Title
Essays in Experimental Economics

Author
Royer, Rebecca Charlotte

Publication Date
2024

UNIVERSITY OF CALIFORNIA SAN DIEGO

Essays in Experimental Economics

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Economics

by

Rebecca C. Royer

Committee in charge:

Professor James Andreoni, Co-Chair
Professor Gordon Dahl, Co-Chair
Professor Sally Sadoff
Professor Anya Samek
Professor Emanuel Vespa

2024

The Dissertation of Rebecca C. Royer is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2024

DEDICATION

I dedicate this dissertation to my dad, who inspired me to pursue a PhD.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to acknowledge Professors James Andreoni and Gordon Dahl for their support as co-chairs of my committee. I also thank my committee members for their guidance throughout the research process: Professors Sally Sadoff, Anya Samek, and Emanuel Vespa. I am grateful to Professor Charles Sprenger for his thoughtful feedback and support throughout graduate school. For preparing me in an invaluable way for graduate school and introducing me to the full research process, I thank Professor Daniel Benjamin.

My family, friends, and partner provided me with extensive support throughout my time as a student. Thank you all for celebrating each win with me and for filling me with confidence when I needed it.

The research in Chapter 1 was funded by NIH grant #SES-2117463 and the Russell Sage Foundation.

For contributions to Chapter 2, we thank Kristin Troutman, Adeline Sutton and research assistants in the Behavioral and Experimental Economics (BEE) Research Group for valuable research assistance. The research in this chapter was funded by the Kenneth and Anne Griffin Foundation, the Arnold Foundation and by NIH grant 1R01DK114238.

Chapter 3, in full, is a reprint of the material as it appears in the American Economic Association Papers and Proceedings. Bernheim, B. Douglas, Rebecca Royer, and Charles Sprenger. 2022. ”Robustness of Rank Independence in Risky Choice.” AEA Papers and Proceedings, 112: 415-20. The dissertation author was a primary investigator and author of this paper.

VITA

2015	Bachelor of Science, American University
2014–2016	Research Assistant, Economics Department, Office of the Comptroller of the Currency
2016–2018	Project Specialist, Center for Economic and Social Research, University of Southern California
2024	Doctor of Philosophy, University of California San Diego

PUBLICATIONS

"Robustness of Rank Dependence in Risky Choice" AEA Papers and Proceedings, vol. 112, pp 415-420, 2022.

ABSTRACT OF THE DISSERTATION

Essays in Experimental Economics

by

Rebecca C. Royer

Doctor of Philosophy in Economics

University of California San Diego, 2024

Professor James Andreoni, Co-Chair
Professor Gordon Dahl, Co-Chair

In Chapter 1, I investigate a novel channel of polarization: divergent interpretations of information. I conduct an online experiment with Democrats and Republicans in the US to study beliefs about racial discrimination in the labor market, a topic on which Democrats and Republicans are polarized. I find that Democrats' beliefs about racial labor market discrimination are responsive to information on racial wage disparities, while Republicans' beliefs are not. As a result, wage gap information fails to reduce (and even increases) the partisan difference in discrimination beliefs. Moreover, even after both groups agree about the extent of racial hiring discrimination, participants change their opinions about whether it is a problem depending on

their political affiliation, enabling disagreement in policy demand. Together, these findings highlight key challenges in using information to reduce polarization.

In Chapter 2, we leverage a randomized evaluation of an early childhood program to study the impact of early life investments on resilience to negative shocks. When the children in our study were 3-5 years old, they were randomized to a preschool program, a parenting program or to a control group. Ten years later, the children were exposed to school shut-downs during the Covid-19 pandemic. With nearly 900 observations, we show that the parenting program had a protective causal impact on the decrease in academic test scores during the year that schools were closed. While the control group saw a 0.31 SD decline in standardized test scores after Covid-19, the parenting group saw only a 0.12 SD decline. We provide a conceptual framework and evidence on potential mechanisms driving this effect.

In Chapter 3, we explore the robustness of rank independence of equalizing reductions with respect to experimental procedures. Bernheim and Sprenger (2020) devise and implement a novel test of rank-dependent probability weighting both in general and as formulated in cumulative prospect theory (CPT). They reject both hypotheses decisively. CPT cannot simultaneously account for the rank independence of ”equalizing reductions” for three-outcome lotteries, which it construes as indicating linear probability weighting, and the relationship between equalizing reductions and probabilities, which it interprets as indicating highly nonlinear probability weighting.

Chapter 1

Belief polarization about racial discrimination in hiring: Evidence from an information experiment

1.1 Introduction

Political polarization has been rising over the past forty years in the US, with Democrats and Republicans exhibiting decreasing overlap in their political views (Canen et al., 2021). Given the adverse effects of polarization, such as political gridlock (Binder, 2014; Mian et al., 2014), theoretical and empirical researchers have sought to investigate its sources.

Literature from political science and economics finds that polarization is driven by Democrats' and Republicans' exposure to distinct information (DellaVigna and Gentzkow, 2010; Zhuravskaya et al., 2020). A natural policy proposal for reducing polarization is therefore through information dissemination. Indeed, information can be effective at reducing belief polarization when the information is unequivocally relevant (Grigorieff et al., 2018; Mu, 2022; Haaland and Roth, 2021).

Less is known about how belief polarization responds to information that is open to interpretation. Information that voters encounter often requires processing, which may depend on one's model of the world. Consider, for example, how information on racial wage gaps affects beliefs about racial discrimination. If one believes that wage gap information reflects labor market discrimination, then this information may move one's beliefs about racial discrimination. If instead, one believes that wage gap information reflects differences in educational attainment, for example, then the information may not move beliefs about racial discrimination. If information is processed differently by Democrats and Republicans, the effect on belief polarization becomes unclear.

These patterns are especially relevant in the context of racial discrimination in the labor market. Democrats and Republicans are polarized on this topic, as Democrats believe there is more labor market discrimination than Republicans do (Alesina et al., 2021). Beliefs about racial discrimination are themselves important in that they drive demand for policies including affirmative action and redistribution, and are relevant for Diversity, Equity, and Inclusion (DEI) training. DEI training uses information to teach people about obstacles minorities face in the labor market, including racial discrimination. Despite their popularity, results on the effectiveness of DEI training are mixed (Chang et al., 2019), potentially because we know little about the types of information that affect beliefs about labor market discrimination.

In this paper, I conduct a pre-registered online information experiment to examine belief polarization about hiring discrimination against Black workers in the US. The experiment consists of a within-subject design with five rounds using a sample of 1100 Democrats and 1100 Republicans. I elicit quantified and incentivized beliefs about racial hiring discrimination using the method from Haaland and Roth (2021). Each round, participants receive potentially useful information and state their updated beliefs about racial hiring discrimination.

The experiment reveals several key patterns and results. At baseline, I first establish that Democrats and Republicans are polarized on this topic. Consistent with the literature, Democrats believe there is more racial hiring discrimination than Republicans do.

I find that Democrats update their beliefs about racial hiring discrimination in response to information on the Black-White wage gap, while Republicans do not. Democrats overestimate the Black-White wage gap at baseline and revise downward their beliefs about hiring discrimination when they learn that the wage gap is smaller than expected. In contrast, Republicans underestimate the wage gap at baseline but do not revise their beliefs about hiring discrimination in response. As a result, the belief gap about hiring discrimination between Democrats and Republicans slightly decreases, but not statistically significantly.

I then provide evidence on the role of educational attainment in explaining the wage gap. That is, I tell participants how much of the wage gap is explained by differences in educational attainment between Black and White workers. Both Democrats and Republicans substantially overestimate the extent to which educational attainment explains the Black-White wage gap, with Republicans overestimating even more than Democrats. Upon finding out that educational attainment explains less of the wage gap than they thought, Democrats revise upwards their beliefs about the extent of racial hiring discrimination. Republicans, on the other hand, do not significantly revise their beliefs about racial discrimination in response. This leads to a marginally significant widening of the belief gap.

A natural question is what drives the observed differences in belief-updating between Democrats and Republicans in response to wage gap information. One explanation could be that Republicans' hiring discrimination beliefs are more difficult to move than Democrats' beliefs in general. This explanation, however, is challenged by one of the five rounds in which Republicans' beliefs move more than Democrats' beliefs. A remaining explanation is that Democrats and Republicans hold different interpretations about the relationship between wage gaps and labor market discrimination. These divergent interpretations could reasonably arise through Democrats and Republicans forming models of the world using distinct sources of news.

At the end of my study, I replicate a finding from Haaland and Roth (2021) that learning the results from an experiment measuring racial discrimination closes the belief gap between Democrats and Republicans. Even though both groups then agree about the extent of hiring discrimination, I find that participants change their opinions about the information depending on their political affiliation.

Relative to their own baseline responses, Republicans are more likely than Democrats to decrease their belief that the observed discrimination is (1) a successful measure of discrimination, and (2) a problem. The difference between Democrats' and Republicans' updating behaviors is consistent with politically motivated reasoning, and highlights a channel through which convergence in beliefs may not yield convergence in policy demand. Recent literature identifies other cases in which information fails to reduce polarization in policy demand and asserts that this may be driven by Democrats' and Republicans' differing beliefs about the role of government (Haaland and Roth, 2021; Marino et al., 2023). My findings demonstrate that this may occur outside of people's beliefs about the government.

This paper highlights crucial limitations of one of the leading proposals for reducing polarization: information dissemination. Contrary to standard economic models that suggest information decreases belief polarization, I find that information may fail to reduce (and even increase) belief polarization when Democrats and Republicans have divergent interpretations of information. Furthermore, even when groups agree on the facts of a political topic, biased reasoning may enable the persistence of polarization in policy demand. As Democrats and Republicans become more polarized in their worldviews, these findings become increasingly relevant.

The paper proceeds as follows. Section 2 summarizes the literature and highlights my contribution. Section 3 describes my experimental design. Section 4 defines my hypotheses. Section 5 reviews my experimental findings in each round. Section 6 discusses, and section 7 concludes.

1.2 Literature on labor market discrimination beliefs

Recent literature explores beliefs about labor market discrimination as a mechanism for how information on labor market disparities affects demand for policy. Settele (2022) finds that exposing participants to a larger gender wage gap increases their demand for policies to combat the wage gap, likely through an increase in beliefs about the extent of gender discrimination in the labor market. Alesina et al. (2021) find that White Republicans are more likely to believe inequities are caused by individual actions, while White Democrats attribute inequities to systemic conditions, including discrimination. Together, these findings highlight that while information on labor market inequities may affect beliefs about discrimination, the relationship may differ for Democrats and Republicans.

In evaluating how people update beliefs about racial discrimination in hiring, we may be concerned that political motivations could lead to biased belief updating (Redlawsk, 2002; Slothuus and De Vreese, 2010). In line with biased updating, Thaler (2019) finds that Democrats believe information more when it suggests there is more racial discrimination in hiring than they thought, relative to information that suggests there is less. Republicans believe information more when it suggests that there is less racial discrimination in hiring than they thought. If motivated reasoning drives belief-updating patterns in my context, then it could dampen the effects of information on belief depolarization.

Haaland and Roth (2021) develop a method of eliciting quantified and incentivized beliefs about racial discrimination in hiring using results from a fake resume study. In fake resume studies, researchers send out fake resumes in response to real job postings. The resumes only differ in whether the applicant's name sounds White or Black, and the researchers measure how often the fake applicants receive callbacks for interviews. Haaland and Roth (2021) measure participants' hiring discrimination beliefs as their predictions of callback rates for applicants with Black-sounding names and applicants with White-sounding names in a fake resume study. I adopt their methodology of eliciting beliefs. The authors find that presenting participants with results from a similar experiment on racial discrimination successfully closes the partisan gap in beliefs. In this paper, I add to our collective knowledge about how belief polarization responds to information that may be interpreted differently by Democrats and Republicans. Because labor market discrimination is notoriously difficult to measure, understanding belief-updating in response to ambiguous information is especially important in this context.

1.3 Experiment

I conduct an online information experiment on the survey platform Prolific, a widely-used survey platform for social science research, using oTree software (Chen et al., 2016). The experiment was preregistered on AsPredicted.org (Project #127316) before data collection began in April 2023. I use a within-subject experimental design consisting of five rounds. Each round, participants receive some information and state their updated beliefs about racial hiring discrimination.

The primary outcome variable across rounds is participants' beliefs about racial discrimination in hiring. Following Haaland and Roth (2021), I measure beliefs about racial discrimination in hiring by asking participants to predict the results of Bertrand and Mullainathan (2004)'s fake resume study. In Bertrand and Mullainathan (2004) (hereafter, "BM"), researchers sent out fake resumes in response to real job postings. Resumes were randomized in terms of education, experience, and other qualifications listed, and systematically differed in whether the name on the resume sounded White or Black. The researchers measured how often employers contacted these fake applicants for an interview. They found that applicants with Black-sounding names needed to send out 50% more resumes than applicants with White-sounding names to receive a callback for an interview.

I elicit participants' predictions of callback rates for applicants with White-sounding names and for applicants with Black-sounding names in BM. This method of eliciting beliefs about racial discrimination in hiring is (1) quantified, which ensures comparability across participants' responses, and (2) incentivized, which increases the likelihood that participants are accurately reporting their beliefs (Gächter and Renner, 2010).

1.3.1 Design overview

At the start of the study, I describe the BM experiment to participants. They are told that researchers ran an experiment to measure racial discrimination in the labor market in which they sent out fake resumes in response to real job postings. The fake resumes had identical qualifications, and differed only in whether the name on the resume sounded White or Black. The researchers measured the callback rates for resumes with Black-sounding names and for resumes with White-sounding names to determine the extent to which employers discriminate.

After presenting participants with this information, I ask how much they agree that a difference in callback rates between applicants with Black-sounding names and White-sounding names would reflect that employers base their callback decisions in part on the race of the applicant. I also ask whether they believe that if BM finds a higher White callback rate than Black callback rate, this would be a problem that should be solved. Similarly, I then ask whether a higher Black callback rate would be a problem that should be solved. Then, the first round begins.

Round 1: Participants state their best guesses of the callback rates for applicants with White-sounding names and applicants with Black-sounding names in BM. That is, they are asked how many times a resume with a Black-sounding name had to be sent out on average to get one callback from an employer for an interview. Then, they are asked how many times they think a resume with a White-sounding name had to be sent out on average to get one callback from an employer for an interview. See Figure 1.1 for a screenshot. From this round, I calculate participants' baseline beliefs about racial discrimination in hiring.

Round 2: Participants are told the callback rate for applicants with Black-sounding names from BM. That is, they are told that a resume with a Black-sounding name had to be sent out 15 times on average to get one callback for an interview. Participants are then asked again for their best guess of the number of times that a resume with a White-sounding name had to be sent out on average to get one callback for an interview in BM. See Figure 1.2 for a screenshot.

Round 3: Participants are told that Black full-time workers in the US earn on average $844 per week, and asked for their best guess of the average weekly earnings for White full-time workers in the US. Then, participants are told that on average, White full-time workers in the US earn on average $1085 per week. Participants are then asked again for their best guess of the

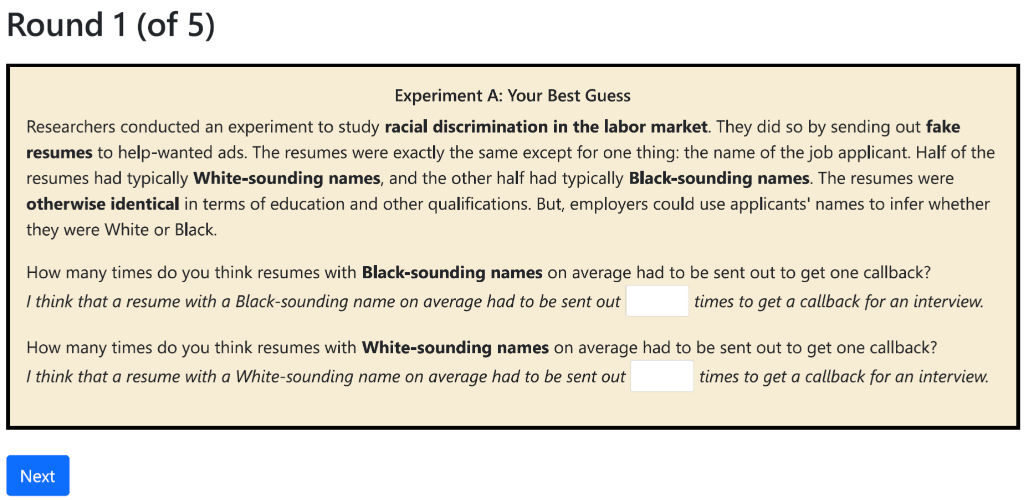

Figure 1.1. Round 1 Screenshot

number of times that a resume with a White-sounding name had to be sent out on average to get one callback for an interview in BM.

Round 4: Participants are asked how much (in %) of the Black-White wage gap they think is driven by (1) differences in educational attainment between Black and White workers and (2) employer discrimination against Black workers. Then, participants are told that statisticians estimate that 12% of the Black-White wage gap is driven by differences in educational attainment. Participants are then asked again for their best guess of the number of times that a resume with a White-sounding name had to be sent out on average to get one callback for an interview in BM. Finally, participants state their updated belief about how much (in %) of the Black-White wage gap they think is driven by employer discrimination against Black workers.

Round 5: Participants are presented with the callback rates for applicants with White-sounding names and Black-sounding names from Jacquemet and Yannelis (2012)'s fake resume study. Participants are then asked again for their best guess of the number of times that a resume with a White-sounding name had to be sent out on average to get one callback for an interview in BM. Participants are then told that a resume with a White-sounding name had to be sent out 10 times on average to get one callback for an interview in BM.

At the end of the study, participants are asked unincentivized questions about their

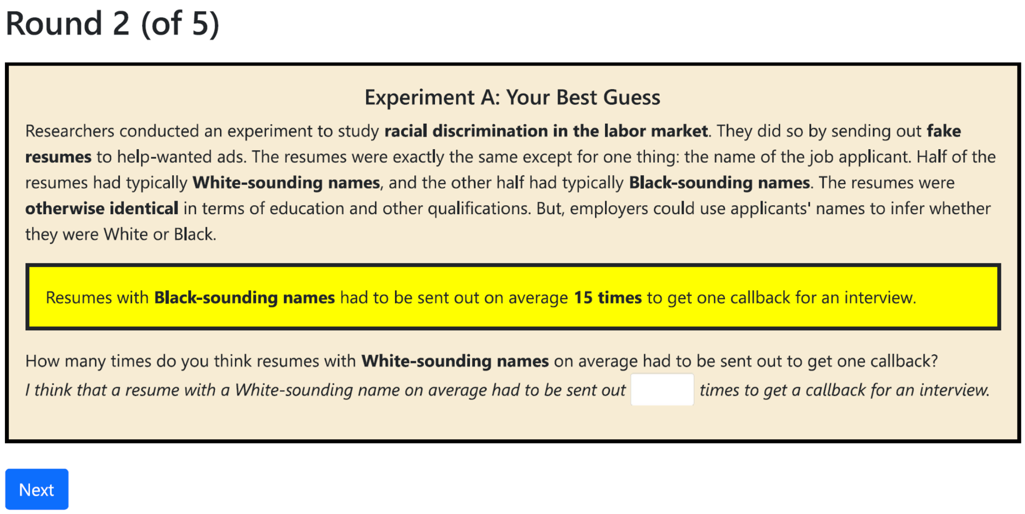

Figure 1.2. Round 2 Screenshot

thoughts on the BM study, their political views, and a couple of math questions.

1.3.2 Incentives

All participants receive a participation payment of at least $3 for finishing this study[1]. In addition, participants have the opportunity to earn a $2 bonus based on their answers. At the end of the study, one of the eligible questions is randomly selected to determine whether the participant receives the bonus. If the participant's guess is close enough to the correct answer on this randomly selected question, then they earn the bonus.

Eight questions are eligible to be selected for the bonus. In each of the five rounds, participants are asked the number of times a resume with a White-sounding name had to be sent out to receive one callback in BM. Participants' answers to this question in each round are eligible to be selected for the bonus, and if selected, participants receive the bonus if they are within one unit of the correct answer. In Round 1, participants are asked the number of times a resume with a White-sounding name had to be sent out to receive one callback in BM. This question is also eligible, and if selected, participants receive the bonus if they are within one

[1]The participation payment was increased from $3 to $3.75 for the final third of data collection due to grant requirements. Within each participation payment amount, the sample is balanced by Democrats and Republicans.

unit of the correct answer. In Round 3, participants are asked their best guess of average weekly earnings for White full-time workers in the US is eligible. This question is also eligible, and if selected, participants receive the bonus if they are within \$100 of the correct answer. In Round 4, participants are asked their best guess of the percent of the Black-White wage gap that is explained by differences in educational attainment. This question is also eligible, and if selected, participants receive the bonus if they are within five percentage points of the correct answer.

The only questions in Rounds 1-5 that are not eligible to be selected for the bonus are about how much of the Black-White wage gap participants believe are driven by employer discrimination against Black workers in Rounds 3, 4, and 5. These questions cannot be incentivized because we do not currently have methods to calculate this number. Each time this question is asked, participants are told that this question is hypothetical and not eligible for a bonus. Analysis of these unincentivized questions is presented in Appendix B.

1.4 Hypotheses

In this section, I outline the main hypotheses on beliefs about racial hiring discrimination for each round.

1.4.1 Round 1

In Round 1, I ask participants their best guesses of the callback rates from BM: the number of times a resume with a Black-sounding name had to be sent out to get one callback and the number of times a resume with a White-sounding name had to be sent out to get one callback. From their responses, I calculate each participant's baseline belief about racial discrimination in hiring as follows.

$$D_{1,i} = log(\widehat{B}_i) - log(\widehat{W}_{1,i}) \tag{1.1}$$

where $\widehat{B}_i$ is participant i's prediction of the number of times resumes with Black-sounding names had to be sent out to receive one callback in BM, and $\widehat{W}_{1,i}$ is participant i's Round 1 prediction of the number of times resumes with White-sounding names had to be sent out to receive one callback in BM.

In this round, I evaluate whether Democrats and Republicans disagree about the extent of discrimination in hiring against Black workers. Findings from the literature (Haaland and Roth, 2021; Alesina et al., 2021) suggest that Democrats believe there is more racial discrimination in hiring than Republicans do. I seek to replicate this finding in Round 1 of my study. I test directly whether Democrats' and Republicans' mean beliefs about racial discrimination in hiring statistically differ. I also test belief differences between Democrats and Republicans across the distribution of responses nonparametrically using a Kolmogorov-Smirnov test.

1.4.2 Round 2

In Round 2, I tell participants the number of times that resumes with Black-sounding names in BM had to be sent out to get one callback for an interview. Given that participants know the callback rate for resumes with Black-sounding names from Round 2 onward, I adjust the calculation of their beliefs about racial discrimination in hiring as follows for Rounds 2-5.

$$D_{j,i} = log(\overline{B}) - log(\widehat{W}_{j,i}) \tag{1.2}$$

where $D_{j,i}$ is participant i's calculated belief about hiring discrimination in Round $j \in \{2,5\}$, $\overline{B}$ is the actual number of times resumes with Black-sounding names had to be sent out to receive one callback from BM (15), and $\widehat{W}_{j,i}$ is participant i's Round $j \in \{2,5\}$ prediction of the number of times resumes with White-sounding names had to be sent out to receive one callback from BM.

In Round 2, I test how participants update their beliefs about racial discrimination in hiring in response to the callback rate for Black-sounding names in BM. If participants interpret this information purely as benchmarking information (i.e., to get a sense of average callback rates in BM), they may not update their beliefs about racial discrimination. That is, they may update their predicted callback rate for applicants with White-sounding names such that their prediction of racial discrimination in hiring is unchanged. Participants may, on the other hand, use the callback rate for Black applicants in BM as a signal of racial discrimination in hiring. Suppose, for example, a participant finds out that Black applicants in BM received fewer callbacks than they anticipated. The participant may interpret this low callback rate as a signal that there is more racial discrimination in hiring than they thought.

To calculate participants' changes in beliefs about racial discrimination in hiring between Round 1 and 2, I calculate the following.

$$\Delta D_{1,2,i} = ihs(D_{2,i}) - ihs(D_{1,i}) \tag{1.3}$$

where $ihs()$ is the inverse hyperbolic sine function. This function approximates the log function, while allowing for zeroes. Therefore, if participants believe there is no racial discrimination in hiring, this function allows me to calculate their belief updates, unlike the log function which would exclude their responses.

I calculate participants' errors on the number of times that resumes with Black-sounding names had to be sent out to receive one callback for an interview as follows:

$$Error\_B_i = log(\overline{B}) - log(\widehat{B}_i) \tag{1.4}$$

where $\overline{B}$ is the true average number of times resumes with Black-sounding names had to be sent out to receive a callback in BM.

Then, I test how participants update their beliefs between Round 1 and 2 in response to the Black callback rate in BM as follows.

$$\Delta D_{1,2,i} = \mu + \delta_1 Rep_i Error\_B_i + \delta_2 Dem_i Error\_B_i + \eta_i \tag{1.5}$$

where $Rep_i = 1$ if participant i is a Republican and 0 otherwise, and $Dem_i = 1$ if participant i is a Democrat and 0 otherwise.

From δ_1 and δ_2, I identify if Republicans and Democrats, respectively, update their beliefs about racial discrimination in hiring in response to the Black callback rate in BM. If $\delta_1 = 0$ ($\delta_2 = 0$), then this would suggest that Republicans (Democrats) do not update their beliefs about racial discrimination and treat the Black callback rate as purely benchmarking information.

1.4.3 Round 3

In Round 3, participants are told (after stating their priors) that White full-time workers in the US earn on average \$1085 per week, while Black full-time US workers earn on average \$844 per week. Participants then state their updated best guess of the number of times that resumes with White-sounding names had to be sent out to receive one callback in BM.

In Round 3, I test whether participants update their beliefs about racial discrimination in hiring in response to information about the Black-White wage gap. Figure 1.3 is a directed acyclic graph (DAG) summarizing how participants may think about the relationship between racial discrimination in hiring and the Black-White wage gap. Arrows indicate the direction of causality.

There are many possible drivers of the Black-White wage gap. I highlight two in Figure 1.3: educational attainment differences between Black and White workers in the US and racial discrimination in hiring. All other drivers of the wage gap that participants may think of are encapsulated in the category “All other drivers.” In Round 3, I test directly whether arrow C holds. That is, I test whether participants believe that racial discrimination in hiring is a driver of the Black-White wage gap.

If participants do not believe that racial discrimination in hiring is a driver of the Black-White wage gap, then we would expect the update in their belief about racial discrimination in hiring between Round 2 and 3 to be uncorrelated with their error on the Black-White wage gap. If, on the other hand, participants believe that racial discrimination in hiring is a driver of the Black-White wage gap, then we may expect participants to update their beliefs about

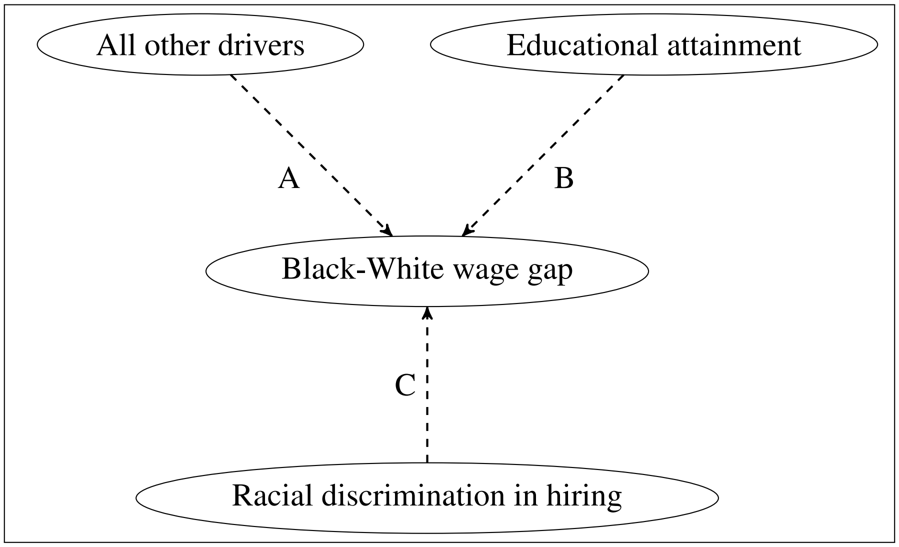

Figure 1.3. Potential drivers of the Black-White wage gap

Notes. The figure above is a directed acyclic graph (DAG) summarizing the relationship between the Black-White wage gap for full-time workers in the US and various potential drivers. Arrows indicate the direction of causality. Dashed arrows represent pathways that participants may or may not believe exist. From participants' responses in my study, I can test whether participants believe arrows B and C are prominent contributors to the Black-White wage gap.

racial discrimination in hiring in response to information on the Black-White wage gap. That is, participants who overestimate the Black-White wage gap may decrease their belief about the extent of racial discrimination in hiring, and those who underestimate the Black-White wage gap may increase their belief about the extent racial discrimination in hiring.

To test whether participants update in response to the Black-White wage gap, I first calculate their belief update about racial discrimination in hiring as follows.

$$\Delta D_{2,3,i} = ihs(D_{3,i}) - ihs(D_{2,i}) \tag{1.6}$$

where $ihs()$ is the inverse hyperbolic sine function, which approximates the natural log function while allowing for zeroes.

I then calculate participants' errors on White average weekly earnings is calculated as follows.

$$Error\_WE_i = log(\overline{WE}) - log(\widehat{WE_i}) \tag{1.7}$$

where $\overline{WE}$ is the true average White weekly earnings (\$1085), and $\widehat{WE}_i$ is participant i's prediction of White weekly average earnings. Because participants are first told Black average weekly earnings, this measure indicates their beliefs about the Black-White wage gap.

I then test whether participants' belief updates are correlated with their error on average weekly earnings for White full-time workers. Findings from Alesina et al. (2021) suggest that Democrats are more likely than Republicans to believe racial disparities are driven more by systemic factors, including racial discrimination. So, I test the relationship separately for Democrats and Republicans, as follows.

$$\Delta D_{2,3,i} = \alpha + \beta_1 Rep_i Error\_WE_i + \beta_2 Dem_i Error\_WE_i + \varepsilon_i \tag{1.8}$$

where $Rep_i = 1$ if participant i is a Republican and 0 otherwise, and $Dem_i = 1$ if participant i is a Democrat and 0 otherwise.

β_1 and β_2 inform whether Republicans and Democrats, respectively, update their beliefs about racial discrimination in hiring in response to the Black-White wage gap.

1.4.4 Round 4

In Round 4, participants are told (after stating their priors) that 12% of the Black-White wage gap for full-time workers in the US is explained by differences in educational attainment between Black and White workers. They then state their updated best guess of the BM White callback rate.

In thinking about how information on the explanatory power of educational attainment may affect participants' beliefs about racial discrimination in hiring, we turn again to Figure 1.3. Because the categories are all-encompassing, all three arrows (A, B, and C) must together explain 100% of the Black-White wage gap. Consider a participant who overestimates the percent of the wage gap that is explained by educational attainment. This participant would then have leftover weight that must be spread between arrows A and C. If the participant believes that racial discrimination in hiring is a driver of the Black-White wage gap, then they may assign some of the weight to arrow C. If the participant does not believe that racial discrimination is a driver, then we would not expect them to add any weight to arrow C.

I calculate participants' Round 4 updates in beliefs about racial discrimination in hiring as follows:

$$\Delta D_{3,4,i} = ihs(D_{4,i}) - ihs(D_{3,i}) \tag{1.9}$$

where $ihs()$ is the inverse hyperbolic sine function, which approximates the log function while allowing for zeroes.

I calculate participants' errors on how much of the wage gap is explained by educational attainment as follows.

$$Error\_EA_i = ihs(\overline{EA}) - ihs(\widehat{EA}_i) \tag{1.10}$$

where $\overline{EA}$ is the calculation of the amount of the wage gap explained by differences in educational attainment (12%), and $\widehat{EA}_i$ is participant i's prediction of this percent.

To investigate whether participants update their beliefs about racial hiring discrimination in response to information about the explanatory power of educational attainment, I run the following regression.

$$\Delta D_{3,4,i} = \phi + \gamma_1 Rep_i Error\_EA_i + \gamma_2 Dem_i Error\_EA_i + \upsilon_i \tag{1.11}$$

where $Rep_i = 1$ if participant i is a Republican and 0 otherwise, and $Dem_i = 1$ if participant i is a Democrat and 0 otherwise.

γ_1 and γ_2 inform whether Republicans and Democrats, respectively, update their beliefs about racial discrimination in hiring in response to the explanatory power of educational attainment. If $\gamma_1 < 0$ and $\gamma_2 < 0$, then this would suggest that Democrats and Republicans, respectively, increase the weight on arrow C in Figure 1.3 upon finding out that educational attainment explains less of the wage gap than they thought.

1.4.5 Round 5

In Round 5, participants are told the callback rates for applicants with White-sounding names and for applicants with Black-sounding names from the fake resume study in Jacquemet and Yannelis (2012). Participants then state their final best guess of the callback rate for applicants with White-sounding names in BM.

The purpose of Round 5 is to show that the belief gap between Democrats and Republicans in my sample can indeed be closed using information. A closing of the belief gap between Democrats and Republicans would replicate a finding from Haaland and Roth (2021) that information on results from experiments designed to measure discrimination can successfully close the partisan belief gap about hiring discrimination.

To examine if the belief gap persists in Round 5, I test directly whether Democrats' and Republicans' mean Round 5 beliefs about racial discrimination in hiring statistically differ. I also test belief differences between Democrats and Republicans across the distribution of responses nonparametrically using a Kolmogorov-Smirnov test.

I also test if Democrats and Republicans significantly update their beliefs about racial discrimination in hiring between Round 4 and Round 5. To do so, I calculate participants' changes in beliefs between Round 4 and 5 as follows.

$$\Delta D_{4,5,i} = ihs(D_{5,i}) - ihs(D_{4,i}) \tag{1.12}$$

To test if Democrats and Republicans significantly update their beliefs, I regress their belief update on their political affiliation.

$$\Delta D_{4,5,i} = \psi + \chi Rep_i + \rho_i \tag{1.13}$$

where Rep_i is a dummy variable indicating if participant i is Republican.

The constant term ψ indicates if Democrats update significantly in Round 5, and $\psi + \chi$

indicates if Republicans update significantly.

1.5 Results

In this section, I review study results. Subsection 1.5.1 describes the sample in terms of demographic characteristics and baseline interpretations of BM. In subsection 1.5.2, I describe participants' baseline beliefs about racial hiring discrimination. In subsection 1.5.3, I investigate belief-updating about hiring discrimination in response to the Black callback rate in BM. Subsection 1.5.4 investigates belief-updating about hiring discrimination in response to information on the Black-White wage gap. In subsection 1.5.5, I investigate belief-updating about hiring discrimination in response to information on the role of educational attainment in explaining the Black-White wage gap. In subsection 1.5.6, I examine how participants update their beliefs in response to results from another fake resume study. Subsection 1.5.7 compares participants' endline interpretations of BM to their baseline interpretations, and compares responses by political affiliation.

1.5.1 Sample

The study was administered on Prolific, a widely used online survey platform among social scientists, from April - August 2023. My sample consists of 1100 self-reported Democrats and 1100 self-reported Republicans in the US with accounts on Prolific. I ensured the sample is split evenly by gender, with 50% female and 50% male participants. Respondents are on average 40 years old, with Democrats being slightly younger (37) than Republicans (43). Overall, the sample is more White than the general US population, with 77% of my sample identifying as White and only 7% identifying as Black. Democrats skew less White and more Black than Republicans. See Table 1.1 for more demographic details. On average, participants took approximately 13 minutes to complete the study, and 21% of participants earned the $2 bonus.

One concern with using an online survey platform for my study is that participants lean more liberal than the general US population. Indeed, the total available sample of self-identified

Democrats on Prolific is approximately 12,000 people, compared to only approximately 3,000 Republicans. While both groups are large enough for my sample size, one may be concerned that Republicans on this platform are more ideologically moderate than those of the general US population. To check, I ask participants two questions at the end of the study. First is their self-reported ideology on a standard seven-point scale from "Extremely Liberal" to "Extremely Conservative". Second is the probability they will vote for the Republican or Democratic candidate in the 2024 presidential election, conditional on voting. Figure 1.4 reports participants' responses to these questions.

Participants' responses to both questions indicate that Republicans are indeed more moderate than Democrats. Approximately 58% of Democrats state there is a 100% probability that they will vote for the Democratic nominee in the 2024 presidential election, while approximately 45% of Republicans in my sample state there is a 100% probability that they will vote for the Republican nominee. A substantial portion of Republicans do self-report being "Extremely Conservative" (16%), but this proportion is significantly smaller than Democrats who self-report being "Extremely Liberal" (33%). Given that my primary analysis is in evaluating differences between Democrats and Republicans, having a more moderate sample of Republicans than the general US population biases me away from finding differences between Democrats and Republicans.

Table 1.1. Sample Demographics by Political Affiliation

	Democrats	Republicans
Male	0.50	0.51
	(0.50)	(0.50)
White	0.70	0.84
	(0.46)	(0.36)
Black	0.10	0.04
	(0.30)	(0.19)
Asian	0.09	0.04
	(0.29)	(0.21)
Other Race	0.10	0.07
	(0.30)	(0.26)
Born in US	0.94	0.95
	(0.24)	(0.22)
Employed Full-time	0.41	0.53
	(0.49)	(0.50)
Employed Part-time	0.13	0.13
	(0.34)	(0.34)
Unemployed	0.11	0.07
	(0.31)	(0.26)
Observations	1100	1100

Before beginning Round 1, I explain BM to participants. I then ask participants if they believe that a difference in callback rates between applicants with Black-sounding names and applicants White-sounding names reflects that employers base their callback decisions in part on the race of the applicant. If participants do not agree with this interpretation, then it would not be appropriate for me to interpret their beliefs about the study findings as their beliefs about racial discrimination in hiring. Figure 1.16 reports the distributions of participants' agreement with this statement, on a scale from "Strongly Agree" to "Strongly Disagree."

In my sample, 94% of Democrats and 79% of Republicans either "somewhat agree" or "strongly agree" that any difference in callback rates between applicants with Black-sounding and White-sounding names would indicate that employers base their callback decisions in part on the race of the applicant. While Democrats are more likely to "strongly agree" than Republicans (54% of Democrats vs. 32% of Republicans), I find it promising that the majority of my sample from both parties generally agree that the results are driven by employers using applicants' races in making their callback decisions.

1.5.2 Round 1: Baseline beliefs about racial discrimination in hiring

In the first round, participants state their baseline beliefs about racial discrimination in hiring. That is, they state the number of times they think a resume with a White-sounding name had to be sent out to get one callback for an interview in BM, and the number of times they think a resume with a Black-sounding name had to be sent out to get one callback for an interview in BM.

Figure 1.6 shows participants' baseline beliefs for White-sounding names, split by political affiliation. Responses greater than 60 are excluded (1.3% of responses) from the graph for visual purposes. The median response among both Democrats and Republicans is that resumes with White-sounding names had to be sent out an average of 3 times to get one callback for an interview. Using a Kolmogorov-Smirnov test for equality of distributions, I cannot reject the null hypothesis that Democrats' responses and Republicans' responses come from the same underlying distribution ($p = 0.46$). Both groups underestimate the number of times resumes with White-sounding names had to be sent out, as BM finds they had to be sent out 10 times to get a callback.

Figure 1.7 shows participants' baseline beliefs on the callback rate for applicants with Black-sounding names in BM, split by political affiliation. Responses greater than 60 are excluded (2% of responses) from the graph for visual purposes. Democrats' median prediction is that resumes with Black-sounding names had to be sent out 8 times to get one callback.

Republicans' median response is 6 times. Average callback beliefs about the Black callback rate are significantly different between Democrats and Republicans ($p = 0.01$), and a Kolmogorov-Smirnov test rejects the null hypothesis that these responses come from the same underlying distribution ($p < 0.001$). Both groups underestimate the number of times resumes with Black-sounding names had to be sent out, as BM finds they had to be sent out 15 times to get one callback.

To calculate participants' baseline beliefs about racial discrimination in hiring, I take the log difference of their predictions of callback rates, as in Equation 1.1. Figure 1.8 reports the cumulative distribution function of participants' baseline beliefs about racial discrimination in hiring in Round 1 split by political affiliation. Across the distribution, Democrats believe there is more racial discrimination in hiring than Republicans do. Distributions are significantly different by political affiliation ($p < 0.001$), according to the Kolmogorov-Smirnov test. Relative to BM results, both Democrats and Republicans overestimate the amount of discrimination.

Result 1: *At baseline, Democrats believe there is more racial discrimination in hiring than Republicans do.*

1.5.3 Round 2: Hiring discrimination beliefs in response to BM Black callback rate

In Round 2, participants are told the number of times that resumes with Black-sounding names had to be sent out to get one callback for an interview in BM. They then state their updated beliefs about the number of times that resumes with White-sounding names had to be sent out for a callback in BM. From Round 2 onward, I calculate participants' beliefs about racial hiring discrimination as in Equation 1.2. That is, I take the log difference between the actual callback rate for applicants with Black-sounding names in BM, and participants' predictions of the callback rate for applicants with White-sounding names.

Figure 1.9 shows participants' beliefs about racial discrimination in hiring in Round 2 split by political affiliation. The dashed lines depict participants' Round 1 beliefs, and the solid lines depict participants' Round 2 beliefs.

Relative to Round 1, both Democrats and Republicans increase their hiring discrimination beliefs in response to the Black callback rate from BM. The gap between Democrats' and Republicans' beliefs about racial discrimination in hiring decreases ($p < 0.001$), as Republicans update more positively than Democrats. This is in line with Bayesian updating, as Republicans overestimated the frequency of callbacks for Black applicants more than Democrats did. I reject that Democrats' and Republicans' beliefs come from the same underlying distribution ($p < 0.001$), according to the Kolmogorov-Smirnov test.

In Table 1.2, I directly test the relationship between participants' errors on the BM Black callback rate and their updating behavior. I calculate participants' belief changes about hiring discrimination between Round 1 and Round 2 as in Equation 1.3, and their error on the BM Black callback rate as in Equation 1.4.

Table 1.2. Hiring Discrimination Belief Updates to BM Black Callback Rate

	Belief Update R2	Belief Update: R2	Belief Update: R2
Error: Black CB	0.214***		
	(0.0132)		
Over × Error: Black CB		0.0769**	
		(0.0374)	
Under × Error: Black CB		0.286***	
		(0.0181)	
Dem × Over × Error: Black CB			0.0768
			(0.0504)
Dem × Under × Error: Black CB			0.256***
			(0.0217)
Rep × Over × Error: Black CB			0.0862*
			(0.0462)
Rep × Under × Error: Black CB			0.299***
			(0.0197)
Constant	0.0469***	-0.0248*	-0.0200
	(0.0109)	(0.0150)	(0.0151)
Observations	2190	2190	2190

Notes. This regression shows participants' Round 2 belief updates about racial hiring discrimination. Belief updates are the inverse hyperbolic sine (IHS) function difference between Round 2 and Round 1 beliefs. "Error: Black CB" is the IHS difference between priors on the Black callback rate and the actual Black callback rate in BM. "Under" ("Over") restricts to participants who underestimate (do not underestimate) the number of times Black resumes had to be sent out to get a callback in BM. "Dem" ("Rep") restricts to participants who identify as Democrats (Republicans). Robust standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Model 1 of Table 1.2 regresses participants' changes in discrimination beliefs between rounds 2 and 1 on their errors on the Black callback rate, as in Equation 1.5. Coefficients can be thought of as elasticities, where a coefficient of 1 would indicate that participants update entirely on racial hiring discrimination, and a coefficient of 0 would indicate that participants update entirely on the White callback rate. The coefficient of 0.214 means that a 100% error in the callback rate for applicants with Black-sounding names translates to a 21.4% increase in beliefs about hiring discrimination.

Model 2 of Table 1.2 interacts participants' errors on the Black callback rate with an indicator for whether participants overestimate or underestimate the Black callback rate. This shows that the relationship between participants' errors and their updating behavior is driven by those who underestimate the frequency of callbacks for resumes with Black-sounding names. That is, participants who find out Black resumes had to be sent out *more* times than they thought update on hiring discrimination, while those who find out Black resumes had to be sent out *fewer* times than they thought do not update on hiring discrimination.

Model 3 of Table 1.2 interacts each of the terms from model 2 with political affiliation. Both groups update positively on discrimination when they find out Black resumes had to be sent out more times to get a callback than they had initially predicted. Among those who overestimate the number of times that Black resumes had to be sent out to get a callback, I find no evidence that Democrats and Republicans update significantly differently from each other ($p = 0.88$). Among those who underestimate the number of times that Black resumes had to be sent out to get a callback, Republicans update their beliefs about racial discrimination in hiring more than Democrats ($p = 0.03$).

1.5.4 Round 3: Hiring discrimination beliefs in response to Black-White wage gap

In Round 3, participants are told average weekly earnings for Black full-time workers in the US (\$844)[2], and asked for their best guess of average weekly earnings for White full-time workers in the US. Participants are then told median weekly earnings for White full-time workers in the US (\$1085). Then, participants again state their best guess of the number of times a resume with a White-sounding name had to be sent out on average to get one callback for an interview in BM.

Figure 1.10 shows Democrats' (left panel) and Republicans' (right panel) distributions of priors on average weekly earnings for White full-time workers in the US. Average weekly earnings among White full-time workers in the US was \$1085 according to the 2021 Current Population Survey. The median prediction among Democrats was \$1142, which was larger than Republicans' median estimate of \$1066 ($p < 0.001$), implying that Democrats think the wage gap is larger than Republicans do. Overall, 46% of Democrats underestimate the wage gap, compared to 59% of Republicans.

On average, Democrats' beliefs about racial discrimination in hiring decrease slightly, but not statistically significantly between rounds 2 and 3 ($p = 0.237$). Republicans' beliefs do not significantly change ($p = 0.833$). As seen in Figure 1.11, Democrats experience a slight distributional shift in beliefs in Round 3 (Kolmogorov-Smirnov test: $p = 0.08$), but Republicans do not (Kolmogorov-Smirnov test: $p = 0.70$). The average gap between Democrats' and Republicans' beliefs decreases slightly but not significantly ($p = 0.11$).

To evaluate how participants respond to wage gap information in this round, I regress their changes in beliefs about hiring discrimination on their error in White average earnings in Table 1.3 as shown in Equation 1.8). The outcome variable, the change in beliefs about hiring discrimination, is calculated as in Equation 1.6. Participants' errors on the average weekly

[2] This statistic was gathered from the Current Population Survey 2021 median earnings for Black full-time workers in the US.

earnings for White full-time workers in the US are calculated as in Equation 1.7. I exclude extreme outliers from participants who submit best guesses of the average weekly White earnings that are an order of magnitude off from the correct answer: observations less than or equal to \$100 or greater than or equal to \$10,000. This includes 21 participants in total (1% of my sample).

As can be seen in model 1 of Table 1.3, there does not seem to be strong relationship between wage gap information and beliefs about hiring discrimination. When we split the sample by Democrats and Republicans in model 2, however, Democrats update their beliefs about racial discrimination in hiring in response to the Black-White wage gap, while Republicans do not.

Result 2: *Democrats update their beliefs about racial discrimination in hiring in response to information about the Black-White wage gap, while Republicans do not.*

Splitting the sample by those who overestimated and underestimated White average earnings in model 3, I find that Democrats who overestimate White earnings seem to be driving the effect, not Democrats who underestimate White earnings. Republicans, on the other hand, do not seem to use the Black-White wage gap to update their beliefs about racial discrimination in hiring regardless of whether they overestimate or underestimate White earnings.

Table 1.3. Hiring Discrimination Belief Updates to Black-White Wage Gap

	Belief Update: R3	Belief Update: R3	Belief Update: R3
Error: White Earn	0.0440		
	(0.0413)		
Dem $\times$ Error: White Earn		0.0895**	
		(0.0434)	
Rep $\times$ Error: White Earn		-0.00274	
		(0.0710)	
Dem $\times$ Over $\times$ Error: White Earn			0.231***
			(0.0452)
Dem $\times$ Under $\times$ Error: White Earn			-0.108
			(0.0864)
Rep $\times$ Over $\times$ Error: White Earn			0.0894
			(0.110)
Rep $\times$ Under $\times$ Error: White Earn			-0.0615
			(0.101)
Constant	-0.0114**	-0.00986*	0.00616
	(0.00556)	(0.00576)	(0.00817)
Observations	2171	2171	2171

Notes. This regression shows participants' Round 3 belief updates about racial hiring discrimination. Belief updates are the inverse hyperbolic sine (IHS) function difference between their Round 3 and Round 2 beliefs. "Error: White Earn" is the log difference between participants' priors and a BLS estimate of average White weekly earnings for full-time US workers. "Dem" ("Rep") includes only participants who identify as Democrats (Republicans). "Under" ("Over") includes only participants who underestimate (do not underestimate) median White weekly earnings. Robust standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

1.5.5 Round 4: Hiring discrimination beliefs in response to role of educational attainment on wage gap

In Round 4, participants state how much (in %) of the Black-White wage gap they think is driven by differences in educational attainment between Black and White full-time workers. Participants are told that statisticians have developed methods of calculating this number, and if this question is selected for a bonus, they earn the $2 bonus if they guess within 5 percentage points of the correct answer (12%).

Participants are then told that statisticians estimate 12% of the Black-White wage gap is explained by differences in educational attainment[3]. The vast majority of the sample (89%) overestimates how much of the wage gap is explained by differences in educational attainment, as shown in Figure 1.12. 91% of Republicans and 86% of Democrats overestimate the role of educational attainment in explaining the wage gap do ($p < 0.001$).

Participants are then asked again their best guess of the callback rate for applicants with White-sounding names in BM. Between Rounds 3 and 4, Democrats' beliefs about racial discrimination in hiring increase ($p = 0.020$), while Republicans' beliefs do not ($p = 0.250$). Democrats update, on average, more in Round 4 than Republicans do, leading to a slight increase in the belief gap about racial discrimination in hiring relative to Round 3 ($p = 0.069$), as shown in Figure 1.15.

In Figure 1.13, I compare belief distributions about racial discrimination in hiring between Rounds 3 and 4. While the distribution for Democrats appear to shift rightward, I cannot reject that Democrats' Round 3 and 4 beliefs about racial discrimination come from the same distributions ($p = 0.16$). I also cannot reject that the distribution of Republicans' beliefs between Round 3 and 4 come from the same distribution ($p = 0.44$).

Table 1.4 reports regression analysis of participants' changes in hiring discrimination beliefs from Rounds 3 to 4 on their error in the explanatory power of educational attainment

[3]This statistic was calculated using the Oaxaca-Blinder decomposition using 2021 earnings and education data from the Census Bureau for full-time workers in the US.

on the Black-White wage gap (Equation 1.11). I calculate participants' updates to their beliefs about racial discrimination in hiring as in Equation 1.9. I calculate participants' errors on how much of the wage gap is explained by educational attainment as in Equation 1.10.

Table 1.4. Hiring Discrimination Belief Updates to Role of Edu on Black-White Wage Gap

	Belief Update: R4	Belief Update: R4	Belief Update: R4
Error: Pct EA	-0.00939		
	(0.00773)		
Under × Error: Pct EA		0.0244	
		(0.0214)	
Over × Error: Pct EA		-0.0208**	
		(0.0100)	
Dem × Under × Error: Pct EA			0.0346
			(0.0250)
Dem × Over × Error: Pct EA			-0.0319***
			(0.0107)
Rep × Under × Error: Pct EA			0.0166
			(0.0339)
Rep × Over × Error: Pct EA			-0.0156
			(0.0113)
Constant	0.0274***	0.0150	0.0128
	(0.00797)	(0.0101)	(0.0101)
Observations	2193	2193	2193

Notes. This regression shows participants' Round 4 belief updates about racial hiring discrimination. Belief updates are the inverse hyperbolic sine (IHS) function difference between their Round 4 and Round 3 beliefs. "Error: Pct EA" is the IHS difference between participants' priors and an estimate of how much educational attainment explains of the Black-White wage gap. "Over" ("Under") includes all participants who overestimate (do not overestimate) the role of educational attainment. "Dem" ("Rep") includes only participants who identify as Democrats (Republicans). Robust standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Model 1 shows an insignificant relationship between participants' beliefs about discrimination and the role of educational attainment on the wage gap. Splitting the sample in model 2 by those who underestimate vs. overestimate the role of educational attainment, however, shows that participants who overestimate the role of educational attainment increase their beliefs about racial discrimination in hiring. Participants who overestimate by 100%, increase their discrimination beliefs by 2%, on average.

In model 3, I split updating behavior by political affiliation. The relationship between hiring discrimination beliefs and overestimating the role of educational attainment holds for Democrats in this regression, but not for Republicans. That is, Democrats tend to increase their beliefs about how much discrimination drives the wage gap after learning that educational attainment explains less than they thought, while Republicans do not.

Result 3: *In response to information on the role of educational attainment on the Black-White wage gap, Democrats update their beliefs about racial discrimination in hiring while Republicans do not.*

1.5.6 Round 5: Hiring discrimination beliefs in response to results from another fake resume study

In Round 5, participants are shown callback rates from another fake resume study, Jacquemet and Yannelis (2012). That is, they are told Jacquemet and Yannelis (2012) found that applicants with White-sounding names had to send out on average 4 applications to receive one callback for an interview, and applicants with Black-sounding names had to send out on average 6 applications to receive one callback for an interview.

The purpose of this final round is to replicate a finding from Haaland and Roth (2021) that results from experiments designed to measure discrimination can successfully close the belief gap between Democrats and Republicans. In Figure 1.14, the distribution of Democrats' beliefs and the distribution of Republicans' beliefs shift to the left in Round 5 ($p < 0.001$ for both groups).

As shown in Figure 1.15, Round 5 is the only round in which the average gap in beliefs about racial discrimination in hiring between Democrats and Republicans is closed ($p = 0.13$). Furthermore, I cannot reject that Democrats' and Republicans' beliefs come from the same distribution ($p = 0.14$).

Table 1.5 reports these findings in a regression, with the outcome variable calculated as in Equation 1.12. Participants did not state their priors for this outcome, so I regress belief updates about hiring discrimination on political affiliation and a constant, as shown in Equation 1.13. Democrats' beliefs about racial discrimination in hiring drop by approximately 19%, while Republicans' drop by approximately 9%. While the gap does close between Democrats and Republicans, beliefs do not converge on the results from BM (shown by the horizontal gray line in Figure 1.15), even though both studies found the same ratio of callback rates between the two groups.

Result 4: *Results from another fake resume study successfully close the gap in beliefs about racial discrimination in hiring between Democrats and Republicans.*

1.5.7 Interpretations of BM results

In this subsection I present results on participants' interpretations of BM. The questions analyzed in subsections 1.5.7 and 1.5.7 were the only questions in the study that were asked at both baseline and endline. These analyses were not incentivized nor pre-registered, and may be considered exploratory.

Do BM results reflect that employers use race in callback decisions?

After introducing BM and before rounds begin (baseline), I ask participants how much they agree on a 5-point scale from "Strongly disagree" to "Strongly agree" with the following statement. "If the researchers find a difference in callback rates between applicants with Black-

Table 1.5. Hiring Discrimination Belief Updates to JY Resume Study Results

	Belief Update: R5-R4
Republican	0.107***
	(0.0168)
Constant	-0.194***
	(0.0107)
Observations	2196

Notes. The table shows regressions of participants' changes in beliefs (between rounds 4 and 5) about racial hiring discrimination in response to results from another fake resume study. Beliefs about racial hiring discrimination are calculated as the log difference between participants' predicted White callback rates and the actual Black callback rate in BM. Changes in beliefs between Round 3 and 4 are calculated using the inverse hyperbolic sine function difference. "Republican" includes only participants who list their political affiliation as "Republican" on their Prolific account. Robust standard errors are in parentheses.
$p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

sounding names and applicants with White-sounding names in Experiment A[4], this would reflect that employers base their callback decisions in part on the race of the applicant."

After all rounds are completed and participants have learned the results from BM (endline), I ask participants how much they agree (on the same scale) with the following statement. "The difference in callback rates between applicants with Black-sounding names and applicants with White-sounding names in Experiment A reflects that employers base their callback decisions in part on the race of the applicant."

Figure 1.16 shows participants' responses at baseline (dashed lines) and endline (solid lines), split by political affiliation. At baseline, Democrats on average score 4.5 on the scale from 1 ("Strongly disagree") to 5 ("Strongly agree"). At baseline, Republicans on average score lower than Democrats ($p < 0.001$) at 4.0. I reject the null hypothesis that responses come from the same underlying distribution, according to the Kolmogorov-Smirnov test ($p < 0.001$).

At baseline, 54% of Democrats and 32% of Republicans ($p < 0.001$) selected the maxi-

[4]BM was referred to as "Experiment A" to participants throughout the study.

mum response ("Strongly agree") and therefore cannot increase their response further at endline. Mechanically, therefore, Republicans have more room to increase their beliefs than Democrats. Only 0.4% of Democrats and 2.5% of Republicans ($p < 0.001$). selected the minimum response ("Strongly disagree") at baseline, so there is less of a concern of participants being unable to decrease their agreement at endline relative to baseline.

At endline, both Democrats' ($p < 0.001$) and Republicans' ($p = 0.002$) agreement levels shift upwards on average relative to their baseline responses. I also calculate the probability within political party that participants change their level of agreement between baseline and endline. As shown in the left panel of Figure 1.17, Democrats (21%) and Republicans (23%) are equally likely to increase their level of agreement relative to their response at baseline ($p = 0.44$). Republicans (17%) are more likely than Democrats (11%) to decrease their agreement at endline relative to their baseline agreement ($p < 0.001$).

Are BM results a problem?

After introducing BM and before rounds begin (baseline), I ask participants how much they agree on a 5-point scale from "Strongly disagree" to "Strongly agree" with the following statement. "If the researchers find that applicants with White-sounding names get callbacks more often than those with Black-sounding names, this would be a problem that should be solved."

After Round 5 is completed and participants have learned the BM results (endline), I ask participants how much they agree on the same scale with the following statement. "The difference in callback rates is a problem that should be solved."

Figure 1.18 shows participants' responses at baseline (dashed lines) and endline (solid lines), split by political affiliation. At baseline, Democrats' average response at 4.6 is greater than Republican' average score of 3.9 ($p < 0.001$). I reject the null hypothesis that these responses come from the same distribution, according to the Kolmogorov-Smirnov test ($p < 0.001$).

Note that 68% of Democrats and 36% of Republicans ($p < 0.001$) selected the maximum response ("Strongly agree") and therefore cannot increase their agreement any further at endline

Mechanically, therefore, Republicans have more room to increase their beliefs than Democrats at endline. Only 0.7% of Democrats and 5.4% of Republicans ($p < 0.001$). selected the minimum response ("Strongly disagree") at baseline, so there is less of a concern of participants being unable to decrease their agreement at endline.

At endline, both Democrats' ($p < 0.001$) and Republicans' ($p = 0.001$) agreement levels shift upwards on average relative to their baseline responses. As shown in the right panel of Figure 1.17, 21% of Republicans and 17% of Democrats increase their agreement at endline, relative to their response at baseline ($p = 0.02$). Given that a strong majority of Democrats could not increase their level of agreement relative to baseline, this difference may be mechanical. That is, Democrats who may have otherwise increased their level of agreement were not able to.

Republicans (13%) are also more likely than Democrats (6%) to decrease their agreement level at endline relative to their agreement at baseline ($p < 0.001$). This is less likely to be mechanical, as the vast majority of both groups had the ability to decrease their agreement level relative to their baseline response.

Result 5: *At endline, Republicans are more likely than Democrats to decrease their agreement (relative to baseline) that (1) BM results reflect that employers base callback decisions in part on applicant race and (2) a higher White callback rate in BM is a problem.*

1.6 Discussion

Across five rounds, I evaluate how participants' beliefs about racial discrimination in hiring respond to various pieces of information. In this section, I synthesize what we learn from the results in each round.

I find that Democrats and Republicans disagree about how informative some signals are about racial discrimination in hiring. Democrats update their hiring discrimination beliefs in response to the Black-White wage gap, while Republicans do not. Similarly, when presented with the role of educational attainment on the wage gap, Democrats update their beliefs about hiring discrimination, while Republicans do not.

One potential explanation for Republicans updating less than Democrats in response to both pieces of wage gap information is that Republicans have stronger priors, and thus less movable beliefs about racial discrimination in hiring than Democrats do. However, Republicans update their beliefs about racial discrimination in hiring *more* than Democrats in Round 2 after presented with the BM callback rate for applicants with Black-sounding names. This suggests that the observed differences in updating patterns between Democrats and Republicans to wage gap information are not due to Republicans having less movable beliefs, and instead may be due to differences in beliefs about the relationship between wage gaps and discrimination.

The findings demonstrate that when groups disagree about the relevance of information, then it may risk increasing belief polarization. In my context, information on the role of educational attainment in explaining the wage gap marginally increased the belief gap between Democrats and Republicans about hiring discrimination. This resulted from Democrats overestimating how much of the wage gap driven by educational attainment and, upon learning the true estimate, subsequently increasing their beliefs about hiring discrimination.

I also find evidence that even when Democrats and Republicans are given information that leads to agreement about the extent of racial hiring discrimination, they may change their beliefs about whether the observed discrimination is a problem. At baseline, I ask participants whether they agree that (1) a difference in callback rates between applicants with Black-sounding names and applicants with White-sounding names would reflect that employers use race in making their callback decisions, and (2) a higher White callback rate in Bertrand and Mullainathan (2004) would be a problem. Then, after presenting participants with the result that the callback rate for White applicants was 50% higher than for Black applicants, I ask participants if they believe the difference in callback rates (1) reflects that employers base their callback decisions on applicant race, and (2) is a problem. Relative to their baseline responses, Republicans are more likely than Democrats to decrease their agreement to both questions.

In thinking about why Republicans decreased their agreement with these statements, perhaps the amount of hiring discrimination found in BM was small enough such that Republicans decrease their concern about the difference in callback rates being a problem. While Republicans did indeed overestimate the gap in callback rates at baseline, Democrats overestimated the gap even more. So, Democrats should be more likely to decrease their agreement at end line if this were the driving mechanism.

A remaining explanation is preference-biased updating (Benjamin, 2019). Because both groups want to believe they live in a world consistent with their political views (as demonstrated in Thaler (2019)), they may update more strongly to information aligning with their preferred state of the world. Even though Democrats learn they overestimated racial hiring discrimination, they remain consistent that discrimination is a problem. Republicans, conversely, use the fact that there is less racial hiring discrimination than they thought to update downward on discrimination being a problem. Biased updating on these unincentivized questions is in line with evidence suggesting that people are more likely to exhibit motivated updating when they face no monetary incentive for accuracy (Prior et al., 2015). This finding highlights an avenue through which belief convergence about the extent of racial hiring discrimination may not translate to a convergence in policy demand.

These findings contribute to empirical literature showing that information dissemination may not be successful at closing polarization in political preferences or policy demand (Haaland and Roth, 2021; Marino et al., 2023). I contribute by demonstrating two channels through which this occurs: (1) differences in information processing that prevent belief convergence, and (2) preference-biased updating on whether the information suggests a problem.

Much of the theoretical literature on belief polarization explores channels through which people may be exposed to differing sets of information, including selective information sharing (Levy and Razin, 2018; Bowen et al., 2023) and biased news sources (Mullainathan and Shleifer, 2005; Levendusky, 2013; Perego and Yuksel, 2022). An exception is Andreoni and Mylovanov (2012) in which the authors identify a channel through which disagreements persist in the face of common information when that information is multi-dimensional. This finding may be particularly relevant when the information is open to interpretation and requires processing based on one's worldview.

Interestingly, Democrats and Republicans throughout the study consistently overestimate racial hiring discrimination relative to BM. In the final round, if participants were to fully update on the extent of racial hiring discrimination based on the results from Jacquemet and Yannelis (2012), then they would correctly predict the BM results, as both studies found the same ratio in callback rates between applicants with White-sounding and Black-sounding names. However, the average callback rates in these two studies were quite different. While BM found that applicants with Black-sounding names had to be sent out 15 times to get one callback, Jacquemet and Yannelis (2012) found that applicants with Black-sounding names had to be sent out only 6 times. Participants may have viewed the low Black callback rate in BM as a signal of hiring discrimination, leading them to overestimate the observed discrimination in BM even in the final round of the study.

1.7 Conclusion

I conduct an experiment to investigate how information affects belief polarization among 1100 Democrats and 1100 Republicans about racial discrimination in hiring. I measure participants' beliefs about racial hiring discrimination using an incentivized and quantified method from Haaland and Roth (2021) in which participants state their predictions of racial differences in callback rates from the fake resume study in Bertrand and Mullainathan (2004).

I first establish that Democrats believe there is more racial discrimination in hiring than Republicans do at baseline. I then explore how beliefs about racial discrimination in hiring respond to varying pieces of information. I find that Democrats update their beliefs in response to the Black-White wage gap and in response to the role of educational attainment in explaining the wage gap. Republicans, on the other hand, do not significantly update their beliefs in either of these cases, likely due to differing views on the relationship between wage gap information and labor market discrimination. The divergent interpretations of this information between Democrats and Republicans prevents the information from reducing belief polarization, and even risks increasing it.

Even when Democrats and Republicans eventually agree about the extent of racial discrimination in a fake resume study, they may exhibit biased reasoning in their interpretations. After learning the results of BM, Republicans are more likely than Democrats to decrease, relative to their baseline response, their beliefs that (1) BM is designed to measure racial discrimination in hiring, and (2) the higher White callback rate in BM is a problem. This finding demonstrates a channel through which convergence in beliefs may not yield convergence in policy demand.

The main findings in this paper expose key drawbacks in using information to decrease belief polarization. When information requires processing based on one's world view, belief polarization may fail to decrease in response, and may even increase. Furthermore, even when groups agree on the facts surrounding a political topic, they may change their beliefs in another dimension: whether the facts reflect a problem. As Democrats and Republicans increasingly view the world through different lenses, these risks may become more prevalent.

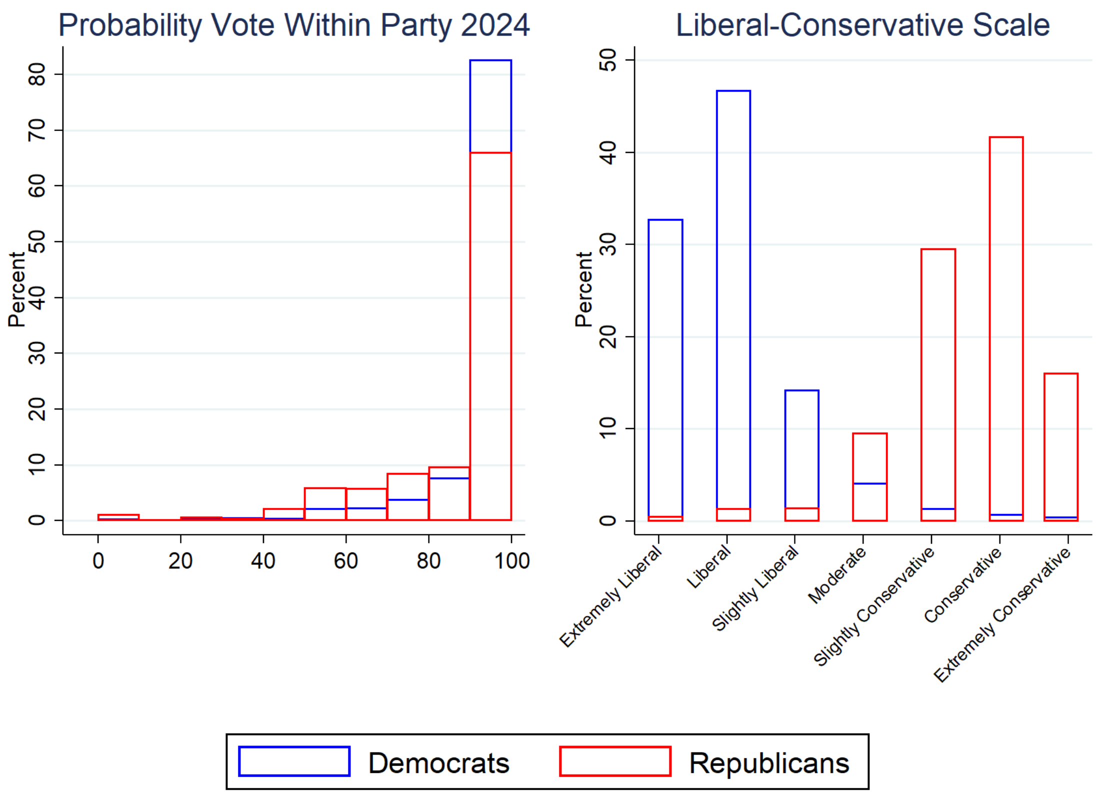

Figure 1.4. Self-Reported Degree of Conservativeness and Liberalness

Notes. The left panel displays the distribution of participants' self-reported probability they will vote for their political party candidate in the 2024 presidential election, conditional on voting. That is, Democrats report the likelihood they will vote for the Democratic presidential candidate in 2024, and Republicans report the likelihood they will vote for the Republican presidential candidate in 2024, conditional on voting in the election. The right panel displays participants' responses about their political ideology, on a seven-point scale from "extremely liberal" to "extremely conservative." For both panels, the distribution of Democrats' responses is shown in blue, and Republican responses are in red.

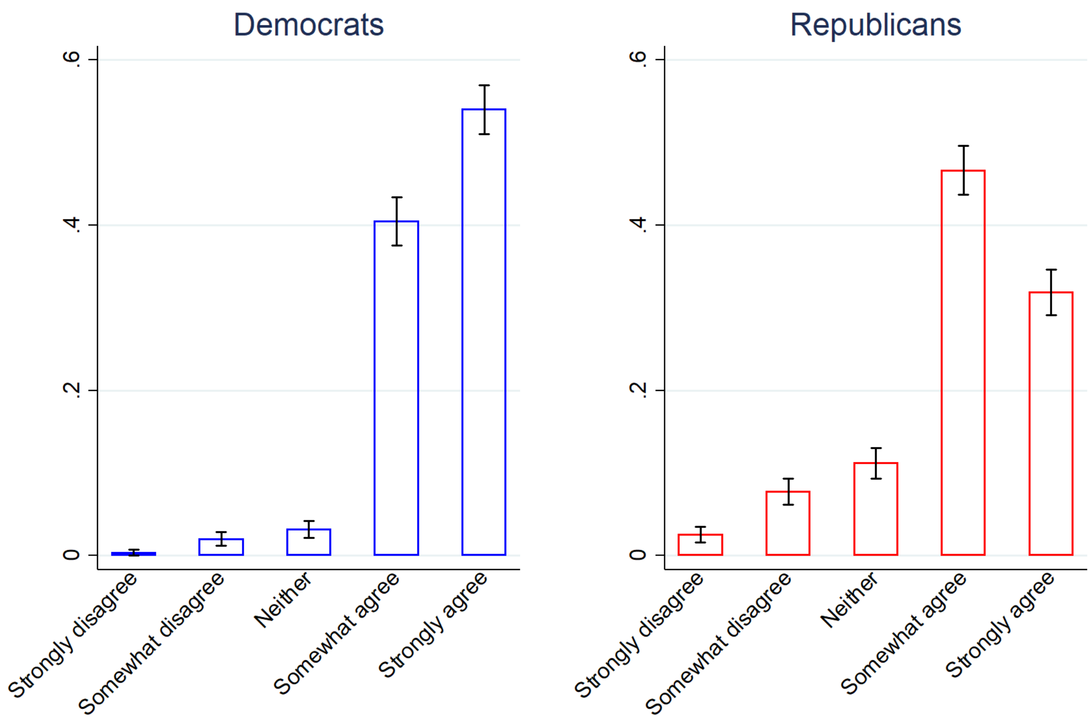

Figure 1.5. Agree BM tests whether employers use race in callback decisions

Notes. The figures show the distributions of participants' responses on a scale from "strongly disagree" to "strongly agree" to the following statement. "If the researchers find a difference in callback rates between applicants with Black-sounding names and applicants with White-sounding names in Experiment A [BM], this would reflect that employers base their callback decisions in part on the race of the applicant." The left panel restricts my sample to Democrats, and the right panel restricts to Republicans.

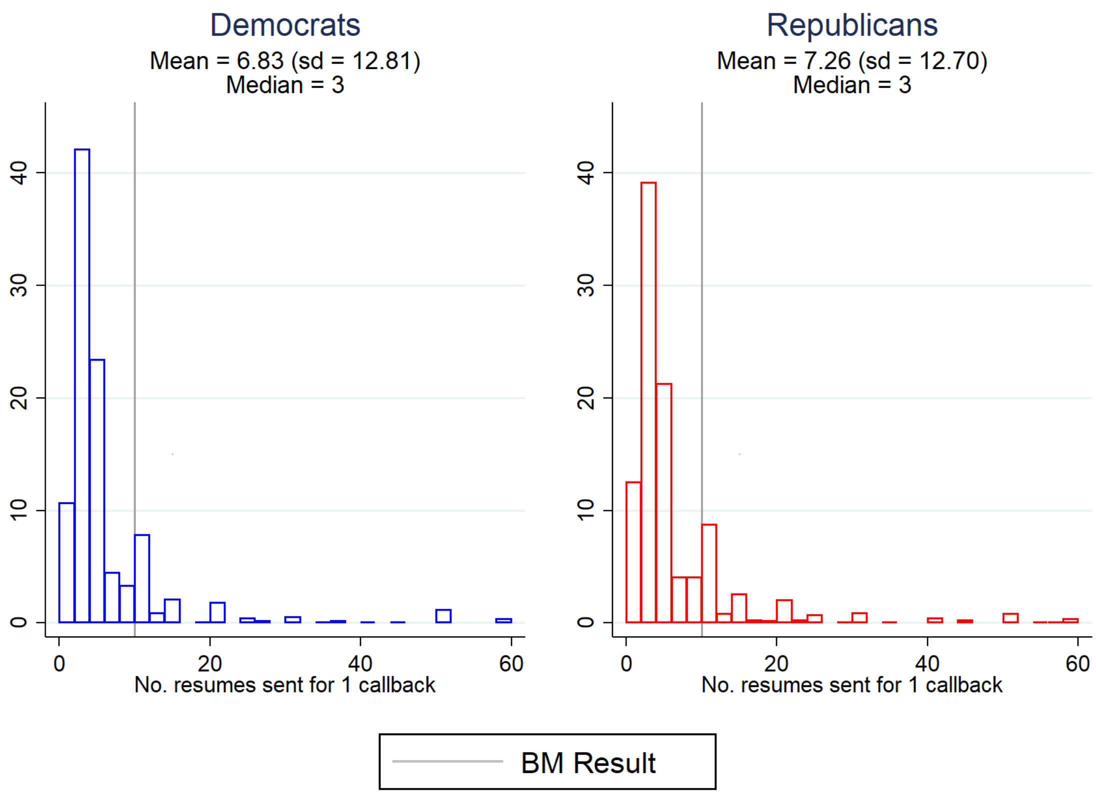

Figure 1.6. Priors on White callback rate in BM

Notes. The figures show the distributions of participants' baseline predictions of the number of resumes applicants with White-sounding names had to be sent out to employers to get one callback for an interview in Bertrand and Mullainathan (2004). The left panel restricts my sample to Democrats, and the right panel restricts to Republicans. The actual number of times resumes with White-sounding names had to be sent out to receive one callback was 10 times, as depicted with the grey vertical line on each panel.

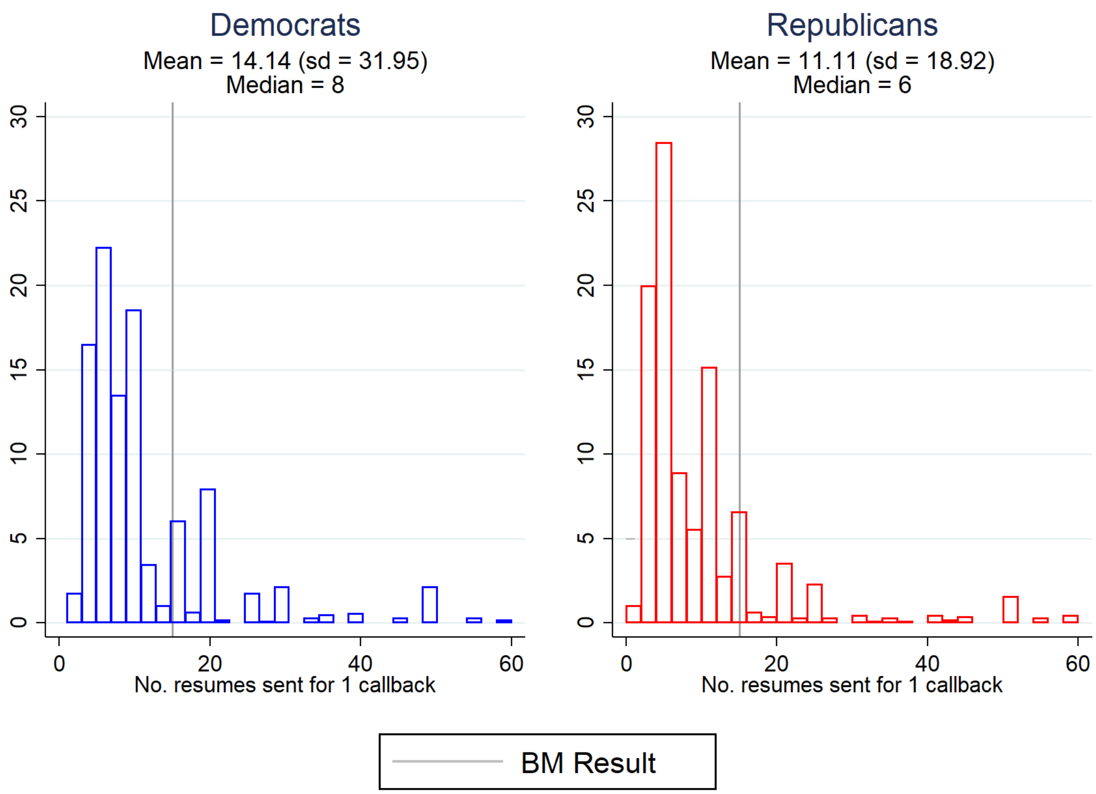

Figure 1.7. Priors on Black callback rate in BM

Notes. The figures show the distributions of participants' baseline predictions of the number of resumes applicants with Black-sounding names had to be sent out to employers to get one callback for an interview in Bertrand and Mullainathan (2004). The left panel restricts my sample to Democrats, and the right panel restricts to Republicans. The actual number of times resumes with Black-sounding names had to be sent out to receive one callback was 15 times, as depicted with the grey vertical line on each panel.

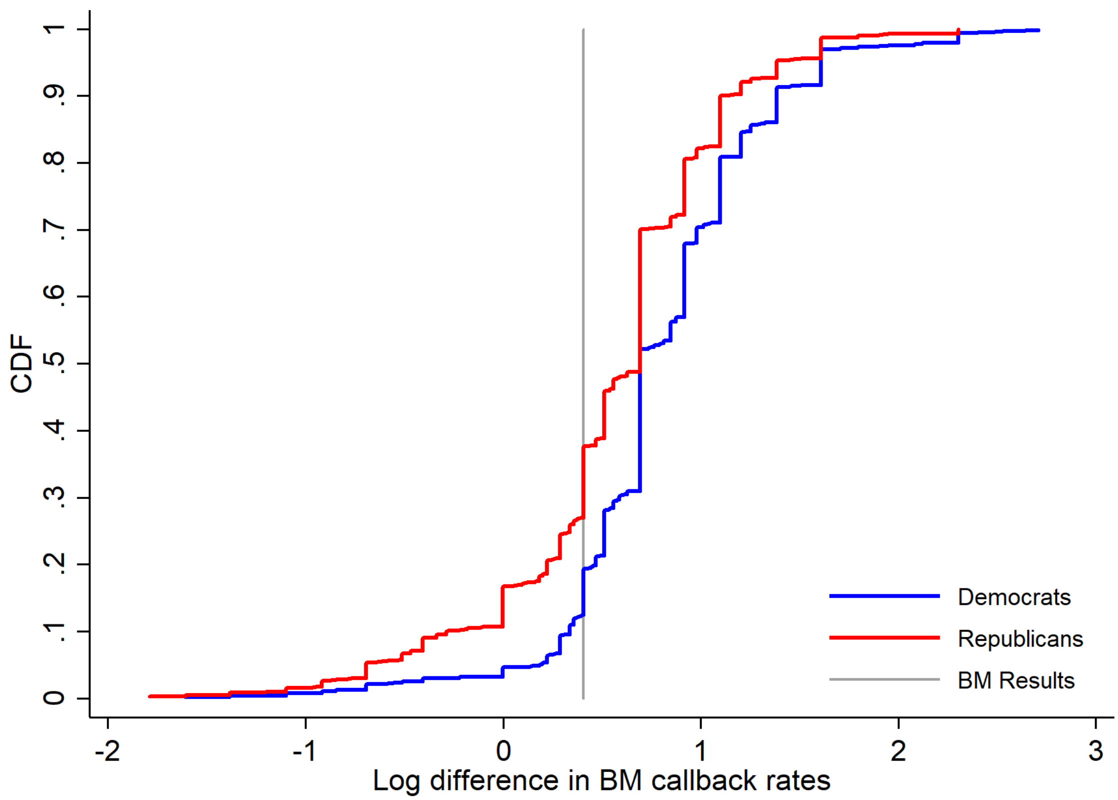

Figure 1.8. Round 1 Beliefs about racial discrimination in hiring

Notes. This figure shows the cumulative distribution functions of participants' baseline beliefs about racial discrimination in hiring in Round 1, split by political affiliation. Beliefs are measured as the log difference between participants' predictions of the number of times resumes with Black-sounding names had to be sent out to receive one callback for an interview in Bertrand and Mullainathan (2004) and the number of times resumes with White-sounding names had to be sent out to receive one callback. The actual log difference in callback rates from BM is depicted by the vertical grey line. Democrats' beliefs are shown in blue, and Republicans' beliefs are in red.

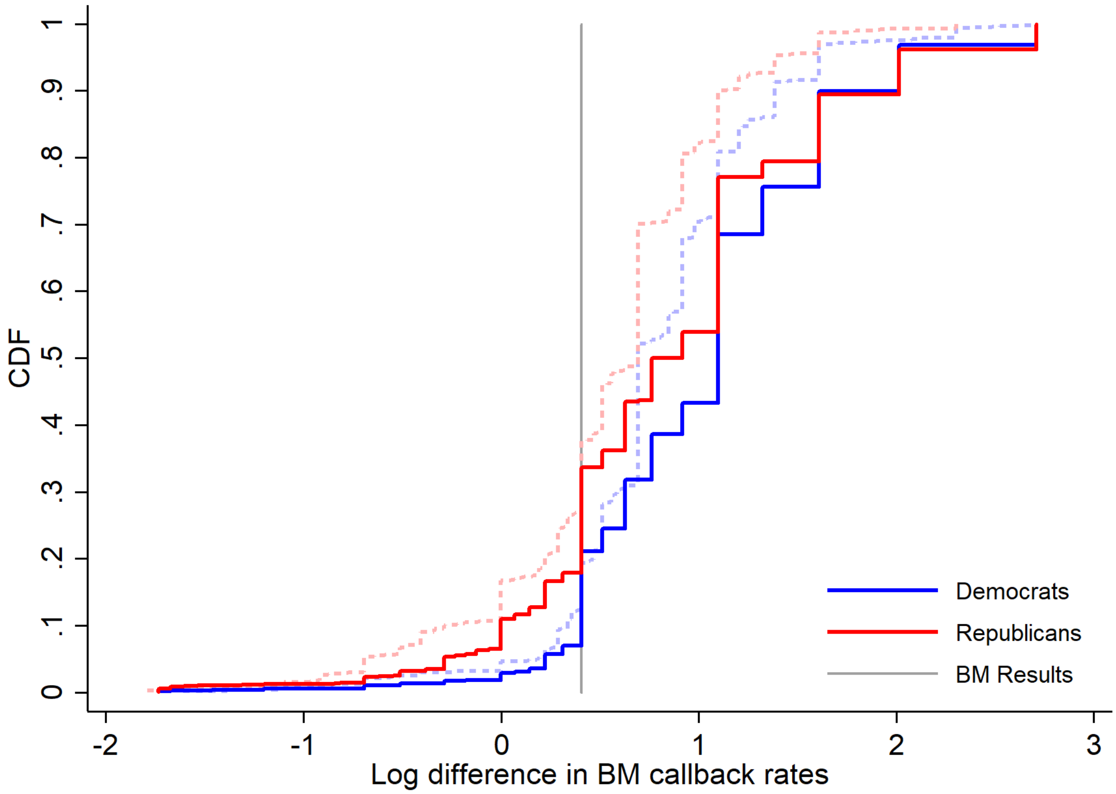

Figure 1.9. Round 2 Beliefs about racial discrimination in hiring

Notes. This figure shows the cumulative distribution functions of participants' beliefs about racial discrimination in hiring in Round 2 (bolded solid lines), split by political affiliation. Participants' Round 1 beliefs are shown as dashed lines for ease of comparison. In Round 2, participants are told the number of times resumes with Black-sounding names had to be sent out to receive one callback for an interview. Beliefs in Round 2 are measured as the log difference between the actual number of times resumes with Black-sounding names had to be sent out to receive one callback (15) in Bertrand and Mullainathan (2004) and participants' predictions of the number of times resumes with White-sounding names had to be sent out to receive one callback in BM. The actual log difference in callback rates from BM is depicted by the vertical grey line. Democrats' beliefs are shown in blue, and Republicans' beliefs are in red.

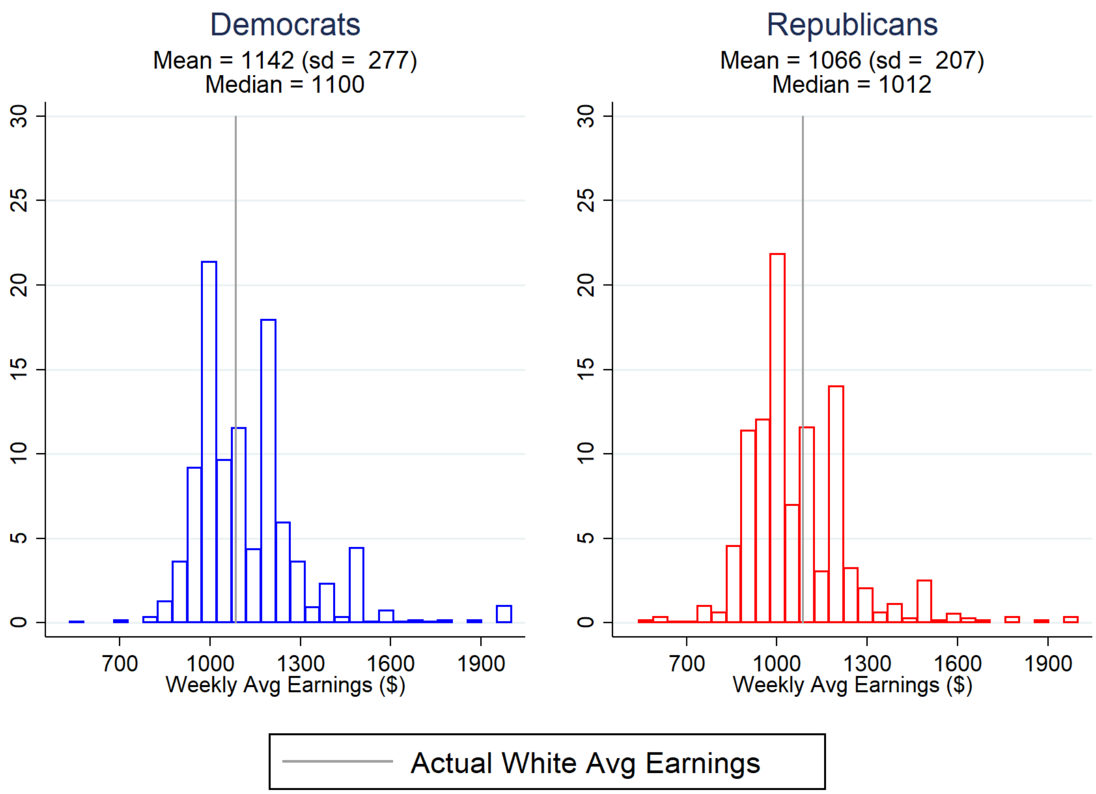

Figure 1.10. Priors on White average weekly earnings

Notes. The figures show the distributions of participants' priors about average weekly earnings among White full-time workers in the US. Before responding, participants are first told the average weekly earnings for Black full-time workers ($844) in the US according to the US Bureau of Labor Statistics. The left panel restricts my sample to Democrats, and the right panel restricts to Republicans. The true average weekly earnings for White full-time workers in the US according to the US Bureau of Labor Statistics is $1085, as depicted by the grey vertical line on each panel.

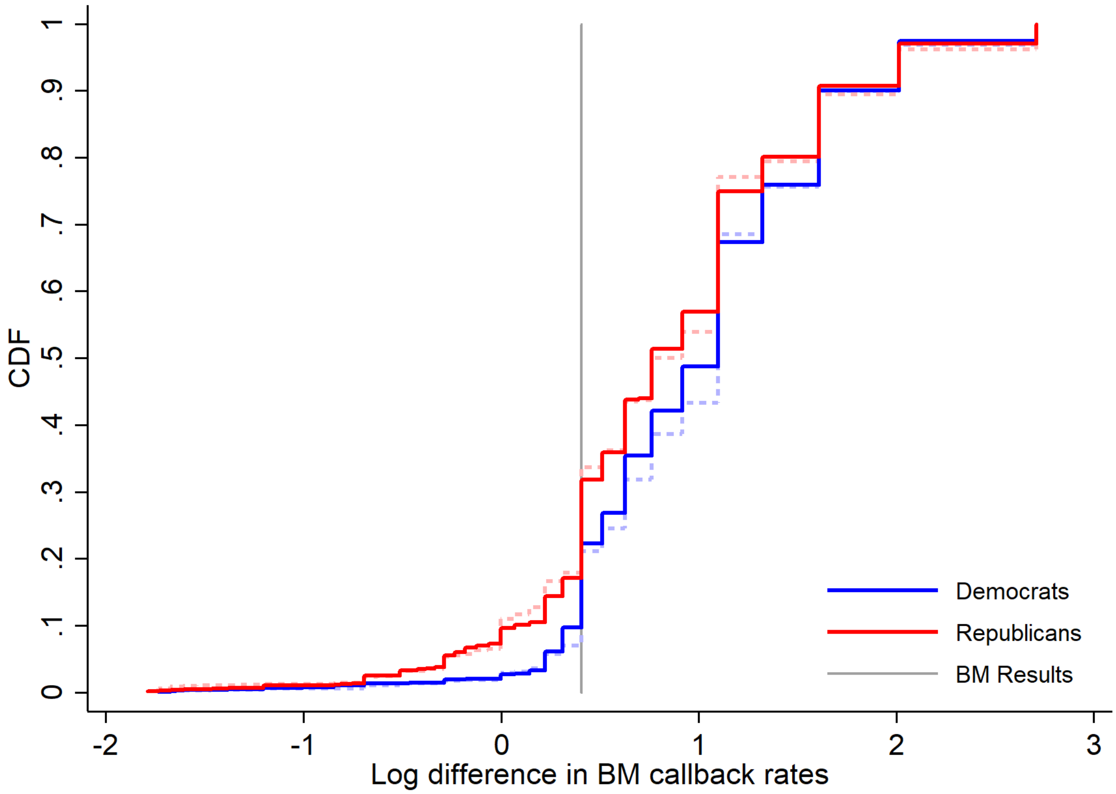

Figure 1.11. Round 3 Beliefs about racial discrimination in hiring

Notes. This figure shows the cumulative distribution functions of participants' beliefs about racial discrimination in hiring in Round 3 (bolded solid lines), split by political affiliation. Participants' Round 2 beliefs are shown as dashed lines for ease of comparison. In Round 3, participants are told the Black-White wage gap for full-time workers in the US. Beliefs are measured as the log difference between the actual number of times resumes with Black-sounding names had to be sent out to receive one callback (15) in Bertrand and Mullainathan (2004) and participants' predictions of the number of times resumes with White-sounding names had to be sent out to receive one callback in BM. The actual log difference in callback rates from BM is depicted by the vertical grey line. Democrats' beliefs are shown in blue, and Republicans' beliefs are in red.

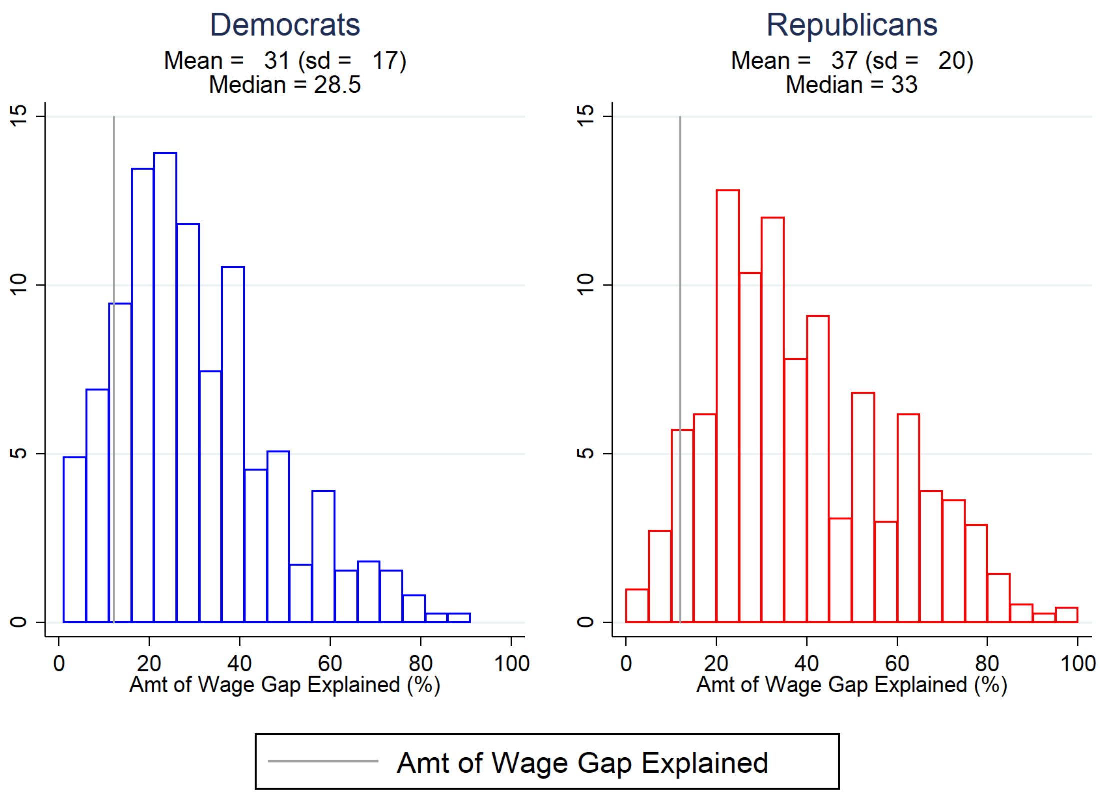

Figure 1.12. Priors on educational attainment

Notes. The figures show the distributions of participants' priors about the percentage of the Black-White wage gap among full-time workers in the US that is explained by differences in educational attainment between Black and White workers. The left panel restricts my sample to Democrats, and the right panel restricts to Republicans. The actual amount of the wage gap explained by differences in educational attainment according to a Oaxaca-Blinder decomposition using data from the Bureau of Labor Statistics is 12%, as depicted by the grey vertical line on each panel.

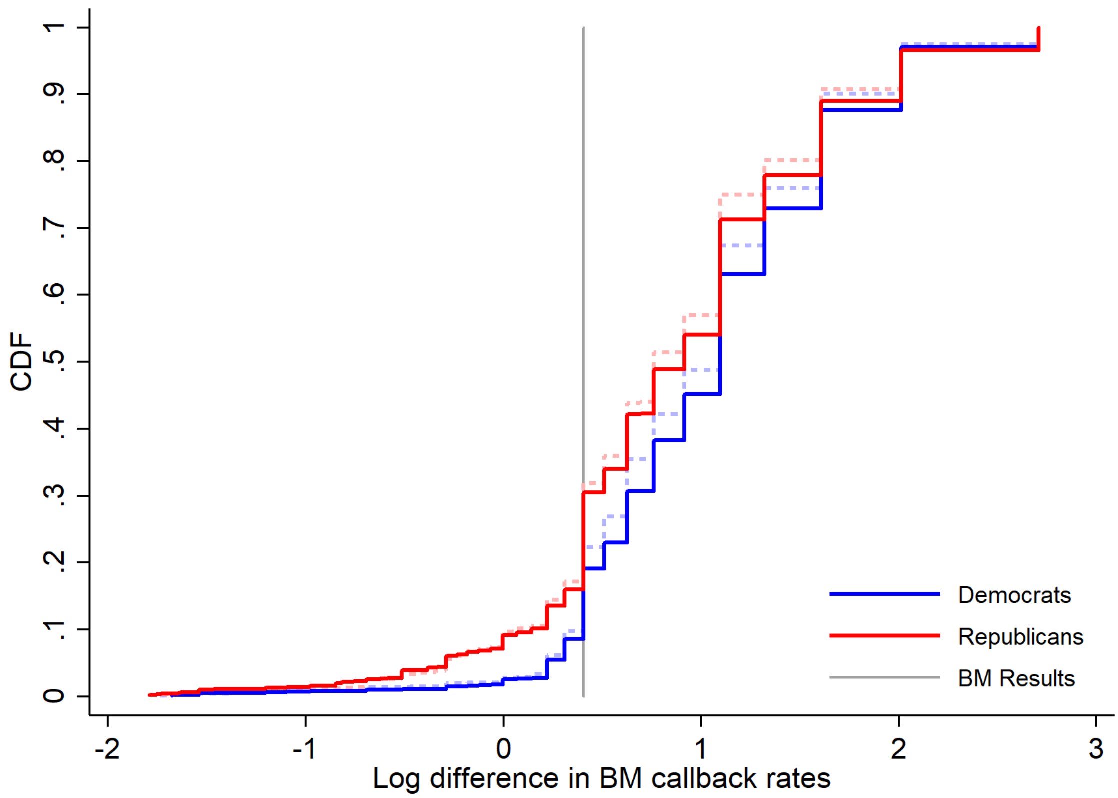

Figure 1.13. Round 4 Beliefs about racial discrimination in hiring

Notes. This figure shows the cumulative distribution functions of participants' beliefs about racial discrimination in hiring in Round 4 (bolded solid lines), split by political affiliation. Participants' Round 3 beliefs are shown as dashed lines for ease of comparison. In Round 4, participants are told the percent of the Black-White wage gap that is explained by differences in educational attainment between Black and White workers. Beliefs are measured as the log difference between the actual number of times resumes with Black-sounding names had to be sent out to receive one callback (15) in Bertrand and Mullainathan (2004) and participants' predictions of the number of times resumes with White-sounding names had to be sent out to receive one callback in BM. The actual log difference in callback rates from BM is depicted by the vertical grey line. Democrats' beliefs are shown in blue, and Republicans' beliefs are in red.

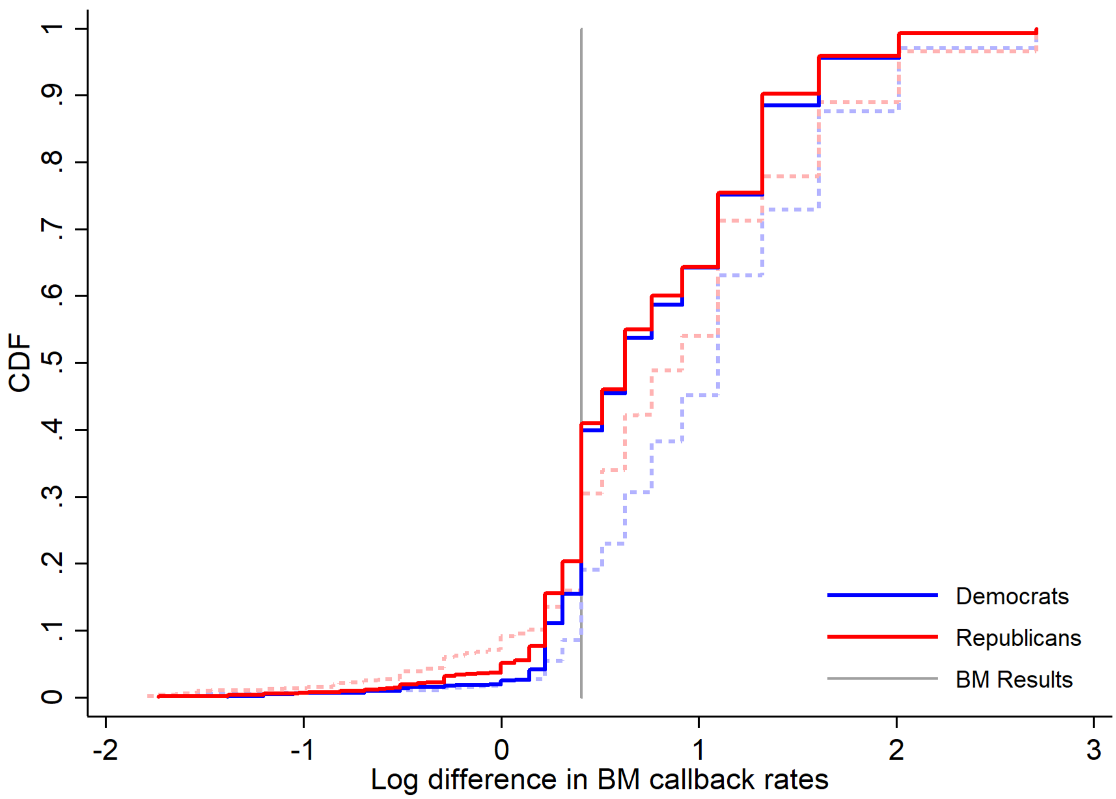

Figure 1.14. Round 5 Beliefs about racial discrimination in hiring

Notes. This figure shows the cumulative distribution functions of participants' beliefs about racial discrimination in hiring in Round 5 (bolded solid lines), split by political affiliation. Participants' Round 4 beliefs are shown as dashed lines for ease of comparison. In Round 5, participants are told the callback rates for applicants with Black-sounding names and for applicants with White-sounding names from the fake resume study in Jacquemet and Yannelis (2012). Beliefs are measured as the log difference between the actual number of times resumes with Black-sounding names had to be sent out to receive one callback (15) in Bertrand and Mullainathan (2004) and participants' predictions of the number of times resumes with White-sounding names had to be sent out to receive one callback in BM. The actual log difference in callback rates from BM is depicted by the vertical grey line. Democrats' beliefs are shown in blue, and Republicans' beliefs are in red.

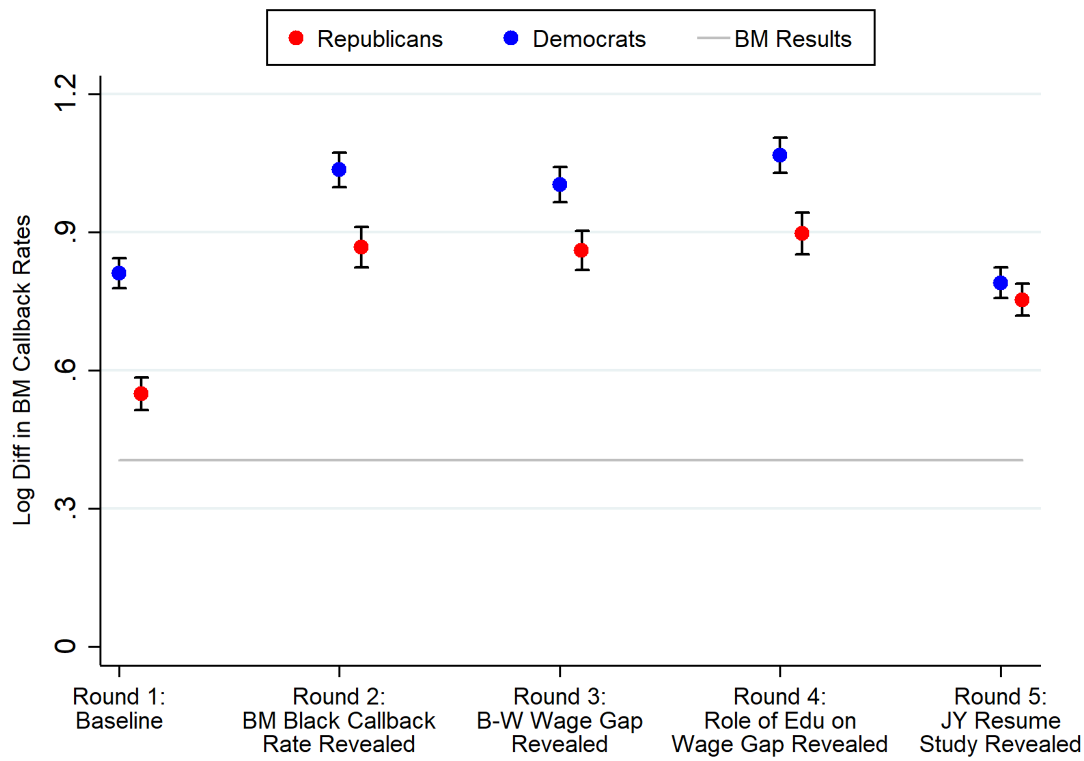

Figure 1.15. Beliefs about racial discrimination in hiring across rounds

Notes. This figure shows participants' average beliefs about racial discrimination in hiring with 95% confidence intervals in each of the five rounds of the study, split by political affiliation. Beliefs in Round 1 are measured as the log difference between participants' predictions of the number of times resumes with Black-sounding names had to be sent out to receive one callback for an interview in Bertrand and Mullainathan (2004) and the number of times resumes with White-sounding names had to be sent out to receive one callback. Beliefs in Round 2 to Round 5 are measured as the log difference between the actual number of times resumes with Black-sounding names had to be sent out to receive one callback (15) in Bertrand and Mullainathan (2004) and participants' predictions of the number of times resumes with White-sounding names had to be sent out to receive one callback in BM. The actual log difference in callback rates from BM is depicted by the horizontal grey line. Democrats' beliefs are shown in blue, and Republicans' beliefs are in red.

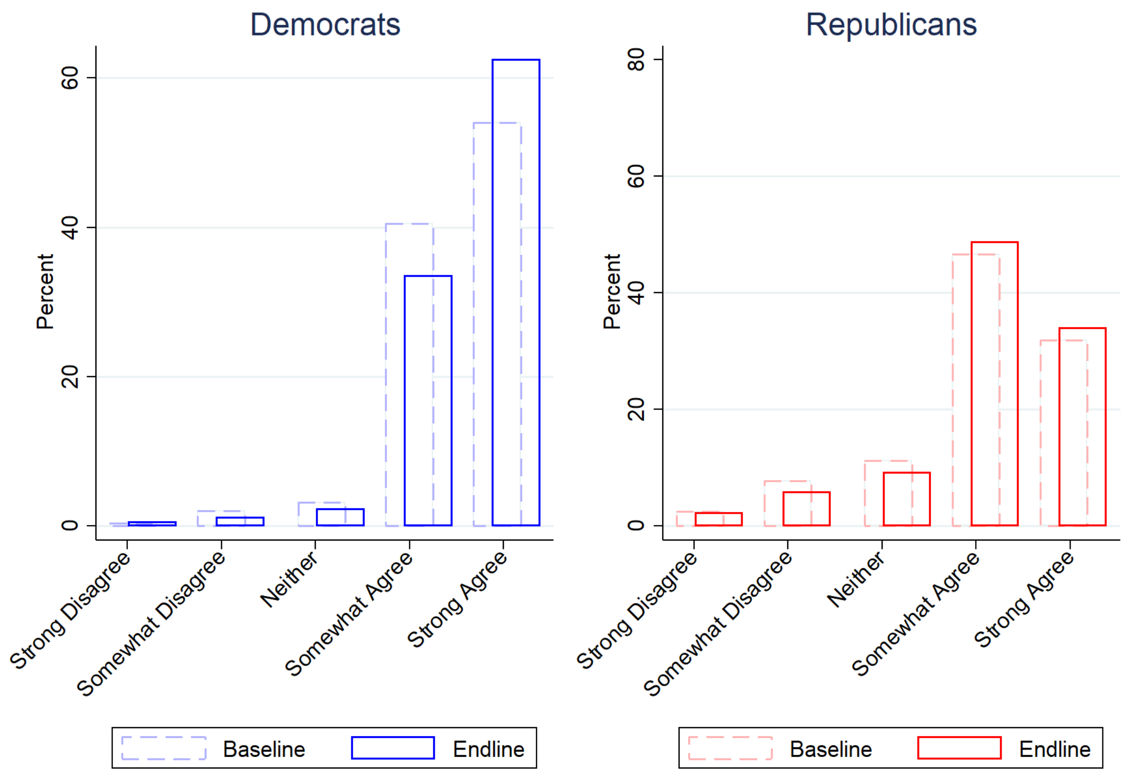

Figure 1.16. Agree BM tests whether employers use race in callback decisions

Notes. The figures show the distributions of participants' responses on a scale from "strongly disagree" to "strongly agree" at baseline and endline about the interpretation of BM. At baseline (depicted by dashed lines), participants are asked how much they agree with the following statement. "If the researchers find a difference in callback rates between applicants with Black-sounding names and applicants with White-sounding names in Experiment A[5], this would reflect that employers base their callback decisions in part on the race of the applicant." At the end of the study (depicted by solid lines), participants are asked how much they agree with the following statement. "The difference in callback rates between applicants with Black-sounding names and applicants with White-sounding names in Experiment A reflects that employers base their callback decisions in part on the race of the applicant." The left panel restricts my sample to Democrats, and the right panel restricts to Republicans.

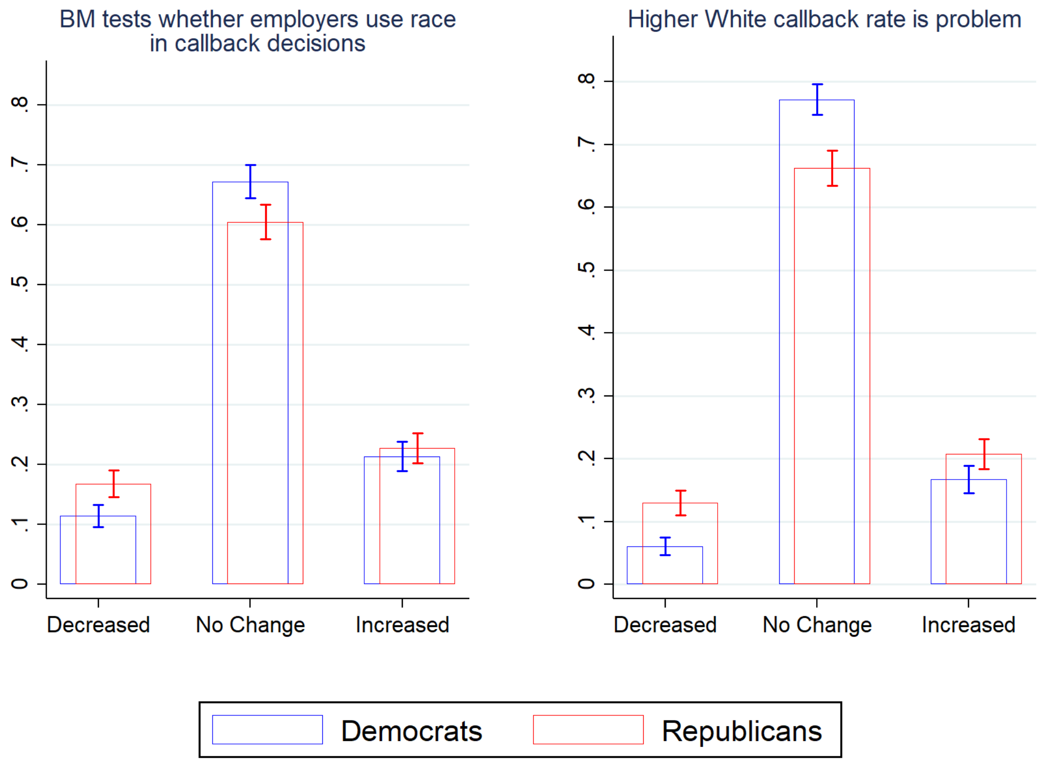

Figure 1.17. Consistency of participants' interpretations: Baseline vs. Endline

Notes. The figures show the proportion of Democrats (blue) and Republicans (red) who increase, decrease, and do not change their agreement on a scale from "strongly disagree" to "strongly agree" at baseline and endline about the interpretation of BM (left panel) and whether the BM results are a problem (right panel). Participants who decrease their agreement (i.e., select a lower level of agreement at endline relative to baseline) are included in the "Decreased" group. Participants who give the same response at baseline and endline are included in the "No change" group. Participants who increase their agreement (i.e., select a greater level of agreement at endline relative to baseline) are included in the "Increased" group. See figure notes in Figure 1.16 and Figure 1.18 for the question wordings about the interpretation of BM and whether the BM results are a problem, respectively.

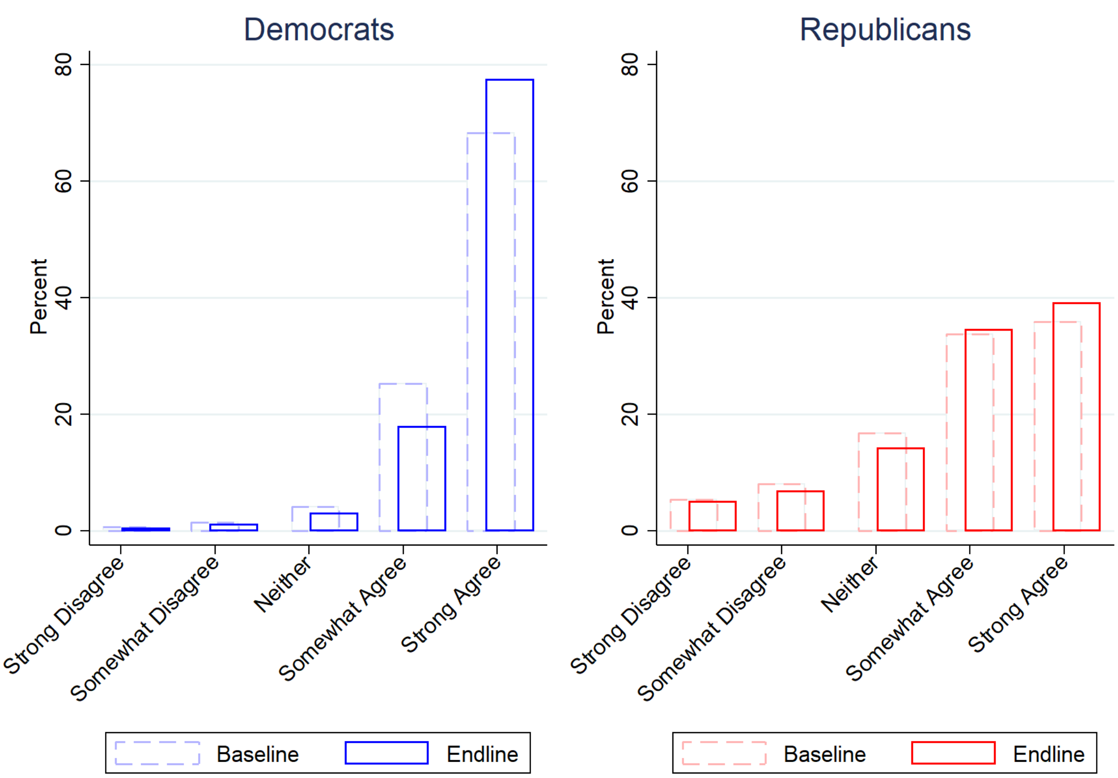

Figure 1.18. Agree that higher White callback rate in BM is a problem

Notes. The figures show the distributions of participants' responses on a scale from "strongly disagree" to "strongly agree" at baseline and endline about whether the BM results are a problem. At baseline (depicted by dashed lines), participants are asked how much they agree with the following statement. "If the researchers find that applicants with White-sounding names get callbacks more often than those with Black-sounding names, this would be a problem that should be solved." At the end of the study (depicted by solid lines), participants are asked how much they agree with the following statement. "The difference in callback rates is a problem that should be solved." The left panel restricts my sample to Democrats, and the right panel restricts to Republicans.

Chapter 2

Parental Investments Reduced Covid-19 Learning Loss: Evidence from a Longitudinal Field Experiment **with Shruti Jha, John A. List, and Anya Samek**

2.1 Introduction

There is growing recognition among economists that parents play a key role in the education production function (Francesconi and Heckman, 2016; List et al., 2018; Almond et al., 2018). Parents may be particularly important in the early years, during which investments in children are thought to have the greatest returns (Heckman, 2012). Programs that provide parenting education or give incentives to parents for investing in their children have resulted in short-term changes in children's academic skills and executive functions (Fryer et al., 2015; Chuan et al., 2021; Cassidy et al., 2017; Lunkenheimer et al., 2008; Sheridan et al., 2011; Schaub et al., 2019; Özler et al., 2018; García and Heckman, 2022). Less is known about the long-term effects of parent interventions on child outcomes (Jeong et al., 2021).

Parental investments became even more relevant during the Covid-19 pandemic, which resulted in worldwide school closures. At the peak of the pandemic, school closures affected nearly 1.6 billion students, or 94% of the world's student population (Doucet et al., 2020). This left most children in the care of their parents during the day. The majority of parents said that they assisted with their child's education during the school closures (Bansak and Starr, 2021).[1] Nevertheless, the school closures resulted in 'learning loss,' decreasing standardized test scores between 0.10 and 0.17 standard deviations (SD) (Betthäuser et al., 2023). The learning loss that children experienced suggests an important role of in-person schooling in the education production function. It also begs the question: what was the role of parents at mitigating such negative shocks, and could parents have been better prepared to compensate for them?

In this paper, we study whether early childhood investments in children - and in particular investments by parents - had a causal impact on reducing Covid-19 learning loss. To do so, we leverage a field experiment from 2010-14 in which households with children ages 3-5 years old were randomized to a parenting program, a preschool program or to a control group for 1-2 years. Ten years later in 2020-21, the children were exposed to school shut-downs during the Covid-19 pandemic. To evaluate the causal impact of these early childhood investments on learning loss, we compare the treatment and control groups on changes to standardized test scores during the Covid-19 school closure year. Randomization allows us to assess the impact of exogenous changes to early childhood investment on Covid-19 learning loss while controlling for other correlates such as socio-economic status (SES) or unmeasured factors.

The field experiment involved several parenting and preschool treatment arms. The parenting program included bi-monthly classes for parents on how to teach to their children at home and provided incentives to parents for doing so. A central tenet of the program was the concept of 'parents as their child's first teacher' - that is, parents were encouraged to think about their investments in their child as central to the education production function. Parents received incentives for completing program activities, and in different treatment arms were either compensated in cash or in money deposited in an account to be used for their child's future post-secondary education. The preschool program provided a free, full-day preschool to the children. The preschool included three treatment arms: a standard, 9-month long program, a

[1] In a survey conducted between April and July 2020 with 200,000 US households and K-12 children, 72% of parents indicated that they spent time assisting their children with schoolwork, and conditional on providing assistance, parents spent about 2.6 hours per day.(Bansak and Starr, 2021).

shorter, 2-month long program delivered in the summer immediately prior to Kindergarten entry and a standard-length program with a low-touch parent intervention component.

As reported in Fryer et al. (2015), the parenting program had short-term impacts on children's executive functioning skills but not on their academic skills[2]. As reported in Fryer Jr et al. (2020), the standard-length preschool program and the shorter program had short-term impacts on the children's academic skills but not on their executive functions. As reported in Castillo et al. (2023), the effects of most programs faded out by the beginning of Covid-19.

When the children were 9-14 years old – by the beginning of the Covid-19 pandemic – the parenting program had no effects on standardized test scores, disciplinary outcomes, or non-cognitive skills, relative to the control group. That is, program effects appear to have faded out, which is commonly seen in related studies (see Almond et al. (2018) for a summary). However, program effects appear to have re-emerged due to the negative shock to schooling investments caused by the Covid-19 school closures. In particular, we find that the parenting program had a protective causal impact on Covid-19 learning loss. While test scores in the control group dropped by 0.30 standard deviations, test scores in the parenting group dropped only by 0.11 standard deviations, a statistically significant difference ($p = 0.01$). Moreover, the effect was driven by the Parent Academy-College program, which not only delivered early childhood parenting advice but also provided incentives upon high school completion. Unlike the parenting program, the preschool program did not have a protective effect. Test scores in the preschool group dropped by 0.25 standard deviations, a drop that is statistically indistinguishable from that of the control group ($p = 0.61$).

Our study provides evidence that interventions in early childhood have long-term impacts, even if they are not immediately apparent. This is similar to work by Heckman et al. (2013) showing that effects of early childhood programs fade out but then re-emerge in adulthood. Little is known about why this happens as there is limited data on middle-childhood and adolescent

[2]Executive functioning skills measured by Fryer et al. (2015) include working memory, inhibitory control and attention.

outcomes - a review refers to this gap as the 'missing middle' (Almond et al., 2018). Our study represents a significant opportunity to bridge the gap in understanding of the impact of early childhood interventions during the missing middle years. No study, that we are aware of, has explored the long-term impact of a parenting program such as ours.

Our study also points to the importance of parents as an input in the education production function. It is surprising that the effect of the parenting intervention was not apparent while schools were in session, and only emerged when one of the inputs to the education production was disrupted. This suggests that the effects of parental investments may have been overshadowed by the impacts of schools. When the pandemic shut down schools, parental investments played a larger role in child academic performances, as shown through the widening of the SES skills gap among children during the pandemic. Two contributing factors to this widening gap include the inability of low SES parents to work from home (Agostinelli et al., 2022) and the lower likelihood of low SES parents seeking alternate resources (such as tutors), to mitigate the effects of school closures on children's skills (Bacher-Hicks et al., 2021). Together, these findings suggest that parent resources played an especially important role on children's learning during the pandemic. [3]

The paper proceeds as follows. In Section 2, we introduce a conceptual framework to investigate the relationships between Covid-19, student test scores, and the role of parents. In Section 3, we discuss our methods, including a description of the original experiment and a discussion of the data collection pre- and post- Covid-19. In Section 4, we provide the results. Section 5 concludes.

[3]The only other study we are aware of that investigates parents' roles in mitigating Covid-19 learning loss is Hassan et al. (2023) who find that a telementoring intervention with parents during Covid-19 reduced learning loss among children in Bangladesh. We contribute by (1) evaluating long-term effects and (2) investigating the relationship in a developed context.

2.2 Conceptual Framework

To investigate the dynamics of Covid-19 learning loss among students, we develop a conceptual framework. We begin with a model of child skill development. We let child skills, M_{it}, be a vector of skills and economic preferences for individual i at time t. Following Heckman (2007), we model child skill development as a function of the child's skill in the prior period, individual characteristics, and investments in the given period. We allow for inputs from the child, the child's parents/guardians, and the child's school.

We model child skill development as follows.

$$M_{it} = f(M_{it-1}, Z_i, C_{it}, P_{it}, S_{it}) \tag{2.1}$$

where M_{it-1} is child i's skills in the prior period and Z_i are individual characteristics. C_{it}, P_{it}, and S_{it} are inputs from the child, parent, and school, respectively.

The production function $f(\cdot)$ is increasing in each argument and quasi-concave. We assume that inputs between the child, parent, and school are substitutable. Consider a constant elasticities of substitution (CES) production function as in Heckman (2007) and Almond et al. (2018), which allows us to capture substitutability of the inputs.

$$M_{it} = g(M_{it-1}, Z_i, A(\gamma_1 C_{it}^{\phi}, \gamma_2 P_{it}^{\phi}, \gamma_3 S_{it}^{\phi})^{\frac{\psi}{\phi}}) \tag{2.2}$$

In this equation, A represents total factor productivity. γ_j for $j \in \{1,2,3\}$ indicates the relative importance of each input, with $\gamma_1 + \gamma_2 + \gamma_3 = 1$. ψ refers to the returns to scale, where $\psi = 1$ denotes constant returns to scale. ϕ indicates the ability for parent, school, and child inputs to compensate for one another. As $\phi \to 1$, the inputs approach perfect substitution.

Every period, each agent (child, parent, and school) chooses their optimal investment decision, given the other agents' input decisions, child characteristics, and child's prior skill level. Investment decisions are subject to a budget constraint $B_{it} = g(p_x, m)$ where p_x is a vector of prices and m is income.

During the Covid-induced school shutdowns, schools were no longer able to maintain their previous levels of inputs to children. In relation to our model, the Covid-19 pandemic led to a large negative shock to schools' inputs, S_{it}. Because the child skill production function is increasing in each argument, this negative shock in school inputs would lead to a drop in potential child skills, holding all else constant. Due to our assumption of the substitutability of inputs, however, parents and children could compensate for the Covid-induced drop in school inputs by increasing their inputs in response.

We return to this conceptual framework in Section 5 to evaluate our results.

In thinking about how the agents in our model respond, let's evaluate the impact on parents' and children's investment decisions. Consider a two-period model $t \in \{0,1\}$, where $t = 0$ refers to the pre-Covid period and $t = 1$ refers to the period immediately after the Covid-19 shut down schools. This introduced a negative shock on schooling levels, so $S_{i1} < S_{i0}$. Given our assumption of substitutability, parent and child inputs can help offset these negative effects of Covid-19 on school inputs.

Consider the parent's optimal investment decision, P_{i1}^{*}, during school closures.

$$P_{i1}^{*} = G(C_{i1}, S_{i1} | M_{i1}^{\prime}, \Omega_{i1}) \tag{2.3}$$

Given our assumption of substitutability of inputs, parents can help offset the reduction in school inputs by increasing their inputs.

Similarly, consider the child's optimal investment decision, C_{i1}^{*}, during school closures.

$$C_{i1}^{*} = G(P_{i1}, S_{i1} | M_{i1}^{\prime}, \Omega_{i1}) \tag{2.4}$$

Given our assumption of substitutability of inputs, children can also help offset the reduction in school inputs by increasing their inputs.

2.3 Methods

2.3.1 Chicago Heights Early Childhood Center

We leverage the Chicago Heights Early Childhood Center (CHECC), a field experiment that was conducted in 2010-2014 (Fryer et al., 2015). CHECC was located in Chicago Heights, IL, a relatively low-income, high-minority south-side suburb of Chicago. The field experiment focused on evaluating the impact of various interventions with 3-5 year old children with the aim of reducing the academic achievement gap. Households who registered for CHECC in 2010 and 2011 were randomized to one of several programs, including a control group and a Parent Academy group aimed at teaching parents how to teach to their children.

Households participated in CHECC for 1-2 years, depending on the year they enrolled and the age of their child. The control group did not receive any educational interventions from CHECC, but did participate in periodic assessments and surveys. The Parent Academy group received educational workshops and incentives for investing in their child. In each year, there were 18 Parent Academy workshops and 17 homework assignments that involved teaching activities parents undertook with kids at home. Parent Academy lessons focused on teaching parents how to help their children with both academic skills (such as spelling and counting) as well as executive functioning skills (such as working memory and inhibitory control). The lessons also emphasized the parent's role as the "child's first teacher."

Parents received incentives for participating in the Parent Academy. They could earn up to $6,900 per year, which amounted to over a quarter (27.6%) of median household income for the analysis sample. These earnings were based partly on attendance and partly on performance. All Parent Academy families received immediate cash incentives ($100) for attending each workshop. In addition, families received performance incentives for completing homework assignments and for their child's scores on assessments. Half of Parent Academy families were randomized to receive these incentives immediately in cash (Parent Academy-Cash), whereas the other half were randomized to receive these incentives in a savings account that would become accessible if and when the treatment child attended post-secondary education years later (Parent Academy-College).

Parents and children participated in an assessment and survey before, during and immediately after treatment. The assessment evaluated child academic skills (reading, writing and math) using the Peabody Picture Vocabulary-III test (PPVT-III, Dunn and Dunn (1965)) and the Woodcock-Johnson-III test (WJ-III, Woodcock, McGrew, and Mather, 2001). The assessment also evaluated child executive functioning skills (working memory, inhibitory control and emotions) using tests developed by Blair and Willoughby (2006a, b) and the Preschool Self-Regulation Assessment (PSRA). The parent survey included questions about household socio-demographics, as well as questions about parent investments (time spent teaching to the child), parent self-efficacy (confidence in teaching) and beliefs about child abilities (see Appendix A).

Related papers have summarized the impact of Parent Academy on child short-term outcomes before Covid-19. Fryer et al. (2015) showed that Parent Academy affected both academic and executive functioning skills among Hispanic and White children. Fryer et al. (2015) also showed that children with above median pre-treatment skills benefited the most from the program. Castillo et. al (2023) showed that Parent Academy affected executive functioning skills immediately after the program, but did not have medium-term effects on test scores, grades or disciplinary referrals in grades K-8. Chuan et al. (2022) discussed the association of early childhood parental investments in reading and math on test scores in middle childhood. No paper has evaluated the impact of Covid-19 on learning loss in this sample.

2.3.2 Experience During Covid-19

The state of Illinois moved public schools to fully remote instruction in March of 2020 and remained mostly remote for the entirety of the 2020-21 academic school year. The remote model included synchronous Zoom instruction and asynchronous components. Some students received Chromebooks, iPads, or other laptop devices from the district to participate in remote instruction. Available technology and school resources varied by districts.

Some districts began a hybrid model before returning to full-in person learning, which included students attending in-person for part of the week and learning asynchronously for part of the week. Attendance was taken as usual, but there was no consequence for extensive absences in most districts. Standardized tests was also cancelled by most districts in spring 2021. In general, schools returned to in-person learning in fall 2021, although some districts employed a delayed start due to a new Covid-19 wave.

2.3.3 Covid-19 Test Score Data

Following participation in CHECC, children matriculated into public or private schools in this and surrounding areas. We obtained administrative data on Illinois Assessment of Readiness (IAR) scores on all children in the Illinois public schools from the state of Illinois. The public schools administered the IAR test each spring when the children were in grades 3-8. IAR is administered to all public school children that are present at the time of testing. Districts are not permitted to allow students to opt out of taking the test, although in practice some students do opt out, i.e., by not showing up on the testing date. The IAR test includes an assessment of skills in English and Language Arts (ELA) and Mathematics (MAT) and is based on the Common Core State Standards in Illinois. Each student's raw score is converted into a scale score which ranges from 650 to 850, by adjusting for slight differences in difficulty among the forms and administrations of the test. Students are also assigned a performance level, which is a categorical level defined by the scale score. Students with scale scores below 700 have not yet met expectations, between 700 and 725 have partially met expectations, between 725 and 750 have approached expectations, and above 750 have either met or exceeded expectations.

Prior to shutdown, in spring 2018-2019, standardized tests were administered as normal. Testing was disrupted in the spring of 2020 when schools shut down. Due to schools providing education in a variety of methods during spring 2021, test administration did not return to normal. ISBE allowed districts two testing windows in either spring or fall 2021, but scores were categorized only as spring 2021. Districts had to decide the best way to execute testing and arrange multiple opportunities to test students. After connecting with multiple districts, it appears that if students were learning remotely during the spring, they requested parents to make an appointment to test, often on Saturdays. This resulted in many students not testing due to parents not showing for their assessment appointment or opting out of testing this year. One district shared families had to opt in to test during this time period. If students were hybrid learning, districts attempted to test when in school. A large number of students did not test in spring 2021 and these scores were reported as "NULL" or "absent from testing". In 2022, all districts returned to typical testing procedures, as students were in attendance and remote learning was not an option. Therefore, we use the 2018-19 data as our measure of pre-Covid scores and the 2020-2021 (when available) and 2021-2022 data as our measure of post-Covid scores.

2.3.4 Analysis Sample

A total of 2,185 children were randomized as part of the CHECC field experiment. The focus in our study is the Parent Academy program and its relevant control group, which includes 863 children (316 in Parent Academy and 547 in Control) who were randomized to the program in 2010 and 2011. We limit our analysis to children who were in the 3rd-8th grades in both 2018-19 (pre-Covid) and 2021-22 (post-Covid), since IAR scores are only collected within this age range. This gives us a total of 676 children (240 in Parent Academy and 436 in Control) who are eligible for analysis. However, data on IAR scores in both the pre-Covid and post-Covid period is only available for 338 (50.0%) of these children (136, 56.7% for Parent Academy and 202, 46.3% for control) either due to students exiting the public school system or due to imperfect administration of the standardized test by the district (especially in the post-Covid period).

Table 2.1 provides demographic summary statistics of our analysis sample by treatment. Similar to the surrounding community in Chicago Heights, our sample is lower-income and have higher representation of racial minorities than the general US population. More than a third of the children in our sample come from families with an annual income of less than $16,000, and less than a third have an annual family income greater than $36,000. Only 10% of our sample is white. The remaining 90% is Black or Hispanic ($< 1\%$ is another race).

We compare students in the Parent Academy treatment and the control group across various demographics to check that our sample is balanced. There are some small differences in characteristics across our treatment and control groups. For example, the control group is slightly more Black ($p = 0.12$) and less Hispanic ($p = 0.22$) than the Parent Academy treatment. However, none of the differences are statistically significant. The lack of significant differences by treatment suggests that the samples are sufficiently comparable for our analysis.

2.4 Results

2.4.1 Overview and Short-Term Effects

We first provide a descriptive analysis of our sample. During the pre-Covid-19 year (2018-19), the students in the control group were in grades 4-6 and had an average IAR score of 723.4 which is within the performance level "Approaching Expectations". Figure 2.1 compares the performance of the CHECC control group to all students taking the IAR in Illinois. In Mathematics (MAT), 13% of CHECC control students had either met or exceeded expectations in 2019, compared to 31.8% of students in the state. In English and Language Arts (ELA), 27.6% of CHECC control students either met or exceeded expectations, compared to 37.8% of students in the state.

Table 2.1. Summary Statistics: Parent Academy vs. Control

	Control	Parent Academy	Pre-K	1v2 (p-val)
Female	0.48	0.52	0.50	0.45
	(0.50)	(0.50)	(0.50)	
Race: Black	0.39	0.31	0.40	0.12
	(0.49)	(0.46)	(0.49)	
Race: Hispanic	0.52	0.59	0.49	0.22
	(0.50)	(0.49)	(0.50)	
Race: White	0.09	0.10	0.11	0.84
	(0.29)	(0.30)	(0.31)	
Mother Edu: Less than HS	0.19	0.24	0.19	0.39
	(0.40)	(0.43)	(0.40)	
Mother Edu: HS or GED	0.26	0.22	0.30	0.51
	(0.44)	(0.42)	(0.46)	
Mother Edu: Some Coll/2-yr Deg	0.37	0.29	0.34	0.15
	(0.48)	(0.45)	(0.48)	
Mother Edu: Bachelor's +	0.18	0.25	0.17	0.14
	(0.39)	(0.44)	(0.38)	
Income: $0-$16000	0.39	0.35	0.34	0.59
	(0.49)	(0.48)	(0.48)	
Income: $16000-$36000	0.32	0.38	0.37	0.38
	(0.47)	(0.49)	(0.49)	
Income: $36000+	0.29	0.27	0.29	0.73
	(0.46)	(0.45)	(0.46)	
Observations	202	136	92	338

Notes. This table provides summary statistics of our analysis sample by CHECC treatment group: Control, Parent Academy, and Pre-K. Column 5 displays *p* values to identify whether the demographic differences between Control and Parent Academy treatments are statistically significant.

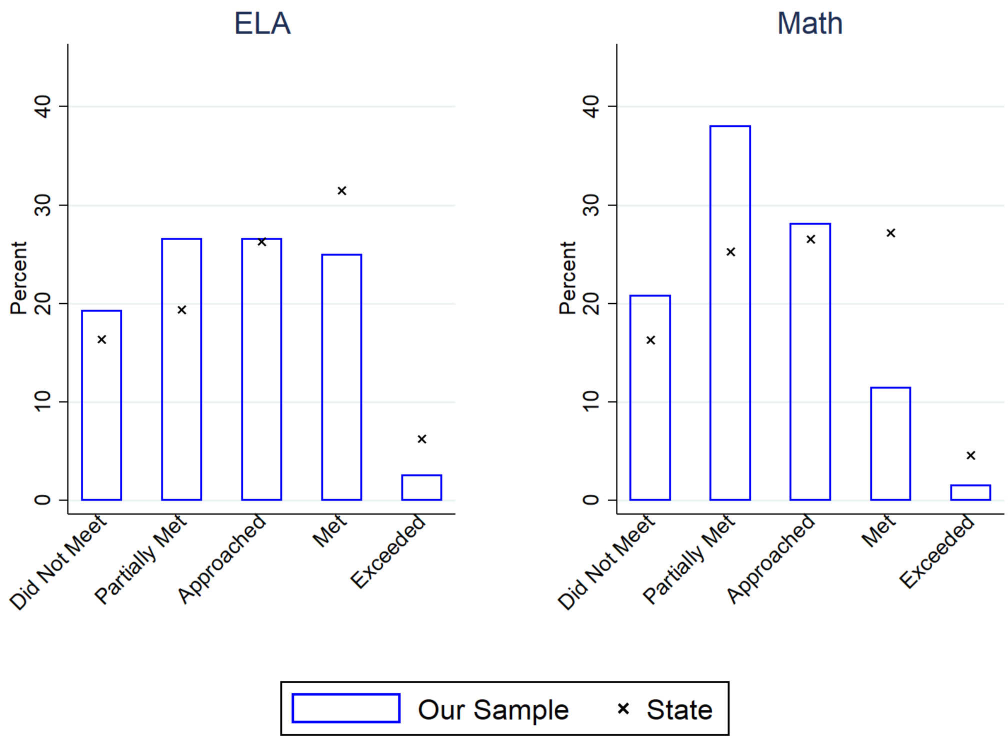

Figure 2.1. 2019 IAR Performance Levels: State vs. Sample

Notes. This figure shows the distribution of 2019 results on the Illinois Assessment of Readiness (IAR) among students in our sample (blue bars) and across Illinois (black Xs). Scale scores are categorized into 5 performance labels: "Did not Meet Expectations", "Partially Met Expectations", "Approached Expectations", "Met Expectations", and "Exceeded Expectations". The left panel depicts results on the English and Language Arts (ELA) exam, and the right panel depicts results on the Math exam.

During the post-Covid-19 year (2020-21), the students in the control group were in grades 6-8 and had an average IAR score of 714.7. This represents a 0.30 SD decrease in the score due to Covid-19 learning loss relative to pre Covid-19 test scores. By comparison, related work shows Covid-19 learning loss between 0.08 and 0.14 SD in the United States (Betthäuser et al., 2023) and 0.08 SD in the Netherlands (Engzell et al., 2021).

Covid-19 learning loss in our sample is substantially larger than the literature has found in other settings. This is likely due to the inequitable effects of Covid-19 on learning loss by race and socioeconomic characteristics of students (Goldhaber et al., 2022). Our sample is heavily

Black, Hispanic, and low-income relative to the US population. This is consistent with the demographic makeup of Chicago Heights and surrounding areas where CHECC took place.

2.4.2 Determinants of Covid-19 Learning Loss

In Table 2.2, we regress Covid-19 learning loss on treatment assignment, performance pre-Covid, and various demographic and socioeconomic characteristics. Within our sample, average learning loss is 6.7 points using the scale score with a standard deviation of 19 points. From this table, we first document several determinants of Covid-19 learning loss. Girls exhibited marginally significantly less learning loss compared to boys in our sample. Relative to Black students, Hispanic students also experienced lower levels of learning loss in our sample.

Mother education is also predictive of Covid-19 learning loss in our sample. Students with less educated mothers tended to experience greater learning loss than students with more educated mothers. Relative to students with mothers who did not have a high school degree, having a mother with a high school degree decreases Covid-19 learning loss by 0.23 standard deviations and having a mother with a Bachelor's degree decreases learning loss by 0.36 standard deviations. After controlling for mother education, we see no impacts of parental income on Covid-19 learning loss.

Students who had higher standardized test scores pre-Covid experienced greater learning loss during Covid-19 ($p = 0.005$). This pattern could arise through a few different pathways. Given our sample of students score on average lower than the general population in Illinois taking the IAR exam, we first consider the possibility that higher performing students experienced a greater drop in test scores than lower-performing students because lower-performing students bottomed-out. That is, perhaps lower-performing students' scores could not decrease much further, whereas higher-performing students had more space for downward movement.

To evaluate this explanation, we graph separately the distributions of pre-Covid IAR scores and post-Covid IAR scores within our sample. Suppose lower-performing students in 2019 had less space for their scores to decrease. If this were the case, then we would expect

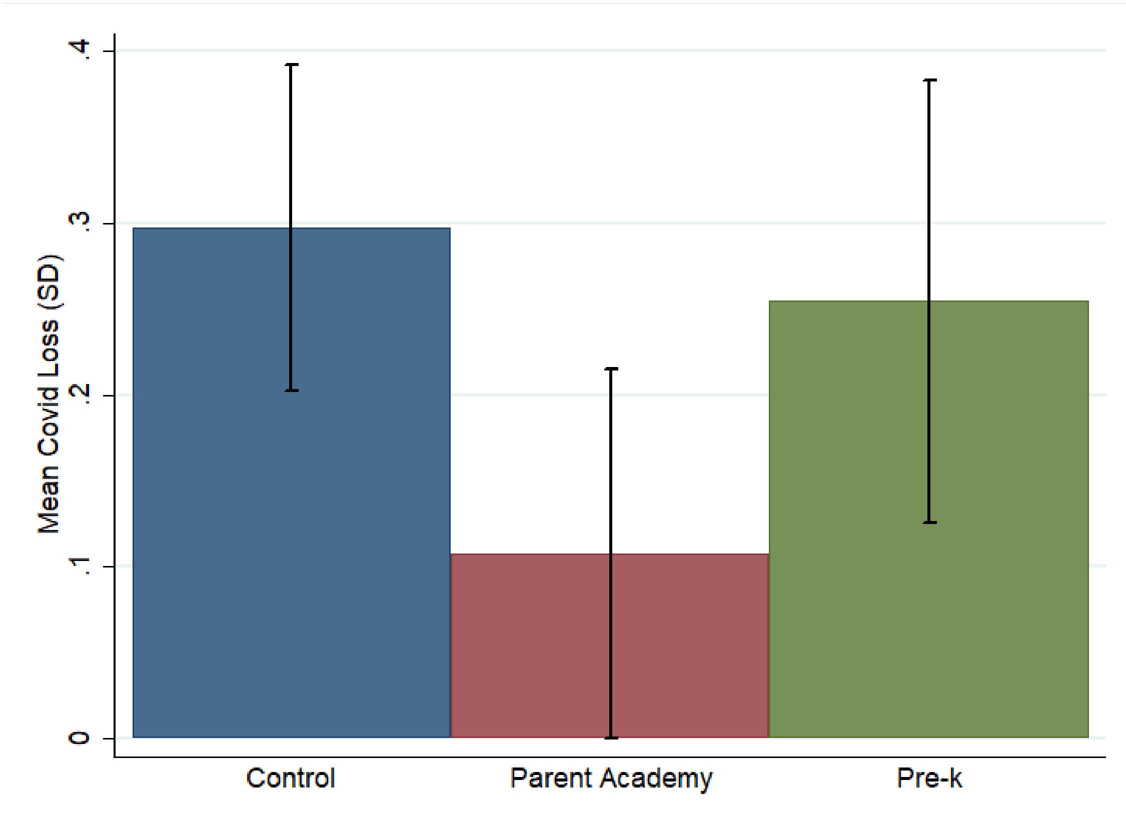

Figure 2.2. Treatment Effects on Covid-19 Learning Loss

Notes. This figure depicts average Covid-19 learning loss by treatment assignment in the Chicago Heights Early Childhood Center (CHECC) in 2010-2011. Covid-19 learning loss is calculated as the difference students' scores on the Illinois Assessment of Readiness (IAR) exam in 2021 (supplemented by 2022 scores if missing) and students' scores in 2019 (supplemented by 2018 scores if missing). Average learning loss is calculated in standard deviations of 2019 scores.

the post-Covid distribution to be shifted to the left for higher scores, and less shifted for lower scores. In Appendix Figure B.1, we see that this is not the case. IAR scores shifted left across the distribution post-Covid. The shift (if anything) appears greater for lower scores than for higher scores. So it does not seem to be the case that lower-performing students bottomed out prior to Covid-19.

Another explanation is that higher-performing students experienced greater learning loss due to mean reversion in test scores. Student test scores are a noisy signal of their true academic ability (Chay et al., 2005). That is, performance in any given period contains random errors such that some students' scores are higher than their true ability, and other students' scores are lower than their true ability. As a result of the randomness of these errors, students who experienced a positive error in one year are more likely to experience a test score drop the following year, and students who experienced a negative error are more likely to experience an increased test score the following year. This is referred to as mean reversion in test scores.

Mean reversion may explain the positive relationship between Covid-19 learning loss and pre-Covid test score. Following Chay et al. (2005), we construct a placebo test. Appendix Table B.1 contains two models. The first is our main model of Covid-19 learning loss restricted to students with scores in 2019 and 2021. The second model is our placebo test, in which we calculate learning loss between 2017 and 2019. For both models, we regress learning loss on a "Normalized Pre Score" (students' 2019 scores in the first model; students' 2017 scores in the second model). The relationship between learning loss and pre-score is stronger in the placebo model than in our Covid-19 learning loss model ($p = 0.006$). We conclude that the greater learning loss among students who scored higher in 2019 is likely a result of mean reversion. That is, high performing students did not experience a greater-than-usual test score drop in 2021.

2.4.3 Treatment Effects on Covid-19 Learning Loss

We now look at how treatment assignment in the CHECC program affected Covid-19 learning loss in our sample. Figure 2.2 shows average Covid-19 learning loss (the difference between post-Covid and pre-Covid standardized test scores) by treatment group. On average, the control group lost 0.30 standard deviations post Covid-19 relative to their pre Covid-19 test scores. Students assigned to the Parent Academy experienced a learning loss of 0.11 standard deviations, which is significantly smaller than the learning loss among the control group ($p = 0.01$). Students assigned to the Pre-K treatment experienced a Covid-19 learning loss of 0.25 standard deviations, which is not statistically significantly different from the learning loss of the control group ($p = 0.61$).

Table 2.2. Treatment Effects on Covid-19 Learning Loss

	Covid-19 Loss (SD)	Covid-19 Loss (SD)	Covid-19 Loss (SD)
Parent Academy	-0.20***	-0.14***	-0.13***
	(0.04)	(0.04)	(0.05)
Pre-k	-0.05	-0.02	-0.02
	(0.07)	(0.09)	(0.10)
Score Pre-Covid	0.11***	0.15***	0.16***
	(0.04)	(0.04)	(0.04)
Female		-0.09**	-0.08*
		(0.04)	(0.05)
Race: White		-0.04	-0.07
		(0.08)	(0.07)
Race: Hispanic		-0.18**	-0.21**
		(0.07)	(0.09)
Age		0.12	0.11
		(0.10)	(0.08)
Mother Edu: HS or GED			-0.16***
			(0.04)
Mother Edu: Some Coll/2-yr Deg			-0.10*
			(0.06)
Mother Edu: Bachelor's +			-0.25**
			(0.10)
Income: $16000-36000			-0.04
			(0.11)
Income: $36000+			0.11
			(0.10)
Constant	0.24	-1.24	-0.84
	(0.16)	(1.41)	(1.24)
Observations	430	430	430
Control Mean	0.30	0.30	0.30

Notes. We regress students' Covid-19 learning loss on CHECC treatment assignment, pre-Covid IAR test score, various demographics, and socioeconomic variables. Covid-19 learning loss is calculated as students' post-Covid (2021) IAR score subtracted from their pre-Covid (2019) IAR score. Pre-Covid scores are supplemented by 2018 scores if 2019 scores are not available, and post-Covid scores are supplemented by 2022 scores if 2021 scores are not available. Learning loss is measured in standard deviations relative to 2019 scores. All models control for CHECC participation year and IAR exam year. Models 2 and 3 also control for baseline cognitive and executive function scores during CHECC participation. Standard errors clustered by district are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

As Table 2.2 shows, Covid-19 learning loss is partially mitigated by prior assignment to Parent Academy. Assignment to Parent Academy decreases learning loss by 0.13 standard deviations after controlling for student and parent-level demographic information. This effect size is approximately half of the effect size of having a mother with a Bachelor's degree compared to having a mother without a high school degree. Prior assignment to Pre-K, on the other hand, had no detectable impact on Covid-19 learning loss. In Appendix Table B.2, we also use inverse probability weights (IPW) to account for socio-demographic and performance differences between our analysis sample and the sample available for analysis.

As discussed in Section 1, all Parent Academy families received participation incentives for attending workshops and performance incentives for completing homework assignments and for their child's scores on assessments. Half of Parent Academy families were randomized to receive performance incentives immediately in cash (Parent Academy-Cash), whereas the other half were randomized to receive these incentives in a savings account that would become accessible if and when the child attended post-secondary education years later (Parent Academy-College).

We test whether the impact of Parent Academy on Covid-19 learning loss differed based on incentive scheme by regressing learning loss on assignment to Parent Academy-Cash and Parent Academy-College separately. Because the students in our sample are in middle school, we may expect the college incentives to be more relevant than the cash incentives because the college incentives are only obtainable if the child attends post-secondary education, while the cash incentives were already distributed years prior. As a result, the college incentives give parents and children an added incentive to perform well academically, while cash incentives do not.

Table 2.3. Cash vs College: Treatment Effects on Covid-19 Learning Loss

	Covid-19 Loss (SD)	Covid-19 Loss (SD)	Covid-19 Loss (SD)
Cash	-0.16**	-0.10	-0.07
	(0.06)	(0.07)	(0.07)
College	-0.24***	-0.18***	-0.18***
	(0.05)	(0.05)	(0.05)
Pre-k	-0.05	-0.02	-0.01
	(0.07)	(0.09)	(0.10)
Normalized Score Pre-Covid	0.11***	0.15***	0.16***
	(0.04)	(0.04)	(0.04)
Female		-0.09**	-0.08*
		(0.04)	(0.04)
Race: White		-0.04	-0.08
		(0.08)	(0.07)
Race: Hispanic		-0.18**	-0.20**
		(0.07)	(0.09)
Age		0.13	0.12
		(0.10)	(0.08)
Mother Edu: HS or GED			-0.16***
			(0.04)
Mother Edu: Some Coll/2-yr Deg			-0.10
			(0.06)
Mother Edu: Bachelor's +			-0.25**
			(0.11)
Income: $16000-$36000			-0.04
			(0.11)
Income: $36000+			0.11
			(0.10)
Constant	0.25	-1.39	-1.00
	(0.15)	(1.43)	(1.25)
Observations	430	430	430

Notes. We regress Covid-19 learning loss on CHECC treatment assignment, pre-Covid IAR score, demographics, and socioeconomic variables. "Cash" and "College" refer to Parent Academy subtreatments in which parents were incentivized either in cash or secondary education funding. Covid-19 learning loss is calculated as post-Covid (2021) IAR score subtracted from pre-Covid (2019) IAR score. Pre-Covid scores are supplemented by 2018 scores if 2019 scores are not available. Post-Covid scores are supplemented by 2022 scores if 2021 scores are not available. Learning loss is measured in 2019 standard deviations. All models control for CHECC participation year and IAR exam year. Models 2 and 3 control for baseline cognitive and executive functioning scores during CHECC. Standard errors clustered by district are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

As reported in Table 2.3, assignment to both Parent Academy-Cash and Parent Academy-College treatments result in less learning loss. After controlling for demographics and socioeconomic indicators, the effect of Parent Academy-Cash remains negative, but is no longer significant, whereas the effect of Parent Academy-College remains negative highly significant. This evidence suggests that the added incentives at stake in the Parent Academy-College treatment increased the effects of the Parent Academy treatment on learning loss.

Given the beneficial impacts of assignment to Parent Academy on Covid-19 learning loss, a natural follow-up question is how Parent Academy affected other outcomes prior to the onset of Covid-19. In Table 2.4, we document how Parent Academy affected other outcomes of interest within our sample of students. To maximize power, we include students who were eligible to be in our sample, but were not included because they did not have Covid-19 learning loss scores. In Columns 1 and 2, we regress students' cognitive and non-cognitive scores (respectively) measured at the end of the CHECC interventions on treatment assignment and controls. We see that Parent Academy assignment did not have significant effects on cognitive scores, but did have a positive and significant effect on non-cognitive (or, executive functions) scores. This replicates the finding from (Fryer et al., 2015). In Column 3, we regress students' IAR test scores in grades 3-8 prior to 2020 on treatment assignment and controls. Here we see that Parent Academy did not significantly affect test scores prior to Covid-19. In the last model, we regress an indicator for whether a student was ever disciplined in school on treatment assignment and controls. Again, we find no impact of assignment to Parent Academy

Table 2.4. Treatment Effect of Parent Academy on other outcomes

	Cog Score	Non-Cog Score	Test Scores	Ever Disciplined
Parent Academy	0.03	0.12	-0.13	-0.51*
	(0.11)	(0.14)	(0.10)	(0.31)
Pre-k	0.20*	0.09	0.01	-0.20
	(0.11)	(0.15)	(0.10)	(0.28)
Female	-0.00	-0.08	0.13*	-0.61***
	(0.08)	(0.11)	(0.08)	(0.23)
Race: White	0.11	0.33*	0.34**	-1.88***
	(0.16)	(0.17)	(0.13)	(0.47)
Race: Hispanic	-0.10	0.47***	0.23**	-0.38
	(0.10)	(0.15)	(0.09)	(0.25)
Mother Edu: Bachelor's +	0.32*	0.47*	0.62***	0.48
	(0.17)	(0.27)	(0.16)	(0.54)
Mother Edu: HS or GED	0.11	0.28*	0.06	0.58
	(0.11)	(0.17)	(0.12)	(0.36)
Mother Edu: Some Coll/2-year Deg	0.23*	0.15	0.22*	0.01
	(0.13)	(0.20)	(0.12)	(0.40)
Income: $16000-%36000	0.11	-0.13	0.13	-0.37
	(0.11)	(0.15)	(0.11)	(0.31)
Income: $36000+	0.35**	0.02	0.24*	0.20
	(0.14)	(0.22)	(0.14)	(0.37)
Constant	-0.62**	-0.59*	-1.35	-2.43
	(0.25)	(0.34)	(1.19)	(4.70)
Observations	244	242	1473	259
Children	244	242	431	144

Notes. In this table, we regress various outcomes on CHECC treatment assignment, demographics, and socioeconomic variables. The outcome variables in Models 1 and 2 are cognitive score and non-cognitive score at the end of students' CHECC participation, respectively. The outcome variable in Model 3 is student IAR scale scores in grades 3-8. The outcome variable in Model 4 is an indicator for each school year (grade 3-8) of whether the child received any disciplinary referral. All models restrict to our sample of students with Covid-19 learning loss scores. All models control for CHECC participation year, cognitive score at CHECC baseline, non-cognitive score at CHECC baseline, and age. Models 3 and 4 also control for exam year and grade level. Robust standard errors are in parentheses in Models 1 and 2. Standard errors clustered by student are in parentheses in Models 3 and 4. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

To better understand how Parent Academy helped students academically during Covid-19 without affecting academic performance prior to Covid-19, we return to our conceptual framework from Section 2.2. As a reminder, we model child skill development as a function of the child's skill in the prior period, child characteristics, and inputs from the child, parent, and school.

When schools shut down in response to Covid-19, schools could no longer provide the same level of inputs into the child skill production function. That is, school inputs experienced a negative shock. Parents then chose their level of investment into the child production function, given child characteristics, child skill in the prior period, and inputs from the school and child. Holding all else equal, child skills will be lower than they would have been if not for the drop in school inputs. Due to the substitutability of inputs in our framework, parents can help offset the effect of the drop in school inputs on child skills by increasing their inputs.

Qualitative evidence from surveys to parents during Covid-19 suggest that parents substantially increased their time and effort in teaching their children as a result of school closures during Covid-19 (Bansak and Starr, 2021). Because parents had a larger role in their child's production function, differences in the quantity and quality of parent inputs became more apparent than before. Parent Academy emphasized to parents the importance of investing in their child and provided them with the tools to make those investments effective. While we may have expected positive effects from Parent Academy to show up in student test scores and behavioral outcomes prior to Covid-19, the effects of school inputs on child skills likely masked the effects of parent inputs. That is, parents may have relied more on school inputs prior to Covid-19, such that the differential impacts of parent inputs were not detectable until schools closed. If so, it is not surprising that the effects of Parent Academy did not become visible until after the onset of school closures.

2.5 Discussion and Conclusion

In this paper, we explore the long-run effects of an early childhood intervention on child academic success during the Covid-19 pandemic. We leverage data from the CHECC field experiment conducted in 2010-2014 in which households with children ages 3-5 years old were randomized to a parenting program ("Parent Academy"), a preschool program ("Pre-K"), or a control group for 1-2 years. Ten years later, children were exposed to the Covid-19 school shutdowns. We link data from the field experiment to standardized Illinois state exams to evaluate the causal impacts of these early childhood interventions on student test scores during Covid-19.

We first document the extent of Covid-19 learning loss in our sample of students. Among the CHECC control group, students in our sample experienced a 0.30 standard deviation drop in test scores after the Covid-19 school closures. This is larger than estimates from the literature of 0.10 to 0.17 standard deviations (Betthäuser et al., 2023). The larger learning loss experienced by our sample is likely due to the inequitable impacts of Covid-19 on learning loss by race and socioeconomic status (Goldhaber et al., 2022).

In our sample, we find various demographic predictors of Covid-19 learning loss. We find suggestive evidence that girls may have experienced less learning loss than boys. This contrasts with evidence from Engzell et al. (2021) and Orlov et al. (2021) which found no gender differences in student learning in the Netherlands and among a sample of US college students, respectively. This also contrasts with evidence from Wu et al. (2022), which found that girls experienced greater learning loss during Covid-19 than boys in North America. We also find that relative to Black students, Hispanic students experience lower levels of learning loss in our sample. Our finding is consistent with Sass and Ali (2022) and Lewis and Kuhfeld (2021) in that Black students seem to experience the most Covid-19 learning loss. We do not replicate the finding that Black students experienced less learning loss than White students, potentially due to our small sample of White students. Evidence from Goldhaber et al. (2022) suggests that within schools, there are no learning loss differences by race. We also find that students with more highly-educated mothers experienced less learning loss. In particular, having a mother with a high school degree or GED was associated with less learning loss relative to having a mother without a high school degree or GED. Engzell et al. (2021) find that having a parent with low levels of educational attainment is associated with an increase of Covid-19 learning loss by 40% in the Netherlands. We find an even greater impact in that having a mother with a high school degree is associated with a 53% less Covid-19 learning loss compared to having a mother without a high school degree.

We compare Covid-19 learning loss among the three CHECC treatment groups run in 2010 and 2011: control, Pre-K, and Parent Academy. Students who were randomized to the Pre-K treatment did not experience a significant difference in learning loss compared to the control group. Students who were assigned to the Parent Academy, on the other hand, experienced a drop of 0.11 standard deviations during this period, which is statistically significantly smaller than the learning loss of the control group. That is, we find that random assignment to Parent Academy during early childhood substantially mitigated learning loss among children ten years later during the Covid-19 school shutdowns.

Our findings suggest that investments in parents may have positive long-term impacts on children academically. Much of the literature on early childhood interventions finds immediate effects on child outcomes that then fade out (Almond et al., 2018). Some papers find evidence that the effects of interventions then reappear later in life (Heckman et al., 2013). We find that the effects of our intervention appear during middle school - a stage in life where early childhood interventions often disappear (Almond et al., 2018).

Our findings also provide evidence about the role of parents in education production function of children. Prior to the Covid-19 school closures, we did not detect differences in child academic performance from the Parent Academy treatment. When Covid-19 shut schools down, however, children whose parents were assigned to the parent training intervention experienced less learning loss. Together, these findings suggest that the role of parents in the education production function was partially offset by schools. When schools were forced to shut down, parents had a larger impact in the education production function.

Chapter 3

Robustness of Rank Independence in Risky Choice with B. Douglas Bernheim and Charles Sprenger

The famous Allais (1953) paradoxes challenge the validity of the independence axiom that lies at the heart of Expected Utility Theory (EUT). As originally formulated, Prospect Theory (OPT, due to Kahneman and Tversky 1979) rationalizes such phenomena by allowing decision weights to vary non-linearly with probabilities, but thereby introduces violations of first-order stochastic dominance. Cumulative Prospect Theory (CPT, due to Tversky and Kahneman 1992) avoids this difficulty by assuming that the decision weight associated with an outcome depends not only on its probability, but also on its rank.

Bernheim and Sprenger (2020) present a novel experimental test of CPT. Specifically, for a given three-outcome lottery $L = (X, Y, Z; p, q, 1-p-q)$ with $Y > Z$, they measure discrete analogs of marginal rates of substitution between Z and Y (*equalizing reductions*). With rank-dependent probability weighting, the tradeoff between Y and Z depends upon whether $X \geq Y > Z, Y > X \geq Z$, or $Y > Z > X$. As it turns out, the percentage change in the equalizing reduction as X crosses from one of these regimes to another identifies the degree of rank dependence in probability weights, and consequently the shape of the CPT probability weighting function, non-parametrically. Bernheim and Sprenger present two main findings based on variation in equalizing reductions for a single set of tasks.

Finding 1: *The probability weights for outcomes Y and Z do not vary meaningfully with* *their ranks (comparing cases with $X \geq Y > Z$ to cases with $Y > X \geq Z$).*

Finding 2: *Responses to variation in probabilities with fixed ranks imply probability weights that match standard estimates based on experiments with two-outcome lotteries.*

Unlike OPT, CPT posits a form of rank-dependent probability weighting that makes these findings irreconcilable. Interpreted through the lens of CPT, Finding 1 implies that the probability weighting function must be linear over the pertinent range. But Finding 2 implies that the probability weighting function is highly non-linear over the very same range.

Since the publication of Bernheim and Sprenger (2020), some proponents of CPT have raised the possibility that Finding 1 may be a consequence of experimental procedures rather than underlying preferences. Possible concerns include the following: 1) comprehension of the three-option lotteries may have been poor; 2) the number of decision tasks may have been overwhelming; 3) the stakes may have been too low; 4) the analysis may have overlooked evidence of rank dependence associated with transitions between the regimes $Y > X \geq Z$ and $Y > Z > X$, which shed light on non-linearities in probability weighting for probabilities near 1, and 5) the structure of the decision tasks may have triggered a heuristic involving the "cancellation" of a common outcome (X).[1]

It is unlikely that the aforementioned considerations account for Bernheim and Sprenger's results. Concerns 1-3 – complexity, fatigue, and low stakes – are not specifically entwined with rank dependence; they would tend to suppress any nuanced feature of decision making. But the setting is not so complex, fatiguing, or inconsequential that it fails to activate conventional probability weighting (Finding 2). Bernheim and Sprenger also addressed Concern 2 (fatigue) by presenting corroborating cross-subject results based on each subject's first task. Concern 4 merely raises the possibility that Finding 1 might not hold on an unexamined portion of the probability domain; it cannot resolve the conflict that CPT implies between Findings 1 and 2 on the examined portion of that domain. Finally, Bernheim and Sprenger address Concern 5

[1] In their critique of Bernheim and Sprenger (2020), Abdellaoui et al. (2020) raise these issues and make a number of other points. See Bernheim and Sprenger (2022) for our full response.

through a supplemental experiment with modified tasks that render the cancellation heuristic inapplicable. The supplemental experiment sacrifices some advantages of the original, but still yields no evidence of rank-dependent probability weighting.

In this paper, we demonstrate that Finding 1 is indeed robust with respect to alternative experimental procedures that address each of the of five concerns articulated above. Naturally, a comprehensive evaluation of CPT must consider other evidence concerning its validity. However, other existing tests suffer from serious confounds that the Bernheim-Sprenger approach avoids; see Bernheim and Sprenger (2020, 2022).[2]

3.1 Review and Extension of Methods

Regardless of whether the applicable theory is CPT, OPT, or Expected Utility Theory (EUT), we can write the indifference condition that defines the equalizing reduction for the lottery $L = (X,Y,Z;p,q,1-p-q)$ as follows:

$$w_X u(X) + w_Y u(Y) + w_Z u(Z) = w_X u(X) + w_Y u(Y+m) + w_Z u(Z-k),$$

where w_s is the decision weight for $s \in \{X,Y,Z\}$. For small m, it follows that $\frac{k}{m} \approx \frac{w_Y u'(Y)}{w_Z u'(Z)}$. For any change from X'' to X', the associated k'' and k' therefore satisfy

$$\log(k') - \log(k'') \approx \log\left(\frac{w'_Y}{w'_Z}\right) - \log\left(\frac{w''_Y}{w''_Z}\right).$$

Thus, the percentage change in the equalizing reduction non-parametrically measures the percentage change in relative decision weights resulting from a change in X.

To determine whether the weights are rank-dependent, we choose values X' and X'' so

[2] This approach improves significantly upon prior tests of Comonotonic and Non-Comonotonic Independence (see, e.g., Birnbaum and McIntosh, 1996; Wu, 1994; Wakker et al., 1994) by neutralizing important confounds and providing quantitative non-parametric measures of non-linearities in the probability weighting function.

that the ranks of the outcomes differ. For $X'' > Y > Z$ and $Y > X' > Z$, CPT implies

$$\log(k') - \log(k'') \approx \log\left(\frac{\pi(q)}{q}\right) - \log\left(\frac{\pi(p+q) - \pi(p)}{q}\right).$$

In other words, under the maintained hypothesis of CPT, the percentage change in the equalizing reduction non-parametrically measures the percentage change in the slope of the probability weighting function between the intervals $[0,q]$ and $[p, p+q]$.

Bernheim and Sprenger (2020) found essentially no difference in equalizing reductions between the regimes $X'' > Y > Z$ and $Y > X' > Z$ for $p \in \{0.1, 0.4, 0.6\}$ (with $q = 0.3$), which means there is no evidence of rank-dependent probability weighting (Finding 1). Treating CPT as a maintained hypothesis, they concluded that the probability weighting function must be linear over the interval $[0, 0.9]$.

If CPT is valid, then one can also recover the probability weighting function by fixing ranks and varying probabilities. Defining $\phi = \left(\frac{q}{1-p-q}\right)\frac{m}{k}$, and using the approximation $\frac{k}{m} \approx \frac{w_Y u'(Y)}{w_Z u'(Z)}$ along with the definitions of w_Y and w_Z for CPT within the regime $Y > X > Z$, we see that, for any change in p, say from p' to p,

$$\log(\phi) - \log(\phi') \approx \log\left(\frac{\pi(1) - \pi(p+q)}{1-p-q}\right) - \log\left(\frac{\pi(1) - \pi(p'+q)}{1-p'-q}\right).$$

In other words, the percentage change in ϕ (which is measurable) is a non-parametric estimate of the percentage change in the average slope of the probability weighting function between the intervals $[1-p'-q, 1]$ and $[1-p-q, 1]$.

Bernheim and Sprenger (2020) found large differences in ϕ for $p \in \{0.1, 0.4, 0.6\}$ (with $q = 0.3$), which means there is evidence of substantial non-linearities in probability weighting. In particular, their estimates imply that the probability weighting function is highly non-linear throughout the interval $[0.4, 1]$ (Finding 2).

Treating CPT as a maintained hypothesis, Finding 1 and Finding 2 clearly have contradictory implications for the properties of the probability weighting function over the interval $[0.4, 0.9]$. Bernheim and Sprenger (2020) therefore reject CPT. Their findings are instead consistent with non-linear rank-independent probability weighting.

We extend these methods in two ways. First, we also examine changes from X' to X satisfying $Y > X' > Z$ and $Y > Z > X$. For CPT, we have

$$\log(k) - \log(k') \approx \log\left(\frac{\pi(1) - \pi(p+q)}{1-p-q}\right) - \log\left(\frac{\pi(1-p) - \pi(q)}{1-p-q}\right).$$

In other words, under the maintained hypothesis of CPT, the percentage change in the equalizing reduction between the regimes $Y > X' > Z$ and $Y > Z > X$ non-parametrically measures the percentage change in the slope of the probability weighting function between the intervals $[q, 1-p]$ and $[p+q, 1]$. Thus, CPT can account for invariance of equalizing reductions across all three regimes only if the slope of the probability weighting function is constant (i.e., the function is linear) over the entire interval $[0, 1]$.

Second, we modify the method by eliciting k_+ and k_-, defined as follows:

$$\begin{aligned}
w_X u(X-1) + w_Y u(Y) + w_Z u(Z) &= w_X u(X) + w_Y u(Y+m) + w_Z u(Z-k_+) \\
w_X u(X) + w_Y u(Y) + w_Z u(Z) &= w_X u(X-1) + w_Y u(Y+m) + w_Z u(Z-k_-)
\end{aligned}$$

For small m, we have $\frac{0.5(k_+ + k_-)}{m} \approx \frac{w_Y u'(Y)}{w_Z u'(Z)}$, so k and $k_M \equiv 0.5(k_+ + k_-)$ measure the same theoretical object. Intuitively, substituting k_M for k immunizes our method against the cancellation heuristic because the lotteries that define k_+ and k_- involve no common outcomes. In contrast to the supplemental experiment in Bernheim and Sprenger (2020), this method preserves all quantitative inferences concerning rank dependence.

3.2 Experimental Design

Our seven conditions all measure equalizing reductions (k) and modified equalizing reductions (k_+ and k_-) for the probability vector $(p, q, 1-p-q) = (0.4, 0.3, 0.3)$. Table 3.1 summarizes the main features of each condition; the appendix includes screenshots. We conducted the experiment in September 2021 on the Prolific platform using Otree software (Chen et al., 2016).

Condition 1 follows Bernheim and Sprenger (2020) in fixing $Y = \$24$, $Z = \$18$, and $m = \$5$. We used price lists to elicit equalizing reductions and modified equalizing reductions in random order for $X = \$31$, $X' = \$22$, and $X'' = \$3$. We display lottery distributions visually using the method of Lopes and Oden (1999). Subjects also receive training to facilitate their comprehension of the probabilities: they draw from the distribution 18 times and report their outcomes. This procedure allows the meaning of the probability distribution to 'sink in' (Heffetz, 2018). Thus, Condition 1 addresses Concern 1 (comprehension of probabilities) through visual presentation and training, Concern 4 (limited scope) by encompassing all three regimes, and Concern 5 (heuristic cancellation) by eliciting modified equalizing reductions.

Condition 2 is the same as Condition 1 except that it employs the titration-BDM mechanism of Mazar et al. (2014), wherein subjects first state a valuation, then review implications for options just below and just above the provisional point of indifference, then (potentially) revise their initial response. This procedure improves upon the original BDM mechanism by walking subjects through the contingent implications of their choices. It creates the same incentives as the corresponding price lists of Condition 1, but subjects make only nine decisions rather than 585 component choices. This condition therefore addresses Concern 2 (decision fatigue) in addition to Concerns 1, 4, and 5.

The remaining five conditions follow the same procedures as Condition 2, but inflate X, Y and Z by a factor of four (Conditions 3, 4, and 5) or 16 (Conditions 6 and 7). The value of m is \$5 in Conditions 3, 4, and 6, \$20 in Condition 5, and \$80 in Condition 7. In other words,

Table 3.1. Conditions for Measuring Equalizing Reductions

	Elicitation	Cancellation Tasks	(X, X', X'')	(Y, Z)	Stakes Multiplier	m	Incentivized	Order	# of Subjects
Condition 1	Price List	Yes	(3, 22, 31)	(24, 18)	1x	5	1/5	Random	93
Condition 2	BDM	Yes	(3, 22, 31)	(24, 18)	1x	5	1/5	Random	109
Condition 3	BDM	Yes	(12, 88, 124)	(96, 72)	4x	5	1/20	Random	103
Condition 4	BDM	Yes	(12, 88, 124)	(96, 72)	4x	5	None	Random	111
Condition 5	BDM	Yes	(12, 88, 124)	(96, 72)	4x	20	1/20	Random	89
Condition 6	BDM	Yes	(48, 352, 496)	(384, 288)	16x	5	None	Random	89
Condition 7	BDM	Yes	(48, 352, 496)	(384, 288)	16x	80	None	Random	103

Notes: All conditions included visual displays and probability training.

we either inflate m by the same factor as the outcomes or leave it fixed at \$5. These conditions address Concern 3 (small stakes).

Conditions 1, 2, 3, and 5 involve real choices. We paid one out of every five subjects based on one of their choices in Conditions 1 and 2, and one out of every 20 subjects in Conditions 3 and 5. Conditions 4, 6, and 7 involve hypothetical choices. This variation provides additional opportunities to test whether incentives induce rank-dependent behavior.

Our procedures prevent subjects' choices from switching back and forth between (X,Y,Z) and $(X,Y+m,Z-k)$ as k increases. This restriction has the advantage of yielding an unambiguous measure of k, and (for Condition 1) of reducing each price list to a single choice, thereby minimizing decision fatigue. A disadvantage is that it sacrifices a potential indicator of poor comprehension (multiple switching). An alternate measure is whether the elicitations yield boundary values. Overall, 2.3% (13.9%) of observations take on the highest (lowest) value. Only 4.6% of subjects provide no interior values, and 80.2% provide two or fewer boundary values. Dropping these responses does not meaningfully change our findings.

3.3 Results

Table 3.2 and Figure 3.1 present our results. Condition 1 replicates the findings of Bernheim and Sprenger (2020). In both cases, the estimated change in decision weights between the regimes $Z < X' < Y$ and $Z < Y < X''$, $\log(k') - \log(k'')$, is close to zero. As before, we fail to reject the null hypothesis of rank independence (i.e., equality between k' and k'').

Condition 1 also extends the prior investigation by examining tasks with $X < Z < Y$. The differences between k, k' and k'' are small (on the order of one to two percent) and statistically insignificant, indicating the virtual absence of rank dependence. Treating CPT as a maintained hypothesis, one would conclude that the average slope of the probability weighting function is essentially unchanged between the intervals $[0, 0.3]$ and $[0.4, 0.7]$, as well as between the intervals $[0.3, 0.6]$ and $[0.7, 1]$.

Results based on k_+ and k_-, which are immune to the cancellation heuristic, corroborate the (near) rank independence of probability weights. Log differences in equalizing reductions imply that probability weights change only slightly (by one to four percent) due to a change in ranks. Critically, this finding does not reflect a tendency to cancel *approximately* common outcomes – i.e., to ignore the difference between X and $X-1$. As shown in Figure 3.1, values of k_+ are generally higher than values of k_-, and the difference is statistically significant ($\chi^2 = 20.94; p = 0.000$). However, the average of k_+ and k_- is statistically indistinguishable from k ($\chi^2 = 1.46; p = 0.228$), which suggests that cancellation is unimportant.

To put the preceding findings in context, Figure 3.1 also displays predicted values of k derived from the parameterized version of CPT due to Tversky and Kahneman (1992). According to this model, equalizing reductions should exhibit discontinuous increases across regimes (moving from $X'' > Y > Z$ to $Y > X' > Z$ to $Y > Z > X$) on the order of 66-176 percent.[3] Subjects display far less sensitivity to ranks than predicted. It is worth emphasizing that Bernheim and Sprenger (2020) found similar patterns of non-linear probability weighting based on fixed-rank variation in probabilities for equalizing reduction tasks.

We similarly find no evidence of significant rank dependence in Condition 2, which aims to reduce decision fatigue by substituting a BDM mechanism for price lists. For Conditions 3 through 7, which vary the size of stakes and the nature of incentives, we generally find that the actual changes in relative probability weights are less than 10%, compared with the CPT prediction of 66-176%. Out of the 42 comparisons in "Rank Dependence" portion of Table 3.2, we reject the null hypothesis of rank-independence at the 5 percent level in only three cases. In two of these three, the actual change in the equalizing reduction and the CPT prediction have opposite signs. We also provide an omnibus measure of rank dependence aggregating all of our conditions, and fail to reject the null hypothesis of rank independence overall. Thus, we confirm that Finding 1 of Bernheim and Sprenger (2020) is robust.

[3]The parameterization assumes that $u(x) = x^{0.88}$ and $\pi(p) = p^{0.61}/(p^{0.61} + (1-p)^{0.61})^{1/0.61}$. The changes in equalizing reductions (107%, 173%, and 66% for the three comparisons in the "Rank Dependence" portion of Table 3.2) closely approximate the changes in probability weights (110%, 176%, and 66%).

Chapter 3, in full, is a reprint of the material as it appears in the American Economic Association Papers and Proceedings. Bernheim, B. Douglas, Rebecca Royer, and Charles Sprenger. 2022. ”Robustness of Rank Independence in Risky Choice.” AEA Papers and Proceedings, 112: 415-20. The dissertation author was a primary investigator and author of this paper.

Table 3.2. Mean Equalizing Reductions and Estimated Rank Dependence

	Mean Equalizing Reduction			Rank Dependence		
	Equalizing Reduction - k $X < Z < Y$	Equalizing Reduction - k' $Z < X' < Y$	Equalizing Reduction - k'' $Z < Y < X''$	$log(k) - log(k')$	$log(k) - log(k'')$	$log(k') - log(k'')$
Bernheim Sprenger (2020) $(p,q,1-p-q) = (0.4,0.3,0.3)$, $(Y,Z) = (24,18)$		4.32 (0.12)	4.34 (0.12)			-0.01 (0.02)
Condition 1 (Price List, 1x stakes, $m = \$5$)	4.25 (0.29)	4.27 (0.28)	4.35 (0.26)	-0.00 (0.06)	-0.02 (0.05)	-0.02 (0.05)
Cancelation Tasks: $0.5(k_+ + k_-)$	4.05 (0.30)	4.21 (0.31)	4.29 (0.30)	-0.04 (0.06)	-0.06 (0.06)	-0.02 (0.07)
Condition 2 (BDM, 1x stakes, $m = \$5$)	3.50 (0.36)	3.57 (0.37)	3.67 (0.39)	-0.02 (0.07)	-0.05 (0.07)	-0.03 (0.06)
Cancelation Tasks: $0.5(k_+ + k_-)$	3.53 (0.37)	3.60 (0.36)	3.55 (0.41)	-0.02 (0.08)	-0.00 (0.08)	0.02 (0.08)
Condition 3 (BDM, 4x stakes, $m = \$5$)	5.17 (0.35)	5.44 (0.39)	4.99 (0.33)	-0.05 (0.06)	0.04 (0.05)	0.09 (0.06)
Cancelation Tasks: $0.5(k_+ + k_-)$	5.19 (0.39)	5.34 (0.43)	4.93 (0.37)	-0.03 (0.08)	0.05 (0.07)	0.08 (0.07)
Condition 4 (BDM, 4x stakes, $m = \$5$, hyp.)	4.67 (0.39)	4.39 (0.34)	4.32 (0.41)	0.06 (0.07)	0.08 (0.09)	0.02 (0.08)
Cancelation Tasks: $0.5(k_+ + k_-)$	4.56 (0.40)	4.47 (0.37)	4.26 (0.44)	0.02 (0.08)	0.07 (0.10)	0.05 (0.09)
Condition 5† (BDM, 4x stakes, $m = \$20$)	3.42 (0.37)	3.52 (0.39)	4.07 (0.40)	-0.03 (0.10)	-0.17 (0.09)	-0.15* (0.07)
Cancelation Tasks: $0.5(k_+ + k_-)$	3.25 (0.38)	3.50 (0.42)	4.11 (0.41)	-0.07 (0.13)	-0.23* (0.11)	-0.16 (0.09)
Condition 6 (BDM, 16x stakes, $m = \$5$, hyp.)	5.37 (0.47)	4.50 (0.46)	4.88 (0.43)	0.18* (0.09)	0.09 (0.08)	-0.08 (0.08)
Cancelation Tasks: $0.5(k_+ + k_-)$	5.14 (0.53)	4.59 (0.49)	4.75 (0.44)	0.11 (0.11)	0.08 (0.08)	-0.03 (0.10)
Condition 7† (BDM, 16x stakes, $m = \$80$, hyp.)	4.85 (0.49)	4.55 (0.44)	4.44 (0.48)	0.06 (0.06)	0.09 (0.08)	0.03 (0.06)
Cancelation Tasks: $0.5(k_+ + k_-)$	4.88 (0.50)	4.60 (0.45)	4.37 (0.47)	0.06 (0.06)	0.11 (0.08)	0.05 (0.06)
Aggregate Values†	4.47 (0.15)	4.34 (0.15)	4.39 (0.15)	0.03 (0.03)	0.02 (0.03)	-0.01 (0.03)
Cancelation Tasks: $0.5(k_+ + k_-)$	4.38 (0.16)	4.34 (0.15)	4.32 (0.16)	0.01 (0.03)	0.01 (0.03)	0.01 (0.03)

Notes: Mean values of k estimated from interval regressions of experimental response on indicators for the rank of X, with titration-BDM data converted to equivalent price list responses. † indicates normalized values, $k \cdot (5/m)$, for ease of comparison across m values. * indicates 5% significance. Rows titled "Cancelation Tasks" are based only on modified equalizing reductions. Standard errors, in parentheses, clustered at individual level and calculated using the delta method.

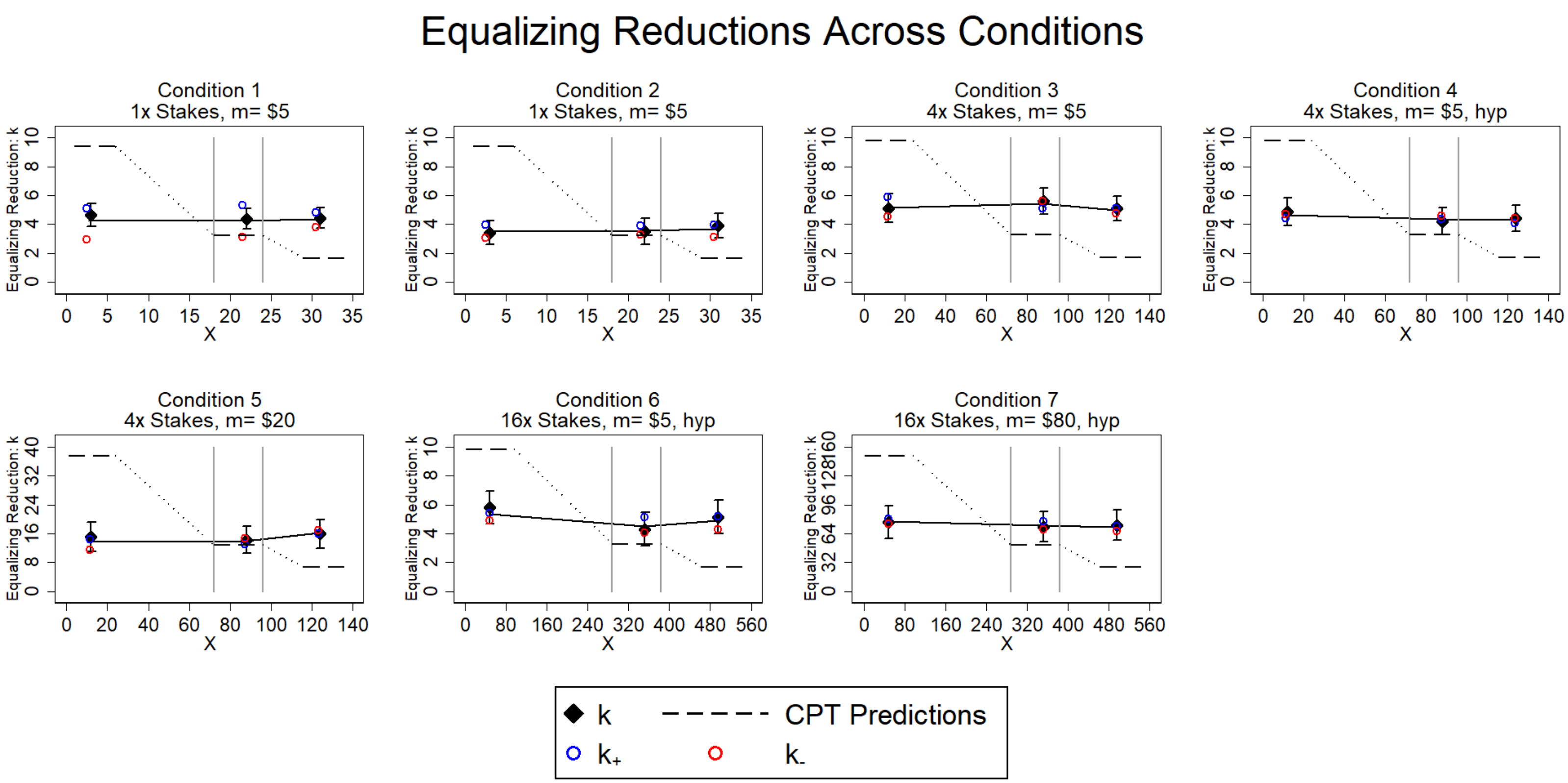

Figure 3.1. Mean Equalizing Reductions and Standard CPT Predictions

Notes: Black diamonds (with corresponding 95% CIs) indicate mean values for k. Blue (red) circles indicate mean values for k_+ (k_-). Grey vertical lines mark values for outcomes Z and Y. Dashed lines indicate predicted values of k in each regime based on standard CPT parametrizations, calculated at X, X', and X''. Dotted lines connect CPT predictions between regimes.

Appendix A

Chapter 1 Appendix

A.1 Round 2 Heterogeneity in Belief-updating

In Round 2, participants are told the callback rate for applicants with Black-sounding names in BM. They subsequently update their beliefs about racial hiring discrimination via their prediction of the callback rate for applicants with White-sounding names in BM. Most participants overestimate how often applicants with Black-sounding names receive a call back for an interview from employers.

At an individual level, we can test whether participants interpret the low callback rate for Black applicants as (1) a signal of discrimination, (2) a signal of low average callback rates in the study, and (3) a combination of the two. Participants who update in line with (1) would treat the information on the callback rate for applicants with Black-sounding names as purely a signal of discrimination. That is, they would not update their best guess of the White callback rate in response to the Black callback rate. Participants who update in line with (2) would change their predicted White callback rate in response to the Black callback rate such that the percentage difference in the their predicted callback rates for Black and White applicants remains constant between Rounds 1 and 2. Participants who update in line with (3) would update their estimate of the callback rate for White applicants between the estimates of (1) and (2). Given that (1) and (2) have predict point estimates while (3) predicts a range of responses, we may expect most participants to fall into (3). However, it is interesting to see if we have a mass of participants at the sharp predictions for either (1) or (2).

Consider, for example, a hypothetical participant named Abby predicts in Round 1 that White applicants need to send out 6 resumes to receive one callback, and Black applicants need to send out 8 resumes. Then Abby finds out in Round 2 that Black applicants need to send out 15 resumes. If Abby maintains that she thinks White applicants need to send out 6 resumes to receive one callback, then her updating procedure is identified as process (1). That is, she uses the Black callback rate as purely a signal of discrimination, and does not update on the White callback rate.

If Abby updates using process (2), then she will update her belief about the White callback rate such that her imputed discrimination belief does not change from Round 1 to Round 2. This would imply an updated Round 2 belief that White applicants need to send out 11.25 resumes to receive one callback. Because more than 99% of participants' submitted callback rates in my study are integers, I count Abby as updating in line with (2) if she correctly rounds to the nearest integer, 11. If Abby uses a combination of these updating processes (process 3), then she would submit an estimate strictly between these two numbers: greater than 6 and less than 11.

For this heterogeneity analysis, I exclude participants who correctly guess the callback rate for applicants with Black-sounding names in BM (6%) and those whose updating patterns cannot be identified because they submitted initial predictions in Round 1 such that updating in line with (1) and (2) yield the same resulting updating behavior ($< 1\%$).

I find that 19% of participants update their beliefs in line with (1), i.e., they update only on discrimination. That is, 19% do not change their beliefs about the White callback rate after finding out the Black callback rate in BM. Democrats and Republicans are equally likely to exhibit this pattern ($p = 0.15$). I find that 18% of participants update in line with (2), i.e., they do not update on discrimination and only use the Black callback rate information as information about the average callback rates in BM. 20% of Democrats and 17% of Republicans exhibit this behavior ($p = 0.07$). 40% of participants exhibit updating patterns that suggest a combination of these two processes. This includes 37% of Democrats and 43% of Republicans ($p = 0.01$).

This leaves 21% of participants who exhibit updating patterns that are not consistent with any of these processes. That is, these participants may find out that Black applicants had to send out more resumes than they thought, and subsequently decreased their beliefs about discrimination. Or, participants find out that Black applicants had to send out fewer resumes than they thought, and subsequently increased their beliefs about discrimination. Democrats and Republicans are equally likely to be in this group ($p = 0.79$).

A.2 Round 4 Unincentivized Questions on Wage Gap Drivers

In Round 4, participants are asked the percent of the Black-White wage gap among full-time workers in the US they think is driven by differences in educational attainment and the percent they think is driven by employer discrimination. Participants are told that statisticians have developed methods of calculating how much of the wage gap is explained by the fact that Black workers have, on average, lower levels of educational attainment than White workers. Their best guess for this number is eligible to be selected for a bonus, in which case participants earn the $2 bonus if they are within 5 percentage points of the correct answer (12%). Participants are also told that there do not currently exist methods of calculating how much of the wage gap is driven by employer discrimination, so this question is hypothetical.

Baseline predictions of how much educational attainment and employer discrimination drive the wage gap are shown in Figure A.1 by political affiliation. For each participant, I assign the remaining weight (out of 100%) to the "all other reasons" category. Democrats on average estimate that 31% of the wage gap is driven by educational attainment, while Republicans estimate that 37% if driven by educational attainment ($p < 0.001$). Democrats tend to put more weight (40%) on employer discrimination than Republicans do (29%) ($p < 0.001$).

In the final question of Round 4, I ask participants their updated beliefs about how much

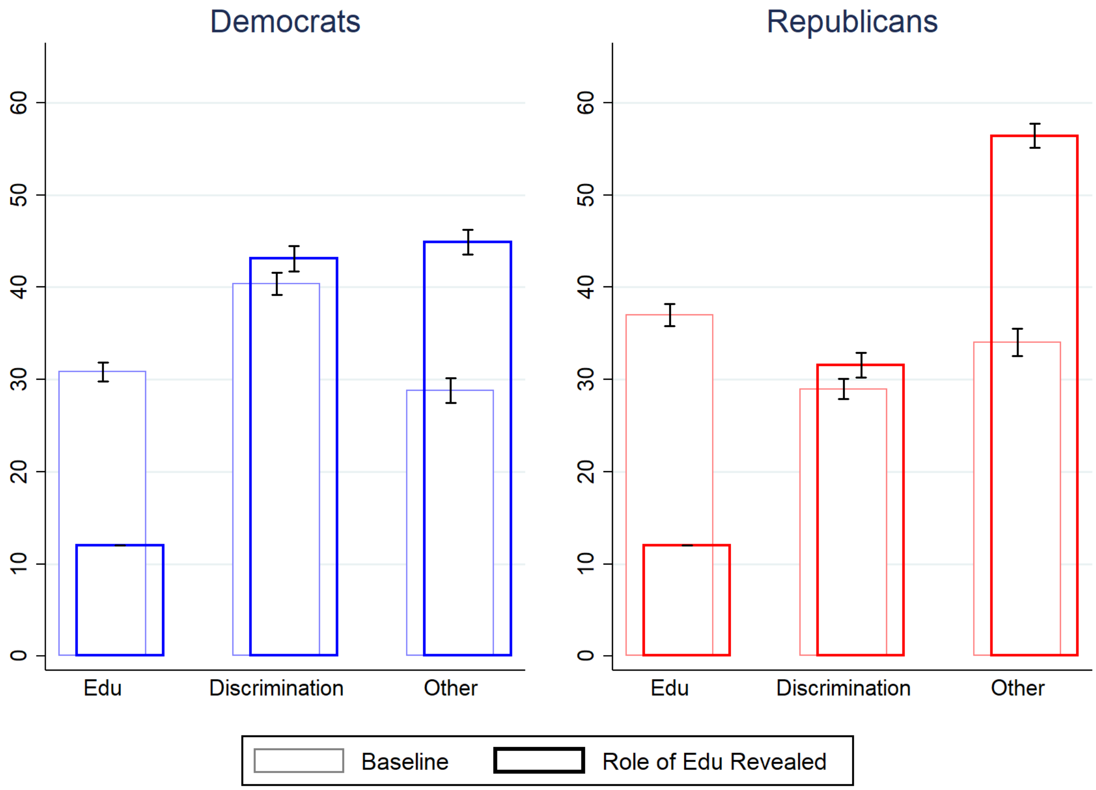

Figure A.1. Beliefs about wage gap drivers

Notes. The figures show participants' average estimates (with 95% confidence intervals) of the explanatory power of potential drivers of the Black-White wage gap among full-time workers in the US. At baseline (depicted by thin lines), participants are asked how much of the wage gap they think is explained by (1) differences in educational attainment, and (2) employer discrimination. The third category ("Other") is calculated as the remaining percentage of the wage gap after accounting for participants' responses to the first two categories. After the percentage of the Black-White wage gap explained by educational attainment is revealed to participants, I again elicit their beliefs about how much of the wage gap is driven by employer discrimination. The thick solid lines show participants' responses after learning this information, with the amount on educational attainment fixed at 12%. The left panel restricts my sample to Democrats, and the right panel restricts to Republicans.

of the Black-White wage gap is driven by employer discrimination, given that 12% is estimated to be driven by educational attainment. Figure A.1 shows participants' decomposition of the wage gap into three categories of drivers for Democrats and Republicans before (baseline) and after finding out how much of the wage gap is driven by differences in educational attainment.

After learning that educational attainment explains 12% of the wage gap, Democrats (left panel) increase their estimates of how much of the wage gap is explained by employer discrimination by approximately 2.8 percentage points, and Republicans increase their estimates of how much employer discrimination explains by on average 2.6 percentage points ($p = 0.83$) even though Republicans overestimated the role of educational attainment more than Democrats. As a result, information on the explanatory power of educational attainment does not affect the gap between Democrats' and Republicans' beliefs about the explanatory power of employer discrimination on the Black-White wage gap.

A.3 Alternative Regression Specifications

A.3.1 Hiring Discrimination Beliefs as Ratios of Callback Rates

In this subsection, I report the main regression results using ratios instead of logs. That is, in Round 1, I calculate participants' beliefs about the extent of racial discrimination in hiring in a given round as follows.

$$D_{1,i} = \frac{\widehat{B}_i}{\widehat{W}_{1,i}} \tag{A.1}$$

In Rounds 2-5, after participants are told the actual callback rate for applicants with Black-sounding names in BM, I calculate beliefs about the extent of racial discrimination in hiring as follows.

$$D_{j,i} = \frac{\overline{B}}{\widehat{W}_{j,i}} \tag{A.2}$$

where $D_{j,i}$ is participant i's calculated belief about hiring discrimination in Round $j \in \{2,5\}$, $\overline{B}$ is the actual number of times resumes with Black-sounding names had to be sent out to receive one callback from BM (15), and $\widehat{W}_{j,i}$ is participant i's Round $j \in \{2,5\}$ prediction of the number of times resumes with White-sounding names had to be sent out to receive one callback from BM.

To calculate participants' updating behavior between rounds, I simply take the difference between their beliefs about hiring discrimination, as follows.

$$\Delta D_{j,j+1,i} = D_{j+1,i} - D_{j,i} \tag{A.3}$$

for $j \in (1,4)$.

Table A.1. Round 2 Hiring Discrimination Belief Updates to BM Black Callback Rate

	Belief Update: R2	Belief Update: R2	Belief Update: R2
Error: Black CB	0.0223***		
	(0.00395)		
Over × Error: Black CB		-0.000899	
		(0.00299)	
Under × Error: Black CB		0.194***	
		(0.0121)	
Dem × Over × Error: Black CB			0.0425
			(0.0393)
Dem × Under × Error: Black CB			0.173***
			(0.0253)
Rep × Over × Error: Black CB			-0.00220
			(0.00461)
Rep × Under × Error: Black CB			0.175***
			(0.0230)
Constant	0.961***	-0.302***	-0.110
	(0.0531)	(0.0700)	(0.200)
Observations	2189	2189	2190

Notes. The table shows regressions of participants' belief updates about racial discrimination in hiring between rounds 1 and 2. Round 1 beliefs are the ratio of participants' predicted callback rate for applicants with Black-sounding names to their predicted callback rate for applicants with White-sounding names in BM. Round 2 beliefs are the ratio of the actual callback rate for applicants with Black-sounding names (15) to participants' predicted callback rate for applicants with White-sounding names in BM. The outcome variable is the difference of these two ratios. "Error: Black CB" is the difference between participants' priors on the Black callback rate and the actual Black callback rate in BM. "Under" ("Over") includes only participants who underestimate (did not underestimate) the number of times Black resumes had to be sent out to get a callback in BM. "Dem" ("Rep") includes only participants who identify as Democrats (Republicans). Robust standard errors are in parentheses.

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

Table A.2. Hiring Discrimination Belief Updates to Black-White Wage Gap

	Belief Update: R3	Belief Update: R3	Belief Update: R3
Error: White Earn	0.000374***		
	(0.0000757)		
Dem × Error: White Earn		0.000495***	
		(0.0000938)	
Rep × Error: White Earn		0.000137	
		(0.000120)	
Dem × Over × Error: White Earn			0.000519***
			(0.000106)
Dem × Under × Error: White Earn			0.000352
			(0.000268)
Rep × Over × Error: White Earn			0.000152
			(0.000181)
Rep × Under × Error: White Earn			0.000110
			(0.000195)
Constant	-0.0103	-0.00407	0.00243
	(0.0175)	(0.0177)	(0.0239)
Observations	2078	2078	2078

Notes. The table shows regressions of participants' belief updates about racial discrimination in hiring between rounds 2 and 3. Round 2 (Round 3) beliefs are the ratios of the actual callback rate for applicants with Black-sounding names (15) to participants' Round 2 (Round 3) predicted callback rate for applicants with White-sounding names in BM. The outcome variable is the difference of these two ratios. "Error: White Earn" is the difference between participants' priors on average earnings for White full-time workers in the US and the actual average earnings for White full-time workers in the US (according to the Bureau of Labor Statistics). "Under" ("Over") includes only participants who underestimate (did not underestimate) the average White full-time earnings. "Dem" ("Rep") includes only participants who identify as Democrats (Republicans). To limit the effects of outliers, observations in which the magnitude of the outcome variable is larger than 3 (4.2% of sample) were dropped from this specification. Robust standard errors are in parentheses.

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

Table A.3. Hiring Discrimination Belief Updates to Role of Edu on Black-White Wage Gap

	Belief Update: R4	Belief Update: R4	Belief Update: R4
Error: Pct EA	-0.00150*		
	(0.000846)		
Dem × Error: Pct EA		-0.00281**	
		(0.00122)	
Rep × Error: Pct EA		-0.000818	
		(0.000903)	
Dem × Under × Error: Pct EA			0.00735
			(0.0142)
Dem × Over × Error: Pct EA			-0.00317**
			(0.00131)
Rep × Under × Error: Pct EA			0.00349
			(0.0144)
Rep × Over × Error: Pct EA			-0.00108
			(0.000972)
Constant	0.112***	0.108***	0.0977***
	(0.0246)	(0.0247)	(0.0283)
Observations	2110	2110	2110

Notes. The table shows regressions of participants' changes in beliefs about racial hiring discrimination between rounds 3 and 4. Round 3 (Round 4) beliefs are the ratios of the actual callback rate for applicants with Black-sounding names to participants' Round 3 (Round 4) predicted callback rate for applicants with White-sounding names in BM. The outcome variable is the difference of these two ratios. "Error: Pct EA" is the difference between participants' priors of how much of the Black-White wage gap is explained by educational attainment and an actual estimate of the amount (12%). "Over" ("Under") includes all participants who overestimate (do not overestimate) the role of educational attainment. "Dem" ("Rep") includes only participants who identify as Democrats (Republicans). To limit the effects of outliers, observations in which the magnitude of the outcome variable is larger than 3 (3.8% of sample) were dropped from this specification. Robust standard errors are in parentheses.

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

Table A.4. Round 5 Hiring Discrimination Belief Updates

	Belief Update: R5
Rep	0.262***
	(0.0995)
Constant	-0.979***
	(0.0674)
Observations	2196

Notes. The table shows regressions of participants' changes in beliefs (between rounds 4 and 5) about racial hiring discrimination in response to results from another fake resume study. Round 4 (Round 5) beliefs are the ratios of the actual callback rate for applicants with Black-sounding names (15) to participants' predictions of the callback rate for applicants with White-sounding names in BM. "Rep" includes only participants who list their political affiliation as "Republican" on their Prolific account. Robust standard errors are in parentheses.

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

Appendix B

Chapter 2 Appendix

B.1 Additional Tables and Figures

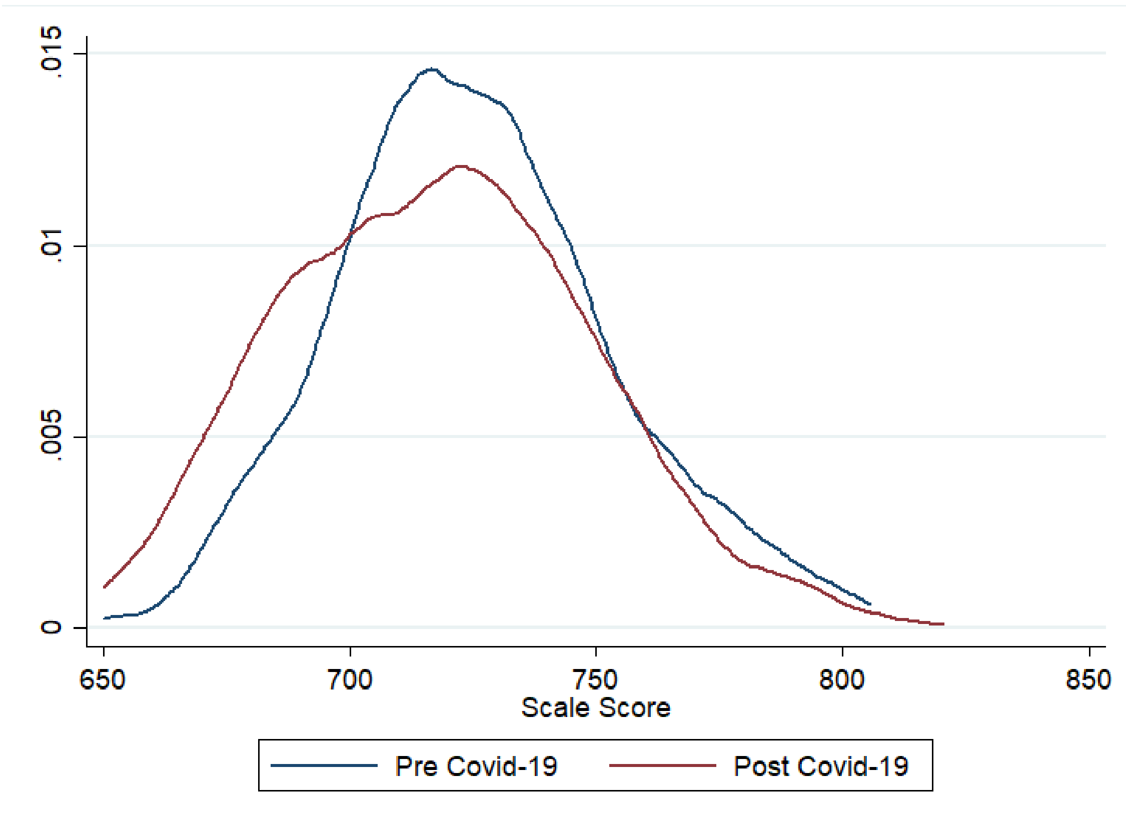

Figure B.1. Distribution of IAR Scores: Pre vs. Post Covid-19

Notes. This figure depicts the probability distribution function of students' scale scores on the Illinois Assessment of Readiness (IAR) exam pre-Covid in 2019 (blue line) and post-Covid in 2021 (red line) among students in our main sample.

Table B.1. Score Changes on Pre-Scores

	Learning Loss (2019-2021)	Learning Loss (2017-2019)
Parent Academy	-0.184**	0.0227
	(0.0908)	(0.0801)
Pre-k	-0.0400	-0.0265
	(0.163)	(0.102)
Normalized Pre Score	0.217***	0.599***
	(0.0410)	(0.0475)
Female	-0.134**	-0.187**
	(0.0658)	(0.0847)
Race: White	-0.0322	-0.202
	(0.134)	(0.175)
Race: Hispanic	-0.313**	-0.432***
	(0.128)	(0.107)
Age	0.176	0.0657
	(0.165)	(0.105)
Mother Edu: HS or GED	-0.141**	0.0591
	(0.0582)	(0.0712)
Mother Edu: Some Coll/2-year Deg	-0.116	-0.0774
	(0.103)	(0.0957)
Mother Edu: Bachelor's +	-0.389**	-0.250**
	(0.155)	(0.122)
Income: 16000-36000	-0.00263	-0.0248
	(0.121)	(0.146)
Income: 36000+	0.193	-0.232
	(0.147)	(0.165)
Constant	-1.753	-0.233
	(2.324)	(1.440)
Observations	350	399

Notes. In this table, we regress student learning loss on CHECC treatment assignment, pre-Covid IAR test score, various demographics, and socioeconomic variables. In Model 1, learning loss is calculated as students' 2021 (post-Covid) IAR score subtracted from their 2019 (pre-Covid) IAR score. In Model 2, learning loss is calculated as students' 2019 IAR score subtracted from their 2017 IAR score. Learning loss is measured in standard deviations relative to the first score. All models control for CHECC participation year, IAR exam year, baseline cognitive and executive function scores during CHECC participation. Standard errors clustered by district are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

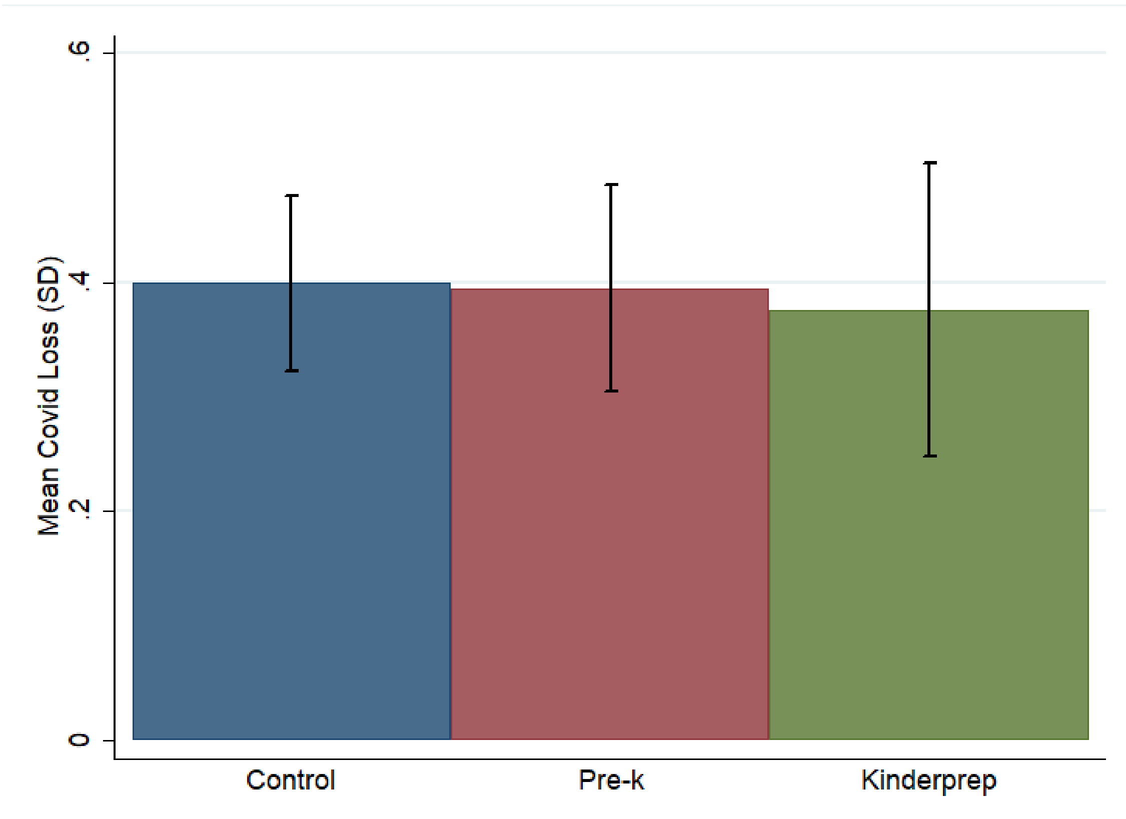

Figure B.2. CHECC Years 2012 & 2013

Notes. This figure depicts average Covid-19 learning loss by treatment assignment in the Chicago Heights Early Childhood Center (CHECC), restricting to students who participated in CHECC in 2012 and 2013. During this period, there was no Parent Academy. Covid-19 learning loss is calculated as the difference students' scores on the Illinois Assessment of Readiness (IAR) exam in 2021 (supplemented by 2022 scores if missing) and students' scores in 2019 (supplemented by 2018 scores if missing). Average learning loss is calculated in standard deviations of 2019 scores.

Table B.2. Treatment Effects on Covid-19 Learning Loss with Inverse Probability Weighting

	Covid-19 Loss (SD)
Parent Academy	-0.12**
	(0.06)
Score Pre-Covid	0.15***
	(0.05)
Female	-0.15**
	(0.06)
Race: White	-0.14
	(0.09)
Race: Hispanic	-0.19*
	(0.10)
Age	-0.01
	(0.12)
Mother Edu: HS or GED	-0.28***
	(0.09)
Mother Edu: Some College/2-year Degree	-0.24**
	(0.11)
Mother Edu: Bachelor's +	-0.34**
	(0.16)
Income: 16000-36000	-0.05
	(0.20)
Income: 36000+	0.17
	(0.15)
Constant	1.20
	(1.80)
Observations	337

Notes. We use inverse probability weights (IPW) as a robustness test for our main finding to account for differences between our analysis sample and the sample of students that was eligible for analysis. We first predict how likely each student is to be included in our sample based on their IAR score prior to Covid-19, an indicator for whether this score is missing, CHECC year, and demographic characteristics. These prediction scores are calculated separately for those in the control group and the Parent Academy treatment. We then run our main specification, weighting observations by the inverse of this likelihood. Standard errors clustered by district in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Bibliography

Abdellaoui, Mohammed, Chen Li, Peter P. Wakker, and George Wu, "A Defense of Prospect Theory in Bernheim & Sprenger's Experiment," *Working Paper*, 2020.

Agostinelli, Francesco, Matthias Doepke, Giuseppe Sorrenti, and Fabrizio Zilibotti, "When the great equalizer shuts down: Schools, peers, and parents in pandemic times," *Journal of public economics*, 2022, *206*, 104574.

Alesina, Alberto, Matteo F Ferroni, and Stefanie Stantcheva, "Perceptions of racial gaps, their causes, and ways to reduce them," Technical Report, National Bureau of Economic Research 2021.

Allais, Maurice, "Le Comportement de l'Homme Rationnel devant le Risque: Critique des Postulats et Axiomes de l'Ecole Americaine," *Econometrica*, 1953, *21* (4), 503–546.

Almond, Douglas, Janet Currie, and Valentina Duque, "Childhood circumstances and adult outcomes: Act II," *Journal of Economic Literature*, 2018, *56* (4), 1360–1446.

Andreoni, James and Tymofiy Mylovanov, "Diverging opinions," *American Economic Journal: Microeconomics*, 2012, *4* (1), 209–232.

Bacher-Hicks, Andrew, Joshua Goodman, and Christine Mulhern, "Inequality in household adaptation to schooling shocks: Covid-induced online learning engagement in real time," *Journal of Public Economics*, 2021, *193*, 104345.

Bansak, Cynthia and Martha Starr, "Covid-19 shocks to education supply: how 200,000 US households dealt with the sudden shift to distance learning," *Review of Economics of the Household*, 2021, *19* (1), 63–90.

Benjamin, Daniel J, "Errors in probabilistic reasoning and judgment biases," *Handbook of Behavioral Economics: Applications and Foundations 1*, 2019, *2*, 69–186.

Bernheim, B. Douglas and Charles Sprenger, "On the Empirical Validity of Cumulative Prospect Theory: Experimental Evidence of Rank-Independent Probability Weighting," *Econometrica*, 2020, *88* (4), 1363–1409.

— **and** — , “Rank-Independent Probability Weighting: A Response to Abdelloui et al. (2021),” *Working Paper*, 2022.

Bertrand, Marianne and Sendhil Mullainathan, “Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination,” *American economic review*, 2004, *94* (4), 991–1013.

Betthäuser, Bastian A, Anders M Bach-Mortensen, and Per Engzell, “A systematic review and meta-analysis of the evidence on learning during the COVID-19 pandemic,” *Nature Human Behaviour*, 2023, *7* (3), 375–385.

Binder, Sarah, “Polarized we govern?,” *Governing in a polarized age: Elections, parties, and political representation in America*, 2014.

Birnbaum, Michael H. and William R. McIntosh, “Violations of Branch Independence in Choices Between Gambles,” *Organizational Behavior and Human Decision Processes*, 1996, *67*, 91–110.

Bowen, T Renee, Danil Dmitriev, and Simone Galperti, “Learning from shared news: When abundant information leads to belief polarization,” *The Quarterly Journal of Economics*, 2023, *138* (2), 955–1000.

Canen, Nathan J, Chad Kendall, and Francesco Trebbi, “Political parties as drivers of us polarization: 1927-2018,” Technical Report, National Bureau of Economic Research 2021.

Cassidy, Jude, Bonnie E Brett, Jacquelyn T Gross, Jessica A Stern, David R Martin, Jonathan J Mohr, and Susan S Woodhouse, “Circle of Security–Parenting: A randomized controlled trial in Head Start,” *Development and psychopathology*, 2017, *29* (2), 651–673.

Chang, Edward H, Katherine L Milkman, Dena M Gromet, Robert W Rebele, Cade Massey, Angela L Duckworth, and Adam M Grant, “The mixed effects of online diversity training,” *Proceedings of the National Academy of Sciences*, 2019, *116* (16), 7778–7783.

Chay, Kenneth Y, Patrick J McEwan, and Miguel Urquiola, “The central role of noise in evaluating interventions that use test scores to rank schools,” *American Economic Review*, 2005, *95* (4), 1237–1258.

Chen, Daniel L, Martin Schonger, and Chris Wickens, “oTree—An open-source platform for laboratory, online, and field experiments,” *Journal of Behavioral and Experimental Finance*, 2016, *9*, 88–97.

Chuan, Amanda, John A List, Anya Samek, and Shreemayi Samujjwala, “Parental Investments in Early Childhood and the Gender Gap in Math and Literacy,” in “AEA Papers and Proceedings,” Vol. 112 American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203 2022, pp. 603–608.

— **, John List, and Anya Samek**, "Do financial incentives aimed at decreasing interhousehold inequality increase intrahousehold inequality?," *Journal of public economics*, 2021, *196*, 104382.

DellaVigna, Stefano and Matthew Gentzkow, "Persuasion: empirical evidence," *Annual Review of Economics*, 2010, *2* (1), 643–669.

Doucet, Armand, Deborah Netolicky, Koen Timmers, and Francis Jim Tuscano, "Thinking about pedagogy in an unfolding pandemic," *Work of Education International and UNESCO*, 2020.

Dunn, Lloyd M and Leota M Dunn, "Peabody picture vocabulary test–," 1965.

Engzell, Per, Arun Frey, and Mark D Verhagen, "Learning loss due to school closures during the COVID-19 pandemic," *Proceedings of the National Academy of Sciences*, 2021, *118* (17), e2022376118.

Francesconi, Marco and James J Heckman, "Child development and parental investment: Introduction," *The Economic Journal*, 2016, *126* (596), F1–F27.

Fryer, Roland G, Steven D Levitt, John A List et al., "Parental incentives and early childhood achievement: A field experiment in Chicago heights," Technical Report, National Bureau of Economic Research 2015.

Gächter, Simon and Elke Renner, "The effects of (incentivized) belief elicitation in public goods experiments," *Experimental Economics*, 2010, *13*, 364–377.

García, Jorge Luis and James J Heckman, "Parenting promotes social mobility within and across generations," *Annual Review of Economics*, 2022, *15.*

Goldhaber, Dan, Thomas J Kane, Andrew McEachin, Emily Morton, Tyler Patterson, and Douglas O Staiger, "The consequences of remote and hybrid instruction during the pandemic," Technical Report, National Bureau of Economic Research 2022.

Grigorieff, Alexis, Christopher Roth, and Diego Ubfal, "Does information change attitudes towards immigrants? Representative evidence from survey experiments," *Representative Evidence from Survey Experiments (March 10, 2018)*, 2018.

Haaland, Ingar and Christopher Roth, "Beliefs about racial discrimination and support for pro-black policies," *The Review of Economics and Statistics*, 2021, pp. 1–38.

Hassan, Md Hashibul, Asad Islam, Abu Siddique, Liang Choon Wang et al., *Telementoring* *and homeschooling during school closures: A randomized experiment in rural Bangladesh*, IZA-Institute of Labor Economics, 2023.

Heckman, James J, "The economics, technology, and neuroscience of human capability formation," *Proceedings of the national Academy of Sciences*, 2007, *104* (33), 13250–13255.

_ , "Invest in early childhood development: Reduce deficits, strengthen the economy," *The Heckman Equation*, 2012, *7* (1-2).

Heckman, James, Rodrigo Pinto, and Peter Savelyev, "Understanding the mechanisms through which an influential early childhood program boosted adult outcomes," *American Economic Review*, 2013, *103* (6), 2052–2086.

Heffetz, Ori, "Are Reference Points Merely Lagged Beliefs Over Probabilities," *NBER Working Paper 24721*, 2018.

Jacquemet, Nicolas and Constantine Yannelis, "Indiscriminate discrimination: A correspondence test for ethnic homophily in the Chicago labor market," *Labour Economics*, 2012, *19* (6), 824–832.

Jeong, Joshua, Emily E Franchett, Clariana V Ramos de Oliveira, Karima Rehmani, and Aisha K Yousafzai, "Parenting interventions to promote early child development in the first three years of life: A global systematic review and meta-analysis," *PLoS medicine*, 2021, *18* (5), e1003602.

Jr, Roland G Fryer, Steven D Levitt, John A List, and Anya Samek, "Introducing cogx: A new preschool education program combining parent and child interventions," Technical Report, National Bureau of Economic Research 2020.

Kahneman, Daniel and Amos Tversky, "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 1979, *47* (2), 263–291.

Levendusky, Matthew S, "Why do partisan media polarize viewers?," *American journal of political science*, 2013, *57* (3), 611–623.

Levy, Gilat and Ronny Razin, "Information diffusion in networks with the Bayesian peer influence heuristic," *Games and Economic Behavior*, 2018, *109*, 262–270.

Lewis, Karyn and Megan Kuhfeld, "Learning during COVID-19: An Update on Student Achievement and Growth at the Start of the 2021-22 School Year. Brief.," *Center for School and Student Progress at NWEA*, 2021.

List, John A, Anya Samek, and Dana L Suskind, "Combining behavioral economics and field experiments to reimagine early childhood education," *Behavioural Public Policy*, 2018, *2* (1), 1–21.

Lopes, Lola L. and Gregg C. Oden, "The Role of Aspiration Level in Risky Choice: A Comparison of Cumulative Prospect Theory and SP/A Theory," *Journal of Mathematical Psychology*, 1999, *43*, 286–313.

Lunkenheimer, Erika S, Thomas J Dishion, Daniel S Shaw, Arin M Connell, Frances Gardner, Melvin N Wilson, and Emily M Skuban, "Collateral benefits of the Family Check-Up on early childhood school readiness: indirect effects of parents' positive behavior support.," *Developmental psychology*, 2008, *44* (6), 1737.

Marino, Maria, Roberto Iacono, and Johanna Mollerstrom, "(Mis-) perceptions, information, and political polarization," 2023.

Mazar, Nina, Botond Koszegi, and Dan Ariely, "True Context-dependent Preferences? The Cause of Market-dependent Valuations," *Journal of Behavioral Decision Making*, 2014, *27*, 200–208.

Mian, Atif, Amir Sufi, and Francesco Trebbi, "Resolving debt overhang: Political constraints in the aftermath of financial crises," *American Economic Journal: Macroeconomics*, 2014, *6* (2), 1–28.

Mu, Ren, "Perceived relative income, fairness, and the role of government: Evidence from a randomized survey experiment in China," *China Economic Review*, 2022, *73*, 101784.

Mullainathan, Sendhil and Andrei Shleifer, "The market for news," *American economic review*, 2005, *95* (4), 1031–1053.

Orlov, George, Douglas McKee, James Berry, Austin Boyle, Thomas DiCiccio, Tyler Ransom, Alex Rees-Jones, and Jörg Stoye, "Learning during the COVID-19 pandemic: It is not who you teach, but how you teach," *Economics Letters*, 2021, *202*, 109812.

Özler, Berk, Lia CH Fernald, Patricia Kariger, Christin McConnell, Michelle Neuman, and Eduardo Fraga, "Combining pre-school teacher training with parenting education: A cluster-randomized controlled trial," *Journal of Development Economics*, 2018, *133*, 448–467.

Perego, Jacopo and Sevgi Yuksel, "Media competition and social disagreement," *Econometrica*, 2022, *90* (1), 223–265.

Prior, Markus, Gaurav Sood, Kabir Khanna et al., "You cannot be serious: The impact of accuracy incentives on partisan bias in reports of economic perceptions," *Quarterly Journal of Political Science*, 2015, *10* (4), 489–518.

Redlawsk, David P, "Hot cognition or cool consideration? Testing the effects of motivated reasoning on political decision making," *Journal of Politics*, 2002, *64* (4), 1021–1044.

Sass, Tim and Salma Mohammad Ali, "Student Achievement Growth During the COVID-19 Pandemic: Spring 2022 Update," 2022.

Schaub, Simone, Erich Ramseier, Alex Neuhauser, Susan CA Burkhardt, and Andrea Lanfranchi, "Effects of home-based early intervention on child outcomes: A randomized controlled trial of Parents as Teachers in Switzerland," *Early Childhood Research Quarterly*, 2019, *48*, 173–185.

Settele, Sonja, "How do beliefs about the gender wage gap affect the demand for public policy?," *American Economic Journal: Economic Policy*, 2022, *14* (2), 475–508.

Sheridan, Susan M, Lisa L Knoche, Kevin A Kupzyk, Carolyn Pope Edwards, and Christine A Marvin, "A randomized trial examining the effects of parent engagement on early language and literacy: The Getting Ready intervention," *Journal of school psychology*, 2011, *49* (3), 361–383.

Slothuus, Rune and Claes H De Vreese, "Political parties, motivated reasoning, and issue framing effects," *The Journal of politics*, 2010, *72* (3), 630–645.

Thaler, Michael, "The "fake news" effect: An experiment on motivated reasoning and trust in news," 2019.

Tversky, Amos and Daniel Kahneman, "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty*, 1992, *5* (4), 297–323.

Wakker, Peter, Ido Erev, and Elke U. Weber, "Comonotonic Independence: The Critical Test Between Classical and Rank Dependent Utility Theories," *Journal of Risk and Uncertainty*, 1994, *9*, 195–230.

Wu, George, "An Empirical Test of Ordinal Independence," *Journal of Risk and Uncertainty*, 1994, *9*, 39–60.

Wu, Mengfan, Qiwei Yu, Sabrina L Li, and Liqiang Zhang, "Geographic and gender disparities in global education achievement during the COVID-19 pandemic," *International Journal of Applied Earth Observation and Geoinformation*, 2022, *111*, 102850.

Zhuravskaya, Ekaterina, Maria Petrova, and Ruben Enikolopov, "Political effects of the internet and social media," *Annual Review of Economics*, 2020, *12*, 415–438.