



MARY McDOUGALL

SURVOX, INC.

EXPERIMENTS WITH MOBILE DATA COLLECTION

EXPERIMENTS WITH MOBILE DATA COLLECTION



**Vinay
Bapna**

Unomer

Embedding Survey Questions Within Non-research Mobile Apps:
A method for collecting high-quality data from smartphone users?



**Michael
Schober**

New School for
Social Research

Hesitation In Socially Desirable Responses In A Mobile Phone Survey



**Jerry
Timbrook**

U. of Nebraska-
Lincoln

Why Do Mobile Interviews Take Longer? A Behavior Coding Perspective



**Rachel
Horwitz**

U.S. Census
Bureau

The Differential Effect of Mobile-Friendly Surveys on Data Quality



Mobile In-App Surveys

Embedding survey questions within non-research apps: A method for collecting high-quality data from smartphone users?

Vinay Bapna, Unomer
Christopher Antoun, U.S. Census Bureau

In-App Survey Distribution



Distribute to specific target audience through a network of partner apps



Instantaneous, in-the-moment survey responses

The Promise In-app Surveys

- For Researchers

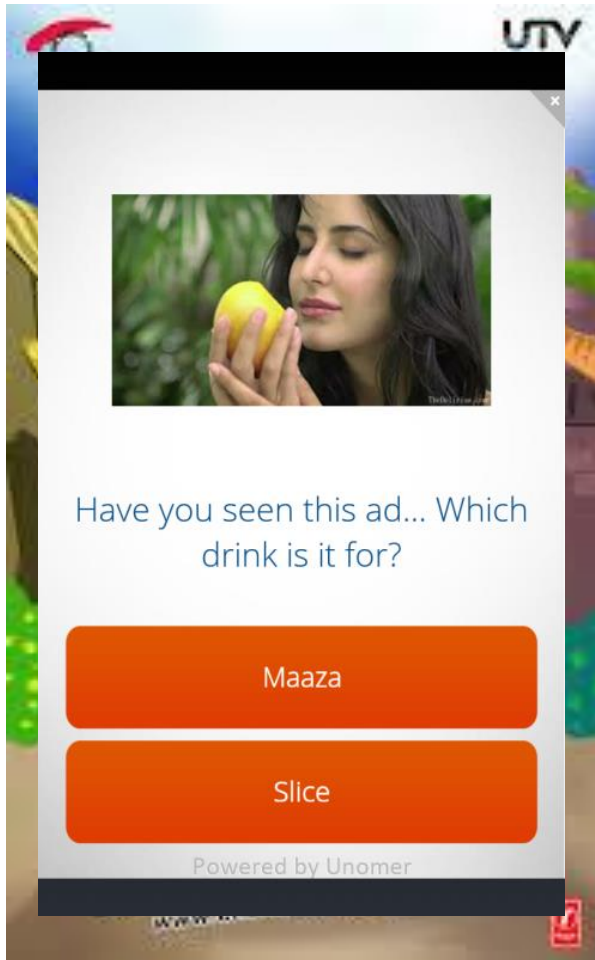
- Deliver simple, visual, interactive in-app surveys
- Embedded and distributed in popular mobile apps
- Supplement survey data with passively collected data
- Gamified with in-app incentives for respondents
- Targeted based on user demographics, profile, behavior
 - Not for general population samples

- For App Publishers

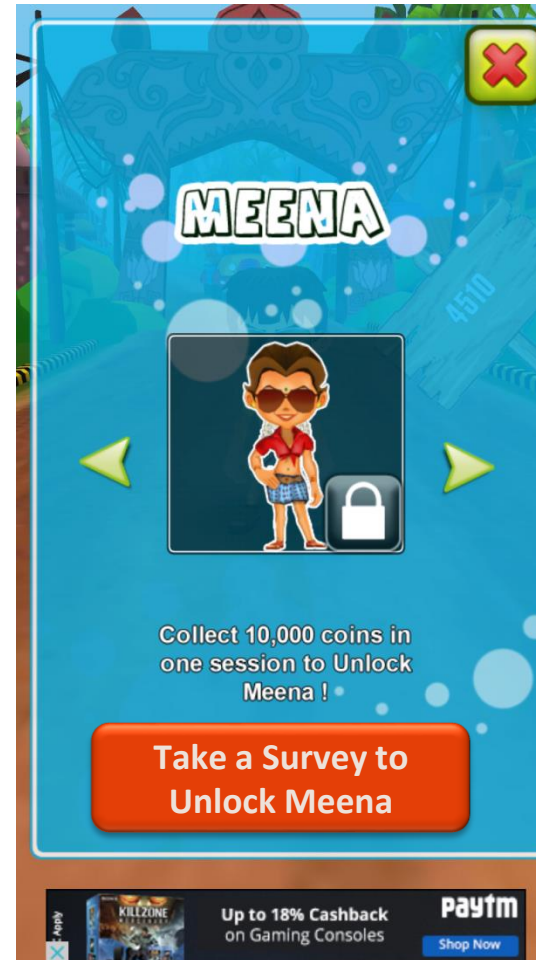
- Users can take short in-app surveys and earn incentives
- Contributes revenues, helps engages users
- Seamless within app flow – interstitials, nifty placement, event triggers



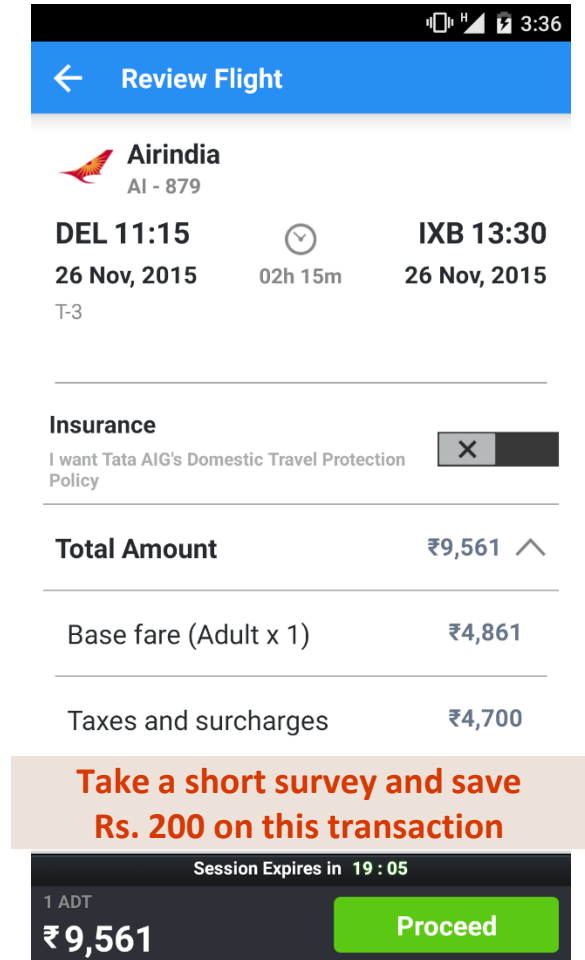
In-app Survey Placements



INTERSTITIALS



STRATEGIC PLACEMENTS



EVENT TRIGGERS



Can In-apps Surveys Disrupt Research?



REACH



TARGETING



ENGAGEMENT

Scalable network

- Each app gives access to millions of potential respondents
- Naturally expanding, growing in multiplies

Apps tell it all

- Behavior, spending, interests
- Plus passive mobile data

Spontaneous

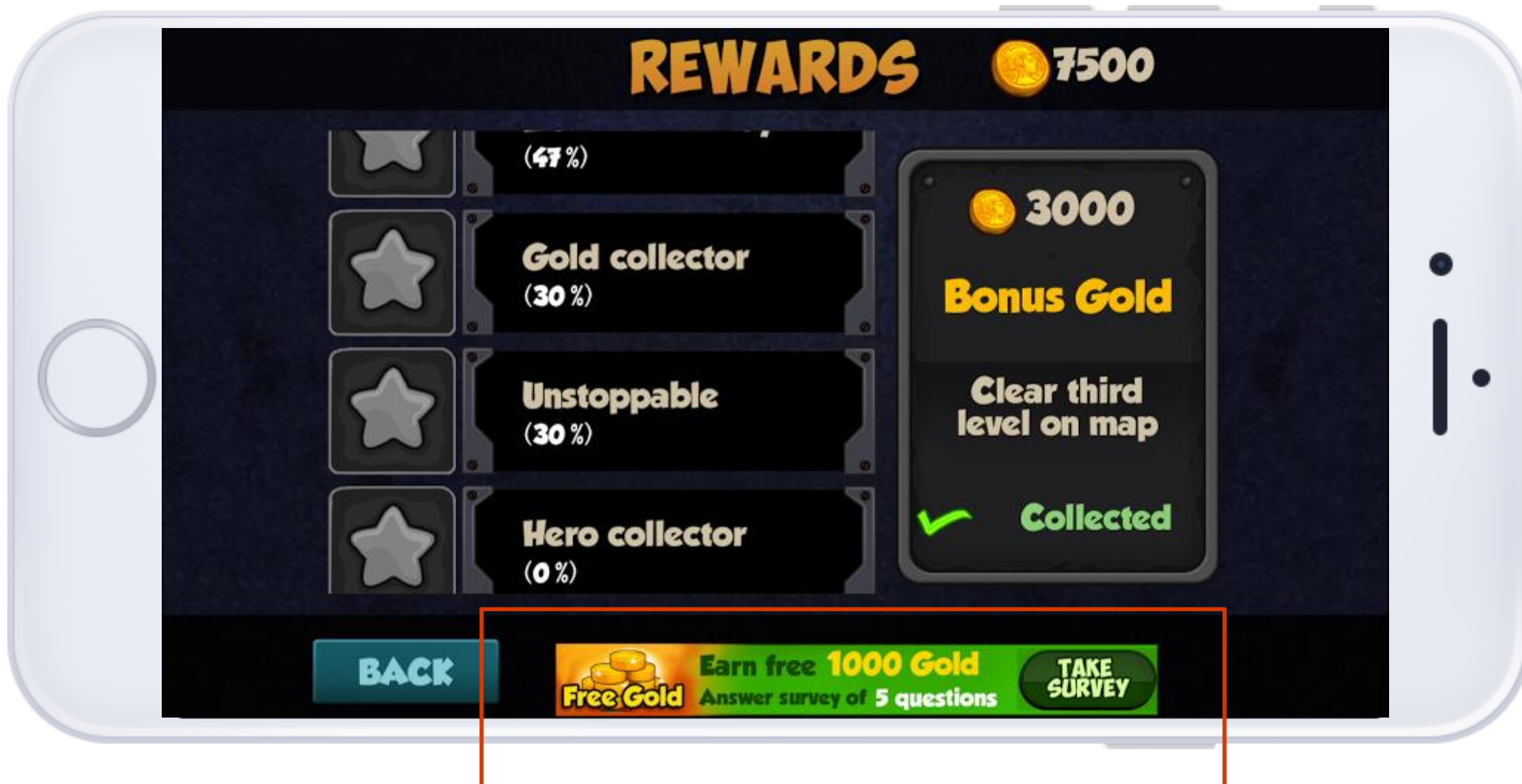
- In-the-moment, real-time
- Interactive, rewarding, gamified



Current Study (Phase 1)

- Carried out in April 2016 (over two week period)
- Survey questions shown in gaming app
 - Answered in exchange for in-app currency (gold coins)
- Surveys (2, 3, 4 questions)
 - Demos + Q about LinkedIn app; $n = 101$
 - Demos + Q about mobile wallet app + Q about phone brand; $n = 100$
 - Demos + 3 Qs about health: $n = 87$
- Passively collected information
 - Country, brand of device, apps installed on phone





Profile of Participants Who Responded (n = 288)

	% of sample		% of sample
Age		Country (51!)	
18-24	35.1	India	13.9
25-34	23.6	U.S.	9.7
35-44	21.2	Russia	8.7
45-54	14.9	Philippines	8.3
55-65	3.1	Brazil	6.3
65+	2.1	Other	53.1
Gender		Devices Used	
Male	71.9	Smartphone	82.3
Female	28.1	Tablet	17.7

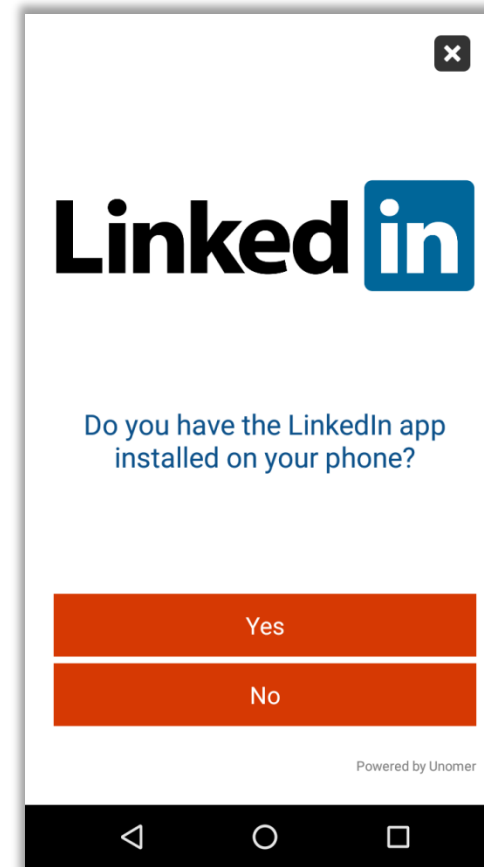
Data Quality Checks

- #1: How accurate are self reports compared to passively collected data?
 - *Compare reported phone brand to the actual phone brand*
- #2: How does predictive validity compare to a traditional sample?
 - *Compute correlations between self-rated health questions and number of doctors visits per year*



Experiment 1 and 2: Do images improve recall?

- Expectation: Images will assist retrieval of information
- #1 *Do you have the LinkedIn app installed on your phone*
 - *Half see image of this app; other half do not*
- #2 *Do you have any mobile wallet apps installed on your phone*
 - *Half see images of such apps; other half do not*



* Do images affect responses in the same way as in conventional web surveys

Data Quality Check #1: Comparing Self Report to Passively Collected Data (n=100)

- *"What brand of phone are you using?"* Motorola, Samsung, Sony, or None of these

Reported Phone Brand	Actual Phone Brand			
	Motorola	Samsung	Sony	None of these
Motorola	3	3	0	3
Samsung	0	22	0	5
Sony	0	2	5	2
None of these	2	5	0	48

Accuracy Rate: 78.0%

Experiment 3: Do images affect responses in the same way as in conventional web surveys

- Expectation: Images serve as standard of comparison affecting judgements
 - *“In general, how would you rate your health?”*
 - *Half see image of sick woman; other half see image of fit woman*



Experiment 3: Do images serve as standard of comparison affecting judgements? (n=87)

- Self-rated Health (SRH-4): "In general, how would you rate your health?" 4. Very good; 3. Good; 2. Fair; 1. Poor*

	Self report		T Test
	Sick	Fit	
Mean	3.22	3.06	$t(85) = 0.89, p = .373$

As we expected, those who saw image of sick woman rated themselves as healthier than those who saw image of fit woman, but the difference was not statistically significant (perhaps due to small sample size).

Experiments 1 & 2: Do Images Improve Recall?

(Benchmark is passively collected data)

- "Do you have the LinkedIn app installed on your phone?"*

Self report	Benchmark	
	Yes	No
No Image (n=51)		
Yes	1	8
No	1	41
Image (n=50)		
Yes	1	11
No	2	36

Accuracy Rates with no image: 82.4%

Accuracy Rates with image: 74.0%

High accuracy, no significant effect of image

- "Do you have any mobile wallet apps installed on your phone?"*

Self report	Benchmark	
	Yes	No
No Image (n=50)		
Yes	0	17
No	1	32
Image (n=50)		
Yes	2	16
No	2	30

Accuracy Rates with no image: 64.0%

Accuracy Rates with image: 64.0%

No significant effect of image

Summary of Results

- Global sample - 51 countries!
- Efficient – 288 responses thru 1 app in less than 2 weeks
- Relatively high accuracy rates – for answers verified with passively collected data
- Lower than expected correlations – related to sample composition, responses option formats
- No significant effect of images
 - Effect on engagement to be investigated



Conclusions

Promise of In-app Surveys

- Ability to conduct surveys with very large sample sizes
- Instantaneous, in-the-moment
- App-based targeting – apps are proliferating
- In-app incentives a new way to engage
- Passive data for deeper analysis

Planned Studies

- Response quality
- Response rates - length of survey, effect of different types of incentives
- Country level studies, comparison with other surveys





Thank you!

Vinay Bapna
vbapna@unomer.com
(650) 331-0136

Hesitation in socially desirable responses in a mobile phone survey

Stefanie Fail
Michael F. Schober



Frederick G. Conrad



Answering on a mobile device

- As survey respondents increasingly answer on mobile devices (Link et al., 2014), what aspects of responding in landline interviews will generalize?
- Quite different dynamics may be at play when respondents are
 - mobile
 - multitasking
 - potentially more affected by ambient noise
 - potentially more affected by presence of other people

Our focus

- *Hesitations* (pauses)—one kind of nonverbal “paradata” that respondents produce along with their answers
 - (along with *ums* and *uhs*, hedges, etc.)
- in surveys that include *sensitive questions* (questions that can lead to socially desirable responding) in a convenience sample of iPhone users

Survey Q's on sensitive topics

(likely to have more and less socially desirable answers)

- “How often do you now smoke cigarettes? Every day, some days, or not at all? ”
- “During the past 30 days, on how many days did you have 5 or more drinks on the same occasion?”
- “How many sex partners have you had in the last 12 months?”

Compare with

- “How many songs do you currently have on your iPhone?”
- “During the last month, how many times did you shop in a grocery store? ”
- “During the past 12 months, how many movies have you seen in movie theaters?”

Suggestive evidence in other domains

- Hesitation linked with pressure to respond in socially desirable ways in online self-reports of traits and behaviors
 - People hesitate longer when told that responses will be used to create a psychological profile (Holtgraves, 2004)

Hesitation linked with lying

- At least when people are instructed to lie in lab
- But different effects for different kinds of questions
 - longer hesitations better predict lying more for yes/no than open ended questions (Walczyk et al., 2005)

→ Will similar patterns be observed in a smartphone survey, where Rs might be around other people, mobile, multitasking, distracted?

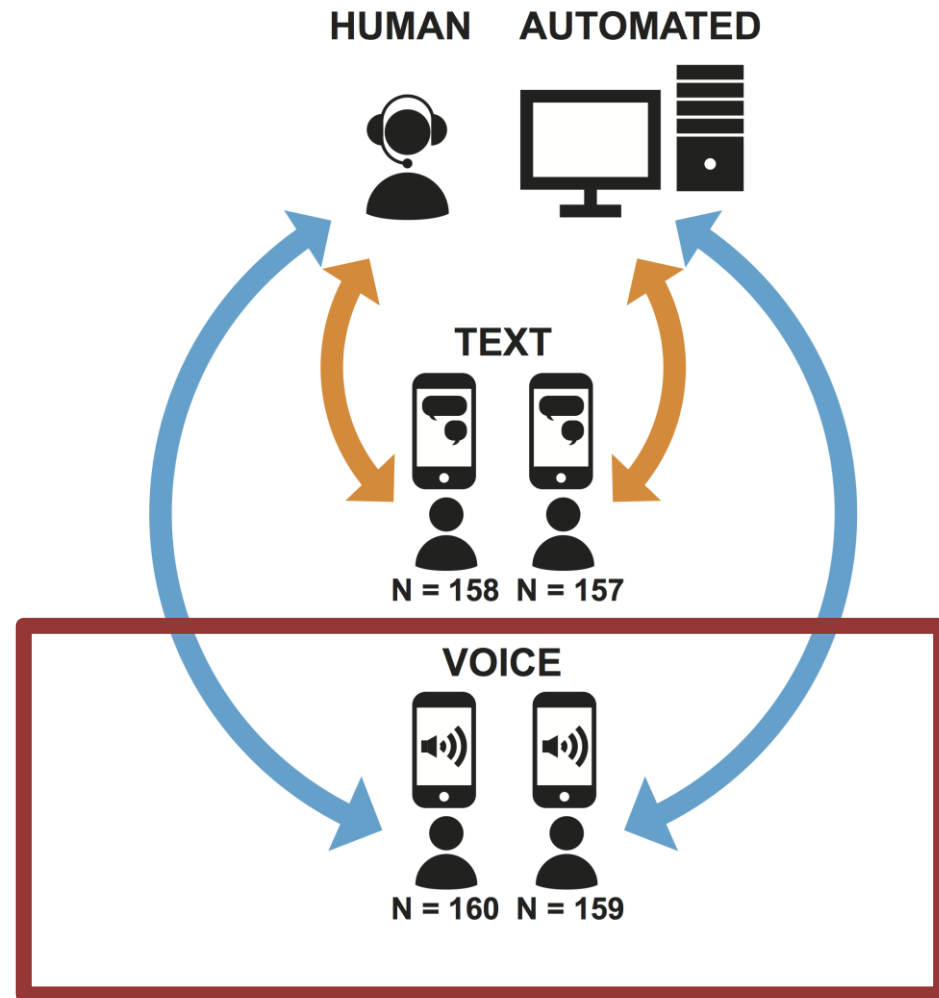
Hesitation with human vs. computer interviewer?

- Respondents reveal sensitive behaviors more when questions asked by a computer than a human
 - e.g., Tourangeau & Smith, 1996; Turner et al., 1998, among many others

→ Will smartphone respondents
hesitate differently when interviewed by
an automated interviewing system (speech IVR)
than by an interviewer?

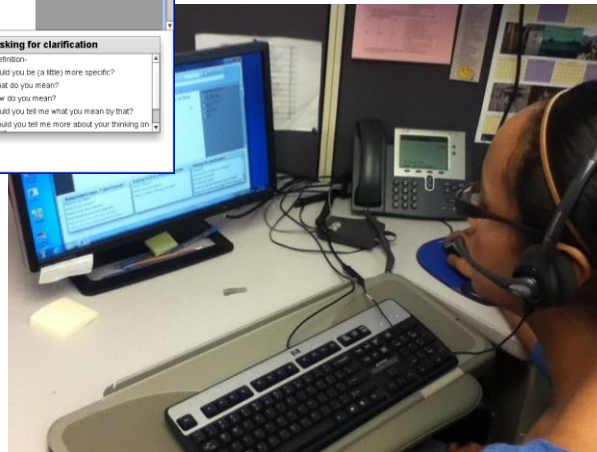
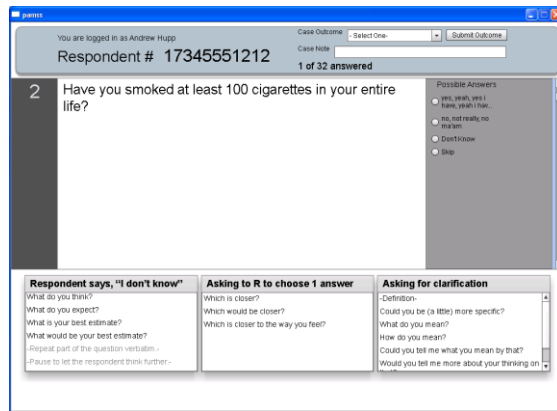
Study

Analyses of audio recordings from corpus of 319 audio-recorded mobile telephone interviews from Schober et al. (2015, PLOS ONE)



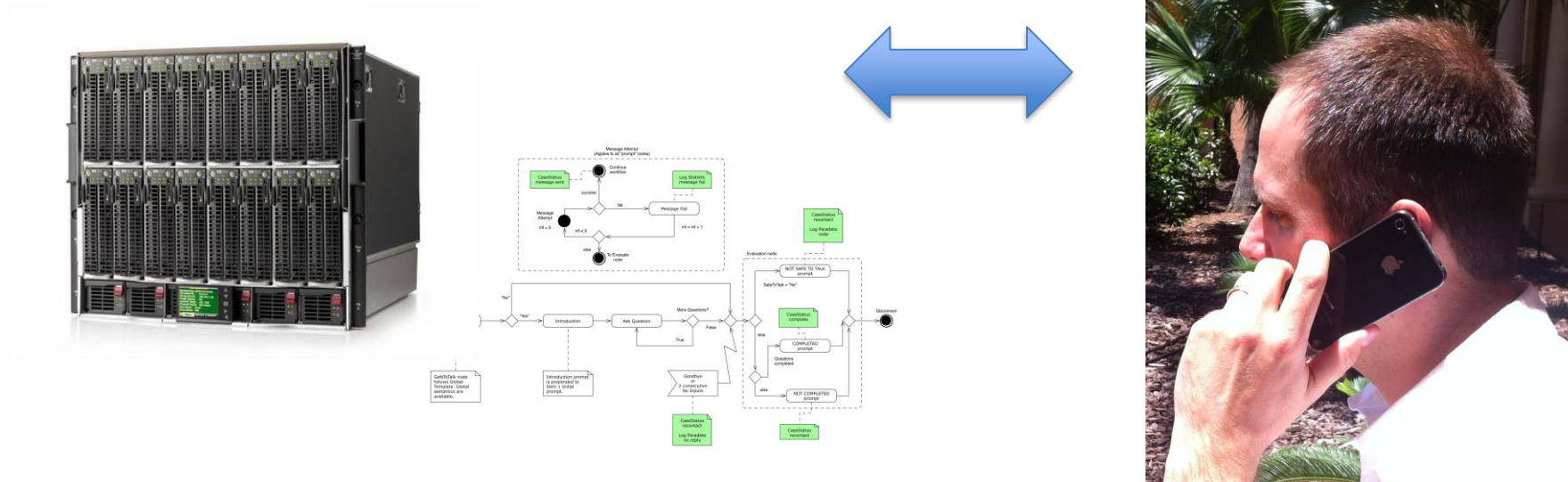
Human Voice interviews

- 10 interviewers (*Is*) from U Mich survey research center
- custom designed CATI interface that supports voice and text interviews (PAMSS)



Automated Voice interviews

- Custom built speech dialogue system
- Uses ATT's Watson speech recognizer, Asterisk telephony gateway
- Recorded human interviewer, speech responses (not touchtone)



Questions

- Respondents answered 32 questions from major US social surveys and methodological studies on their iPhones, March-May 2012
- Questions that we deemed sensitive and non-sensitive that we could (now) match on:
 - Response formats: dichotomous (yes/no), multiple choice, and numerical
 - Recall periods: 1 year, 1 month, no time range

Example matched Qs: 12 month recall, numerical

Nonsensitive

During the past 12 months, how many movies have you seen in movie theaters?

Sensitive

How many sex partners have you had in the last 12 months?

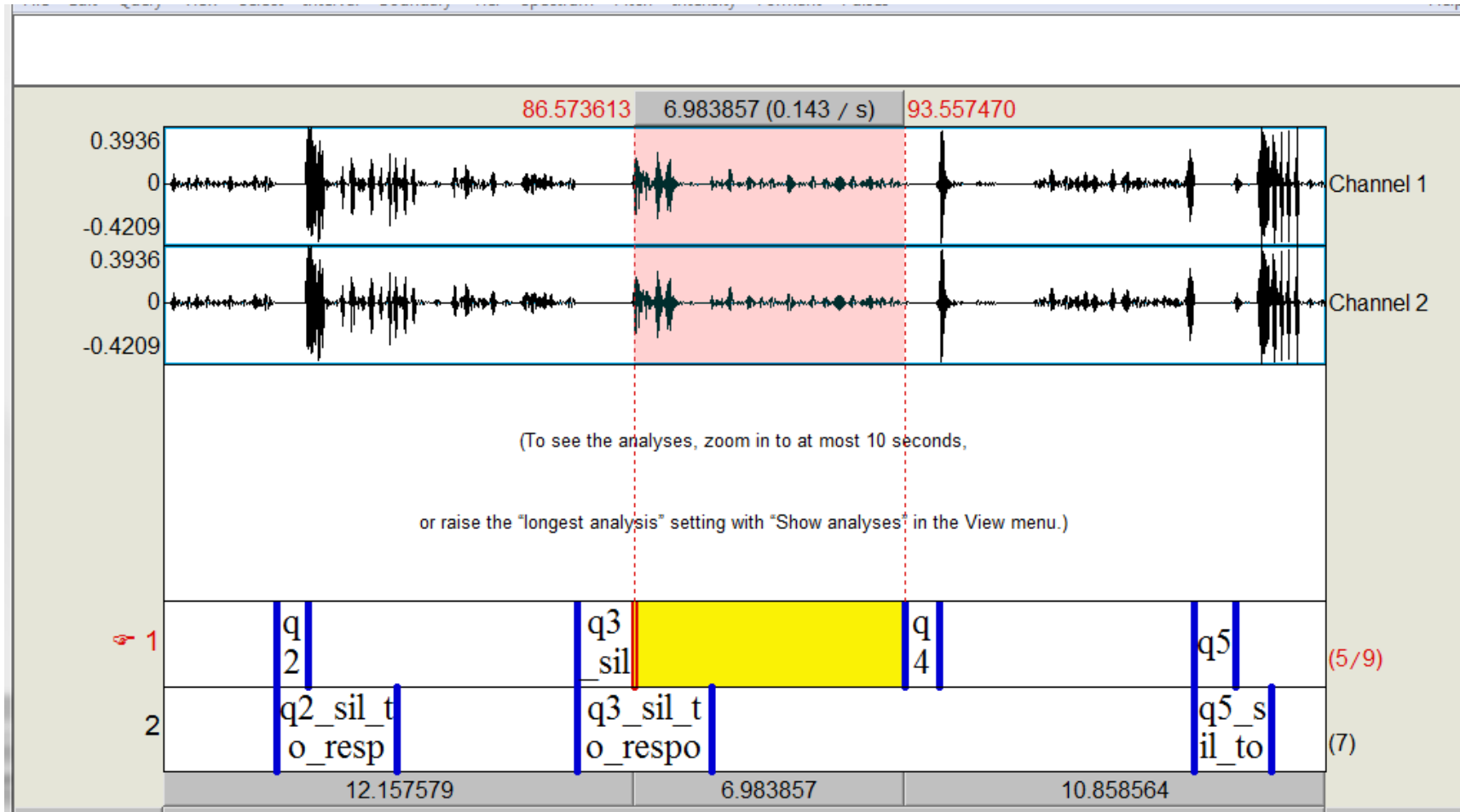
Responses to sensitive questions that we judge as **more** and **less** stigmatized

How often do you now smoke cigarettes: '*every day*', '*some days*' or '*not at all*'?

During the past 30 days, on how many days did you have 5 or more drinks on the same occasion? **>10**

In a typical week, about how often do you exercise? *Less than 1 time per week*, **1 or 2 times per week**, **3 times per week**, or **4 or more times per week**

Annotation of audio recordings of interviews in PRAAT



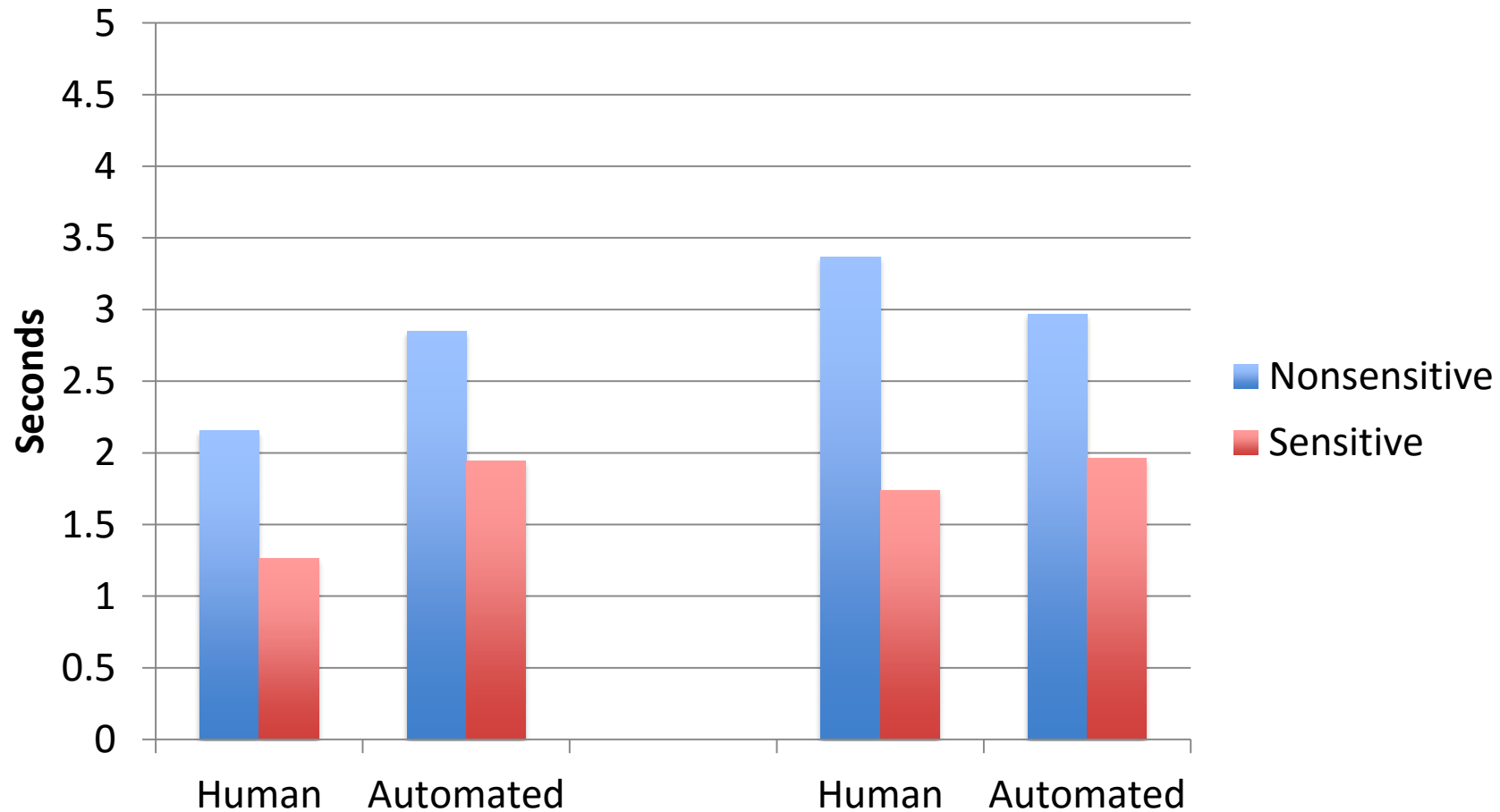
Findings: Sensitive vs. nonsensitive Q's

- Clear pattern across multiple question types and comparisons:
- Rs answered sensitive Qs significantly *more quickly* than nonsensitive Qs, (within-subjects comparison—SAME respondents)
 - different pattern than seen in other domains
- Also, significantly slower with automated than human interviewers

12-month recall Q's

Time to first sound

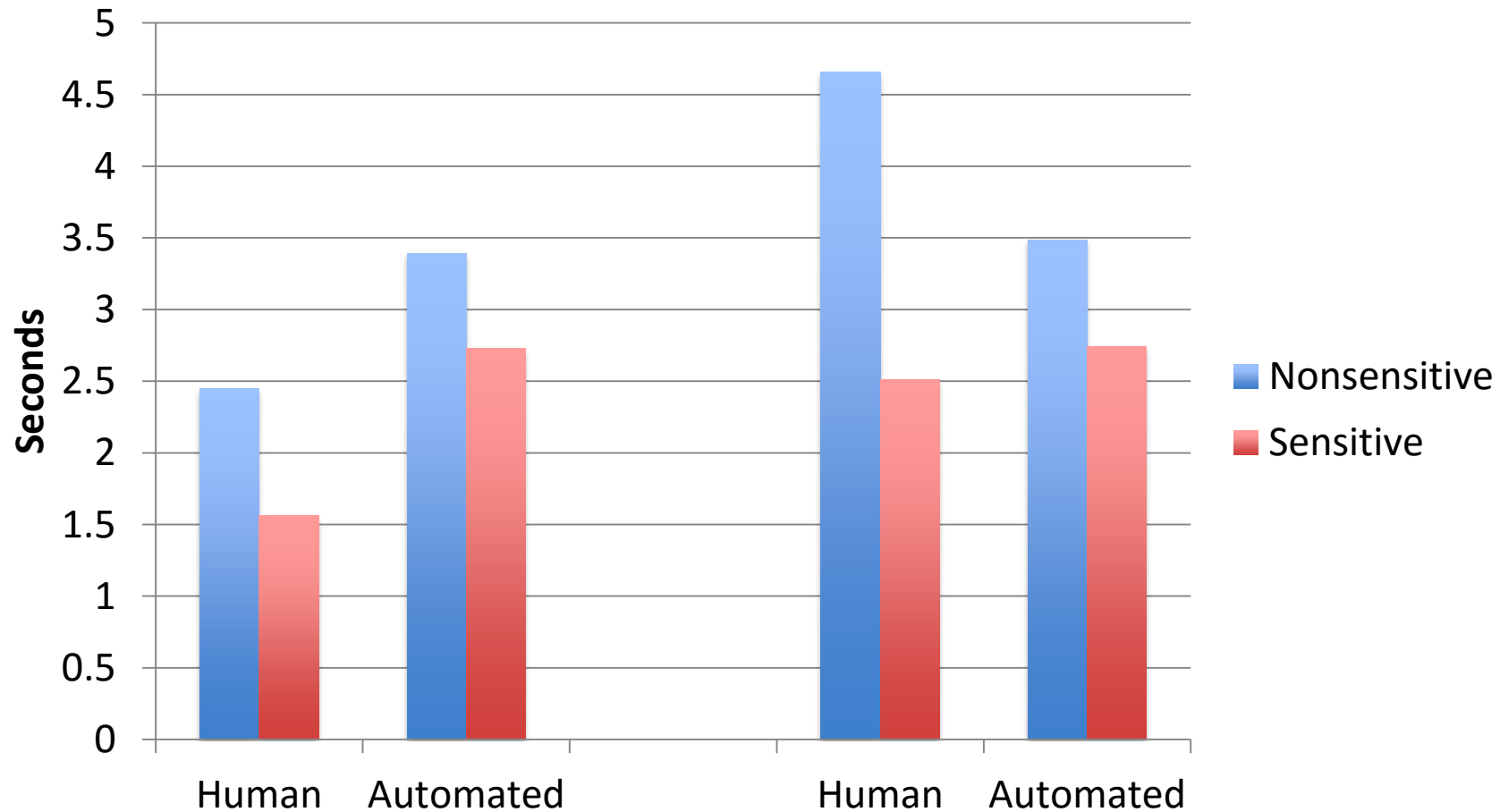
Time to first response



1-month recall Q's

Time to first sound

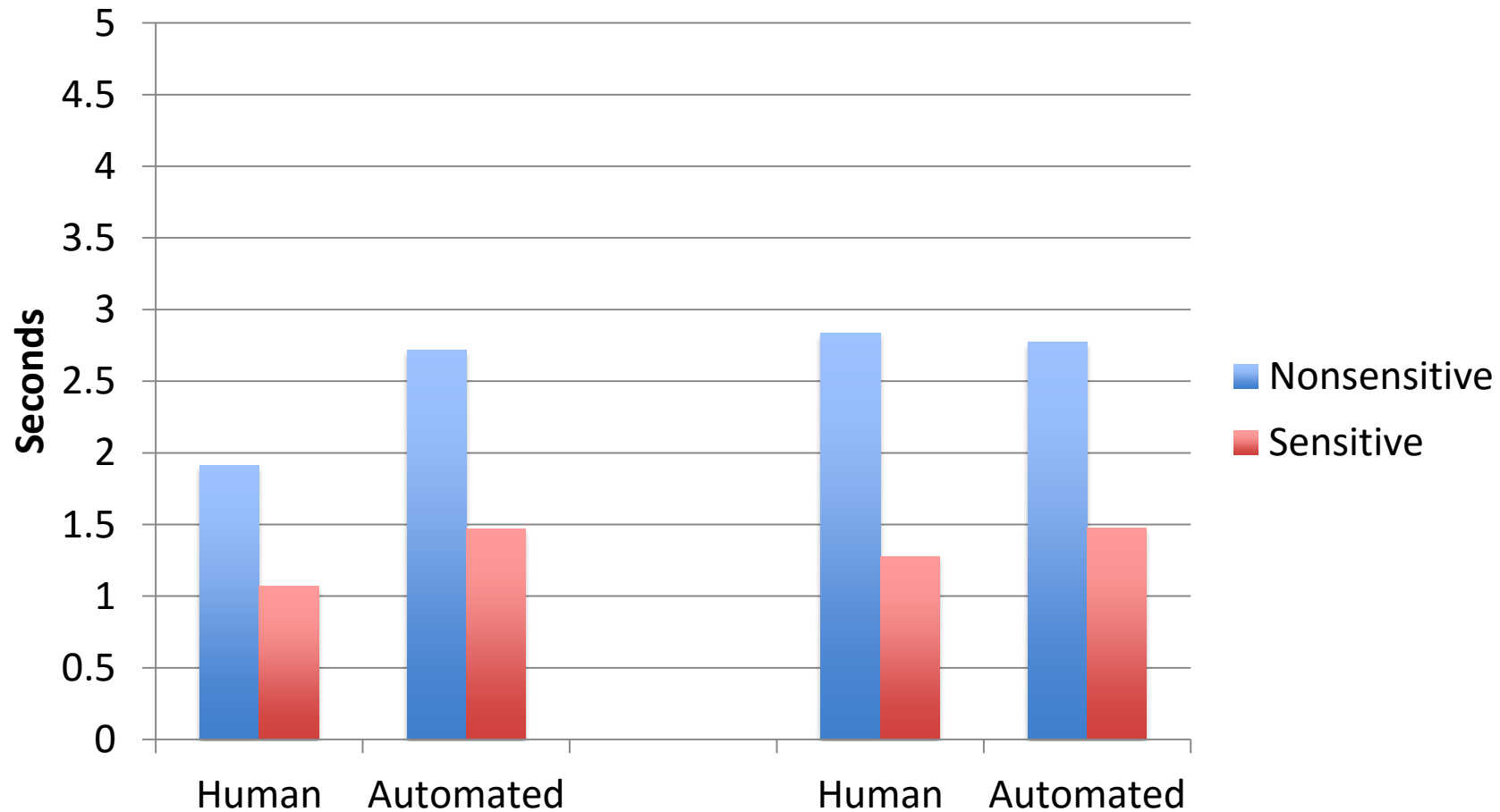
Time to first response



No time range Q's

Time to first sound

Time to first response



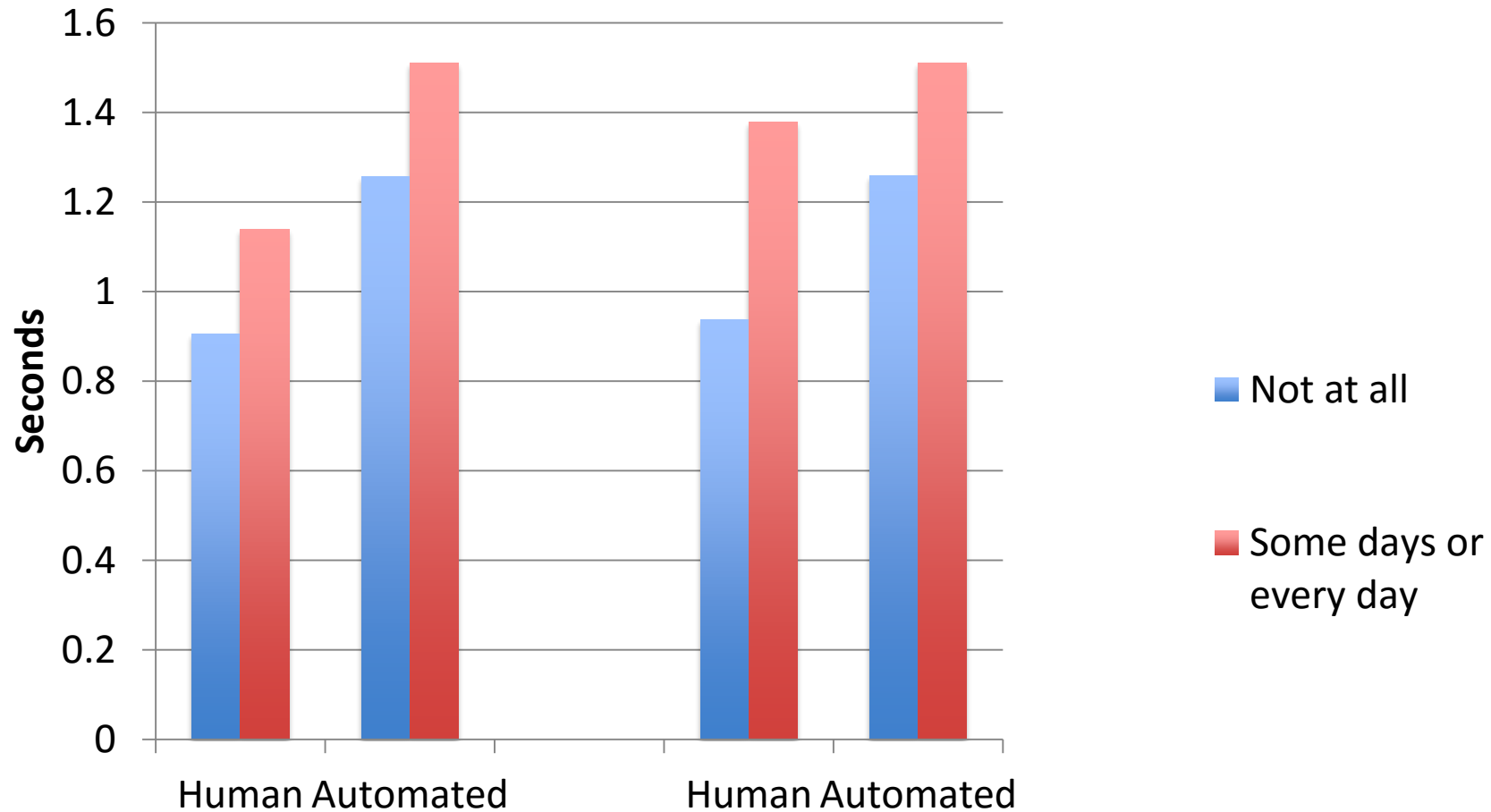
Findings: sensitive vs. nonsensitive *responses*

- Clear evidence that some responses and response options are given significantly more slowly than others
- But two distinct patterns across different questions—so *same* Rs are responsible for BOTH patterns
- Pattern 1: More hesitation for responses we see as more stigmatized

Q1: How often do you now smoke cigarettes:
'every day', 'some days' or 'not at all'?

Time to first sound

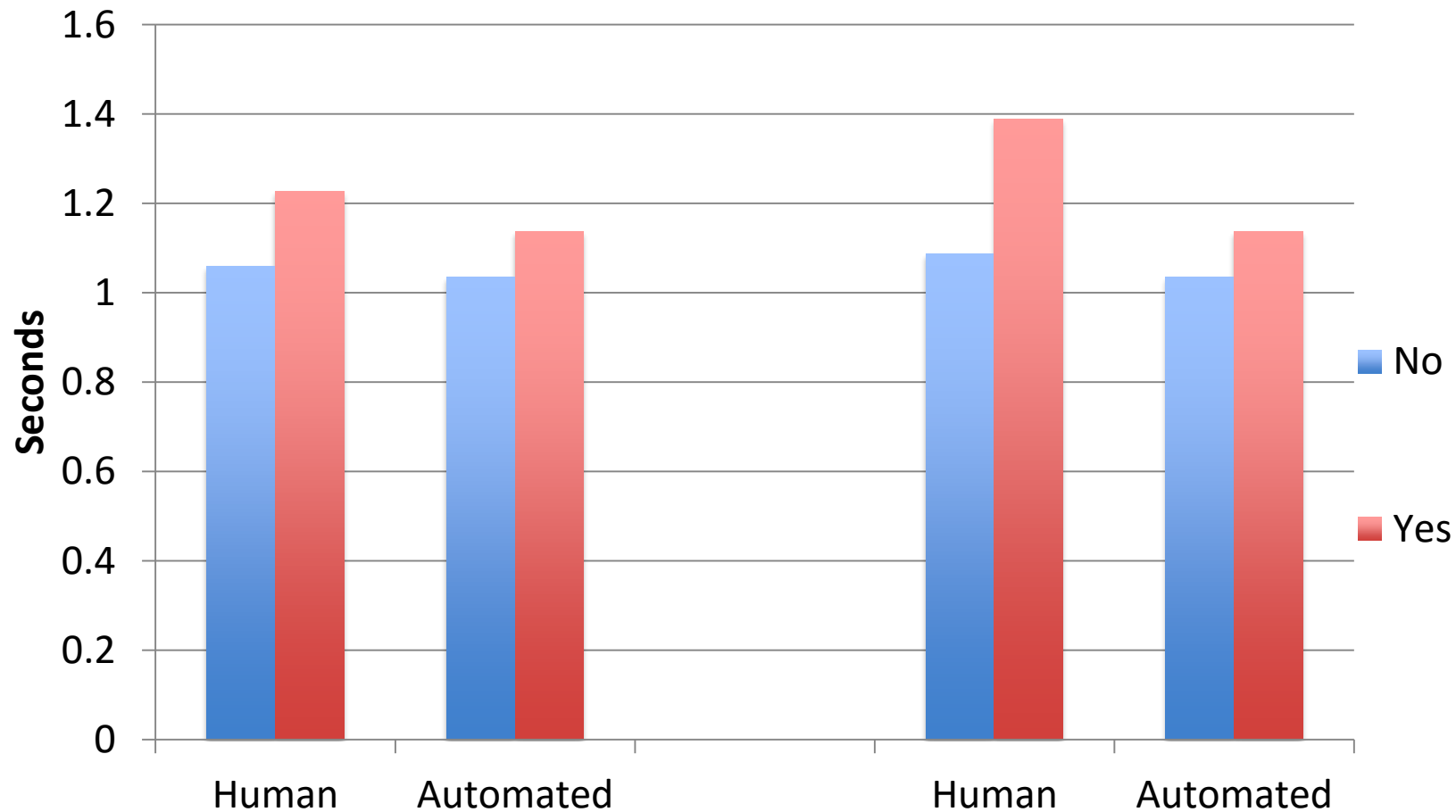
Time to first response



Q3: Have you ever, even once, used marijuana or hashish?

Time to first sound

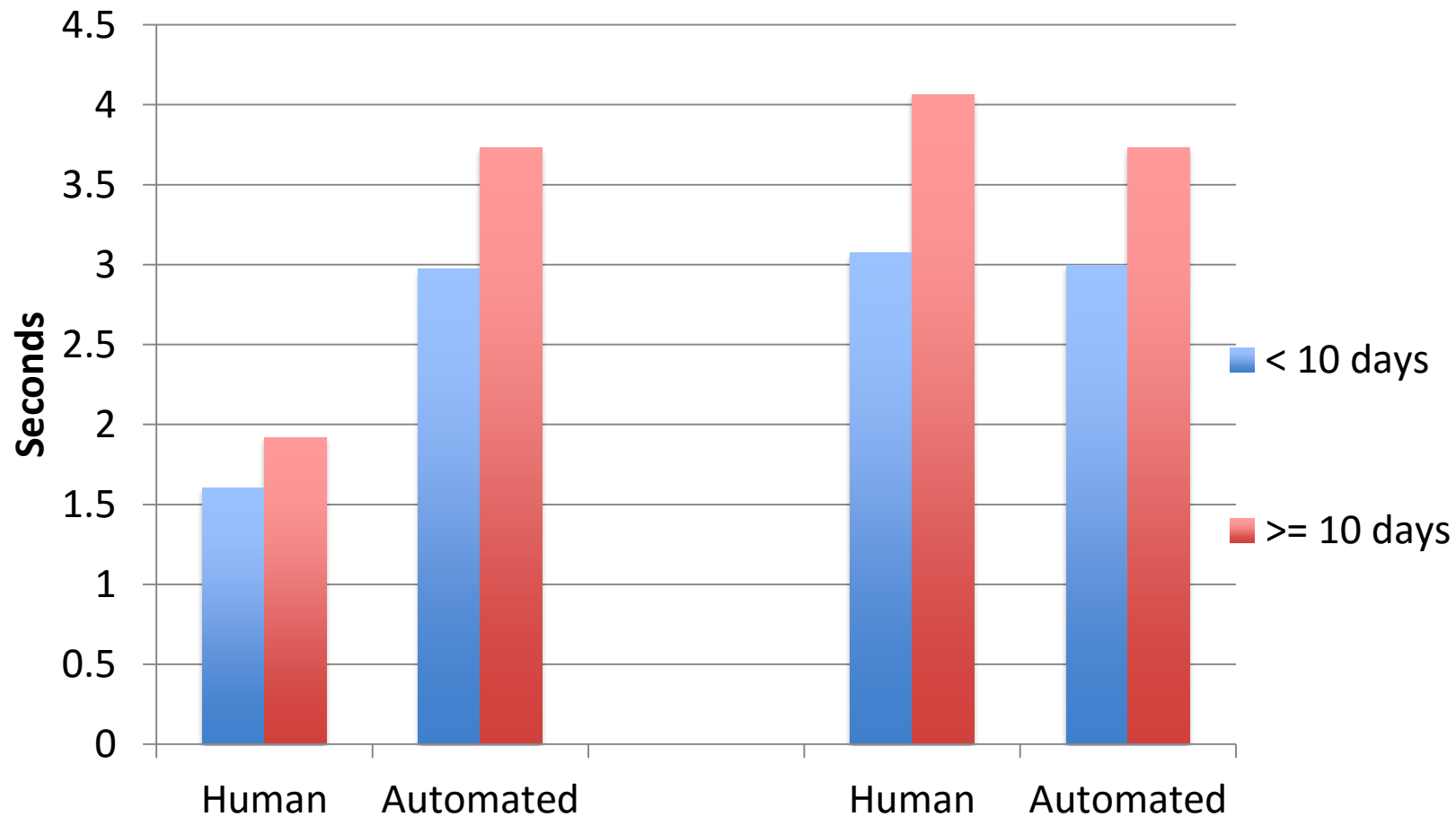
Time to first response



Q4: During the past 30 days, on how many days did you drink one or more drinks of an alcoholic beverage?

Time to first sound

Time to first response



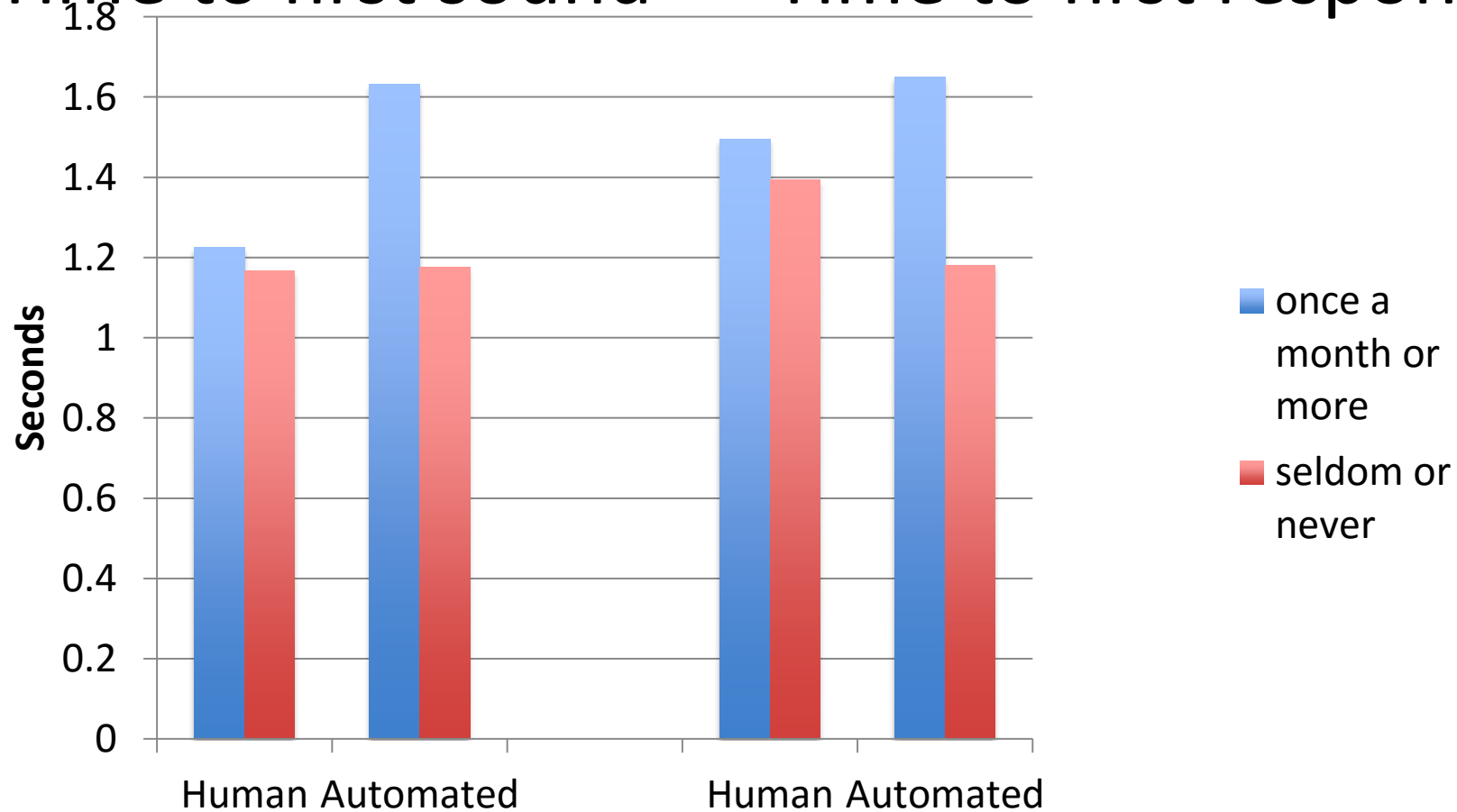
- Pattern 2: responses we deem stigmatized are *faster*
- (assuming we are correct in our judgments-- people's norms and sensitivities may well vary on these topics)
- (and perhaps extreme answers like “never” are easy to give without thinking hard)

Q21: How often do you attend religious services?

'At least once a week', 'almost every week', 'about once a month', '*seldom*', or '*never*'?

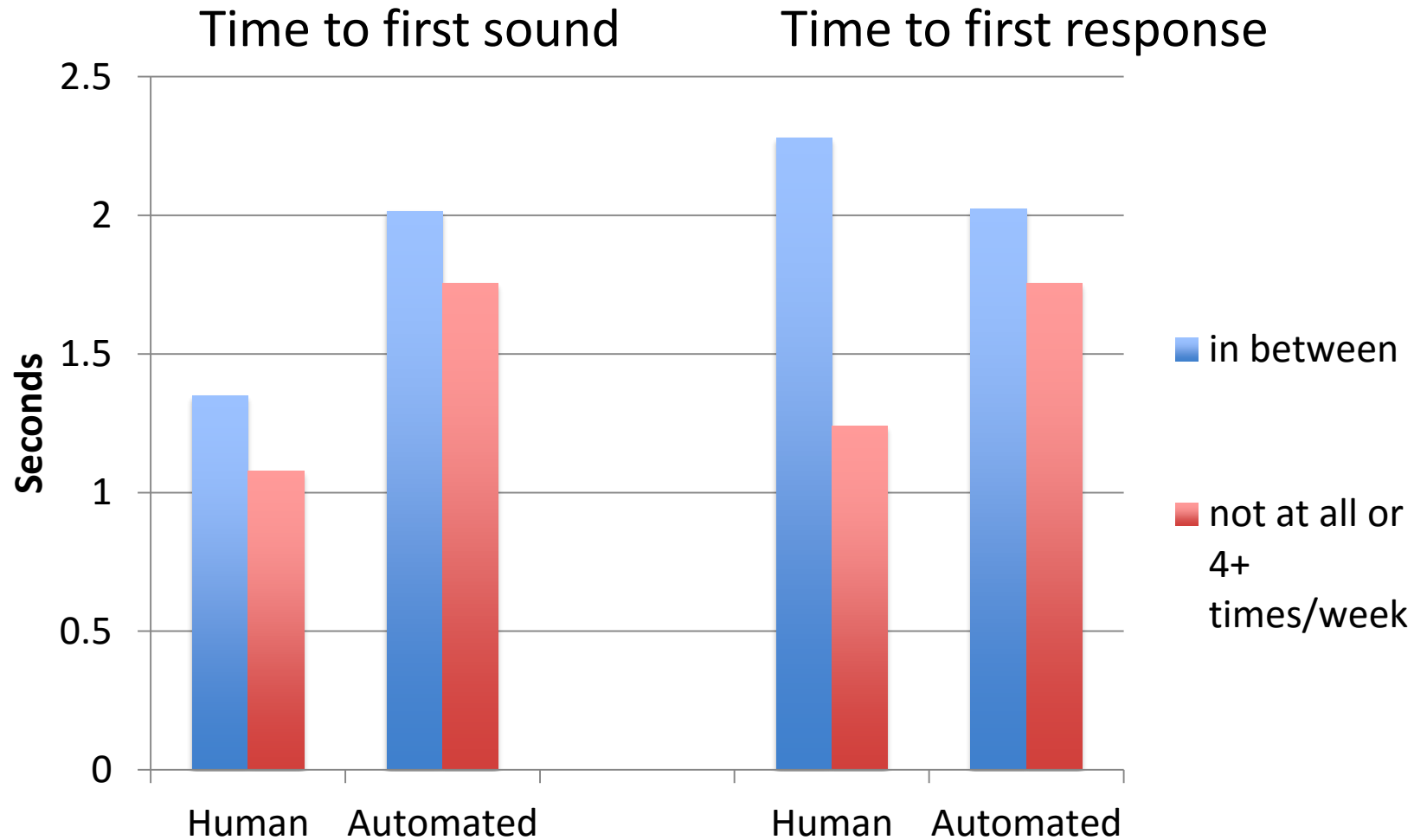
Time to first sound

Time to first response



Q9: About how often did you have sex during the last 12 months?

'Not at all', *'Once or twice'*, *'About once a month'*, *'two or three times a month'*,
'about once a week', *'two or three times a week'*, or *'four or more times a week'*.



Summary

- Clear evidence from matched (fair) comparisons that mobile survey respondents hesitate more when
 - answering *non-sensitive* (vs. sensitive) questions
 - answers to sensitive Qs are stigmatized (less socially desirable responses), for at least some Qs
 - interviewed by an automated system (vs. a human interviewer)
- Hesitation is revealing in a mobile environment

Why more quick with sensitive Q's?

- Many processes could be at play
- Could be being less thoughtful or conscientious because they are offended or embarrassed by the topic
- Perhaps they want to minimize time or effort on answering the question
- OR maybe answers more salient or easily available for this particular set of sensitive Q's, relative to these non-sensitive Q's

Why more quick with interviewer than automated system?

- Hypothesis: Speed might not reflect comfort but rather time pressure of talking to a person
 - → talking with a human is particularly time pressured?
 - (Jefferson: “standard maximum silence of 1 sec” in human conversation)
- Maybe people feel more comfortable keeping an automated system waiting
 - OR maybe people lack experience with speech dialog system, or mistrust that system will accurately recognize their speech
 - (recognition was actually 95.6% accurate)

Why slower with stigmatized responses?

- People may be considering whether to disclose fully
- May be editing for impression management
- May be thinking harder about what's true

Implications

- Unknown how these patterns will generalize to interviews with other questions, or in other modes (FTF, landline telephone, web, etc.)
- But evidence suggests that paralinguistic paradata are indeed significantly associated with sensitivity of both questions and answers in mobile surveys
- Differences in responding to human vs. automated interviewers observed help distinguish mechanisms underlying socially desirable responding
 - E.g., distinguishing between time pressure that results from talking at all vs. pressure that results from having a potentially judgmental human interlocutor

Why do mobile telephone interviews take longer?

A behavior coding perspective

Jerry Timbrook

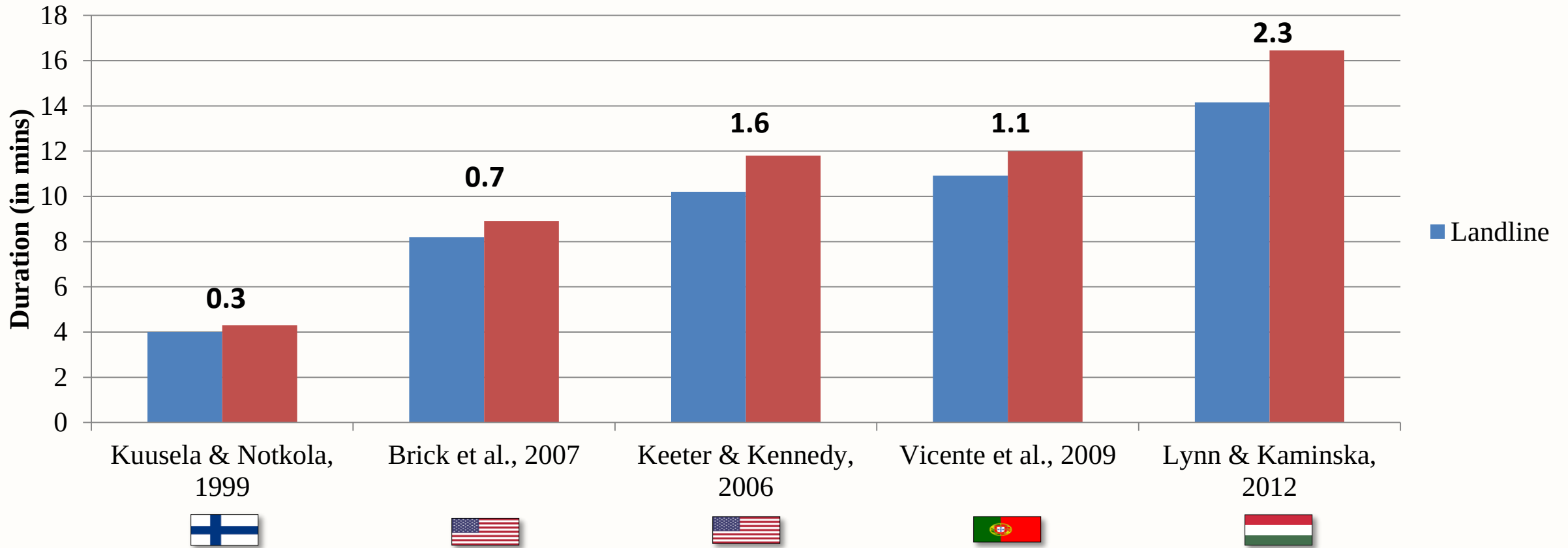
Kristen Olson

Jolene Smyth

University of Nebraska-Lincoln

AAPOR, May 2016

Mobile Telephone Interviews Take Longer Than Landline



Mobile Telephone Interviews Take Longer Than Landline

- Cost implications:
 - (500 mobile interviews)(1 min difference) = 500 extra interviewer minutes
 - Mobile interviews cost 1.5 times more than landline (Pew, 2015)

Why?

Common Explanations

1. Activities outside of the Q-A sequence

- Find a private space to respond (Lynn & Kaminska, 2012)
- Collection of info for mobile reimbursement (Keeter & Kennedy, 2006)
- Skip patterns (Kuusela & Noktola, 1999)

2. Respondent characteristics

- Socio-demographic differences across frames (Nathan, 2001)

Common Explanations

3. Disruption to Perception and Comprehension

- Respondent Perspective
 - Respondents must *hear* and *understand* a question to answer it correctly (Jenkins & Dillman, 1997; Tourangeau et al., 2000)
 - Disruptions may be more prevalent on mobile devices
 - Poorer line quality (Lavrakas et al., 2010)
 - More background noise (Schwarz et al., 1991; Lavrakas, et al., 2010)
 - Multitasking (Lavrakas et al., 2010; Kennedy & Everett, 2011; Lynn & Kaminska, 2012)

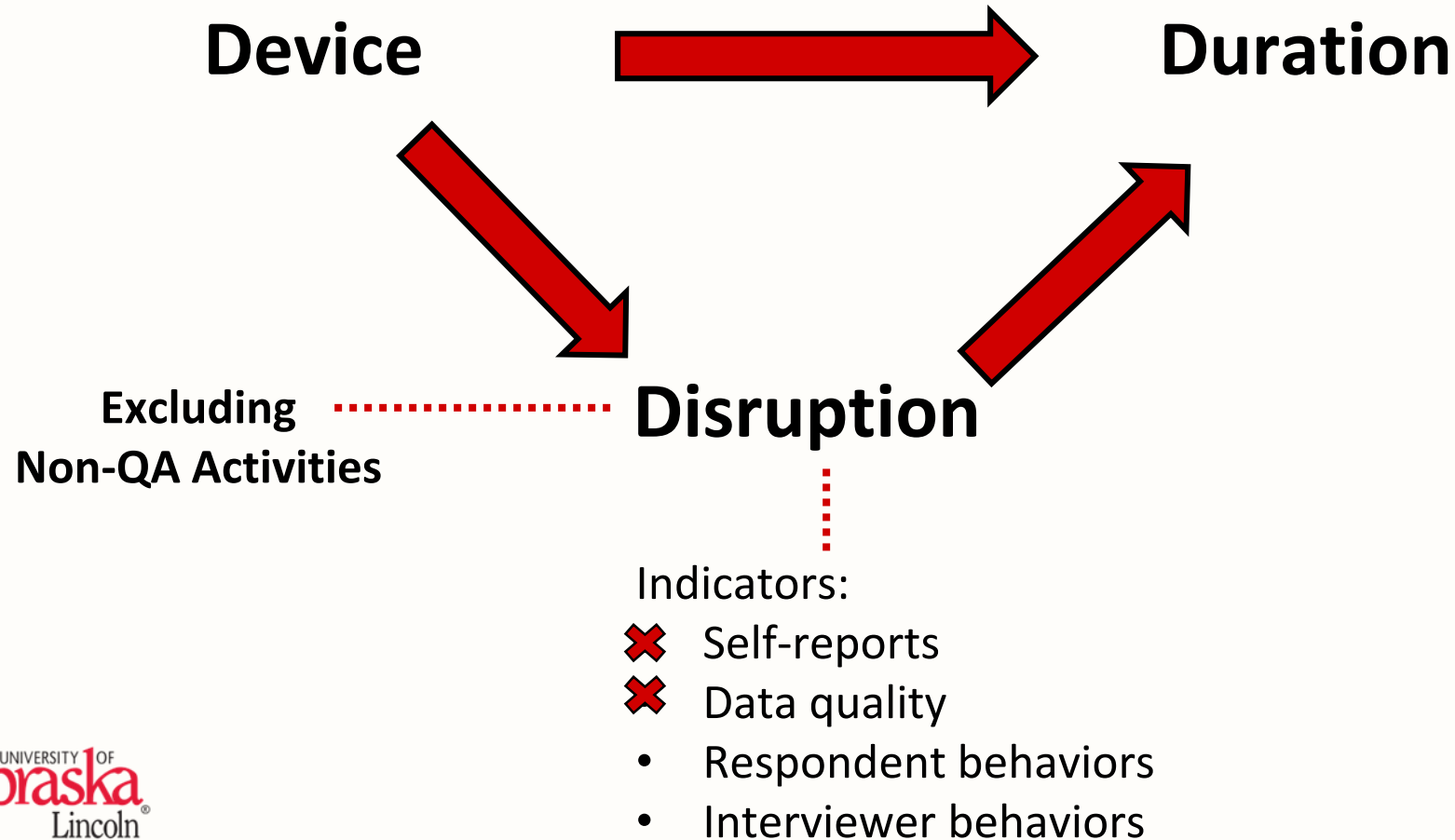
Common Explanations

3. Disruption to Perception and Comprehension

- Interviewer Perspective

- Interviewers must *hear* and *understand* respondents to record correct answers
- Interviewers may react to signs that a respondent is disrupted
 - *Communication accommodation theory* (Giles et al., 1991)
 - Read context cues in conversation, and change communication style to accommodate others
 - Speakers will raise/lower their speed of speech to match that of the listener (Street, 1982, 1983)

Common Explanations



Behavior Coding: A New Perspective

- **Objective, reliable method for coding interaction between interviewers and respondents** (Cannell & Fowler, 1996)
 - Turn by turn description of *actions* during in an interview
 - Provides timestamps

Behavior Coding: A New Perspective

Conversational turns	Behavior Codes			
	Actor	Initial action	Assessment	Disfluency?
I: Regarding North Korea, which issues should the U. S. and Japanese governments, working in cooperation, give priority to resolving? How about normalizing diplomatic relations between the U. S. and North Korea?	Interviewer	Question asking	Read exactly as worded	No

Behavior Coding: A New Perspective

Conversational turns	Behavior Codes			
	Actor	Initial action	Assessment	Disfluency?
I: Regarding North Korea, which issues should the U. S. and Japanese governments, working in cooperation, give priority to resolving? How about normalizing diplomatic relations between the U. S. and North Korea?	Interviewer	Question asking	Read exactly as worded	No
R: Uh, uh, what were the choices again?	Respondent	Clarification	Ask for repeat response options	Yes

Behavior Coding: A New Perspective

Conversational turns	Behavior Codes			
	Actor	Initial action	Assessment	Disfluency?
I: Regarding North Korea, which issues should the U. S. and Japanese governments, working in cooperation, give priority to resolving? How about normalizing diplomatic relations between the U. S. and North Korea?	Interviewer	Question asking	Read exactly as worded	No
R: Uh, uh, what were the choices again?	Respondent	Clarification	Ask for repeat response options	Yes
I: Uh, yes or no. Would it be a priority?	Interviewer	Probing	Repeat response options	Yes
R: Oh. Uh, no.	Respondent	Answer provided	Adequate	Yes

Research Questions

- **RQ1: Does the difference in duration persist across devices for *the same set of questions*?**
 - Excluding all time spent outside of the question-answer sequence and questions unique to a particular device
- **RQ2: Do *respondent characteristics* account for the difference in duration?**
- **RQ3: Do *respondent behaviors indicating disruption* contribute to the difference in duration?**
- **RQ4: Do *interviewer reactions to disruption* contribute to the difference in duration?**

Data

- **US/Japan Newspaper Opinion Poll**

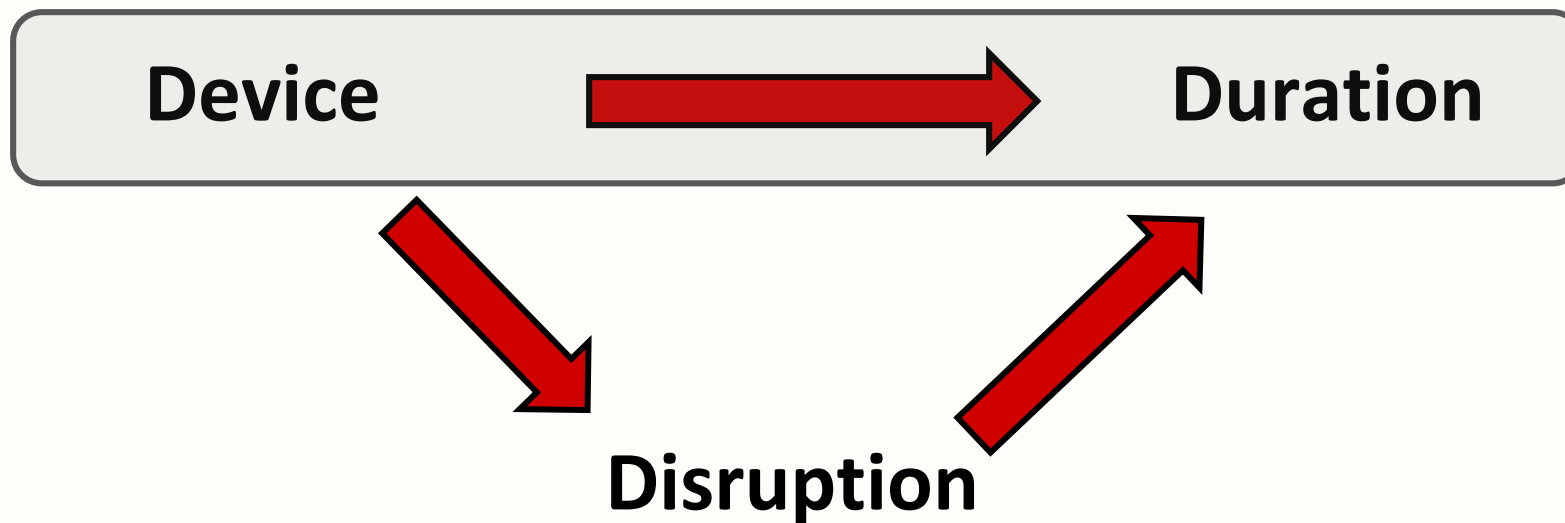
- National telephone survey of U.S. adults conducted by Gallup in November 2013.
- Landline and Mobile, AAPOR RR1 = 7.4%

- **From 1,005 initial interviews, a stratified random subset of 434 interviews were recorded, transcribed and behavior coded at the turn level**

- 249 landline respondents, 185 mobile respondents, 31 interviewers
- Kappas range from 0.38 to 0.98

Differences Across Devices

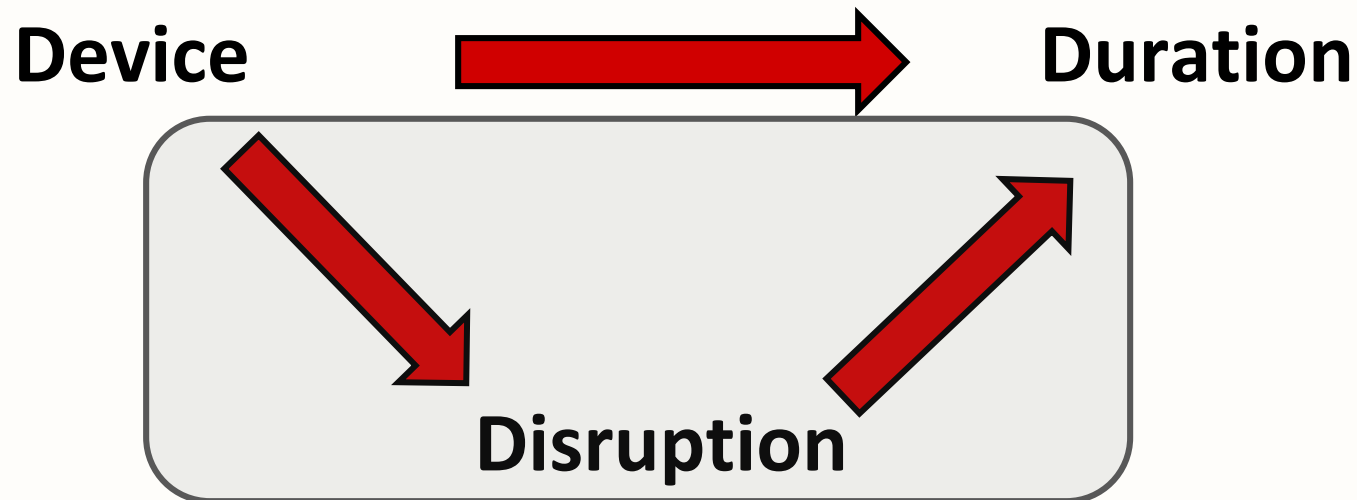
- Start by comparing duration across devices (t-test accounting for clustering within interviewers)



	<u>Landline</u>	<u>Mobile</u>
<u>Duration (in minutes)</u>	11.89	12.99

Differences Across Devices

- Compare disruption behaviors across devices
- t-test (continuous data), ranksum (count), and χ^2 (categorical)
 - Account for clustering within interviewers



Differences Across Devices

- **Significant *respondent* behaviors on *mobile devices***
 - ↑ disfluencies
 - ↑ comments about line quality
 - ↑ unintelligible audio
 - Slower speed of speech
- **Significant *interviewer* behaviors on *mobile devices***
 - ↑ comments about line quality
 - Slower speed of speech

Discussion

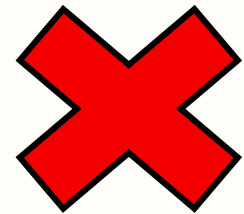
- **RQ1: Does the difference in duration persist for *the same set of response tasks*?**



- QA sequence is longer on mobile devices EXCLUDING:

- Recruitment
- Within-household selection
- Unique skip patterns
- Collection of info for mobile minute reimbursement
- Finding a quiet place to respond

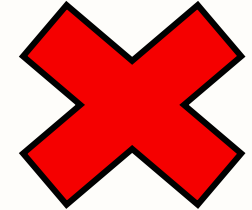
- **RQ2: Do *respondent characteristics* account for the difference in duration?**



- ↑ respondent age = ↑ duration
 - Consistent with Fricker et al. (2005) and Couper & Kreuter (2013)
 - Doesn't account for duration difference between devices
- No other respondent characteristics are predictive

Discussion

- RQ3: Do *respondent behaviors indicating disruption* contribute to the difference in duration?



- Satisficing Behaviors

- ↑ DK/Refuse answers = ↑ duration
 - Landline > Mobile

- Disruptions to Perception/Comprehension

- ↑ unacceptable answers = ↑ duration
 - Landline = Mobile
- Faster respondent speech = ↑ duration
 - Landline respondents talk faster than mobile respondents

Discussion

- **RQ4: Do *interviewer reactions to disruption* contribute to the difference in duration?**
 - Satisficing Behaviors
 - $\uparrow$ comments about duration = $\uparrow$ duration
 - Landline = Mobile
 - Disruptions to Perception/Comprehension
 - $\uparrow$ probes, clarifications, and motivational feedback = $\uparrow$ duration
 - Landline > Mobile
- Slower interviewer speech = $\uparrow$ duration
 - Interviewers in a mobile survey talk slower

Discussion

- RQ4: Do *interviewer reactions to disruption* contribute to the difference in duration?



- Disruptions to Perception/Comprehension

- Interviewers take up 77% of survey dialog (words in a survey)
- *Landline*: 124 wpm vs. *Mobile*: 108 wpm (16 wpm slower)
- Great! But why?
 - *Communication accommodation theory* (Giles et al., 1991) – read context and change communication style to accommodate others
 - Speakers raise/lower their speed of speech to match listener (Street, 1982, 1983)

Summary of Findings

- **Why DO mobile interviews last long?**
- **Interviewers read the context of a mobile survey...**
 - Significant *respondent* behaviors on *mobile devices*
 - Slower speed of speech
 - ↑ disfluencies
 - ↑ comments about line quality
 - ↑ unintelligible audio
- **...and adjust their speed of speech to accommodate.**

Thank you!

Contact:

Jerry Timbrook

jerry.timbrook@gmail.com

The Differential Effect of a Mobile-Friendly Instrument on Data Quality

Rachel Horwitz

Disclaimer: Any views expressed are those of the author and not necessarily those of the U.S. Census Bureau.

Background

- The rise in smartphone ownership and use in surveys is well documented (Pew Research Center; Horwitz, 2015; Baker-Prewitt, 2013)
- As are problems associated with their use in surveys (Baker-Prewitt, 2013; Mavletova, 2013; de Bruijne and Wijnant, 2013)
 - Long completion times (Mavletova, 2013; de Bruijne and Wijnant, 2014; McClain et al, 2012; Peterson, 2012)
 - Higher breakoffs (Baker-Prewitt, 2013; Callegaro, 2013; Mavletova, 2013, Wells et al., 2013)

Background

- Can we identify differential effects in the benefits of optimization by survey length?
 - Optimization helps but does not eliminate longer response times and higher breakoffs (Couper et al., 2015)
 - Meta-analysis includes surveys from different populations, of different lengths, and different types of “optimization”

Mobile-Friendly Census Surveys

2015 National Content Test (NCT)

2016 American Community Survey (ACS)

- First Census Bureau surveys using a mobile-friendly design
 - Use a responsive web design vs. mobile-first
 - Layout of information on the screen may change

Examples

■ Screens

AN OFFICIAL WEBSITE OF THE UNITED STATES GOVERNMENT

**United States™
Census
Bureau** 2015 Census Test

[Instructions](#) [FAQs](#) [Save and Log Out](#)

Where did you live on April 1, 2015? (Help)
Please select the type of address associated with your residence.

Address type: ☒ Street Address ☐ Rural Route ☐ P.O. Box

Address Number:
For example: (5007)

Street Name:
(N Maple Ave)

Apt/Unit:
("Apt. A" or "Lot 3")

City: State: ZIP Code:

[Next](#)

OMB No.: 0607-0981 Approval Expires: 3/31/2016 Accessibility Privacy Security

2015 Census Test

Where did you live on April 1, 2015? (Help)
Please select the type of address associated with your residence.

Address type:

☒ Street Address

☐ Rural Route

☐ P.O. Box

Address Number *(For example: 5007)*

Street Name *(For example: N Maple Ave)*

Apt/Unit *(For example: "Apt. A" or "Lot 3")*

City:

State:

Examples

United States™
Census
Bureau

American Community Survey

InstructionsFAQsSave and Log Out

➔

The following questions are about everyone who is living or staying at 198 Young Rd..

First, create a list of people. Enter one person on each line. Leave any extra lines blank. Enter names until you have listed everyone who lives or stays there, then click Next. [\(Help\)](#)

First Name

John

First Name 2

First Name 3

First Name 4

First Name 5

MI

G

MI 2

MI 3

MI 4

MI 5

Last Name

Smith

Last Name 2

Last Name 3

Last Name 4

Last Name 5

[Click here to add more people](#)

◀ Previous

Next ▶

Contact Us

AccessibilityPrivacySecurity

American Community Survey

☰

➔

The following questions are about everyone who is living or staying at 198 Young Rd..

First, create a list of people. Enter one person on each line. Leave any extra lines blank. Enter names until you have listed everyone who lives or stays there, then continue to the next page [\(Help\)](#)

John

G

Smith

First Name 2

MI 2

Last Name 2

Data

- Surveys

- National Census/Content Test (NCT)

- 2012 – Not optimized (25,103 respondents)
 - 2015 – Optimized (435,951 respondents)

- American Community Survey

- January 2015 – Not optimized (65,846 respondents)
 - January 2016 – Optimized (69,190 respondents)

Survey length

- **NCT ~ 10 minutes**
- **ACS ~ 40 minutes**

- Analysis

- Limited to computer and phone respondents (smartphone and feature phone)

Analysis

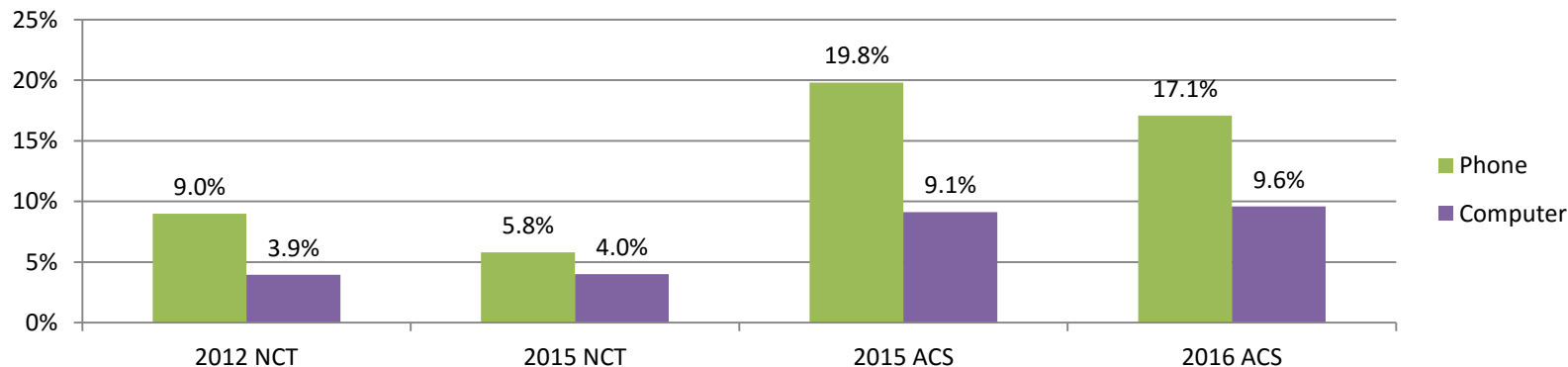
ANALYSIS MEASURES (BY DEVICE)

Logins	% people who successfully logged into survey
Breakoff Rate	% people who logged in but did not submit the survey
Time To Complete	difference between survey submit time & login time
Answer Changes	average # of times a respondent changed an answer
Device Switching	% respondents that started on a phone then switched to a computer

Results – Breakoff Rate

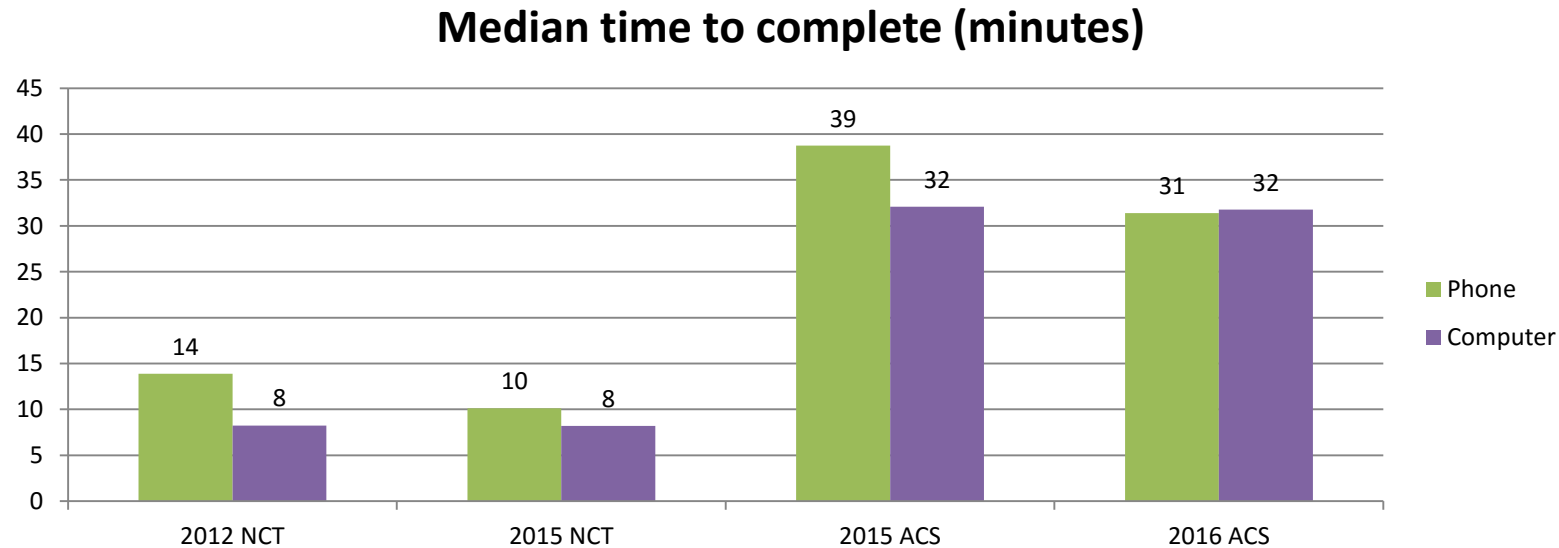
- NCT: 2.3 times higher before optimization, 1.4 times after → differential of 0.9
- ACS: 2.2 times higher before, 1.8 times after → differential of 0.4

Breakoff Rate - Before and after optimization by survey and device



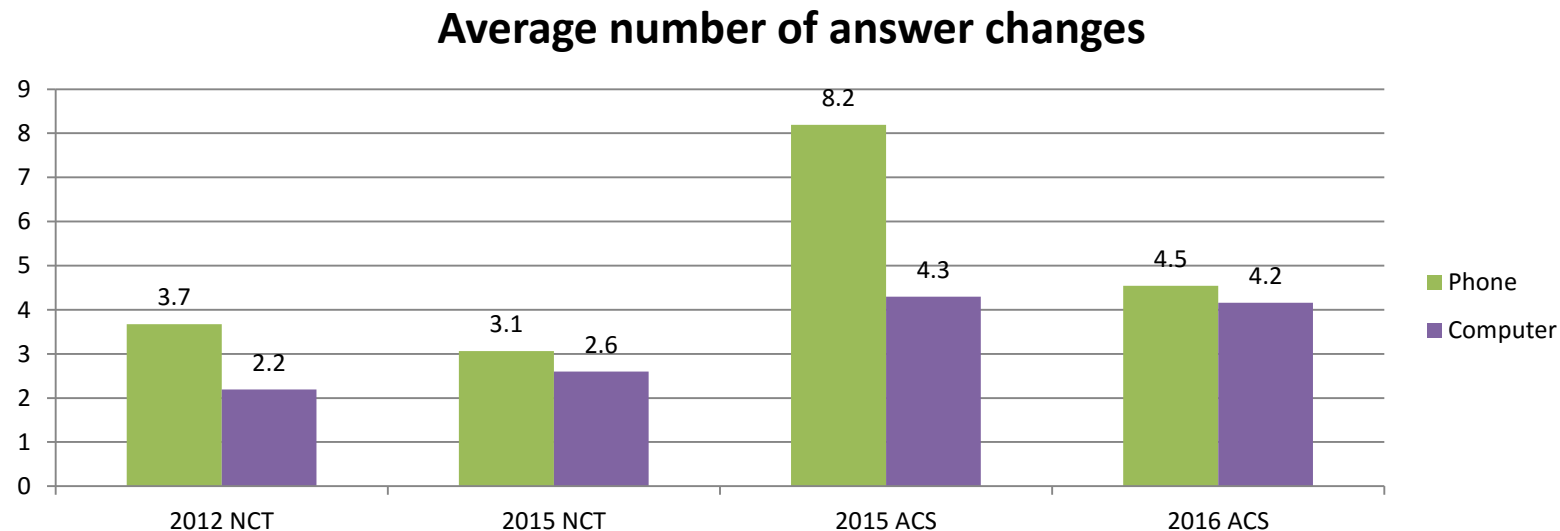
Results – Completion Time

- NCT: 1.8 times longer before, 1.3 times longer after → differential of 0.5
- ACS: 1.2 times longer before, 1.0 times after → differential of 0.2



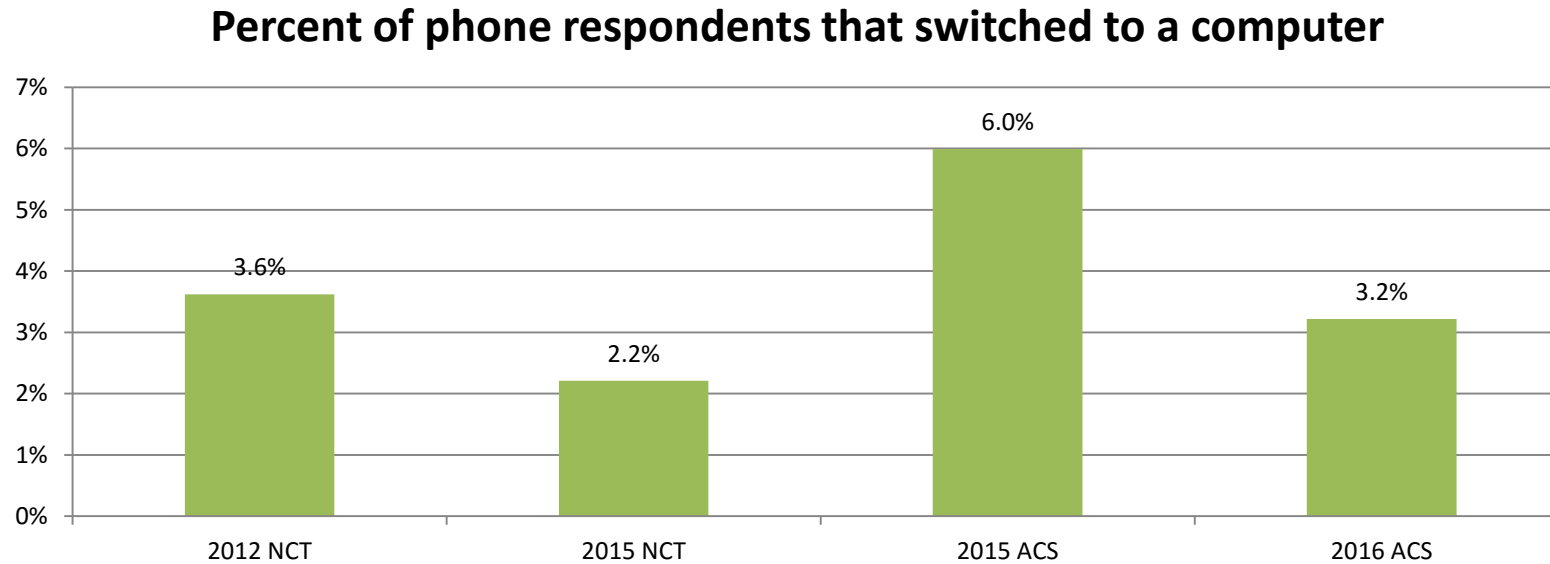
Results – Answer Changes

- NCT: 1.7 times more changes before, 1.2 times more changes after → differential of 0.5
- ACS: 1.9 times more changes before, 1.1 times more changes after → differential of 0.8



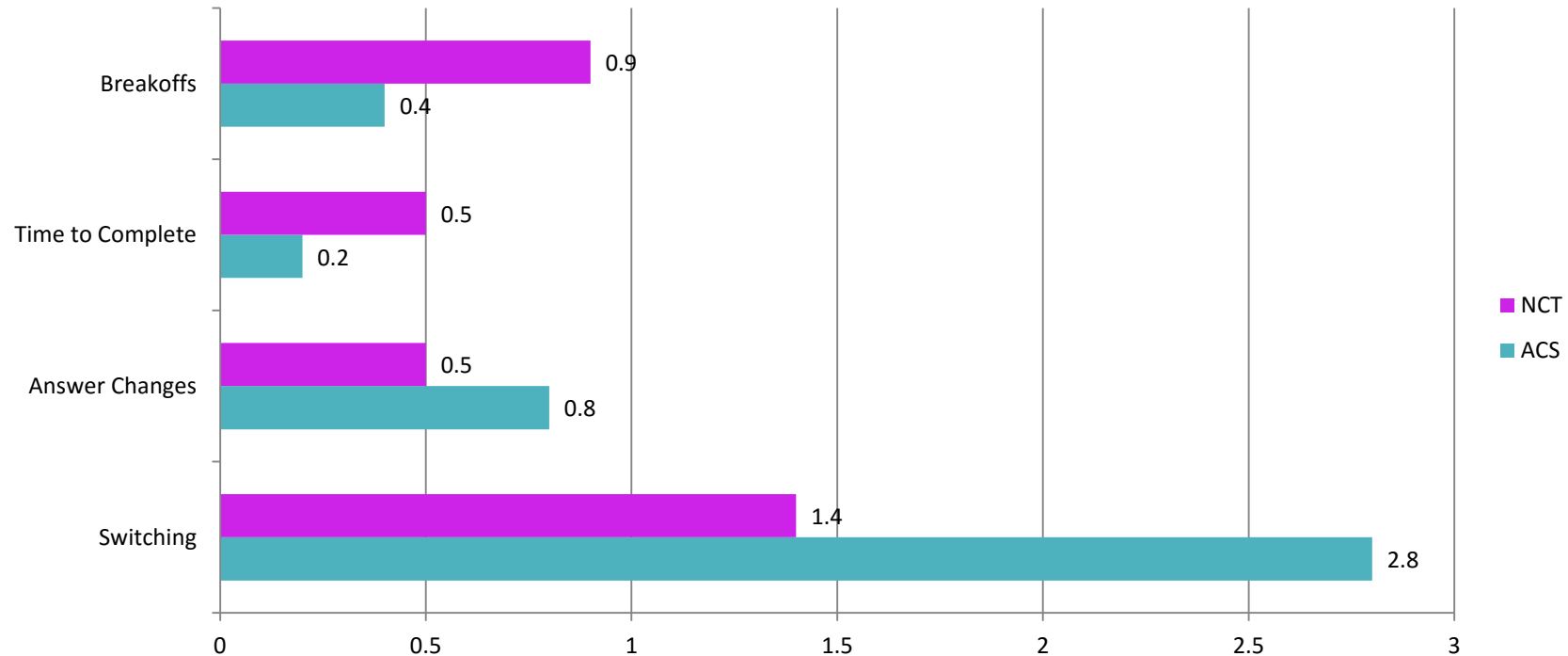
Results – Switching to Computer

- NCT: 1.4 percent decrease after optimization
- ACS: 2.8 percent decrease after optimization



Results – Are there differential gains by survey length?

Comparison of the difference in the phone to computer ratio before and after optimization



What does all this mean?

- Optimization yields an improvement across all measures for both surveys
- There is a difference in the effect of optimization between long and short surveys, but the direction depends on the measure
- Length is likely only one factor, future research...
 - Question Type
 - Connection Speed
 - Phone Quality
 - Population
- Still see gains, even for short surveys

Thank you!

Contact:

Rachel.t.horwitz@census.gov

EXPERIMENTS WITH MOBILE DATA COLLECTION



**Vinay
Bapna**

Unomer

Embedding Survey Questions Within Non-research Mobile Apps:
A method for collecting high-quality data from smartphone users?



**Michael
Schober**

New School for
Social Research

Hesitation In Socially Desirable Responses In A Mobile Phone Survey



**Jerry
Timbrook**

U. of Nebraska-
Lincoln

Why Do Mobile Interviews Take Longer? A Behavior Coding Perspective



**Rachel
Horwitz**

U.S. Census
Bureau

The Differential Effect of Mobile-Friendly Surveys on Data Quality