A Crowdsourcing Methodology to Measure Algorithmic Bias in Black-box Systems: A Case Study with COVID-related Searches

***Binh Le*, Damiano Spina, Falk Scholer, Hui Chia**

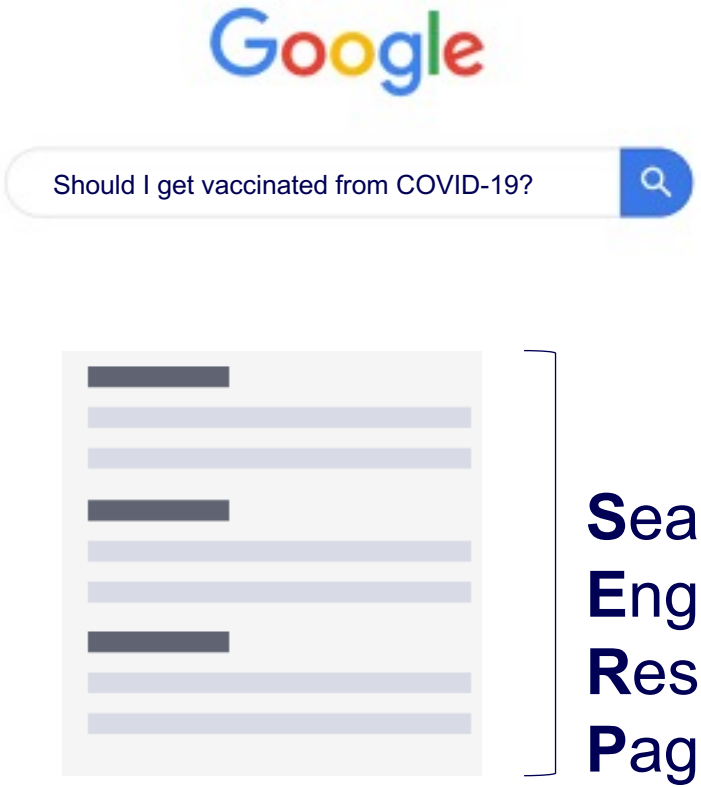
binh.le@rmit.edu.au
damiano.spina@rmit.edu.au
falk.scholer@rmit.edu.au
chia.h@unimelb.edu.au

RMIT UNIVERSITY

CiDDA

THE UNIVERSITY OF MELBOURNE

Motivation



Search
Engine
Results
Page

SERP 1

SERP 2

SERP 3

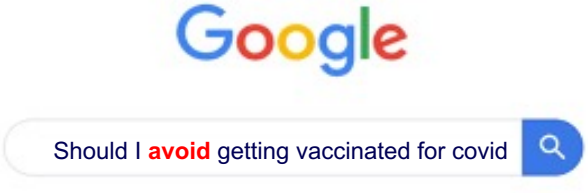
Google

Should I get vaccinated from COVID-19?

RQ1:
Do different users get the same search results?

CiDDA

User



RQ2:
Do results vary between positive and negative query formulations?

Methodology

Crowdsourcing Platform to Audit a Web Search Engine

- Requesters: Publish micro-tasks or Human Intelligent Tasks (HITs)
  - Annotation tasks to obtain labels for machine learning development an evaluation)
- Crowd workers: Can choose to carry out micro-tasks at their convenience

Crowdsourcing Task

Upload the HTML file for the following query: is hydroxychloroquine effective for covid

Click on the following query to open a new page with the corresponding google search results: is hydroxychloroquine effective for covid

Choose file No file chosen

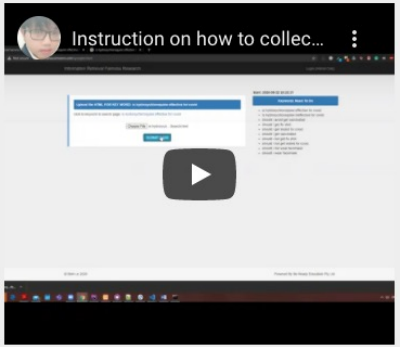
SUBMIT DATA

Start: 2021-08-30 07:27:41

Queries that still need to be completed:

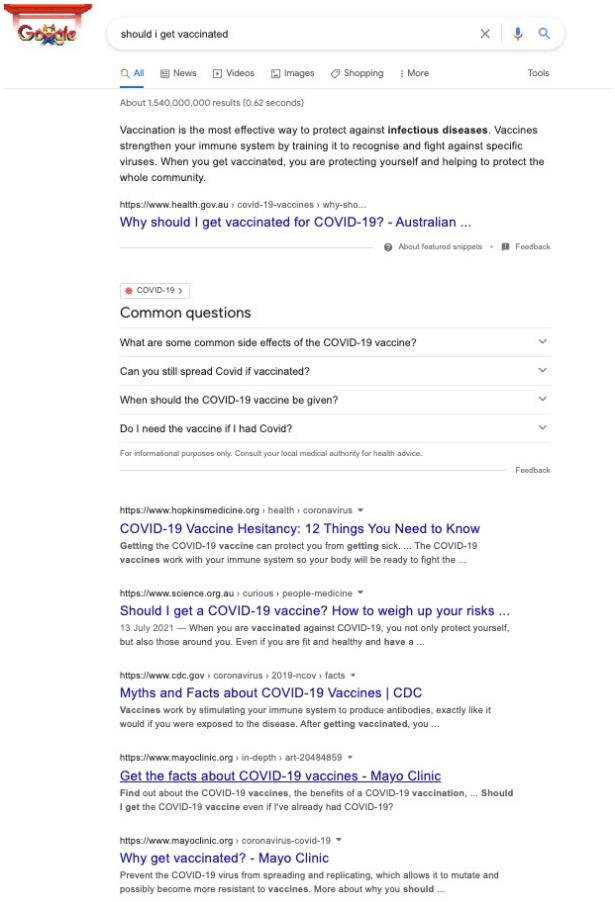
- should i get flu shot
- should i get tested for covid
- should i get vaccinated
- should i not get tested for covid
- should i not get flu shot
- is hydroxychloroquine effective for covid
- is hydroxychloroquine ineffective for covid
- should i avoid get vaccinated
- should i wear facemask
- should i not wear facemask

Instruction



Useful Links

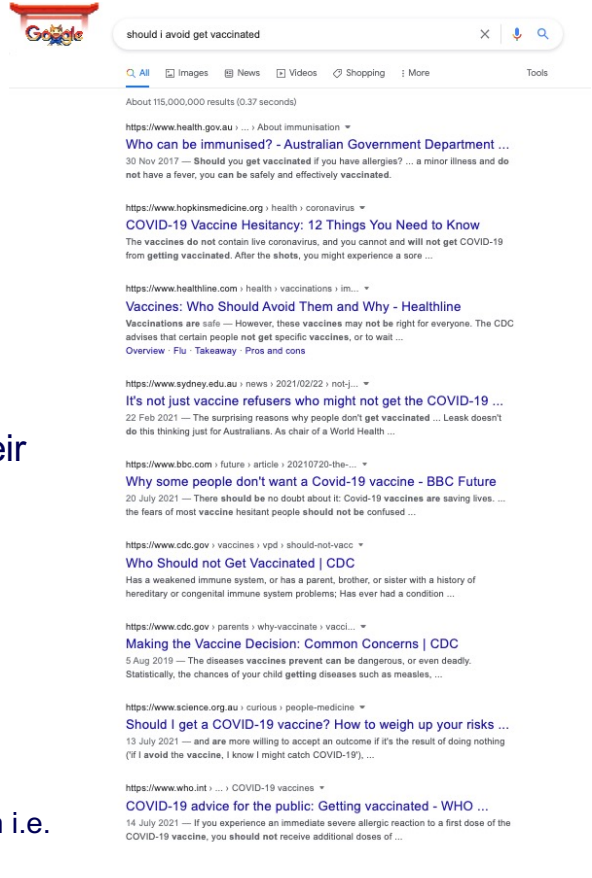
Participant Information Sheet



Participant
(Crowd Worker)

1. Submit the query to Google using their browser
2. Download SERP as HTML file
3. Upload HTML file into the web app

Our system will remove all the personal information i.e. avatar, email address



Queries

10 queries – 5 pairs

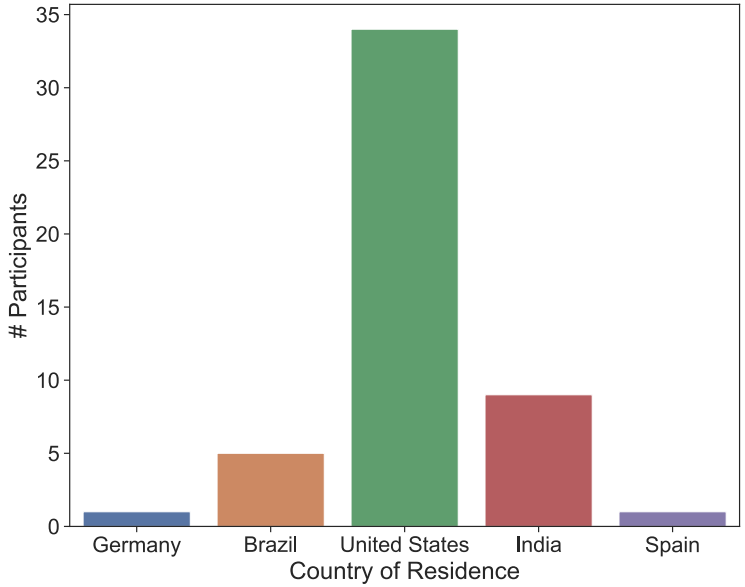
1	should i get tested for covid	Pair 1
2	should i not get tested for covid	
3	should i get flu shot	Pair 2
4	should i not get flu shot	
5	should i get vaccinated	Pair 3
6	should i avoid get vaccinated	
7	should i wear facemask	Pair 4
8	should i not wear facemask	
9	is hydroxychloroquine effective for covid	Pair 5
10	is hydroxychloroquine ineffective for covid	

Queries

- US$ 1.20 per task
- 10 queries per task
- 1 task per participant
- N=50 participants

- Average time: 23’ 30”
- All (50+1) crowdsourcing tasks completed in 2 days. Only 1 incorrect submission.

- Pre-questionnaire: Age, gender, country of origin, country of residence, level of education, …



Comparing SERPs

Rank-Biased Overlap (RBO)

- Non-conjoint, top-weighted, and incomplete ranked lists
- Parameter “p” (patience): probability of the user to continue inspecting the ranking
  - It determines the strength of the weighting to top ranks
  - E.g., p=0.67: the user inspects only the top-3 search results.

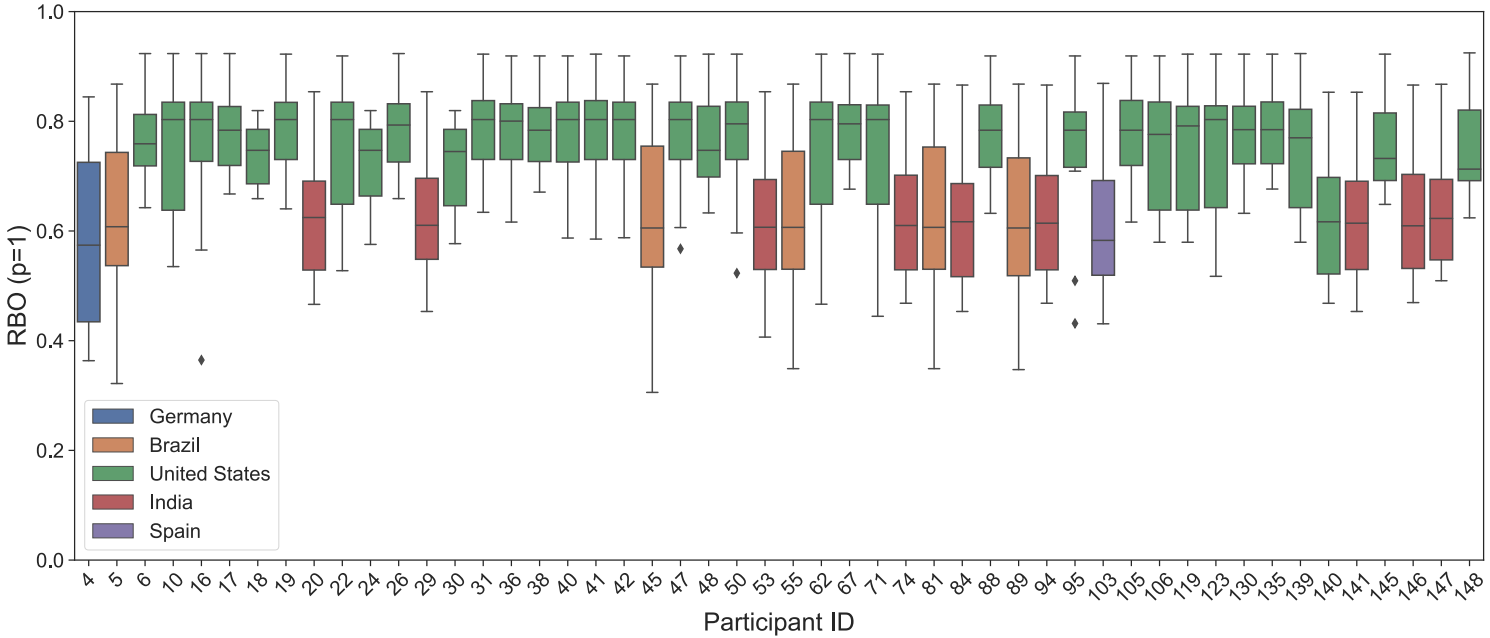
Jaccard Similarity

- Overlap of items in two sets
- It does not consider ranking positions, i.e., it only measures conjointness

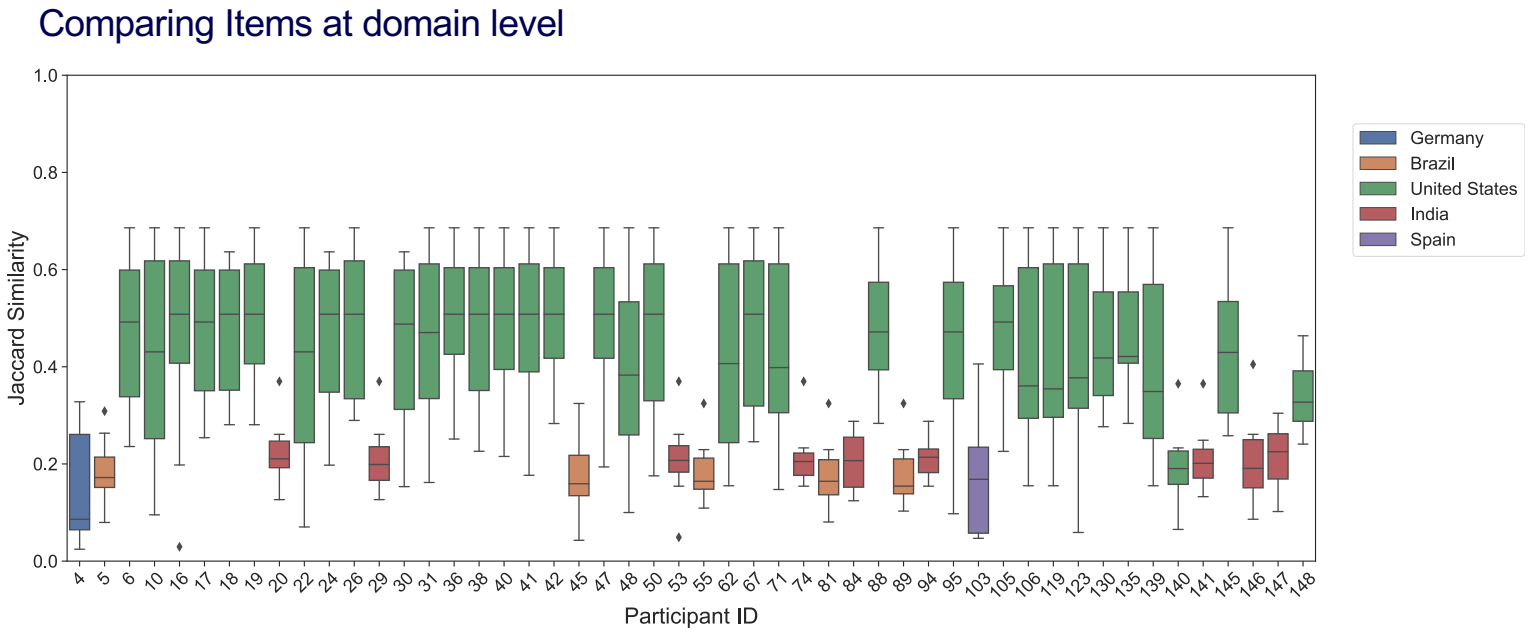
Preliminary Results

Do different users see different SERPs?

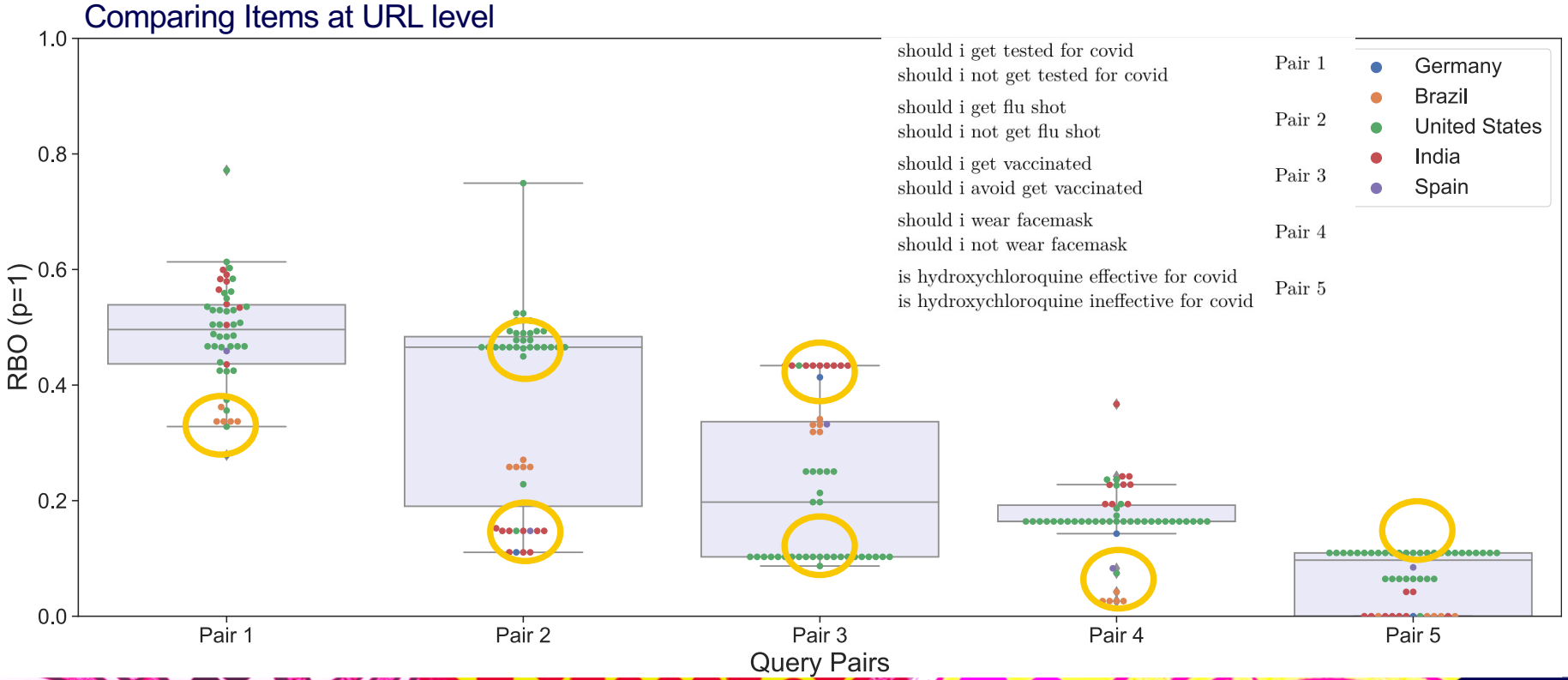
Comparing items at domain level



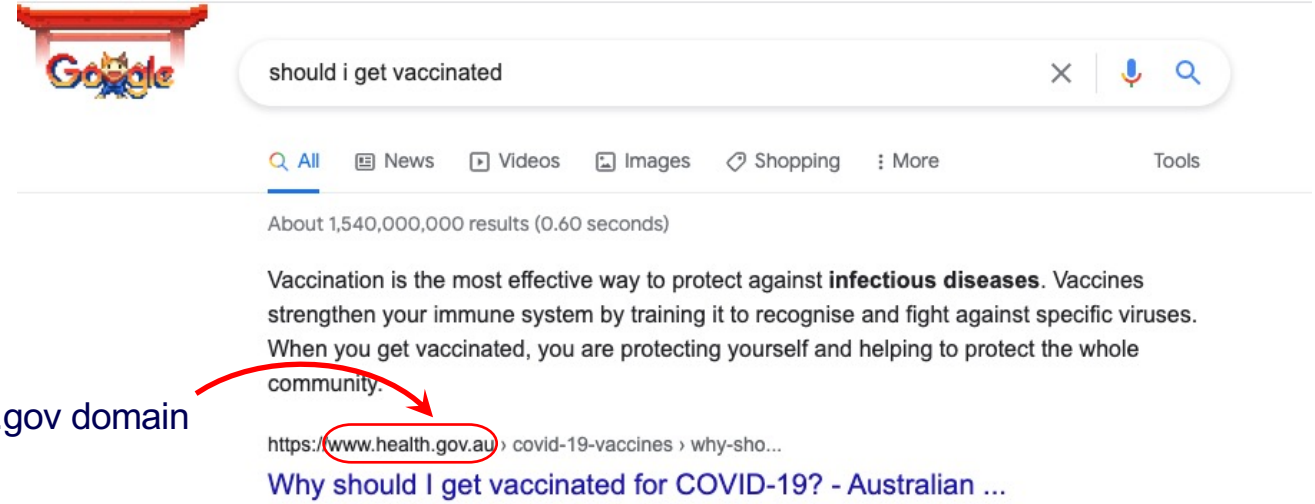
Do different users see different SERPs?



Does Query Formulation Change SERPs?

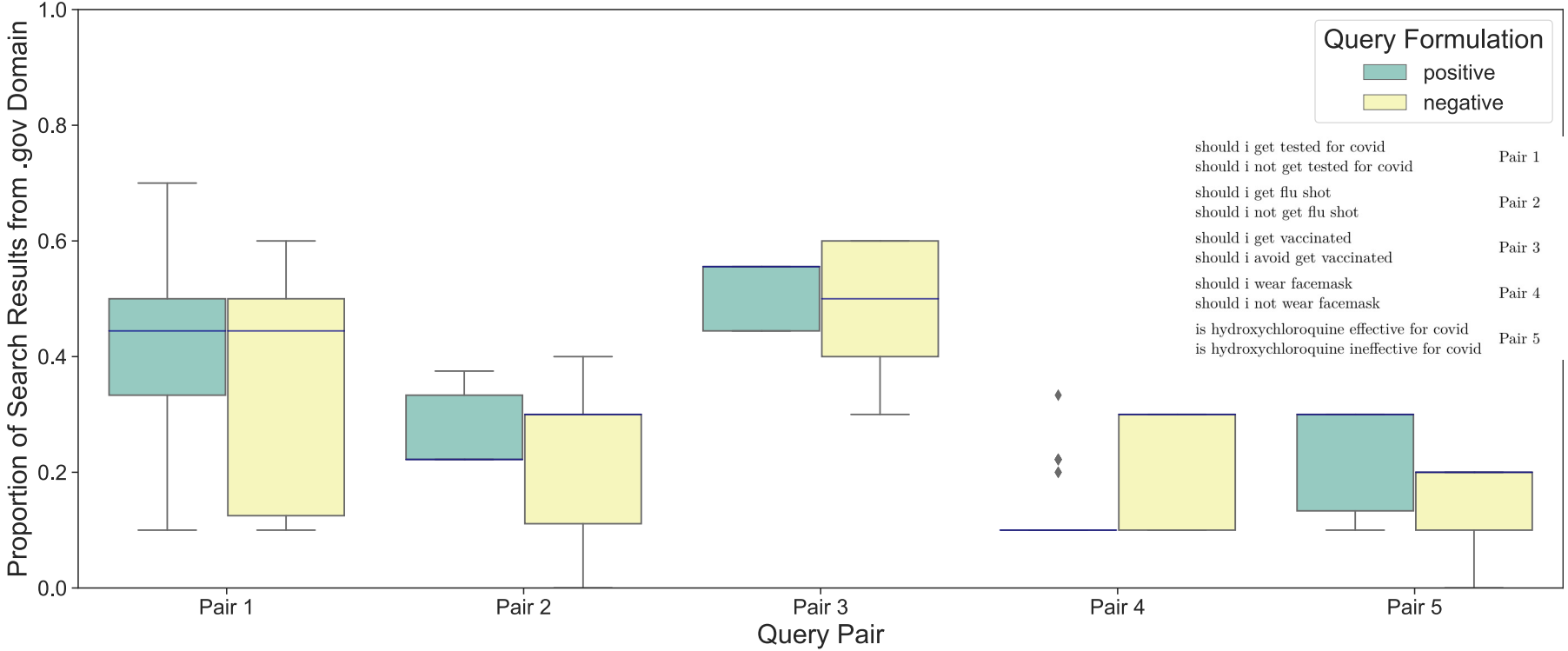


[EXTRA] Quality of Information Sources



.gov domain

Items from .gov Domain in SERPs



Conclusion

- We used a crowdsourcing platform to audit a web search engine (black-box system)
  - Preliminary study
  - We observed differences in the SERPs obtained by crowd workers / participants

- Our study validates the feasibility of this approach
  - Future work: Larger dataset / number of workers
  - A more scalable (but less controlled) methodology than *data donation*

A Crowdsourcing Methodology to Measure Algorithmic Bias in Black-box Systems: A Case Study with COVID-related Searches

github.com/rmit-ir/crowdsourcing-algorithmic-bias

binh.le@rmit.edu.au

damiano.spina@rmit.edu.au

falk.scholer@rmit.edu.au

chia.h@unimelb.edu.au

Acknowledgments:

Australian Government
Australian Research Council

ADM +S ARC Centre of Excellence for Automated Decision-Making and Society

RMIT UNIVERSITY

THE UNIVERSITY OF MELBOURNE

CiDDA