Revisiting the managerial impact of AI algorithmic fairness on business decisions: A replication of Cowgill et al. 2020*

Yingying Zhou

08/04/2021

Abstract

This paper reproduces (Cowgill, Dell'Acqua, and Matz 2020) to reinvestigate the managerial impact of algorithm fairness using a RCT field experiment under two business decision-making scenarios. This paper identifies demographic traits explaining the fundamental diversity in firm manager's attitude towards AI and further examines the effect of interventions on AI adoption through activism arguments on algorithmic fairness. This paper finds that counterfactual advocacy arguments on algorithmic bias are more effective in promoting AI adoption for business use. Besides argument manipulations, race, gender, and knowledge on the status quo fundamentally impact manager's judgement on AI adoption.

Contents

*Code and data are available at: https://github.com/StephaininZ/algorithm_fairness

1 Introduction

Artificial Intelligence algorithmic bias in business decision-making inflicts ethical consequences which mostly materialize to company image damage and revenue loss. While AI technology has expedited the decision-making process in business operations, liberating manual efforts and replacing rule-based models in the age of big data for firms, concerns have been aroused on AI adoption by business decision-makers.

(Cowgill, Dell'Acqua, and Matz 2020) examines the managerial effects of argument intervention through two Randomized Controlled Trial field experiment series in two business cases on hiring and lending decisions via investigations on effect originating from opinion polarity and from scientific veneer. In the first study, (Cowgill, Dell'Acqua, and Matz 2020) randomly assigns subjects to one of three op-ed conditions (fatalistic, counterfactual or no op-ed) to assess participant's adoption decisions and relationship of the argument to the status quo. In the second study, (Cowgill, Dell'Acqua, and Matz 2020) measures the impact of adding scientific authority to arguments on AI ethics using a 2×2 factor design.

(Cowgill, Dell'Acqua, and Matz 2020) finds that fatalism op-ed discourages AI adoption, while the counterfactual encourages it. As for the belief on the fixability of the algorithmic bias, participants tend to have more faith in correcting the fairness problems under fatalistic op-ed conditions. On the other hand, scientific veneer on arguments would reinforce the persuasive effect of the viewpoint on an approximately equal scale for either direction (positive or negative) and thus affect manager's decision to adopt AI.

Using the replication dataset and code provided by the original authors, I re-implement the first study to further the managerial impact of activism arguments through the working channel of opinion polarity in altering business decision-maker's perception and adoption decision on AI. This paper identifies the heterogenous effects from individual demographic characteristics of on their consistent choice set of attitudes towards AI adoption (excluding treatment effects) via a multiple linear regression. Subsequently, it analyses the marginal effect of exerting engineering effort on the belief of AI bias fixability through a panel OLS regression with individual fixed effects.

My findings about the managerial effects by argument polarity are consistent with (Cowgill, Dell'Acqua, and Matz 2020) that arguments stressing the inevitability of algorithmic bias lead managers to abandon AI and op-eds claiming the superiority of AI models relative to human in producing lower bias would encourage AI adoption. In addition, the counterfactual op-ed precondition would sway manager's belief in being able to improve AI bias after the engineering effort (fixability outcomes). For the ordinary linear regression inference about demographic traits and status quo condition, this paper discovers that algorithm models are significantly less favored by female, African Americans, and politically liberal managers compared to other genders, race and political affiliations throughout the study.

Artificial Intelligence technology has a promising prospect in accelerating business decision-making process by delegating data-heavy insight mining and predicting optimal business plans provided that its ethical fairness problem can be mitigated in the future. Once the algorithmic bias can be addressed, AI would be one of the workhorses in the era of big data, especially from the perspective of business operations. Therefore, algorithmic fairness activism should promote the adoption of AI, allow some time for technology refinement despite its current defect in ethical bias. This paper builds on these incentives to investigate which activism intervention influences manager's adoption decision on the AI technology.

The remainder of the paper is constructed as follows. Section 2 describes the dataset, experiment design, and exploratory data analysis on feature visualization. Section 3 outlines the reproduced experiment models, which is designed to discover relationships between features and the target variable. Section 4 summarizes the model results according to evaluation criteria. Finally, Section 5 discusses our research findings and provides directions for future research.

R statistical programming language (R Core Team 2020) is used to replicate the experiment. To be specific, `tidyvers` package is for data preprocessing (Wickham et al. 2019), `kableExtra` package is applied to generate tables (Zhu 2020), `ggplot2` is used to draw diagrams (Wickham 2016), `ggthemes` is for diagram theme changing (Arnold 2021).

2 Data

2.1 Dataset features

Subjects demographic traits (table)

2.2 Experiment in the paper

sampling, methodology, intervention, experiment flow chart (enclosed in appendix, to be redrawn later)

2.3 Descriptive analysis

3 Model

3.1 Multiple linear regression with interaction terms

4 Results

5 Discussion

5.1 Bias and ethical concerns

5.2 Model results

5.3 Real world implications

5.4 Internal validity & external validity of model

5.5 Weakness and opportunities for future work

5.6 Differences and difficulties

A Appendix

```
## here() starts at /Users/stephaniezhou/Desktop/GitRepo/algorithm_fairness
```

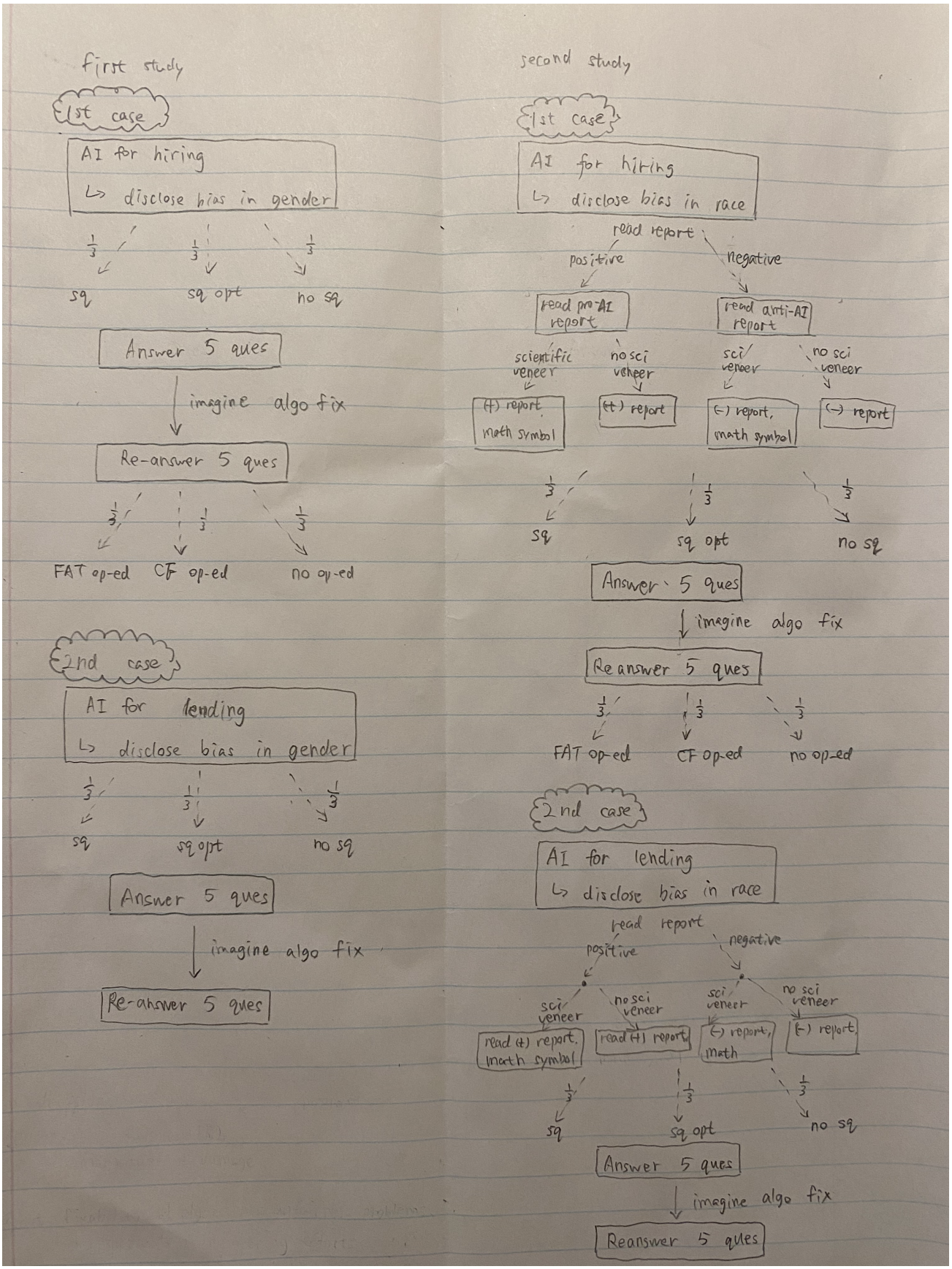

Figure 1: Draft of the overall experiment flows

References

Arnold, Jeffrey B. 2021. *Ggthemes: Extra Themes, Scales and Geoms for 'Ggplot2'.* https://CRAN.R-project.org/package=ggthemes.

Cowgill, Bo, Fabrizio Dell'Acqua, and Sandra Matz. 2020. "The Managerial Effects of Algorithmic Fairness Activism." *AEA Papers and Proceedings* 110 (May): 85–90. https://doi.org/10.1257/pandp.20201035.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Zhu, Hao. 2020. *KableExtra: Construct Complex Table with 'Kable' and Pipe Syntax.* https://CRAN.R-project.org/package=kableExtra.