

# Motivated misreporting in crowdsourcing: Data quality concerns for scientific tasks

Y. Patrick Hsieh, Herschel Sanders, Amanda Smith, Stephanie Eckman Survey Research Division, RTI International

Project funded by NSF #1645307

## INTRODUCTION

- Crowdsourcing have become a popular source of online workforce for assisting scientific and engineering research
  - Classifying images
  - Transcribing audio files
  - Coding texts or social media content
  - Answering surveys
  - Extracting information from the web
- Timeliness
- Cost-effectiveness
- However, what about data quality?

## RESEARCH OBJECTIVE

To explore whether crowd workers may engage in **satisficing** & **motivated misreporting** whereby participants tend to cut corners to reduce their workload of scientific tasks

## EXPERIMENT

- Randomized web survey experiment
- Two formats of question design
  - **Interleafed** – presenting gate questions one at a time, **easier** for participants to learn to skip and avoid follow-up questions
  - **Grouped** – presenting all gate questions on one page, **harder** for participants to learn to skip and avoid follow-up questions
- Three types of common scientific tasks
  - Answering surveys
  - Coding tweets
  - Classifying images (damaged buildings)
  - Benchmarks from literature as coding accuracy for assessing format effect
- Participants randomly assigned to one task in one question format

## Methods

- Design of contemporary lifestyle survey task
  - 30 minutes, 11 sections, total of 267 items
  - 5 experimental question modules (Eckman et al, 2014); the sequence of presentation is randomized for 1<sup>st</sup> – 4<sup>th</sup>; 5<sup>th</sup> is near the end
    - 1. Clothing purchase (6 filters)
    - 2. Ownership of consumer goods (6 filters)
    - 3. Leisure activity (6 filters)
    - 4. Ownership of credit cards (5 filters)
    - 5. Employment (6 filters)
  - Items within module are presented in forward or backward order (randomly assigned to participants)
- Tweet coding task
  - 50 minutes, 45 tweets contains keywords related to marijuana use
  - 2 technical ‘tweet’ questions; 3 filter questions about **sleep, nausea, and pain**, followed by 2 questions about medical conditions/symptoms and tweet’s sentiment
- Damaged building image coding task
  - 50 minutes, 40 building images of the aftermath of Haiti earthquake
  - 1 target-identifying question; 4 filter questions about **beam, column, slab, and wall**, followed by 3 damage assessment items

	Female	Male	Total
M Turk	108 59%	75 41%	183
Panel	119 53.6%	103 46.4%	222
Total	227 56%	178 44%	405

- Crowd workers: average age of 40
- Panelists: average of 50

## Results

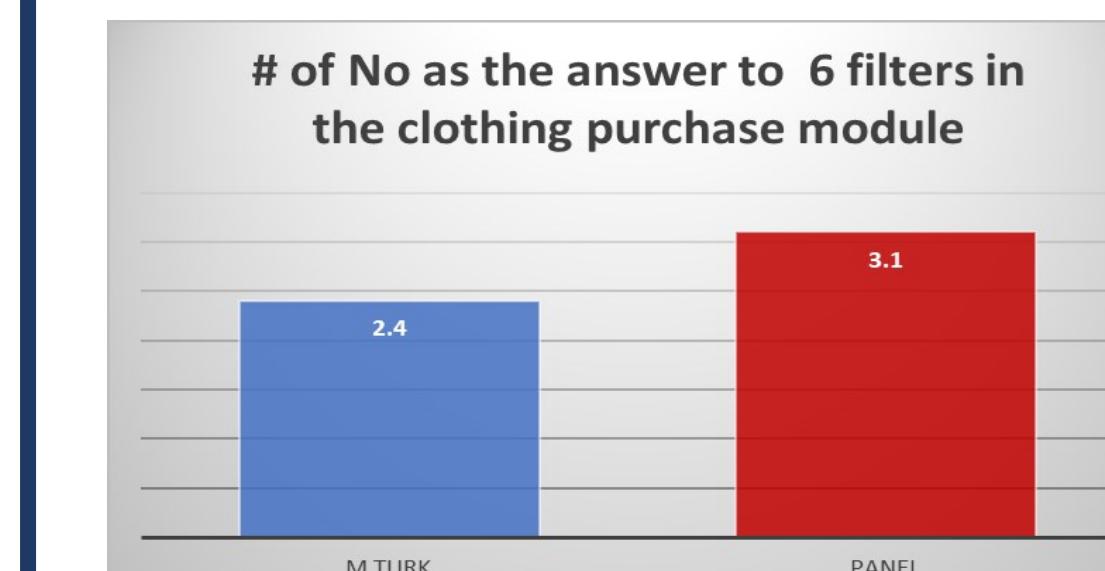
	Complete	Drop out	Screen out	Total
M Turk	184 87.2%	9 4.3%	18 8.5%	211
Panel	204 78.5%	24 9.2%	32 12.3%	260
Total	388	33	50	471

$\chi^2=6.74$ ,  $p=.034$

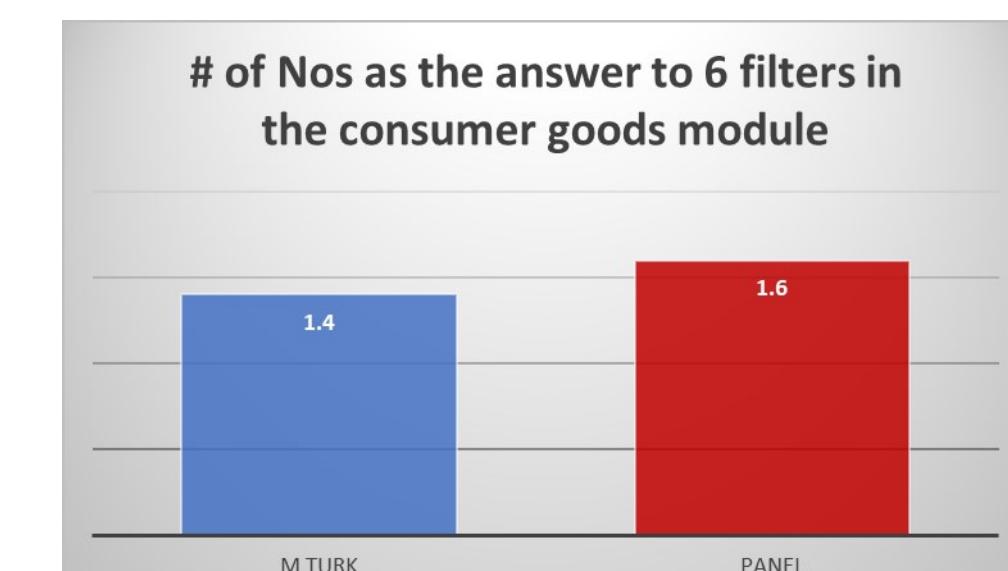
- There is no statistically significant **format effect** on motivated misreporting

	# of Nos as the answer to filters in module				
	Clothing	Goods	Leisure	Credit Card	Employment
Grouped	2.2	1.3	2.7	2.4	2.9
Interleaf	2.5	1.2	2.7	2.4	2.8

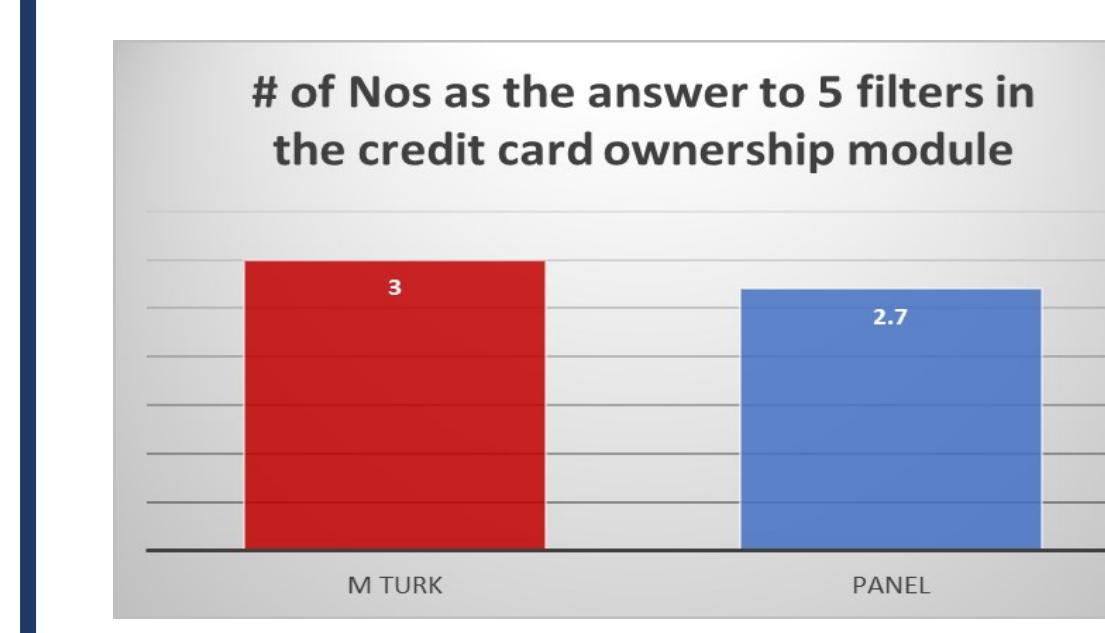
- There is supportive evidence indicating crowd workers are less likely to cut corners (i.e. statistically significant **source effect**)



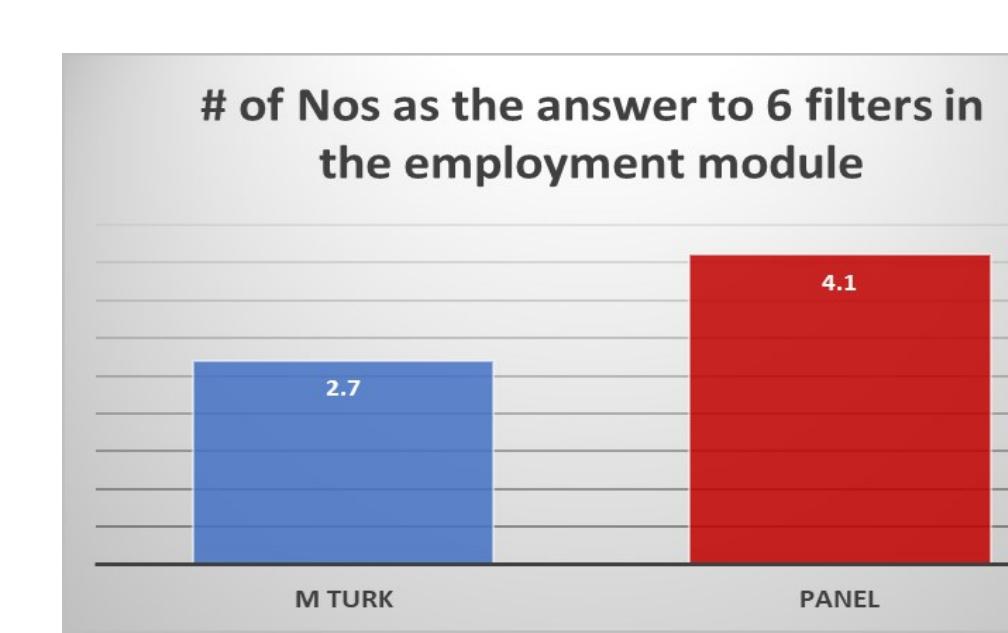
$t=-3.7$ ,  $p=.000$



$t=2.0$ ,  $p=.022$



$t=1.7$ ,  $p=.046$



- Only the item non-response rates of the very **last** module (Mturk 14% vs Panel 18%) are significantly different

## Conclusion

- Crowd workers may be a better source of research participants than online panelists
- Next steps: Analyzing coding accuracy of tweets and images for experimental effects