Three New Survey Modes and Their Impact on Data Quality

Frederick Conrad

Program in Survey and Data Science

University of Michigan

Dedication

- Reg Baker, 1947-2024
- Thought leader, and early (and consistent) champion of judiciously integrating technology into survey and market research
- Chaired AAPOR task forces on Non-probability Sampling and Online Panels
- (Co)edited *CASIC* (1998) volume, impetus for FedCASIC workshops
- Chapter in AAPOR's A Meeting Place and More… (2020) simply titled "Technology"
  - Documents inherent tension – when survey researchers consider a new technology – between increasing efficiency/ reducing costs on the one hand and potentially harming data quality on the other

Introduction

- Survey researchers have adopted new technology, even if skeptically
  - Telephone communication, the web, and mobile devices have transformed survey data collection
- New tech that is well-suited to data collection (and often does not require much coding) has recently exploded on the scene
  - Interactive video, like Zoom or Teams, on all kinds of devices
  - Large language models
- And some older technology has attracted new attention
  - Text messaging

Agenda

- Text message interviews
- Video mediated interviews
- Virtual (animated) interviewers

Acknowledgement

- Collaborators: Michael Schober, Andrew Hupp
- Co-Authors: Chris Antoun, Patrick Ehlen, Stephanie Fail, Makenna Harrison, Matt Jans, Michael Johnston, Kallan Larsen, Rachel Levenstein, Laura Lind, Danny Nielsen, Ai Rene Ong, Rachel Orlowski, Heidi Reichert, Lucas Vickers, Tianheao Wang, Brady West, Yanna Yan, Chan Zhang
- Financial Support:
  - NSF/Methodology, Methods, and Statistics
  - Michigan Center on the Demography of Aging

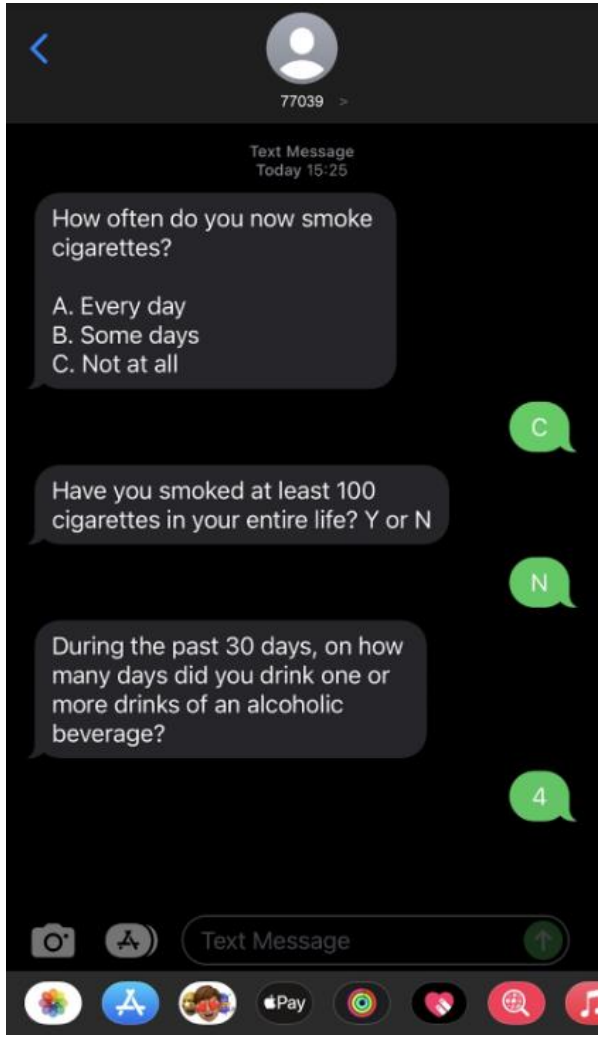
1. Text Messaging for Administering Surveys

- Texting has been used to
  1. invite sample members to respond to a web survey (“text-to-web”)
  2. conduct the interview, i.e., ask questions and capture answers
- Focus today on texting as an interview mode

Texting Has Properties that May be Good for Survey Data Quality and Operational Efficiency …

1. Asynchronous
   - Replying to a text message is generally less time sensitive than replying to what a speaker has said
   - Allows *R*s to take as much time as needed to think about and formulate their answers
2. Noticeable
   - Arrival of a text message is usually signaled by visual, auditory, and haptic notifications
   - Increases *R*s' awareness that they have been contacted (invited to participate or asked a survey question) and/or that they have not responded
3. Persistent
   - Text remains visible indefinitely in contrast to speech, which is ephemeral
   - Allows *R*s to reply when it is convenient and safe because they don't have to remember what was in the message – it is still available

Text Message Survey Interviews



- Each question is a message; each response is a message
- *R* must enter (type) answers – no radio buttons, etc.
- Businesslike interaction
  - Relatively few turns, occasional *Iwer* probes
- Minimal "presence" of interviewing agent
  - Not obvious to *R* whether agent is human or automated

Text Interviewing Experiment

- We* compared satisficing (mental shortcuts) and disclosure of sensitive info in 4 modes (n ≈ 160 per mode)

1. Automated Text
2. Human Text
3. Automated Voice
4. Human Voice

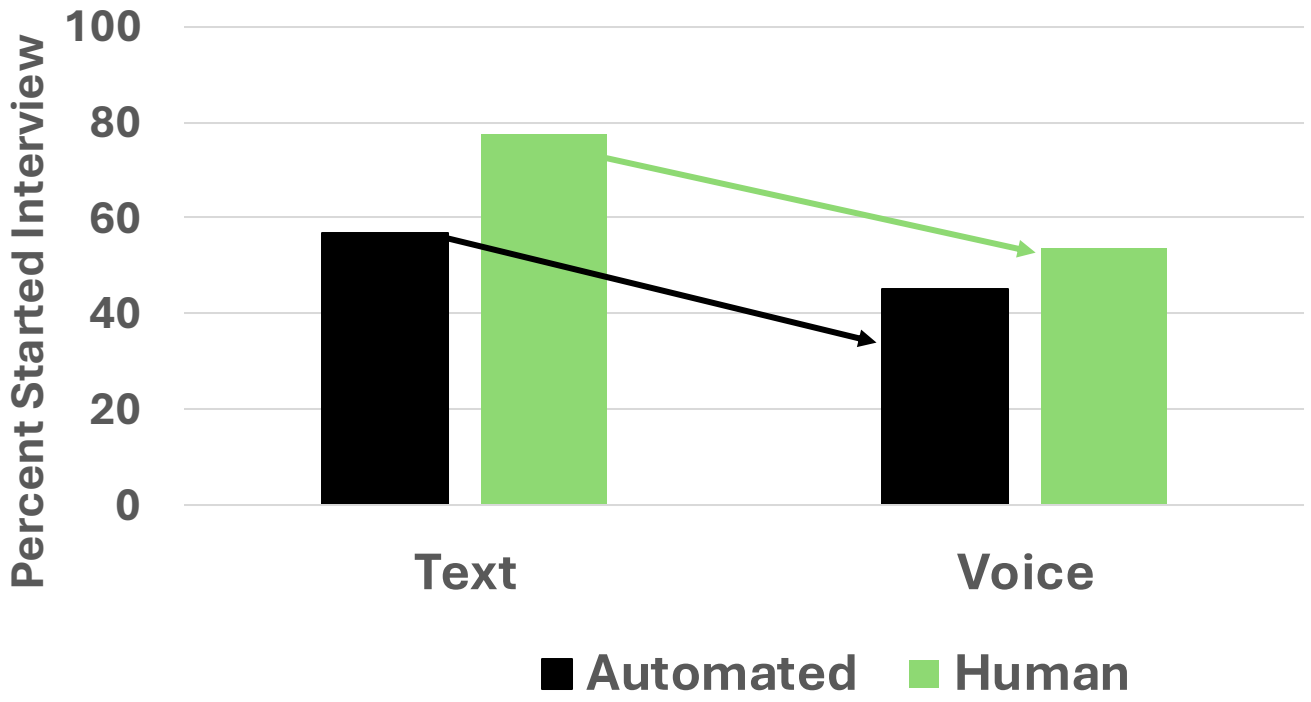
*Schober, M.F., Conrad, F.G., Antoun, C., Ehlen, P., Fail, S., Hupp, A.L., Johnston, M., Vickers, L., Yan, H., & Zhang, C. (2015). Precision and disclosure in text and voice interviews on smartphones. PLOS ONE 10(6): e0128337. doi:10.1371/journal.pone.0128337

Conrad, F.G., Hupp, A.L., Schober, M.F., Antoun, C., Yan, H.Y., & Harrison, M. (under review). Text message interviewing: A deeper look.

Starting the Interview (answering at least 1 *Q*)

- Noticing a texted invitation should increase sample members' awareness they have been invited to a text message interview
- Persistence of invitation enables *R* to start the interview as soon as it is convenient and safe to do so
- Should lead to higher proportion of started Text than Voice interviews

Start Rate (answer at least one *Q*)



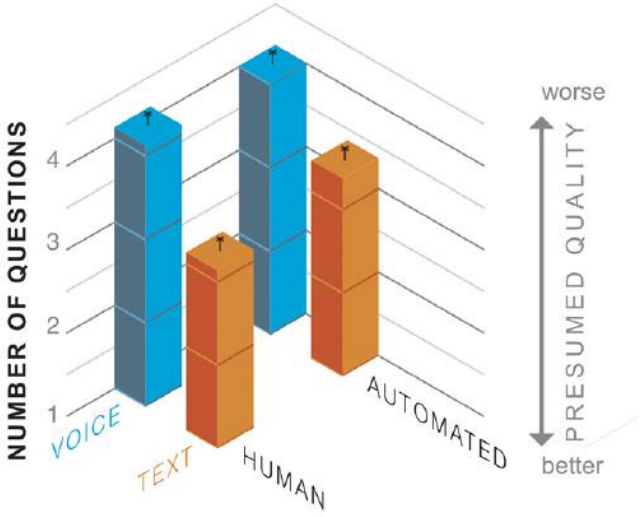
- Presumably due to text invitations being more noticeable and persistent than phone invitations

Data Quality Measures

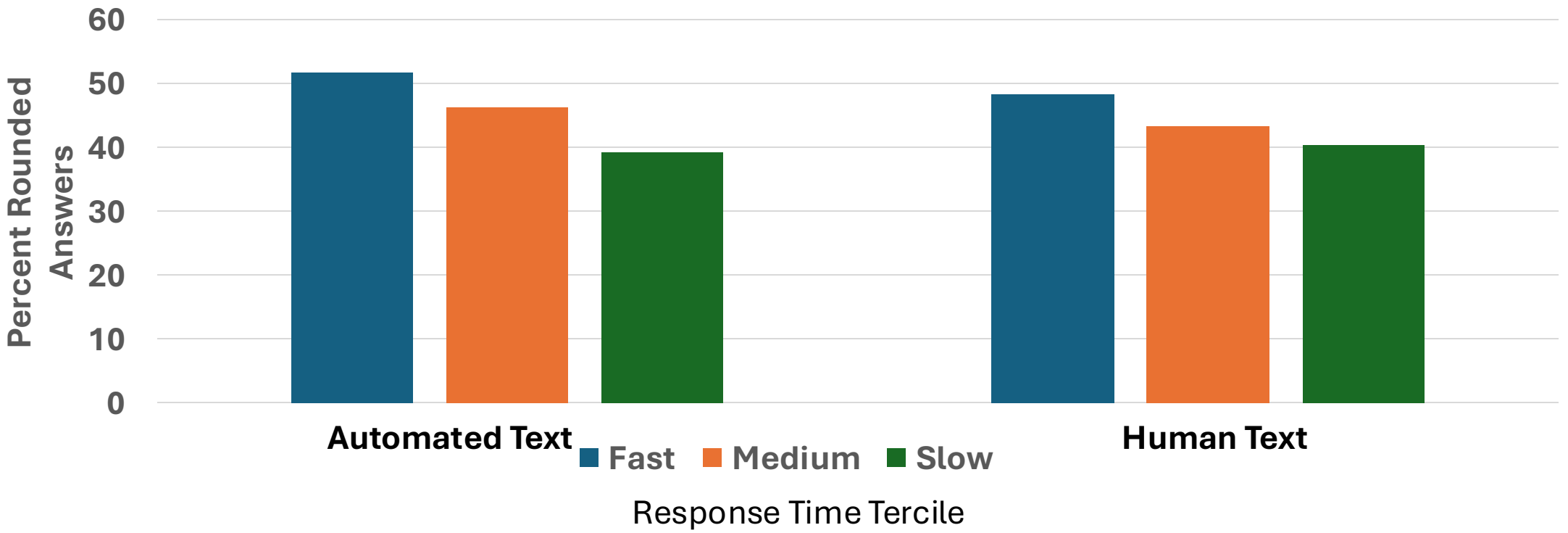
- Rounding:
  - Numerical responses that are round numbers, e.g., 10, can reflect less careful thinking than unrounded responses, e.g., 9,
- Straightlining (non-differentiation):
  - Selecting the same response to all or most items/statements in a battery can reflect less careful thinking than some differentiation
- Disclosure (of sensitive info)
  - More socially undesirable responses (disclosure) can reflect more candor

Data Quality in Text vs. Voice Interviews

A: Rounding
Numerical answers ending in 0 or 5



Higher Quality Data in Texting (i.e., less rounding) Due to Asynchrony

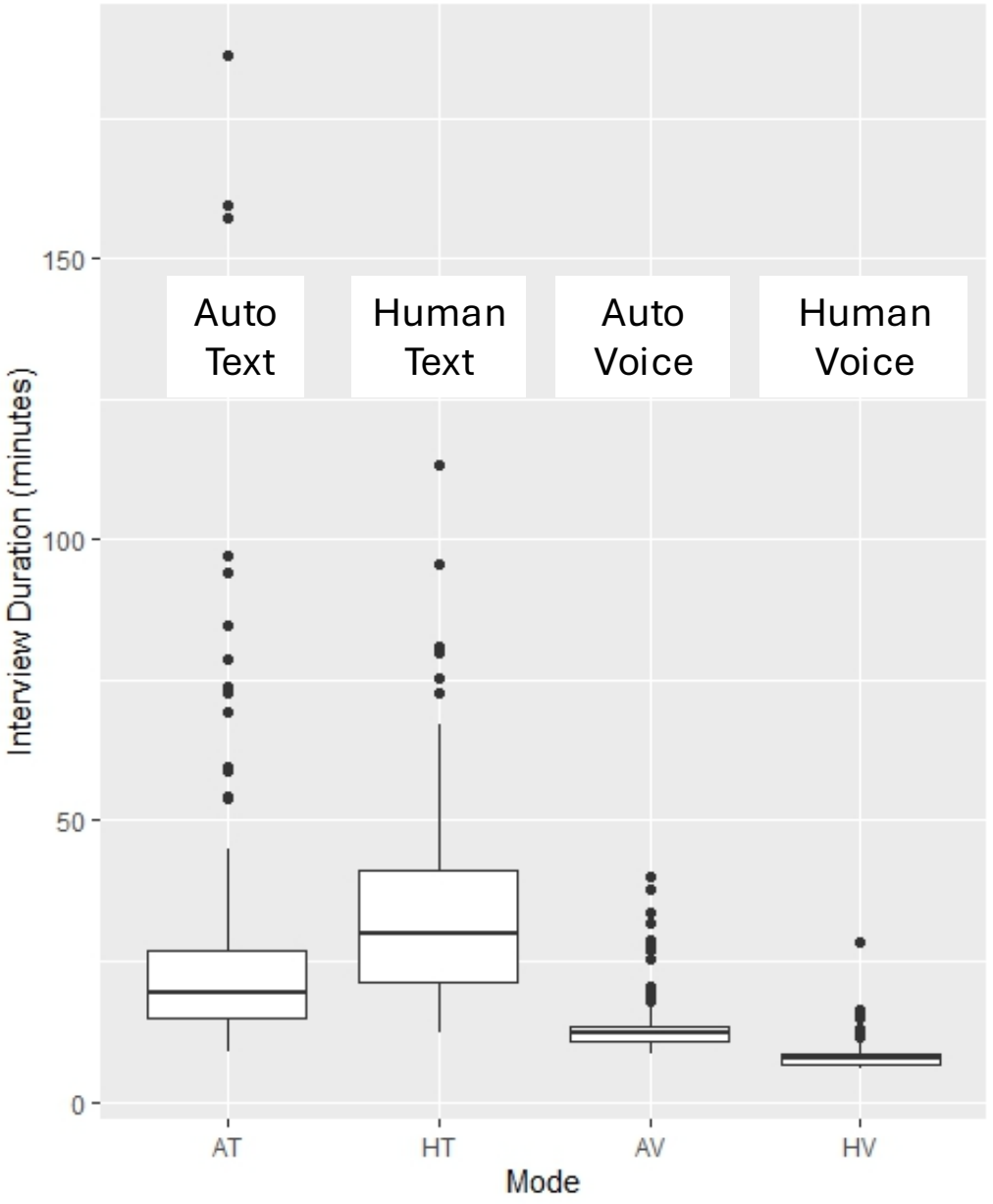


Interview Duration

- Because *R*s can devote as much time to each response as needed, RTs should be longer in Text than Voice interviews
- Because speakers fill silence if > 1 second elapses, Voice RTs should be shorter (faster) in Voice than Text interviews
  - longer pauses are uncomfortable (Jefferson, 1988; Roberts & Francis, 2013)
  - *R*s may tolerate longer silences while formulating an answer, but same principle applies, especially if hear *Iwer* on the phone
- Longer texted responses should add up to longer Text than Voice Interviews

Interview Duration

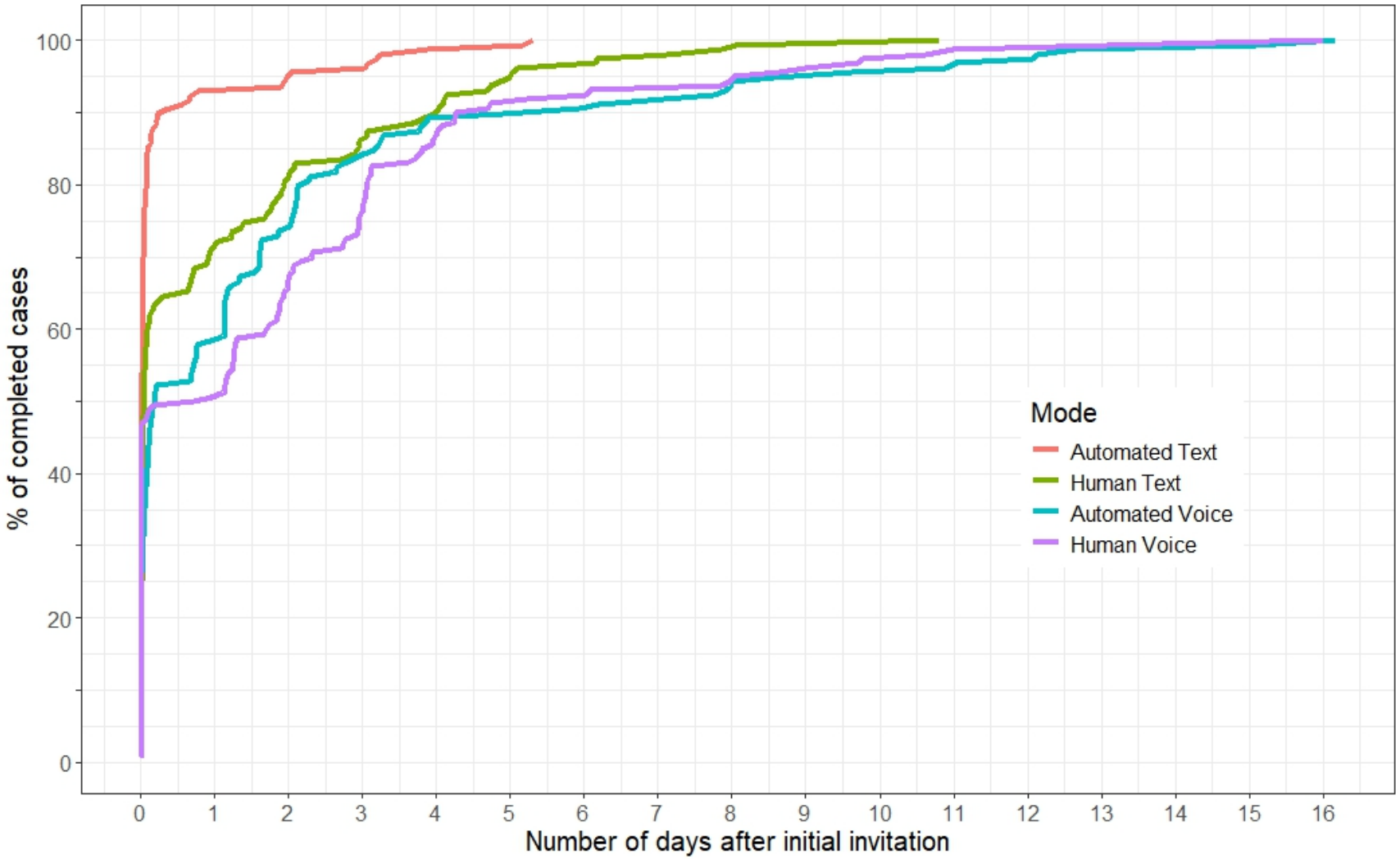
- Put a pin in this result: longer Text than Voice Interviews
- We'll come back to it



Field Work

- Text interviews can be completed when timing is good for *R*, i.e., *R*s can shoehorn individual responses into their schedules
  - *Interstitial* responding
- Voice interviews require *R* to (1) answer their phone, and (2) be available for the entire interview
- So, if contacted, voice *R*s may be less willing to participate until they have sufficient uninterrupted time
- These differences could result in faster completion of Text than Voice cases from start of field period

Time (Days) to Complete Field Work

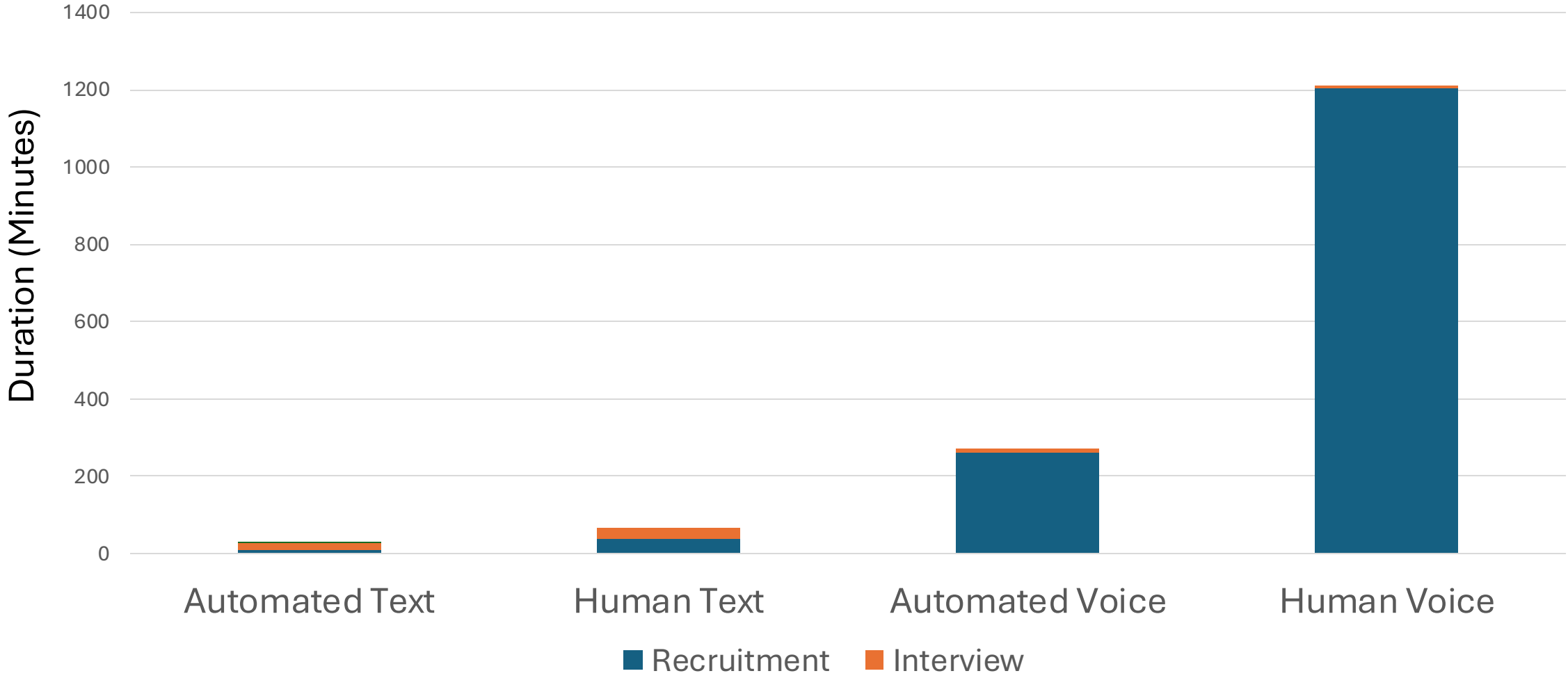


- Automated Text cases completed in 6 days, 93% in one day, most of those in one hour
- Human Text cases completed in 11 days, 72% in one day
- Voice cases were completed in 16 days
- See Spiegelman & Zotti (2024) for a similar finding

How can Field Work be Quicker but Interviews Slower for Text than Voice Data Collection?

- Text *R*s are recruited faster than Voice *R*s
  - Text invites reach and are usually noticed by sample member on first try
  - Voice invites often require multiple contact attempts, typically separated by at least a day to avoid annoying the sample member
- Faster recruitment into Text interviews outweighs faster completion of Voice than Text interviews
  - i.e., very long recruitment duration in Voice interviews swamps shorter interview duration

Case Completion = Recruitment + Interview



Very quick field period for text interviews suggests

- May be well suited to studying experience of sample members in aftermath of a public event like a natural disaster or a political debate
- Can help assure accurate snapshot of volatile phenomena as is common in public opinion
- Presumably this is why CNN and the Washington Post polls use text message interviews for quick turnaround research

Takeaways: Text Message Interviewing

- Compared to phone interviews, texting provides multiple benefits to researchers:
  - higher start rates
  - increased data quality: less satisficing, more disclosure
  - quicker completion of field activities
- Asynchrony, noticeability, and persistence seem largely responsible
- Text message interviews fit naturally into the communication options people have come to expect and may be the only way to collect data from some members of the public
- At least for some projects, text message interviews may be just right

2. Video Interviews

- Two-way, synchronous video communication became familiar to most people (at least in WEIRD* countries) during pandemic
  - especially those whose work or school required it
- Beyond increased familiarity, why might video interviewing be advantageous?

*Western, Educated, Industrialized, Rich, and Democratic

Potential Advantages of Video Interviews?

- Eliminates, at least in principle, most travel costs incurred in in-person interviewing
- Reduces safety and public health concerns about in-person visits
- Allows presentation of visual aids (e.g., showcards) to *R*s
- Facilitates interviewing sample members who may be
  - hard to reach in person (e.g., rural, deployed military)
  - unable to participate in other modes, e.g., low literacy, sensory disability (can turn up volume or brightness)
- Mediated, so may provide some sense of social distance from *Iwer* compared to in-person interviews

Potential Disadvantages of Video Interviews?

- Coverage:
  - Not everyone has access to the technology or experience using it
  - *R*s need a device with a camera and sufficient connectivity, that they can use
  - No master frame of email addresses or other contact info to invite sample members
- Recruitment may be difficult
  - Live video may seem intrusive (at least compared to Phone)
  - Mediated experience may be less engaging than in-person
  - *R*s must schedule interviews; cold calls are impractical
- Start-up and infrastructure costs for research organizations

Mode Comparison Experiment*

- Compared data quality and respondent subjective experience in
  1. Video interviews (n=279)
  2. Prerecorded Video "interviews" (n=385)
  3. Conventional text-based, Web Survey (n=403)

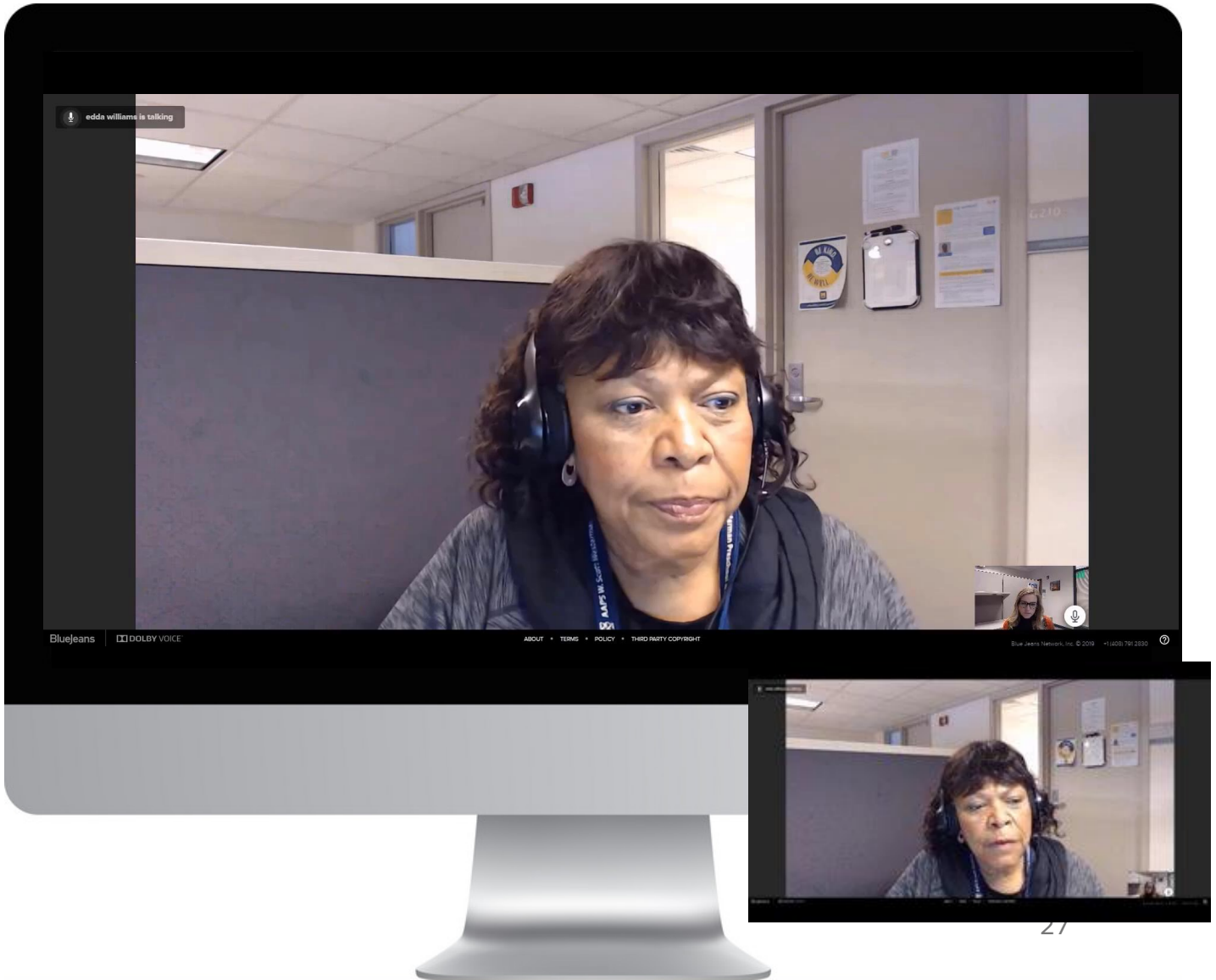
Current Focus: 1 vs. 3

- Why compare Video interviews to Web surveys?
  - In Video interviews, *Iwers* may motivate *R*s to try hard; Web is self-administered, i.e., no *Iwer*
  - In Video, *Iwer* and *R* see and hear each other, maximizing social presence; in Web, *R* likely to feel more private and anonymous
- To what extent is Video like in-person, interviewing?
  - Note: we consider both video and in-person interviews to be FTF

*Conrad, F, G., Schober, M.F., Hupp, A. L., West, B.T., Larsen, K. M., Ong, A. R., & Wang, T. (2022). Video in survey interviews: Effects on data quality and respondent experience. *methods, data, analysis*.

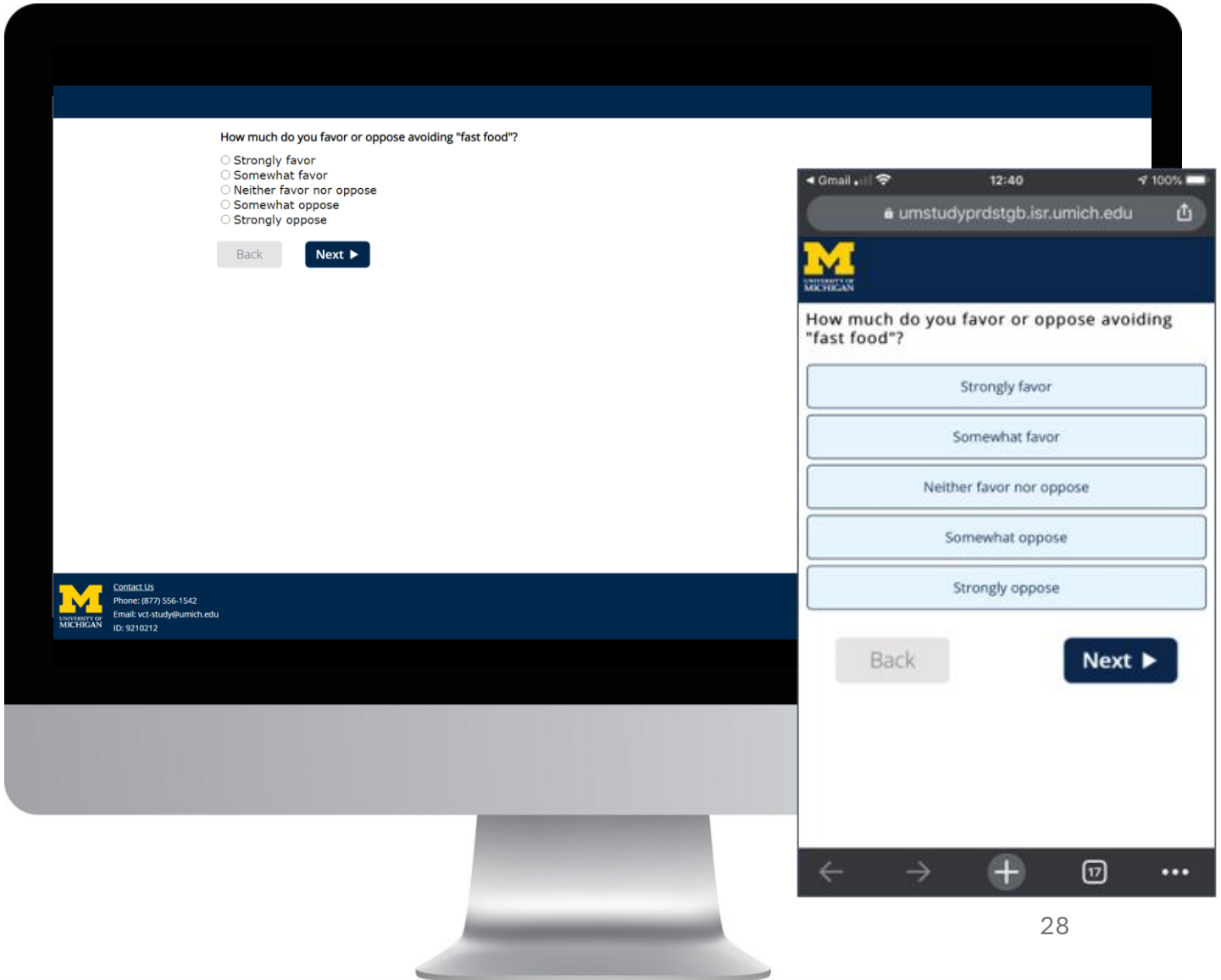
Video Interviews

- BlueJeans video + Blaise 5
- 9 professional *Iwers* (from U-M Survey Research Center) conducted interviews
- *R*s scheduled their own interviews online



Web Survey (WS)

- Textual presentation of *Q*s and response options (Blaise 5)
- *R*s select or type answers



Respondents and Items

- *R*s recruited from two opt-in, online sample sources:
  - CloudResearch and MICHR
  - Randomly assigned to survey mode
- 36 *Q*s – chosen to assess data quality
  - 7 Numeric (to examine rounding)
  - 17 statements in three Batteries (to examine straightlining)
    - implemented as series of individual statements in both modes
  - 9 Categorical, 3 Numeric (to examine disclosure)
- Post-interview debriefing administered textually in both modes

Results: Rounding and Straightlining

- Rounding
  - More *R*s rounded (at least once) in Video (86.9%) than Web (82.0%)
  - Presence of live *Iwer* in Video may have created pressure to respond immediately to avoid awkward silences (e.g., Jefferson, 1988; Roberts & Francis, 2013)
- Straightlining
  - Fewer *R*s straightlined one or more times in Video (1.7%) than Web (7.7%)
  - Presence of a live *Iwer* in Video seems to have motivated *R*s to attend thoughtfully to all statements in the batteries

Results: Disclosure

- Proportion of responses that are sensitive is lower in Video (0.56) than Web (0.58)
  - Small effect but significant
  - Large effect for three items: Volunteer work, Vote in local elections, Visit pornography site
  - Underscores how socially present Video *Iwer* is despite being mediated; as in in-person interviews, seems to inhibit honestly reporting sensitive information (e.g., Tourangeau and Smith 1996)

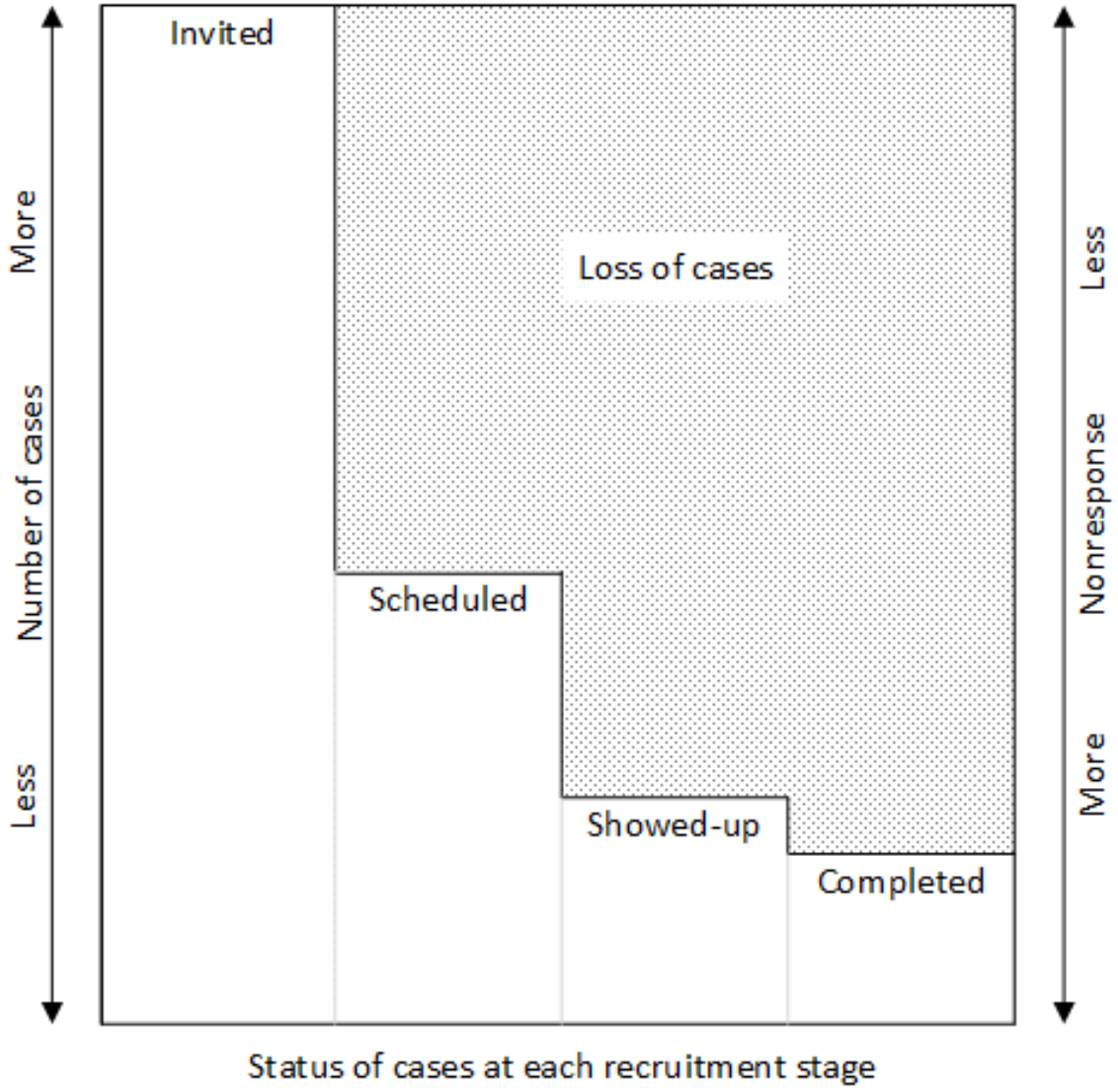
Results: Respondents' experience

- More *R*s "Very Satisfied" with Video (73.9%) than Web (52.1%)
- even though took longer: Video = 9.46 min, Web = 5.85 mins
- Video *R*s asked to compare privacy of just-completed interview to hypothetical in-person interview: same, more private, less private
  - 75.0% said "same" and 23.4% said "more private"
  - So, nearly 1 in 4 found the Video interview more private than in-person
  - Does mediation provide a "protective barrier?"
- How personally connected to Video *Iwer*: 4.59 (out of 5)
- How comfortable with the *Iwer:* 4.70 (out of 5)

How similar are Video and In-person interviews?

- Data quality differences between In-person and Web surveys in the literature mirror our data quality findings, comparing Video and Web
- Rounding:
  - Liu & Wang (2015) report more rounding on feeling thermometer *Q*s when administered in person than on a web survey
- Straightlining:
  - Heerwegh & Looseveldt (2008) report less straightlining in person than on web survey
- Disclosure/ Social Desirability Bias:
  - Endres et al. (2023) report less disclosure, i.e., more socially desirable responding, on feeling thermometer *Q*s in person than on a web survey
  - And, in a direct comparison of the two *Iwer* administered modes, no differences on the same task
- But, despite quite similar effects on data quality, the two *Iwer* modes are operationally very different – *R*s need to schedule an appointment

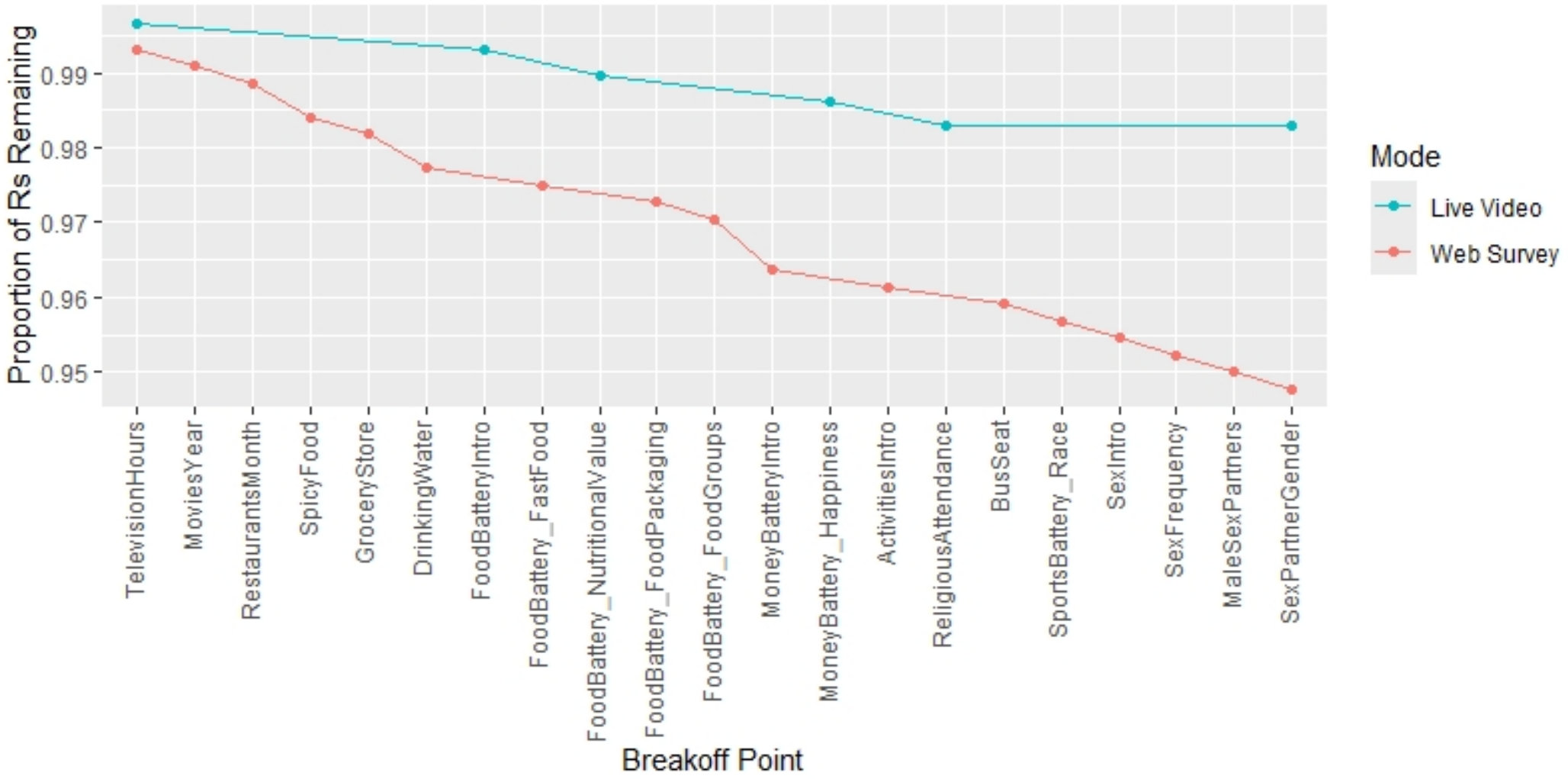
Self-scheduling interview may be obstacle to completion *



Recruitment Stage	n	%
Invited	5,783	100.0
Scheduled	593	10.3
Showed Up	309	52.1
Completed	286	92.6

* Hupp, A.L., Conrad, F.G., Larsen, K., Schober, M.F., West, B.T., Harrison, M. (under review)

More Completion in Video than Web



Take-Aways: Video Interviews

1. Video interviews are a lot like in-person interviews compared to web
   - More time pressure ➡ more rounding
   - More social pressure to be conscientious ➡ less straightlining
   - Less privacy than in web survey ➡ less disclosure
2. *R*s prefer their experience in Video interviews to Web survey
   - Hint that mediation in Video interviews provides some privacy compared to in-person interviews
3. Scheduling interview may introduce challenge to recruitment and completion
4. Video interviewing is viable, but may be more effective as a complementary than stand-alone mode, attracting sample members who will not participate in other modes
   - see forthcoming special issue of *methods, data, analyses* on video interviewing

3. Virtual Interviewers (*VI*s)

- Automated agents that ask survey questions usually by speaking (sometimes via text), are often embodied (have a face), and may interact with *R*s (but sometimes just ask the *Q*s)

1. Do *R*s react socially to *VI*s, for worse or better
2. Can *VI*s help *R*s understand *Q*s as intended

*VI*s (like human *Iwers*) Can Reduce Disclosure

- We* (2013) compared disclosure of sensitive info when *Q*s asked by
  1. Voice only (ACASI)
  2. In person interviewer
  3. *VI* with high facial movement
  4. *VI* with low facial movement
- Lab study with n ≈ 60 in each condition
- How did presence of digital face, i.e., *VI*, affect disclosure?
- Focus on 9 *Q*s for which more disclosure in 1 than 2

* Lind, L. H., Schober, M.F., Conrad, F.G. and Reichert, H (2013). Why do survey respondents disclose more when computers ask the questions? *Public Opinion Quarterly 77*, 888–935.

*VI*s Can Reduce Disclosure (2)

- ○ Not at all in the past year
- ○ Once in the past year
- ○ Two or three times a year
- ○ Once a month
- ○ Once a week
- ○ More than once a week
- ○ Don't know

Repeat question | Next →

- For 8 of the 9 *Q*s, more disclosure in Voice only than either *VI*
- Suggests
  1. *R*s react to *VI* much as to human *Iwer*
  2. a moving face, even if clearly digital, seems to trigger a social reaction which, in this case, leads to less disclosure

Note: this *VI* (Victoria) could only ask *Q*s; *R*s answered by clicking a response option

*VI*s Can Increase Disclosure

- Demographic matching between *R*s and human *Iwers* can increase disclosure
  - If same Gender, *R*s reported more sexual behavior (Catania et al., 1996)
  - The more attributes they share (Gender, Age, Race, Education), the more drug use *R*s report (Johnson et al., 2000)
- Can this be replicated with *VI*s?
- We* matched *VI* and *R*s (n=1735) on race, gender, both, or neither from opt-in, online sample source
- Compared socially undesirable (and presumably more honest) responses in each match condition

*Conrad, F.G., Schober, M.F., Nielsen, D., & Reichert, H. (2020). Social identities of virtual interviewers and their impact on survey responses. In K. Olson, J.D. Smyth, J. Dykema, A.L. Holbrook, F. Kreuter, & B.T. West (Eds.), Interviewer effects from a Total Survey Error perspective (pp. 149-164). Boca Raton, FL: CRC Press.

*VI*s Can Increase Disclosure (2)

- *VI*s asked *R*s:

  How would you describe your weight? Very Underweight, Slightly Underweight, Slightly Overweight, Very Overweight

- More *R*s reported "Slightly Overweight" when same race as *VI* but not under the other match conditions
- Just one *Q* and no increase in reporting the most undesirable category
- but some additional evidence that *VI*s elicit social response which might be judiciously exploited to increase disclosure

More evidence that *R*s react socially to *VI*s

- After main data collection, asked *R*s to choose a *VI* for a hypothetical future interview
- After choosing the *VI*, *R*s asked to provide reason for their choice
  - many *R*s explained their choice in terms of *VI*s' identities and personalities, as well as the feeling they got from the *VI*

Survey

msisurvey.com https://test45.msisurvey.com/mrlWeb/mrlWeb.dll

Getting Started Latest Headlines Most Visited

Survey

MARKETSTRATEGIES INTERNATIONAL

Q36

If you could choose one of these interviewing agents to ask you questions for a future interview, which one would you choose? You will not actually take part in an interview; we just want to know which agent you would choose if you were going to participate in an interview. Please check the box below the interviewing agent you choose to conduct an interview.

◉ Agent 1 ○ Agent 2 ○ Agent 3 ○ Agent 4

○ Agent 5 ○ Agent 6 ○ Agent 7 ○ Agent 8

Transferring data from www.msisurvey.com...

Example Reasons for *VI* Choice

- "Looked & sounded the friendliest"
- "The agent was comforting."
- "She is less eerie"
- “because she is a black woman like myself and she looks young and hip but at the same time very mature"
- "her voice is clear”
- "She looked like a real person and I like her voice tone"
- "I just felt more comfortable with him"
- "She is closer in age and racial background”
- "She's a cutie and she sounds like she's smart."
- "the others are scary looking"
- "laid back and i can relate to him"
- "has a more understanding expressional face also good one you can look at it is pleasant voice allso good"
- "He seems to be very forward and not too impersonal like the rest"
- "He's expressions seemed more natural and the eye color wasn't as errie as some of the others."

Potential Benefits of Giving *R*s a Choice of *VI*s

- Could promote disclosure
  - Catania et al. (1996) found more reports of sexual behaviors when *R*s could choose gender of human *Iwer*
- Choosing a *VI* is easier operationally and less complicated ethically than choosing a human *Iwer*
- May help standardize *R*s' experience if all *R*s answer *Q*s from *VI* they feel warmly toward or otherwise prefer
- which may in turn promote (a kind of) rapport between *R* and *VI*
  - Sun et al. (2020) found that the more rapport *R*s reported with human *Iwer*s the more sensitive behaviors they disclosed

What if *VI*s could help *R*s understand *Q*s?

- Could promote comprehension and data quality w/o human *Iwer*
- A conversational *VI* could clarify *Q* if *R* asked for help
- or could offer to help if *R* seemed confused
  - e.g., long pause or disfluent speech
- This requires human interactional skills not available, at least off-the-shelf, until recently
- We* simulated these skills using a Wizard-of-Oz (WOZ) technique
  - experimenter played video recordings of *VI* utterances to give *R* the impression that *VI* (Derek) was conversing with them

* Conrad, F.G., Schober, M.F., Jans, M., Orlowski, R.A, Nielsen, D., & Levenstein, R. (2015). Comprehension and engagement in survey interviews with virtual agents. *Frontiers in Psychology: Cognitive Science, 6*:1578. doi: 10.3389/fpsyg.2015.01578

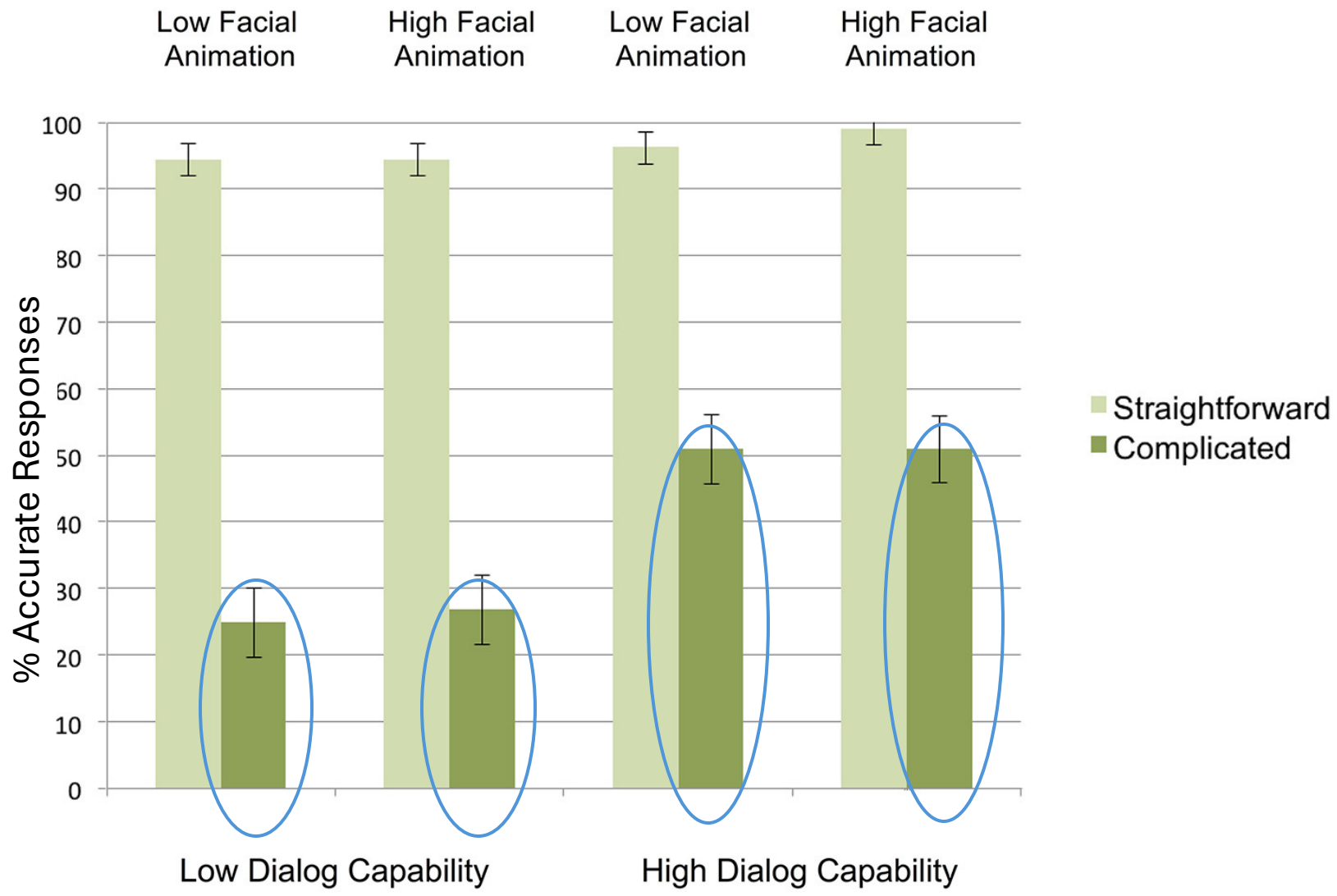
Promoting Better Comprehension with *VI*s

- Laboratory study allowing comparison of response accuracy in interviews with 4 versions of the the *VI*
  1. Dialog Capability: Low or High
  2. Facial Movement: Low or High
- *R*s answered survey questions about fictional characters described in scenarios (vignettes)
- Vignettes were designed to vary the ease of interpreting the questions
  - Straightforward or Complicated (ambiguous)
- Accurately answering *Q*s when scenarios are Complicated requires clarification from *VI*
- Clarification only available when *VI*'s dialog capability is high

VI powered by Human Experimenter (WOZ)

Low Dialog Capability

High Dialog Capability



Response accuracy (percentage of survey responses that matched what the official definition would require) for straightforward and complicated scenarios (error bars represent SE's).

Can High Dialog Capability be Automated?

- Yes, with LLMs
- Text, Voice, or connected to Avatar (possible in principle)
- Would allow the benefits of conversational approach to be implemented at scale
- Demo

*VI*s powered by LLM (voice-only *VI* in this example)

Low Dialog Capability

High Dialog Capability

- Textual prompt given to LLM consisting of (1) the survey *Q*s, (2) the definitions, and (3) instructions about how the Low and High Dialog Capability *VI*s should probe
- LLM provided some incorrect info in response to *R*'s confusion, but mostly got this right
- Driving facial movement in embodied *VI* with LLM output is conceptually straightforward

Take-Aways: *VI*s

- *R*s find it hard to "turn off" their social reaction to *VI*s
  - Could increase engagement and completion but could reduce disclosure of sensitive information
  - Easy to match attributes of *VI* and *R* and *to* allow *R*s to choose *VI,* both of which can potentially increase disclosure
  - *VI* face does not have to be realistic to trigger social response
- *VI*s high in dialog capability can improve *R*s' understanding of *Q*s when circumstances are complicated, increasing data quality
- Technology to deploy *VI*s in production is (mostly) here

Thank You!

fconrad@umich.edu