Comparing new survey modes: Text and voice interviews on smartphones

Michael F. Schober

Acknowledgments

• NSF grant SES-1025645 (Methodology, Measurement, and Statistics program)
• Mick Couper, Lloyd Hemingway, Heather Schroeder, Amanda Schuetz
• Aerialink
Mobile multimodal phones (smartphones)

• People increasingly communicate via smartphones*
  – > 50% of US adults now own one
  – Rapid growth

• People increasingly expect multiple modes (many native to smartphone) for interacting
  – Voice
  – Text (SMS)
  – Email
  – Videochat
  – Web access for communication (e.g., blog posts)
  – Specialized apps for communication (e.g., Facebook)
People increasingly expect capability to:

• communicate while mobile and/or multitasking
• choose a mode that fits their current setting and needs
  – e.g., urgent vs. can wait, public vs. private, noisy vs. quiet, bright vs. dim
• switch modes while communicating
• respond in a different mode than contacted
  – e.g., respond to voicemail with a text
New possibilities and challenges for survey data collection

• As R’s expect multiple modes on the same device, may expect that surveys are multimodal
  – Potential to interact via SMS text when convenient
  – Potential to respond in mode that is appropriate to current setting (e.g., text in noisy environment, voice when there is glare, etc.)
  – More generally, to be able to respond in any mode, anytime, anywhere

• Do people respond to conventional survey modes (e.g., telephone interviews) in the same way on smartphones as on landlines?

• How do people respond to less conventional survey modes that use smartphone capabilities?
New options for survey mode choice

- Now possible to choose a mode on a single device, immediately and conveniently
- Quite different from prior implementations of survey mode choice
  - When R invited by mail completes either on paper or web, this requires extra step of typing URL into browser
  - Can reduce response rates (e.g., Fulton & Medway, 2012)
- Choice on single device may lead to different outcomes
Current study

• examines
  – data quality (satisficing, disclosure)
  – completion rates
  – respondent satisfaction

• four existing or plausible survey modes that work through native apps on the iPhone
  – As opposed to specially designed survey apps
  – As opposed to web survey in phone’s browser
  – Uniform interface for all Rs
    • As opposed to mix of platforms (Android, Windows, etc.)
# Experiment: 4 modes on iPhone

<table>
<thead>
<tr>
<th>Interviewing Agent</th>
<th>Medium</th>
<th>Voice</th>
<th>SMS Text</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Human</td>
<td>Human voice (R speaks with I)</td>
<td>Human text (R texts with I)</td>
</tr>
<tr>
<td>Automated</td>
<td>Speech IVR</td>
<td>Speech IVR (R speaks with system)</td>
<td>Automated Text (R texts with system)</td>
</tr>
</tbody>
</table>
Surveys via text messaging?

• More and more people are embracing text messaging for personal and professional communication
  – on their mobile phones (smartphones or not)
  – on other devices (e.g., tablets, desktops)
• Texting is becoming a potentially important way to reach respondents
  – some may attend to text more than to emails or voicemails
  – respondents may expect to be able to participate in a survey via text
• Some organizations are now including SMS text in their suite of modes for mobile surveys
  – e.g., GeoPoll, Poll Everywhere, iVisionMobile, etc.
Text as a mode of interaction

• Turn-by-turn
  – Threaded (on a smartphone)
• Responses don’t need to be immediate
  – Allows multitasking
• Works even with intermittent network/cell service
  – Unlike voice
• Does not require web capacity on device
  – Unlike mobile web survey
<table>
<thead>
<tr>
<th>Property</th>
<th>Voice</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>Synchrony</td>
<td>Fully synchronous</td>
<td>Less or asynchronous</td>
</tr>
<tr>
<td>Medium</td>
<td>Auditory</td>
<td>Visual</td>
</tr>
<tr>
<td>Language</td>
<td>Spoken/heard</td>
<td>Written/read</td>
</tr>
<tr>
<td>Conversational structure</td>
<td>Turn-by-turn, with potential for simultaneous speech</td>
<td>Turn-by-turn, rarely but possibly out-of-sequence</td>
</tr>
<tr>
<td>Persistence of turn</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Persistence of entire conversation</td>
<td>No</td>
<td>Yes, threaded</td>
</tr>
<tr>
<td>Social presence of partner</td>
<td>Continuous (auditory) presence</td>
<td>Intermittent evidence (when texts arrive)</td>
</tr>
<tr>
<td>Character of multitasking</td>
<td>Simultaneous, especially when hands free, unless other task involves talking</td>
<td>Switching required between texting and other tasks</td>
</tr>
<tr>
<td>Impact of environmental conditions</td>
<td>Potential interference from ambient noise</td>
<td>Potential interference from visual glare</td>
</tr>
<tr>
<td>Impact of nearby others</td>
<td>Others may hear answers; potential audio interference from others’ talk</td>
<td>Others unlikely to see text and answers on screen, though possible</td>
</tr>
</tbody>
</table>
Measures of data quality

- **Conscientious responding (less satisficing)**
  - Rs are known to take shortcuts—to “satisfice”—differently in different survey modes
    - e.g., Chang & Krosnick (2009), Heerwegh & Loosveldt (2008)
  - We examine
    - rounded numerical responses (e.g., multiples of 10)
    - straightlining (nondifferentiation)
      - giving same answer to battery of Qs

- **Disclosure (more is better)**
  - Rs often disclose more sensitive information when they self-administer a questionnaire (web surveys, ACASI)
    - e.g., Kreuter et al. (2009), Tourangeau & Smith (1996)

- **Interview completion**
Possible outcomes: Conscientious responding

• TEXT VS. VOICE
  – Rs might be less conscientious in text because they import “least effort strategy” from how they usually text
  – OR Rs might more conscientious in text because they feel less time pressure to respond than in spoken interviews
    – and can answer when they are ready

• HUMAN VS. AUTOMATED
  – Rs might be less conscientious with automated interviewer (self administration) because there is no human to motivate them to be conscientious
Possible outcomes: Disclosure

• **TEXT VS. VOICE**
  – Rs might disclose **more** in text because of fewer social cues in the interaction
    • less evidence of reaction to answers?
    • more time to be comfortable with answers?
    • no one else can hear the questions or answers?
  – OR Rs might disclose **less** in text because
    • they worry that others might see visually persistent answers?
    • they worry that answers are permanently stored?
    • they can take time to answer in ways that give the best impression?

• **HUMAN VS. AUTOMATED**
  – Rs might disclose more with automated interviewer, as in ACASI or web survey
Items

• First, safe-to-talk question
• 32 Qs taken from major US social surveys and methodological studies
  – E.g., BRFSS, NSDUH, GSS, Pew Internet & American Life Project
  – For most, known to have produced differences in satisficing or disclosure between conventional modes
• Yes/no, numerical, categorical, battery items (series of Qs with same response options)
• Rationale for inclusion
  – Qs with more and less socially desirable answers
    • e.g., sexual history, drug use, newspaper reading
  – Qs for which frequency reports could be precise or estimated (rounded)
    • e.g., number of movies seen last month, number of apps on iPhone
  – Battery Q’s that could produce straightlining (non-differentiation)
Implementation: Human voice

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

- Same 10 I's from U Mich survey research center
- Same custom designed CATI interface
  - I selects, edits, or types questions/prompts, and clicks to send
- Text messages sent through third party (Aerialink)
- Rs can answer with single character: Y/N, letter (a/b/c), or number
Implementation: Speech IVR

• Custom built speech dialogue system
• Uses ATT’s Watson speech recognizer, Asterisk telephony gateway
• Recorded human interviewer, speech responses (not touchtone)
Implementation: Auto-text

- Custom built text dialogue system
- Text messages sent through third party (Aerialink)
- Rs can answer with single character: Y/N, letter (a/b/c), or number
Respondents: 634 iPhone users

- \( n = 157 \) to \( 160 \) randomly assigned to each mode
- Recruited from Craigslist, Facebook, Google Ads, and Amazon Mechanical Turk
  - Web screener verified age (> 21 years) and US area code
  - iPhone usage verified via text message to device and user agent string in response
- $20 iTunes gift code incentive, provided after post-interview web questionnaire
- Age, gender, ethnicity, income, education not reliably different in four modes
- Somewhat younger and less affluent than US national iPhone users
Data collection

• Interviews carried out March-May 2012

• Results based on speech-IVR system recognition
  – 95.6% correct recognition accuracy based on transcripts
  – Same pattern of results if we use human annotations
Conscientious responding:
Rs rounded on fewer Qs in **text** than **voice**

![Bar chart showing the number of questions with rounded answer (divisible by 10) per respondent for Human and Automated responses.](image)

Number of questions with rounded answer (divisible by 10) per respondent.
Rounding: “Movie watching last month”* responses that end in zero

*During the last month, how many movies did you watch in any medium?
Rounding: “Number of songs on your iPhone”* responses that end in zero

*How many songs do you currently have on your iPhone?
Conscientious responding: Straightlining

• Q: support for various dietary practices (eating red meat, limiting fast food, etc.)
  » strongly favor
  » somewhat favor
  » neither favor nor oppose
  » somewhat oppose
  » strongly oppose

• We define answers in battery as “straightlining” when at least 6 of 7 responses are the same

• Significantly less straightlining in text than voice
Conscientious responding: Percent of Rs straightlining

- Human
  - Voice: 9%
  - Text: 4%

- Automated
  - Voice: 9%
  - Text: 5%
Disclosure: Rs produced the most socially undesirable answer for more Qs in **text** than **voice**, and in **automated** than **human**

Number of Qs with most socially undesirable answer (categorical response or numerical answer above cutoff) per R
Disclosure: “Sex partners in last 12 months”, mean

*How many sex partners have you had in the last 12 months?
Disclosure: “Exercising less than 1 day per week”, percent

*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?
Disclosure: “Binge drinking”*, percent any days in last 30

*During the past 30 days, on how many days did you have 5 or more drinks on the same occasion?"
Disclosure: “Attend religious services”*, percent never

*How often do you attend religious services? At least once a week, almost every week, about once a month, seldom, or never?"
Disclosure: “Sex during last 12 months”, percent 4 or more times a week

*About how often did you have sex during the last 12 months?"
What accounts for text vs. voice differences in precision and disclosure?

• Could be any or all of the many differences in timing and behavior between text and voice interviews
  – alone or in combination

• Plausible contributing factors include:
  – Text reduces immediate time pressure to respond, so R has more time to think or look up answers
    → Could explain greater precision (less rounding) in text
  – Text reduces “social presence”
    • Reduced salience of I’s ability to evaluate or be judgmental?
    • No immediate evidence of I’s reaction?
    → Could explain more disclosure in text
Experimental design helps rule in or rule out accounts

• e.g., maybe R’s round less in text because text I’s never laugh (no LOL’s or haha’s)
  – Maybe laughter in voice interviews suggests that casual responses are sufficient
  – But that can’t be it because R’s round just as much in Human and Auto Voice interviews, and automated “interviewer” never laughed
## Examples: Text vs. voice interactions

<table>
<thead>
<tr>
<th>HUMAN TEXT</th>
<th>HUMAN VOICE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>I: During the last month how many movies did you watch in any medium?</td>
</tr>
<tr>
<td>2</td>
<td>R: 3</td>
</tr>
<tr>
<td>3</td>
<td>I: 30.</td>
</tr>
</tbody>
</table>

**Total elapsed time until next Q:**

1:21

0:12
### Examples: Text vs. voice interactions

<table>
<thead>
<tr>
<th></th>
<th>HUMAN TEXT</th>
<th></th>
<th>HUMAN VOICE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>I: During the last month how many movies did you watch in any medium?</td>
<td>1</td>
<td>I: <em>During the last</em></td>
</tr>
<tr>
<td>2</td>
<td>R: Medium?</td>
<td>2</td>
<td>R: Huh?</td>
</tr>
<tr>
<td>3</td>
<td>I: Here’s more information. Please count movies you watched in theaters or any device including computers, tablets such as an iPad, smart phones such as an iPhone, handhelds such as iPods, as well as on TV through broadcast, cable, DVD, or pay-per-view.</td>
<td>3</td>
<td>I: Oh, sorry. Um, during the last month, how many movies did you watch in ANY medium.</td>
</tr>
<tr>
<td>4</td>
<td>R: 3</td>
<td>4</td>
<td>R: Oh! Let’s see, what did I watch. Um, should I say how many movies I watched or how many movies watched me? [laughs] All right let’s-let me think about that. I think yesterday I watched u:m, not in its entirety but you know, coming and going. My kids are watching in. Um, I don’t know maybe 2 or 3 times a week maybe?</td>
</tr>
</tbody>
</table>

**Total elapsed time until next Q:** 2:00
Examples: Text vs. voice interactions

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>HUMAN VOICE</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>I:</td>
<td>Uh, so what would be your best estimate on how many, um, you saw in the whole month.</td>
</tr>
<tr>
<td>6</td>
<td>R:</td>
<td>[pause] Um, I don’t know I’d say maybe 3 movies if that many.</td>
</tr>
<tr>
<td>7</td>
<td>I:</td>
<td>3?</td>
</tr>
<tr>
<td>8</td>
<td>R:</td>
<td>Is that going to the movies or watching the movies on tv. Like you said <em>any medium</em> right?</td>
</tr>
<tr>
<td>9</td>
<td>I:</td>
<td>That’s <em>any movies.</em> Yep.</td>
</tr>
<tr>
<td>10</td>
<td>R:</td>
<td>Maybe 1 or 2 a month I’d say.</td>
</tr>
<tr>
<td>11</td>
<td>I:</td>
<td>1 or 2 a month? [breath] Uh, so what would be <em>closer</em></td>
</tr>
</tbody>
</table>
Examples: Text vs. voice interactions

| Total elapsed time until next Q: | 1:36 |

<table>
<thead>
<tr>
<th>HUMAN VOICE</th>
</tr>
</thead>
<tbody>
<tr>
<td>12 R: <em>Yeah, because</em> I uh, um, occasionally I take the kids on a Tuesday to see a movie, depending on what’s playing. So I’d maybe once or twice a month</td>
</tr>
<tr>
<td>13 I: Which would be closer, once or twice.</td>
</tr>
<tr>
<td>14 R: I would say twice.</td>
</tr>
<tr>
<td>15 I: Twice?</td>
</tr>
<tr>
<td>16 R: R: Mhm. Because it runs 4 Tuesdays which is cheaper to go</td>
</tr>
<tr>
<td>17 I: Right</td>
</tr>
<tr>
<td>18 R: R: so I’d say twice, yah. Because I do take them twice. Not last month but the month before</td>
</tr>
</tbody>
</table>
Need a profile of text vs. voice interviews

- Timing of back-and-forth
- Behavior
  - Cross-modal coding scheme developed for I and R interview “moves” and interactional paradata
    - 25 interviewer moves
      - e.g., ask Q as worded, present response alternatives, no-input (“I didn’t hear that”), no-match (“I didn’t understand that”)
    - 30 respondent moves
      - e.g., answer Q not using exact response alternatives, report behavior instead of answering, ask for clarification
- Additional behaviors
  - e.g., speech disfluencies and typos, laughter, hedges
- High interrater reliability among 3 coders (Cohen’s kappas = .91-.99) on subset of 400 Q-A sequences from 619 interviews
Interview duration (median)

<table>
<thead>
<tr>
<th>Minutes</th>
<th>Text (Human)</th>
<th>Text (Automated)</th>
<th>Voice (Human)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Number of turns per interview (32 Qs)
Median time between turns

- **Text**: 25 seconds
- **Voice**: 5 seconds

**Legend**
- Red: Human
- Teal: Automated
Median time between text turns

--R’s take longer than I’s

[Bar chart showing median time between text turns for Interviewer and Respondent, comparing Human Text and Automated Text.]
Multitasking (self-reported)

**Doing something else on iPhone**

<table>
<thead>
<tr>
<th>Percent of Rs</th>
<th>Text</th>
<th>Voice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>25</td>
<td>5</td>
</tr>
<tr>
<td>Automated</td>
<td>25</td>
<td>5</td>
</tr>
</tbody>
</table>

**Total number of other tasks being carried out during interview**

<table>
<thead>
<tr>
<th>Text</th>
<th>Voice</th>
</tr>
</thead>
<tbody>
<tr>
<td>e.g., working, eating, errands, traveling</td>
<td>e.g., talking, web browsing, texting with others</td>
</tr>
</tbody>
</table>
Text (vs. voice): simpler interaction
(more “paradigmatic”* sequences)

**Respondent**
- Fewer variable and unacceptable answers
- Less reporting of behavior
- Fewer backchannels (“uh-huh”)
- Almost no requests for repeat of survey Q
- Fewer “Don’t Know” answers
- Fewer requests for time to find answer
- Less commentary
- Fewer hedges
- No speech disfluencies, few typos

**Interviewer**
- No misstatements of Q
- Almost no repeats of Q or response alternatives
- Fewer neutral probes
- Almost no laughter (LOL)
- No speech disfluencies (fillers, repairs), few typos
- Less commentary

* Schaeffer & Maynard (1996)
Behaviors and data quality?

• Little evidence (thus far in analyses) for links between R or I behaviors—including those plausibly linked with rapport—and rounding or disclosure

• Clear evidence for link between time pressure and precision in text interviews
Interturn interval in text and rounding

Number of rounded answers

<table>
<thead>
<tr>
<th>Shorter</th>
<th>Longer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>Automated</td>
</tr>
</tbody>
</table>

Effect of interturn interval: F (1,309)=11.79, p<.001
Can nonresponse or noncompletion explain the patterns?

• Were people who were more conscientious (less likely to round their answers) or more willing to disclose sensitive information actually less likely to start or finish interviews in voice modes than in text modes?

• Could our mode effects result not from the contribution of responders and completers, but instead from the non-contribution of non-responders and non-completers?
Study design allows looking at this in a focused way

• all our sample members had already indicated, by screening into the study, interest in and at least some commitment to participating in an interview on their iPhone (in an unspecified interview mode).

• The fact that our participants were randomly assigned to an interviewing mode means that their initiative was unlikely to have differed across the modes.
## Participation, Start Rate, Completion Rate and Response Rate

<table>
<thead>
<tr>
<th>Mode</th>
<th>Invitations</th>
<th>started interview</th>
<th>Start Rate</th>
<th>Completed Interview</th>
<th>Completed Online Debriefing</th>
<th>Completion Rate</th>
<th>AAPOR RR 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human voice</td>
<td>316</td>
<td>169</td>
<td>53.50%</td>
<td>164</td>
<td>160</td>
<td>97.04%</td>
<td>51.90%</td>
</tr>
<tr>
<td>Auto Voice</td>
<td>414</td>
<td>187</td>
<td>45.20%</td>
<td>162</td>
<td>159</td>
<td>86.63%</td>
<td>39.13%</td>
</tr>
<tr>
<td>Human Text</td>
<td>227</td>
<td>176</td>
<td>77.50%</td>
<td>163</td>
<td>158</td>
<td>92.61%</td>
<td>71.81%</td>
</tr>
<tr>
<td>Auto Text</td>
<td>325</td>
<td>185</td>
<td>56.90%</td>
<td>159</td>
<td>157</td>
<td>85.95%</td>
<td>48.92%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1282</strong></td>
<td><strong>717</strong></td>
<td><strong>55.90%</strong></td>
<td><strong>648</strong></td>
<td><strong>634</strong></td>
<td><strong>90.38%</strong></td>
<td><strong>50.55%</strong></td>
</tr>
</tbody>
</table>
Nonresponse?

• no evidence that different kinds of people (age, gender, ethnicity, race, education, income) from our sample were any more or less likely to start the interviews in the different modes
• Implausible that another factor could explain pattern:
  – would require that tendency of Rs to give imprecise answers and reluctance to engage in a text interview (but willingness to engage in a voice interview) would have the same origin
Noncompletion?

- Completion greater in human than automated interviews
- No difference between text and voice
  - → Unlikely to account for voice vs. text differences
- For noncompletion to account for disclosure, would require a systematic reversal of the pattern of disclosure observed for those who completed and those who broke off
  - those who broke off with automated interviewers would have to be those who had less to disclose
  - But one would think that people who break off would be those with more to disclose
Prefer text (vs. voice) for future iPhone interview?

<table>
<thead>
<tr>
<th>Percent</th>
<th>Text</th>
<th>Voice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>80</td>
<td>30</td>
</tr>
<tr>
<td>Automated</td>
<td>90</td>
<td>50</td>
</tr>
</tbody>
</table>

- Text: 80% Human, 90% Automated
- Voice: 30% Human, 50% Automated
Overall, how satisfied were you with the interview?

Most Rs satisfied with interview

<table>
<thead>
<tr>
<th></th>
<th>Voice</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Voice</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automated</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Somewhat satisfied

Very satisfied
Other satisfaction measures

• Most Rs found interview very or somewhat easy
  – More found speech-IVR somewhat hard

• Future interviews:
  – Text Rs overwhelmingly preferred future interview in text vs. voice
  – Voice Rs preferred voice, but less so if speech-IVR
Response Rates, Contact Attempts

Response Rate

<table>
<thead>
<tr>
<th>Percent</th>
<th>Human Voice</th>
<th>Human Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Complete Cases: Contact Attempts per Interview

<table>
<thead>
<tr>
<th>Contact Attempts</th>
<th>Human Voice</th>
<th>Human Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Text took more interviewer time than voice, but less overall time

<table>
<thead>
<tr>
<th></th>
<th>Human Voice</th>
<th>Human Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average time in field</td>
<td>44.4</td>
<td>25.9</td>
</tr>
<tr>
<td>Average interviewer time (HPI)</td>
<td>0.8</td>
<td>1.9</td>
</tr>
</tbody>
</table>

HPI difference would be smaller if interviewers worked more than one text case at a time.
Summary: Voice vs. Text

• Text interviews produce higher data quality: greater disclosure, less satisficing, high satisfaction
• Even though (or because?) they take longer
• Even though data are less secure (more persistent and traceable) than voice
  – Perhaps because of different time pressure than voice?
  – Perhaps because of convenience of answering when and how R wants?
  – Perhaps because of greater social distance with interviewer?
• Caveat: we implemented text interviews in one particular way, with single-character responses
Summary: Human vs. Automated Interviewer

- Automated interviews on a smartphone (in these modes) can lead to data at least as high in quality as data from human interviews in same modes
  - No more satisficing than with human interviewers!
  - More disclosure
- Tradeoffs
  - Field period can be shorter
  - Interviews can take longer
  - Higher break-off
  - Require additional development effort, especially speech-IVR
- Caveat: we implemented one particular version of speech-IVR; others could differ
Many questions remain

• Do different demographic subgroups (e.g., age, income, education) vary in disclosure, effort, preferences?
• Generalizability to other mobile platforms? To less smart mobile phones?
• Generalizability to a non-convenience or non-incentivized sample?
• Desirability of multimode options for smartphones?
  – Ability to respond in a different mode than contacted?
  – Ability to switch modes mid-interview when circumstances change (mobile, noisy, private, etc.)?
• What are psychological and social mechanisms underlying
  – willingness to disclose?
  – conscientious responding?
But

• But this initial evidence suggests the viability of text as an interviewing mode
  – as well as questions about whether data from voice interviews in a smartphone era should be considered the gold standard for validity and data quality
Mode Choice Design

1. No Choice (n=642)
   • Rs randomly assigned to a mode
   • Contact and interview in same mode

2. Choice (n=626)
   • Rs randomly assigned to a contact mode
   • Required to choose interview mode
     – could choose contact mode or any of other three
Mode Choice Design (2)

- **Mode Choice introduction:**
  
  “To get started, we need you to choose how you want to be interviewed -- whatever works best for you. There are four choices and any choice is fine with us. Do you want to ‘talk with a person’, ‘talk with an automated interviewer’, ‘text with a person’, or ‘text with an automated interviewer’?

- **Within each contact mode, order of interview mode options rotated across Rs (16 orders)**
Possible outcomes: Conscientious responding

• If Rs can Choose
  – might provide less conscientious answers because they choose a mode in which it’s easier to take shortcuts
    • e.g., an automated mode because no human interviewer to press them to work hard
  – might provide more conscientious answers because being able to choose may increase their commitment to the task
Possible outcomes: Disclosure

• If Rs can Choose
  – might disclose more because choose more private mode with fewer social cues
    • e.g., Automated text
  – might disclose less because choose more convenient, faster mode with more social cues
    • E.g., human voice
Possible outcomes: Participation

• If Rs can Choose
  – Might **reduce** participation because
    • Increased complexity (Schwartz, 2004; Fulton and Medway, 2012)
    • Break in response process (Fulton & Medway, 2012)
  – Might **increase** participation because
    • Can choose a mode that is suitable given their current environment and other demands (e.g., whether they can talk now)
Possible outcomes: Satisfaction

• If Rs can Choose
  – Might reduce satisfaction because
    • Adding options increases R’s expectations (Schwartz, 2004)
    • Leads to regret over not choosing imagined alternative (Schwartz, 2004)
  – Might increase satisfaction because people perceive the chosen alternative as more attractive (Festinger, 1948; Cooper, 2007)
Respondents: 1260 iPhone users

• **No Choice (Assigned to Mode):** n = 634
  – n = 157 to 160 per mode
  – Interviewed March – May, 2012

• **Choice (Able to Choose Interview Mode):** n = 626
  – n = 149 to 170 per mode of contact
  – Interviewed July – September, 2012

• Recruited from Craigslist, Facebook, Google Ads, and Amazon Mechanical Turk
  – Web screener verified age (> 21 years) and US area code
  – iPhone usage verified via text message to device and user agent string in response

• $20 iTunes gift code incentive, provided after post-interview web questionnaire

• Age, gender, ethnicity, income, education not reliably different between Choice and No Choice groups

• Sample is somewhat younger and less affluent than US national iPhone users
Participation

Overall completion higher without (50.5%) than with choice (46.4%)
Breakoffs (after choice) seem to depend on which modes Rs switch from and to

• Some transitions have no breakoffs at switch
  – Human to Human modes: 0% - 2.8%
  – Human Text to Auto Voice: 0%
  – Auto Text to Auto Voice: 0%

• Others involve more breakoffs
  – Any automated to any human mode: 4.5%-22.2%
    • Probably because continuation is not immediate in our implementation
    • Suggests on-demand human interviewers could substantially reduce these breakoffs
Breakoff before interview not due to *paradox of choice* but to switching costs

% don’t answer Q1 after choosing mode

- Stay in Mode (n=301): 0.7%
- Switch Mode (n=388): 11.1%
Preference for automated text

Number of Rs

- Human Voice
  - Switch into Mode: n = 170
  - Stay in Mode: n = 150

- Human Text
  - Switch into Mode: n = 150
  - Stay in Mode: n = 157

- Automated Voice
  - Switch into Mode: n = 150
  - Stay in Mode: n = 157

- Automated Text
  - Switch into Mode: n = 149
  - Stay in Mode: n = 149

Original Sample Size (before mode choice)
Conscientious responding: Rounding

• We define rounding here as numerical answers divisible by 10
  – How many songs do you currently have on your iPhone?
    • Example rounded answer: 1100
    • Example unrounded answer: 1126
Effect of choice not due to particular choice of mode: less rounding with choice than without after controlling for mode, $p = 0.008$
Rounding: “Number of songs on your iPhone”

*How many songs do you currently have on your iPhone?

\( p = 0.02 \)
Satisficing: Straightlining

• Q: support for various dietary practices (eating red meat, limiting fast food, etc.)
  » strongly favor
  » somewhat favor
  » neither favor nor oppose
  » somewhat oppose
  » strongly oppose

• We define answers in battery as “straightlining” when at least 7 of 8 responses are the same

• Less straightlining in text than voice ($p < .01$)
• Less straightlining with choice than without ($p = .029$)
Satisficing: Percent of Rs straightlining

Effect of choice not due to particular choice of mode: marginally less straightlining with choice than without after controlling for mode, \( p = 0.085 \)
Satisfaction higher with mode choice

Overall, how satisfied were you with the interview?

$p < 0.001$

- Increased satisfaction may result because people perceive the chosen alternative as more attractive (Festinger, 1948; Cooper, 2007)
- Or may better fit their needs
Reasons for choosing modes

• Coded open-ended answers into 29 categories

<table>
<thead>
<tr>
<th>Most common categories</th>
<th>% Providing Reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ease/simplicity</td>
<td>33.8%</td>
</tr>
<tr>
<td>Convenience/flexibility</td>
<td>22.8%</td>
</tr>
<tr>
<td>Quickness (shortest interview time)</td>
<td>10.3%</td>
</tr>
<tr>
<td>Privacy</td>
<td>9.8%</td>
</tr>
<tr>
<td>Like texting</td>
<td>9.0%</td>
</tr>
<tr>
<td>Environment--location</td>
<td>8.8%</td>
</tr>
</tbody>
</table>

Three coders; Agreement= 98.1%

*Why did you choose this interviewing method?
Reasons for choosing modes* (examples)

- **Human voice:**
  - “More comfortable speaking with a real person”

- **Human text:**
  - “I chose to text because I had a small child with me in my home during the interview and could not have concentrated on the questions if it was on the phone.”
  - “To avoid background noise and to clearly understand the question and take my time to answer it.”

- **Automated text:**
  - “I am at work and wouldn't always be able to answer questions if I spoke to someone on the phone.”
  - “Because I didn't want to talk on the phone nor did I want to text a person simply because I knew some of my responses would have been a little late”

- **Speech IVR:**
  - “I didn't want to talk to anyone but, I was driving so I couldn't look at a screen”
  - “Talking to an automated person was less personal”

*Why did you choose this interviewing method?
Summary

• Mode choice produced:
  – less rounding
  – less straightlining
  – fewer breakoffs
  – higher R satisfaction

• Text led to less rounding and straightlining than voice
  – Irrespective of mode choice

• Participation
  – Lower ‘start rates’ with choice than no choice
    • not offset by reduced breakoffs among those who chose
Mode choice on a single device opens up new possibilities

• Potential Rs (with smartphones) are already choosing modes all the time
• Each kind of mode transition may have different properties
  – Immediacy of continuation and convenience differ
  – Need to better understand costs and benefits for survey data
• Our data suggest that choice per se does not suppress participation
  – more likely due to break in response process when Rs switch modes
Implications

• Particular mode transitions seem more and less attractive for survey designers
  – And different design solutions will be needed for different mode transitions
• In any case, higher completion rates, less satisficing, and more satisfaction all argue for further exploring mode choice on a single device