A Conversation about Conversational Interviewing

Frederick Conrad
University of Michigan

January 28, 2011
Center for AIDS Prevention Studies
University of California at San Francisco

Outline

1. Standardization Debate
    1.1. standardized words or meaning?
2. Empirical comparisons of *standardized* and *conversational* interviewing on the telephone
    2.1. response accuracy and interview duration
    2.2. accuracy-duration tradeoff
    2.3. frequency of situations where clarification helps
3. Extensions
    3.1. Face-to-face
    3.2. Web
    3.3. Virtual Interviewers

Imagine this exchange between an interviewer and respondent

I: Last week, did you have more than one job, including part-time, evening or weekend work?

R: Um…I babysit for two different families. How many jobs is that?

Definition: Job

A job exists when there is a definite arrangement for regular work every week, or every month, for pay or other compensation (e.g., profits, anticipated profits, or pay in kind, such as room and board). A formal, definite arrangement with one or more employers to work on a continuing basis for a specified number of hours per week or days per month, but on an irregular schedule during the week or month, is also a job. It is possible for individuals to have more than one employer but only one job. If an individual does the same type of work for more than one employer in an occupation where it is common to have more than one employer, do not consider the individual a multiple jobholder. Examples include private household or domestic workers including baby-sitters, chauffeurs, gardeners, handypersons, cooks, and maids.

Standardized Interviewing

1. Goal:
   - Replicability
     - ideally interviewers are eliminated as source of error
   - Comparability
     - differences in answers reflect true differences in *R*s' circumstances
2. Practice:
   - Questions read exactly as worded
   - Probe non-directively:
     - e.g. re-read all response categories even if *R* seems to be leaning toward one end of the scale
   - Record answers without discretion
   - Be interpersonally non-judgmental about substance of answers

Standardized Interviewing (2)

- In practice:
  - *I* reads questions exactly as worded; provides no other substance
  - probe neutrally (non-directively)
    - "Let me repeat the question"
    - "Would that be yes or no?"
    - "Whatever it means to you"

I: Last week, did you have more than one job, including part-time, evening or weekend work?

R: Um…I babysit for two different families. How many jobs is that?

I: Let me repeat the question...Last week, did you have more than one job, including part-time, evening or weekend work?

Assumptions of standardization

- *R*s are all exposed to exactly the same question stimulus
- Influence of individual *I*s is minimized
- If question presentation doesn't vary, responses should be unbiased
- so long as questions have been demonstrated in pretests to be intelligible

Standardized words vs. standardized meaning

- Rationale for standardized wording: meaning resides in words
  - if Rs get same words, get same meaning
  - *Message model* (Akmajian et al., 1990)
    - assumes that meaning resides on words, i.e., based only on semantics
    - not explicitly recognized by proponents of standardized wording but implicit in approach
  - Pretesting can remove the most serious and widespread comprehension errors

Standardized words vs. standardized meaning (2)

- *Alternative view*: meaning established through conversation
  - participants talk about meaning to assure understand each other
  - utterance must be *grounded* (H.H.Clark and colleagues)
    - speaker and listener talk about what has been said until both agree they understand each *other well enough for current purposes*
    - can take several conversational turns
  - Pretesting cannot anticipate ambiguities for many respondents in a diverse sample; clarification needed in interview between *I* and *R*
- In strictly standardized survey interviews, grounding not possible

Inability to ground meaning

I: Was there ever a period in your life when you considered yourself to be a light drinker.

Mrs.: (smiles) I guess yes.

I: For how many years were you a light drinker.

Mrs.: You know, do you define—How would you define light drinker?

I: Well this is ho—we would have you define it—

Mrs.: Right.

I: Uh huh.

Mrs.: You know, well if having one or two glasses of wine every night with dinner is (laugh) a light drinker, then I guess I would be characterized a light drinker.

Source: Suchman & Jordan, 1990

Proposal for Conversational Interviewing
(Conrad & Schober, 2000; Schober & Conrad, 1997)

- I and R work together to assure R understands question as intended
  - *I* reads question as worded, then
  - *I* says whatever is necessary to assure *R* interprets question as intended, i.e. to ground meaning*
- Goal is to standardize meaning, even if wording varies

*note:“conversational” refers use of everyday conversational resources such as “grounding” -- not lower standard of precision

But is message model appropriate for surveys?

- Perhaps meaning-in-words assumption is sufficient for large-scale surveys with pretested questions
- Perhaps in standardized surveys enough people understand questions as intended without collaboration

I: Last week, did you have more than one job, including part-time, evening or weekend work?

R: Um…I babysit for two different families. How many jobs is that?

I: In this survey we would count that as one job; you have two employers but one job, namely a baby sitter.

Empirical comparisons of the two techniques

- Schober and Conrad (1997)
  - Laboratory experiment compares accuracy for strictly standardized and conversational interviewing techniques
  - when true values are known
  - using telephone interviewers trained to implement one technique
  - asking questions from on-going, large scale, US government surveys

Schober & Conrad, 1997

- Two groups of *I*s trained on survey concepts
- then, one group trained in strictly standardized techniques
  - read as worded, neutral probes, no clarification, etc.
- other group trained to use conversational technique
  - read as worded, then say whatever is needed to assure *R* understands question as intended
  - *I* can provide definition upon request or when thinks it will help
  - *I* can provide definition verbatim or can paraphrase relevant parts

Questions

- All Rs asked same questions
- 12 questions from three US gov't surveys:
  - CPS (Employment)
  - CPI-Housing
  - CPOPS (Purchases)
- Official definitions existed for key concepts in all questions

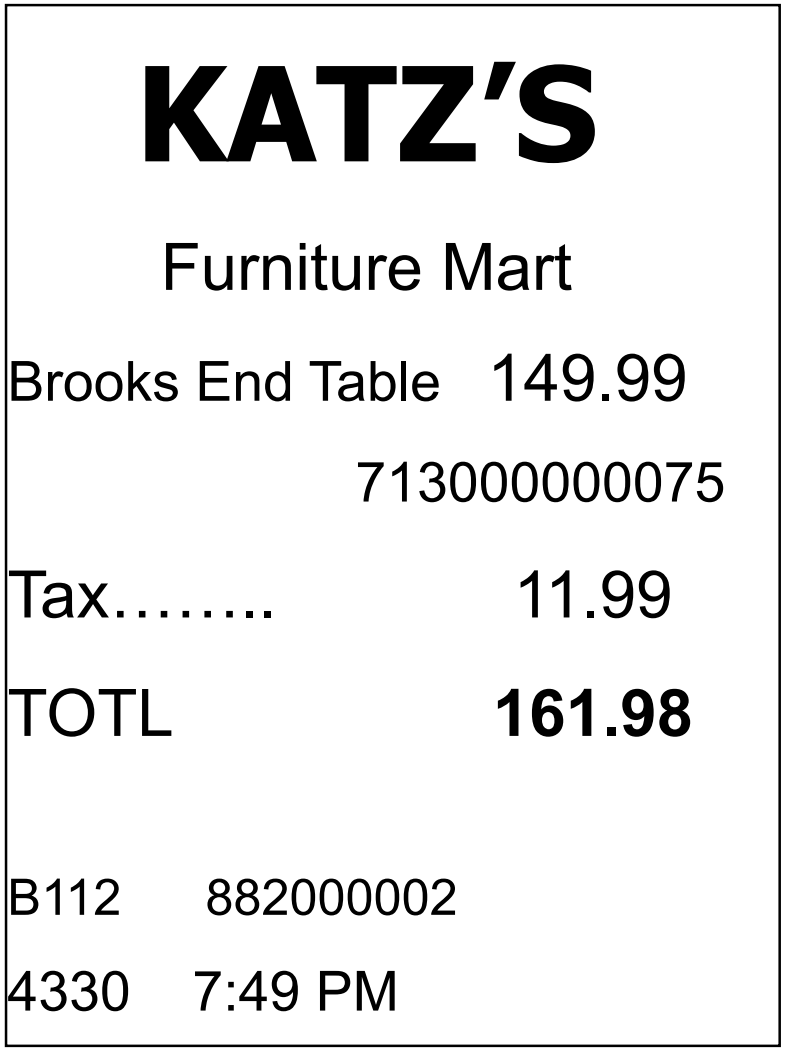
Answers

- *R*s responded on basis of fictional scenarios not available to Interviewers
  - textual vignettes
  - floor plans
  - receipts from purchases
- *I*s never knew correct answer
- *I*s not familiar with content of scenarios
- ★ Use of scenarios enabled direct measurement of response accuracy
  - half complicated, half straightforward

Example question

- *Has Kelly purchased or had expenses for household furniture?*

Example straightforward scenario



KATZ'S

Furniture Mart

Brooks End Table 149.99

713000000075

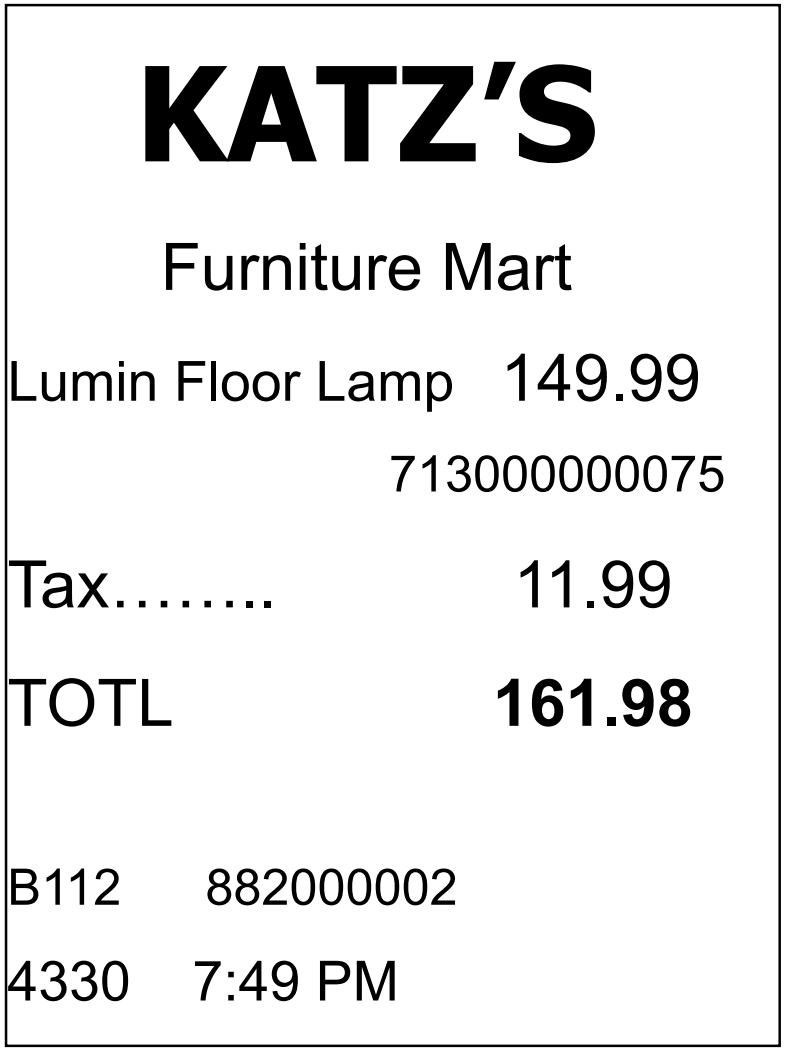
Tax…….. 11.99

TOTL **161.98**

B112 882000002

4330 7:49 PM

Example complicated scenario



KATZ'S

Furniture Mart

Lumin Floor Lamp 149.99

713000000075

Tax…….. 11.99

TOTL **161.98**

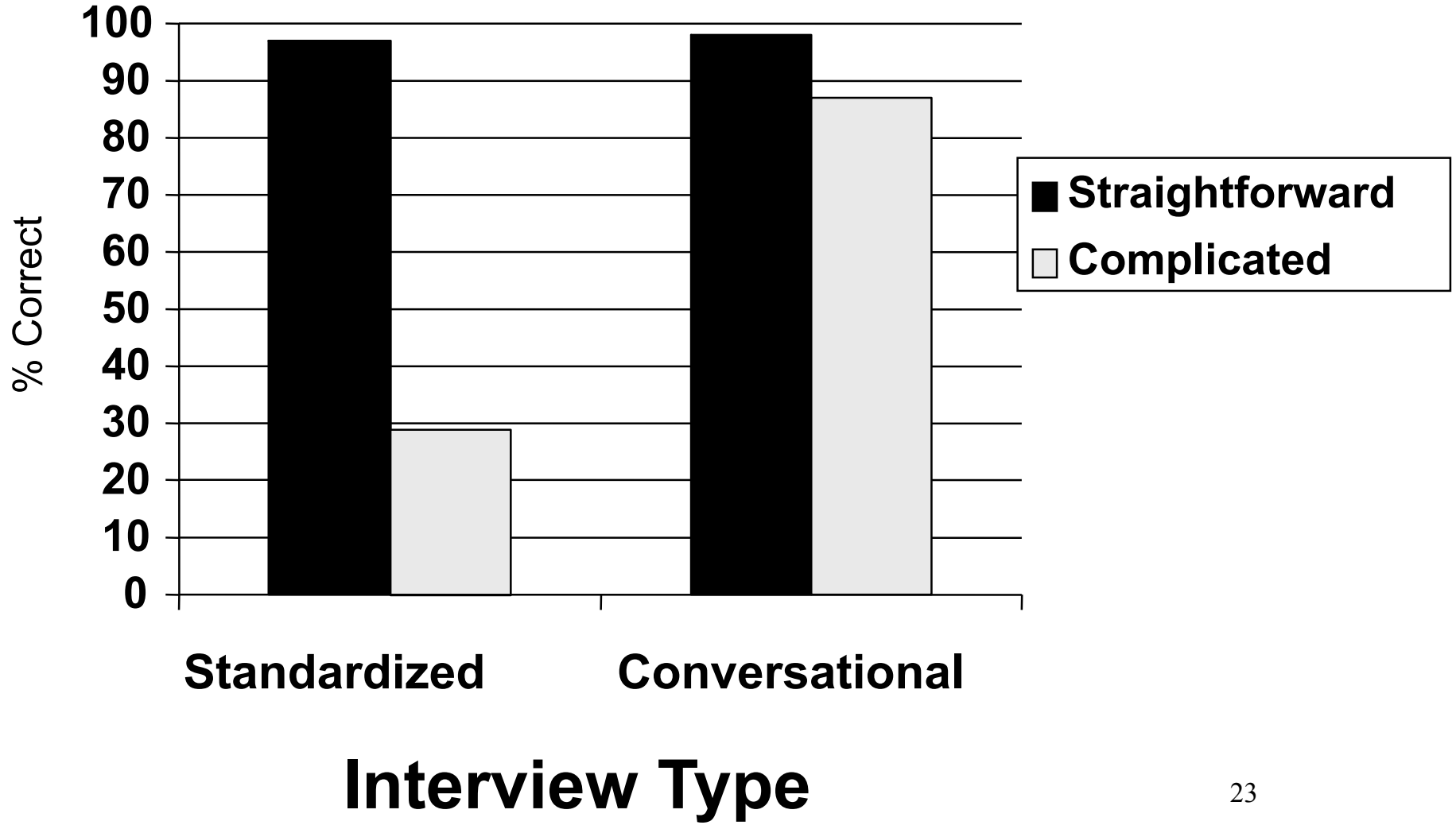
B112 882000002

4330 7:49 PM

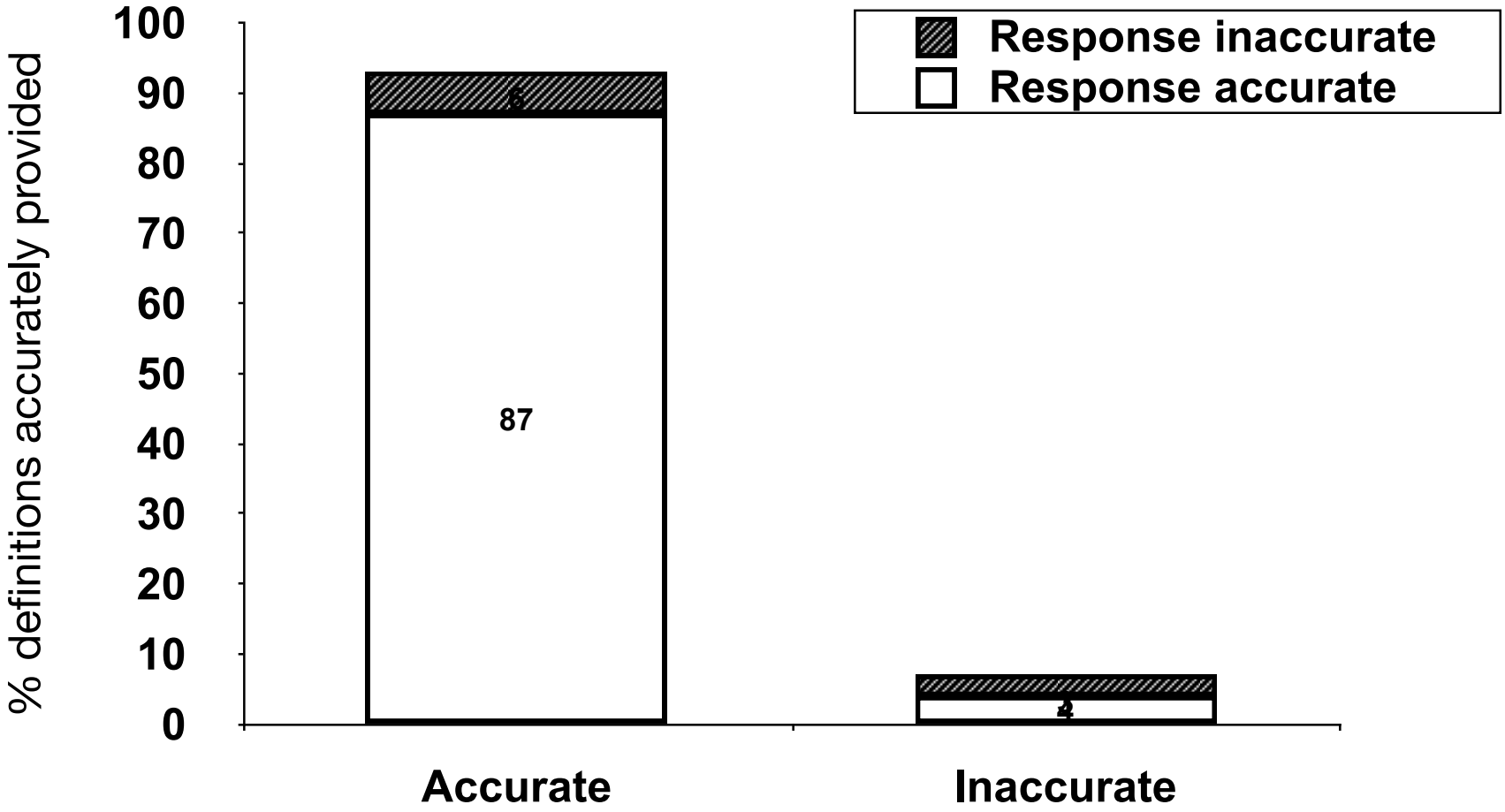
Respondents

- 43 participants, recruited primarily from advertisement in Washington Post
  - 22 assigned to conversational interview condition, 21 to standardized
- Received packets with 12 scenarios before interview and encouraged to study them, but allowed to refer to them during interview
  - 6 led to complicated mappings, 6 to straightforward, counterbalanced across participants

Response Accuracy



Interviewer-provided information



Minutes per Interview

	Standardized	Conversat'al
Median	3.41	11.47
Min	2.48	6.10
Max	5.99	35.44

Speed accuracy trade-off

- Trade-off may not be as extreme (3-fold) in practice

  1. Standardized approach may have underestimated duration: many survey organizations that conduct "standardized" interviews license *I*s to provide discretionary clarification.
  2. Conversational interviews may have overestimated duration: these *I*s not experienced with definitions so often read entire definition instead of providing just the relevant parts.

- But clarification takes time

  - it simply takes more words and more conversational turns to ground meaning than not

Short sequences in Standardized Interviews

I: Has Dana purchased or had expenses for meats and poultry.
R: Yes.
(I goes on to next question)

70% of the sequences in standardized interviews of this form: *I* reads question as worded, *R* provides acceptable answer, *I* moves on to next question; only 10% of sequences had this form in the conversational interviews.

Long sequences in Conversational Interviews

I: Last week did Pat have more than one job, including part-time, evening or weekend work?
R: Um . s- say that again, because *[laughter]*
I: *La-*
R: She has many clients which she . but it's the same kind of job.
I: Okay. U:h *that would-*
R: *In other* words she is um .
I: Well what kind of work *does she do.*
R: *She ba-* she babysits, and she *has*
I: *O-*
R: different clients.
I: Okay, that would be considered as all one job,
R: *All right*
I: *no matter* how many people she- she worked *for.*
R: *Yes* if it's the same type of job, yes, she has one job *and that's all.*
I: *And is this-* this is the only thing that she does. 18
R: Yes, and this is it.
I: Okay, so we'll say no for this. She only has one job?
R: She only has one job.
I: And um . (goes on to next question)

This sort of "directive intervention," in which *I* provides substantive information beyond the question, occurred for 85% of the conversational interviews

Are high levels of response accuracy possible with shorter conversational interviews?

- Schober, Conrad & Fricker (2004)
  - examined variants of conversational interviewing that varied in how and when interviewers provided clarification
    - HOW: *verbatim* or *paraphrased*
    - WHEN: *R-initiated (*when Rs request) or *Mixed Initiative* (when R requests or I thinks it will help)

Scripted

I: Has Kelly purchased or had expenses for household furniture?

R: What do you mean by household furniture?

I: Okay. ah Household furniture eh tables, chairs, footstools, sofas, china cabinets, utility carts, bars, room dividers, bookcases, desks, beds, mattresses, box springs, chests of drawers, night tables, wardrobes, and unfinished furniture. Do not include TV, radio, and other sound equipment, lamps and lighting fixtures, outdoor furniture, infants' furniture, or appliances.

R: Ask the question again *then*.

Paraphrased

I: Has Kelly purchased or had expenses for household furniture.

R: Uh- does a lamp is a lamp considered household furniture.

I: Okay no lamps are not considered to be household uh furniture according to our definition.

R: Okay then no.

I: No? Okay

Respondent-initiated

I: Has Kelly purchased or had expenses for household furniture?

R: What do you mean by household furniture?

I: Okay. ah Household furniture eh tables, chairs, footstools, sofas, china cabinets, utility carts, bars, room dividers, bookcases, desks, beds, mattresses, box springs, chests of drawers, night tables, wardrobes, and unfinished furniture. Do not include TV, radio, and other sound equipment, lamps and lighting fixtures, outdoor furniture, infants' furniture, or appliances.

R: Ask the question again *then*.

Mixed-initiative

I: How many hours per week does Mindy usually work at her job?

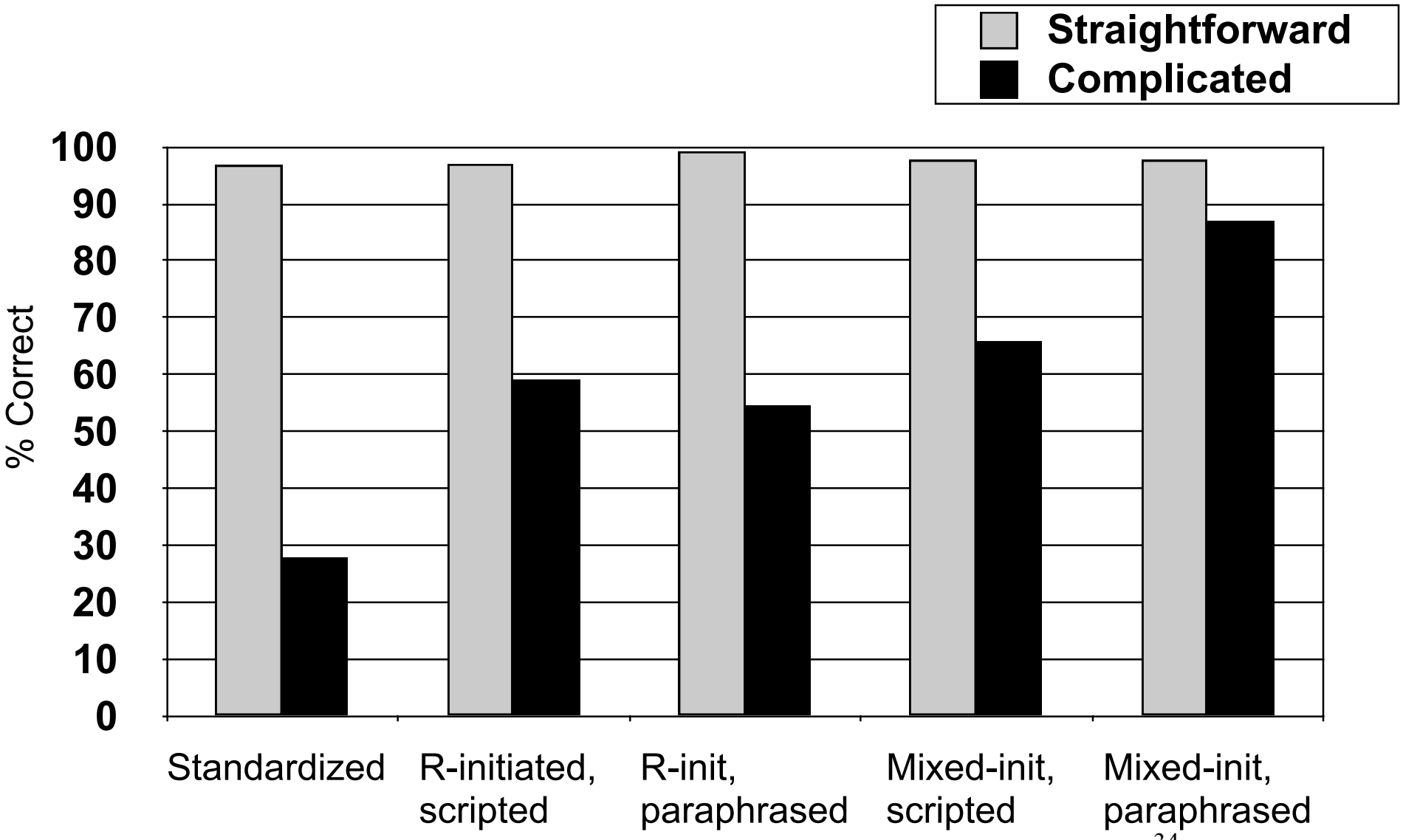
R: (long pause)

I: And by usually I mean fifty percent of the time or more, or the most frequent schedule during the past four or five months.

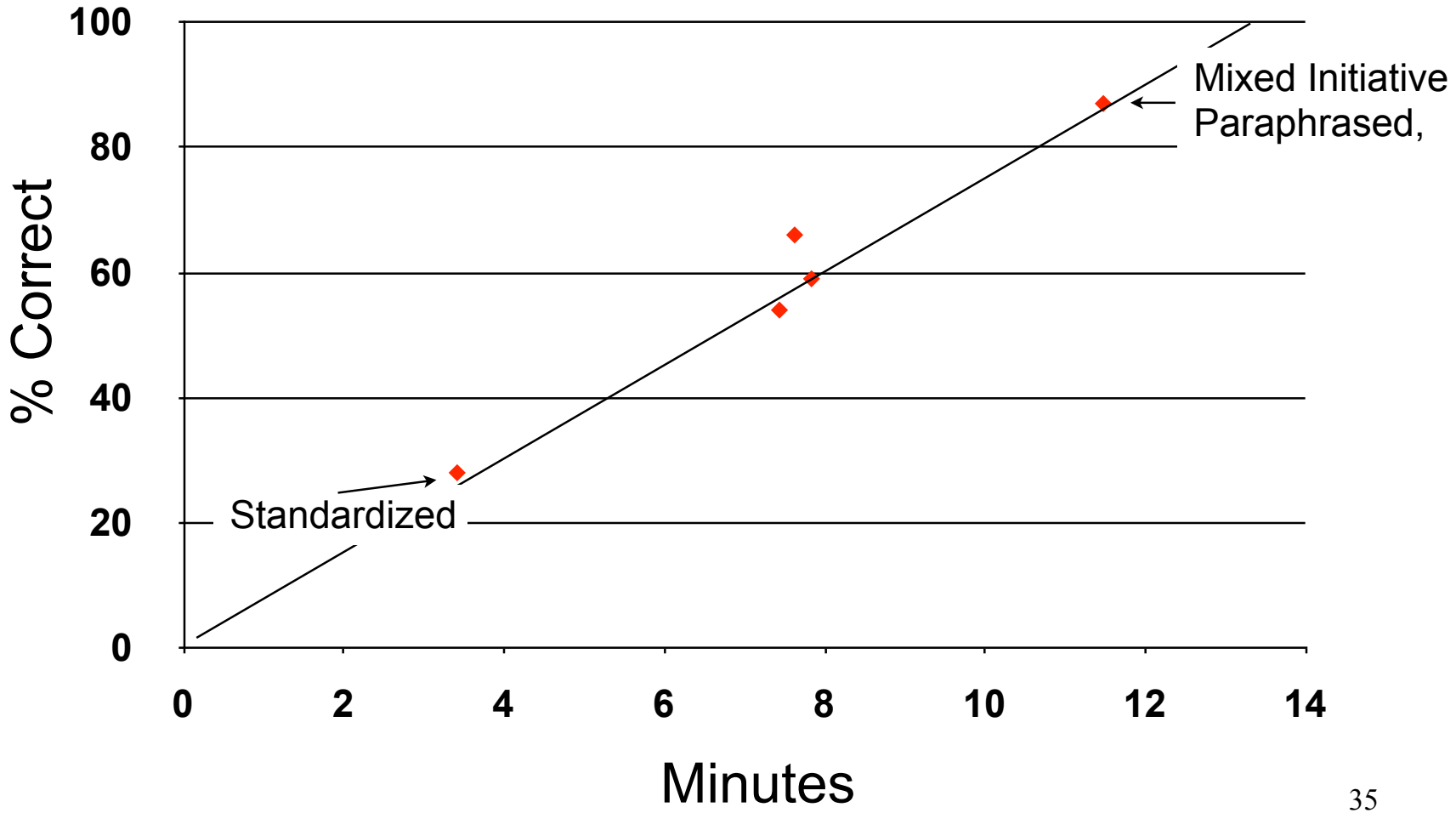
R: Uh, fifty hours.

I: Okay.

Response accuracy



Accuracy x duration
(r = .98)



Implications

- *Some* benefits of conversational interviewing can be gained at lower cost
  - current practice varies, but may be a reasonable trade-off
  - benefit depends on frequency of complicated mappings
- Relying on *R*'s to know when they need help may be insufficient
- In the end, mapping ambiguities are resolved with clarification; the more ambiguities the more clarification helps and the longer the interviews last

Frequency of complicated mappings

- Meaning-in-words assumption won't hold
  - if complicated mappings turn out to be frequent
  - if frequency is unknown but researcher must be confident questions understood as intended
- In lab studies described so far, 50% of questions lead to complicated mappings

Field experiment
(Conrad & Schober, 2000)

- Comparison of understanding and response accuracy in standardized and conversational interviews
- Hard to measure understanding in real survey settings
  - without access to respondents' circumstances, can't tell if responses fit official definitions
  - can't directly measure response accuracy
    - record checks and diaries are expensive
    - may not be accurate themselves

Measuring comprehension

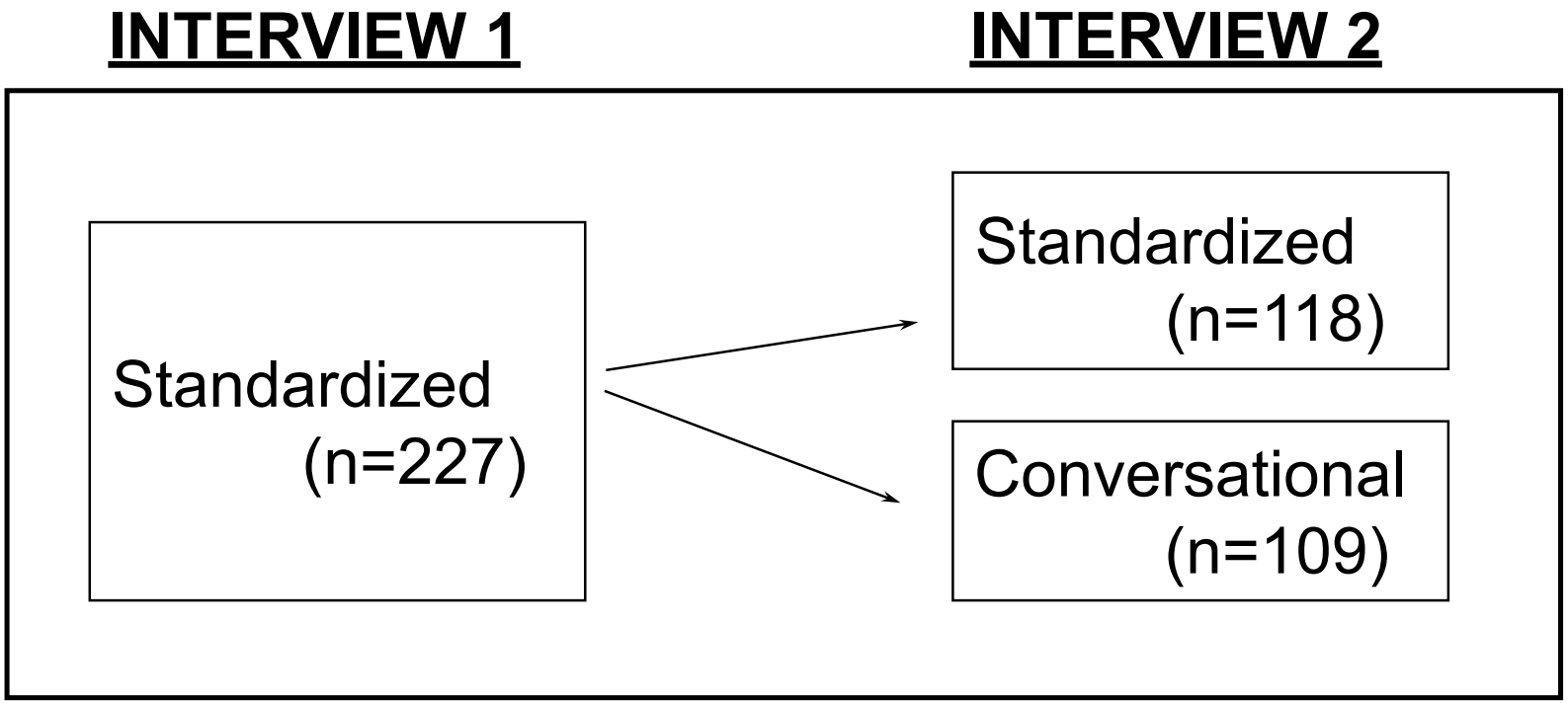
- We *can* determine if conversational interviewing *changes* Rs' understanding

- Logic:
  - if conversational interviewing improves comprehension
  - then Rs in standardized interviews should change their responses in a subsequent conversational interview
  - more than they would in a subsequent standardized interview

Field Experiment (con't)

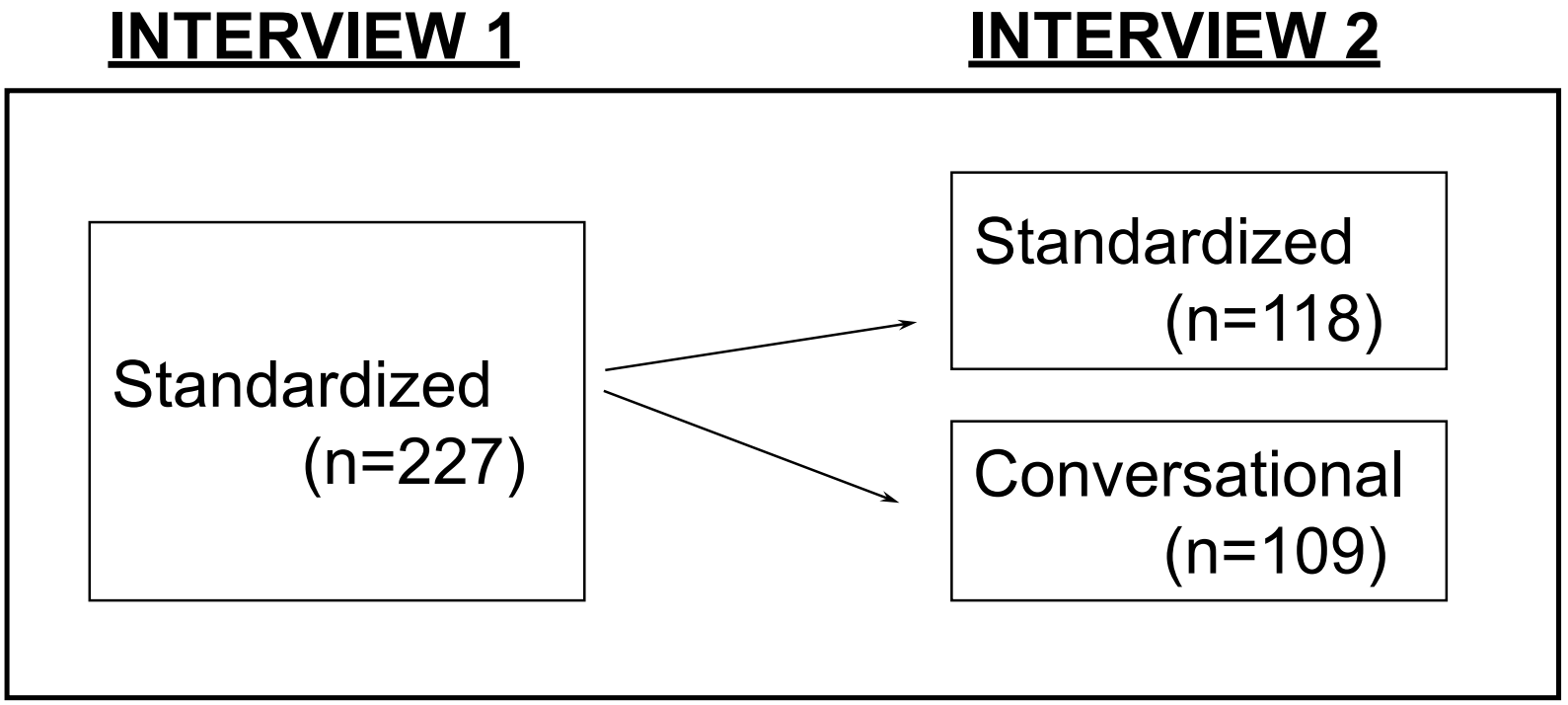
- Respondents (n=227) were interviewed at home by experienced Westat telephone interviewers (n=20)
- Nationally representative sample of residential households with telephones in Continental U.S.
- Asked 5 questions about housing (numerical) and 5 about purchases ("yes"/"no")

- Each *R* interviewed twice by different *I*'s asking same questions
- Standardized and conversational interviews conducted by different I's



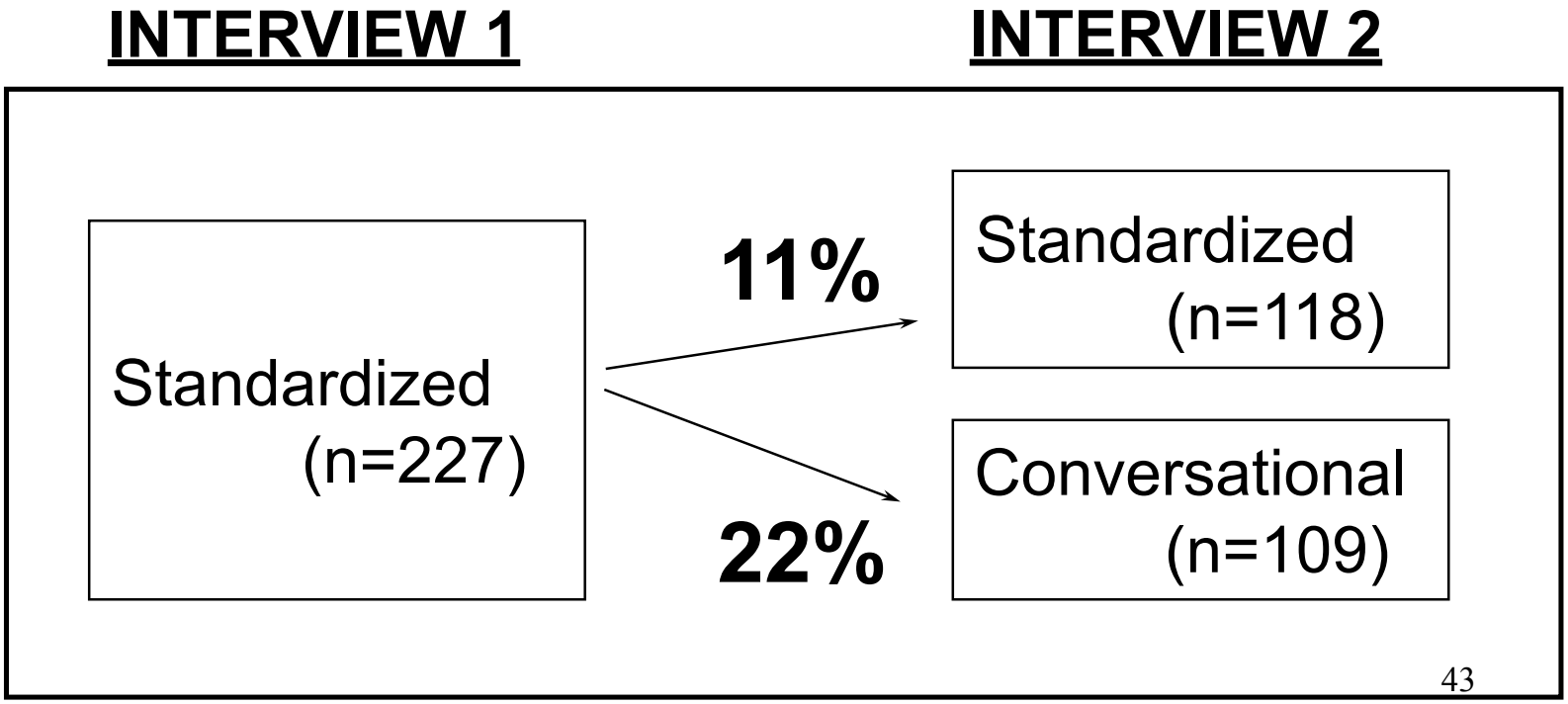
If conversational interviewing improves comprehension:

- Responses should change more from Interview 1 to Conversational Interview 2 than to Standardized Interview 2

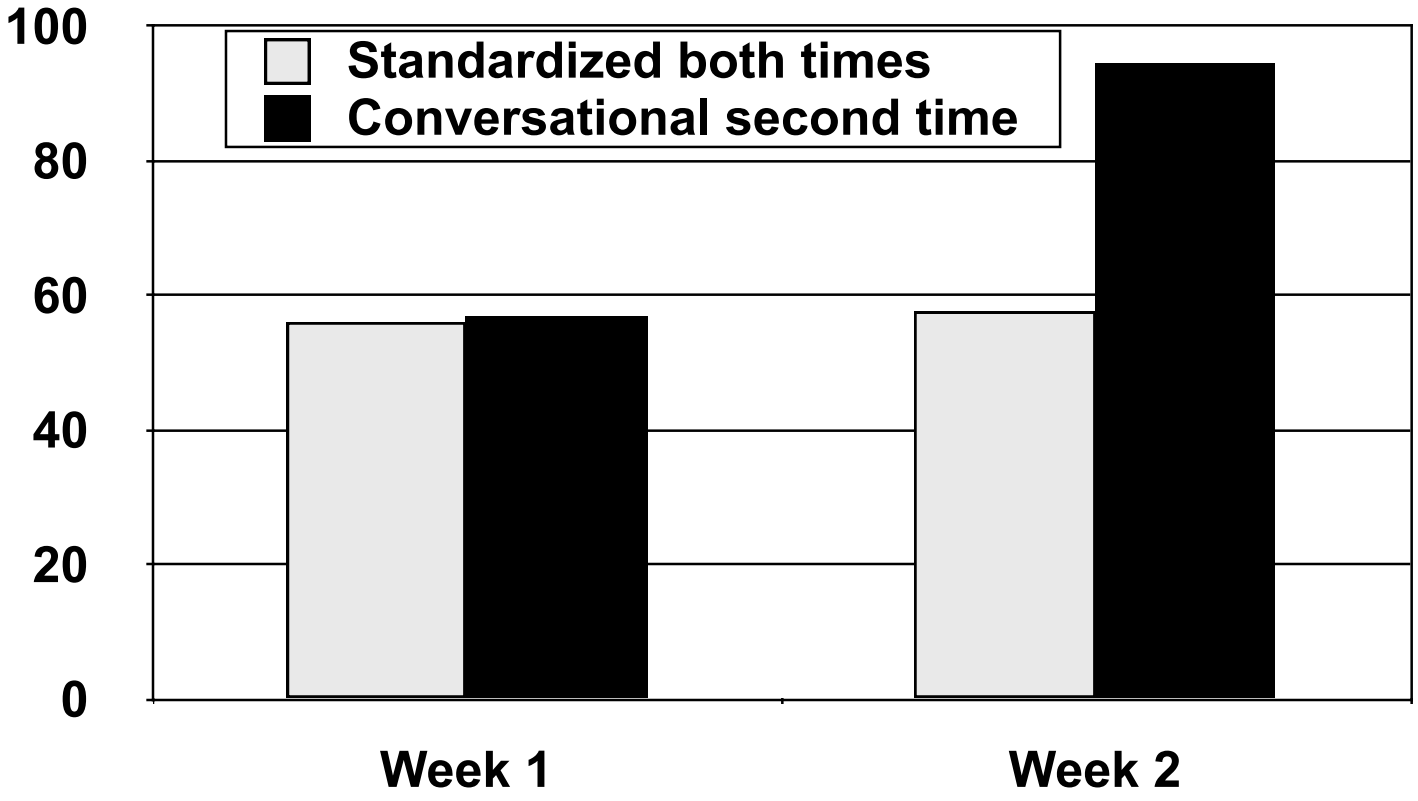


Results: Response Change

- More responses changed when second interview was conversational than standardized
- suggests 11% mappings complicated for these questions in this sample



Results:
% “Legal” listed purchases



Median interview duration

Interview 1 (Standardized)	Interview 2	
5 mins	5 mins	Standardized
	9 mins	Conversational

When people are uncertain about what conversational partners mean

- they don't always explicitly ask for clarification but, instead, indicate their uncertainty implicitly
  - Wait to hear more
  - Pretend all is well
  - Hedge and stammer
- Do people exhibit implicit cues of comprehension difficulty that conversational interviewers may treat as "requests" for help?

I: How many hours per week does Mindy usually work at her job.

R: . Forty-five?

Mode x Interviewing Technique

Conrad, Schober & Dijkstra, 2007; Schober, Conrad, Dijkstra & Ongena, under review

- 2 Interviewing Techniques:
  - Strictly Standardized
  - Conversational
- 2 Modes
  - Face to Face (FTF)
  - Telephone
- Post-interview Q'aire with definitions
  - Interview-Q'aire Δ measures response accuracy

	Stand.	Conv.
FTF	11	10
Telephone	11	10

Number of Interviews

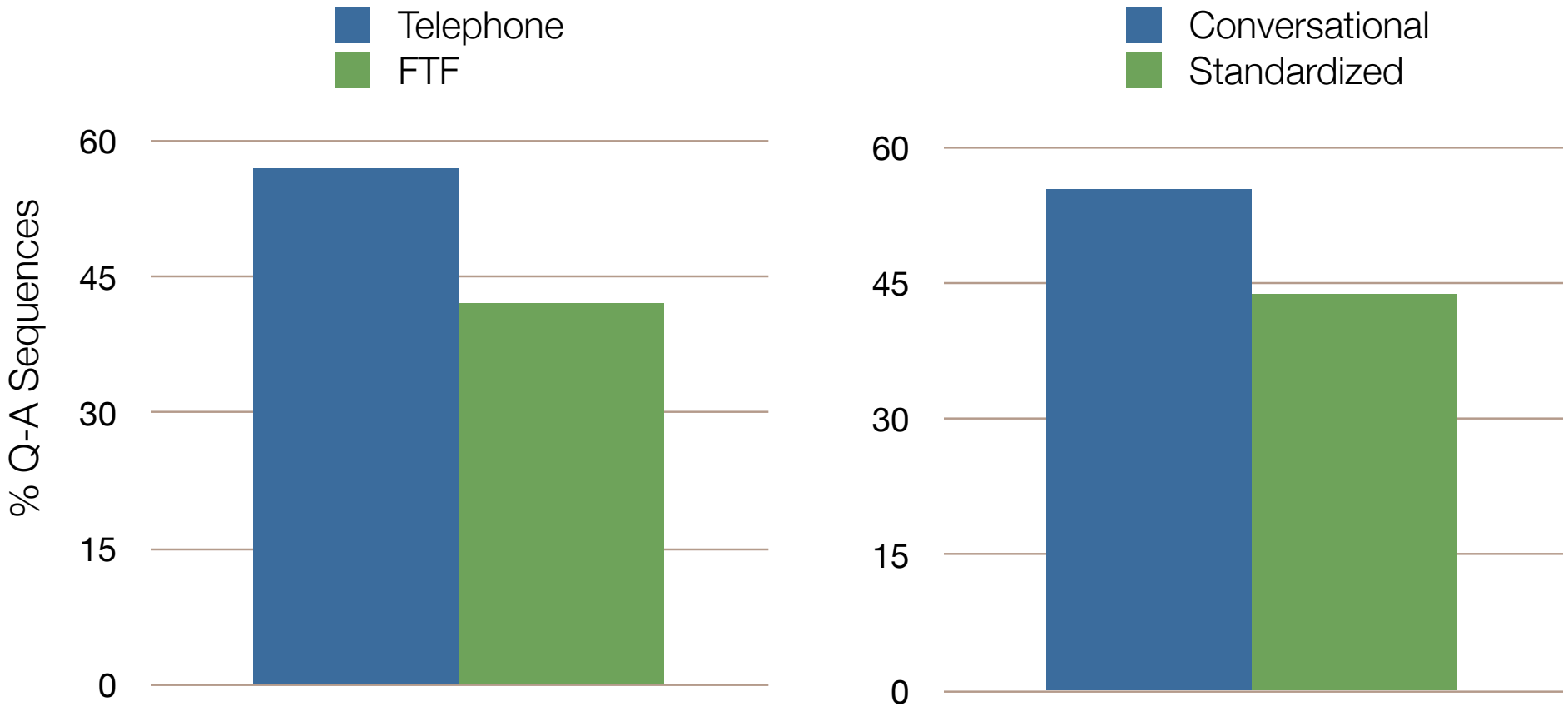
Survey Questions

- 18 questions:
  - Factual questions on student status, membership in clubs, employment history
  - Opinion questions on asylum seekers and illegal aliens

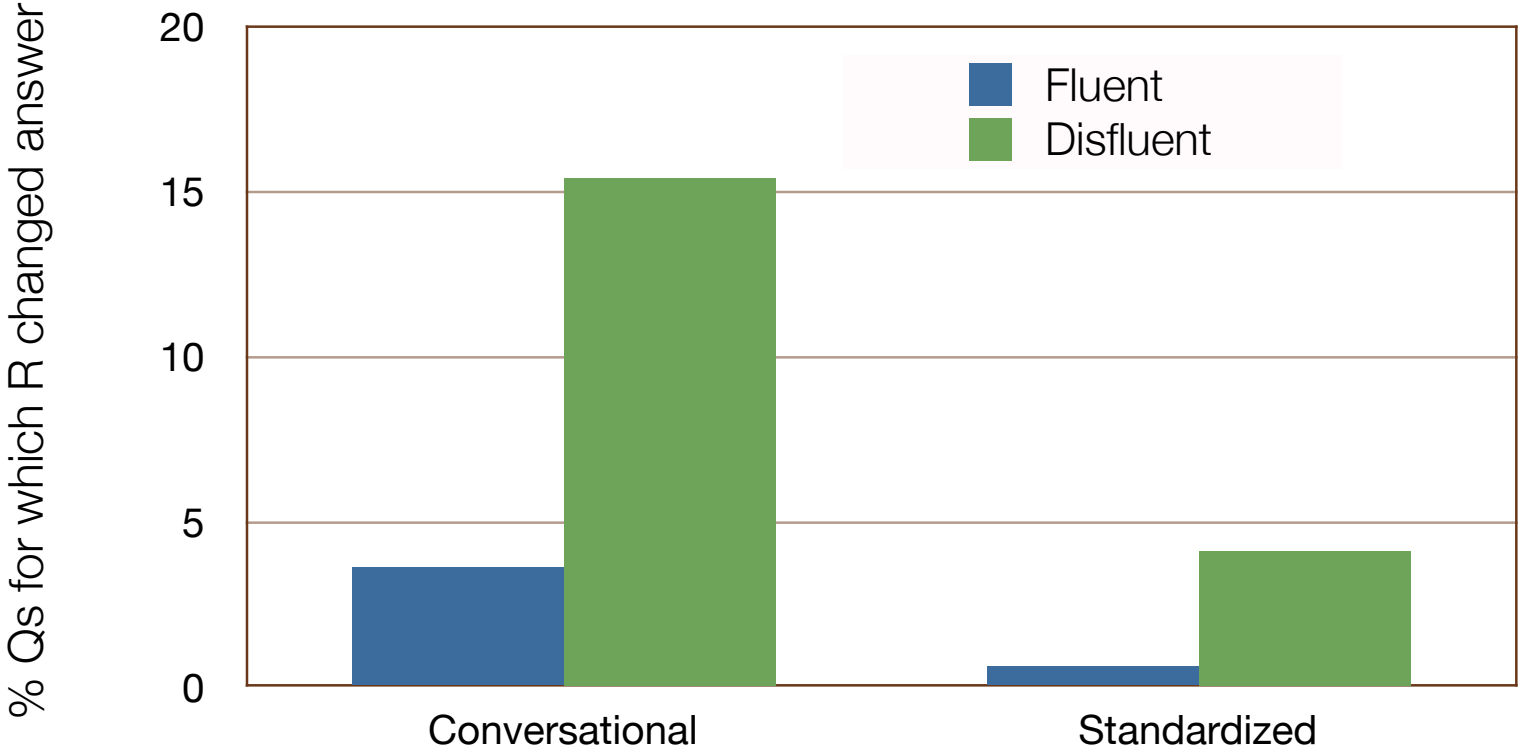
Telephone Interview

Face-to-Face Interview

Disfluencies (*um*s and *uh*s, pauses, repairs, restarts) in answer

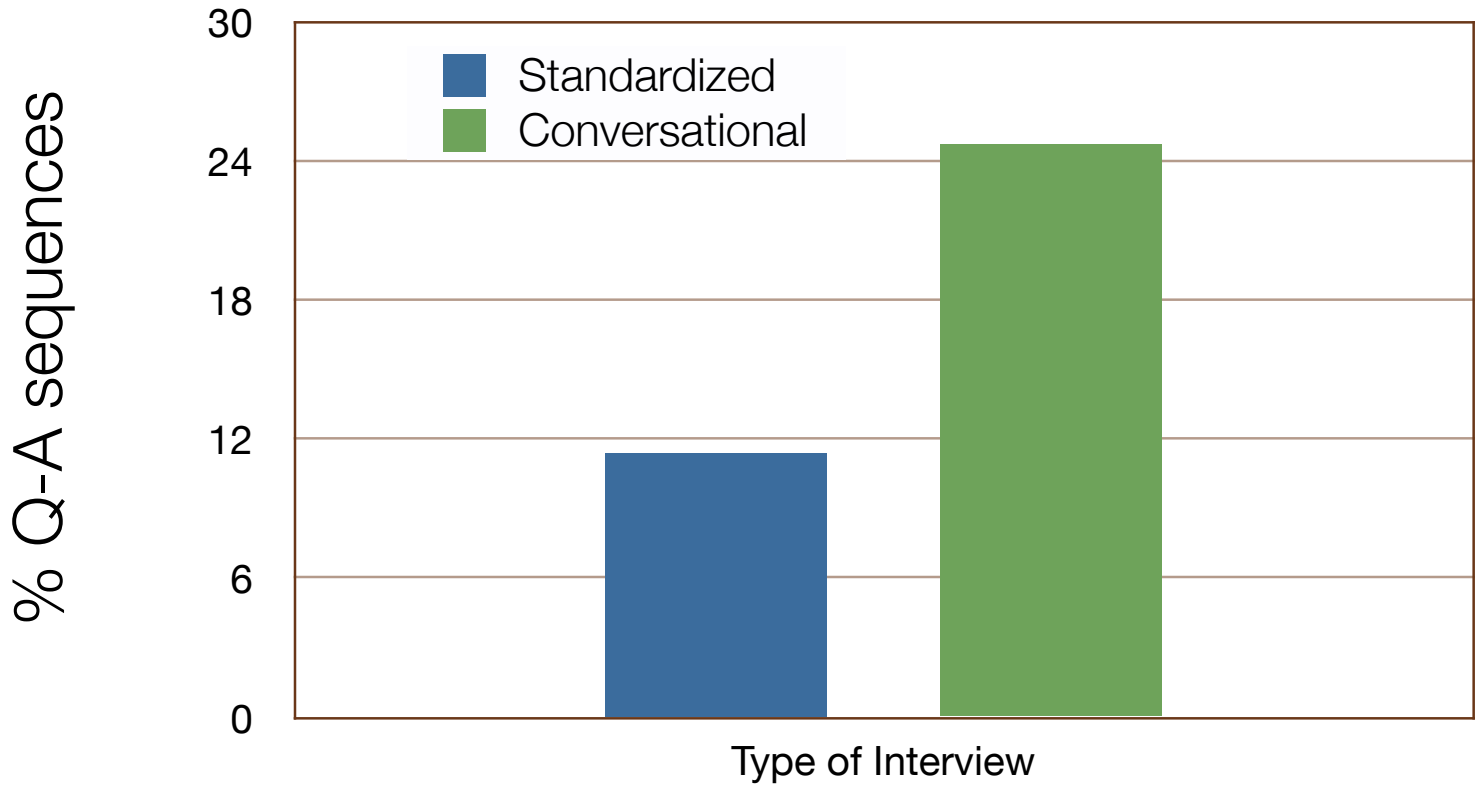


Disfluency and Reliability of Answers



- overall, Rs changed answers more when disfluent
- marginally more for conversational than standardized interview (*p.* = .073)

Gaze Aversion

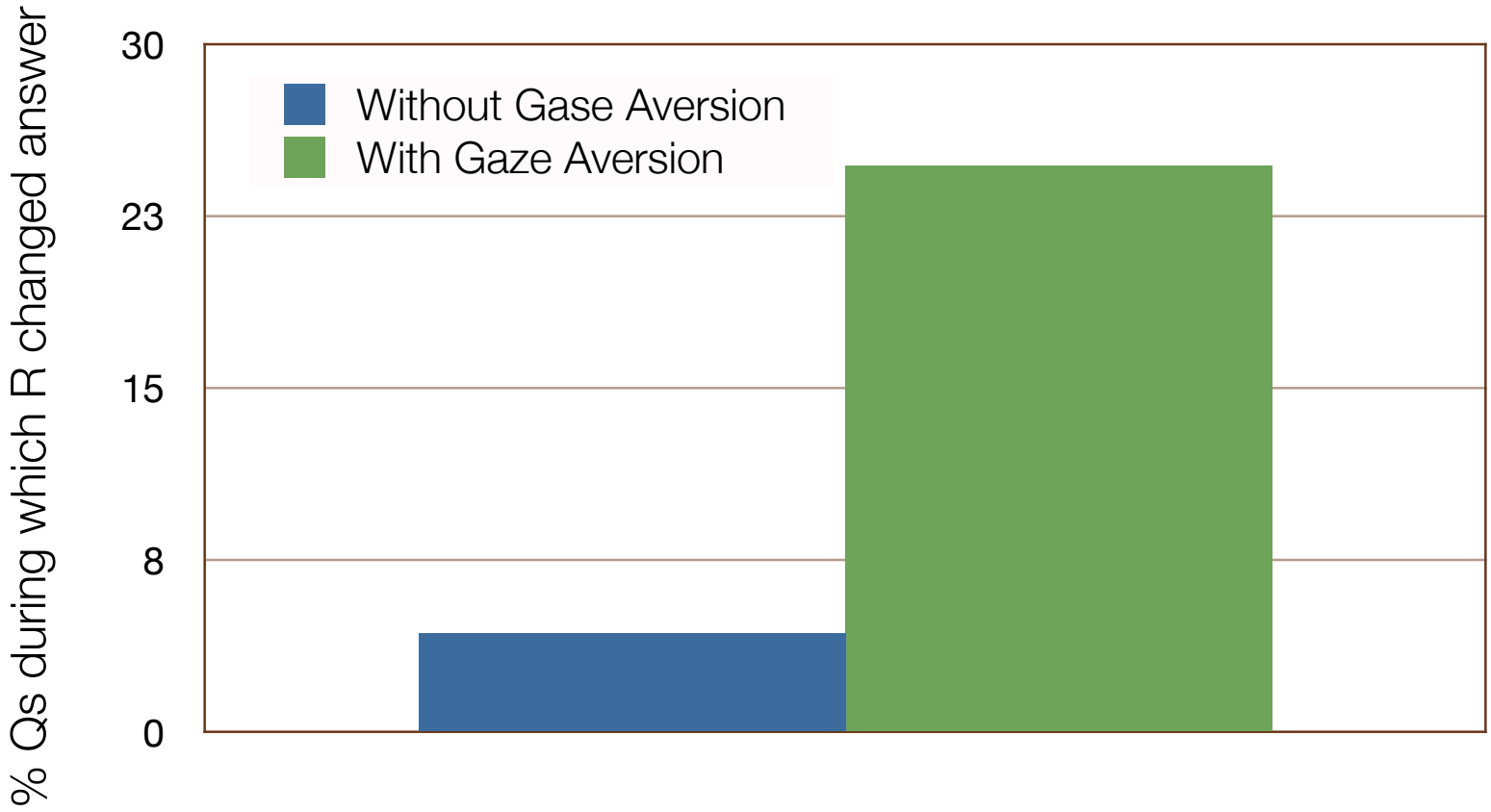


- More gaze aversion in Conversational than Standardized interviews
- Gaze aversion seems to be affected by *I*s ability to act on it

Standardized: Shorter looking away

Conversational: Longer looking away

Gaze Aversion and Reliability of Answers



- overall, Rs changed answers more when disfluent
- gaze aversion equally diagnostic in both interviewing techniques

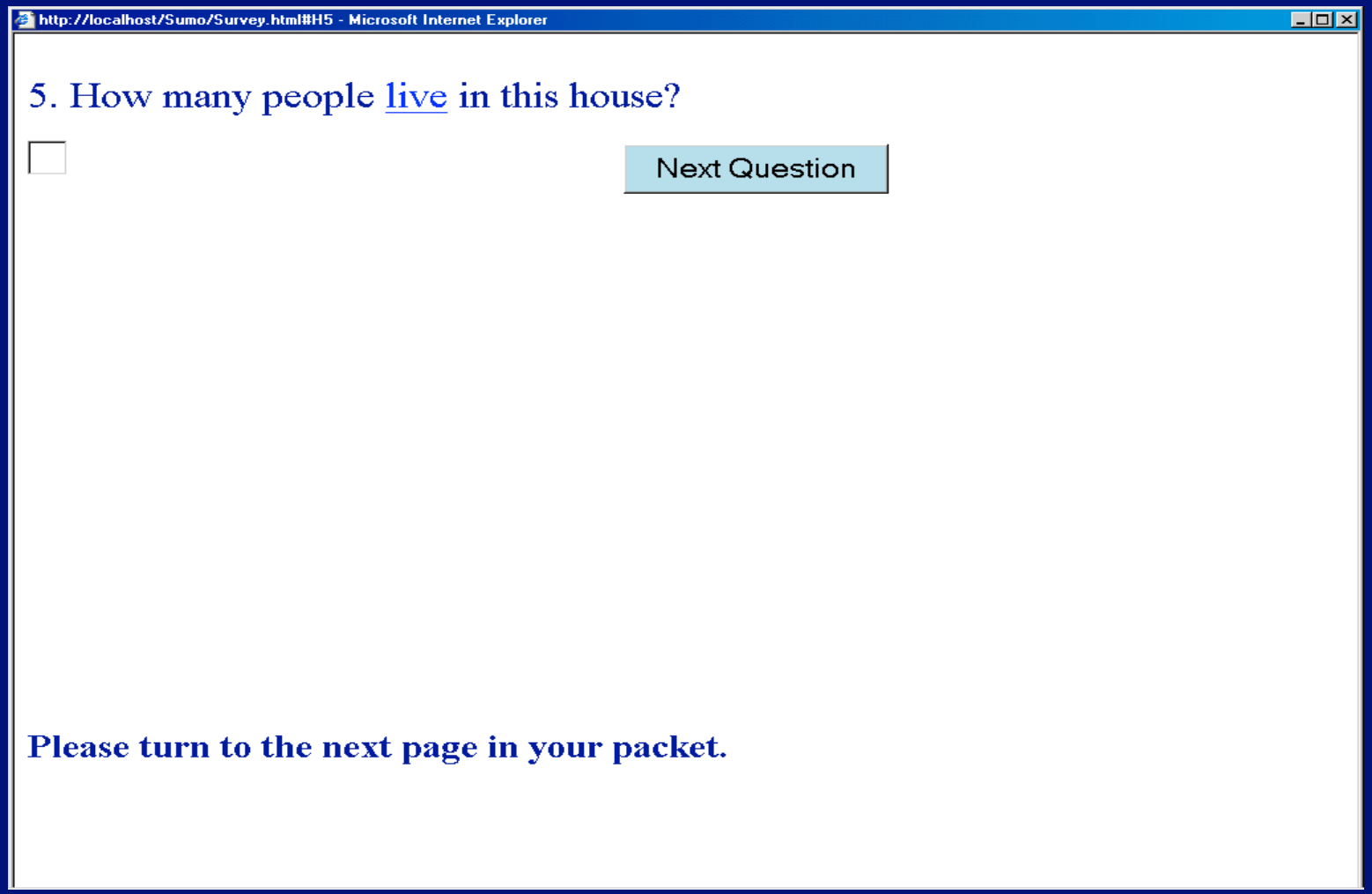
Implications

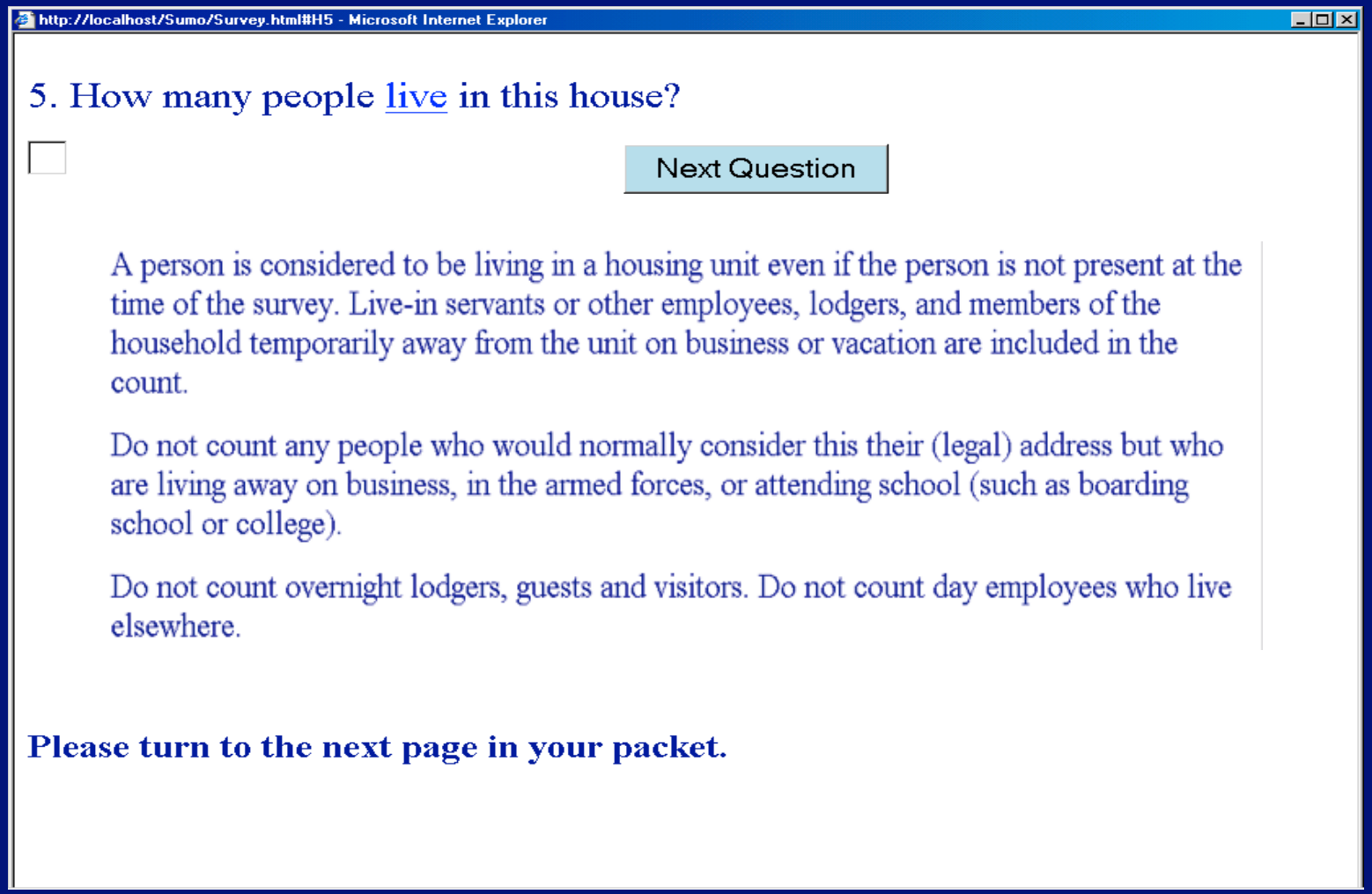
- Telephone Rs substitute audio for visual cues
- Cues differentially displayed depending on *I*'s ability to react to them
  - signal vs. symptom
- *I*s can potentially be trained to offer clarification when *R*s exhibit verbal or visual signs of uncertainty

Clarifying Question Meaning in Web Surveys

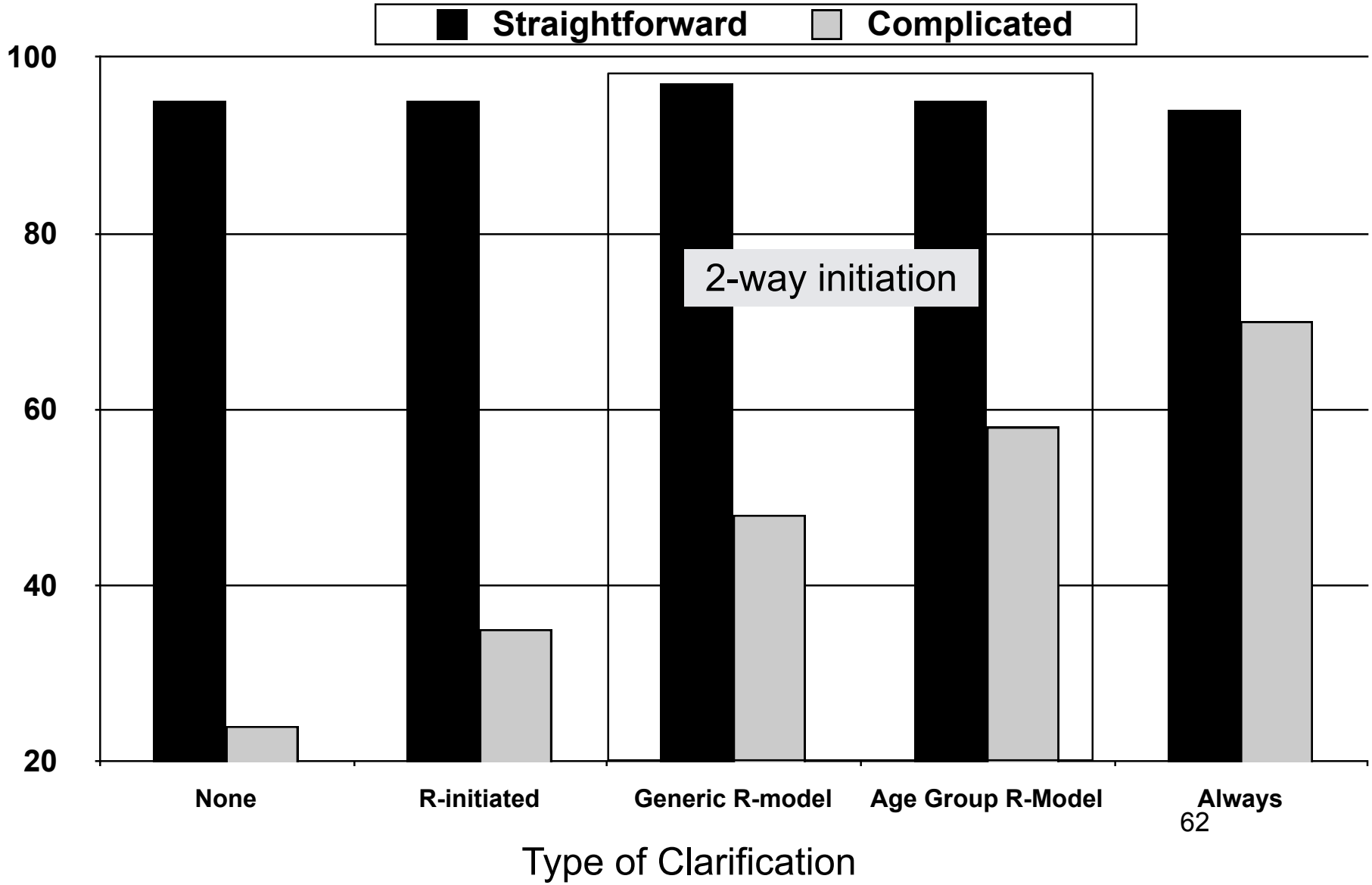
Conrad, Schober & Coiner, 2007

- Clarification provided when
  1. *R* requests it by clicking on relevant question text
  2. displays evidence of comprehension difficulty, specifically inactivity
- Compare different inactivity thresholds
  - generic model: same for everyone
  - group-based model: longer for slower, i.e., older, Rs so we can distinguish inactivity from ordinary, slow thinking
- *R*s answered on basis of scenarios as in previous studies
  - half were *complicated,* half were *straightforward*
  - 10 Rs per age x clarification group = 100 Rs


http://localhost/Sumo/Survey.html#H5 - Microsoft Internet Explorer
5. How many people live in this house?
Next Question
Please turn to the next page in your packet.


http://localhost/Sumo/Survey.html#H5 - Microsoft Internet Explorer
5. How many people live in this house?
Next Question
A person is considered to be living in a housing unit even if the person is not present at the time of the survey. Live-in servants or other employees, lodgers, and members of the household temporarily away from the unit on business or vacation are included in the count.
Do not count any people who would normally consider this their (legal) address but who are living away on business, in the armed forces, or attending school (such as boarding school or college).
Do not count overnight lodgers, guests and visitors. Do not count day employees who live elsewhere.
Please turn to the next page in your packet.

Clarification and % Accuracy



Mixed initiative clarification

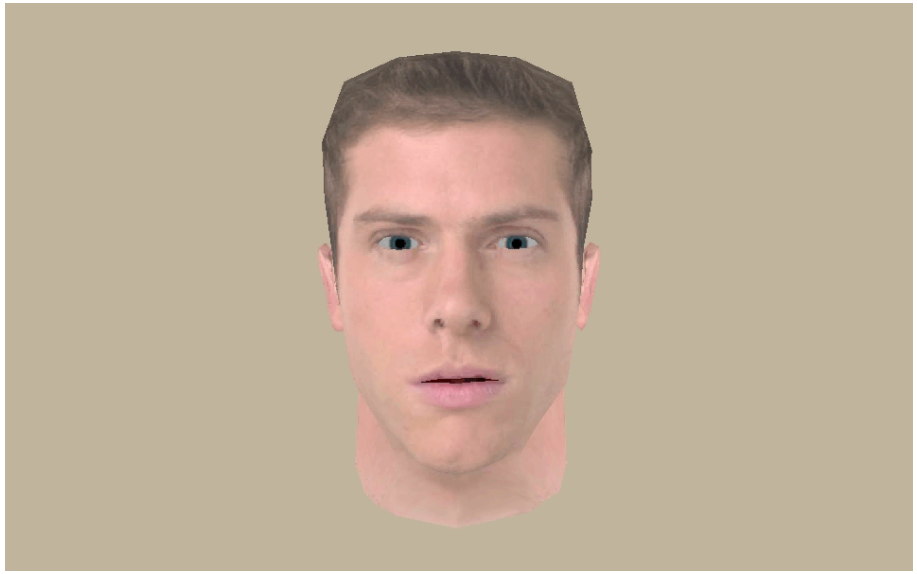
- 2-way initiative of giving clarification leads to greater improvement than when just R-initiated
  - especially when triggering behavior is tailored to different groups
  - tailoring to individuals could help even more
- Possible, in principle, to embody detection of need-for-clarification and clarification-giving behavior in a virtual interviewer

Dialog Capability vs. Visual Realism

(Conrad, Schober, Jans, Nielsen, Orlowski, Levenstein, 2008, in prep.)

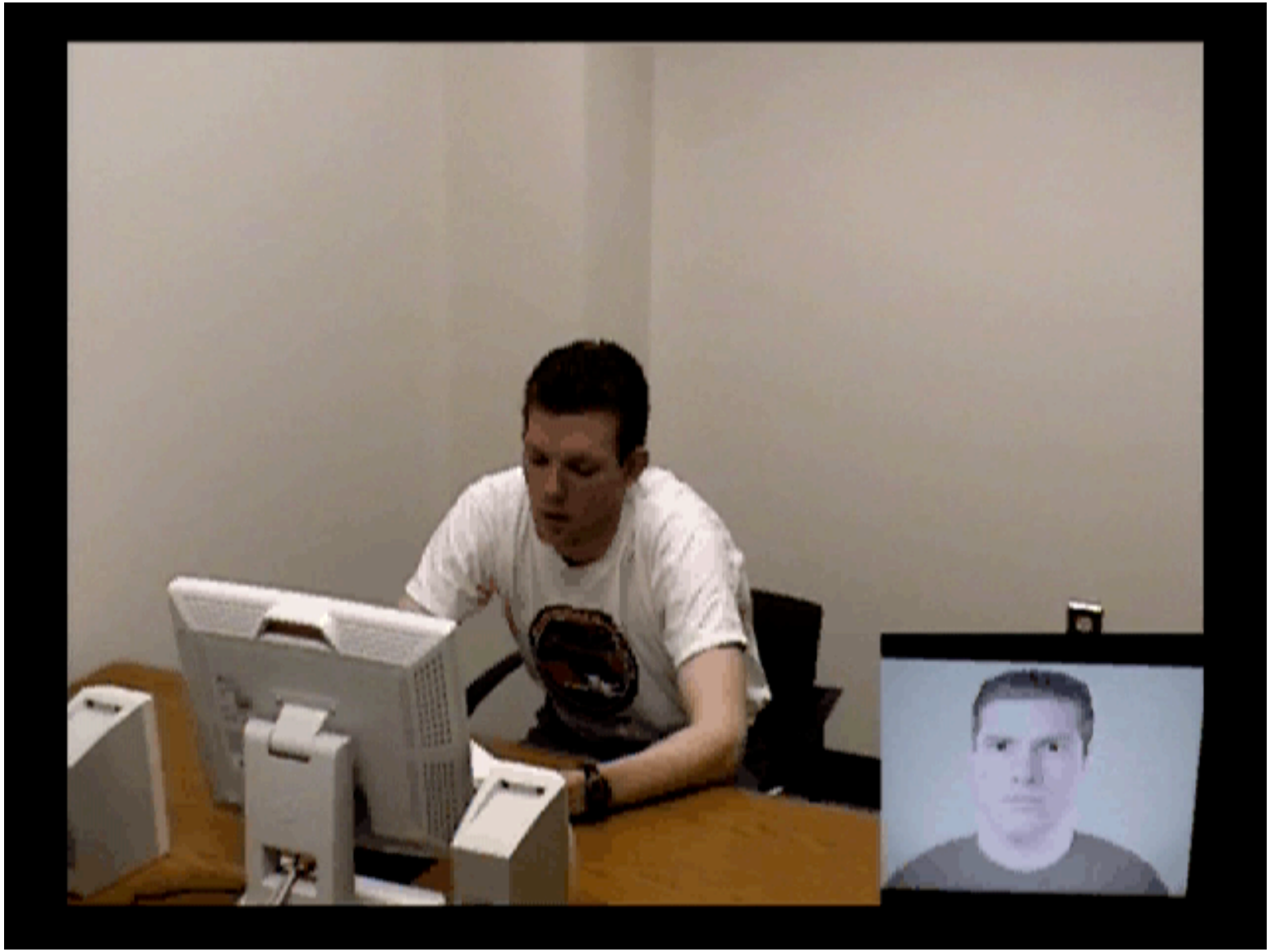
- Lab experiment
  - Questions administered by one of four virtual interviewers
    - vary in visual realism (amount of facial movement)
    - dialog capability (degree to which virtual interviewer understands *R*s' requests and indirect evidence they need clarification)
  - *R*s answer by speaking
  - *R*s base answers on fictional scenarios allowing determination of response accuracy

Low Visual Realism

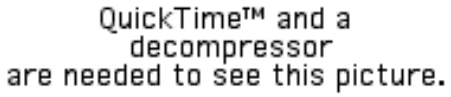


High Visual Realism

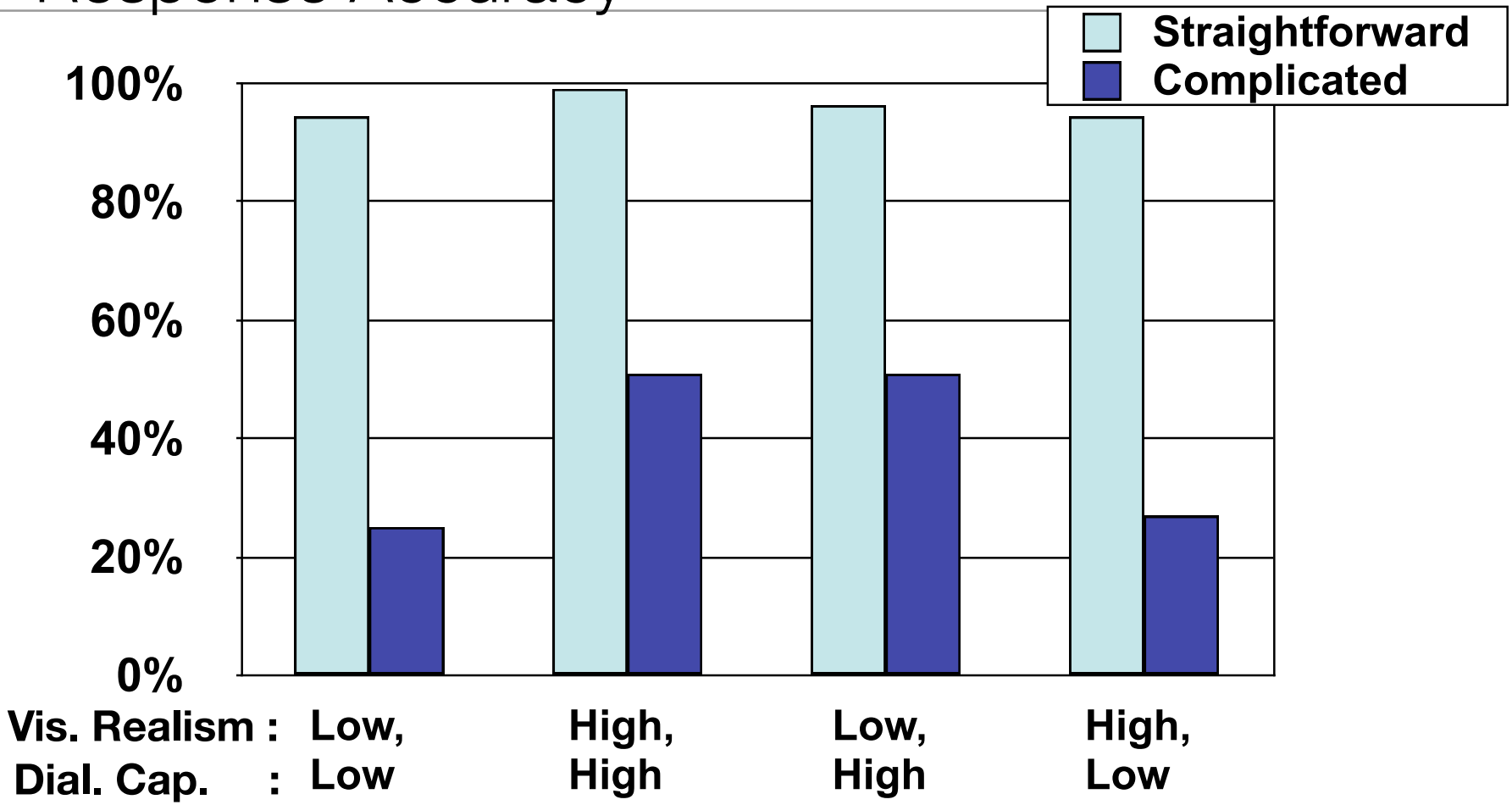
Low Dialog Capability



High Dialog Capability

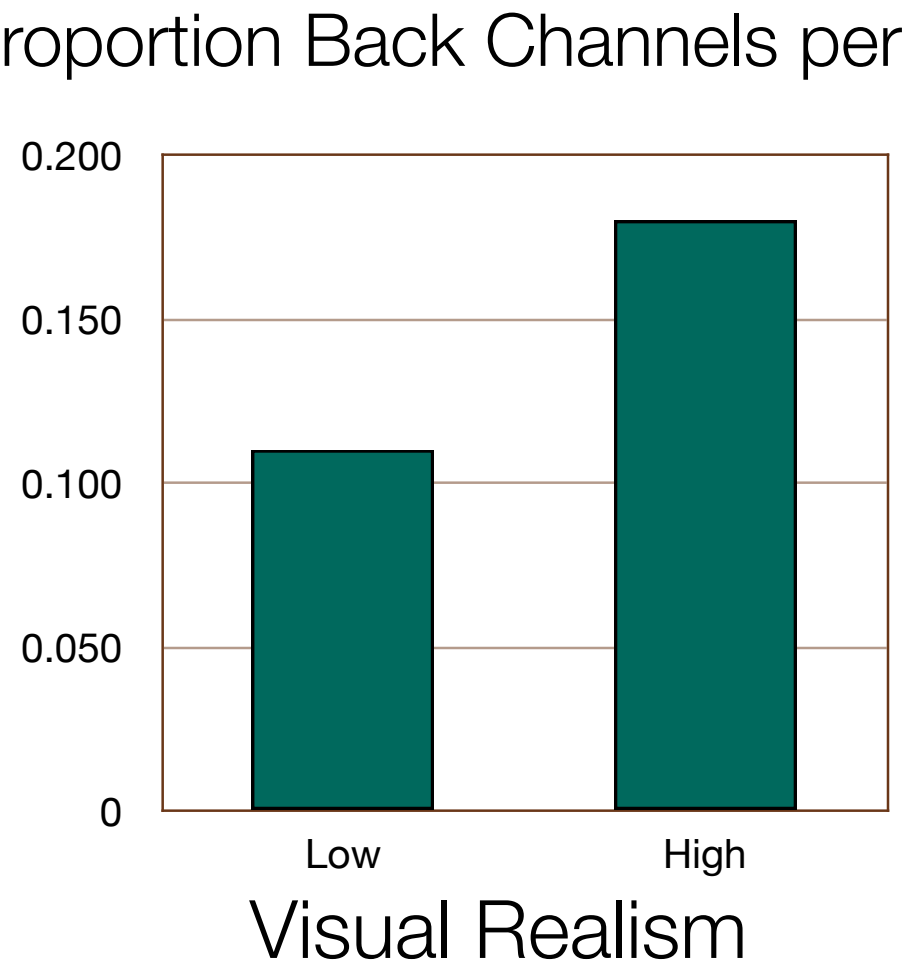


Response Accuracy

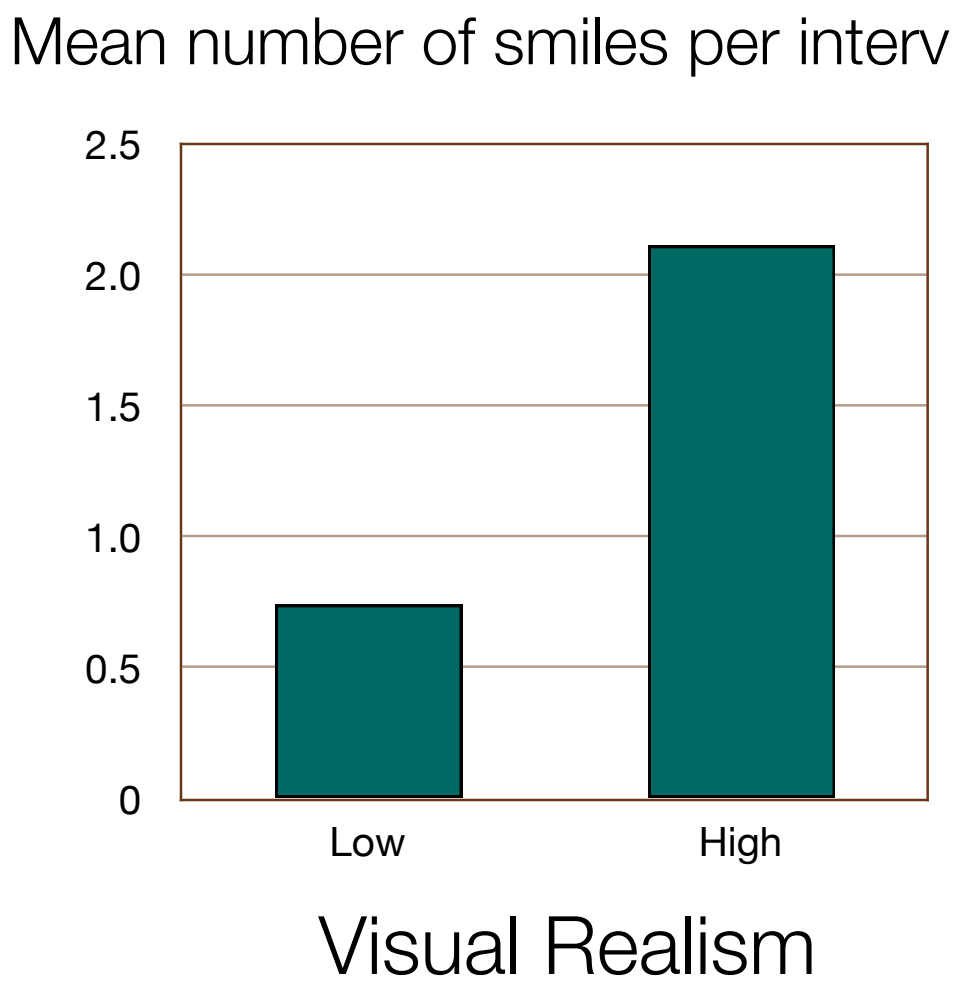


n = 18 or 19 Rs per condition

Visual Realism and Back Channels



Visual Realism and Smiles



- Smiles also last longer with High Visual Realism

What do we now know about Conversational Interviewing?

- Can improve accuracy of answers when mapping between questions and R's circumstances is complicated
- but clarification takes time
- in proportion to frequency of complicated mappings
- Researchers must decide whether it's more valuable to assure accurate comprehension or to complete more cases for a fixed investment
  - measurement error vs. sampling error
- May be possible to gain some of the benefits of this approach in automated systems

Thank You!