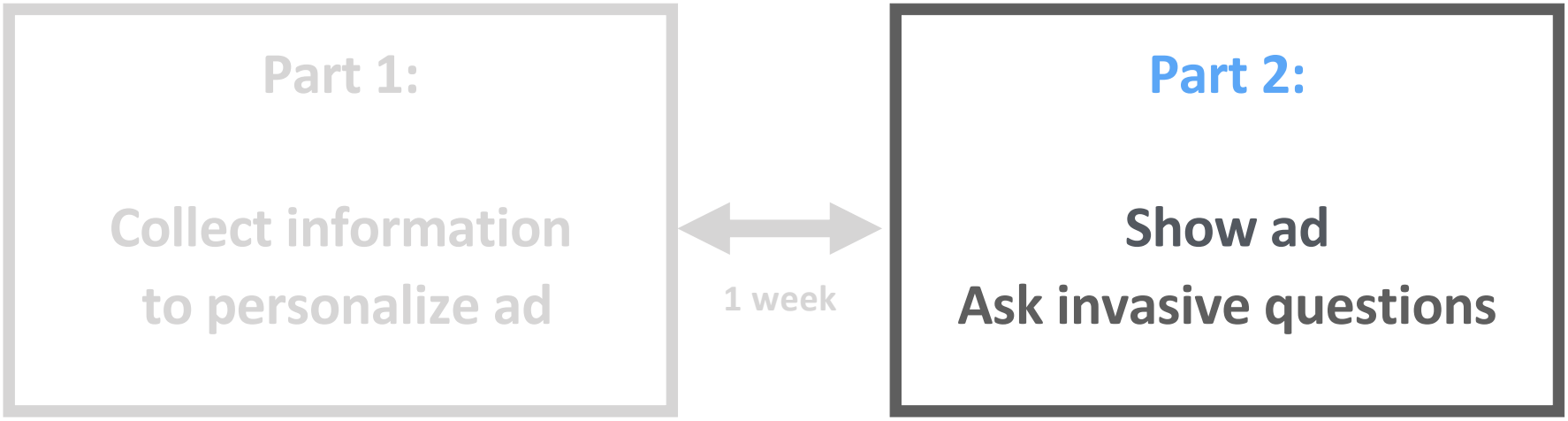




Generic

Study conditions

Deception protocol



Invasive questions

COSMOPOLITAN

series of 43 invasive questions

Invasive questions

How many years of experience do you have with using the Internet?

For what reason did you last cry?

What is the first and last name of your best friend?

What is your home address?

“Prefer not to say” option

Explicitly told no penalty for selecting ‘Prefer not to say’

Debrief

We revealed:

- **not IIDR, but the University of Chicago**
- **we had sent the banner or robotext ad**

Following debrief

We asked:

- do you remember seeing the ad?
- did you suspect it was study related at any point?
- what factors did you consider when answering?
- did you answer any questions inaccurately? (no penalty)

Participant privacy and ethics

Used Prolific's deception filter

Prolific approved our study design

Debriefed participants who dropped out

Deleted PII and answers to invasive questions

RESULTS

Deception effectiveness

92

(65%)

read robotext and not suspicious

85

(62%)

saw banner ad and not suspicious

Reactions to ad

reported feeling scared, concerned, shocked or surprised, creeped out, or uncomfortable

53%

Personalized Banner

44%

Personalized Robotext

0%

Generic Banner

Generic Robotext

Reactions to ad

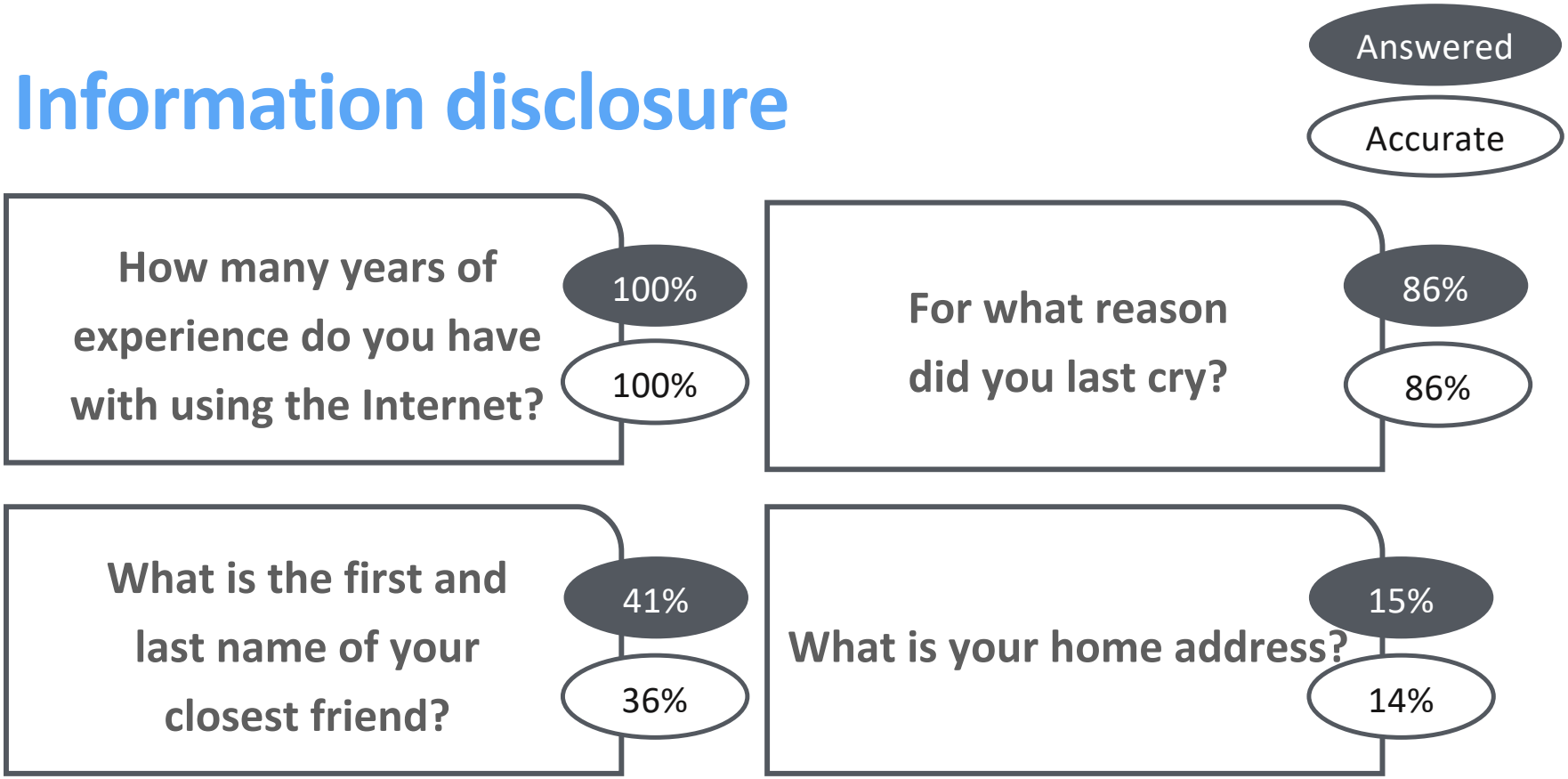
creeped out

determine data origin

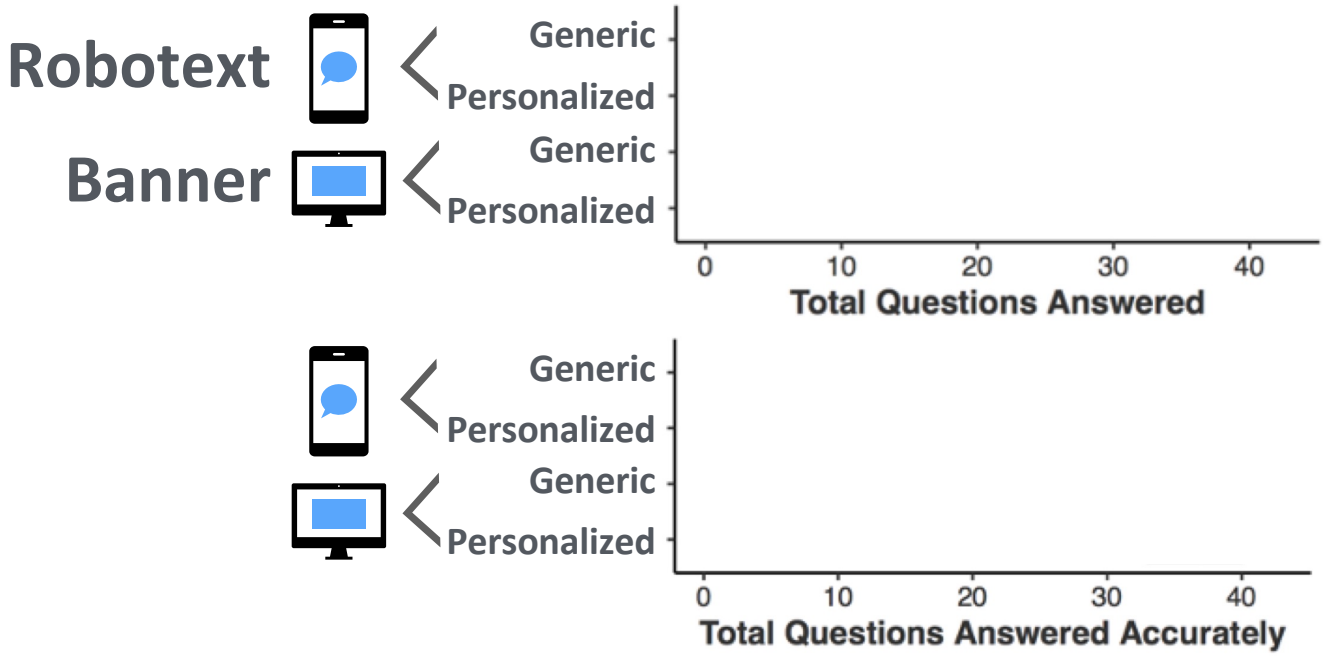
"I had a noticeable reaction of both worry and disgust, and then I realized I had been visiting a lot of restaurant websites recently" (P123, Banner-Personalized)

"I was quite alarmed by the ad, wondering how it knew my spouse's name and my location. I was disturbed and put off by it." (P213, Robotext-Personalized)

Information disclosure



Information disclosure by condition



Part 2: Why select prefer not to say?

personal or embarrassing (42%)

"I selected prefer not to say when I thought it **crossed the line** into my privacy." *(P47, Banner-Generic)*

personally-identifying or security-related (36%)

"…Questions that felt like phishing (mom's maiden name) were a no-go. I cared less about it being invasive and more about it **being bad security**." *(P156, Banner-Generic)*

Information disclosure factors

lack of time/effort

compensation

privacy calculus

“It’s **impossible to really know** where the data goes after you

[illegible]

too much effort to try and keep up with everything.” *(P8, Banner-Personalized)*

“I think there is a lot of risk, and if I wasn’t flat broke I wouldn’t be doing this.

[illegible]

(P153, Robotext-Personalized)

[illegible]

takes all day, so does amazon, so do so many others. I dont get paid for that.” *(P207, Robotext-Generic)*

Crowdworking and trust

[Sannon & Cosley 2019]

mixed trust in IIDR

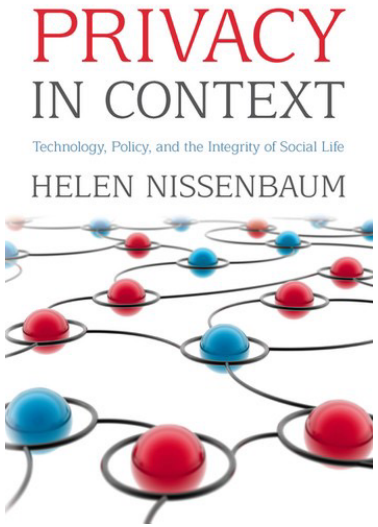
"They are ok; don't know a lot about them" *(P147, Robotext-Personalized)*

high trust in Prolific

"I feel like Prolific weeds out all the bad studies so had faith that Prolifics would keep me safe" *(P170, Robotext-Personalized)*

DISCUSSION

Information flows



Contextual integrity

- Privacy: appropriate flows of information
- Privacy harms: unexpected (re)use of data

Privacy compartmentalization

- Do actors need to be identified for behavior change?

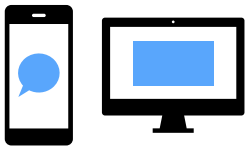
Implications for crowdwork

Crowdworker trust that platforms will help manage privacy

Need privacy by design:

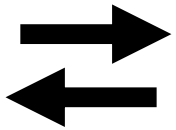
- **Requester identity verification**
- **Collaborations between IRBs and crowdwork platforms**
- **Purpose of data collection**

Directions for future work



Strength of effect?

Single violation of contextual integrity not enough



Competing effects?

Cautionary lesson: don't share data again

Demoralization: privacy hopelessness confirmed

Other contexts?

Repeat with non-crowdworkers

TAKING DATA OUT OF CONTEXT TO HYPER-PERSONALIZE ADS:

Crowdworkers' Privacy Perceptions And Decisions To Disclose Private Information

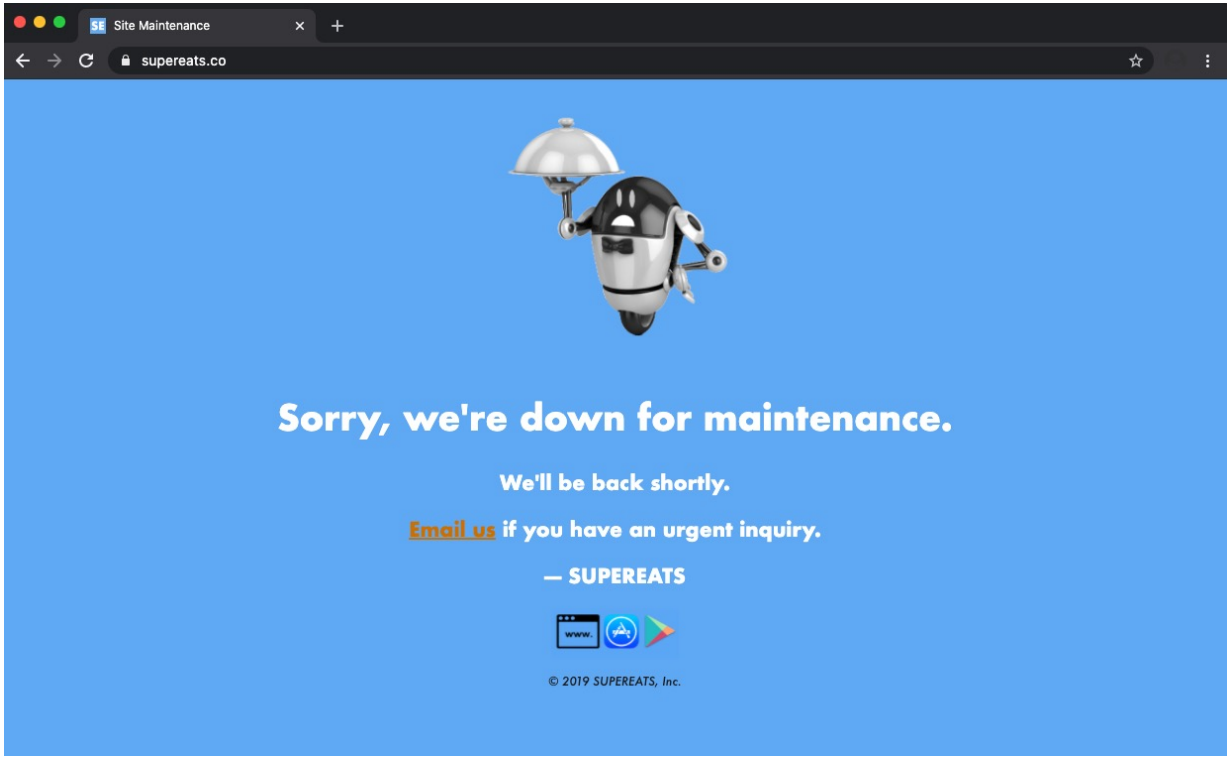
Julia Hanson*, Miranda Wei*, Sophie Veys, Matthew Kugler, Lior Strahilevitz, Blase Ur

* co-lead authors

CHI 2020 May 2020

1. Disclosure will continue despite hyper-personalization
2. Self-regulation of privacy is insufficient
3. Crowdworkers need improved privacy protections

BONUS SLIDES


Site Maintenance
supereats.co
Sorry, we're down for maintenance.
We'll be back shortly.
Email us if you have an urgent inquiry.
— SUPEREATS
© 2019 SUPEREATS, Inc.

pre-study

prospective questions: American Housing Survey; US Census; Cosmopolitan magazine and Facebook quizzes; Pew Research surveys; group discussions

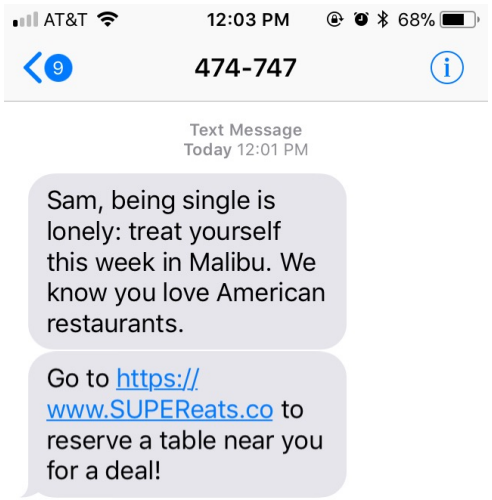
ideal question: consistent invasiveness across answer choices and quickly recallable answer

- "If given the opportunity, I would choose not to answer this question." [Yes / No / Don't Know]

tested with 63 pre-study participants on MTurk

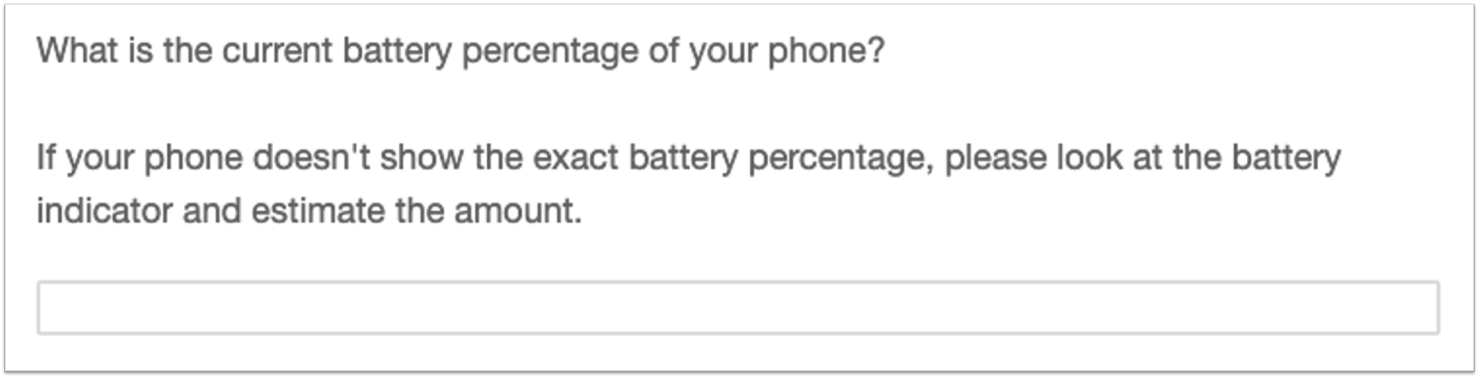
chose set of 43 potentially invasive questions

Personalized ads for single participants



Part 2: following the ad…

General technology use questions to prompt to look at phone:



Part 2: other survey questions

Asked about…

opinions of personalized advertising

knowledge of Cambridge Analytica scandal

part 1:
elicit data

What is your first name?

Sophie

Hi Sophie. Which phone number should we use to contact you?

You will receive up to 3 study-related text messages. Your phone number will be deleted from our records as soon as the study is complete.

123-456-7890

part 1:
elicit data

In this section, you will see some questions about hypothetical everyday situations.

Imagine that you walked into a movie theater that was showing all of the following movies. Which would you be most likely to watch?

- After
- Hellboy
- **A Dark Place**
- Master Z: IP Man Legacy
- Mia and the White Lion
- Us

part 1:
elicit data

In this section, you will see some questions about hypothetical everyday situations.

Imagine that you walked into a movie theater that was showing all of the following movies. Which would you be most likely to watch?

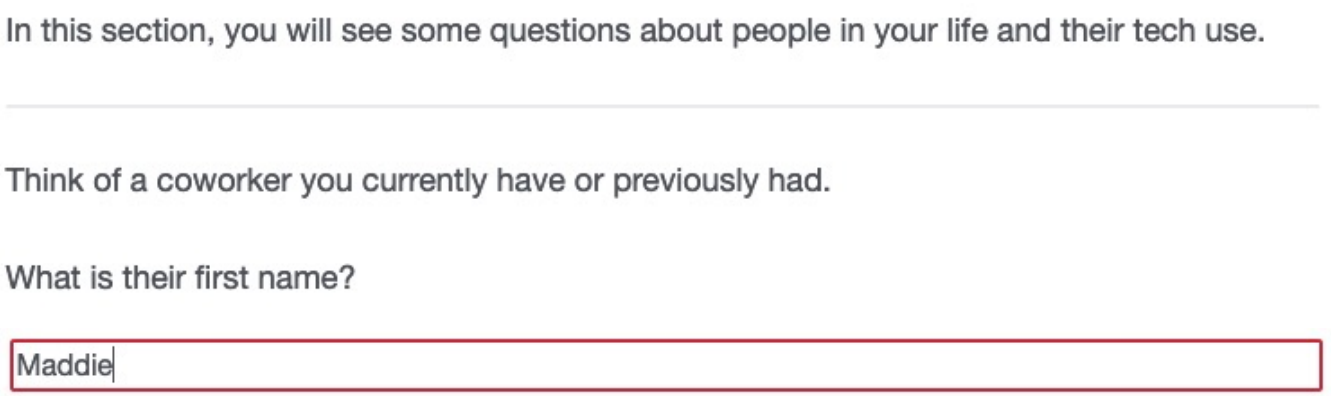
- After
- Hellboy
- **A Dark Place**
- Master Z: IP Man Legacy
- Mia and the White Lion
- Us

part 1:
elicit data

Imagine that you're hungry and walking down a street with the following types of restaurants. Which would you be most likely to go to?

- Italian
- Thai
- Indian
- American
- Japanese
- **Mexican**
- Chinese

part 1:
elicit data



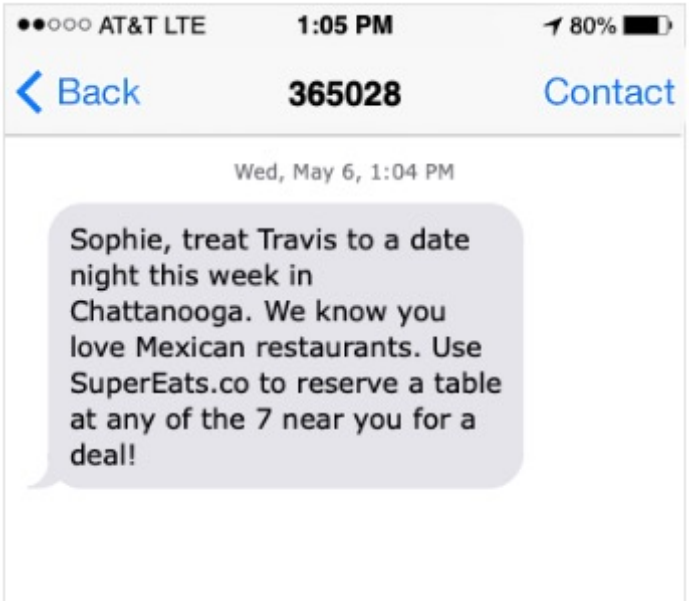
part 1:
elicit data

Think of your current significant other. If you do not currently have a significant other, think of someone you previously had a relationship with, or wish you were in a relationship with.

What is their name?

Travis

part 2:
deliver ad



part 2: deliver ad

This is the first section of this survey. You will be asked questions about your general use of technology.

How many apps do you have on the main home screen of your phone?

What is the current battery percentage of your phone?

If your phone doesn't show the exact battery percentage, please look at the battery indicator and estimate the amount.

11-15

16-20

20 or more

My phone doesn't have a home screen

part 2: measure disclosure

What did you eat for dinner last night?

Prefer not to say

Are you registered to vote at your current place of residence?

Yes

No

Prefer not to say

personalization in advertising

demographics

interests

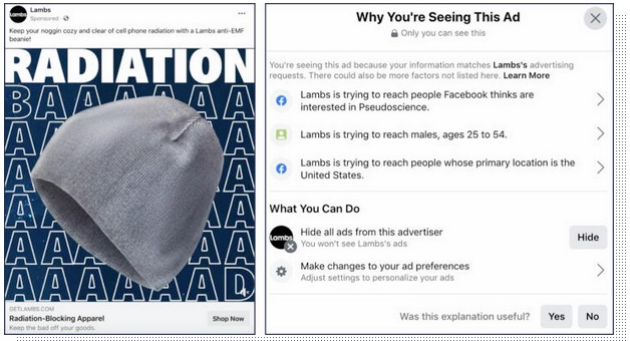
behaviors

device info

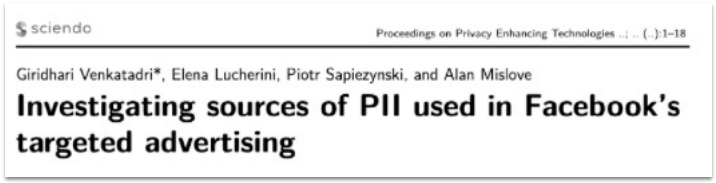
post engagement

location

The Markup



personalization = privacy violation ?



going online —> your data is collected

privacy calculus

privacy paradox

in a variety of contexts, people acknowledge potential privacy implications but take no action