

DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

# CALIFORNIA INSTITUTE OF TECHNOLOGY

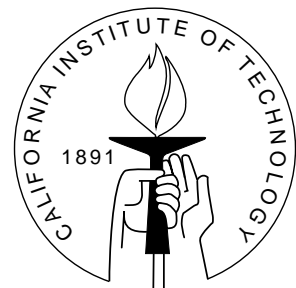
PASADENA, CALIFORNIA 91125

## STATISTICAL DISCRIMINATION AND AFFIRMATIVE ACTION IN THE LAB

Ahrash Dianat  
University of Essex

Federico Echenique  
California Institute of Technology

Leeat Yariv  
Princeton University



**SOCIAL SCIENCE WORKING PAPER 1439**

April 2018

# Statistical Discrimination and Affirmative Action in the Lab\*

Ahrash Dianat

Federico Echenique

Leeat Yariv

## Abstract

We present results from laboratory experiments studying the impacts of affirmative-action policies. We induce statistical discrimination in simple labor-market interactions between firms and workers. We then introduce affirmative-action policies that vary in the size and duration of a subsidy firms receive for hiring discriminated-against workers. These different affirmative-action policies have nearly the same effect and practically eliminate discriminatory hiring practices. However, once lifted, few positive effects remain and discrimination reverts to its initial levels. One exception is lengthy affirmative-action policies, which exhibit somewhat longer-lived effects. Stickiness of beliefs, which we elicit, helps explain the evolution of these outcomes.

---

\*We thank Richard Sander for very helpful comments. Echenique gratefully acknowledges the support of NSF grants SES-1558757 and CNS-1518941. Yariv gratefully acknowledges the support of NSF grant SES-1629613.

# 1 Introduction

Affirmative-action policies have had a tumultuous history since their introduction over 50 years ago in the U.S.<sup>1</sup> Intended to alleviate inequality in employment and pay, increase access to education, and promote diversity, they have been a subject of legal and political controversy. In their various forms, they are often set as a temporary “nudge”, put in place for a limited duration. For instance, in 2003, the Supreme Court ruled the use of race for affirmative action in school admissions as constitutional (*Grutter v. Bollinger*). Justice Sandra Day O’Connor famously stated that:

“We expect that 25 years from now, the use of racial preferences will no longer be necessary to further the interest [in student-body diversity] approved today.”

In this paper, we focus on statistical discrimination, where the source of unequal treatment is a correct statistical evaluation of past performance of different groups of individuals, rather than differential tastes. Our goal is to assess experimentally the efficacy of several affirmative-action policies for generating equal treatment while in place and, especially, after they are lifted.

Our results are threefold. The first is methodological: we are able to induce statistical discrimination in the laboratory, where workers from a disadvantaged group are discriminated against despite the absence of any exogenous differences between them. We consider affirmative-action policies that seek to reverse the induced discriminatory attitudes by rewarding firms for hiring disadvantaged workers. Rewards are of varying amounts and duration. Our second finding is that such policies, while in place, generally succeed in eliminating discrimination and all workers are hired at similar rates. However,

---

<sup>1</sup>The term “affirmative action” was introduced by President John F. Kennedy in an executive order in 1961 as a method for redressing discrimination that had persisted in spite of civil rights laws and constitutional guarantees. It was further developed and enforced under President Lyndon B. Johnson.

our third and arguably most important finding is that when affirmative-action policies are removed, statistical discrimination rears its head. Furthermore, our elicitation of beliefs illustrates that differing group-specific beliefs, which lie at the root of the discrimination we observe, survive the different affirmative-action policies we implement. The main message of the paper therefore suggests more limited optimism than Justice O'Connor's statement—while affirmative action policies are effective tools for combating inequality while in place, their impact is short-lived. Once lifted, behavior reverts back to the unequal treatment that affirmative action was designed to undo.

Our experimental design is a particularly simple operationalization of the statistical discrimination model suggested by Kenneth Arrow (1971; 1973). In Arrow's model, statistical discrimination is the manifestation of a coordination failure.<sup>2</sup> Simplified, the game underlying Arrow's analysis can be described as follows. Agents interact in worker-firm pairs in every period. A worker may invest in a firm-specific skill, but this investment is worthwhile only if she is hired. The firm's incentives are such that hiring a worker is desirable only if the worker has invested in acquiring the skill. However, workers and firms do not know each other's choices when making their own. The resulting game has two Pareto-ranked pure equilibria. In one, the worker invests and the firm hires. In the other, the worker does not invest and the firm does not hire.

In our experiments, workers are of two types: green and purple. Colors are a metaphor for group identity and substitute for observable characteristics of real-world workers (e.g., gender, race). Colors can serve as a coordination device, with different equilibria being followed when different types of workers are involved. Importantly, worker color is observable, both in our experiments and their real-world analogues. We term outcomes in which agents play different equilibria, depending on the color of the worker, as *discriminatory*.

---

<sup>2</sup>Arrow's model assumes a perfectly symmetric situation and generates differences among groups endogenously. Phelps (1972a) also proposes statistical discrimination as an explanation for observed discriminatory outcomes, but his explanation requires the existence of genuine exogenous differences between groups.

In each of our sessions, experimental workers and firms are randomly matched, and play the game multiple times. All participants observe a moving average of historical play—worker investment levels and firm hiring rates pertaining to both green and purple workers. The public history of play is used to generate hysteresis.

Our sessions are comprised of four stages. First, our participants play a version of the game in which purple workers are less productive than green workers—i.e., have higher costs of investment. This first stage of the game “seeds” the discriminatory outcome.

In the second stage, we equalize the costs of investment across workers so that there are no material differences between the two worker groups. The history of play is enough to generate discriminatory outcomes: we find that participants coordinate on the inefficient equilibrium (no-investment and no-hiring) when the participating worker is purple. Indeed, in almost all sessions, firms hire purple workers at significantly lower rates than green workers, as they correctly anticipate purple workers will invest at lower rates. Conversely, purple workers invest less than green workers as they correctly anticipate that they will not be hired as often. These observations illustrate two points. First, our method of inducing discrimination is effective. Second, when we equalize conditions, discrimination remains. In particular, equalization in and of itself is insufficient to redress discrimination due to a historically unequal environment.

The third stage introduces affirmative action. Namely, firms are rewarded for hiring purple workers. We vary the magnitude of the reward and the duration of affirmative action across treatments. Despite a salient public history of inefficient outcomes associated with purple workers, we now see firms more willing to hire purple workers. In turn, purple workers invest at greater rates. Across our treatments, while affirmative action is in place, discrimination essentially goes away.

Finally, in the last stage, we lift the affirmative-action policy. Thus, the last stage

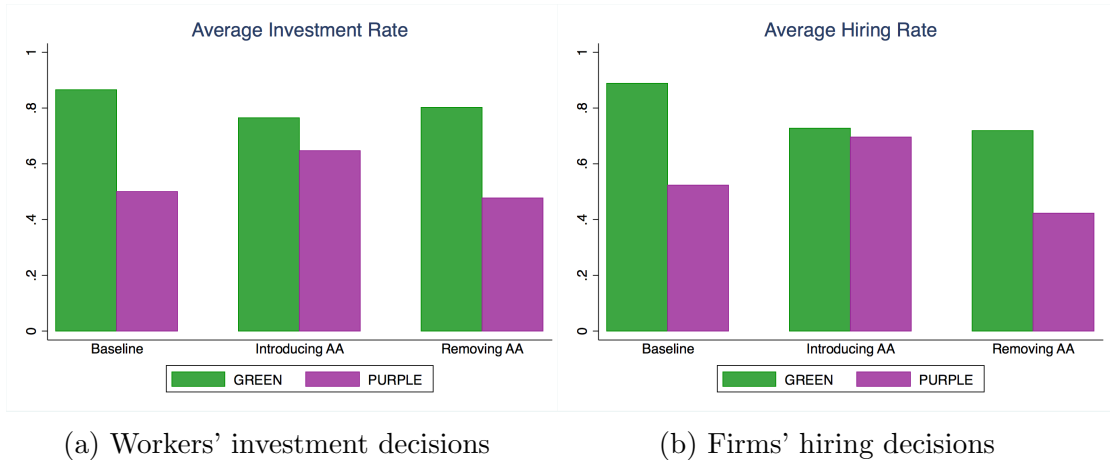


Figure 1

is identical to the second stage, in which the environment is perfectly symmetric across worker types. Across our treatments, participants' play reverts back to the discriminatory equilibrium we observed in the second stage of our experiment: purple workers invest less often than green workers, and are hired less often. In other words, *the effects of affirmative action are short-lived*. One exception is the treatment in which affirmative action was imposed for a relatively long duration, for double the number of periods of each of the other stages. In that case, some improvement is seen even after the policy is lifted, but outcomes are still significantly different from those implied by an equal treatment of workers.

Figure 1 illustrates our main findings through our benchmark treatment, entailing a moderate affirmative-action subsidy for a duration that equals the duration of our other experimental stages. It shows investment and hiring rates for the different stages of our experiment, using green and purple colors to distinguish matches. The figure illustrates the findings described above: the baseline stage exhibits statistical discrimination; affirmative action largely equalizes hiring and investment rates between green and purple workers; and the removal of affirmative action sees participants reverting back to statistical discrimination.

In our experimental design, the aggregate public history of play is used to generate hysteresis, in an attempt to capture discriminatory path-dependent outcomes in the real world. We find that the strength of discrimination survives affirmative action. In one of our treatments, the subsidy for hiring a purple worker is so high that it is a dominant strategy for firms to hire, and therefore a best response for purple workers to invest. This treatment exactly reverses the situation in the seed stage, where it is dominant for purple workers not to invest. Even in this treatment, where we mirror the incentives provided in the seed stage, affirmative action does not have lasting effects. We view this result as particularly telling. Naturally, it is difficult to determine which experimental implementation of affirmative action comes closest to capturing the mechanism operating behind real-world policies. But a policy that is the exact mirror image of the “seed” conditions that induced discrimination to begin with seems like a natural candidate for undoing discrimination. We find, however, that discrimination persists.

Why are the effects of affirmative action so short-lived? The impacts on beliefs provide some insights. Throughout our experiments, we elicit the beliefs of workers regarding their probabilities of being hired, as well as the beliefs of firms regarding the probabilities of the workers they encounter investing in training. In our experiments, firms’ decisions to hire purple workers are consistent with their reported beliefs about whether workers have invested. Affirmative-action policies change (temporarily) the decision to hire a purple workers, but firms’ beliefs do not change significantly—*beliefs are sticky*. These beliefs do not change substantially when affirmative-action policies are introduced or removed. Their level is such that hiring is optimal when subsidies are present, but not when they vanish. As firms appear to be best responding to their beliefs and the incentives they face, this stickiness in firms’ expectations accounts for the quickly-fading impacts of affirmative action.

To summarize, with a growing desire to equalize the playing field across the market,

our results suggest that affirmative-action policies should be thought through carefully. In their frequent forms, to be effective, they need to be both substantial in magnitude and long-lived.

## 2 Related Literature

Arrow (1998) argues that taste-based discrimination cannot explain unequal treatment of workers in the market since there would be arbitrage opportunities that would eventually drive out discriminating economic actors. Arrow concludes that statistical discrimination is the main existing hypothesis for explaining the prevalence of discrimination in market settings.

Statistical discrimination is consistent with a rich set of observations in the field. For instance, Ewens, Tomlin, and Wang (2014) identify statistical discrimination in the US rental market over race by analyzing renters' responses to applications from white-sounding and African American-sounding names in various neighborhoods. Glover, Pal-lais, and Pariente (2017) consider interactions between managers and workers in grocery stores and find that more biased managers are associated with lower performance of minority workers, a pattern similar to that we induce in our seed stage.<sup>3</sup>

Coate and Loury (1993) provide a theoretical investigation of statistical discrimination and affirmative action. They raise the possibility that discrimination may persist after a period of affirmative action. Some of the basic ideas in their theoretical model are present in our design: interactions are essentially bilateral, between workers and firms. Workers can invest, but an informative signal of their decision is available to the firms. Workers' incentives to invest are, in fact, channeled through the informative signal (workers want

---

<sup>3</sup>Bohren, Imas, and Rosenberg (2017) run a field experiment using a large online platform where users post content that is evaluated by others on the platform. They consider the assessment of posts exogenously varied by gender and history on the platform and find evidence for biased beliefs driving discriminatory attitudes. Our seeding stage serves to induce such biased beliefs.



to invest as it directly affects the chances they will be hired). In our design, in contrast, there is more of a pure coordination failure. Workers want to invest even if they are guaranteed to be hired: indeed, that is the case that maximizes workers' incentives to invest.

There are several studies that bring the Coate and Loury (1993) model to the lab. Anderson and Hauptert (1999) and Fryer, Goeree, and Holt (2005) designed classroom experiments that capture the main forces of Coate and Loury (1993). The design of our seeding stage is inspired by these, with two important differences. First, we do not allow (explicit) random signals on workers' investment decisions. Second, workers' incentives to invest respond to whether they are hired or not—we focus on a pure coordination setup.

Kidd, Carlin, and Pot (2008) seek to replicate some of the details in the Coate and Loury model. In particular, their design relies on an informative signal of whether a worker has invested, and much of the analysis tests the comparative-statics results of Coate and Loury (1993). In this setting, workers over-invest. However, the incentive structure in their design is not one of pure coordination failure like ours, in the sense that a worker's incentive to invest is not higher when they know that they will be hired. Furthermore, the Kidd, Carlin, and Pot (2008) design is one-sided: experimental participants play the role of workers, while firms' choices are computerized. In our study, firms' choices are an important object for analysis, and their evolving beliefs, which we elicit, provide hints as to the source of the limited long-run efficacy of affirmative action we observe. Feltovich, Gangadharan, and Kidd (2013) use a similar design to that used by Kidd, Carlin, and Pot (2008), but focus on outcomes following the removal of an affirmative-action policy. They find that workers invest significantly more after the affirmative action has been removed than when it is in place. In their design, since the incentives to invest operate through their effects on the informative signal, workers who

know they will be hired do not have strong incentives to invest. Therefore, affirmative action suppresses investment. Our experiment has the opposite property, and we see that workers invest more when the affirmative-action policy is in place.<sup>4</sup>

Outside of the Coate and Loury (1993) setting, various studies have looked at affirmative-action policies in the lab. For example, in the context of gender discrimination, Niederle, Segal, and Vesterlund (2013) illustrate the effectiveness of affirmative-action policies in inducing women to compete. In contrast, Bracha, Cohen, and Conell-Price (2005) illustrate the potential harmful effects of affirmative action through its production of *stereotype threat*, whereby women are primed with negative stereotypes. Anderson, Fryer, and Holt (2006) present a survey of experiments in psychology and economics dealing with discrimination.

There is important work using field data that suggests potential shortcomings of affirmative action. Sander (2004) illustrates several negative impacts on African-American law students in the US. Sander and Taylor (2012) provide a comprehensive account of the impacts of affirmative-action policies on racial equality in US higher education. They suggest that even while in place, affirmative-action policies may have had detrimental effects on minority students.<sup>5</sup>

To conclude, we believe there are several features of our design whose combination is absent from most of the literature on affirmative action. First, we generate discrimination in the lab, rather than relying on participants' existing prejudices that may vary in strength and be difficult to control. Second, we allow both workers and firms to act strategically and elicit all participants' beliefs throughout the evolution of our experi-

---

<sup>4</sup>Haan, Offerman, and Sloof (2015) study experimentally the impacts of of competition on statistical discrimination. They show that, in the setting of Coate and Loury (1993), competition allows statistical discrimination to emerge more easily and forcefully.

<sup>5</sup>Krueger, Rothstein, and Turner (2006) assess empirically Justice O'Connor's prediction that affirmative action would not be needed 25 years after the Grutter v. Bollinger ruling and find limited support for it.

mental sessions. Beliefs are challenging to elicit in the field and provide insights into the mechanisms driving the successes and limitations of affirmative action. Last, we consider the impacts of affirmative action not only when it is in place, but also after it is lifted.

Methodologically, our technique for seeding discrimination relies on the presence of hysteresis—equilibrium selection at the start of the experiment affects selection later on. Outside the context of discrimination, Romero (2015) illustrates the presence of hysteresis in standard coordination games in which payoff parameters were changed over time.

### 3 Experimental Design

We start with an overview of the basic statistical-discrimination model guiding our design and its motivations. We then set forth a detailed description of our experimental protocol.

#### 3.1 Statistical Discrimination

There are two leading theories of discrimination: taste-based (Becker, 1957) and statistical (Phelps, 1972b; Arrow, 1973). Taste-based discrimination attributes discriminatory decisions to an inherent preference for agents of certain groups over others. Statistical discrimination refers to the idea that, in the absence of direct information about a certain aspect of ability, a decision-maker would substitute group averages or variances corresponding to the individual’s demographics (e.g., gender, race). It is challenging to rule out empirically that discrimination is taste-based, but in the lab one can ensure that there is no inherent preference for one group over another.<sup>6</sup>

---

<sup>6</sup>There are notable econometric efforts to identify a statistical discrimination parameter in a structural model (Knowles, Persico, and Todd, 2001) or in other ways (Glover, Pallais, and Pariente, 2017)). There are also studies that aim to control the source of discrimination in a field experiment, e.g., Agan and Starr (2016) and Ewens, Tomlin, and Wang (2014). However, broadly speaking, it is difficult to completely rule out taste-based discrimination with field data. The lab has the advantage that one can control agents’ payoffs and observe the evolution of their beliefs.

		Firm	
		Hire	Not Hire
Worker	Invest	1800 - c, 1600	1000 - c, 1200
	Not Invest	1400, 400	1200, 1200

Table 1: The investment/hiring game, where  $c$  denotes the cost of investment.

Following Arrow (1971), we focus on statistical discrimination in the labor market, where employers may use observable traits to make hiring decisions. Statistical discrimination is not based on an inherent preference for hiring people of, say, a particular race or gender, but instead on a real or perceived correlation between an observable trait and unobservable productivity, or some unobservable aspect that informs the quality of a hire.

We consider statistical discrimination as a self-fulfilling equilibrium phenomenon. The model guiding our design is a stylized version of the model proposed by Arrow (1971)<sup>7</sup>: a worker can decide to undertake a costly investment in productivity, and a firm has to decide whether to hire the worker. The investment is only worthwhile to the worker if she is hired by the firm, but the worker has to make the decision to invest or not before knowing if she will be hired. The firm only wants to hire a worker if she has invested, but has to decide on hiring the worker without knowing the worker's choice. In our model, summarized in Table 1, for small enough investment costs  $c$ , there are two pure-strategy equilibria: (Invest, Hire) and (Not Invest, Not Hire). The former Pareto dominates the latter; there is therefore the possibility for *coordination failure*, whereby workers fail to invest because they correctly anticipate that they will not be hired.

Suppose there are two kinds of workers: GREEN workers and PURPLE workers—these categories can stand for different genders, races, etc. If worker color is observable, it is possible that firms will coordinate on (Invest, Hire) with, say, GREEN workers and on (Not Invest, Not Hire) with PURPLE workers. This situation corresponds to what

---

<sup>7</sup>See Appendix F in Arrow's paper; see also Coate and Loury (1993).

we term statistical discrimination.

The particular flavor of statistical discrimination we are interested in, based on self-fulfilling expectations about hiring and investment, begs the question of how a society arrives at different hiring and investment rates for GREEN and PURPLE workers. The answer in our paper is *path dependence*. Our experiment starts with a “seed” stage, in which we make investment more costly for PURPLE workers. The seed stage may then affect beliefs in later, symmetric, rounds of play. Firms and PURPLE workers may coordinate on the Pareto dominated equilibrium because their beliefs are anchored in a history in which the outcome was (Not Invest, Not Hire).

The difference in investment costs at the seed stage is certainly a simplification, and is meant to capture a combination of possible real-world differences between people of different genders, races, and ethnicities. One possibility is that it captures literal differences in investment costs for members of disadvantaged and minority groups; for example, differences in the cost of schooling. The difference in investment cost may also reflect other difficulties in accessing the labor market, including those arising from preference-based discrimination—historical hiring rates may have been lower for some groups.

## 3.2 Overview of Treatments

For the purpose of describing the experiment, we define some standard terminology. Our experiment consists of two players playing a simultaneous-move game. Each play of the game is a **round**. Experimental participants play a number of rounds in one roughly one-hour long sitting: each such sitting is a **session**. In a session, the same group of participants plays the game repeatedly, each participant being randomly and anonymously matched with other participants to play the two-player game. The different

sessions vary in how we set the game parameters. A **treatment** is one specification of game parameters.

Our experiment is comprised of three treatments, where each treatment is run in several sessions. In all of our treatments, each session has four stages. Each stage consists of a fixed number of rounds. At the beginning of the session, participants are randomly assigned to the role of either a worker or a firm. Workers are also randomly assigned a color: GREEN or PURPLE. Both a participant's role and color, if applicable, are fixed across all session rounds. This feature of our experiment is important and should be emphasized: participants' color is fixed throughout their participation in the experiment, as is their role in the game.

Across treatments and stages, rounds follow a fixed protocol. In each round, workers and firms are randomly matched in pairs to play an investment/hiring game, shown in Table 1.<sup>8</sup> Each worker decides whether to invest in costly training, where  $c$  represents the cost of investment. Each firm, knowing the color of the worker she is paired with, decides whether to hire the worker. Both worker and firm make their decisions simultaneously, without observing each other's decision. As we describe in the next subsection, the firm's profit from hiring a worker can depend on the stage, the worker's color, and the worker's decision to invest or not. The determination of firms' profits varies across our three treatments. In each round, workers and firms are also asked to report their beliefs about the other player's decision. For instance, a worker is asked to report her belief about how likely it is that the firm chose to hire her. Similarly, a firm is asked to report her belief about how likely it is that the worker chose to invest in training. We use the binarized scoring rule of Hossain and Okui (2013) to incentivize belief elicitation. The binarized scoring rule is incentive compatible even for decision makers who are not risk neutral. Before making their decisions, all participants can observe a public history

---

<sup>8</sup>All payoffs in the experiment are expressed in tokens, where 1 token = \$0.01.

consisting of four moving averages: (1) GREEN workers' average investment rate, (2) PURPLE workers' average investment rate, (3) firms' average hiring rate when paired with a GREEN worker, and (4) firms' average hiring rate when paired with a PURPLE worker. These averages are based on the decisions of all participants across all previous rounds of the session. For maximum clarity, these averages are reported graphically and numerically. At the end of each round, participants observe both players' decisions and payoffs.<sup>9</sup>

At the end of the experiment, participants complete two risk elicitation tasks and one survey questionnaire. Specifically, we use the risk elicitation task from Gneezy and Potters (1997). Each participant is given a token endowment and decides how many tokens to invest in a risky project with a known chance of success. We use duplicate elicitations in order to account for measurement error, see Gillen, Snowberg, and Yariv (2018). The survey questionnaire contains basic demographic, academic, and lifestyle questions.<sup>10</sup>

### 3.3 Stages of the Experiment

We now describe the four stages of each session.

#### Stage 1: Seed Stage

In the first stage of each session (Rounds 1 - 10), GREEN and PURPLE workers face different costs of investment. For a GREEN worker, the cost of investment is 200 tokens ( $c = 200$  in Table 1). For a PURPLE worker, the cost of investment is 600 tokens ( $c = 600$ ). These parameters generate two different games depending on the particular worker-

---

<sup>9</sup>The full set of instructions is available here: [http://www.leeatyariv.com/papers/Discrimination\\_Instructions.pdf](http://www.leeatyariv.com/papers/Discrimination_Instructions.pdf)

<sup>10</sup>Survey details are available at: [http://www.leeatyariv.com/papers/Discrimination\\_Demographics.pdf](http://www.leeatyariv.com/papers/Discrimination_Demographics.pdf)

firm matching. Notably, PURPLE workers have a dominant strategy of not investing. When a firm is paired with a PURPLE worker, iterated elimination of strictly dominated strategies yields (Not Invest, Not Hire) as the unique outcome; (Not Invest, Not Hire) is therefore the unique Nash equilibrium of the game. When a firm is paired with a GREEN worker, however, the strategic environment is a coordination game with two pure-strategy Nash equilibria: (Invest, Hire) and (Not Invest, Not Hire). The (Invest, Hire) equilibrium is Pareto-dominant. For each participant, playing their component of the Pareto-dominant equilibrium is a best response if the other participant is playing their component of the equilibrium with a probability of at least  $\frac{2}{3}$ . Stage 1 serves to “seed” discrimination between GREEN and PURPLE workers.

### **Stage 2: Baseline Stage**

In the second stage of each session (Rounds 11 - 20), investment costs are equalized. Both GREEN and PURPLE workers now face an investment cost of 200 tokens ( $c = 200$ ). Statistical discrimination occurs if participants coordinate on the Pareto-dominant (Invest, Hire) equilibrium when a firm plays against a GREEN worker, while participants coordinate on the Pareto-dominated (Not Invest, Not Hire) equilibrium when a firm plays against a PURPLE worker. The purpose of Stage 2 is twofold. The first is to test whether statistical discrimination emerges in the lab, once “seeded” by Stage 1. The second is to assess the extent to which it is alleviated over time, as the game with symmetric investment cost is played repeatedly.

### **Stage 3: Introducing Affirmative Action**

In the third stage of each session, we maintain equal investment costs for both colors, but we implement an affirmative-action policy to incentivize the hiring of PURPLE workers. The policy takes the form of a subsidy  $s$  for any firm that chooses to hire a PURPLE



worker, regardless of whether or not the worker invested in training. Our experimental treatments vary the size of the subsidy and the length of this stage.

- **Subsidy:** For a period of 10 rounds (Rounds 21 - 30), each firm that hires a PURPLE worker earns an additional payment of 200 tokens ( $s = 200$ ). For a firm, hiring a PURPLE worker is now a best response if the worker is investing with a probability of at least  $\frac{1}{2}$ .
- **High Subsidy:** For a period of 10 rounds (Rounds 21 - 30), each firm that hires a PURPLE worker earns an additional payment of 900 tokens ( $s = 900$ ). For a firm, hiring a PURPLE worker is now a dominant strategy. Observe that by making the hiring of PURPLE workers a dominant strategy, the High Subsidy treatment is the mirror image of the seed stage.
- **Long Subsidy:** For a period of 20 rounds (Rounds 21 - 40), each firm that hires a PURPLE worker earns an additional payment of 200 tokens ( $s = 200$ ). For a firm, hiring a PURPLE worker is now a best response if the worker is investing with a probability of at least  $\frac{1}{2}$ .

#### **Stage 4: Removing Affirmative Action**

In the fourth stage of each session (Subsidy and High Subsidy: Rounds 31 - 40, Long Subsidy: Rounds 41 - 50), the subsidy is removed. The parameters of Stage 4 are identical to those of Stage 2.

### **3.4 Implementation**

All experimental sessions were run at the Experimental Social Science Laboratory (ESSL) at UC Irvine. A total of 268 participants participated in 15 sessions. A summary of the

	Subsidy	High Subsidy	Long Subsidy
Length of Experiment	40 rounds	40 rounds	50 rounds
Length of Stage 3	10 rounds	10 rounds	20 rounds
Subsidy for Hiring GREEN (Stage 3)	$s = 0$	$s = 0$	$s = 0$
Beliefs for Hiring GREEN (Stage 3)	$p_{Invest} \geq \frac{2}{3}$	$p_{Invest} \geq \frac{2}{3}$	$p_{Invest} \geq \frac{2}{3}$
Subsidy for Hiring PURPLE (Stage 3)	$s = 200$	$s = 900$	$s = 200$
Beliefs for Hiring PURPLE (Stage 3)	$p_{Invest} \geq \frac{1}{2}$	$p_{Invest} \geq 0$	$p_{Invest} \geq \frac{1}{2}$
Number of Sessions	5	5	5
Number of Participants	88	84	96

Table 2: A summary of our experimental treatments.

treatments and corresponding sessions appears in Table 2. Each session lasted approximately one hour. Each participant’s earnings were the sum of a \$7 show-up payment, their payoff from one randomly selected experimental round, and their payoff from one randomly selected risk-elicitation task. Average participant earnings were \$24.46 (including the show-up payment). The experiment was programmed and conducted using the oTree software (Chen, Schonger, and Wickens, 2016).

## 4 Experimental Outcomes

We describe our results using the entire set of data. In the Online Appendix, we replicate the analysis using the last five rounds of each stage, in order to account for potential learning effects. The analysis using the last five rounds yields identical conclusions.

### 4.1 Outcomes Across Stages

#### Inducing Statistical Discrimination (Stages 1 and 2)

Stages 1 and 2, in which we seed discriminatory beliefs and then assess the resulting behavior, are shared across our treatments. We first present the data from these two stages, pooled across all three treatments. Figure 2 shows investment and hiring rates

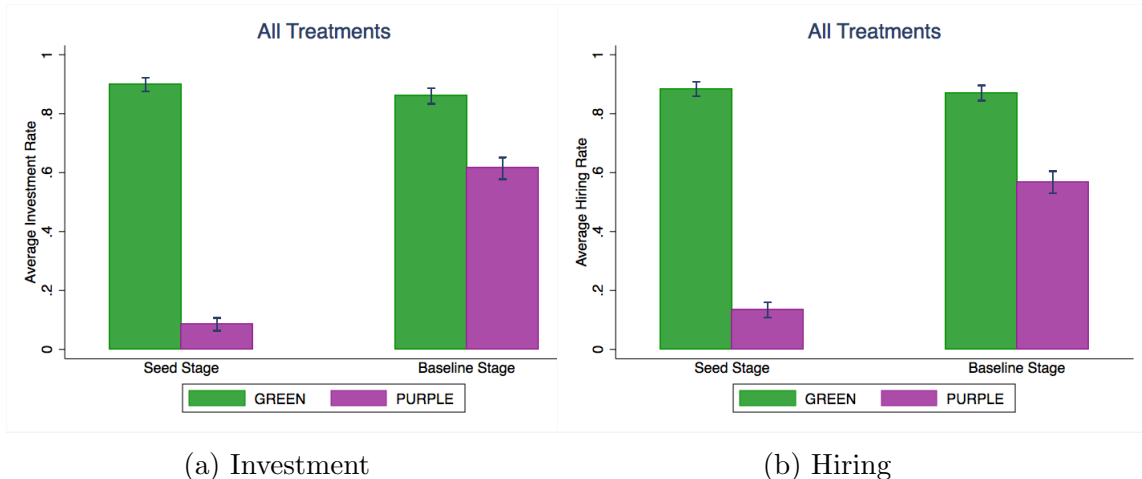


Figure 2: Investment/hiring decisions before affirmative action

by worker color. The sharp differences in Stage 1, the seeding stage, in which investment costs are unequal across worker types, are not surprising. Indeed, in this stage, PURPLE workers have a dominant strategy of not investing. For Stage 2, the baseline stage, in which investment costs are equalized across workers, 86% (576/670) of GREEN workers and 61% (412/670) of PURPLE workers invest in training. Similarly, 87% (583/670) of firms hire GREEN workers while 57% (380/670) of firms hire PURPLE workers. Both differences are statistically significant at conventional levels ( $p < 0.001$ ). These baseline differences in hiring and investment rates for workers of different colors reflect statistical discrimination.

We note that even when restricting attention to the very last round of the baseline stage, Round 20, when participants have the most experience playing the investment/hiring game with equal worker costs, we find similar patterns.<sup>11</sup> The discriminatory outcomes we observe are persistent.

<sup>11</sup>Specifically, in Round 20, 82% (55/67) of GREEN workers invest in training, while only 52% (35/67) of PURPLE workers invest in training. Similarly, 79% (53/67) of firms hire GREEN workers, while only 43% (29/67) of firms hire PURPLE workers. Both of these differences across worker types are statistically significant at conventional levels ( $p < 0.001$ ).

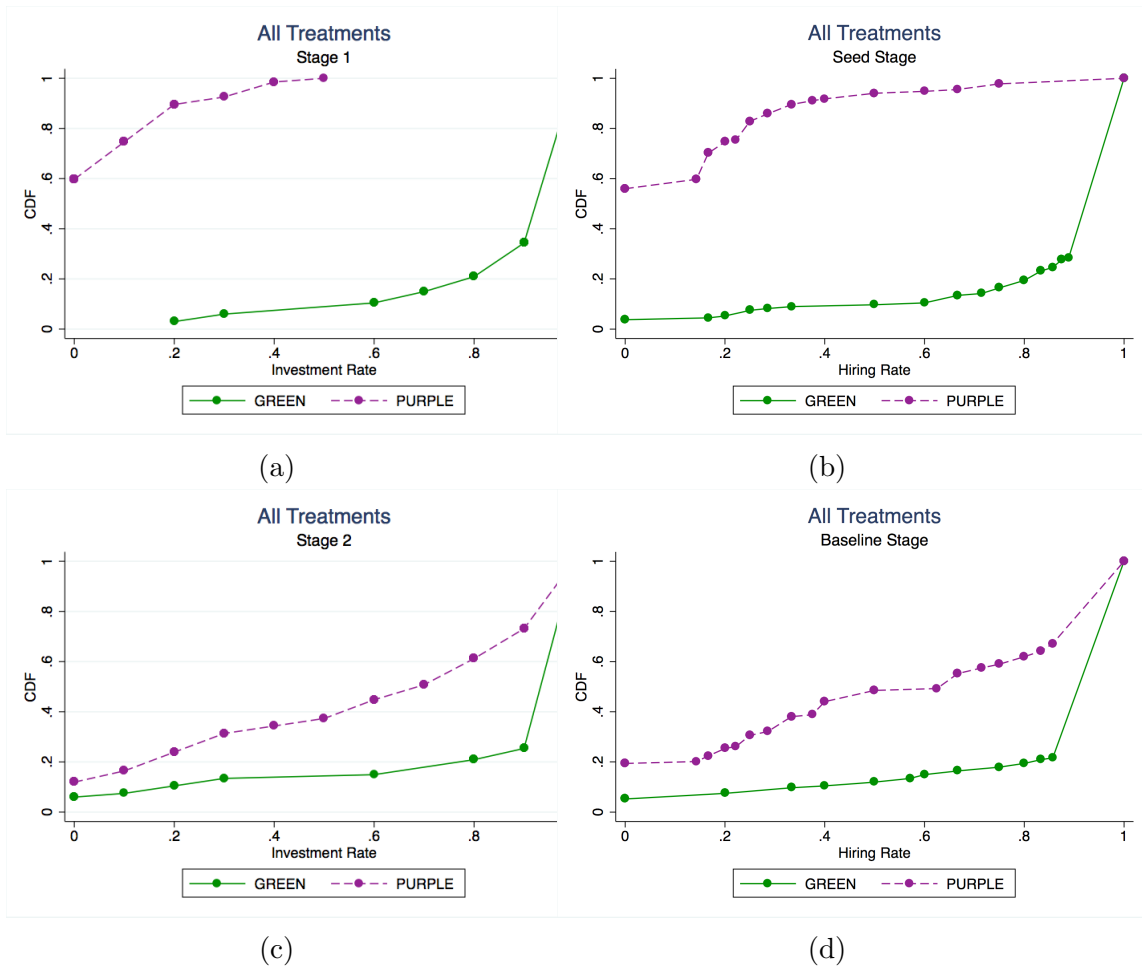


Figure 3: Empirical CDFs of participant-level behavior before affirmative action

We now look more closely at individual workers' investment decisions. Combining the data across the three treatments generates 134 individual-level observations pertaining to workers per stage.<sup>12</sup> In the seed stage, Stage 1, 60% (40/67) of PURPLE workers consistently play their dominant strategy and do not invest in any round, while 66% (44/67) of GREEN workers invest in all rounds. When the investment costs are equalized in the baseline stage, Stage 2, only 27% (18/67) of PURPLE workers invest in all rounds while 75% (50/67) of GREEN workers invest in all rounds. The empirical cumulative distribution functions (CDFs) of participant-level investment rates are shown in Figure 3 (panels a and b). In both stages, it is clear that the empirical distribution for GREEN workers first-order stochastically dominates the empirical distribution for PURPLE workers.<sup>13</sup>

We conduct a similar exercise with respect to firms' hiring decisions. Since firms interact with both worker types in all stages, we calculate two different hiring rates for each firm: their average hiring rate when paired with GREEN workers and their average hiring rate when paired with PURPLE workers. The empirical CDFs of participant-level hiring rates are also shown in Figure 3 (panels c and d). Consistent with our observations pertaining to workers' behavior, the empirical distribution for hiring GREEN workers first-order stochastically dominates the empirical distribution for hiring PURPLE workers in both stages.<sup>14</sup>

To summarize, individual analysis paints a similar picture to that produced by our aggregate analysis above.

The results described so far utilize all sessions in our experiments, but the rest of the paper uses a subset. When considering our data session-by-session, we see that in 3 of

---

<sup>12</sup>For each stage, an observation corresponds to a worker's average investment rate across the ten rounds of the stage.

<sup>13</sup>For each stage, we can also reject the null hypothesis that the average investment rates for GREEN and PURPLE workers come from the same underlying theoretical distribution with a Kolmogorov-Smirnov test yielding  $p < 0.001$ .

<sup>14</sup>We can also reject the null hypothesis that the average hiring rates for GREEN and PURPLE workers come from the same distribution with a Kolmogorov-Smirnov test yielding  $p < 0.001$ .

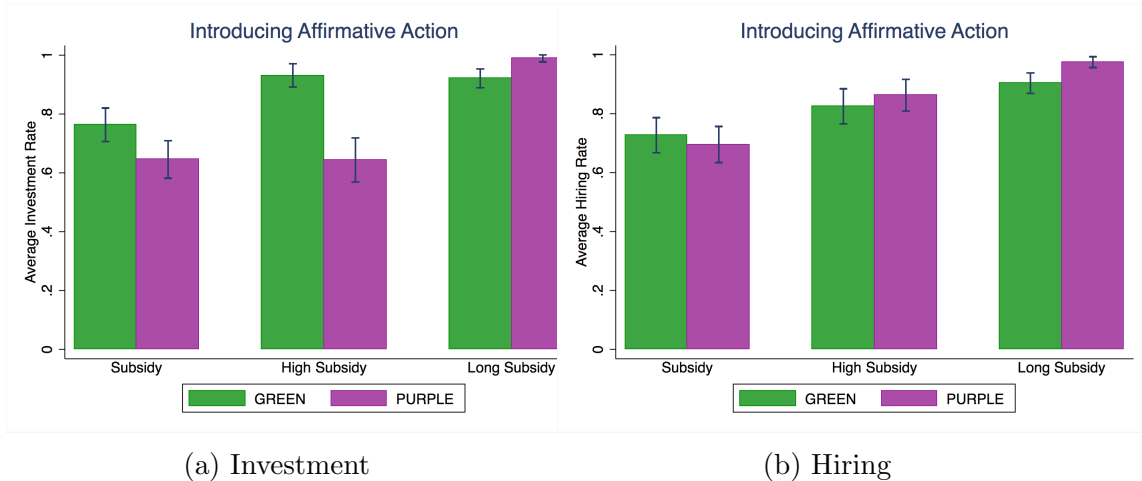


Figure 4: Investment/hiring decisions during affirmative action

our 15 experimental sessions, the differences in hiring rates across GREEN and PURPLE workers in the baseline stage are not statistically significant at the 1% level.<sup>15</sup> That is, in 3 of our sessions, the mere equalization of investment costs across workers is sufficient for generating equitable outcomes (or at least outcomes that are not statistically significantly unequal). Our study focuses on investigating the efficacy of affirmative-action policies in reversing statistical discrimination: For this reason, when we turn to the analysis of affirmative action, we concentrate on sessions in which statistical discrimination is induced. Thus, the analysis that follows concentrates on the 12 sessions in which statistical discrimination is observed and significant in the baseline stage. That said, we emphasize that our qualitative conclusions do not change if we include all sessions in the analysis—the Online Appendix reproduces our main results using all 15 sessions.

## Effects of Affirmative Action

### Introducing Affirmative Action (Stage 3)

In Stage 3 of each experimental treatment, we introduce an affirmative-action policy in which each firm who hires a PURPLE worker is paid an additional subsidy. The size of the subsidy and its duration vary across our three treatments.

Figure 4 shows investment and hiring rates by worker color. Compared to the rates in Figure 2, we see a decrease in discriminatory outcomes, but there are significant differences between our three variants of affirmative action when it comes to workers' investment behavior. A longer period of affirmative action, but not a larger subsidy, is successful in reversing the investment patterns observed in the first two stages of the experiment. Specifically, when the affirmative-action subsidy lasts for 20 rounds in the Long Subsidy treatment, PURPLE workers invest at a higher, and significantly different, rate than GREEN workers ( $p < 0.001$ ). However, in both the Subsidy and High Subsidy treatments, where affirmative action lasts only 10 rounds, GREEN workers invest at higher and significantly different rates than PURPLE workers (Subsidy:  $p = 0.007$ , High Subsidy:  $p < 0.001$ ). In other words, *PURPLE workers do not fully internalize firms' responses to subsidies.*

The three affirmative-action policies we test are effective in manipulating firms' hiring behavior. In the Subsidy treatment (10 rounds,  $s = 200$ ), 73% (160/220) of firms hire GREEN workers and 70% (153/220) of firms hire PURPLE workers. Furthermore, the difference in average hiring rates of GREEN and PURPLE workers is no longer statistically significant ( $p = 0.463$  with a two-sided t-test). Intensifying the affirmative-action policy by increasing the size of the subsidy (from  $s = 200$  to  $s = 900$  in High Subsidy)

---

<sup>15</sup>Of the three sessions in which the hiring differences are not statistically significant, one session belongs to the High Subsidy treatment and two sessions belong to the Long Subsidy treatment.

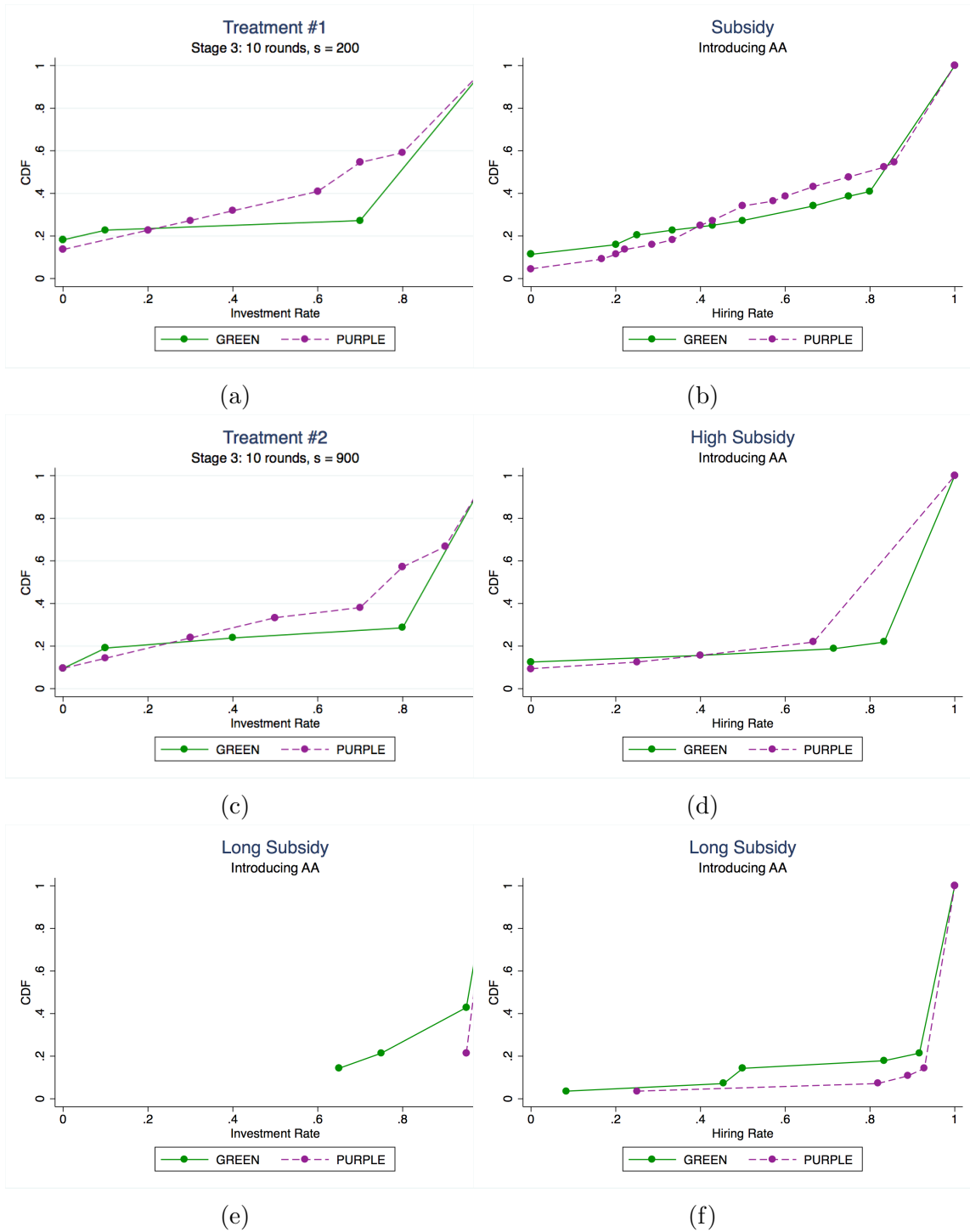


Figure 5: Empirical CDFs of participant-level behavior during affirmative action



increases hiring rates, without exhibiting discrimination of workers. Indeed, in the High Subsidy treatment (10 rounds,  $s = 900$ ), 83% (132/160) of firms hire GREEN workers while 86% (138/160) of firms hire PURPLE workers.<sup>16</sup> The difference in hiring rates is, again, not statistically significant ( $p = 0.357$ ). In the Long Subsidy treatment (20 rounds,  $s = 200$ ), 90% (253/280) of firms hire GREEN workers, while 98% (273/280) of firms hire PURPLE workers. In this case, we are able to reject the null hypothesis that the average hiring rate is the same for GREEN and PURPLE workers ( $p < 0.001$ ).

The individual-level data further confirm that affirmative action is largely successful in reversing the previously-observed differences across worker colors. Figure 5 shows the empirical CDFs of participant-level investment and hiring rates while affirmative action is in place. For two treatments, we cannot reject the null hypothesis that GREEN and PURPLE workers' average investment rates come from the same distribution.<sup>17</sup> Furthermore, for all treatments, we cannot reject the null hypothesis that the average hiring rates of GREEN and PURPLE workers come from the same distribution.<sup>18</sup>

### Removing Affirmative Action (Stage 4)

In Stage 4 of each treatment, we remove the affirmative-action subsidy. The parameters of this stage are then identical to those of our baseline stage, Stage 2. The purpose of this stage is to assess whether the benefits of affirmative action persist after the policy is lifted.

Figure 6 depicts investment and hiring rates by worker color. In terms of both in-

---

<sup>16</sup>With  $s = 900$ , it is a dominant strategy for a firm to hire a PURPLE worker. We can reject the null hypothesis that firms are consistently playing their dominant strategy ( $p < 0.001$  with a one-sided t-test).

<sup>17</sup>Using a Kolmogorov-Smirnov test, we have the following p-values across treatments. Subsidy:  $p = 0.203$ , High Subsidy:  $p = 0.011$ , and Long Subsidy:  $p = 0.863$ .

<sup>18</sup>Using a Kolmogorov-Smirnov test, we have the following p-values across treatments. Subsidy:  $p = 0.814$ , High Subsidy:  $p = 1.000$ , and Long Subsidy:  $p = 0.976$ .

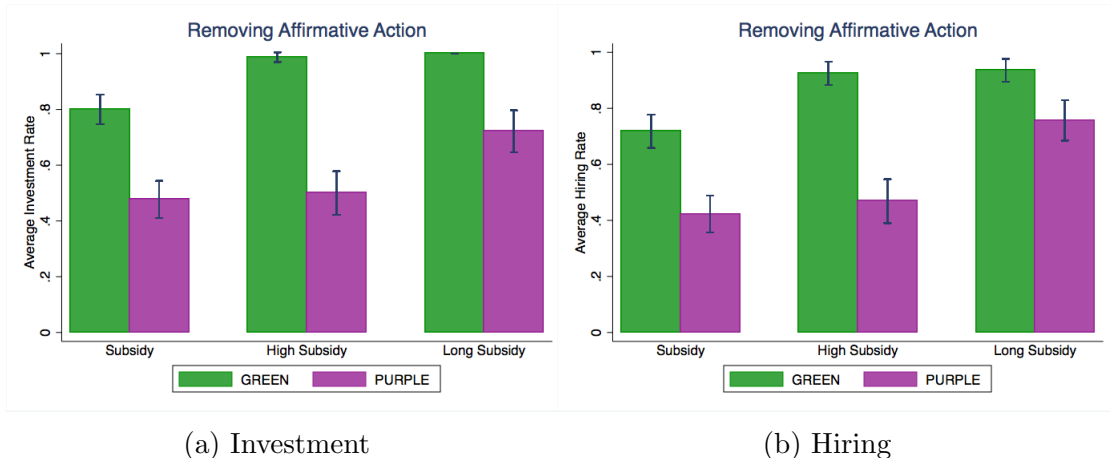


Figure 6: Investment/hiring decisions after affirmative action

vestment and hiring decisions, we observe a surprising reversion to pre-affirmative action patterns of behavior. In the Subsidy treatment, 80% (176/220) of GREEN workers and 48% (105/220) of PURPLE workers invest in training. After a larger affirmative-action subsidy in the High Subsidy treatment, we find that 99% (158/160) of GREEN workers and 50% (80/160) of PURPLE workers invest in training. Both of these differences in average investment rates between GREEN and PURPLE workers are statistically significant ( $p < 0.001$ ). When the duration of affirmative action is increased from 10 rounds to 20 rounds in our Long Subsidy treatment, the reversion to previous discrimination is weaker: 100% (140/140) of GREEN workers and 72% (101/140) of PURPLE workers invest in training. However, we can still reject the null hypothesis that GREEN and PURPLE workers invest in training at the same rate ( $p = 0.001$ ).

Firms' hiring decisions exhibit a similar pattern. In the Subsidy treatment, 72% (158/220) of firms hire GREEN workers while only 42% (93/220) of firms hire PURPLE workers. In the High Subsidy treatment, where the affirmative-action subsidy is substantially higher, 93% (148/160) of firms hire GREEN workers while only 47% (75/160) of firms hire PURPLE workers. Both of these differences in average hiring rates between GREEN and PURPLE workers are statistically significant ( $p < 0.001$ ). With a longer period of affirmative action in the Long Subsidy treatment, we once again see that the

deleterious effects of removing the subsidy are less severe: 94% (131/140) of firms hire GREEN workers and 76% (106/140) of firms hire PURPLE workers. However, we can still reject the null hypothesis that GREEN and PURPLE workers are hired at the same rate ( $p < 0.001$ ).

Figure 7 displays the empirical CDFs of participant-level behavior, both investment and hiring, after the affirmative-action policy is lifted. In all cases, the empirical distribution for GREEN workers either fully or nearly first-order stochastically dominates the corresponding empirical distribution for PURPLE workers. This provides further evidence that discrimination in favor of GREEN workers persists in our experimental labor markets—even long after the explicit advantage of GREEN workers has been eliminated. With regard to investment decisions, we can reject the null hypothesis that GREEN and PURPLE workers’ average investment rates come from the same distribution in the Subsidy and High Subsidy treatments, but not in the Long Subsidy treatment.<sup>19</sup> With regard to hiring decisions, we can also reject the null hypothesis that firms’ average hiring rates for GREEN and PURPLE workers come from the same distribution in the Subsidy and High Subsidy treatments.<sup>20</sup> These observations further testify that a longer duration of affirmative action can mitigate the reversal of the policy gains when the policy is eventually lifted.<sup>21</sup>

We note that there is a striking asymmetry between the effects of the seeding and affirmative action stages. Seeding creates a long-lasting difference in how GREEN and PURPLE workers behave, and how firms treat them. The affirmative-action policy is similar in that it modifies the payoffs to the firms to undo statistical discrimination

---

<sup>19</sup>Specifically, using the Kolmogorov-Smirnov test we get the following p-values across treatments. Subsidy:  $p = 0.006$ , High Subsidy:  $p = 0.003$ , and Long Subsidy:  $p = 0.055$ .

<sup>20</sup>Using the Kolmogorov-Smirnov test we get the following p-values across treatments. Subsidy:  $p = 0.001$ , High Subsidy:  $p < 0.001$ , and Long Subsidy:  $p = 0.944$ .

<sup>21</sup>While we focus on statistical discrimination alone, we suspect that the presence of preference-based discrimination in environments such as ours would only exacerbate the outcomes we report.

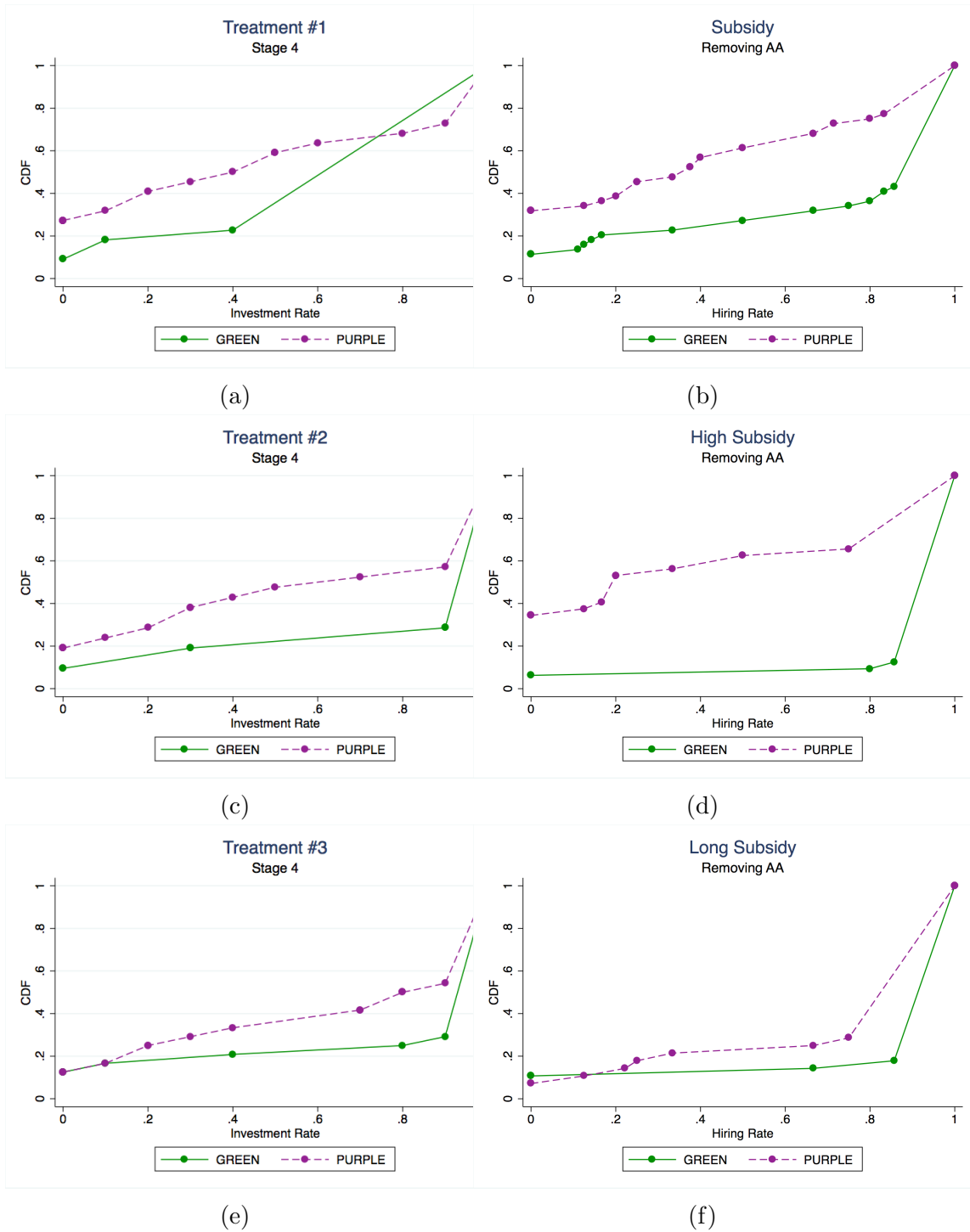


Figure 7: Empirical CDFs of participant-level behavior after affirmative action

by favoring the hiring of PURPLE workers. However, this policy is nowhere close to being as effective as the initial seeding. This observation is particularly interesting with respect to our High Subsidy treatment, which effectively mirrors the seeding stage—it lasts for an identical period of time and involves a dominant strategy for firms of hiring PURPLE workers. Nonetheless, when the High Subsidy affirmative-action policy is lifted, discriminatory outcomes reappear.

## 4.2 Choice from a Menu

In our experiments, each firm is randomly paired with a particular worker and faces a binary hiring decision. In a labor market setting, however, firms are typically confronted with the choice of *which* worker to hire from a menu of workers. The efficacy of an affirmative-action policy would then be measured by its ability to nudge firms to hire certain types of workers from a menu containing their favored workers. Our data allow us to deduce expected choices from such menus of workers with and without affirmative-action policies in place. That is, we can use our data to deduce how often firms would hire a PURPLE worker when given the choice of hiring either type of worker.

In each round of the experiment, each firm is asked to report her belief about the likelihood that the worker she is paired with chose to invest in training. Similarly, each worker is asked to report her belief about the likelihood that the firm she is paired with chose to hire her.<sup>22</sup> Using these belief elicitation, we first calculate each firm’s average reported beliefs for GREEN and PURPLE workers’ investment rates in each stage of the experiment. Using these, we then calculate each firm’s expected payoffs from hiring GREEN and PURPLE workers in each stage of the experiment. In the following section, we show that firms best respond to their reported beliefs. We therefore assume that, when

---

<sup>22</sup>As mentioned earlier, we use the binarized scoring rule of Hossain and Okui (2013) to incentivize belief elicitation. The binarized scoring rule is incentive-compatible even for decision-makers who are not risk-neutral.

	Seed Stage	Baseline Stage	Introducing AA	Removing AA
Subsidy	0.00	0.05	0.48	0.07
High Subsidy	0.06	0.03	1.00	0.03
Long Subsidy	0.07	0.11	0.93	0.25

Table 3: Fraction of firms hiring a PURPLE worker over a GREEN worker

faced with a menu, a firm would hire a PURPLE worker only if her expected payoff of hiring a PURPLE worker is strictly greater than her expected payoff of hiring a GREEN worker. Table 3 shows the fraction of firms hiring a PURPLE worker over a GREEN worker under these assumptions (using the shorthand AA for affirmative action).

Consistent with our previous findings, we see that PURPLE workers would be hired at a significantly higher rate when affirmative action is introduced ( $p < 0.001$  in all three treatments). However, when affirmative action is removed, the hiring rate for PURPLE workers would decline substantially. Furthermore, in all three treatments, the hiring rate for PURPLE workers after the policy intervention would not be significantly different than prior to the policy intervention ( $p = 0.65$  for Subsidy;  $p = 1.00$  for High Subsidy;  $p = 0.17$  for Long Subsidy).

To conclude, by extrapolating our belief elicitation data, we derive predictions for a richer market environment that echo our main findings. Statistically discriminated-against workers, who compete against more initially “desirable” workers, benefit from affirmative-action policies, but only while those policies are in place.

## 5 Beliefs as a Channel for Persistence

So far, we have focused on the binary outcomes of our experimental interactions (invest or not for workers; hire or not for firms). We now investigate the evolution of partici-

	Seed Stage		
	Subsidy	High Subsidy	Long Subsidy
GREEN workers	0.85/0.84	0.74/0.89	0.79/0.89
PURPLE workers	0.93/0.93	0.87/0.87	0.94/0.94
Firms paired with GREEN	0.80/0.80	0.74/0.83	0.81/0.89
Firms paired with PURPLE	0.90/0.86	0.84/0.82	0.89/0.91

	Baseline Stage		
	Subsidy	High Subsidy	Long Subsidy
GREEN workers	0.90/0.86	0.89/0.99	0.92/0.98
PURPLE workers	0.72/0.50	0.73/0.51	0.72/0.26
Firms paired with GREEN	0.88/0.89	0.86/0.96	0.91/0.99
Firms paired with PURPLE	0.69/0.48	0.74/0.47	0.69/0.34

	Introducing Affirmative Action		
	Subsidy	High Subsidy	Long Subsidy
GREEN workers	0.93/0.76	0.88/0.93	0.82/0.92
PURPLE workers	0.81/0.35	0.85/0.36	0.97/0.09
Firms paired with GREEN	0.85/0.76	0.88/0.83	0.89/0.90
Firms paired with PURPLE	0.77/0.30	0.86/0.86	0.94/0.75

	Removing Affirmative Action		
	Subsidy	High Subsidy	Long Subsidy
GREEN workers	0.87/0.79	0.89/0.99	0.96/1.00
PURPLE workers	0.85/0.52	0.91/0.63	0.86/0.68
Firms paired with GREEN	0.85/0.80	0.90/0.93	0.96/0.94
Firms paired with PURPLE	0.80/0.58	0.78/0.61	0.80/0.59

Table 4: Fraction of best-responses to reported beliefs/public histories

pants' beliefs throughout our experiments. As we will see, beliefs are rather sticky, and while agents respond to the pecuniary incentives subsidies introduce, their beliefs are not altered dramatically.

We first inspect the linkages between beliefs and actions in our experiments. We consider the question of whether participants are playing best-response strategies, both with respect to their reported beliefs and with respect to the public history of play. Table 4 shows the breakdown of best-response rates by participant role/color, treatment, and stage. As can be seen, participants' actions are largely optimal given their reported beliefs. In aggregate, 85% (7,534/8,880) of participants' actions are a best response to their reported beliefs.<sup>23</sup> Participants best respond to the public history at far lower rates, standing at 73% (6,483/8,880). This is reasonable—since the public history includes the results from all previous rounds and does not reset between stages, less sensitivity to the public history implies sensitivity to the changing parameters/incentives of the experiment.

While participants' actions are consistent with their beliefs, actions are far coarser. The evolution of beliefs throughout our sessions can provide more insight into the action choices we observe. Figure 8 illustrates workers' average beliefs across rounds and Figure 9 illustrates firms' average beliefs across rounds, along with the average public history that was visible to participants on the experimental interface.

Reported beliefs roughly reflect trends of the public history, but differ in magnitudes and slopes. They are slightly less optimistic for GREEN workers and substantially more optimistic for PURPLE workers. In particular, participants do not simply mimic the public history when reporting beliefs.

---

<sup>23</sup>This is consistent, if somewhat higher, than other reported statistics of best responses, see e.g. Rey-Biel (2009) and references therein.



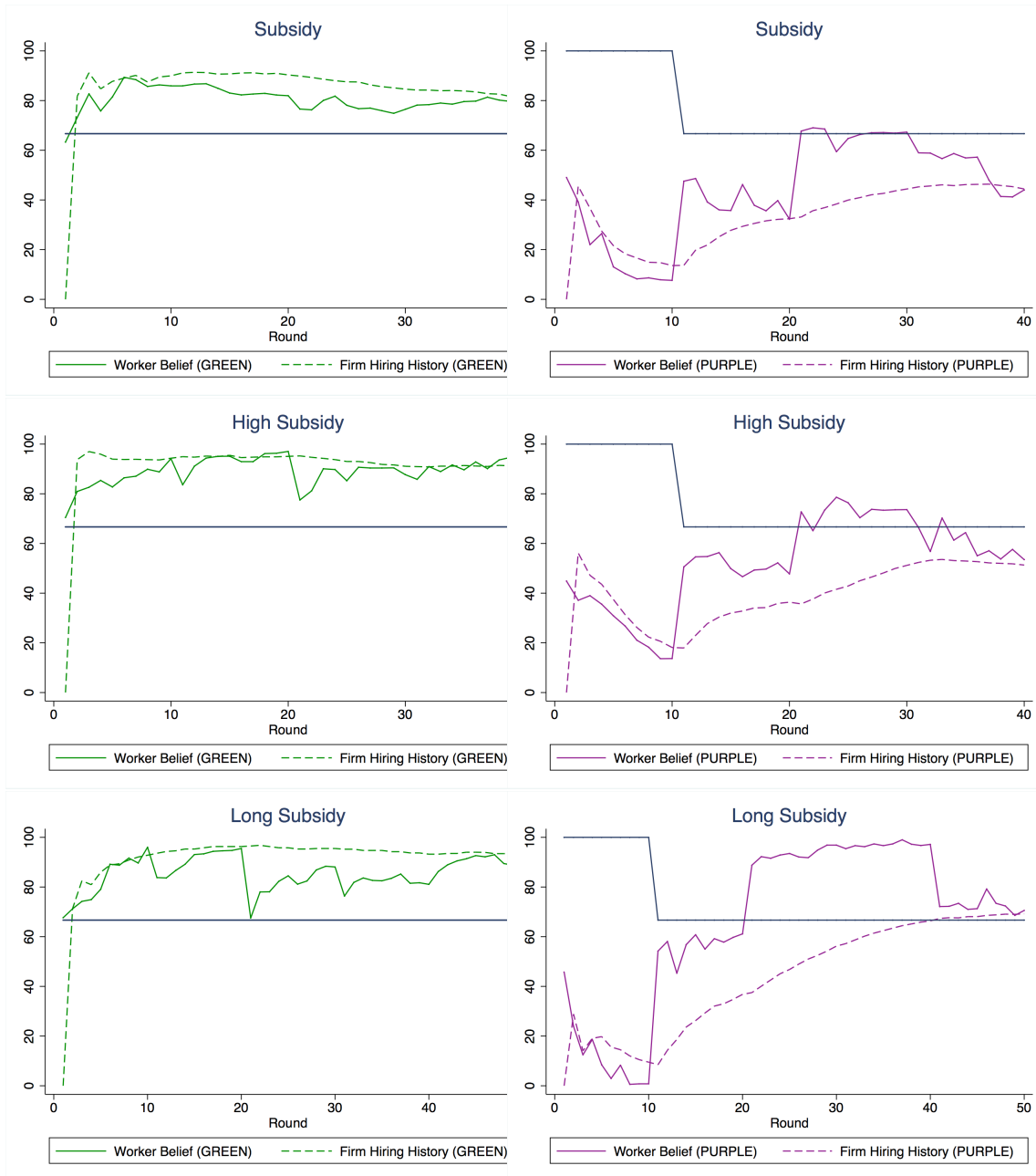


Figure 8: Workers' beliefs and firm hiring history (solid black lines illustrate threshold beliefs for investment to be optimal)

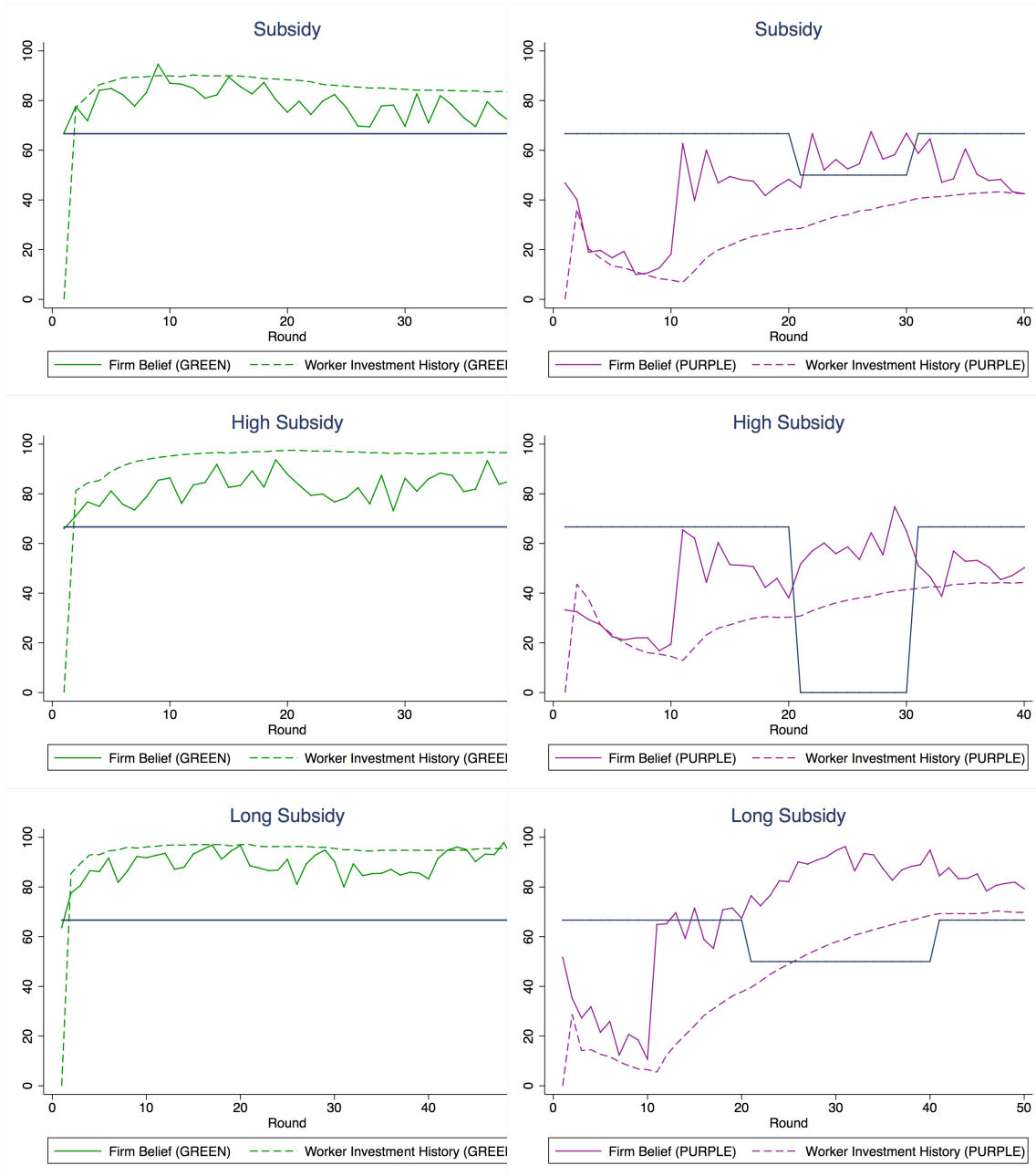


Figure 9: Firms' beliefs and worker investment history (solid black lines illustrate threshold beliefs for hiring to be optimal)

Importantly, while we see a jump in beliefs pertaining to PURPLE workers' investment after the seeding stage, beliefs remain fairly flat afterwards, with only slight increases during the affirmative-action phase, which flattens out during the affirmative-action phase of our Long Subsidy treatment. The level of beliefs is such that, absent subsidies, the decision to hire is only marginally optimal in some cases, whereas it is clearly optimal when affirmative-action incentives are in place.<sup>24</sup>

In contrast, beliefs regarding GREEN workers are slightly below those suggested by their actual investment rates. Naturally, one needs to take with care any impression that beliefs are pessimistic. Indeed, experimental investment rates by GREEN workers are rather high. Therefore, any natural experimental perturbations of those statistical rates that generate beliefs would generally shift average assessments down. Regardless, beliefs are sufficiently high as to guarantee hiring throughout the different experimental stages.

In order to quantify these effects as well as summarize the results of the paper, we estimate the following OLS regressions separately for each treatment:

$$Belief_{it} = \beta_0^B + \beta_1^B * IntroducingAA_t + \beta_2^B * RemovingAA_t + \varepsilon_{it}$$

and

$$Hire_{it} = \beta_0^H + \beta_1^H * IntroducingAA_t + \beta_2^H * RemovingAA_t + \varepsilon_{it},$$

where  $Belief_{it}$  is firm  $i$ 's reported belief (from 0 to 1) about the likelihood that her paired PURPLE worker in round  $t$  chooses to invest in training,  $Hire_{it}$  is a dummy variable

---

<sup>24</sup>One might wonder whether participants are simply coordinating on the mixed-strategy equilibrium of the game in the baseline and final stages, which would be consistent with the reported beliefs. We note that, at the individual level, behavior in these stages is inconsistent with that equilibrium (see Figure 6).

that equals 1 if firm  $i$  hires a PURPLE worker in round  $t$ ,  $IntroducingAA_t$  is a dummy variable that equals 1 if round  $t$  is in Stage 3 of the experiment (when affirmative action is introduced), and  $RemovingAA_t$  is a dummy variable that equals 1 if round  $t$  is in Stage 4 of the experiment (when affirmative action is removed). Last,  $\varepsilon_{it}$  is an error term, which we cluster by participant.

Since we are interested in the impact of affirmative action and its removal, we restrict our analysis to firms interacting with PURPLE workers in Stages 2 - 4 of the experiment.<sup>25</sup> The regression results are shown in Tables 5 and 6. The first three columns correspond to our three treatments. The fourth treatment splits the first 10 and last 10 periods of the affirmative-action stage in the Long Subsidy treatment. The corresponding regression allows us to infer whether the length of the subsidy is internalized by participants early on. Indeed, absent forward-looking behavior, the first 10 periods of the affirmative-action stage in that treatment are equivalent to the full affirmative-action stage in our Subsidy treatment.

Echoing our previous observations, affirmative action is effective in manipulating both firms' beliefs and hiring decisions while the policy is in place. For each treatment, the coefficient estimates  $\hat{\beta}_1^H$  and  $\hat{\beta}_1^B$  are positive and statistically significant. For a given treatment, we can then compare the magnitudes of  $\hat{\beta}_1^H$  and  $\hat{\beta}_1^B$  to capture the *relative* effectiveness of the affirmative-action policy in manipulating firms' actions and beliefs. Across all treatments, we see that  $\hat{\beta}_1^H > \hat{\beta}_1^B$ . That is, firms' beliefs are less responsive to the policy intervention than firms' hiring decisions.

The coefficient estimates  $\hat{\beta}_2^B$  and  $\hat{\beta}_2^H$  then measure the extent to which the benefits of affirmative action persist even after the policy is lifted. In each treatment, we find no statistically significant difference in the hiring of PURPLE workers between the affirma-

---

<sup>25</sup>As before, we only include the sessions in which we generated statistical discrimination in our baseline stage, Stage 2, of the experiment.

VARIABLE: Hire	Subsidy (1)	High Subsidy (2)	Long Subsidy (3)	Long Subsidy (4)
Introducing AA	0.173** (0.066)	0.331*** (0.062)	0.311*** (0.065)	
Removing AA	-0.100 (0.074)	-0.063 (0.060)	0.093 (0.079)	0.093 (0.079)
Introducing AA (first half)				0.307*** (0.064)
Introducing AA (second half)				0.314*** (0.067)
Constant	0.523*** (0.059)	0.531*** (0.067)	0.664*** (0.066)	0.664*** (0.066)
Observations	660	480	560	560
Number of participants	44	32	28	28

Table 5: Results from OLS regressions. Standard errors are shown in parentheses and are clustered at the individual level. Significance levels are indicated as follows: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.10$ .

tive action stage and after affirmative action is lifted. However,  $\hat{\beta}_2^H > 0$  only for the Long Subsidy treatment, albeit insignificantly so. In a similar vein, looking at firms' beliefs about PURPLE workers, the coefficient estimate  $\hat{\beta}_2^B$  is positive and, here, statistically significant only for the Long Subsidy treatment.

We note that the effect of affirmative action is significantly stronger in the first 10 periods of our Long Subsidy treatment than in the phase of 10 analogous periods in our Subsidy treatment. This suggests that participants account for the length of affirmative-action policies even when they are first introduced. This implies that placing a longer-duration affirmative-action policy has two advantages. *A longer duration increases the effectiveness of the policy even in its early introductory phase*, in addition to enhancing its long-run effects after its removal. Going back to Justice Sandra Day O'Connor's 2003 quote, it is possible that a sufficiently long phase of affirmative action, which may or may not correspond to 25 years, would assure some erosion of discriminatory attitudes both

VARIABLE: Firm Belief	Subsidy (1)	High Subsidy (2)	Long Subsidy (3)	Long Subsidy (4)
Introducing AA	0.086** (0.041)	0.085** (0.036)	0.218*** (0.037)	
Removing AA	0.022 (0.055)	-0.019 (0.037)	0.172*** (0.043)	0.172*** (0.043)
Introducing AA (first half)				0.193*** (0.035)
Introducing AA (second half)				0.244*** (0.042)
Constant	0.490*** (0.044)	0.512*** (0.048)	0.655*** (0.049)	0.655*** (0.049)
Observations	660	480	560	560
Number of participants	44	32	28	28

Table 6: Results from OLS regressions. Standard errors are shown in parentheses and are clustered at the individual level. Significance levels are indicated as follows: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.10$ .

during and after affirmative action is in place.

In our experiments, participants also respond to two risk elicitation questions as well as various demographic questions. Importantly, when adding these as controls, neither has a significant effect nor do they alter the coefficients of treatment effects described above.<sup>26</sup> The Online Appendix contains a detailed description of this analysis. In the Online Appendix, we also consider the possibility that firms that are more highly discriminatory in the baseline stage respond differently to affirmative action and its removal. We classify firms into *types* pertaining to their discriminatory tendencies in the baseline stage and illustrate that, indeed, effects are more pronounced for firms that are initially more discriminatory.

<sup>26</sup>One exception is inclusion in a minority group, which significantly reduces hiring when affirmative action is introduced and when it is lifted in our High Subsidy treatment. However, since this is the only significant effect out of our five controls in the three treatments, we suspect it might be spurious.

## 6 Discussion and Conclusion

In 1961, President John F. Kennedy first utilized the term *affirmative action* in its contemporary sense in Executive Order 10925. The intention was to have government contractors “take affirmative action to ensure that applicants are employed, and employees are treated during employment, without regard to their race, creed, color, or national origin.” Today, affirmative action refers to the leading set of laws, policies, guidelines, and administrative practices aimed at alleviating racial and gender-based discrimination. Our goal in this paper is to use an array of lab experiments to assess the effects of affirmative-action policies in combatting statistical discrimination, while in place and after they are lifted.

Our results raise questions about the long-term effectiveness of affirmative-action policies. To level the playing field, our findings suggest that affirmative-action policies need to be activated for substantial periods of time. These implications relate to the empirical regularity that affirmative-action policies tend to become entrenched, and left in place far longer than initially intended, see Sowell (2004).

Any conclusion about the long-term effects of affirmative action should be qualified. We have focused on affirmative action as an instrument for correcting statistical discrimination in labor markets, but there are other sources of discrimination, and other environments in which discrimination occurs. For instance, it is quite possible that exposing people to the rich diversity of humankind may change their beliefs through various channels—role models, exemplars, etc.—and thereby address taste-based discrimination.<sup>27</sup> There is also a fairness argument in favor of affirmative action—it serves as compensation for past discrimination. Affirmative action in education does not only seek to change expectations, its objective includes granting access to minority students

---

<sup>27</sup>For instance, Miller (2017) illustrates some positive effects of temporary affirmative-action policies that he explains through employers improving their hiring practices.

who would not otherwise be able to attend good schools or to obtain higher education (see the landmark study of Bowen and Bok (2016) for an elaborate discussion of affirmative action in education). There are also possible dynamic benefits from affirmative action. Even if affirmative action does not eliminate discrimination today, it may help future generations of minorities access better opportunities. As in our experiments, the fruits of affirmative action might simply take a long time to ripen.



## References

- Agan, Amanda Y and Sonja B Starr. 2016. “Ban the Box, Criminal Records, and Statistical Discrimination: A Field Experiment.” SSRN Working Paper.
- Anderson, Donna M and Michael J Hauptert. 1999. “Employment and Statistical Discrimination: A Hands-On Experiment.” *The Journal of Economics* 25 (1):85–102.
- Anderson, Lisa, Roland Fryer, and Charles Holt. 2006. “Discrimination: Experimental Evidence From Psychology and Economics.” *Handbook on the Economics of Discrimination* :97–118.
- Arrow, Kenneth J. 1971. “Some models of Racial Discrimination in the Labor Market.” Tech. Rep. RM-6253-RC, RAND.
- . 1973. *The Theory of Discrimination. S. 3–33 in: O. Ashenfelter/A. Rees (Hrsg.), Discrimination in Labor Markets.* Princeton University Press.
- . 1998. “What Has Economics to Say About Racial Discrimination?” *The journal of economic perspectives* 12 (2):91–100.
- Becker, Gary S. 1957. *The Theory of Discrimination.* University of Chicago Press.
- Bohren, Aislinn, Alex Imas, and Michael Rosenberg. 2017. “The Dynamics of Discrimination: Theory and Evidence.” *Mimeo* .
- Bowen, William G and Derek Bok. 2016. *The shape of the river: Long-term consequences of considering race in college and university admissions.* Princeton University Press.
- Bracha, Anat, Alma Cohen, and Lynn Conell-Price. 2005. “Affirmative Action and Stereotype Threat.” *Mimeo* .

- Chen, Daniel L, Martin Schonger, and Chris Wickens. 2016. "oTree - An Open-Source Platform for Laboratory, Online, and Field Experiments." *Journal of Behavioral and Experimental Finance* 9:88–97.
- Coate, Stephen and Glenn C Loury. 1993. "Will Affirmative-action Policies Eliminate Negative Stereotypes?" *The American Economic Review* :1220–1240.
- Ewens, Michael, Bryan Tomlin, and Liang Choon Wang. 2014. "Statistical Discrimination or Prejudice? A Large Sample Field Experiment." *Review of Economics and Statistics* 96 (1):119–134.
- Feltovich, Nick, Lata Gangadharan, and Michael P Kidd. 2013. "Implementation and Removal of an Affirmative-Action Quota: The Impact on Task Assignment and Workers' Skill Acquisition." *Canadian Public Policy* 39 (Supplement 1):S123–S140.
- Fryer, Roland G, Jacob K Goeree, and Charles A Holt. 2005. "Experience-based Discrimination: Classroom Games." *The Journal of Economic Education* 36 (2):160–170.
- Gillen, Ben, Erik Snowberg, and Leeat Yariv. 2018. "Experimenting with Measurement Error: Techniques With Applications to The Caltech Cohort Study." *Journal of Political Economy* Forthcoming.
- Glover, Dylan, Amanda Pallais, and William Pariente. 2017. "Discrimination as a Self-Fulfilling Prophecy: Evidence from French Grocery Stores." *Quarterly Journal of Economics* 132 (3):1219–1260.
- Gneezy, Uri and Jan Potters. 1997. "An Experiment on Risk Taking and Evaluation Periods." *Quarterly Journal of Economics* 112 (2):631–645.
- Haan, Thomas, Theo Offerman, and Randolph Sloof. 2015. "Discrimination in the Labour Market: The Curse of Competition between Workers." *The Economic Journal* .

- Hossain, Tanjim and Ryo Okui. 2013. "The Binarized Scoring Rule." *Review of Economic Studies* 80 (3):984–1001.
- Kidd, Michael P, Paul S Carlin, and Jonathan Pot. 2008. "Experimenting with Affirmative Action: The Coate and Loury Model." *Economic Record* 84 (266):322–337.
- Knowles, John, Nicola Persico, and Petra Todd. 2001. "Racial Bias in Motor Vehicle Searches: Theory and Evidence." *Journal of Political Economy* 109 (1):203–229.
- Krueger, Alan, Jesse Rothstein, and Sarah Turner. 2006. "Race, Income, and College in 25 Years: Evaluating Justice O'Connor's Conjecture." *American Law and Economics Review* 8 (2):282–311. URL <http://dx.doi.org/10.1093/aler/ah1004>.
- Miller, Conrad. 2017. "The Persistent Effect of Temporary Affirmative Action." *American Economic Journal: Applied Economics* 9 (3):152–90.
- Niederle, Muriel, Carmit Segal, and Lise Vesterlund. 2013. "How Costly is Diversity? Affirmative Action in Light of Gender Differences in Competitiveness." *Management Science* 59 (1):1–16.
- Phelps, Edmund S. 1972a. "The Statistical Theory of Racism and Sexism." *The American Economic Review* 62 (4):659–661. URL <http://www.jstor.org/stable/1806107>.
- Phelps, Edmund S. 1972b. "The Statistical Theory of Racism and Sexism." *American Economic Review* 62 (4):659–661.
- Rey-Biel, Pedro. 2009. "Equilibrium Play and Best Response to (Stated) Beliefs in Normal Form Games." *Games and Economic Behavior* 65 (2):572–585.
- Romero, Julian. 2015. "The Effect of Hysteresis on Equilibrium Selection in Coordination Games." *Journal of Economic Behavior & Organization* 111:88–105.
- Sander, Richard H. 2004. "A Systematic Analysis of Affirmative Action in Law Schools." *Stanford Law Review* 57 (367-483).

Sander, Richard H. and Stuart Taylor. 2012. *Mismatch: How Affirmative Action Hurts Students Its Intended to Help, and Why Universities Wont Admit It*. Basic Books.

Sowell, Thomas. 2004. *Affirmative Action Around the World: An Empirical Study*. Yale University Press.