Data Quality Concerns in Scientific Tasks

Y. Patrick Hsieh
Stephanie Eckman
Herschel Sanders
Amanda Smith

RTI International is a registered trademark and a trade name of Research Triangle Institute.

Use of Crowdsourcing

- Crowdsourcing popular source of online workforce for scientific research
  - Classifying images
  - Transcribing audio files
  - Coding texts or social media content

These tasks are a lot like surveys

- Fast & inexpensive
- Amazon Mechanical Turk (MTurk)

What about Data Quality?

Crowdsourcing vs Panels

MTurk

- Paid per HIT
- Metrics available
  - # of tasks completed
  - % of tasks approved
- Strong norm:
  - Quality work → fair pay

Online Panel

- Paid per survey
- Few quality metrics available

Do cultures & incentives lead to **data quality differences?**

- In surveys?
- In scientific tasks?

Motivated misreporting

Research Question

- Web survey design

Format	MTurk	Online Panel
Grouped	Filter Filter Filter Follow Up Follow Up Follow Up Follow Up	
Interleafed		

2 tasks:

- Survey
- Image coding

2 Sources of Participants

- **MTurk**
  - 80% prior approval rate
  - In US
- Survey:
  - 185/214 completed
  - 59% female
  - 39 years old
  - 48% >= bachelors
- Image coding:
  - 141/342 completed
  - 62% female
  - 50% bachelors or higher

- **Online panel**
  - Convenience sample in US
  - Balanced to Census
- Survey:
  - 204/260 completed
  - 53% female
  - 48 years old
  - 37% >= bachelors
- Image coding:
  - 141/372 completed
  - 60% female
  - 45% bachelors or higher

Task A: Lifestyle Survey

- 4 filter sections
  - Clothing
  - Consumer goods
  - Leisure activity
  - Credit cards

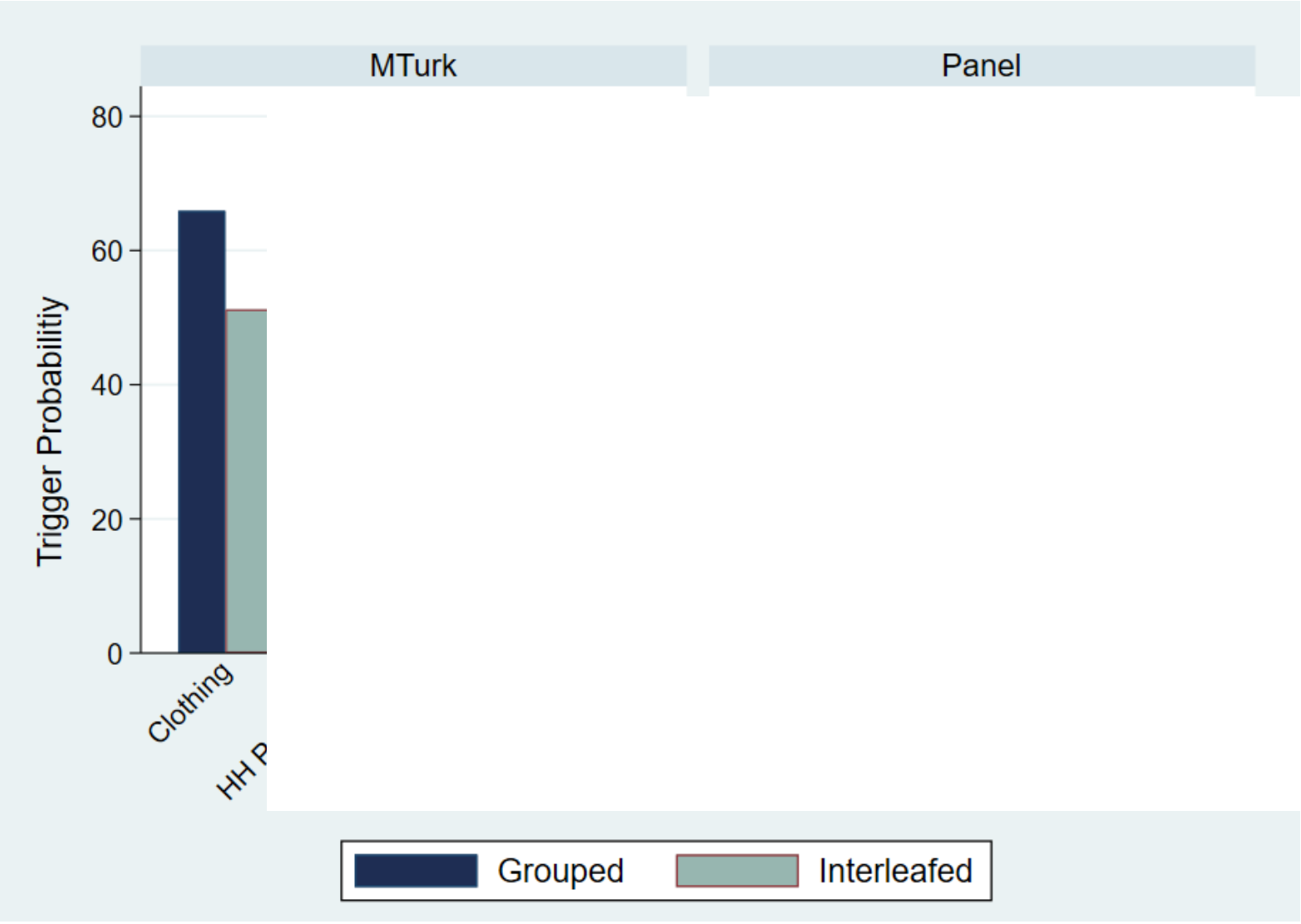
- 30 minutes
- $4 incentive
- Order of sections randomized
- Filters in forward or backward order

Task B: Image Coding

- Image coding task
  - 40 photos of Haiti buildings
  - $6 incentive
  - 50 minutes
- 4 elements
  - Beam
  - Column
  - Slab
  - Wall
- 2 filters
  - Can you see element?
  - Is it damaged?

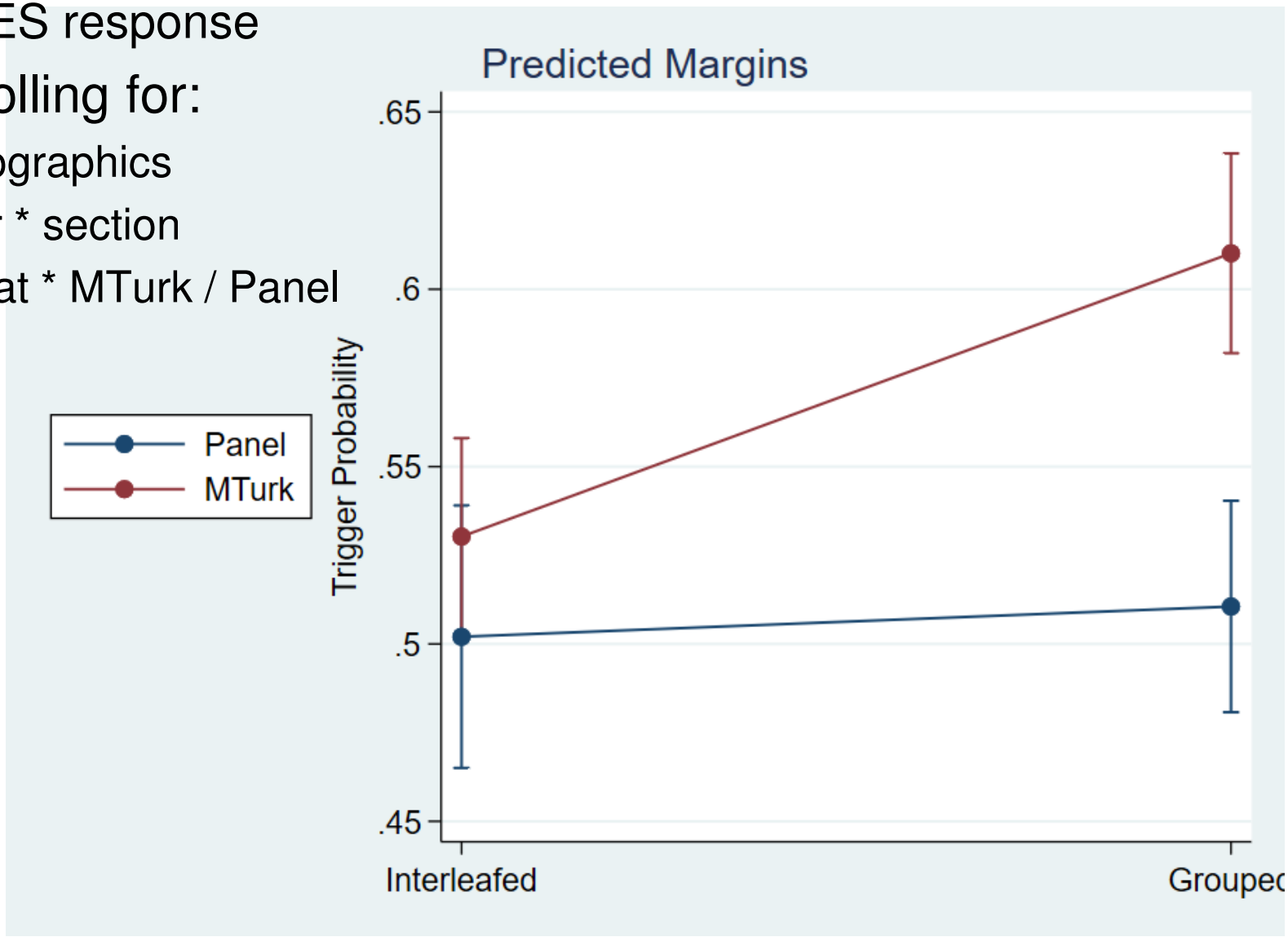
Results: Motivated Misreporting in Survey Questions

- Expected format effect: more YES answers in GROUPED format



Results: Motivated Misreporting in Survey Questions

- DV: YES response
- Controlling for:
  - Demographics
  - Order * section
  - Format * MTurk / Panel



Results: Motivated Misreporting in Image Coding

- Effect in opposite direction: More YES in Interleafed

Average # of YES responses		
	Element visibility	Element damage
Grouped	68.7	49.3
Interleaf	87.1	53.1

- MTurkers answered YES more often

Average # of YES responses		
	Element visibility	Element damage
Panel	65.4	47.1
MTurk	88.9	55.0

Take Aways (preliminary)

- Results not as expected
  - Survey: Format effect only in MTurk
  - MTurkers are similar to other survey respondents
  - Why no format effect in panel? ???
    - No motivated misreporting in Panel?
    - Or misreporting in both formats?
  - Image Coding: Format effect in opposite direction
- Some evidence MTurkers work harder than panelists
  - Survey: less item NR
  - Image Coding: longer time with training materials

Discussion

- Data scientists are doing surveys to make training data

- We know a lot about survey data quality!
  - Measurement error
  - Nonresponse error
  - Coverage error

How do these affect
- **Training data?**
- **Model predictions?**

More Information

Y. Patrick Hsieh

yph@rti.org

@coolpat

Stephanie Eckman

seckman@rti.org

@stephnie