How Well Do My Results Generalize?

Comparing Security and Privacy Survey Results from MTurk, Web, and Telephone Samples

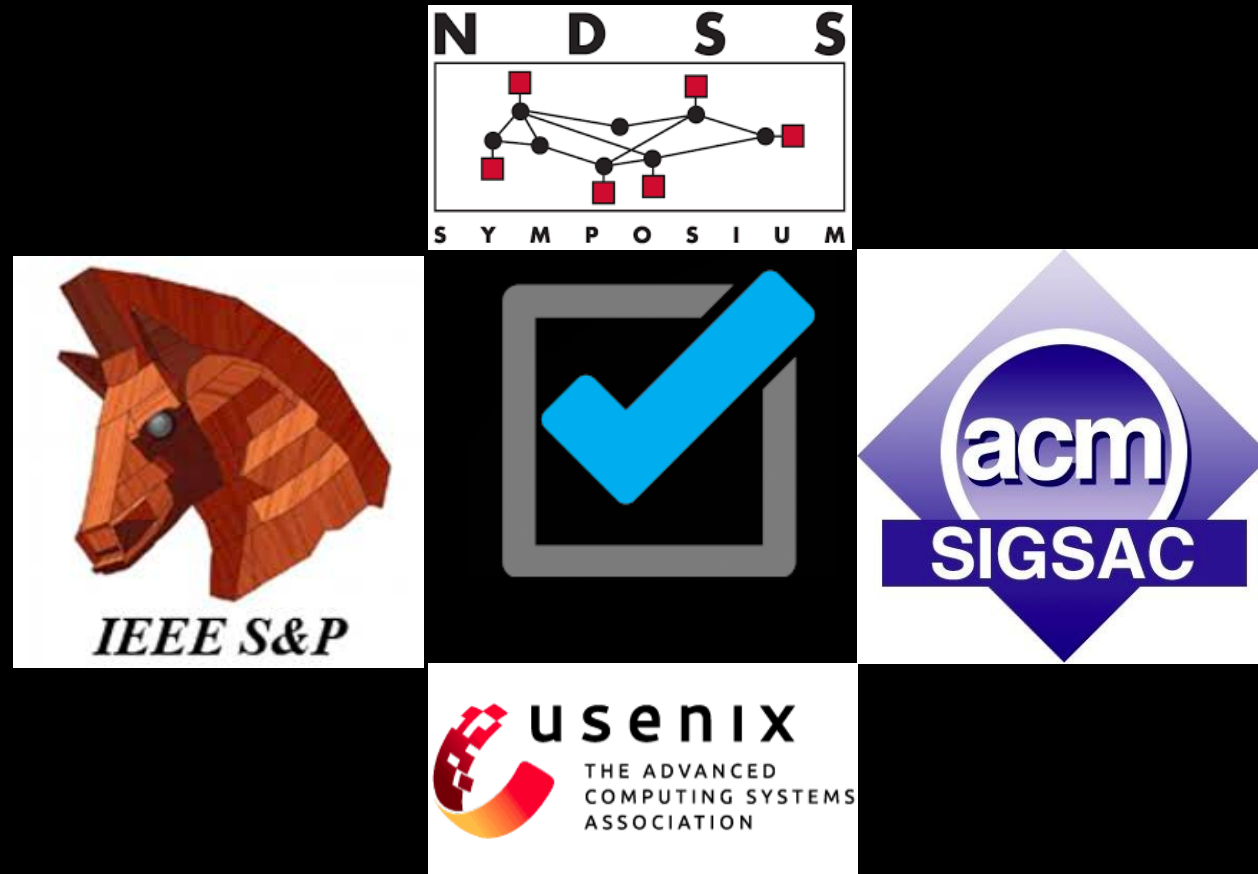Elissa M. Redmiles, Sean Kross, and Michelle L. Mazurek

@eredmil1

eredmiles@cs.umd.edu

30+ papers in the Top 4 security conferences used surveys in the past 5 years

in addition to 100+ security-related papers in SOUPS & CHI

Research question:
How generalizable are security & privacy surveys*?

*in the USA

Ingredients of a Survey

Research Questions (What Do I Want to Know)?

Constructs (What Do I Need to Measure to Answer RQs)?

Questions (How Can I Validly Measure My Constructs?)

Sample (Who Should Answer My Survey?)

Analysis (How Can I Answer My Research Question?)

Surveys vs. log data
Redmiles et al. CCS18

What kinds of survey samples exist?

Cost

Probabilistic

- (only possible with phone or paper)

A quick primer on survey weighting

Jungle Population (n=1000)

🐯 500 🐵 300 🦄 200

Watering Hole Sample (n=100)

🐯 10 🍫
30 🍨

🐵 15 🍫
15 🍨

🦄 20 🍫
10 🍨

A quick primer on survey weighting

$$Weight = \frac{🐯\; Proportion\; in\; Population}{🐯\; Proportion\; in\; Sample}$$

$$Weight = \frac{\frac{500}{1000}}{\frac{40}{100}} = 1.25$$

Jungle Population (n=1000)

🐯   500

Watering Hole Sample (n=100)

🐯   12.5 🍫
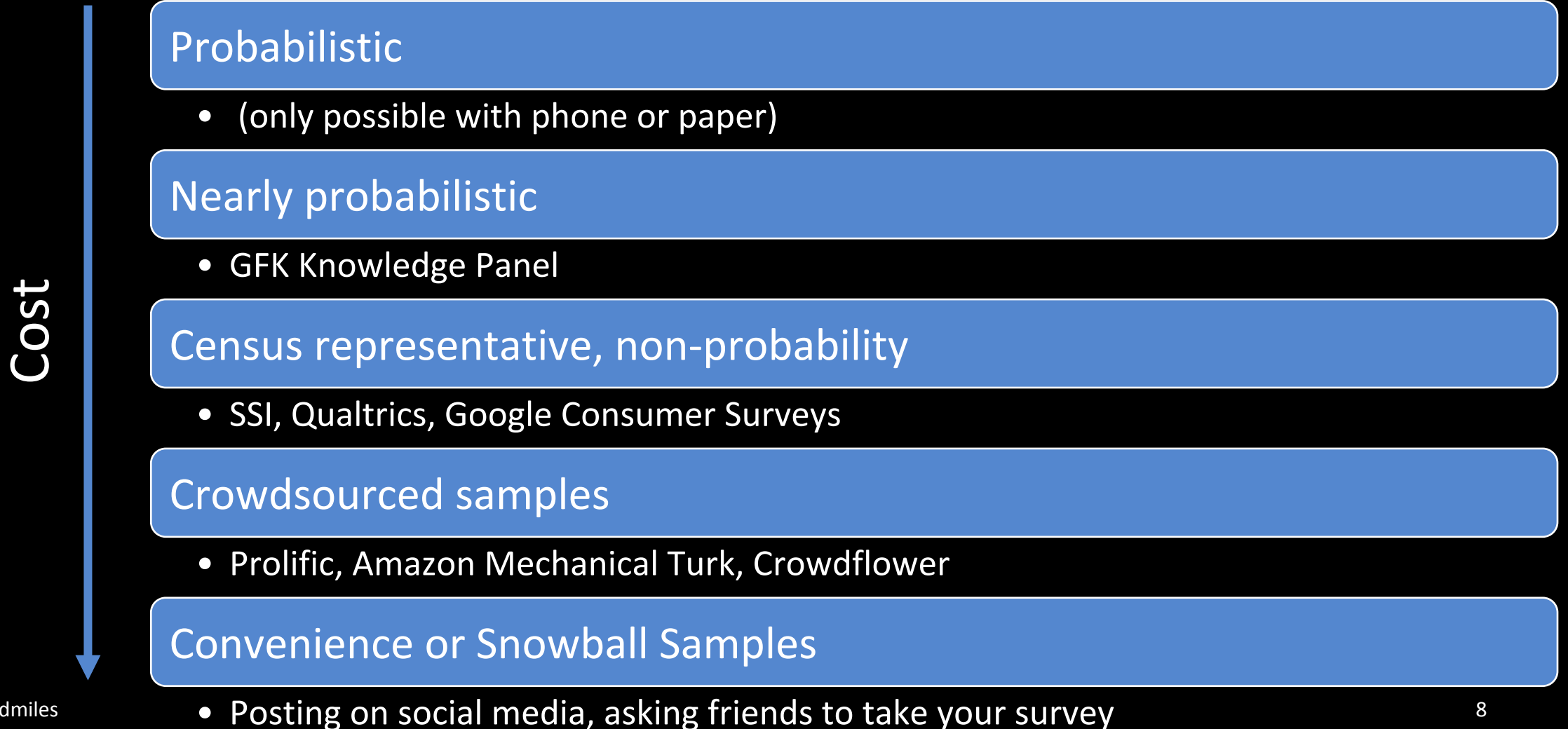   37.5 🍨

🐯   10 🍫
   30 🍨

Without weighting we would reach different conclusions about opinion prevalence

What kinds of survey samples exist?

Cost

Probabilistic

- (only possible with phone or paper)

Nearly probabilistic

- GFK Knowledge Panel

Census representative, non-probability

- SSI, Qualtrics, Google Consumer Surveys

Crowdsourced samples

- Prolific, Amazon Mechanical Turk, Crowdflower

Convenience or Snowball Samples

- Posting on social media, asking friends to take your survey

Elissa M. Redmiles

8

Statistically compare gold standard responses (representative of the US pop. within 2.7%)

Probabilistic telephone sample (gold standard)

- Mode: telephone
- Probabilistic (CI 2.7%)
- n=3,000
- Price: ~$80,000

Web Panel

- Mode: Web
- Census-rep. panel
- n=428
- Price: $1500

MTurk

- Mode: Web
- Crowdsourced
- n=480
- Price: $500

Compared answers to questions about

Internet Behavior

Information Sources: Online Protection

Knowledge: Protective Behaviors

Negative Experiences

Internet Behavior

- Do you ever use the internet to...?
  - Use **social media** such as Facebook, Twitter, or Instagram
  - Apply for a **job**
  - Apply for **government benefits** or assistance
  - Apply for a **loan** or cash advance
  - Search for **sensitive health information**
  - **Buy** a product, such as books, toys, music, or clothing

Internet Behavior

Information Sources: Online Protection

- To which of the following have you turned to for **advice** about how to **protect your personal information** online?
  - Friend or Peer
  - Family Member
  - Co-worker
  - Librarian or resource at library
  - Government website
  - Website run by a private organization
  - Teacher

Internet Behavior

Information Sources: Online Protection

Knowledge: Protective Behaviors

- Do you feel as though you already know enough about...?
  - Choosing **strong passwords** to protect your online accounts
  - Managing **privacy settings** for the information you share online
  - Understanding the **privacy policies** of the websites and applications you use
  - Protecting the security of your devices when **using public WiFi** networks
  - Protecting your computer or mobile devices **from viruses and malware**
  - **Avoiding online scams** and fraudulent requests for your personal information

Internet Behavior

Information Sources: Online Protection

Knowledge: Protective Behaviors

Negative Experiences

- As far as you know have you ever…?
  - **Had important personal information stolen** such as your Social Security Number, your credit card, or bank account information?
  - Had **inaccurate information show up in your credit report**?
  - Had an **email or social networking account of yours compromised** or taken over without your permission by someone else?
  - Been the **victim of an online scam** and lost money?
  - Experienced **persistent and unwanted contact** from someone online?
  - **Lost a job opportunity** or educational opportunity because of something that was posted online?
  - Experienced **trouble in a relationship** or friendship because of something that was posted online?
  - Had **someone post something about you online that you didn't want shared**?

Comparative Sample Analysis

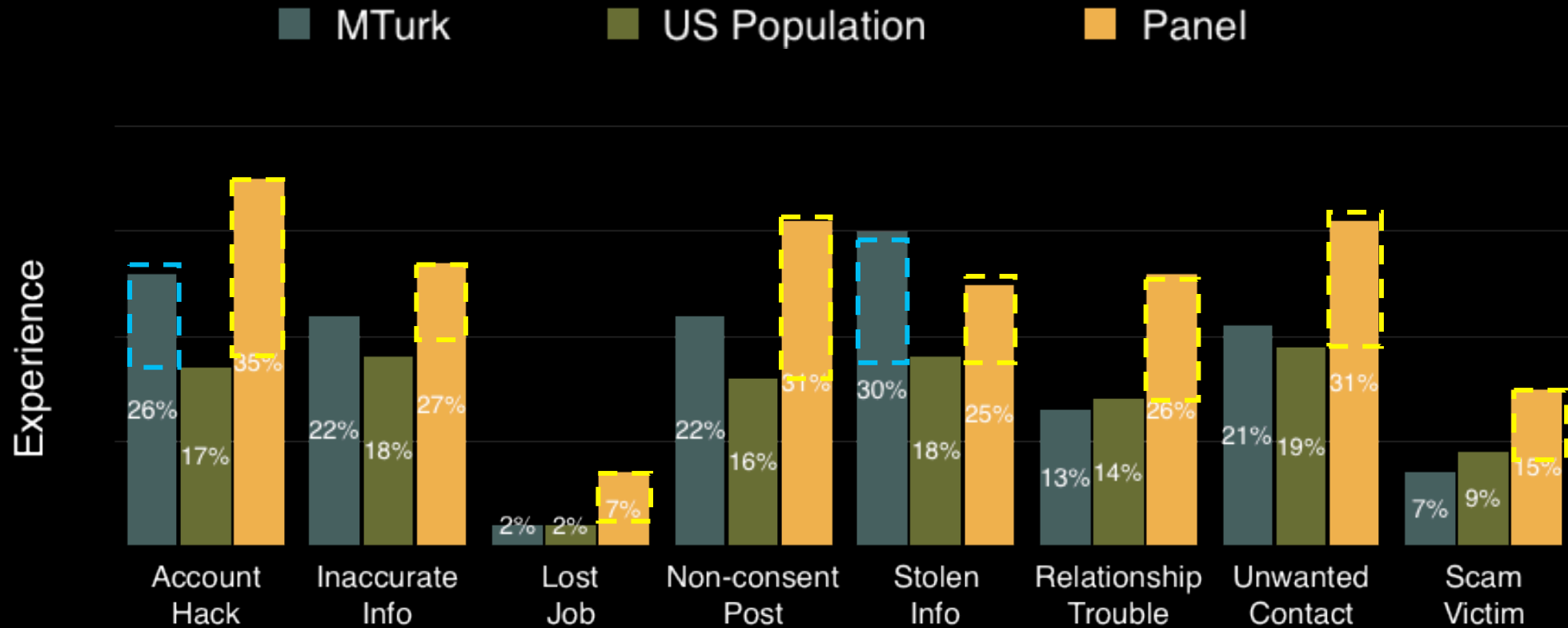Question-by-question X^2 proportion tests (Bonferroni correction)

Check our stats! Analysis code released with the paper

Overall

Age

Education

Both web samples significantly more likely to engage in variety of online behaviors

Legend: MTurk | US Population | Panel

Behavior	MTurk	US Population	Panel
Gov. Benefits	38%	21%	39%
Health	66%	49%	58%
Job	79%	43%	61%
Loan	22%	14%	27%
Product	99%	76%	90%
Social Media	97%	69%	91%
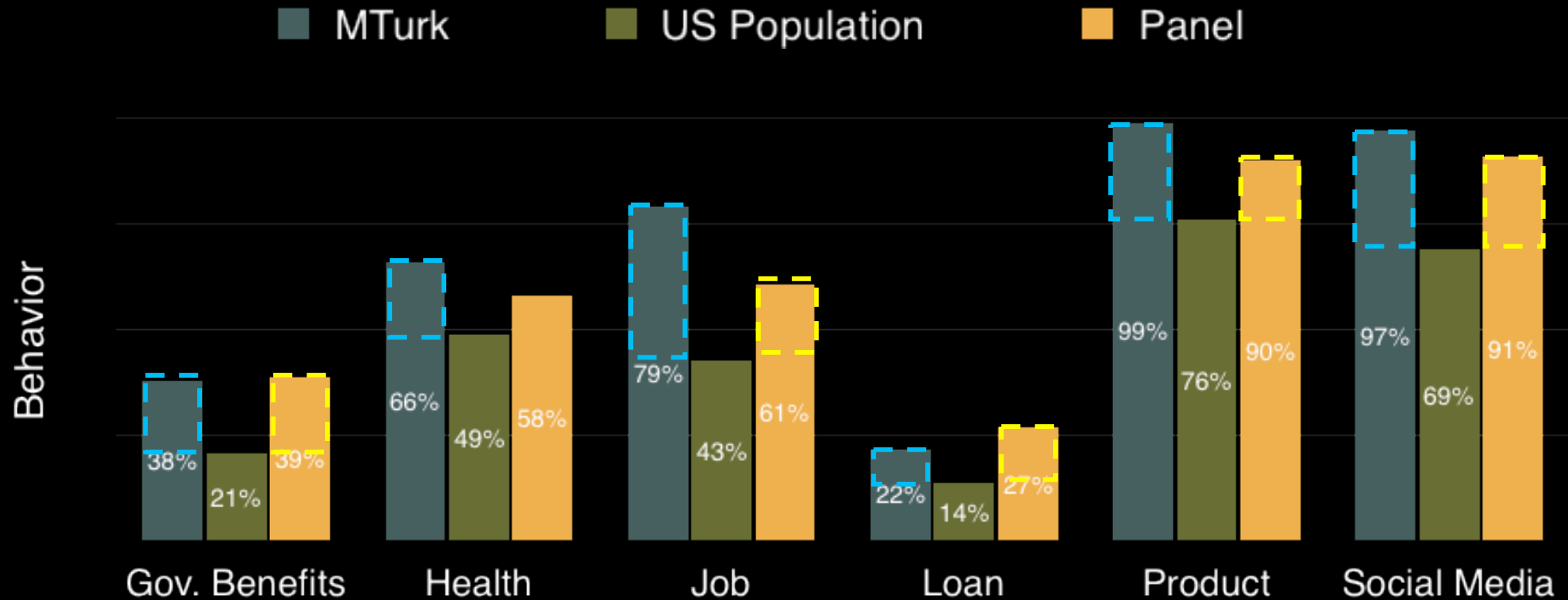
Census-rep. web panel significantly more likely to report negative experiences

Higher reporting of negative experiences may be related to more internet use

Both web samples significantly more likely to report seeking advice from websites

Legend: MTurk | US Population | Panel

	Co-Worker	Friend	Librarian	Teacher	Website
MTurk	16%	43%	3%	3%	58%
US Population	20%	38%	6%	7%	21%
Panel	16%	48%	3%	3%	30%
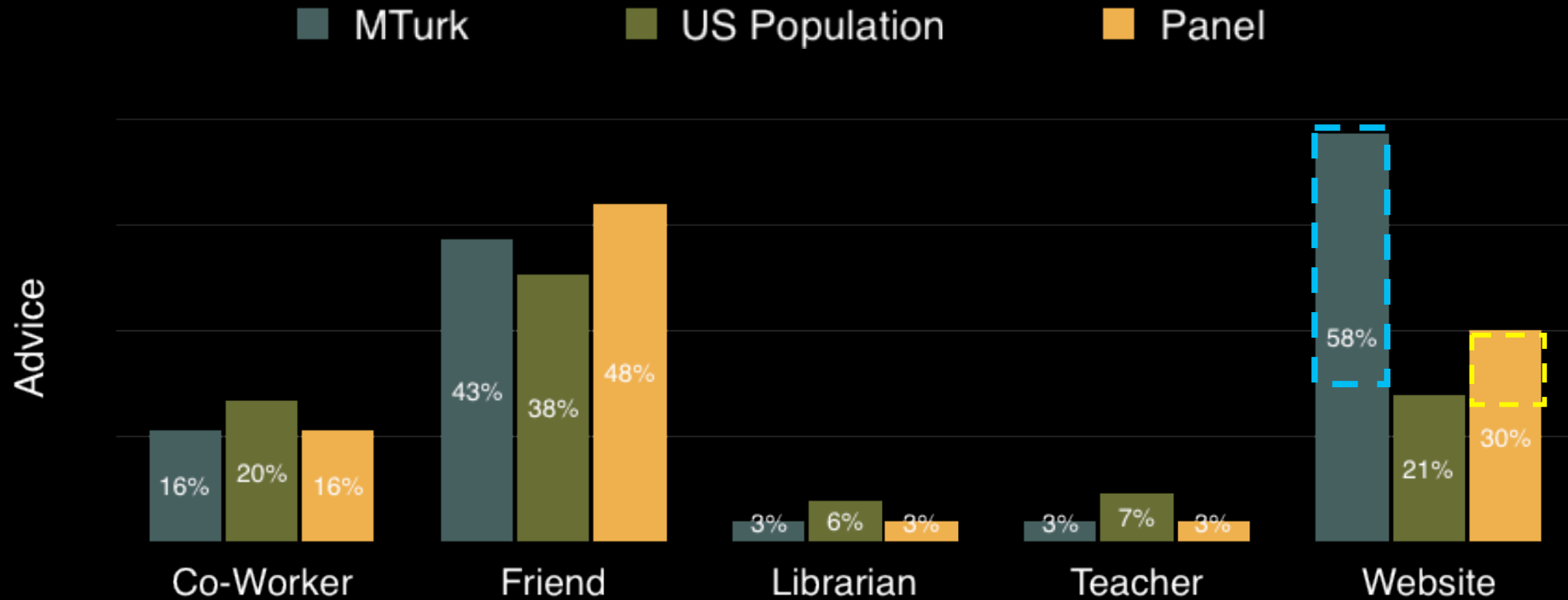
Web sample respondents are more likely to report seeking advice & seek advice from more sources

Census-rep. web panel significantly less likely to feel knowledgeable about security & privacy
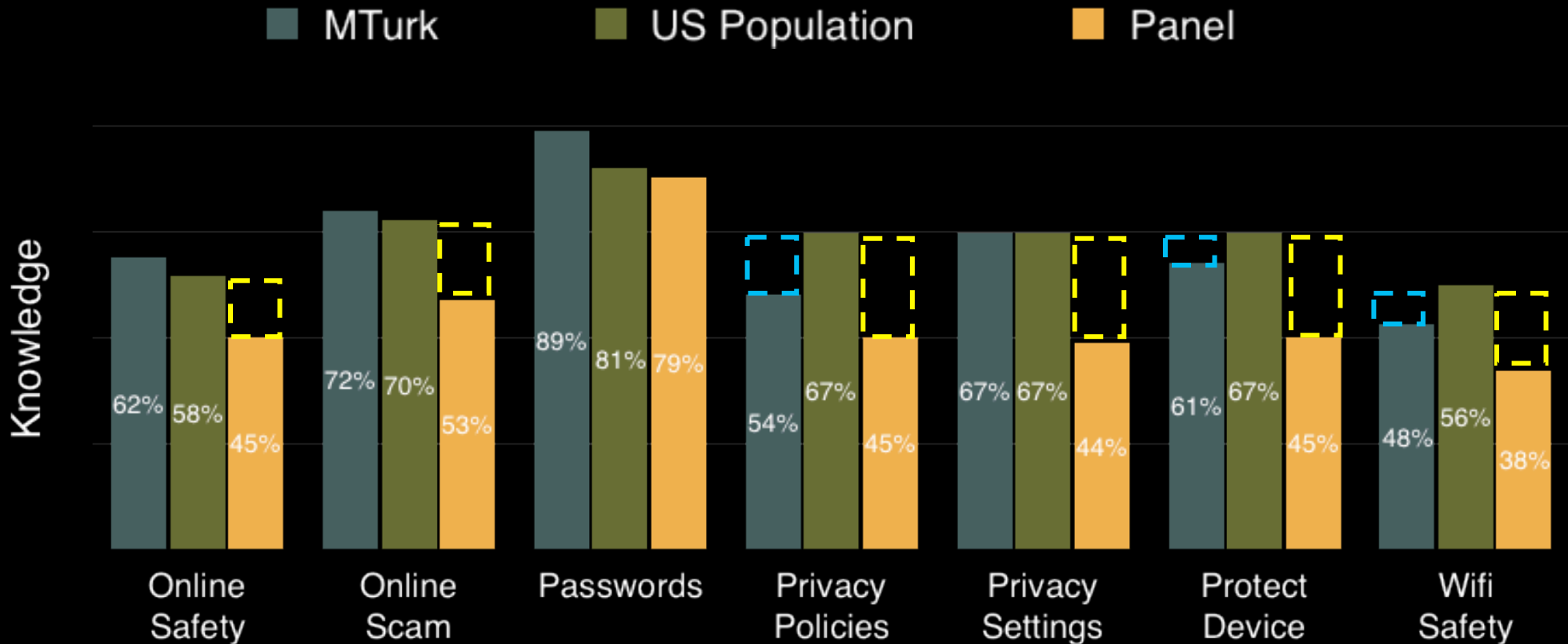
All samples report similarly re: passwords – 80% or more feel like they know enough!

Comparative Sample Analysis: By Age Subgroup

Question-by-question X^2 proportion tests (Bonferroni correction)

Check our stats! Analysis code released with the paper

Overall

Age

Education

18-29 years old

- MTurk differs on only 6 questions

30-49 years old

- MTurk differs on 8 Qs, Panel on 9Qs
- Mturk reports more Behaviors, Panel reports less Knowledge

50+ years old

- Both web samples differ a lot, MTurk more so
- MTurk differs on everything except advice
- Panel reports more Behavior.& Neg. Experiences

Comparative Sample Analysis: By Education Subgroup

Question-by-question X^2 proportion tests (Bonferroni correction)

Check our stats! Analysis code released with the paper

Overall

Age

Education

HS education or less

- Panel is the only sample with enough participants; 10 Qs differ

Some college education or above

- MTurk more similar to US pop.
  - More online behavior for SC & BS+
  - More online behavior & less knowledge for BS+

Proposed mitigation:
demographic weighting of Mturk data

Not much, reduces from 14 differences overall to 11

This has worked in other survey applications, but in security the weighting variables might not be strictly demographic

How do I pick a sample?

Do you need to draw conclusions that generalize to all U.S. users?

Yes

No

Use multiple samples
OR
Try probabilistic or near-probabilistic samples
(e.g., conduct survey manually from a purchased
prob. list or try GCS / KnowledgePanel)
OR
Future: weight survey results to better generalize

For what population would you like
your results to generalize?

Age: 18-49 yrs
Ed.: some college

Age: 50+ yrs
Ed.: H.S. or less

Mturk Sample

Census-representative
web panel

Where do we go from here?

Develop statistical mitigations
Test weighing samples on security-specific variables
Develop custom weights for standard security measures

How Well Do My Results Generalize?
Comparing Security & Privacy Survey Results from MTurk, Web, and Telephone Samples

Elissa M. Redmiles, Sean Kross,, and Michelle L. Mazurek

Questions? eredmiles@cs.umd.edu

Research Question
How generalizable are security & privacy survey results?

Methods
Statistically compare probabilistic sample of US pop. (CI 2.7%) to MTurk and census rep. web panel samples

Findings
MTurk more generalizable for 18-49yr olds w/ some college
Panel or prob. more generalizable for 50+, HS or less

Additional Security Survey Resources
go.umd.edu/survey-meth

Backup

Survey modes != samples

Paper
- Good for low tech populations, rarely used

Phone
- Good for low tech populations
- Allows for "probabilistic" sample
- CATI: computer assisted telephone interviewing

Web
- Often used in security, privacy, HCI studies
- Highest non-response rate of any mode
- Cheapest

Time comparison

- Probabilistic sample collected in December 2015
  - Compared to Rader & Wash 2013 sample
  - Only one significant difference (more respondents reported having information stolen in 2015)

- MTurk and Panel samples collected in January 2017 and in March 2018 (after Cambridge Analytica)
  - Only difference, fewer MTurkers reporting purchasing products online in 2018

Sample Demographics

Metric(%)		MTurk	Panel	Prob.W	Prob.UW	Census
Sex	Male	50	49	49	52	48
	Female	48	51	51	48	52
Race/Ethn.	Caucasian	84	69	63	58	66
	Hispanic	4	12	16	19	15
	African American	10	14	12	14	11
	Other	5	7	7	7	8
Education	LT H.S.	0.4	3	13	13	13
	High School	12	31	28	27	28
	Some college	41	34	30	24	31
	B.S. or above	46	31	29	35	28
Age	18-29 years	20	27	20	16	21
	30-49 years	58	23	33	24	35
	50+ years	22	49	44	56	44
Income	<$30k	25	28	NA*	34	32
	$30k-$50k	24	23.5	NA*	15	19
	$50k-$75k	26	19	NA*	11	18
	$75k-$100k	12	13	NA*	9	11
	$100k-$150k	8	10	NA*	8	12
	$150k+	3	5	NA*	7	10

Why can't we just use existing survey methodology sample literature?

Asking about online behavior on the internet is **different** that asking about e.g., smoking behavior!

A quick primer on survey weighting

Survey raking
Weight iteratively over multiple variables
with known distribution (e.g., age, race, etc.)

CCS18: When to use survey vs. log data

Research Question

How well do survey and log data align for questions regarding user security behavior?

Methods

Compare log (n=517,932) and survey (n=2,092) data about software updating

Findings

Surveys approximate general not detailed constructs

Take Aways

Use surveys for perceptions & broad reactions
Try filtering non-sensical responses
Use observation for assessing detailed variations

Redmiles, E.M., Zhu, Z., Kross, S., Kuchhal, D., Dumitras, T.., and Mazurek M.L.. *Asking for a Friend: Evaluating Response Biases in Security User Studies.* ACM **CCS**

CCS18: Carefully designed survey & selected test cases

Imagine that you see this message appear on your computer.
Would you install the update?

Adobe Reader Updater

Update is ready to install

This update addresses customer issues and security vulnerabilities. Adobe recommends that you always install the latest updates.

You may need to restart your system after the update is installed.

Details

Install   Cancel

- Yes, the first time I saw this message.
- Yes, within a week of seeing this message.
- Yes, within a few weeks of seeing this message.
- Yes, within a few months of seeing this message.
- No.
- I don't know.

Detailed | Application
| Update Cost
| Security-Only
| Message Length

General | Update Risk

| Tendency to Update

Redmiles, E.M., Zhu, Z., Kross, S., Kuchhal, D., Dumitras, T.., and Mazurek M.L.. *Asking for a Friend: Evaluating Response Biases in Security User Studies.* ACM **CCS**