Maastricht University

Essays on Technology and Society

Citation for published version (APA):

Giuberti Coutinho, L. (2022). *Essays on Technology and Society: The Impacts of Algorithms and Online Media on Labor Market Outcomes and Political Dynamics*. [Doctoral Thesis, Maastricht University]. Maastricht University. https://doi.org/10.26481/dis.20221025lg

Document status and date:
Published: 01/01/2022

DOI:
10.26481/dis.20221025lg

Document Version:
Publisher's PDF, also known as Version of record

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy
If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Download date: 20 Apr. 2024

Essays on Technology and Society

The Impacts of Algorithms and Online Media on Labor Market Outcomes and Political Dynamics

Lorena Giuberti Coutinho

Copyright ©Lorena Giuberti Coutinho, 2022
ISBN 978-94-6469-013-2

Cover Design: Ron Zijlmans
Printing: ProefschriftMaken || www.proefschriftmaken.nl

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form, or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission in writing, from the author.

Essays on Technology and Society

The Impacts of Algorithms and Online Media on Labor Market Outcomes and Political Dynamics

DISSERTATION

to obtain the degree of Doctor at Maastricht University,
on the authority of the Rector Magnificus Prof. dr. Rianne M. Letschert,
in accordance with the decision of the Board of Deans,
to be defended in public on Tuesday 25 October 2022, at 16:00 hours.

by

LORENA GIUBERTI COUTINHO

SUPERVISORS:
Prof. Dr. Robin Cowan
Dr. Matthias Wibral

ASSESSMENT COMMITTEE:
Prof. Dr. Pierre Mohnen (Chair)
Dr. Micheline Goedhuys
Prof. Dr. Mauricio Bugarin
Dr. Yphtach Lelkes

SUMMARY

This dissertation investigates how new technologies, particularly algorithms and online media, impact human decision-making and attitudes. The thesis is structured into five chapters, with Chapter 1 providing the roadmap of the dissertation and introducing the motivation, the research questions, the methodology, and the data used. Chapters 2, 3 and 4 investigate, through empirical research, the two topics of the thesis: the impacts of digital technologies on the job market and on the political sphere of democratic countries.

Chapter 2 investigates whether the recommendations made by an algorithm are perceived differently from those made by a human (expert), and potentially lead to a different outcome of the hiring process. Through a preregistered field experiment with law firms, I test whether there is an observable difference in employers' evaluation of candidates recommended by algorithms in comparison to those recommended by human advisors. I further elicit preexisting attitudes and beliefs regarding certain characteristics of algorithms in a labor market context and a general preference for algorithmic vs human advice. This allows me to explore whether and how potential differences in these individual beliefs and preferences might alter how they respond to advice generated by these sources. I take advantage of the setup of a large job fair to collect employers' judgments on CV recommendations without intervening in any other behavior except for randomly labeling the CVs.

Results show no overall effect of the CV label. While characteristics such as work experience or speaking English do affect the rating of a candidate's qualification and the hiring interest, the source of recommendation does not. Findings also show no overall effect on job offers or CV retention. In the analysis of heterogeneity in the treatment effect along preexisting attitudes and beliefs regarding certain characteristics of algorithms, I do not find an interaction of the treatment effect with the labor market specific beliefs about algorithms. This potentially reflect the fact that most participants do not believe that one source of recommendations dominates the other in all three dimensions elicited. It is rather the general preference for algorithms which seems to matter for the qualification rating of a candidate. On one hand, decision makers with a general preference for algorithmic advice give significantly higher ratings if a candidate was recommended by an algorithm compared to when the candidate was recommended by a human resource expert. On the other hand, decision makers with a general preference for human advice give significantly lower ratings to candidates recommended by algorithms when compared with candidates recommended by humans.

Chapter 3 and 4 examine the second topic of the thesis: the impact of internet and social media usage on the political sphere and, more specifically, on the process of political polarization. The Chapters search for a causal link between the internet and social media use and (i) the enhancement of feelings of hostility towards opposing candidates and (ii) greater deviation from centrist positions on political ideology and attitudes. Both studies use instrumental variables (IV) techniques to investigate the effect of internet and social media use on political polarization.

In Chapter 3, I employ an IV approach that follows past studies using exogenous infrastructure variations to identify the internet's impact on political attitudes. I use exogenous variation in the fiber-optic backhauls infrastructure to identify the impact of internet and social media usage on political polarization in Brazil. For the IV analysis, we use a dummy variable indicating the availability of fiber optic backhaul for each municipality in 2018.

In Chapter 4, I employ a system generalized method of moments (System-GMM) estimator applied to a dynamic panel data model to explore the effects of social media use in political decisions in the Netherlands.The System-GMM is acknowledged as the most efficient method to estimate dynamic panel models that suffer from endogeneity by employing (internal) IV techniques. The system-GMM allows for the dynamic nature of political polarization and controls for unobserved, time invariant, and individual-specific effects. Thus, the panel data analysis goes beyond the cross-sectional analysis in Chapter 3.

The findings of Chapters 3 and 4 indicate that, contrary to what is suggested by the mainstream literature, internet and social media use may not be the cause of rising political polarization across countries. Chapter 3 finds that the enhancement of feelings of hostility towards opposing candidates in Brazil, a phenomenon called affective polarization, cannot be attributed to internet or social media use. While the study identifies a positive relationship between social media usage and polarization, when the internet and social media are treated as endogenous variables and a two-equation is estimated using the IV, the relationship between social media usage and polarization disappears.

Chapter 4 goes further in the research of the relationship between social media usage and political polarization and finds that, in the Netherlands, social media use attenuates rather than drives polarization, a result that holds for different measures of social media use - dummy (yes vs. no), intensity (time spent), and frequency. The study identified that reading and viewing social media has a significant and negative effect on polarization. More hours spent reading and viewing social media per week, and greater frequency of social media use are associated with lower polarization.

Finally, Chapter 5 concludes the dissertation by answering the research questions and identifying the policy implications and limitations of the dissertation before suggesting avenues for future research.

ACKNOWLEDGEMENTS

It is a pleasure to thank the many wonderful people who have contributed to the completion of this thesis. First, I am deeply grateful to my supervisor Professor Robin Cowan. I will never forget his patience in developing my blurred ideas into a Ph.D. project in the first year of my Ph.D. His wise advice, selfless attention, and serenity are something truly special that I will always take with me. I was also fortunate to work with Dr. Matthias Wilbral, my second supervisor, whose close support was essential for developing this thesis. Thank you, Matthias, for guiding me with your knowledge and experience and for the many intellectual discussions that have enriched me as a professional.

I would like to extend my gratitude to the members of the dissertation committee Prof. Dr. Pierre Mohnen, Dr. Micheline Goedhuys, Prof. Dr. Mauricio Bugarin, and Dr. Yphtach Lelkes for spending their time providing valuable comments and insightful suggestions, which enriched this dissertation.

Studying at UNU-MERIT was a pleasure, and having constant contact with colleagues and staff from all over the world was truly fun. A special thank you goes to Eveline, whose generosity throughout the Ph.D. process is invaluable (thank you for taking me on your motorcycle during my first week in Maastricht!). My gratitude also goes to the amazing friends and colleagues I have encountered in Maastricht. Special thanks to the Latin community, Flor, Michelle, Bea, Danilo, Mario, Antonio, Cintia, and Stefanie, who made the Ph.D. process lighter and more fun. Double thanks to my paranymphs: Flor, thank you for being my family in Maastricht and now my sister. Michelle, thank you for the constant joy, company to work in cafes, and heterodox conversations. Big thanks to my office mates, Robert and Sabrina, with whom I shared difficult and joyful moments.

My eternal gratitude goes to my family. Mom and Dad, thank you for providing me with a safe harbor to follow my dreams. To my beloved sisters, Janine and Paula, who supported me during the most challenging phases of this Ph.D. To my husband Francisco, without whom this thesis would have been hardly possible to complete. Thank you for sharing with me all of the agonies and joys of this journey. I am always amazed by your rare quality of transforming arduous and complex activities into simple and pleasurable moments. This has allowed us to achieve many great things during this journey, the best of them being the arrival of our son, Benjamin. Son, thank you for sharing with me the emotions of the last moments of this Ph.D. and for bringing me the strength and focus needed to complete it. Let this thesis be dedicated to you.

CONTENTS

1 INTRODUCTION 1
- 1.1 Motivation . . . 1
  - 1.1.1 The use of algorithms in decision-making . . . 3
  - 1.1.2 Human-algorithm interactions in decision-making . . . 5
  - 1.1.3 Human-algorithm interactions in job market decisions . . . 7
  - 1.1.4 Online Media and Democracy . . . 7
  - 1.1.5 Online Media and Political Polarization: the Question of Echo Chambers and Filter Bubbles . . . 9
  - 1.1.6 Political Polarization and Internet and Social Media Usage in Brazil and the Netherlands . . . 11
- 1.2 Research Questions . . . 12
- 1.3 Contribution and Dissertation Outline . . . 13

References 15

2 HUMAN VS ALGORITHMIC RECOMMENDATIONS IN THE LABOR MARKET: EVIDENCE FROM A FIELD EXPERIMENT 21
- 2.1 Introduction . . . 22
- 2.2 Related Literature . . . 24
- 2.3 Research Design . . . 27
  - 2.3.1 Data and Descriptive Analysis . . . 29
- 2.4 Results . . . 31
  - 2.4.1 Preexisting Attitudes Towards Algorithms and Humans . . . 32
  - 2.4.2 Heterogeneous Effects . . . 34
- 2.5 Conclusion . . . 40

References 42

A APPENDIX 45
  - A.0.1 Recruitment Materials . . . 45
  - A.0.2 Experimental Instructions and Questions . . . 46
  - A.0.3 Screenshots . . . 47

3 THE IMPACT OF INTERNET AND SOCIAL MEDIA USE ON AFFECTIVE POLARIZATION: EVIDENCE FROM BRAZIL 57
- 3.1 Introduction . . . 58

3.2 Measuring Affective Polarization 60
3.3 Data 62
3.3.1 Affective Polarization 62
3.3.2 National Trends in Affective Polarization 63
3.3.3 Internet and Social Media Use 63
3.3.4 Broadband infrastructure for Brazilian municipalities 64
3.3.5 Control Variables 65
3.4 Identifying the Effect of internet and Social Media Use on Affective Polarization 65
3.4.1 Estimation Strategy 66
3.5 Results 69
3.6 Conclusion 72

References 74

B APPENDIX 78

4 DEPOLARIZATION THROUGH SOCIAL MEDIA USE? A DYNAMIC PANEL ANALYSIS FOR THE NETHERLANDS 83
4.1 Introduction 84
4.2 Related Literature 85
4.3 Data 88
4.3.1 Political Polarization 88
4.3.2 Social Media Use 89
4.3.3 Control Variables 90
4.4 Identification Strategy 90
4.5 Results 92
4.5.1 Effects of Social Media Use on Polarization 92
4.5.2 Alternative Polarization Measures 94
4.5.3 Robustness Checks 94
4.6 Heterogeneous effects 98
4.7 Conclusion 98

References 100

C APPENDIX 104
C.0.1 Variables 104
C.0.2 Descriptive Statistics 107
C.0.3 Robustness Checks 108

5 SUMMARY AND CONCLUSION 115
5.1 Summary 115

5.2 Limitations and Avenue for Future Research 117
5.3 Conclusion 119

References 122

LIST OF TABLES

Table 2.1 Descriptive Statistics - Outcome Variables 31
Table 2.2 Descriptive statistics - Control Variables 31
Table 2.3 OLS Estimations - Qualification rating 33
Table 2.4 OLS Estimations - Hiring interest 34
Table 2.5 Logit Estimations - Job offer or CV retention 35
Table 2.6 Descriptive Statistics - Perceptions 36
Table 2.7 OLS Estimations - Qualification rating (heterogeneous effects) . 38
Table 2.8 OLS Estimations - Hiring rating (heterogeneous effects) 39
Table 3.1 Percentages and changes in affect towards candidates between 2002 and 2018 63
Table 3.2 Affective Polarization 70
Table 3.3 Internet and Social Media Use 71
Table B.1 Instrumental Variable Relevance 78
Table B.2 Conditional Independence 79
Table 4.1 System-GMM - Ideological Polarization 93
Table 4.2 System-GMM - Opinion Polarization (Immigration) 95
Table 4.3 System-GMM - Opinion Polarization (EU Integration) 96
Table 4.4 System-GMM - Opinion Polarization (Income Inequality) 97
Table C.2 Descriptive Statistics 107
Table C.3 Placebo System-GMM - Ideological Polarization 108
Table C.4 Placebo System-GMM - Opinion Polarization (Income Inequality)109
Table C.5 Placebo System-GMM - Opinion Polarization (EU Integration) . 110
Table C.6 Placebo System-GMM - Opinion Polarization (Immigration) . . 111
Table C.7 Robustness System-GMM Estimations - Ideological Polarization 112
Table C.8 Robustness System-GMM - Opinion Polarization (Immigration) 112
Table C.9 Robustness System-GMM - Opinion Polarization (UE integration)113
Table C.10 Robustness System-GMM - Opinion Polarization (Income Inequality) 113

LIST OF FIGURES

Figure 2.1 Experiment timeline 28
Figure 2.2 Mean outcome variables by treatment (algorithm vs. human recommended) 30
Figure 2.3 Employers prefer humans in most dimensions 36
Figure 2.4 Qualification ratings depending on employers preexisting preferences 37
Figure A.1 Recruiting e-mail 45
Figure A.2 Screenshots Qualtrics 48
Figure A.3 Average qualification rating by firm 55
Figure A.4 Average hiring rating by firm 56
Figure B.1 Changes in average affective polarization between 2002 and 2018 81

1

INTRODUCTION

1.1 MOTIVATION

Over the last decades, digital technologies have transformed almost every aspect of social life and human experience. Digital transformation has upended traditional industries at a remarkable speed, creating new products and services, designing groundbreaking business models, and engendering enormous economic value. The widespread dissemination of automated systems and algorithms allowed for massive efficiency gains and spurred a landscape of constant innovation. Social media caused structural changes in how people interact with each other and how societies communicate, bringing significant developments to markets and other domains, such as the political debate within countries and the international relationships between nations.

On one side, the process of digital innovation has given rise, in a short time, to a large set of products that changed for the better the life of consumers in the whole world. New business models have improved the allocation of resources, broadened the frontiers of markets and generated colossal wealth. Governments have relied on digital technologies to expand and improve the services provided to citizens. Social media has accelerated the dissemination of information, facilitated the coordination of citizens and lowered access barriers to public debate.

On the other, digital transformation also creates a wide assortment of new risks and challenges. Evidence indicates the emergence of large digital divides and their relationship to issues of rising inequality throughout the world (Ho & Tseng, 2006). Wider access to social media has been accompanied by raised concerns regarding the distribution of false or misleading information within society and its possible impacts on the how democracies function (Bradshaw & Howard, 2018; Shu et al., 2020). The complex and ever-evolving dynamics of digital transformation have been followed by the emergence of an extensive body of academic literature that tries to assess and interpret the enormous changes brought by this process. This dissertation seeks to contribute to this debate through the empirical analysis of two topics.

The first one is the impacts of automated and algorithmic solutions on human decision-making processes. Algorithms and automated systems are being increasingly deployed by private companies and government agencies to collect information, analyze data and make decisions in several domains, such as criminal justice, public health, and pension systems. Organizations use algorithms for different goals, such as defining price standards, assessing risks, and predicting human behavior.

The second subject of the thesis relates to one of the most contentious topics in the contemporary public debate of different countries: the impact of digital communication technologies, mainly social media platforms, on the functioning of democracy. The exponential growth of social media in the last decades reshaped the generation, editing, and dissemination of information within societies. Positive expectations and hopes that communication technologies would boost political involvement, expand citizenship and improve the exchange of ideas in the political arena have gradually given way to serious concerns about the risks for democracies brought by digital transformation.

Recently, several voices have pointed out how the digital communication environment can be used to promote, on a large scale, campaigns of disinformation with political purposes (Vosoughi et al., 2018). Electoral disputes that occurred in established democracies have had their integrity tainted with the possibility of manipulation through the strategic usage of social media (Tambini, 2018; Allcott & Gentzkow, 2017). The virtual absence of traditional filters in digital ecosystems, which gave space to the inclusion of new voices in the public sphere, also raises worries about the diffusion of hate speech, violent content, and discriminatory messages towards minorities and marginalized groups (Castaño-Pulgarín et al., 2021).

Far from being mere theoretical issues, the issues analyzed in the thesis have significant practical consequences. As digital transformation becomes pervasive and disrupts almost every aspect of social life, demands for more a solid and coherent regulation framework of digital technologies have risen high on the policy agenda both on the national level and in the international community. In some areas, like taxation and antitrust, there is nowadays almost a consensus that traditional regulation standards are not effective when applied to digital ecosystems and that new forms of public supervision and control are needed to mitigate the risks associated with digital transformation.

In this context, requests for the design and enforcement of more rigid state interventions also have arisen to regulate the employment of algorithms in decision-making and the use of social media within political debate. Policy proposals in different countries have sought to tackle the problems of opaqueness, discriminatory impacts, and lack of accountability that appear with the dissemination of automated decisions. In the United States, the "Algorithmic Accountability Act of 2022" requires companies to assess the impacts of the automated systems they use, with the aim of creating transparency about how and when algorithms are used (Booker et al., 2022). In the EU, the Regulation on Artificial Intelligence (European Commission, 2021) also calls for "conformity assessments" of AI systems used in a number of "high-risk" areas (e.g., education, employment, law enforcement and migration) . Different political solutions have also been suggested to address the challenges brought by digital communication platforms and social networking to the structures of public debate in democratic nations (Balkin, 2020).

As the demand for a more coherent, effective, and proportionate regulatory framework of digital technologies becomes evident, there is a growing need for empirical research that provides useful insights for evidence-based policymaking. Given the multi-faceted and fast-changing nature of digital ecosystems, academia has an important role to play in examining the dynamics of digital transformation and assessing possible harms and risks as well as its causes.

This dissertation aims to improve the current understanding of whether and how new technologies and automated tools affect human decision-making and attitudes. This chapter elaborates on each step of the general framework of the dissertation. First, I review the existing literature on the algorithms in decision-making and particularly on human-algorithm interactions. I then focus on the changes brought by online media on the political domain and particularly on political polarization. I identify the research gaps, research questions, spell out the methods used and lastly present the thesis outline.

1.1.1 *The use of algorithms in decision-making*

Algorithms are hardly a recent invention and have been used to aid decision-making for centuries. They can be defined as "processes or sets of rules to be followed in calculations or other problem-solving operations, especially by a computer" (Oxford Living Dictionary). In the last years, however, we have witnessed an exponential growth in the use of algorithms to perform complex tasks that impact our society.

Driven by the availability of devices that track personal digital data and the advent of complex algorithms using machine learning and artificial intelligence (AI), algorithm tools are drastically revolutionizing decision-making and becoming an integral part of everyday life. Individuals interact with automated systems daily: in face and voice recognition systems, such as those used in smartphones and social media, and in recommender systems, including personalized advertisements, mates on dating platforms, shopping and filtered news and information. Advances in algorithms also provide unprecedented venues for breakthroughs in decision-making processes previously reserved for humans, including making medical diagnoses, predicting judicial decisions, evaluating creditworthiness, and preselection of job candidates.

This algorithmic revolution undoubtedly brings various opportunities and benefits to society and has the potential to greatly improve decision-making. Algorithms are touted for the ability to process large datasets at remarkable speed and to make accurate predictions based on objective parameters, while human predictions, in contrast, are often depicted as inherently inconsistent. People often make different judgments when asked to evaluate the same information twice, as documented in the literature on behavioral economics. Thus, some authors assert that algorithms outperform human decision-making (Kahneman et al., 2021) and even that they will become 'more intelligent' than humans and relegate them to insignificance (Barrat, 2013).

The discussion on whether a set of mathematical rules is superior to human judgment is not new. Already in 1954, Paul Meehl wrote a book titled Clinical Versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence. By reviewing studies on human and numerical predictions on several tasks, Meehl concluded that there was massive and consistent evidence that formalized numerical rules outperform human decision-making (Meehl, 1954). Meehl reaffirmed such a conclusion in a paper published in Science with other co-authors in 1989 (Dawes et al., 1989). Among the reasons pointed out by the authors for the superiority of mathematical rules is the fact that humans are unable to assimilate information consistently.

The literature on behavioral economics has long argued that humans are often influenced by judgment heuristics, biases, or noise, such as intuition, fatigue, recent experience, or minor changes in the ordering of information (Kahneman et al., 1982; Kahneman, 2011). The same information can be inconsequential in one context and critical in another. This can produce random variation in decision-making or judgments. Studies have shown that human professional judgments are extremely variable: individuals often make judgments that differ considerably from their peers, from their own prior conclusions, and from rules that they claim to follow, be it in sentencing criminals (Anderson et al., 1999), valuing commercial properties (Adair et al., 1996), auditing financial statements (Colbert, 1988) or appraising job performance (Taylor & Wilsted, 1974). This makes human predictions inherently noisy and reduces human accuracy and reliability in decision-making (Kahneman et al., 2021). If people often make different judgments when asked to evaluate identical information twice, overall accuracy will be decreased. Numerical mechanisms, in contrast, are able to apply the same rule to all cases through a set of mathematical procedures (Dawes et al., 1989). This is a fundamental difference in the analysis between algorithms and humans.

While the reliance on algorithms has become ubiquitous in decision-making processes across society, several concerns over automated systems' transparency, fairness and accountability have emerged over the last few years (Shin & Park, 2019). Important questions have arisen regarding the opaque nature of algorithmic decision-making and whether the performance of the algorithms comes at the expense of other goals, such as fairness. It is now known that algorithms run the risk of replicating systematic discrimination. Algorithms, just like humans, are not neutral. Such bias can arise either from software developers' unconscious perspectives, which will affect the algorithm's approach, or from the training data. Given a historically biased dataset, machine learning algorithms are likely to yield discriminative outcomes, particularly for minority groups. Whether sensitive traits, such as gender or ethnicity, are omitted or explicitly considered in the prediction models and inputs, machines learn from a number of attributes in the data that correlate with these traits (Williams et al., 2018).

Thus, building algorithms that perpetuate racial and gender discrimination is certainly a risk, and many examples of discriminating algorithms were reported in the media and academia (Caliskan et al., 2017; O'neil, 2016). Completely eliminating bias from algorithms may be extremely difficult given that they are trained in biased datasets, but recent research showed that developing algorithms that reduce human discrimination exhibited in historical training is feasible (Cowgill, 2019). Other scholars (Kleinberg et al., 2018) went further and showed how algorithms could simultaneously increase accuracy and reduce discrimination. By training a machine-learning algorithm with 758,027 bail decisions and evaluating its performance compared with human judges' performance when predicting bail decisions, the study showed that the algorithm performed better than human judges and that it would jail 41% fewer people of color. Despite the developments on the topic, the question of whether algorithms reduce or augment the perpetration of discrimination remains unsolved and controversial (Beer, 2017).

These and other topics, including transparency regarding how algorithms are designed, knowledge, and trust regarding the ability of algorithms to perform different tasks, will potentially affect how individuals perceive and use algorithms. In most cases, replacing human decisions with an algorithm stumble over ethical barriers or is simply impractical. It is unlikely that human judgments are eliminated from any decision process completely. Thus, algorithms rarely take the role of final decision-makers and are often used as an intermediate source of information for human specialists, who make the final judgments. Thus, it seems reasonable to consider that the successful use of algorithms in decision-making also depends on how individuals react to them. While many scholars believe that algorithms have the potential to enhance decision-making by fixing inherent human limitations and the errors and biases related it, one must not overlook the potential effects of the interactions between human and nonhuman agents in decision-making.

1.1.2 *Human-algorithm interactions in decision-making*

While there is a large body of research on algorithms' accuracy and efficiency, increasing attention from multiple disciplines has also been given to whether humans are willing to accept the involvement of algorithms in decision-making - and, if so, how individuals perceive and react to algorithmic recommendations.

The literature on this topic is not conclusive. Many studies come to different conclusions concerning human responses to algorithms in decision-making. While many studies find that humans are averse to delegating or sharing decision tasks with algorithms, a phenomenon called algorithm aversion, other studies show find that individuals prefer automated advice to human advice, what is called algorithm appreciation (Burton et al., 2020; Jussupow et al., 2020). Some papers even refer to an overreliance on automated systems, leading humans to uncritically trust algorithmic outputs without recognizing their limitations (Dijkstra, 1999; Parasuraman & Riley, 1997). For example, many individuals believe that a machine's decision is completely unbiased, unlike human decisions that could be biased. Thus, there is a risk of overreliance on machine decisions in this regard.

Many possible causes have been discussed in the literature to explain aversion or preference for algorithmic recommendations. Prior research suggests that trust in the advisor is a crucial factor influencing advice-taking. Multiple studies show that people tend to distrust automated systems (Underhaug & Tonning, 2019), which affects the utilization of automated advice (Dietvorst et al., 2015, 2018; Prahl & Van Swol, 2017; Muir, 1987).

Studies have also suggested that aversion to algorithms can vary depending on the type of the task and the perceived abilities required to perform such tasks. Castelo et al. (2019), for example, suggest that algorithms are less trusted for subjective or judgmental tasks, which have no demonstrably best answer, primarily because of a belief that algorithms are ineffective at analyzing subjective data. Other studies suggest that algorithm aversion is particularly noticeable in moral domains (Gogoll & Uhl, 2018). Bigman & Gray (2018) showed that people are more averse to machines on moral tasks because of the perception that algorithms are unable to account for human characteristics. Another example of algorithm aversion in what could be considered a moral domain is given by Longoni et al. (2019). The author showed that users were skeptical to artificial medical intelligence because they perceived that algorithms were less able to deal with unique characteristics and circumstances when compared to humans.

In this context, it is important to consider that before reacting to algorithmic recommendations for a particular task, an individual will have developed beliefs on what an algorithm can do and how it will do it. These beliefs can be acquired by an individual's experience with algorithms in a particular domain or by information picked elsewhere. These beliefs will influence how individuals will react to algorithm advice. Studies show that experience with automated recommendations in a particular domain is associated with a preference for automated systems, while having substantial knowledge in a specific field is associated with an aversion to using automated aids (Montazemi, 1991; Whitecotton, 1996). As a result, those with more experience with algorithm aids and who know less about a specific domain tend to make more accurate decisions when aided by automated systems.

1.1.3 *Human-algorithm interactions in job market decisions*

In this context of growing interest in the effects of interactions between humans and algorithms in decision-making, this part of the dissertation analyzes the effect of using algorithm aids in the context of the job market, a domain where algorithmic recommendations are increasingly being used to make employment decisions. Organizations increasingly use automated systems to assess the performance of employees, filter applications (Horton et al., 2021) and recruit talents (Bogen & Rieke, 2018). A survey conducted by LinkedIn with 9,000 hiring managers and recruiters shows that 76% of them believe artificial intelligence will have a significant impact on recruiting (Spar & Pletenyuk, 2018). Despite the widespread use of algorithms in labor market processes, evidence shows that there are still particularly low levels of awareness that algorithms are often used to preselect candidates in hiring procedures (Grzymek & Puntschuh, 2019).

Emerging work suggests several benefits of introducing algorithmic recommendations in labor market processes, such as increasing the likelihood being interviewed (Barach et al., 2019) and rising hiring levels (Horton, 2017). This part of the dissertation covers an understudied aspect in this literature - the potential differences in the effect of algorithmic and human recommendations. A central objective is to analyze if and how human vs. algorithmic recommendations could lead to different employment outcomes and contribute to a better comprehension of the relationship between recommendations made by automated systems and human reactions to these inputs in the labor market.

1.1.4 *Online Media and Democracy*

The evolution of communication technologies over the last decades has completely remodeled how politicians and citizens interact in democratic countries. More particularly, online media allowed the distribution of political messages through a different structure than traditional media, like radio, newspapers, and TV (Nulty et al., 2016). Through social media, political actors can reach their audiences without the intermediation of third parties and with almost no moderation of content (Engesser et al., 2017). Due to the large user base social networks have acquired recently, and because of its low relative cost compared to traditional media, social media has enabled the transmission of political messages at a speed and an unparalleled scale.

After a period of optimism and hopes that online media could boost democratic participation, create open platforms for public debate and bring politicians and citizens closer, strong concerns have arisen in the last years regarding the potential damages that these technologies may have upon the functioning of democracy (Kruse et al., 2018). On the one hand, online media provides an affordable manner for political communication and fundraising for grassroots movements. They allow audiences to hear opinions that were historically ignored by traditional media. The internet and social media were crucial for messages of movements like the MeToo and Black Lives Matter to be heard and spread (Mundt et al., 2018). On the other, online media has also proven to be a tool of incredible power to develop massive electoral campaigns and change the course of relevant political disputes, raising fears of manipulation of public opinion through the use of digital technologies (Ferrara, 2015).

Over the last years, high-profile investigations raised concerns about the perils of online media abuse in electoral disputes. In 2018, a major controversy erupted when different media outlets reported that Cambridge Analytica, a political consulting firm, relied on data obtained from Facebook through irregular sources to build profiles of voters and develop political communication strategies. According to several reports, Cambridge Analytica collected data from up to 87 million Facebook users, which had been previously gathered through a third-party app that allowed the unwarranted extraction of users' data (Isaak & Hanna, 2018). After a yearlong investigation by the United States (US) Federal Trade Commission, Facebook adopted a significant change of posture in electoral campaigning, implementing a comprehensive list of measures to enhance electoral security and curb the spread of misinformation and political manipulation.

In 2019, an inquiry conducted by the US Senate Intelligence Committee found that a Russian company, the Internet Research Agency (IRA), used social media to target hundreds of millions of American citizens with political messages, trying to sway the 2016 US presidential election in favor of one candidate (Howard et al., 2019). According to the investigation, IRA also developed a sustained propaganda campaign through social networks to exacerbate political divisions in the United States (DiResta et al., 2019).

Nowadays, the relationship between the internet, social media, and democracy continue to be a subject of debate, both in the academia, media, and public policies institutions, leading to the production of a broad range of reports and studies that seek to assess the risks brought by these technologies to the political sphere.

A central topic of analysis in this context is the impact of online media on the distribution of false and misleading information. Due to the decentralized structure of social media and the virtual absence of content moderation or filtering, there is a growing concern that social platforms may be used to create and spread false content, creating opportunities for manipulating public opinion and elections. Over the last years, different public authorities have repeatedly pointed out social media's risks to the democratic process.

A subject that has also gained attention relates to the usage of automated technologies to influence public opinion on online media. Through the employment of the so-called bots, political strategists can gain scale and speed in their communication, artificially creating viral trends through orchestrated interventions in the political discourse (Boichak et al., 2021). The use of "bots" to steer the public debate, through the creation of artificial support for some candidates and the discrediting of others, has been identified in important disputes, like the 2016 US Presidential elections and the 2017 French Presidential dispute (Ferrara, 2020).

Within the broader debate on the impact of the internet on democracy, the second part of this thesis deals with a specific topic: the impact of online media on the polarization in democratic countries.

1.1.5 *Online Media and Political Polarization: the Question of Echo Chambers and Filter Bubbles*

Political polarization has become a central topic of discussion in the twenty-first century. Much of the literature is centered in the United States, where after decades of relative stability and moderation in democratic politics, polarization has grown wider, and substantial challenges to traditional common grounds between political actors have emerged (Pierson & Schickler, 2020; Fiorina et al., 2008). Over the last years, different authors have stressed the implications of the enhancement of polarization in American society, such as the erosion of trust, the exacerbation of internal tensions, and the growing hardships of building basic consensus in the democratic arena (Iyengar et al., 2019).

The concerns with the heightening of political polarization are not limited to the United States, and researchers have identified and analyzed this phenomenon in several countries (Boxell et al., 2020). Nations with different backgrounds and with diverse democratic frameworks are facing the challenges of growing collisions between political actors and the weakening of primary common grounds. Well-established democracies in Europe and consolidating democratic nations in Latin America and Asia are also experiencing rising political polarization (Carothers & O'Donohue, 2019).

In this context, a recurrent explanation for this phenomenon associates the enhancement of polarization with the redesign of public debate brought by the diffusion of the internet and social media access. According to this widespread view, the structure of online spaces would resemble an "echo chamber", where individuals would only be able to access information that reinforces their previous beliefs, distancing themselves from contradictory opinions (Sunstein, 2001, 2018). This would occur due to the filtering systems of online platforms and social media, which seek to create a personalized experience for each user, exposing individuals to the content they already enjoy and support (Pariser, 2011).

These dynamics of online platforms would present users with an unbalanced set of information tailored to reflect their beliefs. It would also create isolated communities of like-minded individuals, reducing their contact with different points of view and divergent ideas. This pattern of communication would lead to a strengthening of preexisting opinions, the amplification of partisanship to extreme levels, and the adoption of extreme positions.

Many scholars have found evidence of echo chamber effects on online platforms. Gillani et al. (2018) identified an impact of twitter use on the reinforcement of group identities and on the decrease of interaction with diverse political views. Researchers also detected that users of social media networks tend to interact more with like-minded individuals than with users with different political beliefs (Yardi & Boyd, 2010). Cinelli et al. (2021) also found that social media users are inclined to consume information that matches their worldviews and create polarized communities and hemophilic clusters around shared beliefs.

These empirical findings show that the use of online platforms could be indeed linked to a movement of amplification of political polarization. Other authors even argue that political leaders use online platforms as a permanent megaphone to attack opponents and disseminate misinformation without filters, rather than a tool of horizontal dialogue with citizens. Social media facilitates the creation of direct connection between politicians and their audience and the development of homogenous networks, which are prone to the flourishing of divisive narratives that isolate social groups (Engesser et al., 2017). This pattern of communication would turn online media into a tool to disseminate radical messages that would foster the broadening of political divides and intolerance in political discourse (Waisbord & Amado, 2017).

However, the literature around this discussion is far from being unanimous. Many studies have questioned the existence of echo chambers in online media. For instance, Hargittai et al. (2008) identified that American political bloggers were frequently exposed to cross-cutting points of view and thus found no evidence of an increasing isolating role of the internet. Another study indicated that, except for some specific partisan groups, individuals communicate with others on Twitter regardless of political ideologies, questioning the emergence of echo chambers or filter bubbles (Bruns, 2017). Researchers also found that users who access ideological news outlets do not avoid news sites with alternative worldviews and suggested that the concerns with selective insulation may be exaggerated (Garrett et al., 2013).

Thus, many questions concerning the link between the growing access to online media and the evolving dynamics of political polarization remain open. While the growing attention to this topic has vastly increased our understanding of the effects of online media on political polarization around the world, most empirical efforts have concentrated on the United States and there is limited evidence on other regions. The second part of this dissertation seeks to provide other perspectives to the American-centric studies on the association between internet and social media usage and political polarization. For this purpose, I examine this relationship in two different multi-party democracies: Brazil and the Netherlands.

1.1.6 *Political Polarization and Internet and Social Media Usage in Brazil and the Netherlands*

Brazil is the fourth-largest democratic country in the world, with more than 150 million voters. Since the end of the military regime and the advent of the 1988 Constitution, Brazil has enjoyed a period of democratic stability and regular partisan competition, with the alternation in power of two major parties and the development of substantial political common grounds. In the last years, however, Brazilian politics have developed a much more contentious and feistier dynamics. The political discourse has been marked by the radicalization of different competing forces in the electoral arena (Layton et al., 2021). In the last general elections, which occurred in 2018, the electorate rejected centrist candidates and strongly supported candidates seen as political outsiders who relied on a radical conservative discourse (Renno, 2020). In a scenario of increasing divisions and weakening of political consensus, different observers have raised concerns regarding "Brazil's descent into destructive polarization" (Stuenkel, 2021).

Among the different causes that may have contributed to these developments, several voices have stressed the impacts of online media and algorithmic technologies on the Brazilian public sphere. The population with access to the internet in Brazil has grown substantially in the last few years. The country is now the fifth nation with the highest number of internet users. Online platforms nowadays play a vital role in the flow of communication within Brazilian society: a recent survey indicated that more than 150 million Brazilians have internet access, of which 64% use online platforms to read news and inform themselves and 72% use social networks (NIC, 2020).

In this context, the connection between the rising polarization in Brazilian politics and the working dynamics of online platforms became a recurrent topic. Much attention has been given to the use of online media to diffuse false information and radical discourses and its association with the polarization of Brazilian society (Ribeiro & Ortellado, 2018). Studies have also analyzed the evolution of polarizing narratives in social networks during the campaigns for the 2018 general elections (Fernandes et al., 2020), the use of bots and automated technologies to interfere with the electoral dispute (Ruediger et al., 2017), and the relationship between computational propaganda, consumption of political information, and rising polarization in Brazil (Machado et al., 2018). As Internet usage grows and online platforms become a crucial space in the country's public sphere, it is clear that the question of whether algorithms affect Brazilian democracy is a problem of utmost relevance (do Nascimento Silva & Silva, 2019).

The Netherlands is also an interesting case for testing the relationship between social media use and political polarization. The country has a well-established digital infrastructure and widespread use of internet and social media among its population

(Jeroense et al., 2021). Furthermore, it is a multiparty democracy, with various parties, in which typically no party assures a majority of votes, hence several parties must cooperate to form a coalition government. Furthermore, it has historically been a highly consensus-oriented democracy (Lijphart et al., 1975).

With an intensely fragmented partisan landscape, and the lack of two alternating governing powers, Dutch voters have relatively weak partisan identities (Bankert et al., 2017). Reiljan (2020) shows that, when measured by partisan polarization, the Netherlands are the least polarized country in Europe. Despite its fragmented political landscape, and low levels of party affective polarization, studies show that Dutch citizens are quite divided in terms of broader ideological positions (having a left- or right-wing political position) (Silva, 2018) and on concrete issue positions. Harteveld (2021) shows that Dutch citizens are particularly divided about cultural issues, such as immigration, religion, gender roles, and in regard to the support of populist radical parties.

Recent research has attributed the upward trend in ideological polarization (left wing) and attitude polarization (divide on cultural and economic issues) to the rise of the populist radical parties in the Netherlands (Silva, 2018). Other scholars show also that the Dutch populist radical right party incited a political disagreement around topic of the immigration (Berkhout et al., 2015). Moreover, the leaders of the populist radical parties the Party for Freedom (PVV) and the Socialist Party (SP), particularly the PVV's Geert Wilders, are known to use social media aggressively and in ways that stimulate the development of echo chambers and radical discourses (Jacobs et al., 2020).

The objective of the second part of this dissertation is to analyse the causal relationship between the internet and social media usage and the levels of political polarization in Brazil and the Netherlands.

1.2 RESEARCH QUESTIONS

This dissertation aims to improve the current understanding of whether and how new technologies and automated tools affect human decision-making and attitudes in the labor market and in the political domain. It provides empirical evidence on (i) the effect of algorithmic recommendations on hiring decisions and the role of pre-existing attitudes towards algorithms and human recommendations, and (ii) the effect of internet and social media use on political polarization. More specifically, the chapters in this dissertation address the following research questions.

1. Are recommendations made by algorithms perceived differently from those made by a human (expert), and do they lead to a different outcome of the hiring process? How firms' previous attitudes and beliefs regarding algorithms and humans might alter how they respond to advice generated by these sources? (Chapter 2)

2. Is there an effect of internet and social media use on affective polarization, e.g., enhanced feelings of hostility towards opposing candidates in Brazil? (Chapter 3)

3. Is there an effect of social media use, and particularly of the frequency and intensity of social media use, on greater deviation from centrist positions on political ideology and attitudes in the Netherlands? (Chapter 4)

1.3 CONTRIBUTION AND DISSERTATION OUTLINE

This dissertation is structured into five chapters. Chapter 1 provides the road-map of the dissertation.

As mentioned above, digital technologies have transformed many aspects of human life and social experience. Automated systems and social platforms have proved to be valuable tools for dealing with information overload in several domains. They can provide individuals with suggestions for information that are likely to be of interest to them and thus assist human decision-making. This thesis explores how technologies and automation through algorithms can influence human decision-making and attitudes in different domains.

Chapter 2 explores whether recommendations made by an algorithm are perceived differently from those made by a human (expert), and potentially lead to a different outcome of the hiring process. Algorithms are increasingly being applied to tasks that were previously reserved for and performed by humans in decision-making processes that range from medical analysis, jail-or-release decisions, forecasting employee performance, and streamlining the screening of applications to job openings. Within this literature strand, one area that has received comparatively little attention is the labor market. Exploring the factors affecting decision-making in hiring processes has long been of key concern for policy-makers, given the consequential nature of such decisions for the labor force structure in the long term.

In Chapter 2, I first ask whether there is a gap in employers' evaluation of candidates recommended by algorithmic systems relative to those recommended by human advisors. I further explore how employers' previous attitudes regarding algorithms and humans might alter how they respond to advice generated by these sources. Results show no overall effect of the CV label. While characteristics such as work experience or speaking English do affect the rating of a candidate's qualification and the hiring interest, the source of recommendation does not. I also do not find an overall effect in job offers or CV retention. Also, I do not find an interaction of the treatment effect with the labor market specific beliefs about algorithms, potentially reflecting the fact that most participants do not believe that one source of recommendations dominates the other in all three dimensions we elicited.

It is rather the general preference for algorithms which seems to matter for the qualification rating of a candidate. On one hand, decision makers with a general preference for algorithmic advice give significantly higher ratings if a candidate was recommended by an algorithm compared to when the candidate was recommended by a human resource expert. On the other hand, decision makers with a general preference for human advice give significantly lower ratings to candidates recommended by algorithms when compared with candidates recommended by humans.

Chapter 3 and 4 investigate the impact of internet and social media usage on political polarization. More specifically, the thesis searches for a causal link between the internet and social media use and (i) the enhancement of feelings of hostility towards opposing candidates and (ii) greater deviation from centrist positions on political ideology and attitudes. Despite the increased scholarly attention to the topic, evidence on the effects of internet and social media use on political polarization remains inconclusive. Empirical attempts to examine causal effects have been limited by identification challenges as they rely on self-reported usage of internet and social media – which typically result in biased outcomes. Many studies document pure correlations and are unable to make claims about causality (Boxell et al., 2017; Liang & Nordin, 2013; Boulianne et al., 2020; Lawrence et al., 2010).

The findings in these chapters suggest that, contrary to what is suggested by the mainstream literature, internet and social media use may not be the cause of rising political polarization across countries. Chapter 3 finds that the enhancement of feelings of hostility towards opposing candidates in Brazil, a phenomenon called affective polarization, cannot be attributed to internet or social media use. Chapter 4 goes further and shows that, in the Netherlands, social media use attenuates rather than drives political polarization, a finding that holds for different measures of social media use - dummy (yes vs. no), intensity (time spent), and frequency of SM use.

Finally, Chapter 5 concludes the dissertation by answering the research questions, identifying its policy implications, and limitations of the dissertation before suggesting avenues for future research.

REFERENCES

Adair, A., Hutchison, N., MacGregor, B., McGreal, S. & Nanthakumaran, N. (1996), 'An analysis of valuation variation in the uk commercial property market: Hager and lord revisited', *Journal of property valuation and Investment* .

Allcott, H. & Gentzkow, M. (2017), 'Social media and fake news in the 2016 election', *Journal of economic perspectives* **31**(2), 211–36.

Anderson, J. M., Kling, J. R. & Stith, K. (1999), 'Measuring interjudge sentencing disparity: Before and after the federal sentencing guidelines', *The Journal of Law and Economics* **42**(S1), 271–308.

Balkin, J. M. (2020), 'How to regulate (and not regulate) social media', *Knight Institute Occasional Paper Series* (1).

Bankert, A., Huddy, L. & Rosema, M. (2017), 'Measuring partisanship as a social identity in multi-party systems', *Political behavior* **39**(1), 103–132.

Barach, M. A., Kaul, A., Leung, M. D. & Lu, S. (2019), 'Strategic redundancy in the use of big data: Evidence from a two-sided labor market', *Strategy Science* **4**(4), 298–322.

Barrat, J. (2013), *Our final invention: Artificial intelligence and the end of the human era*, Macmillan.

Beer, D. (2017), 'The social power of algorithms', *Information, Communication & Society* **20**(1), 1–13.

Berkhout, J., Sudulich, L. & van der Brug, W. (2015), The politicisation of immigration in the netherlands, *in* 'The politicisation of migration', Routledge, pp. 113–134.

Bigman, Y. E. & Gray, K. (2018), 'People are averse to machines making moral decisions', *Cognition* **181**, 21–34.

Bogen, M. & Rieke, A. (2018), 'Help wanted: An examination of hiring algorithms, equity, and bias'.

Boichak, O., Hemsley, J., Jackson, S., Tromble, R. & Tanupabrungsun, S. (2021), 'Not the bots you are looking for: Patterns and effects of orchestrated interventions in the us and german elections', *International Journal of Communication* **15**, 26.

Booker, C., Wyden, R. & Clarke, Y. (2022), Algorithmic accountability act of 2022, Available at https://www.wyden.senate.gov/news/press-releases/wyden-booker-and-clarke-introduce-algorithmic-accountability-act-of-2022-to-require-new-transparency-and-accountability-for-automated-decision-systems.

Boulianne, S., Koc-Michalska, K. & Bimber, B. (2020), 'Right-wing populism, social media and echo chambers in western democracies', *New Media & Society* **22**(4), 683–699.

Boxell, L., Gentzkow, M. & Shapiro, J. M. (2017), 'Greater internet use is not associated with faster growth in political polarization among us demographic groups', *Proceedings of the National Academy of Sciences* **114**(40), 10612–10617.

Boxell, L., Gentzkow, M. & Shapiro, J. M. (2020), 'Cross-country trends in affective polarization', *The Review of Economics and Statistics* pp. 1–60.

Bradshaw, S. & Howard, P. N. (2018), 'The global organization of social media disinformation campaigns', *Journal of International Affairs* **71**(1.5), 23–32.

Bruns, A. (2017), Echo chamber? what echo chamber? reviewing the evidence, *in* '6th Biennial Future of Journalism Conference (FOJ17)'.

Burton, J. W., Stein, M.-K. & Jensen, T. B. (2020), 'A systematic review of algorithm aversion in augmented decision making', *Journal of Behavioral Decision Making* **33**(2), 220–239.

Caliskan, A., Bryson, J. J. & Narayanan, A. (2017), 'Semantics derived automatically from language corpora contain human-like biases', *Science* **356**(6334), 183–186.

Carothers, T. & O'Donohue, A. (2019), *Democracies divided: The global challenge of political polarization*, Brookings Institution Press.

Castaño-Pulgarín, S. A., Suárez-Betancur, N., Vega, L. M. T. & López, H. M. H. (2021), 'Internet, social media and online hate speech. systematic review', *Aggression and Violent Behavior* **58**, 101608.

Castelo, N., Bos, M. W. & Lehmann, D. (2019), 'Let the machine decide: When consumers trust or distrust algorithms', *NIM Marketing Intelligence Review* **11**(2), 24–29.

Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W. & Starnini, M. (2021), 'The echo chamber effect on social media', *Proceedings of the National Academy of Sciences* **118**(9), e2023301118.

Colbert, J. L. (1988), 'Inherent risk: An investigation of auditors' judgments', *Accounting, Organizations and society* **13**(2), 111–121.

Cowgill, B. (2019), 'Bias and productivity in humans and machines', *Columbia Business School Research Paper Forthcoming* .

Dawes, R. M., Faust, D. & Meehl, P. E. (1989), 'Clinical versus actuarial judgment', *Science* **243**(4899), 1668–1674.

Dietvorst, B. J., Simmons, J. P. & Massey, C. (2015), 'Algorithm aversion: People erroneously avoid algorithms after seeing them err.', *Journal of Experimental Psychology: General* **144**(1), 114.

Dietvorst, B. J., Simmons, J. P. & Massey, C. (2018), 'Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them', *Management Science* **64**(3), 1155–1170.

Dijkstra, J. J. (1999), 'User agreement with incorrect expert system advice', *Behaviour Information Technology* **18**(6), 399–411.

DiResta, R., Shaffer, K., Ruppel, B., Sullivan, D., Matney, R., Fox, R., Albright, J. & Johnson, B. (2019), 'The tactics & tropes of the internet research agency'.

do Nascimento Silva, V. & Silva, R. H. A. (2019), Are algorithms affecting the democracy in brazil?, *in* 'International Symposium on Ethical Algorithms'.

Engesser, S., Ernst, N., Esser, F. & Büchel, F. (2017), 'Populism and social media: How politicians spread a fragmented ideology', *Information, communication & society* **20**(8), 1109–1126.

European Commission (2021), *Proposal For a Regulation of The European Parliament and of The Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)*, COM/2021/206 final, 2021/0106(COD).
URL: *https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206*

Fernandes, C. M., Ademir de Oliveira, L., Motta de Campos, M. & Gomes, V. B. (2020), 'Political polarization in the brazilian election campaign for the presidency of brazil in 2018: an analysis of the social network instagram', *Int'l J. Soc. Sci. Stud.* **8**, 119.

Ferrara, E. (2015), '" manipulation and abuse on social media" by emilio ferrara with ching-man au yeung as coordinator', *ACM SIGWEB Newsletter* (Spring), 1–9.

Ferrara, E. (2020), 'Bots, elections, and social media: a brief overview', *Disinformation, Misinformation, and Fake News in Social Media* pp. 95–114.

Fiorina, M. P., Abrams, S. J. et al. (2008), 'Political polarization in the american public', *ANNUAL REVIEW OF POLITICAL SCIENCE* **11**, 563.

Garrett, R. K., Carnahan, D. & Lynch, E. K. (2013), 'A turn toward avoidance? selective exposure to online political information, 2004–2008', *Political Behavior* **35**(1), 113–134.

Gillani, N., Yuan, A., Saveski, M., Vosoughi, S. & Roy, D. (2018), Me, my echo chamber, and i: introspection on social media polarization, *in* 'Proceedings of the 2018 World Wide Web Conference', pp. 823–831.

Gogoll, J. & Uhl, M. (2018), 'Rage against the machine: Automation in the moral domain', *Journal of Behavioral and Experimental Economics* **74**, 97–103.

Grzymek, V. & Puntschuh, M. (2019), 'What europe knows and thinks about algorithms results of a representative survey. bertelsmann stiftung eupinions february 2019'.

Hargittai, E., Gallo, J. & Kane, M. (2008), 'Cross-ideological discussions among conservative and liberal bloggers', *Public Choice* **134**(1), 67–86.

Harteveld, E. (2021), 'Fragmented foes: Affective polarization in the multiparty context of the netherlands', *Electoral Studies* **71**, 102332.

Ho, C.-C. & Tseng, S.-F. (2006), 'From digital divide to digital inequality: the global perspective', *International Journal of Internet and Enterprise Management* **4**(3), 215–227.

Horton, J. J. (2017), 'The effects of algorithmic labor market recommendations: Evidence from a field experiment', *Journal of Labor Economics* **35**(2), 345–385.

Horton, J. J., Vasserman, N. S. & Stanford, G. (2021), 'Job-seekers send too many applications: Experimental evidence and a partial solution'.

Howard, P. N., Ganesh, B., Liotsiou, D., Kelly, J. & François, C. (2019), 'The ira, social media and political polarization in the united states, 2012-2018'.

Isaak, J. & Hanna, M. J. (2018), 'User data privacy: Facebook, cambridge analytica, and privacy protection', *Computer* **51**(8), 56–59.

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N. & Westwood, S. J. (2019), 'The origins and consequences of affective polarization in the united states', *Annual Review of Political Science* **22**, 129–146.

Jacobs, K., Sandberg, L. & Spierings, N. (2020), 'Twitter and facebook: Populists' double-barreled gun?', *New Media & Society* **22**(4), 611–633.

Jeroense, T., Luimers, J., Jacobs, K. & Spierings, N. (2021), 'Political social media use and its linkage to populist and postmaterialist attitudes and vote intention in the netherlands', *European Political Science* .

Jussupow, E., Benbasat, I. & Heinzl, A. (2020), Why are we averse towards algorithms? a comprehensive literature review on algorithm aversion., *in* 'ECIS'.

Kahneman, D. (2011), *Thinking, fast and slow*, Macmillan.

Kahneman, D., Sibony, O. & Sunstein, C. R. (2021), *Noise: A flaw in human judgment*, Little, Brown.

Kahneman, D., Slovic, S. P., Slovic, P. & Tversky, A. (1982), *Judgment under uncertainty: Heuristics and biases*, Cambridge university press.

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J. & Mullainathan, S. (2018), 'Human decisions and machine predictions', *The quarterly journal of economics* **133**(1), 237–293.

Kruse, L. M., Norris, D. R. & Flinchum, J. R. (2018), 'Social media as a public sphere? politics on social media', *The Sociological Quarterly* **59**(1), 62–84.

Lawrence, E., Sides, J. & Farrell, H. (2010), 'Self-segregation or deliberation? blog readership, participation, and polarization in american politics', *Perspectives on Politics* **8**(1), 141–157.

Layton, M. L., Smith, A. E., Moseley, M. W. & Cohen, M. J. (2021), 'Demographic polarization and the rise of the far right: Brazil's 2018 presidential election', *Research & Politics* **8**(1), 2053168021990204.

Liang, C.-Y. & Nordin, M. (2013), 'The internet, news consumption, and political attitudes–evidence for sweden', *The BE Journal of Economic Analysis & Policy* **13**(2), 1071–1093.

Lijphart, A. et al. (1975), *The politics of accommodation: Pluralism and democracy in the Netherlands*, Vol. 142, Univ of California Press.

Longoni, C., Bonezzi, A. & Morewedge, C. K. (2019), 'Resistance to medical artificial intelligence', *Journal of Consumer Research* **46**(4), 629–650.

Machado, C., Kira, B., Hirsch, G., Marchal, N., Kollanyi, B., Howard, P. N., Lederer, T. & Barash, V. (2018), 'News and political information consumption in brazil: Mapping the first round of the 2018 brazilian presidential election on twitter', *The computational propaganda project. Algorithms, automation and digital politics. https://comprop. oii. ox. ac. uk/research/brazil2018* .

Meehl, P. E. (1954), 'Clinical versus statistical prediction: A theoretical analysis and a review of the evidence.'.

Montazemi, A. R. (1991), 'The impact of experience on the design of user interface', *International journal of man-machine studies* **34**(5), 731–749.

Muir, B. M. (1987), 'Trust between humans and machines, and the design of decision aids', *International Journal of Man-Machine Studies* **27**(5-6), 527–539.

Mundt, M., Ross, K. & Burnett, C. M. (2018), 'Scaling social movements through social media: The case of black lives matter', *Social Media+ Society* **4**(4), 2056305118807911.

NIC (2020), 'Ict households survey – executive summary'.

Nulty, P., Theocharis, Y., Popa, S. A., Parnet, O. & Benoit, K. (2016), 'Social media and political communication in the 2014 elections to the european parliament', *Electoral studies* **44**, 429–444.

O'neil, C. (2016), *Weapons of math destruction: How big data increases inequality and threatens democracy*, Broadway books.

Parasuraman, R. & Riley, V. (1997), 'Humans and automation: Use, misuse, disuse, abuse', *Human Factors* **39**(2), 230–253.

Pariser, E. (2011), *The Filter Bubble: What The Internet Is Hiding From You*, Penguin Books Limited.

Pierson, P. & Schickler, E. (2020), 'Madison's constitution under stress: A developmental analysis of political polarization', *Annual Review of Political Science* **23**, 37–58.

Prahl, A. & Van Swol, L. (2017), 'Understanding algorithm aversion: When is advice from automation discounted?', *Journal of Forecasting* **36**(6), 691–702.

Reiljan, A. (2020), ''fear and loathing across party lines'(also) in europe: Affective polarisation in european party systems', *European journal of political research* **59**(2), 376–396.

Renno, L. R. (2020), 'The bolsonaro voter: issue positions and vote choice in the 2018 brazilian presidential elections', *Latin American Politics and Society* **62**(4), 1–23.

Ribeiro, M. M. & Ortellado, P. (2018), 'Fake news: What it is and how to deal with it', *International Journal on Human Rights* **15**(27), 69–81.

Ruediger, M. A. et al. (2017), 'Bots, social networks and politics in brazil: a study on illegitimate interferences with the public debate on the web, risks to the democracy and the 2018 elections'.

Shin, D. & Park, Y. J. (2019), 'Role of fairness, accountability, and transparency in algorithmic affordance', *Computers in Human Behavior* **98**, 277–284.

Shu, K., Bhattacharjee, A., Alatawi, F., Nazer, T. H., Ding, K., Karami, M. & Liu, H. (2020), 'Combating disinformation in a social media age', *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **10**(6), e1385.

Silva, B. C. (2018), 'Populist radical right parties and mass polarization in the netherlands', *European Political Science Review* **10**(2), 219–244.

Spar, B. & Pletenyuk, I. (2018), Global recruiting trends, Technical report.

Stuenkel, O. (2021), 'Brazil's polarization and democratic risks', *Carnegie Endowment for International Peace* .
URL: *https://carnegieendowment.org/2021/02/17/brazil-s-polarization-and-democratic-risks-pub-83783*

Sunstein, C. R. (2001), *Republic. com*, Princeton University Press.

Sunstein, C. R. (2018), # *Republic: Divided democracy in the age of social media*, Princeton University Press.

Tambini, D. (2018), 'Social media power and election legitimacy'.

Taylor, R. L. & Wilsted, W. D. (1974), 'Capturing judgment policies: A field study of performance appraisal', *Academy of Management Journal* **17**(3), 440–449.

Underhaug, M. & Tonning, H. (2019), In bots we (dis) trust?, Master's thesis, University of Stavanger, Norway.

Vosoughi, S., Roy, D. & Aral, S. (2018), 'The spread of true and false news online', *Science* **359**(6380), 1146–1151.

Waisbord, S. & Amado, A. (2017), 'Populist communication by digital means: presidential twitter in latin america', *Information, communication & society* **20**(9), 1330–1346.

Whitecotton, S. M. (1996), 'The effects of experience and a decision aid on the slope, scatter, and bias of earnings forecasts', *Organizational Behavior and Human Decision Processes* **66**(1), 111–121.

Williams, B. A., Brooks, C. F. & Shmargad, Y. (2018), 'How algorithms discriminate based on data they lack: Challenges, solutions, and policy implications', *Journal of Information Policy* **8**(1), 78–115.

Yardi, S. & Boyd, D. (2010), 'Dynamic debates: An analysis of group polarization over time on twitter', *Bulletin of science, technology & society* **30**(5), 316–327.

2

HUMAN VS ALGORITHMIC RECOMMENDATIONS IN THE LABOR MARKET: EVIDENCE FROM A FIELD EXPERIMENT

ABSTRACT

Especially in the sourcing and screening of potential candidates, algorithms are increasingly working alongside human HR staff or replacing them altogether. In view of findings of "algorithm aversion" or "algorithm appreciation" in other domains, the question arises whether the recommendations made by an algorithm are perceived differently from those made by a human (expert), and potentially lead to a different outcome of the hiring process. We study this question in a preregistered field experiment with law firms. Specifically, we test whether there is observable difference in employers' evaluation of candidates recommended by algorithms relative to those recommended by human advisors. We also elicit preexisting attitudes and beliefs about certain characteristics of algorithms in a labor market context and a general preference for algorithmic vs human advice. This allows us to investigate whether and how potential differences in the evaluation of candidates depending on their label are related to these individual beliefs and preferences. Results show no overall difference in employer's response to resumes recommended by algorithms and humans. In the analysis of heterogeneity of preexisting attitudes and beliefs towards algorithms, we do not find an interaction of the treatment effect with the labor market specific beliefs about algorithms. It is rather the general preference for algorithms which seems to matter for the qualification rating of a candidate. Decision-makers with a general preference for algorithmic advice also give significantly higher ratings if a candidate was recommended by an algorithm compared to when the candidate was recommended by a human resource expert.

JEL Classification: C93, M51, M50, O33
Keywords: Algorithms; Artificial Intelligence; Field Experiment; Firm Employment Decisions; Labor Market; Personnel Economics; Hiring

2.1 INTRODUCTION

Algorithms are becoming ever more important in many domains of life. They are increasingly being applied to tasks that were previously reserved for and performed by humans in decision-making processes that range from medical analysis, jail-or-release decisions, forecasting employee performance and streamlining the screening of applications to job openings.

While there is considerable literature exploring algorithm efficiency in performing such tasks, a burgeoning strand of research is now focused on investigating the human aspects of algorithms, i.e., acceptance or rejection of algorithmic generated insights by individual users of decision aids. Findings of such interdisciplinary research remain inconclusive, suggesting that the acceptance of algorithmic advice depends on aspects of the decision environment.

Within this literature strand, one area that has received comparatively little attention is the labor market. Exploring the factors affecting decision-making in hiring processes has long been of key concern for policy-makers, given the consequential nature of such decisions for the labor force structure in the long term. Even minor aspects, such as the sequencing of candidates evaluated in a hiring process, can have substantial impact in the evaluation of candidates (Radbruch & Schiprowski, 2020).

The adoption of algorithms in various stages of the hiring process is now widespread. The technology is widely used for streamlining the screening of applications to job openings in online environments (Horton et al., 2021) and for recruiting and tracking employees' performance in influential firms such as Google, Microsoft, and SAP (Walker, 2012). In a recent industry survey (Spar & Pletenyuk, 2018), 76% of respondents stated that artificial intelligence will have a significant impact on recruiting. The CEO of ZipRecruiter, a large platform matching job seekers and employers, estimates that at least three-quarters of all CVs submitted for job positions in the US are screened by algorithms (Schellmann, 2022).

As the adoption of algorithms in hiring expands, its effects on the labor market also started to be of concern to policy-makers. Recent regulatory proposals in the United States (US) and the European Union (EU) call for algorithmic impact assessments of employment decisions. In the United States, the "Algorithmic Accountability Act of 2022" requires impact assessments of automated decisions or judgments that have any significant effect on employment (Booker et al., 2022). In the EU, the Regulation on Artificial Intelligence (AI) also calls for "conformity assessments" of AI systems used for screening or filtering applications in employment decisions (European Commission, 2021).

Despite the prevalence of algorithms in hiring decisions and the relevance of its effects in the labor market in the coming years, few are the studies that have explored it. In view of findings of "algorithm aversion" or "algorithm appreciation" in other domains, the question arises whether the recommendations made by an algorithm are perceived differently from those made by a human (expert), and potentially lead to a different outcome of the hiring process.

We study this question in a preregistered field experiment with law firms. Field experiments, including resume audit studies, have become traditional methods to explore factors affecting employer's decisions in the hiring process and have generated robust findings on how employers respond to resumes characteristics.[1] Specifically, we test whether there is an observable difference in employers' evaluation of candidates recommended by algorithms relative to those recommended by human experts. We further explore how employers' previous attitudes regarding algorithms and humans might alter how they respond to advice generated by these sources.

The setting of the experiment is a big job fair for undergraduate and graduate students in Brazil organized by an educational institution. Students send their CVs to this institution whose HR personnel then decide which students will be matched with which firm for an interview and forward the respective CV to the respective firm. The firm then interviews these candidates at the job fair.

Our intervention affects the labelling of the CVs. We randomized whether potential candidates were presented to the law firms as recommended by a human resource advisor of the institution or an algorithm. Decision makers at the firms were then invited to participate in a survey before and after the job fair. In the survey before the job fair, they were asked to rate the qualification and their interest in hiring each candidate after evaluating their CVs. We also elicit preexisting attitudes and beliefs about certain characteristics of algorithms in a labor market context and a general preference for algorithmic vs human advice (i.e., across all domains of life) in this survey. This allows us to investigate whether and how potential differences in the evaluation of candidates depending on their label are related to these individual beliefs and preferences. In a follow-up survey after the job fair, employers were asked for each candidate whether they had offered a job or an internship to this candidate or kept the resume for future hiring after the interviews.

Results show no overall effect of the CV label. While characteristics such as work experience or speaking English do affect the rating of a candidate's qualification and the hiring interest, the source of recommendation does not. We also do not find an overall effect in our follow-up survey on job offers or CV retention.

Concerning labor market specific beliefs about algorithms, decision makers on average believe that human resource specialists are better at taking non-standard profiles in terms of academic or professional background into account and more trustworthy

1 Resume audit studies are a specific type of field experiment primarily used to test for discriminatory behavior. They explore if recruiters respond differently to identical resumes, with minor differences associated with a treatment. For a comprehensive literature review on field experiments on discrimination, see Bertrand & Duflo (2017). For resume audit studies, see Bertrand & Mullainathan (2004); Jacquemet & Yannelis (2012); Riach & Rich (1991).

than algorithms. However, respondents also believe that human resource specialists are more prone to bias in their selection of candidates. Concerning recommendations in general across all domains of life, our respondents prefer human over algorithmic advice.

In the analysis of heterogeneity in the treatment effect along these dimensions, we do not find an interaction of the treatment effect with the labor market specific beliefs about algorithms, potentially reflecting the fact that most participants do not believe that one source of recommendations dominates the other in all three dimensions we elicited. It is rather the general preference for algorithms which seems to matter for the qualification rating of a candidate. Decision makers with a general preference for algorithmic advice also give significantly higher ratings if a candidate was recommended by an algorithm compared to when the candidate was recommended by a human resource expert. Our findings suggests that firms introducing algorithms into their hiring process should consider that preexisting attitudes concerning algorithms can affect the evaluation of candidates depending on the source of recommendation.

The remainder paper is organized as follows. Section 2 discusses the related literature. Section 3 describes the study's data and research design. Section 4 presents the main results, and section 6 discusses our findings and concludes.

2.2 RELATED LITERATURE

This work relates to two different strands of literature. First, we contribute to research investigating the human side of algorithms, i.e., how individuals perceive algorithmic recommendations more broadly. Research examining individuals' perceptions of automated outputs is not recent.[2] Early studies already reported that people reacted to mathematical or machine problem-solving with skepticism compared to human specialists in medical predictions and forecasting tasks (Dawes et al., 1989; Meehl, 1954). More recently, this skepticism to mathematical and computational approaches has been labeled as algorithm aversion (Dietvorst et al., 2015; Yeomans, 2019). While there is by now a very large literature exploring the human side of algorithms, findings remain inconclusive. Many studies suggest that, depending on aspects of the decision environment, individuals can exhibit anything ranging from extreme aversion to appreciation for algorithms.[3] We contribute to this strand of research in two different ways. First, we explore how individuals perceive algorithm as opposed to human recommendations in an field experiment in the labor market. Second, we investigate the

2 For the purpose of this study, an algorithmic recommendation is considered an umbrella term for related paradigms like automated decision aids, decision support systems, expert systems, decision formulas, computerized aids, and diagnostic aids. Likewise, variations of decision making, judgment, forecasting, and prediction are considered equivalent.

3 For comprehensive literature reviews on algorithm aversion or appreciation, see Burton et al. (2020); Chugunova & Sele (2020); Jussupow et al. (2020).

role of previously suggested channels influencing algorithm acceptance or rejection in hiring decisions.

Burton et al. (2020) emphasized that individuals' perceptions and previous expectations on algorithms and humans might influence how they utilize the recommendation of such intermediary. Rarely, if ever, human decision-makers will make use of an algorithm recommendation without bringing their preexisting perceptions and attitudes regarding what algorithms and humans are capable of doing, given their attributes. What follows is that decision-makers may respond to a recommendation made by algorithms differently than to a recommendation made by humans, even if the recommendation is otherwise identical.

We thus explore how previous attitudes towards algorithms and human advisors might influence the use of recommendations in hiring decisions. There are several aspects of human decision-making for which either an algorithm or a human might be more suited. We draw on the literature to identify relevant dimensions for which firms' previous attitudes towards algorithmic and humans could interact with the evaluation of algorithm and human recommended candidates.

A crucial question is if firms perceive humans as more trustworthy than algorithms at selecting candidates than human resources specialists. Previous studies show that people tend to distrust automated systems (Underhaug & Tonning, 2019; Muir, 1987; Prahl & Van Swol, 2017) and often choose human over algorithm advisors (Dietvorst et al., 2015, 2018). Moreover, trust in algorithms is particularly low for subjective or judgmental tasks, which have no demonstrably best answer (Castelo et al., 2019). A recent survey found that the majority of Americans find the use of algorithms for CV screening unacceptable (Smith, 2018). Trust in algorithms or humans is thus a key aspect potentially influencing the use of algorithmic and human aids to select candidates in hiring procedures.

Another important dimension is the perceived susceptibility of algorithms and humans to bias. Fairness in algorithmic decision-making is a key concern for individuals, particularly when evaluating, selecting and hiring personnel.[4] Algorithms have created hopes for overcoming human advisor biases. In the heuristics and biases literature, algorithms are perceived as a cognitive fix for inherent human limitations of data processing and the errors and biases related to this (Sundar & Nass, 2001; Kahneman et al., 2021). The ability to process data and perform accurate and objective predictions in comparison to human analysis is seen as a motivation for reliance on algorithms. Kleinberg et al. (2020) argue that because of its greater level of specificity when compared to human decision-making, algorithms could potentially enhance the detection of biases and the prevention of discrimination in markets such as labor hiring.

4 For a systematic literature review of discrimination and fairness by algorithmic decision-making in the context of human resources recruitment and development, see Köchling & Wehner (2020).

However, while some assert that algorithms may overcome human bias in human resource decisions, others understand that people may also perceive the decision-making as too simplistic - as if some background or information is not being considered (Newman et al., 2020). Related to this, algorithm aversion is particularly noticeable in moral domains (Gogoll & Uhl, 2018). Bigman & Gray (2018) showed that people are more averse to machines on moral tasks because of the perception that algorithms cannot account for human characteristics. In particular, it has been suggested that algorithms are ineffective at analyzing outliers (Germann & Merkle, 2019) and are less able to deal with unique characteristics and circumstances when compared to humans (Longoni et al., 2019). A straightforward question is thus whether people perceive algorithms as capable of dealing with exceptional cases in resumes when compared to humans.

It is also important take into consideration that although many companies are increasingly using algorithms in the hiring process, there still seems to be particularly low levels of knowledge regarding the use of algorithms for such task. A study in the United Kingdom, Germany, France, Poland, Spain and Italy showed that only 31 percent of the population knew that algorithms are often used to select candidates in hiring processes (Grzymek & Puntschuh, 2019). Such low levels of knowledge prevail regardless of the country. What manifests from these low levels of awareness regarding the use of algorithms on the labor market is that people might be influenced by expectations created by the experience with algorithmic aids in domains that are subject of extensive media coverage and where people are more likely to notice the consequences in their everyday lives.

It is thus important to take a step aside from the labor market and investigate previous attitudes towards algorithms or humans in areas beyond the decision domain, as noted by Burton et al. (2020). We have thus also elicited measures of overall preference for algorithms or human recommendations and investigated how this general preference can influence the use of algorithmic and human aids to select candidates in hiring procedures.

Our paper also contributes to a growing literature exploring the introduction of algorithmic recommendations in the labor market.[5] Many studies examine the effect of introducing automated recommendations on the firm's decision to interview or hire an applicant. Barach et al. (2019) analyses the effect of introducing algorithmic recommendations for improving matching efficiency in various stages of the hiring process, using quasi-experimental methods. Abebe et al. (2017) goes further and explores if job fairs improve employment outcomes using a randomized control trial. They randomize invites to both workers and firms to participate in the job fair, and then an algorithm is used to match them. Horton (2017) conducts an experiment in a online platform and finds that recommendations increase hiring. In a small field experiment,

5 For literature reviews on algorithms in hiring, see Cameron (2020); Bogen & Rieke (2018).

Cowgill (2018) shows that by replacing the decision of a human CV screener by an algorithm, selected candidates were more likely to pass interviews and receive job offers.

All of these studies, however, overlook the role of traditional human recommendations in the labor market. They overlook, in particular, potential differences in the effect of algorithmic and human recommendations. In other words, they all analyze the effect of introducing algorithmic recommendations in the hiring process, by using as a counterfactual no recommendation at all. To address this gap, our work pursues a complementary aspect to this literature, using a randomized control trial to analyse if and how human vs. algorithmic recommendations could lead to different employment outcomes. To the best of our knowledge we are the first to conduct a field experiment to study human vs. algorithmic recommendation in the labor market.

2.3 RESEARCH DESIGN

The study was conducted in conjunction with a large job fair at an educational institution ("the Institute" from now on) in Brasilia, Brazil.[6] Promoted annually since 2017, the job fair is the first initiative of its kind in the Midwest region of Brazil. It is an important opportunity for undergraduate and graduate students to meet with relevant law firms for internship and job interviews. Students send their CVs to this institution whose HR personnel then decide which students will be matched with which firm for an interview and forward the respective CV to the respective firm. The firm then interviews these candidates at the job fair. The institution thus facilitates job interviewing and hiring by matching both sides of the labor market. In the 2020 edition of the job fair, when the data for the study was also collected, 182 students and 42 law firms participated. Figure 1 illustrates a timeline of the experiment.

We employed a within-subject experiment design, in which each law firm was exposed to control and treatment conditions. Our intervention affects the labelling of the CVs. We randomized whether potential candidates were presented to the law firms as recommended by a human resource advisor of the institution or an algorithm. Each firm received CV recommendations labeled as coming from i) human resource advisors of the institute and ii) an algorithm based on previous human decisions. The human resources experts of the Institute decided which candidates were sent to each firm. This implies that we did not change which candidates were sent to each firm, only the labeling. The randomization within each firm and student participating in the study balanced the labels as much as possible.[7] Firms were truthfully told that rec-

6 The experiment was approved by the Ethics Review Committee of Maastricht University Inner City faculties in October 2020 and preregistered at OSF (https://osf.io/v7tr8).

7 The randomization was designed to minimize the absolute sum of the difference between the number of times a student or a firm was assigned (randomly) to treatment and control. Out of the 182 CVs, 180 were labeled as human and algorithm recommended equally often (one time each) and 2 were

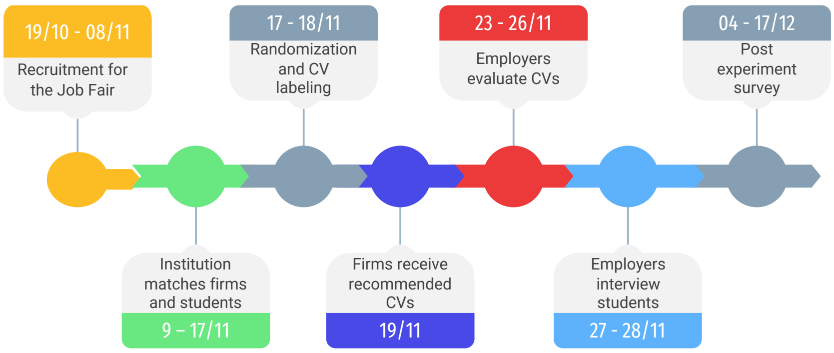

Figure 2.1: Experiment timeline

ommendations either came from a human expert or an algorithm based on previous human decisions.

Decision makers at the firms were invited to participate in surveys before and after the job fair. More specifically, one week before the job fair, law firms received an e-mail with the resumes of the students to be interviewed in the fair and a date to take the survey. In the e-mail, they were provided with a folder where each candidate's resume indicated whether it was picked by the algorithm tool or by the human resource team.[8] The survey before the job fair asked employers to rate the qualification and their interest in hiring each one of the candidates after evaluating their resumes.[9]

We also elicit preexisting attitudes and beliefs about certain characteristics of algorithms in a labor market context and a general preference for algorithmic vs human advice (i.e., across all domains of life) in this survey. This allows us to investigate whether and how potential differences in the evaluation of candidates depending on their label are related to these individual beliefs and preferences. We draw on previous work to identify relevant aspects for which previous attitudes towards algorithmic and humans could interact with the evaluation of algorithm and human recommended candidates.

We explore how trust (Underhaug & Tonning, 2019; Muir, 1987; Dietvorst et al., 2015, 2018), perceived susceptibility to bias (Sundar & Nass, 2001; Kahneman et al.,

labeled as algorithm or human recommended only. Out of the 42 firms participating in the job fair, 16 firms received an equal number of CVs labeled as human or algorithm recommended, 14 received 4 algorithm recommended CVs and 5 human recommended CVs, 11 received 5 algorithm recommended CVs and 4 human recommended CVs and only one received 6 algorithm recommended and 3 human recommended CVs. The codes used for the randomization are provided in the preregistration.

8 See the Appendix for e-mails sent before and after the interviews and for examples of the labeled resumes.

9 See the Appendix for the instructions and screenshots of the experiment.

2021; Köchling & Wehner, 2020), perceived inability to deal with exceptional cases (Germann & Merkle, 2019; Longoni et al., 2019) and general preference could influence the utilization of algorithm vs. human advice. We formulated statements affirming that algorithms or human advisors were better than the other on these four dimensions and asked law firms to express their agreement on a five-point scale ranging from "strongly disagree" to "strongly agree". To avoid acquiescence bias, we produced two versions of each one of the four statements.[10] We presented one of the two versions at random. Also, each statement was presented in a random order.

In a follow-up survey after the job fair, employers were asked for each candidate whether they had offered a job or an internship to this candidate or kept the resume for future hiring after the interviews.

2.3.1 *Data and Descriptive Analysis*

Through our partnership with the Institute, 182 students and 41 out of 42 firms participating in the job fair participated in the study. 41 firms participating in the job fair provided their interest in hiring and their judgment on the qualification of 182 students recommended as potential matches. This allowed us to rely on 357 observations in the first survey. 21 firms participated in the second survey providing 201 observations on 146 different CVs.

We present in table 2.1 and figure 2.2 the descriptive analysis of the outcome variables by treatment (human vs. algorithm recommended). Our first outcome variable is the employer's qualification rating. We measured this variable by asking: "Independent of the fit with your company, how do you judge the overall qualification of this candidate?" on a Likert scale of 0-10. The second outcome variable is employers' interest in hiring. We measure this variable by asking: "How interested would you be in hiring [name]?" on a Likert scale of 0-10. The third outcome variable is a follow-up on the interest in hiring. We measure this variable one week after the interviews by asking whether firms had offered a job/internship to one of the candidates or kept any of the resumes for future hiring on a binary scale. In figures A.3 and A.4 in the appendix, we also provide a description of the average qualification and hiring rating for each firm.

In table 2.2, we present the descriptive statistics of the variables indicating the characteristics of candidates, namely: continuous variable for age, dummy indicating if the candidate has work or internship experience, ordinal variable indicating their proficiency in English, Spanish, German and French (0 - No proficiency, 1 - Elementary, 2 - Limited proficiency, 3 - Professional proficiency, 4 - Full professional proficiency), dummy describing if candidate is a graduate student and dummy for female.

10 See the Appendix for a list of the statements. The two versions are indicated by the numbers 1 and 2.

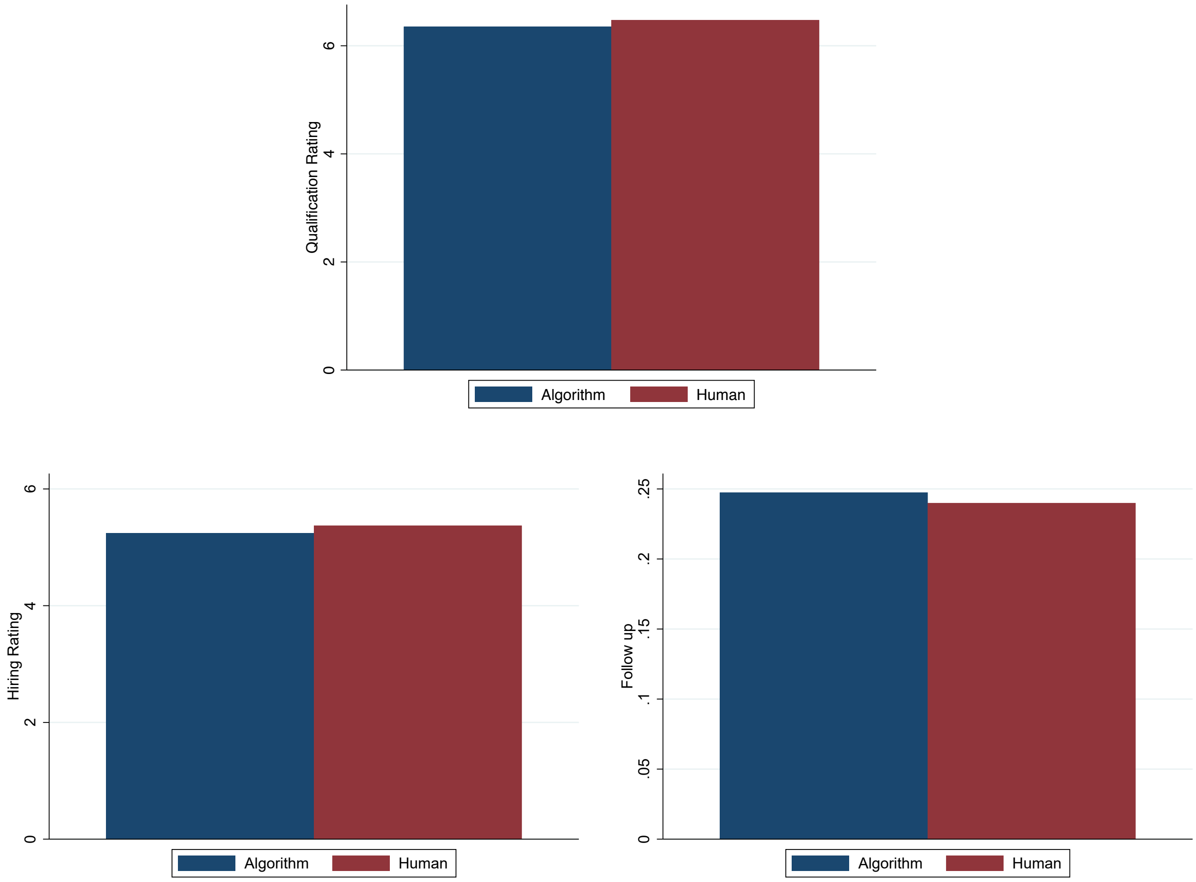

Figure 2.2: Mean outcome variables by treatment (algorithm vs. human recommended)

Table 2.1: Descriptive Statistics - Outcome Variables

Variables	N	mean	sd	min	max
		Human			
Qualification rating	18	6.48	2.19	0	10
Hiring interest	178	5.37	2.66	0	10
Follow-up	100	0.24	0.43	0	1
		Algorithm			
Qualification rating	179	6.35	2.28	0	10
Hiring interest	179	5.24	2.77	0	10
Follow-up	101	0.25	0.43	0	1

Table 2.2: Descriptive statistics - Control Variables

Variables	N	mean	sd	min	max
Age	357	24.18	6.83	17	52
Work experience	357	0.12	0.32	0	1
Internship experience	357	0.41	0.49	0	1
English	357	2.54	1.21	0	4
Spanish	357	1.08	0.97	0	3
German	357	0.18	0.48	0	3
French	357	0.31	0.78	0	4
Graduate	357	0.82	0.39	0	1
Female	357	0.58	0.49	0	1

2.4 RESULTS

The effect of our treatment in our main outcome variables (Y_{ij}) reflect the estimation of equation 2.1. For the effect of the source of recommendation on employers' hiring interest and quality rating, we used linear models. For the effect of the treatment on interest in hiring after the interviews, we used a logit regression. The treatment is the source of recommendation, which is a dummy variable, equaling 1 if the source of recommendation for employer i and resume j is an algorithm and 0 if it is a human advisor.

$$\begin{aligned} Y_{ij} = \beta_0 + \beta_2 Source_{ij} + \beta_2 Age_j + \beta_3 Gender_j + \beta_4 Graduate_j + \beta_5 Experience_j \\ + \beta_6 Languange_j + \mu_i + \epsilon_{ij} \end{aligned} \tag{2.1}$$

To control for students' characteristics, we included: i) a linear measure for the indicated age of the candidate; ii) a gender dummy equal to 1 if the candidate is a female; iii) a dummy equal to 1 if the candidate is a graduate student; iv) a dummy equal to 1 for work experience; and v) an ordinal variable indicating their proficiency in English, Spanish, German and French. To account for any source of constant variation between firms, we included firms' fixed effects μ_i. These fixed effects capture any constant source of variation among firms, such as differences in size, or strictness when evaluating the candidates.

Tables 2.3, 2.4 and 2.5 show the estimations of the models where the outcome variable is the qualification rating, hiring interest, and the job offers or CV retention. The first three columns report OLS regressions with slightly different specifications. The first column includes all candidate characteristics and controls for the order each resume was presented to firms. The second column adds firms' fixed effects. As expected, results are robust to the addition of these controls.

There is no evidence that employers have different qualification ratings, or hiring interest on candidates recommended by algorithms and humans respectively before or after the interviews. As expected, results also show that human capital predicts qualification ratings. On average, employers value work experience on candidate's resumes. In particular, the coefficient on work experience ranges from 0.7–1.6 Likert-scale points. Results also show that employers value knowledge in other languages. As shown in Tables 2.3, 2.4 and 2.5 the coefficients on the variable female are not significantly different from zero, suggesting no evidence of gender discrimination in our data.

2.4.1 *Preexisting Attitudes Towards Algorithms and Humans*

We now provide a descriptive analysis of firms' attitudes towards algorithms and humans recommendations, along the dimensions explained in Section 2. A value of 5 expresses that human recommendations are strongly favored, and a value of 1 that algorithms are strongly favored. Thus, a value of 3 indicates a neutral perception.

Overall, the descriptive analysis of the questions eliciting firms' general preference for algorithms, perception of trust, and capacity to deal with outliers suggests a preference for human recommendations both in the labor market and in general. The findings that emerge from the descriptive analysis are partly aligned with the literature on algorithm aversion. Concerning labor market specific beliefs about algorithms, decision-makers on average believe that human resource specialists are better at selecting non-standard profiles in terms of academic or professional background and more trustworthy than algorithms. Concerning recommendations in general across all domains of life, our respondents prefer human over algorithmic advice.

Table 2.3: OLS Estimations - Qualification rating

	qualification_rating (1)	qualification_rating (2)	qualification_rating (3)
treatment	-0.198	-0.250	-0.242
	(0.19)	(0.16)	(0.16)
age		0.048***	0.046***
		(0.02)	(0.02)
female		0.154	0.080
		(0.17)	(0.17)
graduate		0.430	0.381
		(0.38)	(0.39)
workexperience		0.805*	0.967**
		(0.43)	(0.44)
internshipexperience		1.467***	1.461***
		(0.19)	(0.19)
english		0.360***	0.374***
		(0.08)	(0.08)
spanish		0.131	0.139
		(0.11)	(0.11)
german		-0.192	-0.227
		(0.21)	(0.21)
french		0.249**	0.225**
		(0.11)	(0.11)
order_resume			-0.098***
			(0.03)
_cons	6.514***	3.411***	3.931***
	(0.13)	(0.51)	(0.52)
N	357	357	357

Standard errors in parentheses

* $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

When it comes to the perception of bias, however, respondents believe that human resource specialists are more prone to bias in their selection of candidates. This is in line with recent research that argues that algorithmic decisions are perceived as a cognitive fix for humans' inherent limitations of data processing and biases related to it (Sundar & Nass, 2001; Kahneman et al., 2021).

Table 2.4: OLS Estimations - Hiring interest

	hiring_interest (1)	hiring_interest (2)	hiring_interest (3)
treatment	-0.172	-0.231	-0.228
	(0.24)	(0.22)	(0.22)
age		-0.048**	-0.048**
		(0.02)	(0.02)
female		0.249	0.221
		(0.24)	(0.24)
graduate		-0.361	-0.380
		(0.57)	(0.58)
workexperience		1.484**	1.546**
		(0.64)	(0.65)
internshipexperience		1.491***	1.488***
		(0.27)	(0.27)
english		0.194	0.199
		(0.12)	(0.12)
spanish		0.082	0.085
		(0.13)	(0.13)
german		-0.154	-0.167
		(0.25)	(0.26)
french		0.332**	0.323**
		(0.16)	(0.16)
order_resume			-0.038
			(0.05)
_cons	5.391***	5.054***	5.252***
	(0.17)	(0.81)	(0.84)
N	357	357	357

Standard errors in parentheses

* $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

2.4.2 *Heterogeneous Effects*

Our general results indicate no difference in employers' evaluation of candidates recommended by algorithms relative to those recommended by human experts. Next, we explore if firms' previous attitudes towards algorithms and humans might drive employers to evaluate candidates recommended by algorithms and humans differently.

Table 2.5: Logit Estimations - Job offer or CV retention

	job offer/CV retention (1)	job offer/CV retention (2)	job offer/CV retention (3)
treatment	0.177	0.428	0.478
	(0.50)	(0.63)	(0.69)
age		0.014	0.001
		(0.06)	(0.06)
female		0.703	0.607
		(0.66)	(0.67)
graduate		-0.138	-0.135
		(2.07)	(2.21)
workexperience		0.919	1.423
		(1.70)	(1.81)
internshipexperience		-0.148	0.124
		(0.72)	(0.70)
english		0.747**	0.753*
		(0.34)	(0.38)
spanish		0.340	0.330
		(0.29)	(0.31)
german		-0.613	-0.710
		(0.62)	(0.62)
french		-0.521*	-0.530*
		(0.30)	(0.27)
order_resume			-0.195*
			(0.11)
_cons	-2.161**	-5.068**	-4.020*
	(1.03)	(2.27)	(2.17)
N	113	113	113

Standard errors in parentheses

* $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

We begin with a descriptive analysis of the differences in employers' average qualification ratings depending on their preexisting overall preference toward algorithm or human advice. A value of 1 represents employers who have a strong overall preference for algorithms, and the value of 5 represents employers who have a strong overall preference for human recommendations. A value of 3 indicates that employers are neutral. We can observe in Figure 2.4 that employers who have an overall preference for algorithms – indicated by the values 1 and 2 - give higher ratings to

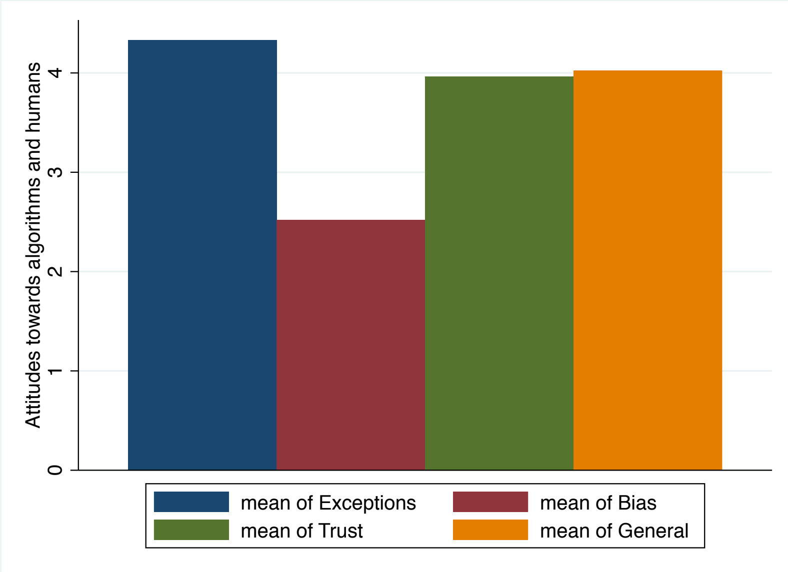

Figure 2.3: Employers prefer humans in most dimensions

Variables	N	mean	sd	min	max
Exceptional Cases	41	4.31	0.81	2	5
Bias	41	2.58	1.16	1	5
Trust	41	3.90	0.73	2	5
General	41	4.07	0.94	1	5

Table 2.6: Descriptive Statistics - Perceptions

candidates recommended by algorithms in comparison to candidates recommended by humans. While employers who have an overall preference for human advice – indicated by the values 4 and 5 – give higher ratings to candidates recommended by human advisors in comparison to candidates recommended by algorithms.

To verify whether firms' preexisting attitudes towards algorithms and humans could drive them to rate algorithm and human recommended candidates, we include interactions between the treatment variable and firms' preexisting attitudes regarding algorithmic and human advice.[11] Results are reported on tables 2.7 and 2.8.

The coefficients representing the effects of employers' perceptions are partly consistent with previous studies testing the relation of preexisting attitudes towards algorithms and humans and algorithm aversion or preference. As expected, perceiving algorithms as less trustworthy, as less able to react to outliers is related to lower rates to algorithm recommended candidates in comparison to human recommended candidates. However, we do not find an interaction of the treatment effect with the labor market specific beliefs about algorithms, potentially reflecting the fact that most par-

11 The interaction terms between the source of recommendation (treatment) and the variables eliciting employers' preexisting attitudes towards algorithms were specified in analysis plan of the preregistration.

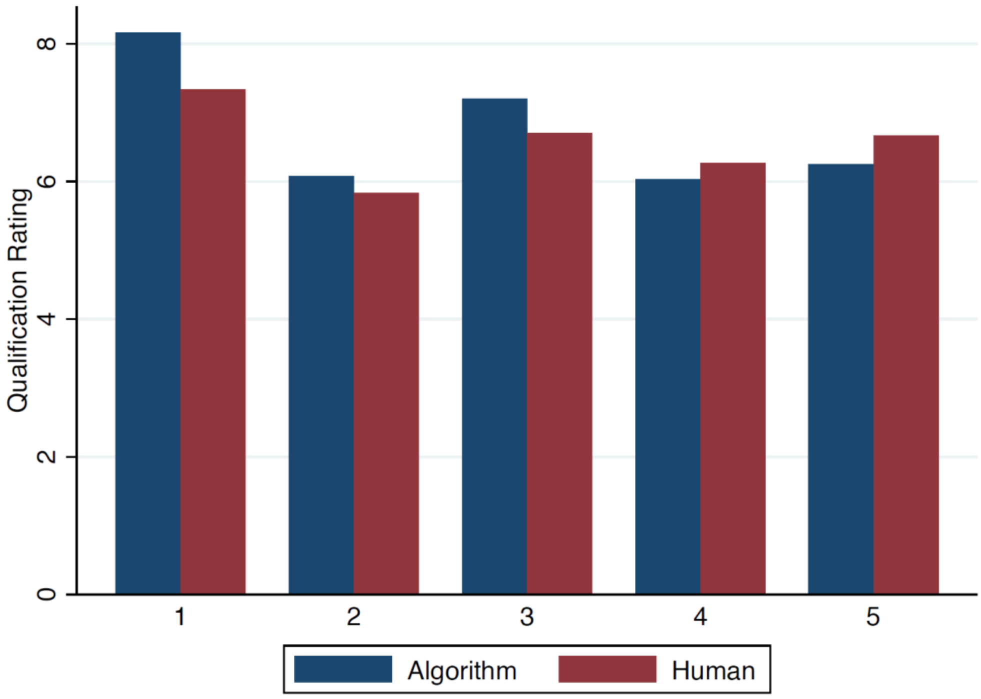

Figure 2.4: Qualification ratings depending on employers preexisting preferences

Table 2.7: OLS Estimations - Qualification rating (heterogeneous effects)

	qualification (1)	qualification (2)	qualification (3)	qualification (4)
treatment	1.636***	0.777	-0.407	0.498
	(0.59)	(0.85)	(0.40)	(0.82)
age	0.052***	0.046***	0.046***	0.047***
	(0.02)	(0.02)	(0.02)	(0.02)
female	0.048	0.063	0.085	0.079
	(0.17)	(0.17)	(0.17)	(0.17)
graduate	0.423	0.400	0.376	0.379
	(0.39)	(0.39)	(0.39)	(0.39)
workexperience	0.815*	0.935**	0.966**	0.977**
	(0.44)	(0.44)	(0.44)	(0.44)
internshipexperience	1.467***	1.472***	1.456***	1.470***
	(0.19)	(0.19)	(0.19)	(0.19)
english	0.391***	0.370***	0.372***	0.375***
	(0.08)	(0.08)	(0.08)	(0.08)
spanish	0.112	0.133	0.137	0.142
	(0.11)	(0.11)	(0.11)	(0.11)
german	-0.215	-0.225	-0.211	-0.231
	(0.21)	(0.21)	(0.21)	(0.21)
french	0.256**	0.223**	0.222**	0.225**
	(0.11)	(0.11)	(0.11)	(0.11)
order_resume	-0.105***	-0.100***	-0.100***	-0.100***
	(0.03)	(0.03)	(0.03)	(0.03)
tr_perception_general	-0.466***			
	(0.14)			
tr_perception_exceptions		-0.235		
		(0.19)		
tr_perception_bias			0.066	
			(0.14)	
tr_perception_trust				-0.187
				(0.20)
_cons	3.828***	3.970***	3.949***	3.911***
	(0.52)	(0.53)	(0.53)	(0.53)
N	357	357	357	357

Table 2.8: OLS Estimations - Hiring rating (heterogeneous effects)

	hiring (1)	hiring (2)	hiring (3)	hiring (4)
treatment	1.239	-1.265	-0.664	-0.734
	(0.93)	(1.23)	(0.58)	(1.17)
age	-0.044*	-0.048**	-0.049**	-0.049**
	(0.02)	(0.02)	(0.02)	(0.02)
female	0.196	0.239	0.234	0.222
	(0.24)	(0.24)	(0.24)	(0.24)
graduate	-0.346	-0.399	-0.392	-0.378
	(0.57)	(0.58)	(0.57)	(0.58)
workexperience	1.427**	1.578**	1.541**	1.539**
	(0.66)	(0.65)	(0.65)	(0.65)
internshipexperience	1.493***	1.477***	1.474***	1.483***
	(0.27)	(0.27)	(0.27)	(0.27)
english	0.212*	0.203*	0.194	0.198
	(0.12)	(0.12)	(0.12)	(0.12)
spanish	0.064	0.092	0.079	0.083
	(0.13)	(0.13)	(0.13)	(0.13)
german	-0.158	-0.169	-0.124	-0.165
	(0.25)	(0.26)	(0.26)	(0.26)
french	0.347**	0.325**	0.315*	0.323**
	(0.16)	(0.16)	(0.16)	(0.16)
order_resume	-0.043	-0.036	-0.041	-0.036
	(0.05)	(0.05)	(0.05)	(0.05)
tr_perception_general	-0.364			
	(0.23)			
tr_perception_exceptions		0.239		
		(0.28)		
tr_perception_bias			0.173	
			(0.21)	
tr_perception_trust				0.128
				(0.31)
_cons	5.172***	5.214***	5.298***	5.266***
	(0.84)	(0.85)	(0.83)	(0.84)
N	357	357	357	357

ticipants do not believe that one source of recommendations dominates the other in all three dimensions we elicited. It is rather the general preference for algorithms which seems to matter for the qualification rating of a candidate. Decision-makers with a general preference for algorithmic advice give significantly higher ratings if a candidate was recommended by an algorithm compared to when the candidate was recommended by a human resource expert.

2.5 CONCLUSION

This chapter explores whether there is difference in the assessment made by employers of job-seekers recommended by algorithms in comparison to candidates recommended by human choice. We also examine if the previous attitudes of recruiters in relation to algorithms solutions and human decision-making affected the reactions to the recommendations provided by these two types of sources.

Our results show no overall effect of the CV label. While characteristics such as work experience or speaking English do affect the rating of a candidate's qualification and the hiring interest, the source of recommendation does not. We also do not find an overall effect in our follow-up survey on job offers or CV retention.

The study did not find a significant difference in the responses of employers to recommendations made by algorithms and humans. This null effect, however, seems to be concealed by substantial heterogeneity on employers' preexisting attitudes and beliefs towards algorithm and human advice. When confronted with identical recommendations produced by algorithms or humans, employers' reactions varied according to their preexisting attitudes towards algorithmic and human advice in areas beyond the labor market.

Scholars well emphasized that decision-makers rarely make use of an algorithm recommendation with a blank slate. When using an algorithmic aid, individuals bring their preexisting perceptions and attitudes regarding what algorithms and humans are capable of doing, given their attributes. Thus, previous attitudes towards algorithms and humans are expected to interact with the evaluation of algorithms and human-recommended candidates. In other words, individuals can respond differently to recommendations generated by algorithms or humans, depending on their previous perceptions of algorithmic and human aid. These perceptions can be based on their experience and knowledge in the decision domain or their experiences and knowledge in other domains. Various aspects through which employers' previous perceptions towards algorithmic and humans could affect the use of algorithmic and human aids are identified in the literature. Many scholars suggested that the perceived inability of algorithms to deal with exceptional cases, susceptibility to bias, and lack of trust in algorithms to select a candidate could influence individuals to respond to an algorithmic recommendation differently than to a human recommendation.

Our findings suggest, however, that none of these perceptions drive employers to evaluate candidates recommended by algorithms and humans differently. Contrary to common assumptions, we find that, when using algorithmic aids in hiring procedures, employers are not influenced by specific reservations on the capacity of an algorithm to select a candidate but rather by a general feeling towards algorithms in spheres other than the labor market. On one hand, decision-makers with a general preference for algorithmic advice give significantly higher ratings if a candidate is recommended by an algorithm compared to when the candidate was recommended by a human resource expert. On the other hand, decision-makers with a general preference for human advice give significantly lower ratings to candidates recommended by algorithms when compared with candidates recommended by humans.

The findings of this paper shed light on an important, albeit overlooked, aspect regarding the widespread use of algorithm solutions in the labor market - the interaction between human perceptions and algorithms. As organizations rely increasingly on algorithms in hiring decisions, it is crucial to understand, correct and prevent difficulties related to effective use of algorithms. While algorithmic governance becomes an ever more present reality, the findings of this paper indicate the relevance of policies aimed at optimizing the algorithm-human relationship. An important issue here is making investments in the enhancement of algorithmic literacy, geared to instruct decision-makers on the rationale of algorithm systems, their strengths, and their limitations, as previously suggested by Burton (2020). There still seems to be particularly low levels of knowledge regarding the use of algorithms in the labor market.[12] Another conceivable policy solution involves the design and implementation of more transparent algorithms solutions. Although transparency often comes as a trade-off with the algorithm's complexity and performance, the disclosure of algorithms into intelligible multistep procedures would afford more opportunities for alignment on human-algorithm decision processes and could be thus crucial for the effective use of algorithms.

12 A study in the United Kingdom, Germany, France, Poland, Spain and Italy showed that only 31 percent of the population knew that algorithms are often used to select candidates in hiring processes (Grzymek & Puntschuh, 2019).

REFERENCES

Abebe, G. T., Caria, S., Fafchamps, M., Falco, P., Franklin, S., Quinn, S. & Shilpi, F. (2017), 'Job fairs: Matching firms and workers in a field experiment in ethiopia', *World Bank Policy Research Working Paper* (8092).

Barach, M. A., Kaul, A., Leung, M. D. & Lu, S. (2019), 'Strategic redundancy in the use of big data: Evidence from a two-sided labor market', *Strategy Science* **4**(4), 298–322.

Bertrand, M. & Duflo, E. (2017), 'Field experiments on discrimination', *Handbook of Economic Field Experiments* **1**, 309–393.

Bertrand, M. & Mullainathan, S. (2004), 'Are emily and greg more employable than lakisha and jamal? a field experiment on labor market discrimination', *American Economic Review* **94**(4), 991–1013.

Bigman, Y. E. & Gray, K. (2018), 'People are averse to machines making moral decisions', *Cognition* **181**, 21–34.

Bogen, M. & Rieke, A. (2018), 'Help wanted: An examination of hiring algorithms, equity, and bias'.

Booker, C., Wyden, R. & Clarke, Y. (2022), Algorithmic accountability act of 2022, Available at https://www.wyden.senate.gov/news/press-releases/wyden-booker-and-clarke-introduce-algorithmic-accountability-act-of-2022-to-require-new-transparency-and-accountability-for-automated-decision-systems.

Burton, J. W., Stein, M.-K. & Jensen, T. B. (2020), 'A systematic review of algorithm aversion in augmented decision making', *Journal of Behavioral Decision Making* **33**(2), 220–239.

Cameron, L. (2020), The rise of algorithmic work: Implications for organizational control and worker autonomy, PhD thesis.

Castelo, N., Bos, M. W. & Lehmann, D. R. (2019), 'Task-dependent algorithm aversion', *Journal of Marketing Research* **56**(5), 809–825.

Chugunova, M. & Sele, D. (2020), 'We and it: An interdisciplinary review of the experimental evidence on human-machine interaction', *Max Planck Institute for Innovation & Competition Research Paper* (20-15).

Cowgill, B. (2018), 'Bias and productivity in humans and algorithms: Theory and evidence from resume screening', *Columbia Business School, Columbia University* **29**.

Dawes, R. M., Faust, D. & Meehl, P. E. (1989), 'Clinical versus actuarial judgment', *Science* **243**(4899), 1668–1674.

Dietvorst, B. J., Simmons, J. P. & Massey, C. (2015), 'Algorithm aversion: People erroneously avoid algorithms after seeing them err.', *Journal of Experimental Psychology:*

General **144**(1), 114.

Dietvorst, B. J., Simmons, J. P. & Massey, C. (2018), 'Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them', *Management Science* **64**(3), 1155–1170.

European Commission (2021), *Proposal For a Regulation of The European Parliament and of The Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)*, COM/2021/206 final, 2021/0106(COD).
URL: *https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206*

Germann, M. & Merkle, C. (2019), 'Algorithm aversion in financial investing', *Available at SSRN 3364850* .

Gogoll, J. & Uhl, M. (2018), 'Rage against the machine: Automation in the moral domain', *Journal of Behavioral and Experimental Economics* **74**, 97–103.

Grzymek, V. & Puntschuh, M. (2019), 'What europe knows and thinks about algorithms results of a representative survey. bertelsmann stiftung eupinions february 2019'.

Horton, J. J. (2017), 'The effects of algorithmic labor market recommendations: Evidence from a field experiment', *Journal of Labor Economics* **35**(2), 345–385.

Horton, J. J., Vasserman, N. S. & Stanford, G. (2021), 'Job-seekers send too many applications: Experimental evidence and a partial solution'.

Jacquemet, N. & Yannelis, C. (2012), 'Indiscriminate discrimination: A correspondence test for ethnic homophily in the chicago labor market', *Labour Economics* **19**(6), 824–832.

Jussupow, E., Benbasat, I. & Heinzl, A. (2020), Why are we averse towards algorithms? a comprehensive literature review on algorithm aversion., *in* 'ECIS'.

Kahneman, D., Sibony, O. & Sunstein, C. R. (2021), *Noise: A flaw in human judgment*, Little, Brown.

Kleinberg, J., Ludwig, J., Mullainathan, S. & Sunstein, C. R. (2020), 'Algorithms as discrimination detectors', *Proceedings of the National Academy of Sciences* **117**(48), 30096–30100.

Köchling, A. & Wehner, M. C. (2020), 'Discriminated by an algorithm: a systematic review of discrimination and fairness by algorithmic decision-making in the context of hr recruitment and hr development', *Business Research* pp. 1–54.

Longoni, C., Bonezzi, A. & Morewedge, C. K. (2019), 'Resistance to medical artificial intelligence', *Journal of Consumer Research* **46**(4), 629–650.

Meehl, P. E. (1954), 'Clinical versus statistical prediction: A theoretical analysis and a review of the evidence.'.

Muir, B. M. (1987), 'Trust between humans and machines, and the design of decision aids', *International Journal of Man-Machine Studies* **27**(5-6), 527–539.

Newman, D. T., Fast, N. J. & Harmon, D. J. (2020), 'When eliminating bias isn't fair: Algorithmic reductionism and procedural justice in human resource decisions', *Organizational Behavior and Human Decision Processes* **160**, 149–167.

Prahl, A. & Van Swol, L. (2017), 'Understanding algorithm aversion: When is advice from automation discounted?', *Journal of Forecasting* **36**(6), 691–702.

Radbruch, J. & Schiprowski, A. (2020), Interview sequences and the formation of subjective assessments, Technical report, ECONtribute Discussion Paper.

Riach, P. A. & Rich, J. (1991), 'Testing for racial discrimination in the labour market', *Cambridge Journal of Economics* **15**(3), 239–256.

Schellmann, H. (2022), 'Finding it hard to get a new job? robot recruiters might be to blame', *The Guardian* .
URL: *https://www.theguardian.com/us-news/2022/may/11/artitifical-intelligence-job-applications-screen-robot-recruiters?CMP=share$_b$tn$_l$ink*

Smith, A. (2018), 'Public attitudes toward computer algorithms', *Pew Research Center* **16**.

Spar, B. & Pletenyuk, I. (2018), Global recruiting trends, Technical report.

Sundar, S. S. & Nass, C. (2001), 'Conceptualizing sources in online news', *Journal of Communication* **51**(1), 52–72.

Underhaug, M. & Tonning, H. (2019), In bots we (dis) trust?, Master's thesis, University of Stavanger, Norway.

Walker, J. (2012), 'Meet the new boss: big data. wall street journal'. Last accessed August 2021.
URL: *https://online.wsj.com/article/SB10000872396390443890304578006252019616768.html*

Yeomans, M. (2019), 'Some hedonic consequences of perspective-taking in recommending', *Journal of Consumer Psychology* **29**(1), 22–38.

A

APPENDIX

We provide in this appendix the details of the design of our experiment, including recruitment materials and experimental instructions and questions.

A.0.1 *Recruitment Materials*

The institution sent e-mails recruiting firms participating in the job fair to take part in the study, as shown in the screenshot (Figure 4). The e-mail thanked the firm for participating in the job fair and proposed a date and time for participating in a quick study. In the e-mail, firms could click on a link that would give them access to a folder where they could find the CV of the candidates, labeled either as algorithm or human recommended.

Prezado
Agradecemos pela participação na **IV Job Fair**.

Solicitamos ao **responsável pela entrevista** que nos confirme o horário abaixo para darmos mais detalhes a respeito da edição da **IV Job Fair** e para realizarmos uma **pequena pesquisa**. A conversa de orientação será realizada pelo zoom e tem a duração prevista de 15 minutos.

📅 **Data:** 2020/11/23
⏰ **Horário:** 14h30 às 14h45
🖱️ **Link de Acesso da Reunião**: https:/ edu-br.zoom.us/j/97214891988

Compartilhamos os currículos dos candidatos selecionados para entrevistas e os horários agendados.

Clique no link abaixo e acesse os currículos:
https://drive.google.com/drive/folders/1gu9DZ_EnuhqgkKHwgkam2D6FFbB4D5uD?usp=sharing

Cordialmente,

Figure A.1: Recruiting e-mail

A.0.2 *Experimental Instructions and Questions*

The experiment was conducted by the platform zoom. The researcher would share the screen with the potential employers and read the instructions and questions, making sure the participants understood all stages. Bellow we report the translated version of the experiment.

Introduction. This year, the institute is participating in a study to understand what firms' value in hiring. We assure you that no one outside the research team will have access to this information and that the answers will be anonymized for the analysis. This means the answers will be linked to a random number and no longer to your firm. In any case, we inform you that you may revoke the use of the data in this form two weeks after completing the survey, simply by sending an email to pesquisajobfair@xxx.edu.br. If you have any questions, contact the researchers at the same email.

Part A.

At this point in the experiment, the recommended CVs were displayed to the potential employers. The CVs had the following information of each candidate: name, telephone number, address, age, education, work experience, knowledge on foreign languages and areas of interest. After the display of each CV, the potential employers was asked the following questions:

1. How interested would you be in hiring this candidate?
0 = not interested at all; 10 = very interested

2. Independent of the fit with your company, how do you judge the overall qualification of this candidate?
0 = not interested at all; 10 = very interested

Part B.

After evaluating all of the CVs recommended for each firm, the potential employers were asked questions regarding their perceptions on algorithm and human recommendations. We report here two versions of each question. One of the two versions was selected at random to each one of the participants.

Finally, we would like to ask you to express if you totally agree, agree, do not agree or disagree, disagree or totally disagree with the following statements regarding the selection of *candidates in the job market.*

1.1 Human resources specialists are better able to react to exceptional cases related to the academic or professional background than algorithms.
1.2 Algorithms are better able to react to exceptional cases related to the academic or professional background than human resources specialists.

2.1 Human resources specialists are less prone to bias at selecting candidates than algorithms.
2.2 Algorithms are less prone to bias at selecting candidates than human resources specialists.

3.1 Algorithms are more trustworthy at selecting candidates than human resources specialists.
3.2 Human resources specialists are more trustworthy at selecting candidates than algorithms.

For the last question, we would like you to answer taking into account recommendations in general, such as financial advice or shopping.

4.1 In general, I prefer to receive recommendations from algorithms than human experts.
4.2 In general, I prefer to receive recommendations from human experts than algorithms.

Part C.

One week after the job fair, the potential employers received the following message by e-mail:

Dear [firm], thank you for participating in the Job Fair. Could you please inform if you offered a job/internship to one of the candidates or kept one of the resumes for future hiring? If so, which resume?

A.0.3 *Screenshots*

We provide here the screenshots of all experimental instructions and questions. For illustrative purposes, we report, only two of the CVs recommended to a firm.

(a)

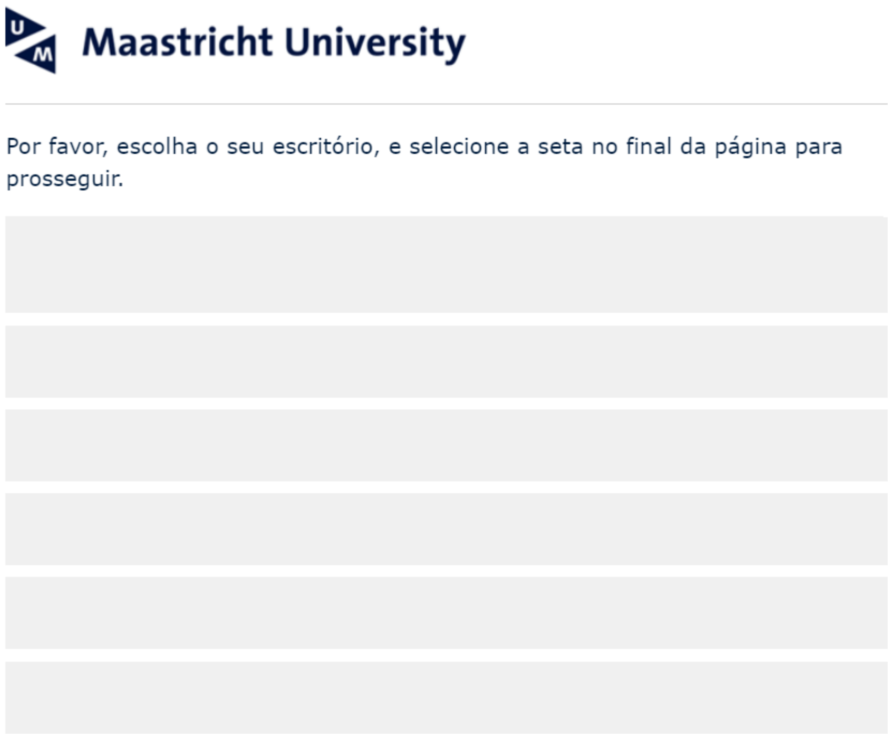

Figure A.2: Screenshots Qualtrics

(b)

Maastricht University

Este ano o está participando de uma estudo acadêmico sobre as motivações dos empregadores no momento da contratação. Para isso, gostaríamos de solicitar que você faca uma rápida avaliação a respeito dos currículos selecionados para o seu escritório.

Lhe asseguramos que ninguém fora da equipe de pesquisa terá acesso a essas informações e que será realizada a anonimização das respostas para análise. Isso significa as respostas serão vinculadas a um número aleatório e perderão a possibilidade de vinculação ao seu escritório. Em qualquer caso, o(a) sr(a) pode revogar o uso dos dados neste formulário até duas semanas após o término da job fair, simplesmente entrando em contato comigo ou enviando um e-mail para pesquisajobfair@ Podemos prosseguir?

Sim

Nao

(c)

Maastricht University

O(a) sr(a) realizará as entrevistas com os candidatos?

Sim

Não

→

(d)

Maastricht University

Considere as informações sobre FERNANDA

SELECIONADO PELA EQUIPE DE RH

FERNANDA

Celular:
E-mail:
Endereço residencial:
Idade:
Escolaridade:

Possui outra formação?

Possui cursos de extensão?

Possui experiência prévia relevante na área?

Áreas de Interesse:
Direito Administrativo
Direito de Família
Direito Civil
Direito Constitucional

Idiomas
Inglês - Fluente
Espanhol - Básico
Alemão -
Francês -
Outro idioma / nível -

(e)

Maastricht University

Independentemente da adequação do candidato ao seu escritório, como você avalia a qualificação de FERNANDA de 0 a 10?

Nada qualificado — Muito qualificado

0 1 2 3 4 5 6 7 8 9 10

Com base nas informações do currículo, quão interessado você estaria em contratar FERNANDA para uma potencial vaga de 0 a 10?

Nada interessado — Muito interessado

0 1 2 3 4 5 6 7 8 9 10

→

(f)

Maastricht University

Considere as informações sobre GABRIEL

SELECIONADO PELO ALGORITMO

GABRIEL

Celular:
E-mail:
Endereço residencial:
Idade:
Escolaridade:

Possui outra formação?

Possui cursos de extensão?

Possui experiência prévia relevante na área?

Áreas de Interesse:
Direitos Humanos
Direito Internacional
Direito Penal
Direito Constitucional
Direito Empresarial

Idiomas
Inglês - Avançado
Espanhol -
Alemão -
Francês -
Outro idioma / nível -

(g)

Maastricht University

Independentemente da adequação do candidato ao seu escritório, como você avalia a qualificação de GABRIEL de 0 a 10?

Nada qualificado — Muito qualificado

0 1 2 3 4 5 6 7 8 9 10

Com base nas informações do currículo, quão interessado você estaria em contratar GABRIEL para uma potencial vaga de 0 a 10?

Nada interessado — Muito interessado

0 1 2 3 4 5 6 7 8 9 10

→

(h)

Maastricht University

Por último, pensando no contexto da seleção de candidatos, gostaria que o(a) sr(a) dissesse se concorda totalmente, concorda parcialmente, nem concorda, nem discorda, discorda parcialmente ou discorda totalmente com as seguintes afirmações.

→

(i)

Maastricht University

A3

	Concordar totalmente	Concordar parcialmente	Nem concordar nem discordar	Discordar parcialmente	Discordar totalmente
Algoritmos são mais confiáveis ao selecionar candidatos do que especialistas em recursos humanos.	○	○	○	○	○

A2

	Concordo totalmente	Concordo parcialmente	Nem concordo nem discordo	Discordo parcialmente	Discordo totalmente
Especialistas em recursos humanos são menos sujeitos a vieses ao selecionar candidatos do que os algoritmos.	○	○	○	○	○

(j)

A1

	Concordo totalmente	Concordo parcialmente	Nem concordo nem discordo	Discordo parcialmente	Discordo totalmente
Especialistas em recursos humanos são mais capazes de reagir a casos excepcionais relacionados ao histórico profissional ou acadêmico, quando comparados a algoritmos.	○	○	○	○	○

→

(k)

Maastricht University

Para última frase, gostaríamos o(a) sr(a) dissesse se concorda totalmente, concorda parcialmente, nem concorda, nem discorda, discorda parcialmente ou discorda totalmente com a seguinte frase, levando em consideração recomendações em geral - como para compras ou investimentos financeiros - e não apenas para seleção de candidatos no mercado de trabalho.

	Concordar totalmente	Concordar parcialmente	Nem concordar nem discordar	Discordar parcialmente	Discordar totalmente
Em geral, prefiro receber recomendações de especialistas humanos do que de algoritmos.	○	○	○	○	○

→

(l)

Maastricht University

Muito obrigada pela sua participação na pesquisa. Por favor fique à vontade para nos dar qualquer feedback ou fazer qualquer comentário.

→

(m)

Maastricht University

Agradecemos o tempo que você dedicou respondendo a esta pesquisa.
Sua resposta foi registrada.

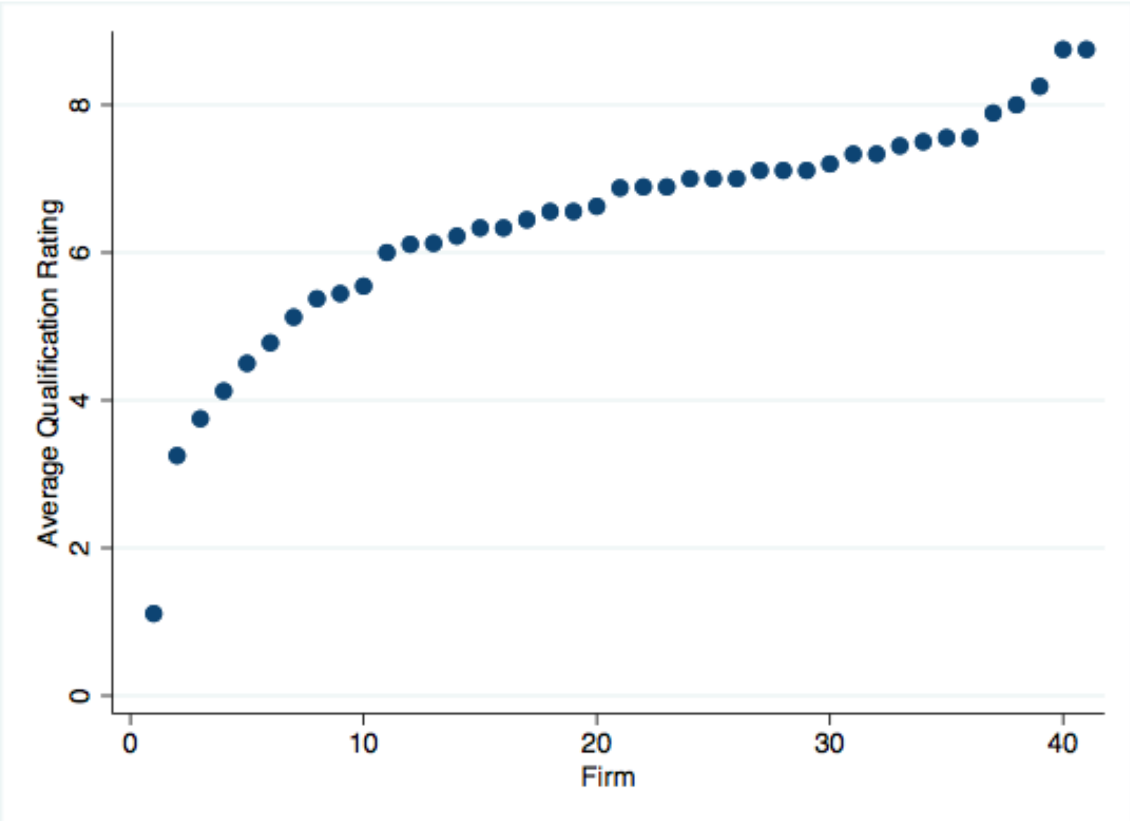

Figure A.3: Average qualification rating by firm

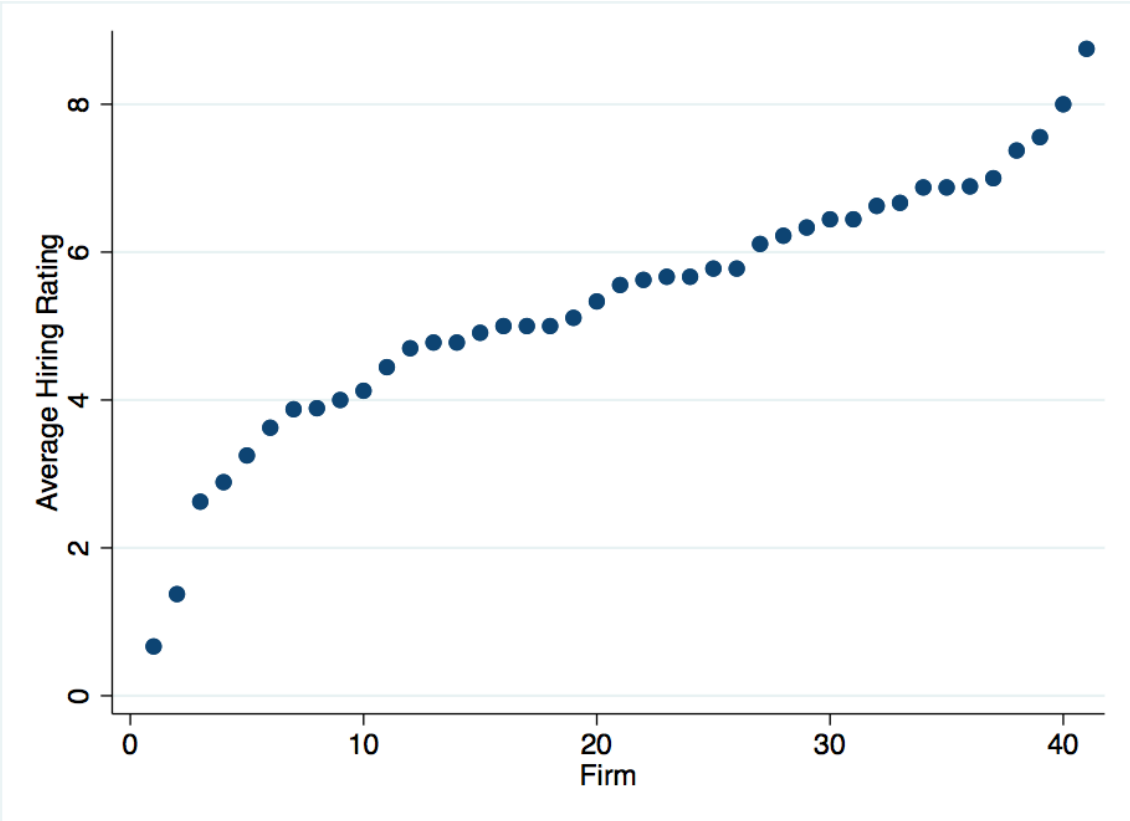

Figure A.4: Average hiring rating by firm

3

THE IMPACT OF INTERNET AND SOCIAL MEDIA USE ON AFFECTIVE POLARIZATION: EVIDENCE FROM BRAZIL

ABSTRACT

Influential scholars have pointed to the internet and social media as a reason for the recent political divide in many countries. Greater exposure to imbalanced information in these environments could reinforce previous political positions leading voters to develop more extreme positions or greater animosity towards candidates of the opposing political group, a phenomenon known as affective polarization. This study investigates the impact of internet and social media use on affective polarization in Brazil. We employ an instrumental variable approach using exogenous infrastructure variations to identify the internet's impact on political behavior. We do not find evidence that access to this new media environment explains affective polarization within the population under study.

JEL Classification: D12, D72, L82, L86
Keywords: affective polarization, broadband internet, social media, Brazil

An earlier version of the chapter has been published as UNU-MERIT working paper #2021-032.

3.1 INTRODUCTION

The debate on the effects of technological tools on ideological fragmentation is not a new one. As early as the Internet was invented, Van Alstyne & Brynjolfsson (1996) already expressed concerns about information technologies leading to "cyber-balkanization". The Internet would "shrink geographic distances and facilitate information exchange", making it easier for like-minded individuals to associate with one another and strengthen communities with a common ideology.

Another influential scholar (Sunstein, 2001) has argued that the Internet would make it easier for individuals to isolate themselves into like-minded groups, which would lead to the creation of "echo chambers". Greater exposure to imbalanced information within like-minded groups would strengthen one's own confidence in a preferred ideological identity and increase the distance from opposing ideological views.[1] More recently, the concern has gained new momentum by the diffusion of social media. These online communication tools could exacerbate selective exposure by filtering out certain information with algorithmic rankings (Flaxman et al., 2013; Pariser, 2011; Bakshy et al., 2015; Sunstein, 2018). Some scholars have also investigated the existence of homophily - the tendency of like-minded individuals to interact with one another - on online social media (Halberstam & Knight, 2016; Barberá et al., 2015; Bakshy et al., 2015; Barberá, 2014; Bakshy et al., 2015).

Despite the increased scholarly attention to the topic, empirical evidence on the effects of Internet and social media use on political polarization remains inconclusive. Empirical attempts to examine causal effects have been limited by identification challenges as they rely on self-reported usage of internet and social media – which typically result in biased outcomes. Self-selection into internet and social media use is one of the reasons for potential endogeneity. It happens because individuals who self-report using internet and social media to inform themselves about politics are potentially different from those who choose not to use it. One may assume that individuals that turn to internet and social media to inform themselves about politics have more extreme positions when compared to individuals that do not. Thus, these individuals may differ from the average individual in their political polarization. These differences may invalidate causal comparisons of outcomes by treatment status, possibly even after adjusting for observed covariates. Establishing causal inference requires finding an exogenous source of variation in the use of information technologies or conducting randomized controlled experiments. Many empirical studies document pure correlations, and are unable to make claims about causality (Boxell et al., 2017; Liang & Nordin, 2013; Boulianne et al., 2020; Lawrence et al., 2010). To our knowledge, the

1 Studies on the tendency of like-minded individuals to associate with one another and expose themselves to information that simply confirms their preexisting opinions date back even further. For a review of the literature on selective exposure to information, see Sears & Freedman (1967).

only academic papers that aim to explore a causal effect of Internet or social media use in political polarization are Lelkes et al. (2017) and Levy (2021).

Moreover, most of the related studies have focused on the US, and there is limited empirical evidence on other regions.[2] There are valuable scientific gains to be made from exploring the phenomena outside the US. For one, most advanced democracies have multiparty systems, in contrast to the two-party political system in the US, which demands exploring the phenomena through novel political polarization measures. At the same time, there is a need to further examine how online media can contribute to polarization in young democracies, where the disruptive role of the internet and social media may arguably be larger than in established Western democracies (Tucker et al., 2018) with relatively stronger institutions.

Brazil, a relatively new democracy with one of the highest party fragmentation of the world (Clark et al., 2006), constitutes an interesting case to study. Brazilian politics is characterized by a recent divide between the left and right ideological spectrum. Ranking on a left-right scale reveals a widening gap between centrist and extreme positions, especially from 2014 to 2018.[3] Adding to this, Brazil is the country with the fifth largest online population in the world, only behind China, India, US and Indonesia (Kemp, 2022). The media and academia have stressed the rising use of the internet and social media for electoral purposes[4] and its association with recent political outcomes in Brazil.[5] However, to the best of our knowledge, the relationship between the use of online media and political polarization is still unexplored from a quantitative perspective.

This paper contributes to the literature by investigating the causal impact of internet and social media use on Brazil's recent political polarization. We employ an instrumental variable (IV) approach that follows past studies using exogenous infrastructure variations to identify the internet's impact on political behavior (Falck et al., 2014; Schaub & Morisi, 2019). We use exogenous variation in the fiber-optic backhaul infrastructure to identify the impact of internet and social media usage on political polarization. Data on fiber optic availability in each one of the 5,570 municipalities - the lowest level of political division in Brazil - is provided by the National Telecommunications Agency. For the IV analysis, we use a dummy variable indicating the availability of fiber optic backhaul for each municipality in 2018.

2 For a detailed review of the literature on the effects of the Internet and social media on political outcomes, see Zhuravskaya et al. (2019) and Tucker et al. (2018).

3 The third section of this paper provides descriptive statistics on the trends of political identification in Brazil, which shows that the percentage of Brazilian voters who declare as extreme left or extreme right has increased since 2002 - and especially from 2014 to 2018.

4 For a discussion of the role of social media in the last Brazilian presidential elections, see Nicolau (2020), chapter 4.

5 For a discussion on the use of social media and political instability, desinformation and polarization in Brazil, see Evangelista & Bruno (2019) and Santos (2019).

Another contribution is the development of a novel way to measure the degree of affective polarization in multiparty systems by exploring a political identity different from parties. Previous work has measured affective polarization mostly along partisan lines. We develop a measure that captures the extent to which citizens feel more positively toward candidates representing their ideological group and negatively toward opposing groups.

The remainder of the paper is organized as follows. Section 2 discusses the theories and reasoning behind the variable developed to measure affective polarization. Section 3 describes the study's data and presents descriptive data on the recent trends in political identification and affective polarization in Brazil. Section 4 describes the empirical estimation strategy and discusses the instrument. Section 5 presents the main results, and section 6 discusses our findings in the context of the literature and concludes.

3.2 MEASURING AFFECTIVE POLARIZATION

Affective polarization refers to a type of political polarization with its roots in Social Identity and Self-Categorization theories. According to such theories (Tajfel et al., 1979; Terry & Hogg, 1996; McGarty et al., 1992; Turner et al., 1987), individuals instinctively form attachments which produce favoritism towards groups they perceive as similar to themselves and antipathy towards groups they do not identify with. The first group is identified as in-group and the second as out-group. Thus, affective polarization measures the extent to which citizens feel more negatively toward other political groups than toward their own by taking the distance between feelings towards in-group and out-group members (Iyengar et al., 2012).

The US is a major case study of affective polarization. In recent years, the rise in partisan animus is a broad consensus among scholars, illustrated by an increasing attachment to co-partisans and animosity towards opposing partisans (Westwood & Peterson, 2019; Iyengar & Westwood, 2015; Iyengar et al., 2019; Lelkes et al., 2017). While social scientists have measured affective polarization mainly along partisan lines, affective polarization is not merely a partisan matter. Social Identity Theories conceive the emergence of in-group favoritism as a result of cognitive and motivational factors related to a broad range of social identities such as real-world social cleavages, ethnic/religious groups, and arbitrary researcher-generated divisions (Mason, 2016; Huddy, 2001; Deaux et al., 1995).

Measuring partisan affect is central in long-established democracies where parties are clearly sorted into salient groups, such as the US. Less attention has been paid, however, to measuring affective polarization along other political identities,[6] which is particularly important in democracies with multi-party systems.

In political systems exhibiting various parties, voters are less likely to rely on party identifications to make political decisions (Lau & Redlawsk, 2001, 2006) - the proliferation of parties makes it hard for the electorate to self-identify with a party or even understand which parties stand for positions that are similar to their own.

Scholars have long conceived partisan loyalties as unlikely to take root in Brazil. Political scientists do not envisage Brazilian individuals developing deep attachments to parties. One of the facts that illustrates this is that the populist president Bolsonaro, elected in 2018, governed for almost 2 years without being affiliated to a political party. One of the explanations relates to the fact that Brazil's party system is a relatively new phenomenon — free and multiparty elections only began after a long military regime in the 1980s (Kinzo, 2005; Samuels & Zucco, 2018; Fiorina, 1981; Huber et al., 2005). Brazil also exhibits one of the highest degrees of partisan fragmentation in the world (Clark et al., 2006). The Brazilian political institutions not only foster party fragmentation but also make it hard for voters to attach to a party ideology or even have a clear understanding of the positions they stand for. For instance, it has an open-list system for legislative elections that promotes intraparty competition, minimizes the importance of party reputation, and strengthens individual candidacy (Samuels & Zucco Jr, 2014).

While on the one hand, the weak mass partisanship in Brazil hinders measuring affective polarization along party lines,[7] on the other hand, the clear recognition of the left-right ideological dimension by voters offers an opportunity to construct a novel measure of affective polarization. We propose a measure that builds on a strand in the political psychology literature that advocates adopting ideological labels as political identities (Malka & Lelkes, 2010; Mason, 2018; Kinder & Kalmoe, 2017).

In this study, affective polarization is represented by a greater attachment to the right-left ideological identities. The measure captures the extent to which voters have a positive sentiment for the candidates representing their ideological group and a negative sentiment against the candidate that represents the opposing group. In the next section, we will specify the construction of the variable.

6 A notable exception is given by Hobolt et al. (2018), who examine social identities formed during Britain's 2016 referendum on European Union membership and show that strong emotional attachments have emerged because of identification with opinion-based groups related to Brexit, demonstrating a strong affective polarization beyond partisan identity.

7 Since 2005, the proportion of voters who identify with any party in Brazil has been hovering around 40% to 45%, which is below the international average of 46% (Huber et al., 2005). Samuels & Zucco Jr (2014) also highlight that the three most salient parties - PSDB, PT, PMDB - "are the only ones to have obtained more than 5% of partisan preferences on average over this period".

3.3 DATA

We use multiple datasets for the analysis. Individual-level data comes from the Brazilian National Election Study (Estudo Eleitoral Brasileiro – ESEB, in Portuguese). ESEB is a nationally representative survey of the voting-age population conducted shortly after the elections by the Center for Public Opinion Studies (Cesop) of the State University of Campinas (Unicamp) and contains numerous demographic variables and political measures. It asks for self-reported votes, political and ideological preferences, internet usage, and social media for political information with a lag of one month from the election. To date, it has been undertaken in five election cycles, since 2002. For the causal analysis, we rely on the fifth wave conducted with home interviews between November 10 and 24 of 2018, which comprises 2506 observations. The data is representative of the five different regions in Brazil and comprise 172 municipalities, including all state capitals. Its margin of error is two percentage points, and the confidence index is 95%.

We also use data on the availability of fiber optic backhaul at the municipality level provided by the National Telecommunications Agency (Agência Nacional de Telecomunicações - Anatel). Anatel annually collects data from the telecommunications providers of broadband internet to map Brazilian telecommunications infrastructure. Since 2016, internet providers inform the National Telecommunications Agency whether each one of the 5,570 municipalities - the lowest level of political division in Brazil - is covered by fiber optic infrastructure. For the IV analysis, we use a dummy variable indicating the availability of fiber optic backhaul for each municipality in 2018. Population density for each one of the municipalities was obtained from the Brazilian Institute of Geography and Statistics (Instituto Brasileiro de Geografia e Estatística - IBGE). Data on fiber optic availability in a municipality provided by the National Telecommunications Agency is combined with individual-level data provided by ESEB at the municipality of residence of the respondents.

3.3.1 *Affective Polarization*

To construct the measure of affective polarization, we rely on data from the ESEB. Individuals were asked to rate feelings towards candidates on an 11-point scale, that ranges from 0 (“strongly dislike”) to 10 (“strongly like”). We take the rate given to the presidential candidate representing the political spectrum with which the individual self-identify (in-group candidate) and subtract the rate given to the opposing candidate (out-group candidate).[8] The greater the result given by this subtraction, the higher the affective polarization.[9]

3.3.2 *National Trends in Affective Polarization*

Table 3.1 uses data of the fives waves of ESEB to show national trends of affective polarization over time. Those with affective polarization scores from 0-3 are defined as neutrals, those with scores from 4-6 as moderates, and those with scores from 7-10 as extremes. The data shows a strong affective polarization in recent years. In 2018, roughly 40% of the individuals declared feeling extreme animosity towards ideological groups other than their own; in previous years, the extreme share was always under 20 per cent of respondents.[10]

Table 3.1: Percentages and changes in affect towards candidates between 2002 and 2018

Year	Neutral	Moderate	Extreme
2002	0.69	0.14	0.17
2006	0.73	0.13	0.14
2010	0.68	0.13	0.18
2014	0.68	0.12	0.2
2018	0.46	0.16	0.38
$\triangle 2002-2018$	-0.23	0.02	0.21
$\triangle 2014-2018$	-0.22	0.04	0.18

3.3.3 *Internet and Social Media Use*

Our measure of the internet and Social Media use is based on the following question of ESEB's fifth wave (2018). Respondents were asked: "Which of the following sources do you use the most to inform yourself about politics?" Possible answers includes TV, personal contacts (family or work), or internet blogs, social media and google search. We coded the variable as a dummy, assigning a value of 1 to those who used internet blogs, social media, or Google search as their primary source of information, and 0 to all other responses.

8 Only two presidential candidates were running for the second round of the 2018 elections in Brazil. We followed Lelkes et al., (2017) and measured affect toward the candidates in the second round of the elections.

9 We followed Lelkes et al., 2017 and rescaled the measure to lie between 0 (out-group candidate rated at 10 and in-group candidate rated at 0) and 10 (in-group candidate rated at 10 and out-group candidate rated at 0). Responses such as "Do not know," "Did not answer" and "Do not know what it means" were coded as missing.

10 In the appendix (figure B.1), we also report changes in average affective polarization between 2002 and 2018.

3.3.4 *Broadband infrastructure for Brazilian municipalities*

The internet provision in a country involves many infrastructure elements: regional backbones, backhauls, and access points. Every country has a regional fiber-optics backbone (in some cases, it can also use satellite or microwaves) that distributes the signal across the territory. Backhauls are intermediate links between the backbones and the peripheral data access points, which connect the final consumers to the rest of the infrastructure. The technology used in this part of the network usually is xDSL or cable technology. Thus, a backhaul is an essential piece of infrastructure that supports high-speed broadband internet services in a municipality.

Previous evidence suggests that the availability of broadband infrastructure is a crucial factor determining the time that individuals spend online. For example, **?** show that having access to broadband internet increases internet usage by over 1300 min per month. Using data from the United States, **?** show that households with high-speed internet access are uniformly more likely to be online.

In this paper, we exploit exogenous variation in the fiber-optic backhauls infrastructure to identify the impact of internet and social media usage on political polarization. Fiber-optic backhauls are normally placed on the basis of pre-existing infrastructures, such as electricity, gas, and oil distribution systems (Knight et al., 2016). Variations of infrastructure are commonly used to construct IVs to measure the internet's effect on various outcomes. For example, the exogenous variation in fiber optic backhauls' availability was recently exploited in a paper identifying the impact of the internet on educational outcomes in Brazil (Henriksen et al., 2022).

The strategy used in this setup is similar to other identification strategies that explore variation in technology dissemination to identify the effects of the internet on voting behavior and political participation. For instance, to analyse the effect of information disseminated over the internet on the voting behavior in Germany, Falck et al.(2014) use regional and technological peculiarities of the preexisting telephony network that hindered the roll-out of fixed-line infrastructure for high-speed Internet. To demonstrate the causal relationship between internet use and voting for populist parties in Italy and Germany, Schaub and Morisi (2019) instrument internet use with broadband coverage at the municipality level. In the US, Lelkes et al. (2017) identify the causal impact of broadband access on polarization by exploiting differences in broadband availability brought about by variation in state right-of-way regulations, which they assume to affect the cost of building internet infrastructure and the price and availability of broadband access.

3.3.5 *Control Variables*

We further used a number of individual socio-demographic control variables, previously used in the literature on the effect of the internet on political polarization and voting preferences (Falck et al., 2014; Schaub & Morisi, 2019; Lelkes et al., 2017). We included respondent's age, level of income, gender, race, religion, occupation sector, residence region (Northeast, South, Southeast, or Midwest) and a dummy variable indicating whether the respondent is unemployed.

3.4 IDENTIFYING THE EFFECT OF INTERNET AND SOCIAL MEDIA USE ON AFFECTIVE POLARIZATION

Identifying the internet and social media effects on polarization is complicated due to endogeneity concerns. A simple regression with the key independent variable being individuals that use the internet or social media would suffer from a potential bias, which would likely arise from a correlation between the observed predictors and the unobserved residual. One of the first reasons for endogeneity is the issue of individual self-selection into internet and social media use. Such endogeneity could arise from the differences in the individuals that self-report using internet or social media to inform themselves about politics from the ones that do not. One may assume that individuals that turn to internet and social media to inform themselves about politics may have more extreme positions when compared to individuals that do not. Thus, these individuals may differ from the average individual in their political polarization. A related concern is of reverse causation. Polarized individuals may be more prone to seek out information about politics online, which would create a causal arrow running from polarization to internet and social media use and would bias estimates. Thus, even if we observe a correlation between political polarization and internet or social media use it is not possible to affirm whether what we observe is a causal relationship.

To address this problem, we follow recent contributions in the field and adopt an IV approach. Variations of infrastructure are commonly used when constructing IVs to measure the internet's effect on voting and political behavior (Falck et al., 2014; Schaub & Morisi, 2019). In this setup, we pursue a similar strategy. We exploit the exogenous variation in the availability of fiber optic backhauls – an infrastructure that allows access to broadband internet – to identify its effect on political polarization.

Suppose the availability of fiber optic backhaul is a valid instrument. In that case, it must be (i) a relevant predictor of the potential endogenous variable and (ii) it should not be a determinant of political polarization. We show below that our instrument meets both conditions. We find that the availability of fiber optic backhaul is significantly associated with internet and social media use. In a probit model that explains the probability of an individual using the internet or social media as the primary source of political information, the coefficient on the fiber optic variable has a p-value of 0.03, indicating that it is significantly different from 0. In a linear model, the F-statistic exceeds the conventional benchmark of 10 set out in Stock et al. (2002) for tests for weak instruments. Both the probit and the linear estimations are reported on table B1 in the appendix.[11]

Furthermore, we show that the exclusion restriction is plausibly met. Although the exclusion restriction requirement cannot be tested itself, we run placebo tests to provide confidence to our instrument. We show that conditional on population density and household income, fiber optics' availability is not correlated with unobserved individual-level characteristics potentially related to political polarization. If the exclusion restriction is met, fiber optic availability should not predict behaviors linked to political or extreme preferences. Results of regressions of variables capturing political preferences on our instrument, fiber optics availability are reported on table B.2 in the appendix. Reassuringly, our instrument is not correlated with political preferences (left or right-wing self-placement), interest in politics, or variables capturing extreme perceptions of corruption scandals, economy, and minorities in Brazil.

3.4.1 *Estimation Strategy*

Given that the dependent variable takes the form of an ordinal variable, an ordered probit or logit model is appropriate in this setup (Long, 1997). As the standard two-stage procedure produces inconsistent estimators for ordered probit models with endogenous variables, we use a maximum likelihood estimation, which, according to Wooldridge (2010), is more efficient than any two-step procedure. The extended ordered probit model accommodates endogenous and instrumental variables while implementing the maximum likelihood estimator (Wooldridge, 2002; Drezner, 1978).

Let y_i be the ordinal variable that measures political polarization. Our extended model consists of equations (3.1) and (3.2):

$$y^* = x\beta + w\beta_c + \epsilon \tag{3.1}$$

$$\begin{aligned} w^* &= z\gamma + r \\ w &= 0 \quad if \quad w^* \leq \zeta \\ w &= 1 \quad if \quad \zeta < w^* \end{aligned} \tag{3.2}$$

11 The F-test is based on Angrist (1991), who state that using linear regression for the first-stage estimates generates consistent second-stages estimates even with a dummy endogenous variable. The same strategy was used in similar studies with binary endogenous variables (Schaub & Morisi, 2019).

where **x** represents the exogenous covariates, and **w**, internet or social media use (the potentially endogenous variable) determined by a binary probit model. **z** contains the instrumental variable that affects **w** and exogenous variables from **x**.

y^* is determined by a threshold model, where α_j, $j = 1, ..., J$ are threshold parameters to be estimated:

$$\begin{aligned} y = 0 \qquad & if \quad y^* \leq \alpha_1 \\ y = 1 \qquad & if \quad \alpha_1 \leq y^* \leq \alpha_2 \\ & \vdots \\ y = J \qquad & if \quad \alpha_J < y^* \end{aligned}$$

Thus the conditional distribution of y depends on **x** and is determined by:

$$\begin{aligned} P(y = 0|\mathbf{x}) &= P(y^* < \alpha_1|\mathbf{x}) = P(\mathbf{x}\beta + \epsilon < \alpha_1|\mathbf{x}) = \Phi(\alpha_1 - \mathbf{x}\beta) && (3.3) \\ P(y = 1|\mathbf{x}) &= P(\alpha_1 < y^* < \alpha_2|\mathbf{x}) = P(\alpha_1 < \mathbf{x}\beta + \epsilon < \alpha_2|\mathbf{x}) \\ &= \Phi(\alpha_2 - \mathbf{x}\beta) - \Phi(\alpha_1 - \mathbf{x}\beta) && (3.4) \\ &\vdots \\ P(y = J|\mathbf{x}) &= 1 - P(y = J - 1|\mathbf{x}) - ... - P(y = 0|\mathbf{x}) \\ &= 1 - P(\alpha_{J-1} < y < \alpha_J|\mathbf{x}) - ... - P(y < \alpha_1|\mathbf{x}) \\ &= 1 - P(\alpha_{J-1} < \mathbf{x}\beta + \epsilon < \alpha_J|\mathbf{x}) - ... - P(\mathbf{x}\beta + \epsilon < \alpha_1|\mathbf{x}) \\ &= 1 - \Phi(\alpha_J - \mathbf{x}\beta) && (3.5) \end{aligned}$$

The argument presented at the beginning of the section suggests that the correlation between ϵ_i and r_i is nonzero. That is, we expect that there are omitted covariates predicting both the probability of being polarised and of using the internet or social media for political purposes. In order to allow for this possibility, we assume that the covariance between ϵ_i and r_i is represented by the matrix Σ.

Considering the category j of the dependent variable y, we have that the upper and lower limits of ϵ given values y and **x** is:

$$y = j \Longrightarrow \alpha_{j-1} < \mathbf{x}\beta + \epsilon < \alpha_j \Longrightarrow \alpha_{j-1} - \mathbf{x}\beta < \epsilon < \alpha_j - \mathbf{x}\beta$$

$$l_{iy} = \alpha_{j-1} - x_i\beta \tag{3.6}$$

$$u_{iy} = \alpha_j - x_i\beta \tag{3.7}$$

l_{iy} e u_{iy} represent the lower and upper limits of ϵ given $y = j$ and $\mathbf{x}$. For the *probit* which describes the dummy variable, the conditional probability is described by:

$$\begin{aligned} P(w=0|\mathbf{z}) &= P(w^* < \zeta|\mathbf{z}) = P(\mathbf{z}\gamma + r < \zeta|\mathbf{z}) = \Phi(\zeta - \mathbf{z}\gamma) && (3.8) \\ P(w=1|\mathbf{z}) &= 1 - P(w^* < \zeta|\mathbf{z}) = 1 - P(\mathbf{z}\gamma + r < \zeta|\mathbf{z}) \\ &= 1 - \Phi(\zeta - \mathbf{z}\gamma) = \Phi(\mathbf{z}\gamma - \zeta) && (3.9) \end{aligned}$$

From (3.8) e (3.9), the lower and upper limits of r given w and $\mathbf{z}$ are:

$$w = 0 \Longrightarrow \mathbf{z}\gamma + r < \zeta \Longrightarrow r < \zeta - \mathbf{z}\gamma$$
$$w = 1 \Longrightarrow \mathbf{z}\gamma + r > \zeta \Longrightarrow r > \zeta - \mathbf{z}\gamma$$
$$l_{iw} = \begin{cases} -\infty & w = 0 \\ \zeta - \mathbf{z}\gamma & w = 1 \end{cases} \tag{3.10}$$
$$u_{iw} = \begin{cases} \zeta - \mathbf{z}\gamma & w = 0 \\ \infty & w = 1 \end{cases} \tag{3.11}$$

Equations (3.6), (3.7), (3.10), and (3.11) together imply that vectors $\mathbf{l}_i$ and $\mathbf{u}_i$ are defined as:

$$\mathbf{l}_i = [l_{iy}, l_{iw}]$$
$$\mathbf{u}_i = [u_{iy}, u_{iw}]$$

Thus, the log likelihood function of the two equation model that takes the covariance matrix Σ into account is:

$$\ln L = \sum_{i=1}^{N} \ln \Phi_2(l_i, u_i, \Sigma) \tag{3.12}$$

where Φ_2 represents the bivariate normal distribution, determined by:

$$\Phi_2(l_i, u_i, \Sigma) = \frac{1}{\sqrt{(2\pi)^2|\Sigma|}} \int_{l_{iy}}^{u_{iy}} \int_{l_{iw}}^{u_{iw}} \exp(-\frac{1}{2}\mathbf{h}^T\Sigma^{-1}\mathbf{h}) \tag{3.13}$$

where $\mathbf{h} = [w \quad y]$. Thus, the conditional probability of $y = j$ is determined by:

$$Pr(y = j|\mathbf{x}, \mathbf{z}, w) = \Phi_2(l_i, u_i, \Sigma) \tag{3.14}$$

Standard errors were clustered by municipalities to allow for spatial correlation in error terms for individuals with residency in the same municipality.[12]

3.5 RESULTS

The estimates of the effects of the internet and social media use in affective polarization are reported in table 3.2. In the first column, we report the standard single ordered probit model estimates, where we find a positive and statistically significant effect (at the 1% level) of the internet and social media. Individuals who choose to use the internet or social media as their primary source of information for political purposes have a higher probability of being polarized than those who choose other sources of information, given their observed characteristics.

Nevertheless, we find that the standard single model and the IV estimations yield different pictures on the effect of the internet and social media in affective polarization. In the extended model, the error terms of the equations that determine an individual's level of affective polarization and the probability of using the internet or social media as a primary source of information for political purposes are positively and significantly correlated. When these error terms are correlated, single-equation models may be biased, as they may attribute part of the impact of unobservable individual characteristics to internet or social media use. In the second column of table 3.2, we report the IV estimations' results using the extended ordered probit model.[13] The correlation of the error terms is reported at the end of table 3.3.

A comparison of the estimates of the coefficients of internet and social media use in the two models suggests that endogeneity is a problem in this setting. If we ignored the correlation, one might conclude that internet and social media use play a positive and significant role in influencing affective polarization. Taking the correlation into account, the internet and social media's significant effect vanishes.

The comparison of the single and the extended equation models also show that the effect of the remaining variables in political polarization is robust. Results are roughly similar in the two models. We find that the estimated coefficients for gender, religion, race, and income variables are significant and do not change signs in both models. The absolute values of most of the coefficients are roughly the same in the two models. They indicate that being a man increases the predicted probability of being polarized. Being catholic compared to other religions decreases the probability of being polarized. And finally that higher levels of income increase the probability of being polarized.

12 We checked the robustness of our results to not clustering. All the coefficients remain with the same value and sign.

13 To estimate the extended model, we employ eoprobit Stata command, which estimates an ordered probit regression model, accommodating endogenous covariates.

We explore the robustness of our findings to changes in the specification of our model. In the third column of table 3.2, we ignore that the dependent variable is ordinal and present the estimation of the two-stage least squares model (2SLS). Such estimation is justified by Angrist (1991).[14] The linear estimates confirm previous findings: although standard regressions show a positive correlation among internet and social media use and political polarization, when the endogeneity is taken into account, the effect disappears. The effect of other variables is also similar for the linear and non-linear IV estimations. None of the covariates change sign, and almost all variables that had a significant effect in the linear model also have a significance in the non-linear model - except for religion and income, which showed significant effects on the non-linear model, but not on the linear one. This comparison confirms that the extended ordered probit model is most appropriate for the dataset.

To test whether there are heterogenous effects across different groups of individuals that use the internet or social media as a main source of information, we generated the IV estimates with interaction effects between our variable of interest and various covariates in the previous model, namely age, level of income, gender, race, religion, and residence region. The estimates suggest that effects are uniform across different socio-demographic groups and regions. Estimated coefficients for the models with the interactions are not presented in this paper for parsimony, but full results are available upon request.

Table 3.2: Affective Polarization

	Standard Model	**IV estimation**	**2SLS**
Internet/Social Media Use	0.317***	-0.469	3.394
	(4.21)	(-1.23)	(0.73)
Age	0.003557	-0.00344	0.0301
	(1.52)	(-0.79)	(0.66)
Male	0.237**	0.225**	0.727**
	(3.24)	(2.88)	(2.63)
Black/Indigenous	-0.135	-0.137	-0.427
	(-1.86)	(-1.90)	(-1.66)
Unemployed	-0.104	-0.179	-0.104
	(-0.94)	(-1.51)	(-0.16)
Protestant	0.0646	0.0153	0.347
	(0.65)	(0.14)	(0.75)
Catholic	-0.230*	-0.296**	-0.694
	(-2.43)	(-3.00)	(-1.18)
Household income			
Near poverty level	0.0958	0.157	0.0137
	(0.97)	(1.50)	(0.03)
Low income	0.299**	0.365**	0.779
	(3.05)	(3.24)	(1.24)
Middle class	0.395**	0.509***	1.146
	(3.03)	(4.04)	(1.16)

14 Using Monte Carlo simulations, he shows that linear IV estimations often have similar results to more sophisticated non-linear models.

Upper middle class	0.515*	0.554*	1.777
	(2.00)	(2.42)	(1.72)
High income	0.477	0.742*	0.772
	(1.06)	(2.22)	(0.34)
Highest tax brackets	-0.431	-0.278	-2.766
	(-0.97)	(-0.83)	(-1.53)
Occupation sector	YES	YES	YES
Region			
Northeast	-0.371**	-0.349*	-1.313**
	(-2.90)	(-2.55)	(-2.74)
Southeast	-0.193	-0.172	-0.660
	(-1.55)	(-1.36)	(-1.43)
South	-0.193	-0.169	-0.758
	(-1.36)	(-1.19)	(-1.42)
Midwest	-0.0983	-0.0765	-0.195
	(-0.54)	(-0.37)	(-0.73)
Constant			2.588
			(0.64)
Observations	1136	1136	1136

Standard errors in parentheses
* $p<0.05$; ** $p<0.01$; *** $p<0.001$

Table 3.3: Internet and Social Media Use

Optical fiber	0.276*
	(2.05)
Age	-0.0306***
	(-8.77)
Male	0.0286
	(0.32)
Black/Indigenous	-0.0102
	(-0.12)
Unemployed	-0.266*
	(-2.01)
Protestant	-0.184
	(-1.57)
Catholic	-0.294**
	(-2.81)
Household income	
Near poverty level	0.316*
	(2.49)
Low income	0.382**
	(3.20)
Middle class	0.650***
	(4.09)
Upper middle class	0.447
	(1.45)
High income	1.228*
	(2.51)
Highest tax brackets	0.576
	(1.25)

Population density	-0.0000157
	(-1.04)
Occuppation sector	YES
Constant	0.717
corr.e Internet/Social Media	0.525*
e.Affective Polarization)	(2.15)
Observations	1136

Standard errors in parentheses
* $p<0.05$; ** $p<0.01$; *** $p<0.001$

3.6 CONCLUSION

This study shows a rising affective polarization in Brazil in recent years using data from the Brazilian National Election Study. Popular accounts point to the internet and social media as a reason for political divide through the expansion of environments that resemble "echo chambers", where citizens are exposed to selective information (Sunstein, 2018; Parisier, 2011). Greater exposure to imbalanced information in these environments would reinforce previous political positions, leading voters to develop more extreme positions or greater animosity towards candidates of the opposing political group (Lelkes et al., 2017; Iyengar et al., 2019).

In contrast to popular assumptions, we find that affective polarization in Brazil cannot be attributed to internet or social media use. Although the estimation of a straight-forward single-equation model shows a statistically significant relationship between online media use and polarization, we find evidence suggesting bias, e.g. it is not possible to disentangle the effect of internet or social media from other unobservable individual factors. Previous levels of political polarization, for example, could be determining the current level of political polarization, but are unobserved in cross-sectional data sets. To overcome the bias, this study exploited exogenous variation in internet and social media usage by using the differences in the layout of pre-existing infrastructure which permit access to broadband internet. When we treat internet and social media use as endogenous variables and estimate a two-equation model using an IV, the effect disappears. Findings are consistent with other empirical studies (Boxell et al., 2017; Liang & Nordin, 2013; Flaxman et al., 2013; Barberá, 2014) suggesting that internet and social media use may not be the cause of rising political polarization within countries. Such findings should attenuate the widespread concern that internet and social media use is a significant motive of political polarization in society nowadays.

One of the possible interpretations for the result is that the so-called "echo chambers" and "filter bubbles" in online media may not be as strong as previously expected. It has now been suggested that exposure to diverse ideological views on online platforms - such as Facebook and Twitter - are more frequent than commonly believed (Bakshy et al., 2015) and that online media users are more likely to be exposed to diverse news than those who use traditional media (Barnidge, 2017; Fletcher and Nielsen, 2018).

Many questions and limitations remain for future research. A crucial issue for a deeper understanding of the mechanisms at stake is the analysis of qualitative aspects of online communication, such as the content users are exposed to. First, it would be interesting to further explore if online media users in Brazil are exposed to diverse and cross-cutting point views rather than selective information in echo chambers, given that we do not find evidence that internet or social media use explains political polarization. Second, it seems important to investigate how people react to diverse and cross-cutting information. On the one hand, it is expected that exposure to cross-cutting information will increase deliberation and understanding among individuals, decreasing polarization (Mutz, 2006). On the other hand, exposure to diverse and cross-cutting news may exacerbate political polarization. Some studies suggest that people who are exposed to information that conflict with their own beliefs are prone to be incivil, which would increase differences between groups and thus polarization (Schumann, 2014; Whitty, 2016; Kim & Kim, 2019). Exploring qualitative data on the web and social media platforms, as suggested by Barberá et al. (2015) can provide an opportunity to understand further the mechanisms behind political polarization and online media.

REFERENCES

Angrist, J. D. (1991), Instrumental variables estimation of average treatment effects in econometrics and epidemiology, Technical report, National Bureau of Economic Research.

Bakshy, E., Messing, S. & Adamic, L. A. (2015), 'Exposure to ideologically diverse news and opinion on facebook', *Science* **348**(6239), 1130–1132.

Barberá, P. (2014), 'How social media reduces mass political polarization. evidence from germany, spain, and the us', *Job Market Paper, New York University* **46**.

Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A. & Bonneau, R. (2015), 'Tweeting from left to right: Is online political communication more than an echo chamber?', *Psychological science* **26**(10), 1531–1542.

Boulianne, S., Koc-Michalska, K. & Bimber, B. (2020), 'Right-wing populism, social media and echo chambers in western democracies', *New Media & Society* **22**(4), 683–699.

Boxell, L., Gentzkow, M. & Shapiro, J. M. (2017), 'Greater internet use is not associated with faster growth in political polarization among us demographic groups', *Proceedings of the National Academy of Sciences* **114**(40), 10612–10617.

Clark, W. R., Gilligan, M. J. & Golder, M. (2006), 'A simple multivariate test for asymmetric hypotheses', *Political Analysis* pp. 311–331.

Deaux, K., Reid, A., Mizrahi, K. & Ethier, K. A. (1995), 'Parameters of social identity.', *Journal of personality and social psychology* **68**(2), 280.

Drezner, Z. (1978), 'Computation of the bivariate normal integral', *Mathematics of Computation* pp. 277–279.

Evangelista, R. & Bruno, F. (2019), 'Whatsapp and political instability in brazil: targeted messages and political radicalisation', *Internet Policy Review* **8**(4), 1–23.

Falck, O., Gold, R. & Heblich, S. (2014), 'E-lections: Voting behavior and the internet', *American Economic Review* **104**(7), 2238–65.

Fiorina, M. P. (1981), *Retrospective voting in American national elections*, Yale University Press.

Flaxman, S., Goel, S. & Rao, J. M. (2013), 'Ideological segregation and the effects of social media on news consumption', *Available at SSRN* **2363701**.

Halberstam, Y. & Knight, B. (2016), 'Homophily, group size, and the diffusion of political information in social networks: Evidence from twitter', *Journal of public economics* **143**, 73–88.

Henriksen, A. L., Zoghbi, A. C., Tannuri-Pianto, M. & Terra, R. (2022), 'Education outcomes of broadband expansion in brazilian municipalities', *Information Economics and Policy* **60**, 100983.

Hobolt, S. B., Leeper, T. & Tilley, J. (2018), 'Divided by the vote: Affective polarization in the wake of brexit', *American Political Science Association, Boston* .

Huber, J. D., Kernell, G. & Leoni, E. L. (2005), 'Institutional context, cognitive resources and party attachments across democracies', *Political Analysis* **13**(4), 365–386.

Huddy, L. (2001), 'From social to political identity: A critical examination of social identity theory', *Political psychology* **22**(1), 127–156.

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N. & Westwood, S. J. (2019), 'The origins and consequences of affective polarization in the united states', *Annual Review of Political Science* **22**, 129–146.

Iyengar, S., Sood, G., Lelkes, Y. et al. (2012), 'Affect, not ideology: a social identity perspective on polarization', *Public Opinion Quarterly* **76**.

Iyengar, S. & Westwood, S. J. (2015), 'Fear and loathing across party lines: New evidence on group polarization', *American Journal of Political Science* **59**(3), 690–707.

Kemp, S. (2022), Digital 2022: Global overview report, Technical report, Datareportal.

Kim, Y. & Kim, Y. (2019), 'Incivility on facebook and political polarization: The mediating role of seeking further comments and negative emotion', *Computers in Human Behavior* **99**, 219–227.

Kinder, D. R. & Kalmoe, N. P. (2017), *Neither liberal nor conservative: Ideological innocence in the American public*, University of Chicago Press.

Kinzo, M. D. (2005), 'Os partidos no eleitorado: percepções públicas e laços partidários no brasil', *Revista Brasileira de Ciências Sociais* **20**(57), 65–81.

Knight, P., Feferman, F. & Foditsch, N. (2016), *Broadband In Brazil: past, present, and future*, Google Access.

Lau, R. R. & Redlawsk, D. P. (2001), 'Advantages and disadvantages of cognitive heuristics in political decision making', *American Journal of Political Science* pp. 951–971.

Lau, R. R. & Redlawsk, D. P. (2006), *How voters decide: Information processing in election campaigns*, Cambridge University Press.

Lawrence, E., Sides, J. & Farrell, H. (2010), 'Self-segregation or deliberation? blog readership, participation, and polarization in american politics', *Perspectives on Politics* **8**(1), 141–157.

Lelkes, Y., Sood, G. & Iyengar, S. (2017), 'The hostile audience: The effect of access to broadband internet on partisan affect', *American Journal of Political Science* **61**(1), 5–20.

Levy, R. (2021), 'Social media, news consumption, and polarization: Evidence from a field experiment', *American economic review* **111**(3), 831–70.

Liang, C.-Y. & Nordin, M. (2013), 'The internet, news consumption, and political attitudes–evidence for sweden', *The BE Journal of Economic Analysis & Policy* **13**(2), 1071–1093.

Malka, A. & Lelkes, Y. (2010), 'More than ideology: Conservative–liberal identity and receptivity to political cues', *Social Justice Research* **23**(2-3), 156–188.

Mason, L. (2016), 'A cross-cutting calm: How social sorting drives affective polarization', *Public Opinion Quarterly* **80**(S1), 351–377.

Mason, L. (2018), *Uncivil agreement: How politics became our identity*, University of Chicago Press.

McGarty, C., Turner, J. C., Hogg, M. A., David, B. & Wetherell, M. S. (1992), 'Group polarization as conformity to the prototypical group member', *British Journal of Social Psychology* **31**(1), 1–19.

Mutz, D. C. (2006), *Hearing the other side: Deliberative versus participatory democracy*, Cambridge University Press.

Nicolau, J. (2020), *O Brasil dobrou à direita: Uma radiografia da eleição de Bolsonaro em 2018*, Zahar.

Pariser, E. (2011), *The Filter Bubble: What The Internet Is Hiding From You*, Penguin Books Limited.

Samuels, D. & Zucco, C. (2018), 'Partisans, anti-partisans, and voter behavior 1', *Routledge Handbook of Brazilian Politics* pp. 269–290.

Samuels, D. & Zucco Jr, C. (2014), 'The power of partisanship in brazil: Evidence from survey experiments', *American Journal of Political Science* **58**(1), 212–225.

Santos, M. (2019), Desarranjo da visibilidade, desordem informacional e polarização no Brazil entre 2013 e 2018, PhD thesis.

Schaub, M. & Morisi, D. (2019), 'Voter mobilization in the echo chamber: Broadband internet and the rise of populism in europe', *European Journal of Political Research* .

Schumann, S. (2014), *How the Internet shapes collective actions*, Springer.

Sears, D. O. & Freedman, J. L. (1967), 'Selective exposure to information: A critical review', *Public Opinion Quarterly* **31**(2), 194–213.

Stock, J. H., Wright, J. H. & Yogo, M. (2002), 'A survey of weak instruments and weak identification in generalized method of moments', *Journal of Business & Economic Statistics* **20**(4), 518–529.

Sunstein, C. R. (2001), *Republic. com*, Princeton University Press.

Sunstein, C. R. (2018), # *Republic: Divided democracy in the age of social media*, Princeton University Press.

Tajfel, H., Turner, J. C., Austin, W. G. & Worchel, S. (1979), 'An integrative theory of intergroup conflict', *Organizational identity: A reader* **56**, 65.

Terry, D. J. & Hogg, M. A. (1996), 'Group norms and the attitude-behavior relationship: A role for group identification', *Personality and social psychology bulletin* **22**(8), 776–

793.

Tucker, J. A., Guess, A., Barberá, P., Vaccari, C., Siegel, A., Sanovich, S., Stukal, D. & Nyhan, B. (2018), 'Social media, political polarization, and political disinformation: A review of the scientific literature', *Political polarization, and political disinformation: a review of the scientific literature (March 19, 2018)* .

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D. & Wetherell, M. S. (1987), *Rediscovering the social group: A self-categorization theory.*, Basil Blackwell.

Van Alstyne, M. & Brynjolfsson, E. (1996), 'Could the internet balkanize science?', *Science* **274**(5292), 1479–1480.

Westwood, S. & Peterson, E. (2019), 'Compound political identity: How partisan and racial identities overlap and reinforce', *Available at SSRN 3417476* .

Whitty, M. T. (2016), *Cyberpsychology: The study of individuals, society and digital technologies*, John Wiley & Sons.

Wooldridge, J. M. (2002), 'Econometric analysis of cross section and panel data mit press', *Cambridge, MA* **108**.

Zhuravskaya, E., Petrova, M. & Enikolopov, R. (2019), 'Political effects of the internet and social media', *Forthcoming, Annual Review of Economics. DOI/10.1146/annurev-economics-081919-050239* .

B

APPENDIX

Table B.1: Instrumental Variable Relevance

	Internet/Social Media use	Internet/Social Media use
	Probit Model	Linear Model
Fiber optic	0.299*.	0.0903*
	(0.032)	(0.019)
Age	-0.0311***	-0.00912***
	(0.000)	(0.000)
Male	0.0430	0.0164
	(0.579)	(0.496)
Black/Indigenous	-0.116	-0.0388
	(0.126)	(0.097)
Unemployed	-0.158	-0.0635
	(0.172)	(0.063)
Protestant	-0.296**	-0.103**
	(0.006)	(0.006)
Catholic	-0.435***	-0.141***
	(0.000)	(0.000)
Household income		
Near poverty level	0.277*	0.0747*
	(0.013)	(0.017)
Low income	0.375***	0.104***
	(0.000)	(0.000)
Middle class	0.684***	0.202***
	(0.000)	(0.000)
Upper middle class	0.673*	0.188
	(0.036)	(0.097)
High income	1.044*	0.343*
	(0.017)	(0.035)
Highest tax brackets	0.465	0.158
	(0.311)	(0.358)
Pop density	-0.0000227	-0.00000701

	(0.164)	(0.162)
Occupation sector	YES	YES
Northest	0.163	0.0495
	(0.339)	(0.335)
Southeast	0.0433	0.0151
	(0.795)	(0.762)
South	-0.0581	-0.0172
	(0.745)	(0.747)
Midwest	-0.0667	-0.0177
	(0.782)	(0.816)
Constant	0.903*	0.775***
Observations	1527	1527
R2		0.156
F		15.98

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B.2: Conditional Independence

	Political Position	Interest in Politics	Perception Corruption
Optical fiber	0.241	-0.0894	-0.0138
	(0.89)	(-1.01)	(-0.40)
Age	0.0162**	0.00103	0.00132*
	(2.71)	(0.61)	(2.16)
Dummy to male	0.156	0.148**	-0.0357
	(0.86)	(2.89)	(-1.95)
Dummy to black			
and indigenous	-0.175	-0.017	0.0115
	(-0.92)	(-0.29)	(0.55)
Dummy to protestant	0.905**	-0.125	-0.0191
	-2.77	(-1.45)	(-0.75)
Dummy to catholic	0.505	-0.180*	-0.0083
	(1.74)	(-2.25)	(-0.31)
Household income			
Near poverty level	-0.363	0.156*	0.0115
	(-1.34)	(2.21)	(0.39)
Low income	-0.508	0.374***	0.0571*
	(-1.86)	(5.62)	(1.98)
Middle class	-0.788*	0.522***	0.00790
	(-2.24)	(5.59)	(0.23)
Upper middle class	0.437	0.939***	0.144
	(0.80)	(4.35)	(3.03)
High income	0.00206	0.453	0.0672

	(0.00)	(1.23)	(0.58)
Highest tax brackets	-0.619	0.202	-0.0448
	(-0.58)	(0.57)	(-0.30)
Population density	-0.0000655*	0.00000199	0.000000265
	(-2.14)	(0.20)	(0.07)
Constant	5.707***	2.182***	0.794***
	(11.46)	(14.39)	(15.07)
Observations	1530	1524	1520
	Perception Government	Perception Minorities	Perception Economy
Optical fiber	-0.0406	0.0632	0.00875
	(-0.71)	(1.39)	(0.20)
Age	-0.00357**	0.00384***	0.00342***
	(-2.62)	(4.80)	(3.57)
Dummy to male	-0.0881*	0.0553*	-0.0845**
	(-2.02)	(2.04)	(-3.31)
Dummy to black and indigenous	0.0378	-0.0168	0.0573*
	(0.93)	(-0.60)	(2.27)
Dummy to protestant	-0.0869	0.0552	-0.000450
	(-1.49)	(1.05)	(-0.01)
Dummy to catholic	-0.0200	0.0382	-0.0140
	(-0.34)	(0.77)	(-0.36)
Household income			
Near poverty level	0.108	-0.0404	-0.0863*
	(1.86)	(-1.04)	(-2.27)
Low income	0.138*	-0.0677	-0.0642
	(2.39)	(-1.73)	(-1.86)
Middle class	0.0741	-0.140**	-0.131*
	(1.07)	(-2.74)	(-2.60)
Upper middle class	0.181	-0.139	-0.152
	(1.19)	(-1.47)	(-1.71)
High income	0.581***	-0.159	-0.915
	(9.86)	(-0.83)	(-0.47)
Highest tax brackets	-0.00820	-0.395**	0.0490
	(-0.04)	(-3.07)	(0.28)
Population density	0.00000454	-0.000000284	0.00000338
	(0.58)	(-0.07)	(0.61)
Constant	0.670***	0.273**	0.351***
	(6.25)	(3.23)	(4.79)
Observations	614	1522	1512

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

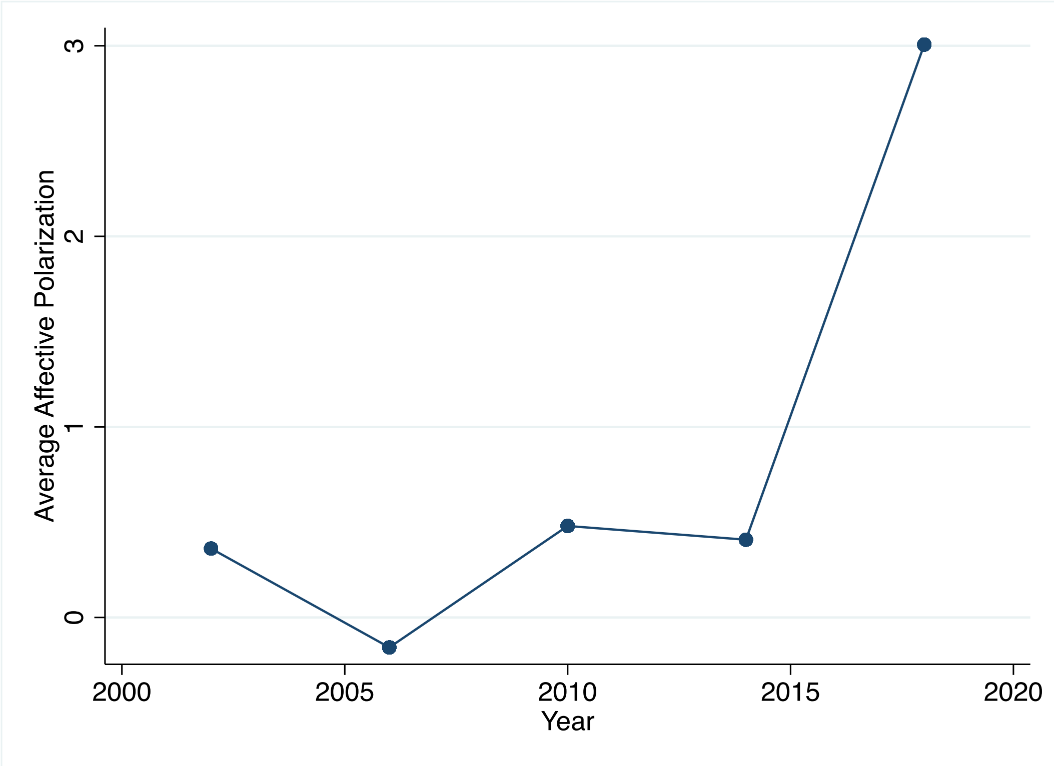

Figure B.1: Changes in average affective polarization between 2002 and 2018

4

DEPOLARIZATION THROUGH SOCIAL MEDIA USE? A DYNAMIC PANEL ANALYSIS FOR THE NETHERLANDS

ABSTRACT

The debate on whether social media use leads to political polarization has gained a prominent position in the academic discussion. Despite the increased attention to the topic, empirical evidence remains inconclusive. We employ a system generalized method of moments (System-GMM) estimator applied to a dynamic panel data model to identify the effects of social media use on political polarization in the Netherlands. The system-GMM allows for the dynamic nature of political polarization and rigorously addresses the likely endogeneity of the relationship between social media and polarization by employing (internal) instrumental variable (IV) techniques. Findings suggest that - contrary to popular assumptions - social media use attenuates rather than drives political polarization, a result that holds for different measures of social media usage. Reading and viewing social media has a significant and negative effect on polarization. Also, more hours spent reading and viewing social media per week, and greater frequency of social media use are associated with lower polarization.

JEL Classification: C23, D72, L82, L86
Keywords: Political polarization, social media, dynamic panel analysis, Netherlands

4.1 INTRODUCTION

The rise of online social media has changed how people are exposed to information and news. Facebook, Instagram, Twitter, YouTube, and others provide users with instant and varied sources of information, which is also often filtered by algorithmic systems. The selective nature of information flows on online social media has attracted scholarly attention.

On the one hand, prominent scholars have argued that the use of online social media use would fuel insularity in political communication. People would end up communicating within "echo chambers" and "filter bubbles", in which individuals are selectively exposed to content that reinforces previously held beliefs (Sunstein, 2018; Pariser, 2011).[1] On the other hand, scholars such as Mutz (2006) and Benkler (2008) have argued that the lower costs of accessing information and increased choice promoted by online social networks would actually lead to greater exposure to diverse ideas.

With the escalation of the ideological divide worldwide, polarization is often attributed to the rise of social media use. In recent decades, political polarization has been at the forefront of popular and scientific debate. While some level of political division is important to incite public debate (Mouffe, 2011), increasing polarization can impede not only compromise in the design and implementation of public policies but also the effective functioning of democracies (Fishkin, 2009; Sunstein, 2001). Thus, there is an increased concern that social media use would exacerbate political polarization.

Evidence, however, is still mixed. While some studies found that social media use can reinforce polarization (Levy, 2021; Quattrociocchi et al., 2016), others have found no effect and even that social media use would actually help to alleviate polarization (Bakshy et al., 2015; Dubois & Blank, 2018; Barberá, 2014; Beam et al., 2018).

As literature reviews on the topic have highlighted (Tucker et al., 2018; Kubin & von Sikorski, 2021), explanations for such disparate evidence are related to the fact that the relationship between social media and political polarization is likely heterogeneous and complex to estimate. Observational studies struggle with the fact that social media use measures are based on self-reports, leaving open the possibility of endogeneity and biased estimates. Endogeneity may arise from unobserved factors that may correlate with political polarization and SM use or reverse causation. One may assume that polarized individuals are more likely to use SM to seek information, which would create a causal arrow running from polarization to SM use and would bias estimates (Tucker et al., 2018). Many have reverted to experiments due to endogeneity issues.

1 This idea is not entirely new to the social media environment - it can be traced back to 1996, when Van Alstyne & Brynjolfsson (1996) expressed concerns about information technologies leading to a "cyber-balkanization".

Adding to the complexity is that polarization is seen as both a state and a dynamic process. Polarization is, as DiMaggio et al. (1996) put it, "both a state and a process". One of the difficulties in accounting for such a dynamic process is the lack of longitudinal data on political attitudes. Although a great deal has been written about the dynamics of political polarization[2] most of the studies exploring the effects of social media use on political polarization are based on static analyses.[3] To our knowledge, the only empirical studies exploring the causal relationship between political polarization and social media over time are Lee et al. (2018); Nordbrandt (2021); Barberá (2014).

The contribution of this paper lies in combining the advantages of a rich panel data set with an advanced econometric identification method to explore the effects of social media use on political polarization in the Netherlands. We employ a system generalized method of moments (System-GMM) estimator applied to a dynamic panel data model. The System-GMM is acknowledged as the most efficient method to estimate dynamic panel models that suffer from endogeneity. The method has been widely applied in areas that typically suffer from endogenous explanatory variables, for example, in economic growth (Bayraktar-Sağlam, 2016), finance (Levine et al., 2000), and education (Castelló-Climent & Mukhopadhyay, 2013). It has also been applied to examine political polarization and its relation to globalization (Fang et al., 2021), political instability (Alt & Lassen, 2006) and energy consumption (Apergis & Pinar, 2021).

The system-GMM allows for the dynamic nature of political polarization and rigorously addresses the likely endogeneity of the relationship between social media and polarization by employing (internal) instrumental variable (IV) techniques. In addition, it controls for unobserved, time-invariant, and individual-specific effects. Thus, the panel data analysis we conduct has numerous benefits over the cross-sectional analyses used on social media and political polarization (Iyengar et al., 2012). Findings suggest that - contrary to popular assumptions - social media use attenuates rather than drives political polarization.

4.2 RELATED LITERATURE

Our work relates to two strands of literature. The first is the empirical analysis of "echo chambers" and "filter bubble" theories. Theories of why SM would fuel polarization are related to the information flow and network configurations shaped by this new environment. The second is the body of empirical literature that allows for the dynamic nature of political polarization.

2 For recent studies on the dynamics of political polarization, see Levin et al. (2021).

3 For observational studies, see Cho et al. (2018); Chang & Park (2021); Johnson et al. (2017); Dubois & Blank (2018). For experimental studies, see Banks et al. (2021); Bail et al. (2018); Heiss et al. (2019); Johnson et al. (2020).

The rise of SM has led to concerns that this information environment populated by personalized recommendation features and algorithmic filtering would reinforce political polarization. Popular scholars have theorized this environment where information would flow in the so-called "echo chambers" and "filter bubbles" would reinforce the preconceived beliefs of users (Pariser, 2011; Sunstein, 2018).

> The new generation of Internet filters looks at the things you seem to like – the actual things you've done, or the things people like you like – and tries to extrapolate. They are prediction engines, constantly creating and refining a theory of who you are and what you'll do and want next. Together, these engines create a unique universe of information for each of us – what I've come to call a filter bubble – which fundamentally alters the way we encounter ideas and information. (Pariser, 2011, p.9)

Thus, according to Pariser, while traditional media allows individuals with random viewpoints, SM's algorithmic filtering limits exposure to differing perspectives.[4] While offering a narrower view of the political debate, "bubble filters tend to dramatically amplify confirmation bias - in a sense, they are designed to do just that" (Pariser, 2011, p. 88). From the author's perspective, an increase in political polarization would be thus a direct consequence.

Such argumentation is in line with Sustein's echo chamber theory. According to Sustein, a central factor behind polarization is the existence of a "limited argument pool" - one that is skewed in a particular direction.

> If your Twitter feed consists of people who think as you do, or if your Facebook friends share your convictions, the argument pool will be sharply limited. Indeed, shifts should occur with individuals not engaged in the discussion but instead consulting only ideas — on radio, television, or the Internet — to which they are predisposed. Such consultations will tend to entrench and reinforce preexisting positions — often resulting in extremism. (Sustein, 2017, p.135)

Considering these arguments, an increasing number of studies have tested empirical support for these theories. To date, evidence using SM data remains mixed. Examinations of selective exposure have shown that Facebook's algorithm may indeed limit exposure to counter-attitudinal news (Levy, 2021). Indeed, Quattrociocchi et al. (2016) shows that the spreading of information on Facebook tends to be confined to like-minded communities. However, a number of studies found that SM users are actually frequently exposed to diverse and cross-cutting viewpoints (Bakshy et al., 2015).

4 This idea is not entirely new and can be traced back to 1996 when Van Alstyne and Brynjolfsson (1996) expressed concerns about information technologies leading to a "cyber-balkanization".

Barberá (2014), in particular, shows that SM users are embedded in diverse ideological networks and that exposure to political diversity reduces political polarization.

Analysis using surveys is also mixed. While some scholars found that SM use indirectly contributed to polarization Lee et al. (2018), more recent evidence did not find any evidence that SM use affects polarization over time (Nordbrandt, 2021). Beam et al. (2018) and Johnson et al. (2017), in particular, identified negative effects of social media use polarization.

A recent review of literature on the topic suggested that one of the explanations for such mixed results is the heterogeneity of the countries, platforms, and political issues analyzed in such studies (Kubin & von Sikorski, 2021).

Our work also relates to a body of research that allows for the dynamic nature of political polarization. Many aspects of political attitudes are theorized as persistent.[5] However, not many empirical studies accounted for the impacts of political polarization over time. Green & Yoon (2002) was the first empirical study to apply a dynamic panel model to analyze political behavior at an individual level. They were interested in exploring the persistence of party identification over time. In other words, they focused on testing if the process of party identification had a memory, i.e., if past levels had reverberating effects on current party identification at an individual level. This was modeled by including lags of the dependent variable (party identification) as a regressor in the model.[6] They found that the coefficient of the lagged dependent variable was not statistically different from zero and thus concluded that party identification was not persistent at the micro-level, in contrast to what was proposed by previous theories.

Green and Yoon (2002) represented a major advance over cross-sectional studies. However, it is now known that the Anderson–Hsiao first-difference estimator used for the dynamic panel model is biased and imprecise. Arellano (1989) shows that estimators that use instruments in levels are preferred to Anderson–Hsiao estimators that use instruments in differences. Thus, the Anderson–Hsiao estimator is inefficient compared with other more recently developed estimators that employ a system of equation framework - namely the System-GMM estimator[7].

In a replication exercise, Wawro (2002) has drawn attention to the costs of using less efficient estimators for a dynamic panel data analysis. He found that Green and Yoon (2002) were possibly wrong in dismissing that party identification had a memory over time. Some of the model's specifications of Green and Yoon (2002) were unreliable after applying the System-GMM estimator with the appropriate specification tests.

5 For a review of recent studies on the dynamics of political polarization, see Levin et al. (2021).

6 Scholars have argued that the inclusion of lags of dependent variable is adequate to account for any dynamics where some variable has an impact that is distributed over time, rather than only immediately Beck & Katz (2011); Wawro (2002).

7 The System-GMM will be further explained in section 4 of this paper (Arellano & Bond, 1991; Arellano & Bover, 1995; Arellano & Honoré, 2001).

It has now been demonstrated that the System-GMM is the most efficient method to estimate dynamic panel data models. The method was applied recently to analyse political polarization and its relation to globalization (Fang et al., 2021), political instability (Elbahnasawy et al., 2016), renewable energy consumption (Apergis & Pinar, 2021) and transparency (Alt & Lassen, 2006).

Our study adds to the literature on the causal effects of SM use on political polarization by applying a dynamic panel model. Most of the studies exploring the effects of SM use in political polarization are based on static analyses[8]. To our knowledge, the only empirical studies exploring the causal relationship between political polarization and SM over time are Lee et al. (2018); Nordbrandt (2021); Barberá (2014).

Our work pursues a complementary aspect to these studies by applying a method that allows for the dynamic nature of political polarization and rigorously addresses the likely endogeneity of the relationship between SM and polarization by employing instrumental variable (IV) techniques and individual-specific effects.

4.3 DATA

Our data comes from the Longitudinal Internet studies for the Social Sciences panel (LISS Panel), a national representative panel study of the Dutch population. It includes about 5000 households and started in 2008, but only in 2013 measures of social media use and frequency were included. It is repeated yearly and covers a wide range of topics. This study comprises the analysis of three modules – Social Integration and Leisure (that includes measures of social media use), Politics and Demographics. The dataset used for our study comprises the years between 2013 and 2020. For a descriptive overview of the survey questions used to construct the variables, see Appendix (C.0.1).

4.3.1 *Political Polarization*

This study follows (Barberá, 2014; Abramowitz & Saunders, 2008) and measures political polarization as the absolute deviation from the center position in the political ideology or opinion space. We use four different variables of political polarization based on questions on political ideology and three policy issues, measured on Likert scales.

Ideological polarization is constructed by respondents' answers to the following question: "In politics, a distinction is often made between the left and the right. Where would you place yourself on the scale below, where 0 means left and 10 means right?".

8 For observational studies, see (Cho et al., 2018; Chang & Park, 2021; Johnson et al., 2017; Dubois & Blank, 2018). For experimental studies, see (Banks et al., 2021; Bail et al., 2018; Heiss et al., 2019; Johnson et al., 2020)

To measure opinion polarization about specific policy issues, we use the three following questions about immigration, income inequality, and European Union integration: i) "In the Netherlands, some people believe that immigrants are entitled to live here while retaining their own culture. Others feel that they should adapt entirely to Dutch culture. Where would you place yourself on a scale of 1 to 5, where 1 means that immigrants can retain their own culture and 5 means that they should adapt entirely?"; ii) "Some people believe that differences in income should increase in our country. Others feel that they should decrease. Still others hold an opinion that lies somewhere in between. Where would you place yourself on a scale from 1 to 5, where 1 means that differences in income should increase and 5 means that these should decrease?"; iii) "Some people and political parties feel that European unification should go a step further. Others think that European unification has already gone too far. Where would you place yourself on a scale from 1 to 5, where 1 means that European unification should go further and 0 means that it has already gone too far?".

Measures are standardized to ensure valid distance comparisons. Thus, our measure of polarization is the absolute value of the standardized measure of the positions on political ideology and opinion questions.

4.3.2 *Social Media Use*

Our measures of Social Media (SM) use are based on three different questions. First, our dummy measure, which we call SM use is based on the following question: "Do you spend time on the following on-line activities? reading and viewing social media (e.g. Facebook, Instagram, Twitter, YouTube, LinkedIn, Google+, Pinterest, Flickr, or similar services)". Answer categories were 1= yes, 2 = no.

Secondly, our measure on the intensity of SM use is based on the following question: "Can you indicate the mean number of hours per week you spend time on these online activities?". Answers were given in integer. Finally, our measure on the frequency of SM use is based on the following question: "How often did you make use of social media in the past 2 months?" Possible answers included 1 = never; 2 = less than once a month; 3 = 1-3 times per month; 4 = once a week; 5 = several times per week; 6 = every day; 7 = several times per day.

As the distribution of the values of intensity and frequency of use was severely skewed to the right, we log-transform the variables to correct the skewness. Thus, overall, our analyses include three measures of SM use: dummy SM use (yes vs. no) and the log-transformed intensity and frequency of SM use. We disregard missing values on each one of the variables.

4.3.3 *Control Variables*

Following the literature on the effect of SM use on political polarization and ideological preferences, we further use a number of individual socio-demographic control variables (Lee et al., 2018; Barberá, 2014; Boulianne et al., 2020). We include respondent's age, gender, dummies on the highest level of education, civil status, urban character of place of residence and dummies on ethnicity.

4.4 IDENTIFICATION STRATEGY

We estimate the following dynamic panel data models, with a panel dataset that covers the period of 2013 to 2020:

$$\begin{aligned} PolPolarization_{i,t} &= \beta_1 \cdot SM_{use_{i,t}} + \beta_2 \cdot PolPolarization_{i,t-1} \\ &\quad + \delta' \cdot \mathbf{x_{i,t}} + u_i + \mathbf{v_{i,t}} \end{aligned} \tag{4.1}$$

$$\begin{aligned} PolPolarization_{i,t} &= \beta_1 \cdot lnSM_{hours_{i,t}} + \beta_2 \cdot PolPolarization_{i,t-1} \\ &\quad + \delta' \cdot \mathbf{x_{i,t}} + u_i + \mathbf{v_{i,t}} \end{aligned} \tag{4.2}$$

$$\begin{aligned} PolPolarization_{i,t} &= \beta_1 \cdot lnSM_{frequency_{i,t}} + \beta_2 \cdot PolPolarization_{i,t-1} \\ &\quad + \delta' \cdot \mathbf{x_{i,t}} + u_i + \mathbf{v_{i,t}} \end{aligned} \tag{4.3}$$

where $PolPolarization_{i,t}$ denotes the political polarization of individual i in year t, $SMuse_{i,t}$ is a dummy variable indicating if the individual i has reported using SM in year t, in the first equation, $lnSMhours_{i,t}$ is the natural logarithm of the number of hours individual i has reported using SM per week in year t, in the second equation, and $lnSMfrequency_{i,t}$ is the natural logarithm of the frequency individual i has reported using SM in year t. $PolPolarization_{i,t-1}1$ is the lagged political polarization for individual i. $\delta' \cdot x_{i,t}$ is the vector of other explanatory variables for individual i in year t, u_i is an unobservable time-invariant individual effect, and $v_{i,t}$ denotes the error term.

As highlighted by a literature review on the topic (Tucker et al., 2018) identifying SM effects on political polarization is complicated due to endogeneity concerns. Endogeneity may arise from unobserved time-invariant individual's characteristics (fixed-effects) that may be correlated with both political polarization and SM use. That is, one may assume that polarized individuals are more likely to use SM to seek out for information, which would create a causal arrow running from polarization to SM use and would bias estimates.

In the literature, one prominent way to address these problems has been through generalized method of moments (GMM) estimators applied to dynamic panel data models (Hill et al., 2021). Dynamic panel models allow for the persistence of the political polarization over time by including in their specification its lagged values as a regressor and unobserved individual-specific effects. These models have important advantages. First, they address potential bias through any omitted variables that are constant over time (unobserved individual-specific or 'fixed' effects). Secondly, as we discuss below, they use internal instrumental variables that allow parameters to be estimated consistently in models with endogenous variables. Finally, by including lagged variables that account for persistent effects, they allow researchers to explore if an individual's behavior in the past might have an impact on her behavior in current periods.

We employ the System-GMM estimator, which produces efficiency gains over the first-difference GMM estimator. The first-difference GMM (Diff-GMM) deals with the endogeneity problem by first differencing the data and using lagged values of the dependent variables as instruments (Arellano and Bond, 1991). However, it is now well known that sample biases can occur when instrumental variables are weak. Therefore, Blundell and Bond (1998) suggest adding another set of orthogonality conditions between the levels of the error term and the first differences in the exogenous variables. The System-GMM simultaneously estimates the equations in differences and in levels, using distinct instruments for each equation. The lagged variables in levels instrument the first-difference equation, and the lagged differences instrument the levels equation. This technique adequately deals with suspected endogeneity and fixed-effects (Roodman, 2009*b*,*a*).

We consider SM use (dummy variable) and the intensity and frequency of SM use as endogenous variables and the remaining covariates as exogenous regressors. This means that they are treated as instrumental variables (IVs) in the GMM estimator.

Based on the second order autocorrelation test and the Hansen J statistics on overidentifying restrictions, we find that adjusting the proliferation of instruments used was needed for the model to perform adequately. To avoid overfitting we use a Principal Component extraction condition from the instrument matrix.[9] We used the two-step estimator of the System-GMM, which allows for heteroskedastic. For the one-step estimator, it is assumed that the error terms are independent and homoscedastic. For the two-step estimator, the residuals obtained in the first step are used to construct a consistent estimate of the variance-covariance matrix, relaxing the hypotheses of independence and homoscedasticity. This dynamic system-GMM strategy thoroughly

9 For more detailed information in the method, see Mehrhoff (2009); Kapetanios & Marcellino (2010).

addresses the endogeneity concerns noted earlier and gives the estimates for this paper.

4.5 RESULTS

4.5.1 *Effects of Social Media Use on Polarization*

The System-GMM estimates of the effects of SM use on our main measure of polarization - namely the ideological polarization - are reported in Table 4.1. We run separate models for each measure of SM use. In the first column, we report estimates of the model measuring the effect of using SM, in the second column, the model measuring the effect of the intensity of SM use, and in the third column the model measuring the effect of the frequency of SM use.[10] In every regression, we use the two-step system-GMM method with Windmeijer (2005) robust standard error, and Principal Component Analysis to control for the proliferation of the instruments (Mehrhoff, 2009; Kapetanios & Marcellino, 2010).

To check the validity of the instruments and the model's fit we analyze Hansen's J-statistic specification tests, Arellano-Bond test for first-order and second-order autocorrelation, and the Kaiser-Lawyer-Olkin measuring of sample adequacy (KMO). The test of overidentification is based on the Hansen J statistic. The Hansen test results do not reject the null hypothesis that the instruments are valid for all the specifications used (Hansenp $p > 0.05$ for all estimates), indicating that instruments are valid for all regressions. As for the Arellano-Bond test for autocorrelation, the results reject the null hypothesis of the absence of first order autocorrelation (AR(2) $p < 0.00$ for all estimates) and do not reject the null hypothesis of the absence of second order autocorrelation ($p > 0.05$ for all estimates). The insignificance of all the second order autocorrelation test results implies that there was no second-order serial correlation of the error term for all the regression models. Finally, the Kaiser-Meyer-Olkin (KMO) measure of sample adequacy for PCA shows values higher than 0.5. That is, we have confidence that the factor analysis used is adequately adjusted to the data. In sum, the tests indicate good specification quality.

Findings lend support to the argument that using SM attenuates ideological polarization. In Table 4.1, reading and viewing SM has a significant and negative effect on polarization, controlling for demographic, individual characteristics and prior polarization (dummy variable β = - 0.06, $p < 0.001$). Furthermore, spending more hours per week reading and viewing SM and using SM with a greater frequency is also associated with lower ideological polarization (intensity variable β = - 0.04, $p < 0.001$ and frequency variable β = - 0.05, $p < 0.001$).

10 In order to detect a potential correlation between predictors in the models we run the Variance Inflation Factor, which indicates no concerns of collinearity.

Table 4.1: System-GMM - Ideological Polarization

	Pol_Polarization	Pol_Polarization	Pol_Polarization
Use_SM	-0.061***		
	(0.02)		
ln_Hours_Per_Week_SM		-0.043*	
		(0.02)	
ln_Frequency_Use_SM			-0.047*
			(0.03)
L.Pol_Polarization	0.109***	0.075**	0.111***
	(0.03)	(0.03)	(0.03)
ln_Age	-0.001	0.021	0.012
	(0.03)	(0.03)	(0.03)
D_No_Schooling	1.220	20.611	13.443
	(7.24)	(18.91)	(17.58)
D_Primary_School	-0.031	-0.080	-0.055
	(0.08)	(0.09)	(0.09)
D_High_School	-0.036*	-0.023	-0.027
	(0.02)	(0.03)	(0.02)
D_College	0.013	0.021	0.016
	(0.02)	(0.03)	(0.02)
D_Male	0.056***	0.065***	0.055***
	(0.02)	(0.02)	(0.02)
D_Married	-0.059***	-0.068***	-0.066***
	(0.02)	(0.02)	(0.02)
Urban	0.017***	0.026***	0.020**
	(0.01)	(0.01)	(0.01)
D_Dutch	-0.023	-0.025	-0.021
	(0.03)	(0.03)	(0.03)
D_Foreign_Non_Western	-0.022	-0.040	-0.021
	(0.05)	(0.06)	(0.05)
Year	YES	YES	YES
_cons	0.795***	0.672***	0.758***
	(0.12)	(0.12)	(0.14)
Hansen	0.314	0.649	0.454
AR(1)	0.000	0.000	0.000
AR(2)	0.409	0.665	0.621
KMO	0.869	0.874	0.877
N instruments	29	29	29
N	10744	7733	9144

Standard errors in parentheses
* $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

The comparison of the three specifications also suggests that the model is well-specified. The covariates are overall consistent in the three models. We find that the estimated coefficients for gender, civil status, having completed high school and urban character of the place of residence indicate the expected signs and are mostly significant in the three models. And the absolute values of most of the coefficients are roughly similar. Also, the estimation of the lagged level of political polarization complements the correct specification of the dynamic panel data models. The estimated lagged coefficients of political polarization are positive and significant at the 1 % level of confidence, suggesting that political polarization at the individual level is persistent: past levels of political polarization have reverberating effects on current ones.

4.5.2 *Alternative Polarization Measures*

We also investigated the effect of SM use in alternative polarization measures, namely opinion polarization. Tables 4.2, 4.3 and 4.4 report the System-GMM estimates of the effects of SM use in opinion polarization on immigration, EU integration and income inequality, with the same specifications of the models as in the previous subsection.

Taken together, findings also lend support to the argument that using SM attenuates polarization. Reading and viewing SM was found to have a significant and negative effect on opinion polarization on immigration and EU integration, controlling for demographic, individual characteristics and prior political polarization. Furthermore, more hours spent reading and viewing SM use per week was also found to have a negative effect on polarization on immigration opinions. Using SM had no effect on polarization on income inequality opinions.

4.5.3 *Robustness Checks*

In this section we verify if the results previously found are robust. First, we test the results by considering placebo tests and secondly, we make changes in the specifications of the final model. To perform the placebo tests, we perform an additional System-GMM estimation using "fake" SM use responses. To create the "fake" SM use responses, we assign the responses corresponding to SM use (dummy, intensity and frequency variables) to random individuals in each one of the survey waves. The results reported in the Appendix corroborate our previous findings. Reassuringly, we find that polarization is not affected by SM use when we run the model with the same specifications for the placebo responses of SM use (tables C.3, C.4, C.5 and C.6). We also replicate the final model including only the age and gender covariates (tables C.7, C.8, C.9 and C.10). As expected, results are robust to the exclusion of controls.

Table 4.2: System-GMM - Opinion Polarization (Immigration)

	pol_immigration	pol_immigration	pol_immigration
Use_SM	-0.035*		
	(0.02)		
ln_Hours_Per_Week_SM		-0.028**	
		(0.01)	
ln_Frequency_Use_SM			-0.037
			(0.02)
L.pol_immigration	0.045**	0.051**	0.053**
	(0.02)	(0.02)	(0.02)
ln_Age	-0.057***	-0.057***	-0.051**
	(0.02)	(0.02)	(0.02)
D_No_Schooling	0.361***	0.306**	0.340***
	(0.12)	(0.12)	(0.12)
D_Primary_School	0.158***	0.190***	0.167***
	(0.04)	(0.05)	(0.05)
D_High_School	0.051***	0.055***	0.050***
	(0.01)	(0.02)	(0.01)
D_College	0.036**	0.023	0.032*
	(0.02)	(0.02)	(0.02)
D_Male	0.035***	0.035***	0.035***
	(0.01)	(0.01)	(0.01)
D_Married	0.004	-0.003	-0.005
	(0.01)	(0.01)	(0.01)
Urban	0.008**	0.009*	0.007
	(0.00)	(0.00)	(0.00)
D_Dutch	-0.010	-0.017	-0.019
	(0.02)	(0.02)	(0.02)
D_Foreign_Non_Western	0.283***	0.265***	0.260***
	(0.04)	(0.04)	(0.04)
Year	YES	YES	YES
_cons	0.940***	0.945***	0.961***
	(0.08)	(0.09)	(0.11)
Hansen	0.256	0.129	0.123
AR(1)	0.000	0.000	0.000
AR(2)	0.522	0.758	0.611
KMO	0.904	0.907	0.908
N Instruments	30	30	30
N	17806	12581	14325

Standard errors in parentheses
* $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table 4.3: System-GMM - Opinion Polarization (EU Integration)

	pol_Opinion_UE	pol_Opinion_UE	pol_Opinion_UE
Use_SM	-0.129***		
	(0.05)		
ln_Hours_Per_Week_SM		0.009	
		(0.02)	
ln_Frequency_Use_SM			0.021
			(0.05)
L.pol_Opinion_UE	0.141***	0.105**	0.131***
	(0.03)	(0.04)	(0.03)
ln_Age	0.023	0.084***	0.086***
	(0.03)	(0.02)	(0.03)
D_No_Schooling	-0.021	0.016	0.042
	(0.08)	(0.08)	(0.07)
D_Primary_School	0.036	0.022	0.027
	(0.03)	(0.04)	(0.04)
D_High_School	0.007	0.023*	0.016
	(0.01)	(0.01)	(0.01)
D_College	0.111***	0.115***	0.110***
	(0.02)	(0.02)	(0.02)
D_Male	0.115***	0.138***	0.134***
	(0.01)	(0.01)	(0.01)
D_Married	0.021*	0.036***	0.029**
	(0.01)	(0.01)	(0.01)
Urban	0.010**	0.014***	0.010**
	(0.00)	(0.00)	(0.00)
D_Dutch	-0.010	-0.008	-0.011
	(0.02)	(0.02)	(0.02)
D_Foreign_Non_Western	0.031	0.016	0.024
	(0.03)	(0.03)	(0.03)
Year	YES	YES	YES
_cons	0.652***	0.296***	0.267
	(0.13)	(0.09)	(0.17)
Hansen	0.599	0.179	0.060
AR(1)	0.000	0.000	0.000
AR(2)	0.986	0.714	0.796
KMO	0.901	0.907	0.910
N instruments	23.000	23.000	22.000
N	16885	11836	13516

Standard errors in parentheses
* $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table 4.4: System-GMM - Opinion Polarization (Income Inequality)

	pol_inequality	pol_inequality	pol_inequality
Use_SM	-0.034		
	(0.03)		
ln_Hours_Per_Week_SM		0.011	
		(0.01)	
ln_Frequency_Use_SM			-0.007
			(0.03)
L.pol_inequality	0.025	-0.004	-0.002
	(0.02)	(0.03)	(0.02)
ln_Age	-0.012	0.022	0.005
	(0.02)	(0.02)	(0.02)
D_No_Schooling	0.380**	0.396**	0.420***
	(0.18)	(0.16)	(0.16)
D_Primary_School	0.124***	0.111***	0.120***
	(0.04)	(0.04)	(0.04)
D_High_School	0.026**	0.031**	0.035**
	(0.01)	(0.01)	(0.01)
D_College	0.007	0.012	0.006
	(0.02)	(0.02)	(0.02)
D_Male	0.058***	0.072***	0.070***
	(0.01)	(0.01)	(0.01)
D_Married	0.006	0.004	-0.002
	(0.01)	(0.01)	(0.01)
Urban	0.004	0.007	0.005
	(0.00)	(0.00)	(0.00)
D_Dutch	-0.001	0.015	0.006
	(0.02)	(0.02)	(0.02)
D_Foreign_Non_Western	0.011	0.008	0.012
	(0.03)	(0.03)	(0.03)
Year	YES	YES	YES
_cons	0.832***	0.669***	0.778***
	(0.08)	(0.09)	(0.11)
Hansen	0.059	0.175	0.211
AR(1)	0.000	0.000	0.000
AR(2)	0.243	0.418	0.701
KMO	0.904	0.907	0.909
N instruments	25	25	25
N	17806	12581	14325

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.6 HETEROGENEOUS EFFECTS

To test whether SM use effects varies across different group of individuals, we estimate models where we include the interactions between SM use and various control variables, namely level of education, age, urban character of the place of residence and gender. We did not find evidence of heterogeneous effects among different group of individuals.[11] Estimated coefficients for the models with the interactions are not presented in this paper for parsimony, but full results are available upon request.

4.7 CONCLUSION

Returning to our research question — the effect of social media use on political polarization - there are two competing theories. On one hand, various authors have argued that social media usage would reinforce polarization, because they would facilitate communication between like-minded individuals, the strengthening of communities with same ideologies and the development of imbalanced flows of information. On the other, some voices have argued that social networks could have the effect of exposing individuals to opposing views and to a more pluralistic set of information, what could lead to the adoption of more centrist positions and the alleviation of political polarization.

This study lends support to the second strand of research. Our findings suggest that social media use attenuates rather than drives polarization, a result that holds for different measures of social media usage - dummy (yes vs. no), intensity (time spent), and frequency. Reading and viewing social media has a significant and negative effect on polarization. More hours spent reading and viewing social media per week, and greater frequency of social media use are also associated with lower polarization.

Our findings seem to support Mutz's argument of cross-cutting social networks (Mutz, 2006, 2002) according to which social media is associated with weaker opinions and identities because it would promote a greater understanding of other's people perspectives.

The research developed in this paper has important limitations and several questions remain open for future research. First, the research focused on the quantitative aspects of online communication and skipped any discussion regarding the qualitative characteristics of messages disseminated through social media. Careful qualitative analysis is important to illuminate the types of content that circulate in social networks and assess the possibilities these technologies open to the dissemination of extreme messages, hate speech and misinformation. Second, the study did not

11 It is worth mentioning that the number of observations of individuals with no education or only primary education is very low in the sample, which might have restrained us from finding a different effect for these different set of groups.

take into account the heterogeneity across different social media platforms. Because each social network has a specific method of disseminating information and reaches a particular audience, it is important to investigate if some platforms might promote different effects on political polarization and, if so, what are the reasons that cause these effects.

REFERENCES

Abramowitz, A. I. & Saunders, K. L. (2008), 'Is polarization a myth?', *The Journal of Politics* **70**(2), 542–555.

Alt, J. E. & Lassen, D. D. (2006), 'Transparency, political polarization, and political budget cycles in oecd countries', *American Journal of Political Science* **50**(3), 530–550.

Apergis, N. & Pinar, M. (2021), 'The role of party polarization in renewable energy consumption: Fresh evidence across the eu countries', *Energy Policy* **157**, 112518.

Arellano, M. (1989), 'A note on the anderson-hsiao estimator for panel data', *Economics letters* **31**(4), 337–341.

Arellano, M. & Bond, S. (1991), 'Some tests of specification for panel data: Monte carlo evidence and an application to employment equations', *The review of economic studies* **58**(2), 277–297.

Arellano, M. & Bover, O. (1995), 'Another look at the instrumental variable estimation of error-components models', *Journal of econometrics* **68**(1), 29–51.

Arellano, M. & Honoré, B. (2001), Panel data models: some recent developments, *in* 'Handbook of econometrics', Vol. 5, Elsevier, pp. 3229–3296.

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. F., Lee, J., Mann, M., Merhout, F. & Volfovsky, A. (2018), 'Exposure to opposing views on social media can increase political polarization', *Proceedings of the National Academy of Sciences* **115**(37), 9216–9221.

Bakshy, E., Messing, S. & Adamic, L. A. (2015), 'Exposure to ideologically diverse news and opinion on facebook', *Science* **348**(6239), 1130–1132.

Banks, A., Calvo, E., Karol, D. & Telhami, S. (2021), '# polarizedfeeds: Three experiments on polarization, framing, and social media', *The International Journal of Press/Politics* **26**(3), 609–634.

Barberá, P. (2014), 'How social media reduces mass political polarization. evidence from germany, spain, and the us', *Job Market Paper, New York University* **46**.

Bayraktar-Sağlam, B. (2016), 'The stages of human capital and economic growth: Does the direction of causality matter for the rich and the poor?', *Social Indicators Research* **127**(1), 243–302.

Beam, M. A., Hutchens, M. J. & Hmielowski, J. D. (2018), 'Facebook news and (de) polarization: reinforcing spirals in the 2016 us election', *Information, Communication & Society* **21**(7), 940–958.

Beck, N. & Katz, J. N. (2011), 'Modeling dynamics in time-series–cross-section political economy data', *Annual Review of Political Science* **14**, 331–352.

Benkler, Y. (2008), *The wealth of networks*, Yale University Press.

Boulianne, S., Koc-Michalska, K. & Bimber, B. (2020), 'Right-wing populism, social media and echo chambers in western democracies', *New Media & Society* **22**(4), 683–699.

Castelló-Climent, A. & Mukhopadhyay, A. (2013), 'Mass education or a minority well educated elite in the process of growth: The case of india', *Journal of Development Economics* **105**, 303–320.

Chang, K. & Park, J. (2021), 'Social media use and participation in dueling protests: The case of the 2016–2017 presidential corruption scandal in south korea', *The International Journal of Press/Politics* **26**(3), 547–567.

Cho, J., Ahmed, S., Keum, H., Choi, Y. J. & Lee, J. H. (2018), 'Influencing myself: Self-reinforcement through online political expression', *Communication Research* **45**(1), 83–111.

DiMaggio, P., Evans, J. & Bryson, B. (1996), 'Have american's social attitudes become more polarized?', *American journal of Sociology* **102**(3), 690–755.

Dubois, E. & Blank, G. (2018), 'The echo chamber is overstated: the moderating effect of political interest and diverse media', *Information, communication & society* **21**(5), 729–745.

Elbahnasawy, N. G., Ellis, M. A. & Adom, A. D. (2016), 'Political instability and the informal economy', *World Development* **85**, 31–42.

Fang, J., Gozgor, G. & Yan, C. (2021), 'Does globalisation alleviate polarisation?', *The World Economy* **44**(4), 1031–1052.

Fishkin, J. (2009), *When the people speak: Deliberative democracy and public consultation*, Oup Oxford.

Green, D. P. & Yoon, D. H. (2002), 'Reconciling individual and aggregate evidence concerning partisan stability: Applying time-series models to panel survey data', *Political Analysis* **10**(1), 1–24.

Heiss, R., von Sikorski, C. & Matthes, J. (2019), 'Populist twitter posts in news stories: Statement recognition and the polarizing effects on candidate evaluation and anti-immigrant attitudes', *Journalism Practice* **13**(6), 742–758.

Hill, A. D., Johnson, S. G., Greco, L. M., O'Boyle, E. H. & Walter, S. L. (2021), 'Endogeneity: A review and agenda for the methodology-practice divide affecting micro and macro research', *Journal of Management* **47**(1), 105–143.

Iyengar, S., Sood, G., Lelkes, Y. et al. (2012), 'Affect, not ideology: a social identity perspective on polarization', *Public Opinion Quarterly* **76**.

Johnson, B. K., Neo, R. L., Heijnen, M. E., Smits, L. & van Veen, C. (2020), 'Issues, involvement, and influence: Effects of selective exposure and sharing on polarization and participation', *Computers in Human Behavior* **104**, 106155.

Johnson, T. J., Kaye, B. K. & Lee, A. M. (2017), 'Blinded by the spite? path model of political attitudes, selectivity, and social media', *Atlantic Journal of Communication* **25**(3), 181–196.

Kapetanios, G. & Marcellino, M. (2010), 'Factor-gmm estimation with large sets of possibly weak instruments', *Computational Statistics & Data Analysis* **54**(11), 2655–2675.

Kubin, E. & von Sikorski, C. (2021), 'The role of (social) media in political polarization: a systematic review', *Annals of the International Communication Association* **45**(3), 188–206.

Lee, C., Shin, J. & Hong, A. (2018), 'Does social media use really make people politically polarized? direct and indirect effects of social media use on political polarization in south korea', *Telematics and Informatics* **35**(1), 245–254.

Levin, S. A., Milner, H. V. & Perrings, C. (2021), 'The dynamics of political polarization', *Proceedings of the National Academy of Sciences of the United States of America* **118**(50), e2116950118.

Levine, R., Loayza, N. & Beck, T. (2000), 'Financial intermediation and growth: Causality and causes', *Journal of monetary Economics* **46**(1), 31–77.

Levy, R. (2021), 'Social media, news consumption, and polarization: Evidence from a field experiment', *American economic review* **111**(3), 831–70.

Mehrhoff, J. (2009), 'A solution to the problem of too many instruments in dynamic panel data gmm', *Available at SSRN 2785360* .

Mouffe, C. (2011), *On the political*, Routledge.

Mutz, D. C. (2002), 'Cross-cutting social networks: Testing democratic theory in practice', *American Political Science Review* **96**(1), 111–126.

Mutz, D. C. (2006), *Hearing the other side: Deliberative versus participatory democracy*, Cambridge University Press.

Nordbrandt, M. (2021), 'Affective polarization in the digital age: Testing the direction of the relationship between social media and users' feelings for out-group parties', *new media & society* p. 14614448211044393.

Pariser, E. (2011), *The Filter Bubble: What The Internet Is Hiding From You*, Penguin Books Limited.

Quattrociocchi, W., Scala, A. & Sunstein, C. R. (2016), 'Echo chambers on facebook', *Available at SSRN 2795110* .

Roodman, D. (2009*a*), 'How to do xtabond2: An introduction to difference and system gmm in stata', *The stata journal* **9**(1), 86–136.

Roodman, D. (2009*b*), 'A note on the theme of too many instruments', *Oxford Bulletin of Economics and statistics* **71**(1), 135–158.

Sunstein, C. R. (2001), *Republic. com*, Princeton University Press.

Sunstein, C. R. (2018), # *Republic: Divided democracy in the age of social media*, Princeton University Press.

Tucker, J. A., Guess, A., Barberá, P., Vaccari, C., Siegel, A., Sanovich, S., Stukal, D. & Nyhan, B. (2018), 'Social media, political polarization, and political disinformation: A review of the scientific literature', *Political polarization, and political disinformation: a review of the scientific literature (March 19, 2018)* .

Van Alstyne, M. & Brynjolfsson, E. (1996), 'Could the internet balkanize science?', *Science* **274**(5292), 1479–1480.

Wawro, G. (2002), 'Estimating dynamic panel data models in political science', *Political Analysis* **10**(1), 25–48.

Windmeijer, F. (2005), 'A finite sample correction for the variance of linear efficient two-step gmm estimators', *Journal of econometrics* **126**(1), 25–51.

C

APPENDIX

C.0.1 *Variables*

The list below gives details on questions used to construct some of the variables used.

VARIABLES	**QUESTION**
Social Media Use (Dummy)	Do you spend time on the following online activities? reading and viewing social media (e.g. Facebook, Instagram, Twitter, YouTube, LinkedIn, Google+, Pinterest, Flickr, or similar services). 1 = Yes/ 0 = No.
Social Media Use (Intensity)	Can you indicate the mean number of hours per week you spend time on these online activities? – Integer.
Social Media Use (Frequency)	How often did you make use of social media in the past 2 months? – 1 = never; 2 = less than once a month; 3 = 1-3 times per month; 4 = once a week; 5 = several times per week; 6 = every day; 7 = several times per day.
Ideological Polarization	In politics, a distinction is often made between "the left" and "the right". Where would you place yourself on the scale below, where 0 means left and 10 means right?

Opinion Polarization - Income inequality	Some people believe that differences in income should increase in our country. Others feel that they should decrease. Still others hold an opinion that lies somewhere in between. Where would you place yourself on a scale from 1 to 5, where 1 means that differences in income should increase and 5 means that these should decrease?
Opinion Polarization - Immigration	In the Netherlands, some people believe that immigrants are entitled to live here while retaining their own culture. Others feel that they should adapt entirely to Dutch culture. Where would you place yourself on a scale of 1 to 5, where 1 means that immigrants can retain their own culture and 5 means that they should adapt entirely?
Opinion Polarization - EU integration	Some people and political parties feel that European unification should go a step further. Others think that European unification has already gone too far. Where would you place yourself on a scale from 1 to 5, where 1 means that European unification should go further and 0 means that it has already gone too far?
Education	Highest level of education irrespective of diploma – 1 = primary school; 2 = vmbo (intermediate secondary education, US: junior high school); 3 = havo/vwo (higher secondary education/preparatory university education, US: senior high school); 4 = mbo (intermediate vocational education, US: junior college); 5 = hbo (higher vocational education, US: college); 6 = wo (university); 7 = other; 8 = Not yet completed any education; 9 = Not (yet) started any education.

Age	Age of the household member – integer.
Male	Gender – 1 = Male; 0 = Female.
Urban	Urban character of place of residence. – 5 = Extremely urban; 4 = Very urban; 3 = Moderately urban; 2 = Slightly urban; 1 = Not urban. Urban character: Surrounding address density per km2; extremely urban = 2,500 or more; very = 1,500 to 2,500; moderately = 1,000 to 1,500; slightly = 500 to 1,000; not = less than 500.
Married	Civil status. – 1 = Married; 0 = Separated; 0 = Divorced; 0 = Widow or widower; 0 = never been marred.
Ethnicity	Origin. – 0 = Dutch background; 101 = First generation foreign, Western background; 102 = First generation foreign, non-western background; 201 = Second generation foreign, Western background; 202 = Second generation foreign, non-western background; 999 = Origin unknown or part of the information unknown (missing values).

C.0.2 *Descriptive Statistics*

Table C.2: Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Use Social Media	32637	0.665	0.472	0.000	1.000
D Male	62535	0.488	0.500	0.000	1.000
D Married	62535	0.466	0.499	0.000	1.000
Urban	61951	2.928	1.319	1.000	5.000
D Primary School	62535	0.094	0.291	0.000	1.000
D High School	62535	0.282	0.450	0.000	1.000
D College	62535	0.132	0.339	0.000	1.000
D No Schooling	62535	0.038	0.190	0.000	1.000
D Dutch	62535	0.616	0.486	0.000	1.000
D Foreign Non Western	62535	0.049	0.216	0.000	1.000
Pol Polarization	31512	0.809	0.581	0.093	2.385
ln Hours Per Week SM	22549	1.178	0.782	0.000	4.796
ln Frequency SM	25287	1.743	0.416	0.693	2.079
ln Age	62463	3.579	0.745	0.000	4.663
pol immigration	35333	0.831	0.557	0.346	2.793
pol Opinion UE	33393	0.860	0.511	0.377	2.214
pol inequality	34837	0.849	0.528	0.189	2.904

Source: LISS PANEL

C.0.3 *Robustness Checks*

Table C.3: Placebo System-GMM - Ideological Polarization

	Pol_Polarization	Pol_Polarization	Pol_Polarization
Placebo_Use_SM	0.009		
	(0.01)		
Placebo_ln_Hours		0.000	
		(0.01)	
Placebo_ln_Frequency			0.010
			(0.02)
L.Pol_Polarization	0.108***	0.096***	0.100***
	(0.03)	(0.03)	(0.03)
ln_Age	0.022	0.035	0.023
	(0.03)	(0.03)	(0.03)
D_No_Schooling	-7.872	-0.362***	-0.355***
	(10.45)	(0.02)	(0.02)
D_Primary_School	-0.006	-0.077	-0.045
	(0.08)	(0.09)	(0.09)
D_High_School	-0.034	-0.033	-0.038*
	(0.02)	(0.02)	(0.02)
D_College	0.010	0.010	0.011
	(0.02)	(0.02)	(0.02)
D_Male	0.065***	0.083***	0.071***
	(0.02)	(0.02)	(0.02)
D_Married	-0.057***	-0.053***	-0.056***
	(0.02)	(0.02)	(0.02)
Urban	0.014**	0.020***	0.019***
	(0.01)	(0.01)	(0.01)
D_Dutch	-0.019	-0.021	-0.008
	(0.03)	(0.03)	(0.03)
D_Foreign_Non_Western	-0.023	-0.043	0.021
	(0.05)	(0.06)	(0.06)
Year	YES	YES	YES
_cons	0.641***	0.574***	0.597***
	(0.11)	(0.11)	(0.11)
Hansen	0.824	0.941	0.738
AR(1)	0.000	0.000	0.000
AR(2)	0.388	0.375	0.306
N instruments	29.000	29.000	29.000
KMO	0.874	0.907	0.925
N	10744	7733	8535

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.4: Placebo System-GMM - Opinion Polarization (Income Inequality)

	pol_inequality	pol_inequality	pol_inequality
Placebo_Use_SM	-0.013		
	(0.01)		
Placebo_ln_Hours		0.007	
		(0.01)	
Placebo_ln_Frequency			-0.020
			(0.01)
L.pol_inequality	0.008	-0.200	0.001
	(0.02)	(0.26)	(0.02)
ln_Age	-0.000	0.013	0.009
	(0.02)	(0.02)	(0.02)
D_No_Schooling	0.417**	0.452**	0.447***
	(0.17)	(0.19)	(0.16)
D_Primary_School	0.112***	0.129**	0.115***
	(0.04)	(0.05)	(0.04)
D_High_School	0.027**	0.043**	0.037***
	(0.01)	(0.02)	(0.01)
D_College	0.007	0.011	0.005
	(0.02)	(0.02)	(0.02)
D_Male	0.061***	0.091***	0.070***
	(0.01)	(0.03)	(0.01)
D_Married	0.007	0.004	-0.004
	(0.01)	(0.02)	(0.01)
Urban	0.004	0.009	0.004
	(0.00)	(0.01)	(0.00)
D_Dutch	-0.003	0.017	0.006
	(0.02)	(0.02)	(0.02)
D_Foreign_Non_Western	0.009	0.013	0.016
	(0.03)	(0.04)	(0.03)
Year	YES	YES	YES
_cons	0.792***	0.862***	0.781***
	(0.07)	(0.22)	(0.07)
Hansen	0.023	0.973	0.029
AR(1)	0.000	0.208	0.000
AR(2)	0.346	0.720	0.540
N instruments	30	30	30
KMO	0.912	0.951	0.961
N	17867	12609	14367

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.5: Placebo System-GMM - Opinion Polarization (EU Integration)

	pol_Opinion_UE	pol_Opinion_UE	pol_Opinion_UE
Placebo_Use_SM	0.014		
	(0.01)		
Placebo_ln_Hours		0.001	
		(0.00)	
Placebo_ln_Frequency			-0.002
			(0.01)
L.pol_Opinion_UE	-0.000	-0.001	-0.135
	(0.35)	(0.26)	(0.30)
ln_Age	0.094**	0.089***	0.106***
	(0.04)	(0.03)	(0.04)
D_No_Schooling	-0.004	0.013	0.017
	(0.08)	(0.07)	(0.08)
D_Primary_School	0.050	0.025	0.023
	(0.04)	(0.04)	(0.05)
D_High_School	0.013	0.026	0.020
	(0.01)	(0.02)	(0.02)
D_College	0.134**	0.132***	0.150***
	(0.05)	(0.04)	(0.05)
D_Male	0.142***	0.152***	0.173***
	(0.05)	(0.04)	(0.05)
D_Married	0.023	0.041**	0.038**
	(0.01)	(0.02)	(0.02)
Urban	0.008*	0.014***	0.012**
	(0.00)	(0.01)	(0.01)
D_Dutch	-0.015	-0.013	-0.023
	(0.02)	(0.02)	(0.03)
D_Foreign_Non_Western	0.034	0.016	0.021
	(0.03)	(0.03)	(0.04)
Year	YES	YES	YES
_cons	0.385***	0.337***	0.398***
	(0.14)	(0.12)	(0.13)
Hansen	0.191	0.679	0.443
AR(1)	0.198	0.082	0.239
AR(2)	0.741	0.540	0.399
N instruments	30	30	30
KMO	0.913	0.951	0.961
N	17028	11933	13630

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.6: Placebo System-GMM - Opinion Polarization (Immigration)

	pol_immigration	pol_immigration	pol_immigration
Placebo_Use_SM	-0.002		
	(0.01)		
Placebo_ln_Hours		-0.001	
		(0.00)	
Placebo_ln_Frequency			-0.020
			(0.01)
L.pol_immigration	0.046**	0.045*	0.044**
	(0.02)	(0.02)	(0.02)
ln_Age	-0.042**	-0.035*	-0.038**
	(0.02)	(0.02)	(0.02)
D_No_Schooling	0.419***	0.413***	0.432***
	(0.11)	(0.11)	(0.12)
D_Primary_School	0.135***	0.152***	0.148***
	(0.04)	(0.05)	(0.05)
D_High_School	0.051***	0.058***	0.054***
	(0.01)	(0.02)	(0.01)
D_College	0.035*	0.027	0.030
	(0.02)	(0.02)	(0.02)
D_Male	0.038***	0.044***	0.042***
	(0.01)	(0.01)	(0.01)
D_Married	0.002	-0.003	-0.005
	(0.01)	(0.01)	(0.01)
Urban	0.008**	0.009*	0.008*
	(0.00)	(0.00)	(0.00)
D_Dutch	-0.015	-0.023	-0.027
	(0.02)	(0.02)	(0.02)
D_Foreign_Non_Western	0.270***	0.248***	0.256***
	(0.04)	(0.04)	(0.04)
Year	YES	YES	YES
_cons	0.866***	0.843***	0.893***
	(0.07)	(0.08)	(0.08)
Hansen	0.164	0.092	0.033
AR(1)	0.000	0.000	0.000
AR(2)	0.260	0.323	0.531
N instruments	30	30	30
KMO	0.911	0.951	0.960
N	18223	12861	14647

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.7: Robustness System-GMM Estimations - Ideological Polarization

	Pol_Polarization	Pol_Polarization	Pol_Polarization
Use_SM	-0.069***		
	(0.02)		
ln_Hours_Per_Week_SM		-0.034**	
		(0.02)	
ln_Frequency_Use_SM			-0.053*
			(0.03)
L.Pol_Polarization	0.110***	0.095***	0.122***
	(0.03)	(0.03)	(0.03)
ln_Age	-0.060**	-0.038	-0.050*
	(0.03)	(0.03)	(0.03)
D_Male	0.061***	0.083***	0.066***
	(0.02)	(0.02)	(0.02)
_cons	0.991***	0.897***	0.983***
	(0.11)	(0.12)	(0.13)
hansenp	0.012	0.551	0.227
AR(1)	0.000	0.000	0.000
AR(2)	0.454	0.794	0.553
N instruments	18	18	18
KMO	0.869	0.874	0.877
N	10807	7777	9196

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.8: Robustness System-GMM - Opinion Polarization (Immigration)

	pol_immigration	pol_immigration	pol_immigration
Use_SM	-0.026		
	(0.02)		
ln_Hours_Per_Week_SM		-0.023*	
		(0.01)	
ln_Frequency_Use_SM			-0.047*
			(0.03)
L.pol_immigration	0.046**	0.055**	0.060***
	(0.02)	(0.03)	(0.02)
ln_Age	-0.071***	-0.065***	-0.076***
	(0.02)	(0.02)	(0.02)
D_Male	0.030**	0.029**	0.028**
	(0.01)	(0.01)	(0.01)
_cons	1.060***	1.034***	1.129***
	(0.08)	(0.09)	(0.11)
Hansen	0.024	0.035	0.012
AR(1)	0.000	0.000	0.000
AR(2)	0.558	0.688	0.553
N instruments	18.000	18.000	18.000
KMO	0.904	0.907	0.908
N	16684	11705	13371

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.9: Robustness System-GMM - Opinion Polarization (UE integration)

	pol_Opinion_UE	pol_Opinion_UE	pol_Opinion_UE
Use_SM	-0.133***		
	(0.05)		
ln_Hours_Per_Week_SM		0.008	
		(0.02)	
ln_Frequency_Use_SM			-0.043
			(0.05)
L.pol_Opinion_UE	0.151***	0.100**	0.120***
	(0.03)	(0.04)	(0.03)
ln_Age	0.002	0.076***	0.048*
	(0.03)	(0.02)	(0.03)
D_Male	0.122***	0.145***	0.141***
	(0.01)	(0.01)	(0.01)
_cons	0.759***	0.404***	0.582***
	(0.12)	(0.09)	(0.19)
Hansen	0.005	0.007	0.001
AR(1)	0.000	0.000	0.000
AR(2)	0.945	0.696	0.770
N instruments	11.000	14.000	10.000
KMO	0.901	0.907	0.908
N	16684	11705	13371

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.10: Robustness System-GMM - Opinion Polarization (Income Inequality)

	pol_inequality	pol_inequality	pol_inequality
Use_SM	-0.037		
	(0.03)		
ln_Hours_Per_Week_SM		0.007	
		(0.02)	
ln_Frequency_Use_SM			-0.008
			(0.03)
L.pol_inequality	0.034	0.005	0.007
	(0.03)	(0.03)	(0.02)
ln_Age	-0.022	0.014	-0.007
	(0.02)	(0.02)	(0.02)
D_Male	0.052***	0.064***	0.059***
	(0.01)	(0.01)	(0.01)
_cons	0.896***	0.746***	0.849***
	(0.09)	(0.09)	(0.12)
Hansen	0.001	0.044	0.014
AR(1)	0.000	0.000	0.000
AR(2)	0.078	0.280	0.328
N instruments	13.000	13.000	13.000
KMO	0.904	0.906	0.909
N	16684	11705	13371

5

SUMMARY AND CONCLUSION

5.1 SUMMARY

This dissertation aims to contribute to understanding the impacts of new technologies on human attitudes and decision-making processes. It pursues this objective by exploring, through empirical research, two topics.

First, it investigates the potential impacts of algorithms on the job market. More precisely, Chapter 2 examines whether the recommendations made by an algorithm are perceived differently from those made by a human (expert) and potentially lead to different outcomes of the hiring process. Automated recommendations are being used more and more in different domains of social life, and the job market is one field where algorithms play an ever-increasing role, from analyzing workers' performance to selecting applicants to opening positions. While there is already vast literature regarding the inner functioning, the accuracy, and the efficiency of these algorithms solutions, little attention has been given to the human involvement in hiring decisions supported by algorithms.

Chapter 2 contributes to this body of research by analyzing – through a field experiment – if there is a difference in the assessment made by employers of job-seekers recommended by algorithms in comparison to their evaluation of candidates recommended by human choice. The Chapter also asks whether the previous attitudes of recruiters in relation to algorithms solutions and human decision-making affected the reactions to the recommendations provided by these two types of sources. The study does not find a significant difference in the responses of hirers to recommendations made by algorithms and humans. While characteristics such as work experience or speaking English affect the rating of a candidate's qualification and the hiring interest, the source of recommendation does not. We also do not find an overall effect in our follow-up survey on job offers or CV retention.

Concerning labor market specific beliefs about algorithms, decision-makers, on average, believe that human resource specialists are better at taking non-standard profiles in terms of academic or professional background into account and more trustworthy than algorithms. However, respondents also believe that human resource specialists are more prone to bias in their selection of candidates. Concerning recommendations in general across all domains of life, our respondents prefer human over algorithmic advice.

In the analysis of heterogeneity in the treatment effect along these dimensions, this study does not find an interaction of the treatment effect with the labor market specific beliefs about algorithms, potentially reflecting the fact that most participants do not believe that one source of recommendations dominates the other in all three dimensions we elicited. It is rather the general preference for algorithms that seems to matter for the qualification rating of a candidate. On the one hand, decision-makers with a general preference for algorithmic advice give significantly higher ratings if a candidate was recommended by an algorithm compared to when the candidate was recommended by a human resource expert. On the other hand, decision-makers with a general preference for human advice give significantly lower ratings to candidates recommended by algorithms when compared with candidates recommended by humans. In sum, Chapter 2 showed that hirers are influenced not by their views regarding the capacity of an algorithm to find adequate candidates but by their general preference for algorithms or humans in spheres other than the labor market.

The second topic analyzed by this dissertation is the impact of internet and social media usage in the realm of political attitudes in democratic countries. With the rise in access to the internet worldwide, social media platforms have demonstrated enormous potential to transform how communities interact and communicate, reshaping the public sphere and playing a fundamental role in the political debate of different nations. After a period of optimism with the new possibilities brought by the enhanced access to the internet, which could purportedly improve the flow of information and boost political participation, there is nowadays a general concern with the effects of information technologies on the functioning of democratic systems.

A major topic of debate in this context is the relationship between the enhancement of polarization within democratic societies, a process that has been observed in different countries over the last years, and the increase in internet and social media usage. In this context, important studies have argued that online media is the main cause of the heightening of political polarization because they expose individuals to unbalanced and selective information, reinforcing previous beliefs and leading to the development of more extreme positions. On the other hand, some authors have questioned the existence of a causal link between online media use and greater levels of polarization.

Chapters 3 and 4 of this dissertation shed light upon some overlooked aspects of the dynamics of online media and their effects on the political debate, helping to make a more accurate diagnosis of this complex phenomenon. They contribute to this discussion by empirically testing the causal relationship between internet and social media use and political polarization in different countries.

Chapter 3 provides an important contribution to the literature on affective polarization in multiparty democracies. The study examines whether enhanced feelings of hostility towards opposing candidates were caused by internet and social media use in Brazil. Contrary to widespread assumptions in the literature, the study does not find evidence that access to the internet and social media leads to the intensification of affective polarization in Brazil. The relationship between social media usage and political polarization is also the subject of Chapter 4. Different from common belief, the study finds that social media use does not enhance but rather alleviates political polarization. In sum, Chapter 4 identified a significant negative effect of accessing social media on polarization in the Netherlands, controlling for demographic and individual characteristics and previous political polarization. This finding holds for different measures of social media usage — dummy (yes vs. no), intensity (time spent), and frequency.

5.2 LIMITATIONS AND AVENUE FOR FUTURE RESEARCH

The research developed in this thesis has different limitations and blind spots. Findings are informative and stimulate further research. More work is needed to illuminate other aspects of the multifaceted impacts of new technologies on human decision-making and attitudes.

Chapter 2 contributes to the emerging literature on the human side of algorithms in the labor market using field experiment data. One of the critical concerns of field experiment data is the generalizability (also known as external validity) of inferences. The question of generalizability is key for policymakers for it determines whether one can extend the results for populations beyond those studied in the experiment (Gertler et al., 2016). Given that the study took place in the normal circumstances of a job fair, using previous implementation routines of the host institution, the results obtained represent an advance to previous lab experiments that studied the response to human vs. algorithmic recommendations (e.g., Logg et al. (2019), Castelo et al. (2019), Dietvorst et al. (2015)).

The investigation shows that employers respond differently to recommendations given by automated systems and humans, according to their pre-existing preferences on algorithms and humans. These preferences were constructed based on experiences with algorithms across all domains of life (e.g., financial advice, shopping). This finding is likely general, for there is no apparent reason to think that hirers are idiosyncratic in this regard. However, it is important to test whether it remains valid in other settings. Thus, an avenue for future research is to explore whether one can replicate this experiment finding in contexts other than the labor market.

More research is also needed to understand better some aspects of the complex relationship between humans and algorithms. Given the ubiquity of algorithmic use in decision-making, it is still necessary to understand which features of automated systems can help overcome unfounded preconceptions when using algorithmic advice.

Chapters 3 and 4 evaluate the impact of internet and social media use on political polarization in Brazil and the Netherlands. The major restrictions and blind spots of these chapters refer to data limitations — both of them relied on secondary data. Chapter 3 uses cross-sectional data from Brazil and Chapter 4 longitudinal data from the Netherlands. First, information provided by the surveys restricted further analysis on the mechanisms behind the impacts of online media on political attitudes. In order to test the mechanisms behind the echo chamber and filter bubble theories, it would be important to obtain panel data measuring whether and the extent to which online media users are exposed to different points of view, news and cross-cutting communities. Literature on this topic will surely benefit from increased data availability.

Second, it would be interesting to analyze qualitative aspects of online communication, such as the nature of the content users are exposed to. It seems important to investigate how people react to diverse and cross-cutting information. On the one hand, it is expected that exposure to cross-cutting information will increase deliberation and understanding among individuals, decreasing polarization (Mutz, 2006). On the other hand, exposure to diverse and cross-cutting news may exacerbate political polarization. Some studies suggest that people who are exposed to information that conflict with their own beliefs are prone to be uncivil, which would increase differences between groups and thus polarization (Schumann, 2014; Whitty, 2016; Kim & Kim, 2019). As suggested by Barberá et al. (2015), exploring qualitative data on the web and social media platforms, rather than only quantitative surveys, is an opportunity to understand further the mechanisms behind political polarization and online media.

Another critical question is how the impacts of social networking on polarization vary across different types of citizens and social media platforms. As noted by Bugarin & Portugal (2021), there may be significant heterogeneity in citizens' behavior and attitudes in reaction to social media information. Informed and uninformed voters may react differently to information disseminated by means of social network technologies. Thus, it would be interesting to search for data that allows testing the potentially different effects of online media use across the different types of citizens. In addition to that, each platform (e.g., Twitter, Facebook, YouTube, WhatsApp, Telegram, TikTok) has its functioning methods, design features, content guidelines, and moderation policies, which define how users interact with each other, which information will be shared, and what stories will gain traction in the online community.[1] The size, composition, and demographics of their audiences also vary substantially. In this context, it is important to identify whether some social media platforms are more prone

1 For example, Facebook recently announced it would overhaul its algorithms to address racial, gender, and other discriminations caused by the use of personalization systems (The Guardian, 2022).

to impact political polarization than others and, if so, which factors influence these effects. Another important point of investigation in this scenario is if smaller and more segmented platforms (e.g., Gab, Parler, and 4chan) may have a larger capacity to disseminate messages of extreme political content and generate more intense responses from users.

Another central topic relates to the effect of laws, regulations, and policies adopted both by public bodies and private companies to moderate content and speech on social media platforms. Governments of several democratic nations have implemented legal measures to ban certain types of content deemed offensive or dangerous. Social media platforms, in turn, have taken steps – such as the suspension of accounts and the use of warning labels – to set boundaries to the dissemination of political messages considered false or violent. Given the relevance of social networking in today's political debate, it is vital to investigate the effects of these regulations and measures on the online public sphere.

5.3 CONCLUSION

As a whole, this dissertation contributes to the vast literature that analyses changes brought by new technologies and automation through algorithms. Automated systems have proved to be a valuable tool in several domains to deal with information overload and inherent human limitations in assessing information consistently. They provide individuals with suggestions that are likely to be of interest to them, including personalized shopping, filtered news, and information. Furthermore, they make recommendations for medical diagnoses, judicial decisions, and preselection of job candidates. Hence, these tools could potentially improve human decision-making and experiences but also pose challenges to society. In this context, this dissertation provides insights into how individuals process and react to information filtered and recommended by automated systems.

First, this study calls attention to the fact that the successful use of automated systems in any decision-making process will also depend on how individuals perceive and react to them. Potential interactions between human and nonhuman agents will ultimately affect decision-making assisted by algorithms. The importance of this is by now well established. Kahneman, Sibony and Sustein, well-known scholars, recently published a book about the limitations of human judgments (Kahneman et al., 2021). Among the topics raised, the authors noticed that a potential aversion to the use of algorithms in decision-making, a well-studied phenomenon in the literature, represented an impediment to the more frequent use of algorithms in various domains and the improvement of decision-making powered by automated systems. A lot remains to be done to fully uncover the implications of human-algorithm interactions to decision-making.

The first part of the dissertation contributes to this understanding and represents a concrete finding in the domain of the labor market. I show that when confronted with identical recommendations produced by algorithms or humans, employers' reactions differ depending on their previous preference for these two sources of advice. The study showed that individuals are influenced not by perceptions regarding the capacity of an algorithm to the particular task of selecting adequate candidates but rather by a general feeling toward algorithms and humans, which were based on their experiences with algorithm recommendations in spheres other than the labor market.

This finding suggests that firms introducing algorithms into their hiring process should consider that pre-existing attitudes concerning algorithms can affect the evaluation of candidates. Thus, they should ponder that hiring processes mediated by algorithms should now confront a bias (towards the source of the advice) that has no relation to any of the characteristics of the candidate. Here it is worth drawing a parallel with the seminal work of Bertrand & Mullainathan (2004), who showed that merely assigning African-American- or White-sounding names to CVs of fictitious candidates affected the outcomes of hiring processes because of racial biases. Similarly, this study shows that substituting the human advice with an algorithm in a hiring process introduces a new bias - based on a general preference towards the source of the advice - that will potentially change how hirers judge candidates.

The findings of Chapter 2 provide important implications for policies aimed at enhancing decision-making — in particular those aimed at optimizing the algorithm-human relationship. As organizations rely increasingly on algorithms to assist hiring decisions, it is crucial to understand and prevent potential unintended effects related to the use of algorithms. The study showed that pre-existing preferences on algorithms and humans could affect an individual evaluation of candidates in a hiring process. This preference is established by experiences with algorithms in spheres other than the selection of candidates on a hiring process. Evidence shows that even in developed countries, there still seems to be particularly low levels of knowledge regarding the use of algorithms in the labor market. A study in the United Kingdom, Germany, France, Poland, Spain, and Italy showed that only 31 percent of the population knew that algorithms are often used to select candidates in hiring processes (Grzymek & Puntschuh, 2019).

In this context, promoting algorithmic literacy geared to instruct decision-makers on algorithm systems' rationale, strengths, and limitations may play a significant role in optimizing algorithm-human interactions in decision-making. Another intermediate goal in terms of policies involves designing and implementing more transparent algorithm solutions. Although transparency can come as a trade-off with the algorithm's complexity and performance, the disclosure of the algorithms' procedures would afford more opportunities for an understanding of how algorithms perform tasks and could be thus crucial for its effective use in decision-making.

The second part of this dissertation deepens our understanding of how and whether information filtered by automated systems on the internet and social media influences political polarization processes. Political polarization within deliberative democracies is a general trend reshaping the public spheres of several nations, as Chapter 3 has shown regarding Brazilian politics. Different indicators report an accentuation of the fissures in the political debate within numerous countries over the last decades. The problems and risks of this process of exacerbation of polarization are well-known. Excessive polarization may lead to widening political divides in such a dimension that makes compromises impossible, eroding the capacity of cooperation within society and undermining the effective functioning of political systems (Fishkin, 2009; Sunstein, 2001).

Over the last years, different accounts have described a strengthening of political segregation and partisanship in different countries and a radicalization of political discourse, associating this scenario with the massive diffusion of new technologies, particularly social media. Many scholars have pointed out that the complex algorithms behind internet searching tools and social media platforms rely on leveraging individual users' data to provide highly personalized information to maximize scroll time. This would often expose users to a more biased and unbalanced set of information than they would have had they chosen the content on their own, which would then promote more extremist positions and enhance hostility towards opposite ideas and figures.

The results of the second part of this dissertation recollect the role that internet and social media may have in promoting a pluralistic environment of exchange of ideas. Finding that internet and social media use are not the cause of recent affective polarization in Brazil, and even that social media use can promote depolarization in the Netherlands should attenuate the concerns that internet and social media are a significant threat to the well-functioning of democracy nowadays. Furthermore, it suggests that the enhancement of the political divide experienced recently by different democracies may have much deeper roots, like the rising economic inequality (Voorheis et al., 2015), rising trade exposure (Dorn et al., 2020) and the weakening of cross-cutting social identities.[2]

In sum, this dissertation adds to the literature exploring the transformations led by new technologies and automated systems. First, my work highlights how behavioral economics can provide useful insights into the use of algorithms in decision-making. It shows how psychological factors, specifically people's general preferences on algorithms or human advice, determine the judgment on how an individual is qualified for a job. It shows that to effectively use algorithms in human decision-making, it is essential to consider how boundedly rational individuals make judgments and react

2 The decline of cross-cutting identities is at the root of affective polarization, according to Iyengar et al. (2019).

to algorithm solutions. Second, this thesis offers a different point of view on how individuals process filtered information by internet search tools and social media algorithms in the political domain. For influential scholars, these new technological tools undermine democratic ideals by enhancing social fragmentation and weakening mutual understanding between different communities. The study demonstrated that the concerns regarding the isolating role of online platforms are not verifiable in some contexts and that social media may well play a fostering political moderation in society.

REFERENCES

Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A. & Bonneau, R. (2015), 'Tweeting from left to right: Is online political communication more than an echo chamber?', *Psychological science* **26**(10), 1531–1542.

Bertrand, M. & Mullainathan, S. (2004), 'Are emily and greg more employable than lakisha and jamal? a field experiment on labor market discrimination', *American Economic Review* **94**(4), 991–1013.

Bugarin, M. & Portugal, A. (2021), 'Dormant voters, social media, and strategic radicalism – preliminary results', *Anais do Simpósio Interdisciplinar sobre o Sistema Político Brasileiro e XI Jornada de Pesquisa e Extensão da Câmara dos Deputados* pp. 191–204.

Castelo, N., Bos, M. W. & Lehmann, D. R. (2019), 'Task-dependent algorithm aversion', *Journal of Marketing Research* **56**(5), 809–825.

Dietvorst, B. J., Simmons, J. P. & Massey, C. (2015), 'Algorithm aversion: People erroneously avoid algorithms after seeing them err.', *Journal of Experimental Psychology: General* **144**(1), 114.

Dorn, D., Hanson, G., Majlesi, K. et al. (2020), 'Importing political polarization? the electoral consequences of rising trade exposure', *American Economic Review* **110**(10), 3139–83.

Fishkin, J. (2009), *When the people speak: Deliberative democracy and public consultation*, Oup Oxford.

Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B. & Vermeersch, C. M. (2016), *Impact evaluation in practice*, World Bank Publications.

Grzymek, V. & Puntschuh, M. (2019), 'What europe knows and thinks about algorithms results of a representative survey. bertelsmann stiftung eupinions february 2019'.

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N. & Westwood, S. J. (2019), 'The origins and consequences of affective polarization in the united states', *Annual Review of Political Science* **22**, 129–146.

Kahneman, D., Sibony, O. & Sunstein, C. R. (2021), *Noise: A flaw in human judgment*, Little, Brown.

Kim, Y. & Kim, Y. (2019), 'Incivility on facebook and political polarization: The mediating role of seeking further comments and negative emotion', *Computers in Human Behavior* **99**, 219–227.

Logg, J. M., Minson, J. A. & Moore, D. A. (2019), 'Algorithm appreciation: People prefer algorithmic to human judgment', *Organizational Behavior and Human Decision*

Processes **151**, 90–103.

Mutz, D. C. (2006), *Hearing the other side: Deliberative versus participatory democracy*, Cambridge University Press.

Schumann, S. (2014), *How the Internet shapes collective actions*, Springer.

Sunstein, C. R. (2001), *Republic. com*, Princeton University Press.

The Guardian (2022), 'Facebook to axe 'discriminatory' algorithm in us government settlement, 22 of june'.
URL: *https://www.theguardian.com/technology/2022/jun/21/facebook-lawsuit-settlement-advertising-lookalike-audience-doj?CMP=share_btn_link*

Voorheis, J., McCarty, N. & Shor, B. (2015), 'Unequal incomes, ideology and gridlock: How rising inequality increases political polarization', *Available at SSRN: https://ssrn.com/abstract=2649215 or http://dx.doi.org/10.2139/ssrn.2649215* .

Whitty, M. T. (2016), *Cyberpsychology: The study of individuals, society and digital technologies*, John Wiley & Sons.

IMPACT PARAGRAPH

The impact paragraph of this doctoral dissertation is added in compliance with article 22.5 of the "Regulations for obtaining the doctoral degree at Maastricht University" decreed by resolution of the board of deans, dated 1 October 2020.

The findings of this dissertation have important consequences for the debate regarding the development of sound policies in the context of digital transformation. As digital technologies become more and more ubiquitous and disrupt practically every field of human experience, solid policy action is of utmost necessity to address the challenges and tap the potential associated with the accelerated process of technological progress. The design and implementation of a coherent policy agenda depend on evidence-based analysis, which provides a robust foundation for the conception of policy interventions. In this context, this dissertation contributed to understanding the impacts of new technologies on human decision-making and attitudes in two domains.

First, it examined the impacts of algorithms on the job market. More precisely, Chapter 2 examined the effect of algorithmic advice on hiring decisions made by employers. In a world where algorithmic systems are being applied in various stages of the hiring process, understanding its effects in the labor market will continue to be of particular relevance in the coming years. Algorithms are now widely used for streamlining the screening of applications to job openings in online environments, and for recruiting and tracking employees' performance in influential firms. Chapter 2 contributes to this body of research by analyzing – through a field experiment – if there is an observable difference in the assessment made by employers of jobseekers recommended by algorithms in comparison to their evaluation of candidates recommended by human choice. The study also examines if the previous attitudes of recruiters in relation to algorithms solutions and human decision-making affected the reactions to the recommendations provided by these two types of sources.

The primary finding of this chapter is that employers respond differently depending on their attitudes towards algorithmic and human advice in domains beyond the labor market. Contrary to common assumptions, I find that, when using algorithmic aids in hiring procedures, employers are not influenced by specific reservations on the capacity of an algorithm to select a candidate but rather by a general feeling towards algorithms in spheres other than the labor market. On one hand, decision makers with a general preference for algorithmic advice give significantly higher ratings if a candidate was recommended by an algorithm compared to when the candidate was recommended by a human resource expert. On the other hand, decision makers with a general preference for human advice give significantly lower ratings to candidates recommended by algorithms when compared with candidates recommended by humans.

The findings of this paper shed light on an important, albeit overlooked, aspect regarding the widespread use of algorithm solutions in the labor market - the interaction between human perceptions and algorithms. As organizations rely increasingly on algorithms in hiring decisions, it is crucial to understand, correct and prevent difficulties related to algorithm aversion. While algorithmic governance becomes an ever more present reality, the findings of this paper indicate the relevance of policies aimed at optimizing the algorithm-human relationship. An important issue here is making investments in the enhancement of algorithmic literacy, geared to instruct decision - makers on the rationale of algorithm systems, their strengths, and their limitations. There still seems to be particularly low levels of knowledge regarding the use of algorithms in the labor market. Another conceivable policy solution involves design and implementation of more transparent algorithms solutions. Although transparency often comes as a trade-off with the algorithm's complexity and performance, the disclosure of algorithms into intelligible multistep procedures would afford more opportunities for alignment on human- algorithm decision processes and could be thus crucial for the effective use of algorithms.

Second, in Chapters 3 and 4 this dissertation contributed to one of the most controversial subjects in the public debate today: the role of technology and automated tools in the political arena of democratic systems. Over the last years, different accounts have described a strengthening of political segregation and partisanship in different countries and a radicalization of political discourse, associating this scenario to the massive dif- fusion of new technologies, particularly social media. According to this view, social media exposes individuals to a biased and imbalanced set of information, consolidating pre-existing political beliefs and enhancing hostility towards opposite ideas and figures. Furthermore, the exchange of information through social media would facilitate the diffusion of more extreme and intense messages than those encountered in physical interactions.

Despite the increased scholarly attention to the topic, empirical evidence on the effects of Internet and social media use on political polarization remains inconclusive. Empirical attempts to examine causal effects have been limited by identification challenges induced by self-reporting usage of Internet and social media – which typically result in biased outcomes. Many empirical studies document pure correlations, and are unable to make claims about causality. Moreover, most of the related studies have focused on the US context, and there is limited empirical evidence on other regions. The findings in these chapters suggest that contrary to what is suggested by the mainstream literature, internet and social media use may not be the cause of rising political polarization across countries. Chapter 3 finds that political polarization in Brazil cannot be attributed to internet or social media use. Chapter 4 goes further and shows that that social media use attenuates rather than drives polarization, a finding that holds for different measures of social media use - dummy (yes vs. no), intensity (time spent) and frequency. Such findings attenuate the widespread concern that internet and social media use is a significant motive of political polarization in society nowadays.

From the perspective of design and implementation of public policies, these findings have significant impacts since they indicate that the efforts to reduce the polarization of political environments should be directed to other issues rather than the regulation of social media. Indeed, there are strong arguments that suggest that the enhancement of the political divide experienced recently by different democracies may have much deeper roots, like the rising economic inequality, rising trade exposure and the weakening of cross-cutting social identities.

Political polarization within deliberative democracies is a general trend reshaping the public spheres of several nations, as Chapter 3 has shown regarding Brazilian politics. Different indicators report an accentuation of the fissures in the political debate within numerous countries over the last decades. The problems and risks of this process of exacerbation of polarization are well-known. Excessive polarization may lead to widening political divides in such a dimension that makes compromises impossible, eroding the capacity of cooperation within society and undermining the effective functioning of political systems. Extreme polarization also has important nonpolitical impacts, affecting labor markets, immigration fluxes, and economic perspectives. Therefore, adequate comprehension of the causes under- pinning the polarization process is vital for designing policies that may attenuate this trend. Chapters 3 and 4 of this dissertation shed light upon some overlooked aspects of the dynamics of social media and their effects on the political debate, helping to make a more accurate diagnosis of this complex phenomenon.

In recapitulation, this dissertation provides empirical evidence on the impacts of new technologies on human behavior, attitudes, and decision-making processes. Over the last decades, digital technologies have transformed almost every aspect of social life and human experience. Digital transformation has upended traditional industries at a remarkable speed, creating new products and services, designing groundbreaking business models, and engendering enormous economic value. The widespread dissemination of automated systems and algorithms allowed for massive efficiency gains and spurred a landscape of constant innovation. Social media caused structural changes in how people interact with each other and how societies communicate, bringing significant developments to markets and other domains, like the political debate within countries and the international relationships between nations. Overall, the findings in this dissertation may be of interest to researchers, policymakers, development practitioners, and the society at large.