American Economic Journal: Applied Economics 2024, 16(4): 288–321
https://doi.org/10.1257/app.20220466

Experience-Based Discrimination†

By Louis-Pierre Lepage*

I study discrimination arising from individual experiences of employers with worker groups. I present a model in which employers are uncertain about the productivity of one of two groups and learn through hiring. Positive experiences lead to positive biases, which correct themselves by leading to more hiring and learning. Negative experiences decrease hiring and learning, preserving negative biases, which can cause persistent discrimination. The model explains prejudice as incorrect statistical discrimination and generates novel predictions and policy implications. I then illustrate experience-based discrimination in an experimental labor market, finding support for key model predictions. (*JEL* D83, J23, J24, J31, J71, M51)

Evidence across the social sciences documents pervasive negative employer perceptions against certain groups of workers.[1] In economics, a growing literature studies the role of negative perceptions—as potentially biased or incorrect beliefs about groups—in generating discrimination.[2] Yet little work focuses on understanding how biased beliefs arise in the first place and why they seemingly persist over time. In this paper, I propose and present evidence that discrimination can arise from experience as employers develop biased beliefs about the productivity of worker groups from their market interactions with them. In contrast to the two classes of models typically considered in economics, discrimination is neither the product of preferences (Becker 1957) nor inferences from true group differentials (Phelps 1972; Arrow 1973; Aigner and Cain 1977; Coate and Loury 1993).

Rather, I posit that when employers enter the market, they are not only uncertain about the productivity of individual workers as in the statistical discrimination

* Lepage: SOFI, Stockholm University (email: louis-pierre.lepage@sofi.su.se). Christopher Walters was coeditor for this article. I thank John Bound, Charlie Brown, David Miller, Tanya Rosenblat, and Jeff Smith for continued guidance and feedback. I thank Martha Bailey for valuable advice and Dan Black, Aislinn Bohren, Kerwin Charles, Ash Craig, Ellora Derenoncourt, Pierre Deschamps, Alex Imas, Kevin Lang, Steve Lehrer, Erik Lindqvist, Conrad Miller, Jose Montalban, Nicolas Morales, Mike Mueller-Smith, Mandy Pallais, Rosina Rodriguez Olivera, Dan-Olof Rooth, Mel Stephens, Justin Sydnor, Joanne Tan, Emma von Essen, Justin Wolfers, Steve Woodbury, as well seminar participants at various conferences and institutions for their comments. I gratefully acknowledge funding from the Michigan Institute for Teaching and Research in Economics and from the Russell Sage Foundation Small Grants in Behavioral Economics G-2103-31077. IRB approval was obtained from the University of Michigan for this project (HUM00146993). This paper supersedes material included in papers previously circulated as "Endogenous Learning, Persistent Employer Biases, and Discrimination" and "Bias Formation and Hiring Discrimination."

† Go to https://doi.org/10.1257/app.20220466 to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

[1] See, for example, Kirschenman and Neckerman (1991), Wilson (1996), and Pager and Karafin (2009).

[2] See Fershtman and Gneezy (2001), Reuben, Sapienza, and Zingales (2014), Bordalo et al. (2016), Glover, Pallais, and Pariente (2017), Arnold, Dobbie, and Yang (2018), Bohren et al. (2019), Bordalo et al. (2019), Sarsons (2019), Bohren et al. (2023), and Benson and Lepage (2023).

literature but also the productivity of their group. Since productivity may differ across groups, for example due to historical or social factors, employers value learning about groups to inform their hiring. Previous experiences with workers of a group not only shape an employer's beliefs about that group's productivity but also their subsequent decisions to hire from the group and, indirectly, learn about its productivity. Learning about minority or disadvantaged groups is particularly important if there is less initial information about them in the labor market, making employers more reliant on their own experiences to assess their productivity. In practice, we know from surveys that employers routinely make group associations informed by their experience (Pager and Karafin 2009).[3]

I present a model that captures these intuitive insights, highlighting how biased beliefs that arise and persist endogenously from experience can generate discrimination. In a dynamic setting, employers have noisier initial information on one group's productivity relative to another (Lundberg and Startz 1983; Lang 1986; Cornell and Welch 1996; Morgan and Várdy 2009) and trade off learning about its productivity against current-period profit maximization.[4] Part of the information observed through hiring is privately observed by the hiring employer (Schönberg 2007; Pinkston 2009; Kahn 2013), such that their own hiring history influences their subsequent hiring and learning. Positive experiences with a group create positive biases about its productivity, which endogenously correct themselves by leading to more hiring and learning. Negative experiences, however, create negative biases, which persist by decreasing hiring of the group and therefore learning. The persistence of negative biases results in a negatively skewed belief distribution about the productivity of the group whose productivity is initially more uncertain.

The model helps to (i) understand conditions under which biased beliefs can arise and generate persistent discrimination, (ii) compare this type of discrimination to classical theories, and (iii) inform policies to mitigate discrimination. Each period, beliefs determine market clearing wages, pinned down by the marginal employer's beliefs. Optimal hiring follows a cutoff rule in beliefs: employers below the marginal employer do not hire from the group, preserving their negative biases. The model's key prediction is that, over time, negatively biased beliefs can cause the wage of the group about whose productivity employers have noisier initial information to fall and remain below their expected productivity in the long run. Further, since discrimination arises endogenously from expected profit maximization, it is possible for it to survive some forms of market competition. Moreover, while information from outside of an employer's own hiring can help mitigate biased beliefs, discrimination can persist as long as employers put nonzero weight on their own experiences.[5] In summary, individually biased beliefs can persist within a statistical discrimination framework.

[3] Pager and Karafin (2009) on page 87 document this response of an employer to a negative experience with a Black female worker: "You know, everyone has a couple of bad hires. And you remember those very vividly. And who that person is can really impact. That person just stuck in my head. . . . And I could see her. It was hard to not see her in other people that you meet."

[4] The general trade-off that firms face between exploration and extraction has long been recognized as a key element of organizational learning (March 1991).

[5] In practice, this is likely to arise both because specific hiring contexts vary across employers, implying that own experiences hold valuable information specific to an employer, and because a large body of evidence documents the

Next, I create a controlled environment to test the endogenous learning mechanism that underpins the model. I consider two equally productive arbitrary worker groups that complete a real-effort task corresponding to their productivity.[6] Employers repeatedly hire one worker per period, choosing from one of the two groups, and observe their hire's productivity. They are incentivized to hire the most productive workers available, requiring them to identify which group is more productive, if any. I give employers better initial information on the productivity of one group and study how an employer's hiring history with the other group shapes their hiring and learning. Specifically, by eliciting employer beliefs, I track biased beliefs resulting from an employer's previous hires and how they impact subsequent hiring and therefore learning.

I find support for the mechanism's main testable hypotheses. Negative experiences with the uncertain group, captured through the hiring of relatively low productivity workers, lead to negatively biased beliefs about the group's productivity, which persist specifically by decreasing hiring of the group and therefore learning. In contrast, I find that positive experiences create positive biases, which increase hiring and learning, in turn mitigating these biases. Across employers, differential hiring and learning result in a persistent negatively skewed distribution of beliefs about the group's productivity.

I then test some of the model's predictions regarding changes in the hiring setting and the use of policy tools to mitigate discrimination by varying the experimental design. I find that policies incentivizing learning or providing additional information on groups reduce bias formation, with implications for real world policies like hiring subsidies and affirmative action as well as algorithmic hiring tools. In addition, I document evidence that bias formation against the uncertain group is especially strong when it is labeled as a minority. Lastly, the mechanism operates similarly with gender worker groups, where female workers represent a minority.

Like taste-based discrimination, experience-based discrimination generates differences between average performance and average pay of a group. In fact, the model generates steady state predictions analogous to Becker (1957) with endogenous beliefs replacing exogenous preferences. Apparent taste-based discrimination can result from incorrect statistical discrimination, providing a new way to understand prejudice as the result of experiences shaping beliefs in distortionary ways. Biased beliefs from endogenous learning still differ starkly from a preference. They lead to distinct dynamic predictions and implications for welfare and policy while highlighting that the insights of prejudice-based models for labor market discrimination can be generated from uncertainty, without reliance on a utility function or biased updating.[7]

tendency of agents to put substantial weight on their own experiences (Moore, Tenney, and Haran 2015; Guenzel and Malmendier 2020; Malmendier 2021).

[6] Defining groups based on an arbitrary characteristic (group color) is valuable to isolate the mechanism of interest by abstracting from existing biases and discrimination (Charness, Rigotti, and Rustichini 2007; Chen and Li 2009).

[7] Individuals appear quick to form group perceptions and act on these in a way that shapes future views, consistent with the notion of prejudice from psychology (Bertrand and Duflo 2017). My model shows (i) how biases can micro-found the reduced-form notion of prejudice in economics and (ii) how biases affect decision-making in statistical discrimination models.

Like statistical discrimination, biased beliefs arise from uncertainty. In classical models, employers learn about individual productivity but are assumed to know the productivity of groups or at least to have correct equilibrium beliefs about it. In contrast, I model learning about groups from the extrapolation of experiences with individuals, showing that employers can hold negatively biased equilibrium beliefs because they decide to stop learning. The model predicts discrimination even with equally productive groups and without prior biases, self-fulfilling prophecies, or equilibrium multiplicity.[8] It does so while relying on a standard setting of information asymmetry between worker groups, which naturally results from market interactions when one group is a minority. The model shows how learning about some groups can be slow, complementing work on learning about individuals within groups (Farber and Gibbons 1996; Altonji and Pierret 2001). It also contributes to research combining insights from bandit problems with discrimination in other settings (Denrell 2005; Le Mens and Denrell 2011; Bardhi et al. 2020; Li et al. 2020; Fershtman and Pavan 2021; Komiyama and Noda 2024).[9]

This paper contributes to the growing literature on biased beliefs and stereotypes by proposing a microfoundation of bias with novel dynamic predictions and policy implications. Individual biases arise and evolve from employers conducting inference on an endogenously selected sample of observations about group productivity rather than true group differentials (Bordalo et al. 2016), biased updating (Sarsons 2019), worker evaluation and supervision (Bartoš et al. 2016; Glover, Pallais, and Pariente 2017), or implicit group associations (Bertrand, Chugh, and Mullainathan 2005). Endogenous learning provides a rationale for how employers who are willing to give workers from any group a fair chance can develop persistent negative biases about some groups, suggesting that biased beliefs may be more pervasive and persistent than typically understood. The model rationalizes emerging evidence on hiring decisions being influenced by personal experience (Leung 2017; Benson and Lepage 2023) and is consistent with existing work on experiences shaping beliefs and behavior in other contexts (Malmendier 2021). It also provides a new lens to analyze policies like affirmative action or hiring subsidies, which, consistent with evidence from the experiment, induce learning by increasing minority hiring, mitigating longer-term disparities (Miller 2017). Similarly, the model and empirical evidence highlight that providing information on groups mitigates discrimination *on average*, as can encouraging intergroup interactions, consistent with evidence reviewed in Lang and Kahn-Lang Spitzer (2020) as well as the contact hypothesis (Pettigrew and Tropp 2006).

[8] Arrow (1973) mentions that biased priors could lead to a self-fulfilling prophecy if employers ignore subsequent information or worker responses confirm employer beliefs, but these models have no learning.

[9] See Bergemann and Välimäki (2008) for a review of bandit problems. Komiyama and Noda (2024) studies homogeneous employer bias from a failure of social learning: myopic short-lived firms have little initial information on minority workers and find it difficult to estimate their productivity, discouraging hiring and information accumulation. Li et al. (2020) presents evidence that recruiters and algorithms underestimate the value of learning about the productivity of workers with less common characteristics through hiring. In contrast, employers in my model are forward-looking expected-profit maximizers who fully internalize the value of learning, highlighting that biased beliefs can still arise and generate discrimination.

I. Labor Market Model

A. *Employer Information and Beliefs*

Consider a large number of employers hiring workers from two observably different groups A and B (e.g., race). Through hiring, employers learn about the productivity of worker groups, which may differ due to historical or social factors. Employers know the productivity distribution of group A but are initially uncertain about that of group B. The important feature is that initial information about group B's productivity is noisier, but assuming complete information on group A simplifies the analysis. Information asymmetries across groups are a common feature in the literature, with the distinction that I focus on the dynamic implications of an initial asymmetry for hiring and learning (Lang 1986; Cornell and Welch 1996; Morgan and Várdy 2009). In a majority/minority context, as is frequent in discrimination settings,[10] an information asymmetry naturally results from repeated market interactions: most employers observe more information about the majority, leaving them with relative uncertainty about the minority group.[11]

Each individual worker from either group has productivity drawn from $X|\mu \sim G(x)$, where G is a one-parameter family of distributions characterized by their mean μ, finite variance, and density function $g(x)$ with full support on an interval of real numbers $\mathbf{X}$.[12] Each worker is endowed with a fixed productivity and inelastically provides a unit of labor each period.[13] Employers know that group A's mean productivity is μ and have common priors about the mean productivity of group B, $\mu_B = E_G[x]$, distributed according to the density function $h(\cdot)$ with mean μ_0.[14] I focus on the case where $\mu_0 = \mu$, such that employers have unbiased priors, to highlight that prior bias is unnecessary to generate discrimination. Employers have no hiring experience, but each of them hires one worker per period, updates their beliefs when they hire from group B, and the match dissolves after one period.[15] Employers observe no individual worker signal prior to hiring and only condition their expectation of productivity on group membership, leading to homogeneous wages within groups. Allowing for noisy individual signals of productivity has no impact on key predictions but leads to interesting auxiliary implications as discussed in online Appendix 1.[16]

[10] In the context of gender where women account for nearly half of total employment, clear majority/minority settings are still prevalent given occupational gender segregation.

[11] Asymmetry could also arise if employers have better information about workers of their group. The distinction is of little consequence for the model's predictions but has implications for the learning problem faced by group B employers, tying into a broader literature on in-group/out-group biases.

[12] For simplicity, I assume that employers know other potentially relevant moments of the distribution, for example, variance, to focus attention on the mean.

[13] While a formal model of endogenous worker responses is beyond the scope of this paper, online Appendix 1 discusses how worker responses may exacerbate discrimination.

[14] Employers have misspecified beliefs in the sense that groups are equally productive and the true mean productivity of group B, μ, is a fixed constant, but employers treat it as a random variable due to uncertainty. Employer priors could result from experiences or information observed outside the labor market, but I take these beliefs as given.

[15] One-period contracts focus on group learning by studying employers repeatedly choosing between groups. Multi-period contracts may slow down learning but do not change relative incentives to hire and learn about group B, determined by μ_B. See online Appendix 1 for a discussion of firm size in the model.

[16] For example, it predicts that discrimination may vary across occupation or skill level based on the ease with which productivity is observed.

Workers hired from group B determine the information set of employer j, $\mathcal{S}_{jt}$, composed of one i.i.d. signal drawn from X for each hire. In the baseline model, signals of productivity are private and only available through an employer's own hiring—an employer does not learn about group B unless they hire workers from the group. The cumulative number of signals employer j has observed by time t is $K_{jt} = \sum_{n=1}^{t} \mathbf{1}\{L_{Bnj} = 1\}$, where L_{Bnj} indicates whether a group B worker was hired in period n. Under Bayesian updating on the mean, the distribution of posterior beliefs conditional on $\mathcal{S}_{jt}$ corresponds to

$$\mu_B | \mathcal{S}_{jt} = \frac{\prod_{k \in \mathcal{S}_{jt}} g_{\mu_B}(x_k) h(\mu_B)}{\int \prod_{k \in \mathcal{S}_{jt}} g_{\mu_B}(x_k) h(\mu_B) d\mu_B}. \tag{1}$$

B. *Hiring Decision*

Consider a frictionless labor market, which clears each period. I first consider infinitely lived employers learning about one cohort of workers, abstracting from product-market competition through dynamic entry and exit of firms. Employers are risk neutral wage takers and maximize the present value of lifetime profits. They consider the value of learning about group B, leading to a dynamic optimization problem in which they are initially incentivized to hire from group B in order to learn.[17] An employer's posterior beliefs are characterized by $\psi_{\mathcal{S}_{jt}}$ and $\boldsymbol{\Psi}_t$ is a list of posterior beliefs across employers. Group A's wage, w_A, is time invariant and equal to its expected productivity μ. Group B's wage, $w_{Bt}(\boldsymbol{\Psi}_t)$, is set competitively through market clearing each period and evolves under the influence of $\boldsymbol{\Psi}_t$. The current-period payoff from hiring a worker is equal to their productivity, x_i, with expected value μ for group A and $E[\mu_B | \mathcal{S}_{jt}]$ for group B. Conditional on beliefs and wages at time t, employer j hires from group A or B to maximize their expected profits,

$$\begin{aligned} V\left(\psi_{\mathcal{S}_{jt}}, w_{Bt}(\boldsymbol{\Psi}_t)\right) = \max\Big\{ & \mu - w_A + \beta E_t\left[V\left(\psi_{\mathcal{S}_{jt+1}}, w_{Bt+1}(\boldsymbol{\Psi}_{t+1})\right)\right], \\ & E_t[\mu_B | \mathcal{S}_{jt}] - w_{Bt}(\boldsymbol{\Psi}_t) + \beta E_t\left[V\left(\psi'_{\mathcal{S}_{jt+1}}, w_{Bt+1}(\boldsymbol{\Psi}_{t+1})\right)\right] \Big\} \end{aligned} \tag{2}$$

where β is a discount factor. The continuation value $V(\cdot)$ includes updated beliefs $\psi'_{\mathcal{S}_{jt+1}}$ when group B is hired and $\psi_{\mathcal{S}_{jt+1}} = \psi_{\mathcal{S}_{jt}}$ otherwise. $E_t\left[V\left(\psi'_{\mathcal{S}_{jt+1}}, \cdot\right)\right] \geq E_t\left[V\left(\psi_{\mathcal{S}_{jt+1}}, \cdot\right)\right]$ since hiring group B yields information that cannot decrease expected profits.

Endogenizing group B's wage is key because it is an outcome of interest and because intuition suggests that it should act as a counterbalancing force to bias. By

[17] The dynamic decision problem I study has intuitive similarities with self-confirming equilibrium models for noncooperative games (Fudenberg and Levine 1993). Both study a learning process in which agents learn from experience, beliefs are not contradicted along the equilibrium path, and issues arise from insufficient learning. My model focuses on learning about the environment rather than other players' strategies, showing that some employers optimally stop learning.

bias, I refer to any case in which an employer's posterior mean differs from μ. If the group's wage falls because of negatively biased employer beliefs, then group B becomes "cheaper," which should induce employers to hire them and learn, correcting biases. I study market outcomes accounting for these adjustments. One consideration is whether employers learn about group B from its wage. The baseline model rules this out by assuming static wage expectations: employers expect the wage next period to equal the current one, $E\left[w_{Bt+1}|\mathcal{S}_{jt}\right] = w_{Bt}$, and place zero probability on w_{Bt+1} taking any other value, which is correct in the long run. While the wage in theory could carry relevant information, wages in practice summarize decentralized decisions that depend on factors unobserved by any given employer. Relative wages are also a function of many factors (changing skill and education, macroeconomic shocks, industry mixes, demographics, etc.) such that isolating the impact of other employers' subjective beliefs about group B on residual wages appears implausible.[18] Nevertheless, Section IE presents an extension in which employers noisily learn from sources outside of their hiring, including wages.

Optimal hiring in the current period is determined by contrasting expected profits hiring from group B versus A. The difference is positive whenever

$$\beta E_t\left[V\left(\psi'_{\mathcal{S}_{jt+1}}, w_{Bt+1}(\Psi_{t+1})\right) - V\left(\psi_{\mathcal{S}_{jt+1}}, w_{Bt+1}(\Psi_{t+1})\right)\right] > \mu - E_t\left[\mu_B|\mathcal{S}_{jt}\right] - \left(w_A - w_{Bt}(\Psi_t)\right). \tag{3}$$

Equation (3) compares the expected learning value from hiring group B on the left with expected foregone profit on the right. The perceived value of learning depends on the likelihood that it leads to changes in hiring and higher profits. In the case of negative bias, group B becomes less attractive from both a learning and production standpoint. Thus, when prior experience suggests that group B is less productive, there is a trade-off between expected learning benefits and expected foregone profits from hiring less productive workers. This trade-off corresponds to a contextual one-armed bandit problem where employers choose each period between a safe arm (group A) yielding a payoff from a known distribution and a risky arm (group B) with an unknown payoff distribution. As standard in these problems, obtaining comparatively low payoffs from the risky arm can eventually lead the employer to stop experimenting and choose the safe arm. One distinction from the classical bandit setup is that I consider a market in which wages and therefore payoffs are endogenous to the beliefs of other employers.

[18] Economists themselves have had long-standing unresolved debates about decomposing wage gaps into discriminatory components (Lang and Lehmann 2012). Further, existing work posits that agents often neglect the informational content of prices in contexts of voting, trading, investing, and auctions (Eyster, Rabin, and Vayanos 2019) and documents the imperviousness of agents to information that is not experience based (Malmendier 2021). Similarly, recent developments in modeling firm behavior surveyed in Aguirregabiria and Jeon (2019) focus on how uncertainty and learning in complex environments can lead firms to have biased beliefs, for example, about demand, costs, or the behavior of other firms.

C. *Hiring Cutoff and the Group B Wage*

Define λ_{jt} as the relative willingness to pay (WTP) of employer j for a group B worker,

$$\lambda_{jt}(\mathcal{S}_{jt}) = \beta E_t\Big[V\big(\psi'_{\mathcal{S}_{jt+1}}, w_{Bt+1}(\boldsymbol{\Psi}_{t+1})\big) - V\big(\psi_{\mathcal{S}_{jt+1}}, w_{Bt+1}(\boldsymbol{\Psi}_{t+1})\big)\Big] - \big(\mu - E_t[\mu_B|\mathcal{S}_{jt}]\big). \tag{4}$$

The trade-off between learning and foregone profit, ignoring wage considerations, is captured by λ_{jt}. For notational simplicity, I write λ_{jt} rather than $\lambda_{jt}(\mathcal{S}_{jt})$, understanding that it is a function of an employer's information set. It can be positive even if $E[\mu_B|\mathcal{S}_{jt}]$ falls below μ, highlighting that employers want to avoid future losses from incorrect beliefs.

Each period, labor market clearing implies that, at current wages, the fraction of employers who prefer to hire group B is equal to the fraction of workers from the group. The group B wage each period is thus determined by the marginal employer m: the employer with the lowest λ_{jt} who must hire from the group to clear the market. Specifically, the wage is set such that the marginal employer is indifferent between hiring from either group, $\lambda_{mt} = w_{Bt}(\boldsymbol{\Psi}_t) - w_A$, characterizing the optimal hiring strategy stated in Proposition 1.

PROPOSITION 1 (Optimal Hiring): *The optimal hiring strategy follows a cutoff rule where employer j hires group B at time t if and only if $\lambda_{jt} \geq \lambda_t^c$. Moreover, $\lambda_t^c = w_{Bt}(\boldsymbol{\Psi}_t) - w_A$.*

PROOF:

See Appendix A. ■

Proposition 1 characterizes the cutoff below which it is optimal for employers to avoid hiring group B at a given wage, preserving their beliefs. Since the wage gap is determined by $\lambda_t^c = \lambda_{mt}$, optimal hiring of other employers follows: those with λ_{jt} above the marginal employer hire group B and others group A—clearing the market. Market clearing thus implies

$$\nu_{\boldsymbol{\Psi}_t}\Big(\big\{\psi_{\mathcal{S}_{jt}} : \lambda_{jt} \geq \lambda_t^c\big(w_{Bt}(\boldsymbol{\Psi}_t)\big)\big\}\Big) = F_B \tag{5}$$

and

$$\nu_{\boldsymbol{\Psi}_t}\Big(\big\{\psi_{\mathcal{S}_{jt}} : \lambda_{jt} < \lambda_t^c\big(w_{Bt}(\boldsymbol{\Psi}_t)\big)\big\}\Big) = F_A ,$$

where $\nu_{\boldsymbol{\Psi}_t}$ is a measure over $\boldsymbol{\Psi}_t$, F_g is the fraction of workers from group g, and each worker-employer pair has no incentive to deviate.

An equilibrium, as formally defined in Appendix A, is a stochastic process over beliefs and a mapping from beliefs to wages, which are governed by three conditions each period. First, employers maximize their expected profits according to their Bellman equation and the optimal hiring rule. Second, the labor market clears. Third, employers below the hiring cutoff for group B don't update beliefs, while those above update beliefs based on their hire's productivity according to Bayes' rule.

D. *Biased Beliefs and Discrimination*

As a result of Proposition 1 and Equation (1), it is straightforward to characterize the asymptotic distribution of posterior beliefs described in Proposition 2.

PROPOSITION 2 (Asymptotic Beliefs and Persistent Negative Biases): *As* $t \rightarrow \infty$, *beliefs of employers who remain above the hiring cutoff converge in distribution to* μ. *Others hold a range of beliefs such that* $E\left[\mu_B | \mathcal{S}_{jt}\right] < \mu$. *The limiting fraction of employers with* $E\left[\mu_B | \mathcal{S}_{jt}\right] < \mu$ *equals the fraction of group A workers.*

PROOF:

See Appendix A. ∎

Standard Bayesian reasoning implies that posterior beliefs converge to the truth as the number of signals goes to infinity. On the other hand, employers below the cutoff (implying $E\left[\mu_B | \mathcal{S}_{jt}\right] < \mu$ in the long run given a strictly positive value of learning) don't hire group B, preserving negative biases. In the long run, since unbiased (biased) employers hire B (A), the fraction biased is equal to the fraction of group A workers.[19] Proposition 2 highlights that a subset of employers hold negatively biased beliefs, even asymptotically.

Learning from experience generates a plausible belief distribution for discrimination to arise. First, beliefs exhibit sustained heterogeneity across employers. Second, differential learning across employers results in beliefs being negatively skewed. This belief distribution arises without relying on group differentials, prejudice, biased priors, or biased updating, providing a novel way to understand persistent, heterogeneous, negatively biased beliefs.

The next consideration is whether biased beliefs generate a wage gap. Proposition 3 characterizes the evolution of group B's wage.

PROPOSITION 3 (Wage Gap and Persistent Discrimination): $w_{Bt}\left(\Psi_t\right)$ *is strictly decreasing in t and converges to a constant* $c < w_A$.

PROOF:

See Appendix A. ∎

[19] The Becker (1957) taste-based model requires that the fraction of prejudiced employers be at least as large as the fraction of group A workers to generate a wage gap. Both models thus require a majority of biased or prejudiced employers to generate a wage gap if group A is larger than group B. The fraction biased in my model is endogenously determined to be exactly equal to that of group A rather than being assumed. Widespread biased beliefs may be more plausible than widespread animus, and there is evidence that a large share of employers hold negative perceptions in the context of race (Lang and Lehmann 2012).

The belief distribution becomes negatively skewed because only negative bias can be stable. With hiring experience, supramarginal values of λ_{jt} become concentrated around 0 as $E\left[\mu_B|\mathcal{S}_{jt}\right]$ becomes concentrated around μ. By definition, λ_{mt} lies below supramarginal values of λ_{jt} and thus eventually falls below 0, leading $w_{Bt}(\Psi_t)$ to fall below w_A. By market clearing, the wage cannot increase or remain constant over time. Given a continuum of employers, some employers above the cutoff are expected to have a negative experience hiring from B in any given period such that their λ_{jt} falls below the current cutoff. Then the fraction of employers who want to hire from group B at the current wage is lower than the fraction of group B workers. The wage must thus decrease to induce employers to hire group B and clear the market. Lastly, since beliefs are fixed asymptotically, there is virtually no updating, so the wage converges to a constant. In online Appendix 2, I provide model simulations of the wage and how it changes with model parameters.

Since both groups are equally productive, the wage gap implies that group B is paid below its expected productivity. The predicted wage gap depends on relative group productivity, but the prediction that group B is paid below its expected productivity does not. The model predicts that negatively biased beliefs about group B arise and persist endogenously through individual experiences, generating discrimination against the group.

E. *Market Exit, Competition, and Outside Learning*

I augment the model with dynamic employer entry and exit, providing a simplified reduced-form way to introduce competition through differential exit rates based on beliefs and analyze a setting in which employers hold finite information sets.

Employers exit and are replaced with entrants who hold unbiased priors at aggregate rate δ each period.[20] The exit rate depends on profits and therefore hiring, determined by $E_t\left[\mu_B|\mathcal{S}_{jt}\right]$. Employers who hire group B earn higher expected profits and should have a lower exit rate, $\delta_B < \delta_A$ with $\delta = \delta_B F_B + \delta_A F_A$. If the only determinant of exit is beliefs about group B ($\delta_B = 0$), a differential exit rate eventually eliminates discrimination. In contrast, if firms who hire group B also exit (Audretsch 1991; Schary 1991), it is possible for discrimination to persist under certain parameter values as summarized in Remark 1. That is because entrants can develop biased beliefs just as incumbents did since biased beliefs arise endogenously rather than reflecting a model primitive.[21]

[20] Prior variance may decrease if employers learn from previous cohorts. This is unlikely to eliminate the problem since it would require employers to eventually completely ignore their experience, while the learning problem in practice changes across cohorts. The relative education and experience of women and minority workers was not the same decades ago, and employment contexts have changed substantially.

[21] In taste-based models, firm growth is important since prejudiced firms remain in the market earning lower profits to indulge in their taste for discrimination. Then discrimination is mitigated because unprejudiced firms grow more quickly. In my model, firms do not accept a lower return for their mistaken beliefs, so growth is not conceptually necessary for discrimination to potentially be competed away. See online Appendix 1 for a discussion of the implications of firm size for the model.

Remark 1 (Persistent Discrimination with Market Competition): For some values of δ_A and δ_B with $\delta_A > \delta_B$, there exists a period $\bar{t}$ in which $w_{Bt}(\boldsymbol{\Psi}_t)$ falls below w_A, remains below for all $t > \bar{t}$, and converges to a constant $c < w_A$.

For exit rates near zero, Remark 1 follows from Proposition 3. It is illustrated through simulation in online Appendix 2. All else equal, higher aggregate exit rates and higher competition (differential exit rates) decrease the extent of the wage gap, consistent with empirical evidence (Ashenfelter and Hannan 1986; Black and Strahan 2001) and also illustrated in online Appendix 2. At the extensive margin, this type of competition does not necessarily eliminate the wage gap entirely. In fact, by preventing belief convergence, market exit can help sustain discrimination in some settings, for example when employers learn from sources outside of their hiring as I discuss next.

In many cases, labor markets may provide few salient signals to an employer who has formed beliefs based on their own experience. Even at similar firms, there is mismatch between employment contexts and hiring decisions as well as performance depending on many factors. For example, Benson and Lepage (2023) report in the context of a large national retailer that a manager's hiring of Black workers is influenced by their own previous hiring experiences with the group but not those of other managers at the same store. Still, if employers observe noisy information about group *B*'s productivity from outside their hiring, such as competitors or wages, then they may learn in the absence of hiring. Consider a case in which employers observe one outside signal each period irrespective of hiring. As long as employers put nonzero weight on their own signals, those who hire from group *B* learn faster since they observe both private and outside signals.[22] The belief distribution remains negatively skewed in any finite period, at a minimum creating discrimination along the equilibrium path and reducing the group's lifetime income. In the long run, if beliefs converge, then the wage gap is eliminated. If beliefs do not fully converge, for example because there is market entry and exit or the learning problem evolves over time, then the wage gap can remain following the intuition from Remark 1. In practice, these two conditions appear plausibly satisfied: employers routinely enter and exit the market with finite information sets, and the relative productivity of worker groups has been evolving with changes in demographics and education.

Accordingly, an intuitive interpretation of the model is a cohort of employers learning about a cohort of workers with imperfect transfer across cohorts. Overall, outside learning suggests that discrimination may differ based on the observability of competitors, wages, and productivity and that there is scope for information provision.

[22] Even making hiring outcomes public within employer networks may not conceptually solve the issue that employers learn too little because it could lower incentives for employers to hire group *B* and learn from their own signals, leading to free riding (Keller, Rady, and Cripps 2005).

II. Relationship with Other Models and Empirical Implications

The model generates steady state predictions analogous to Becker (1957), replacing preferences with endogenous beliefs:

- An employer hires group *A* if the wage gap is smaller than λ_{jt} and group *B* otherwise.
- If enough employers have (approximately) correct beliefs to hire all of group *B*, there is effective segregation without a wage gap.
- Otherwise, there is a wage gap determined by the marginal employer.

The model thus generates a difference between average productivity and average pay of a group without deviating from a statistical discrimination framework. This is key since taste-based discrimination has been criticized for the arbitrariness of relying on preferences. The predictions of prejudice-based models for labor market discrimination do not in fact rely on preferences or behavioral primitives but can be understood as arising from uncertainty. Biased beliefs capture context-dependent aspects of discrimination such as skill or education differentials and may be more widespread than animus, which evidence suggests has steadily decreased over past decades (Lang and Lehmann 2012). Preferences and biased beliefs still lead to very distinct predictions regarding welfare and how discrimination arises, evolves, and can be mitigated, as discussed below.

The model complements statistical discrimination by studying learning about groups. In many contexts, the assumption that employers know the productivity of groups or instantly learn it in equilibrium seems implausible. Discrimination in my model does not arise from objective group differences but potentially incorrect perceived differences at the individual level: employers with the same prior beliefs hiring from the same worker pool in the same hiring setting hold different beliefs based on their specific hires. The distinction is important even when worker groups are unlikely to have equal productivity because my model predicts that closing productivity gaps would not eliminate biased beliefs, while it could eliminate statistical discrimination or stereotypes based on a "kernel of truth" (e.g., Bordalo et al. 2016). Further, while statistical discrimination is generally regarded as efficient, a social planner concerned with inequality or equality of opportunity in my model could improve group *B* outcomes at no efficiency cost through increased learning.

The model also complements work on biased beliefs and discrimination. Biased beliefs in the model are (i) endogenous, (ii) dynamic, (iii) individual, and (iv) driven by experience. These features make for a fairly self-contained theory rather than a mechanism through which existing biases are transformed or preserved. They highlight that biased beliefs can evolve in the face of new information, but learning does not necessarily eliminate discrimination when learning itself is endogenous. Biased beliefs also do not necessarily reflect a common feature of the environment but can still be widespread and negatively skewed. Other mechanisms typically create discrimination from non-Bayesian updating in static contexts without learning (Bordalo et al. 2016; Sarsons 2019; Campos-Mercade and Mengel 2024). These other mechanisms focus on employers failing to learn correctly when given information, while

I focus on employers failing to learn because they optimally decide to acquire too little of it.[23] Lastly, grounding biased beliefs in experience gives them a clear origin and predictable evolution, also distinguishing between types of information that could mitigate discrimination.

Central to the model is the idea that employers learn about groups through interaction and exposure, consistent with the contact hypothesis (Pettigrew and Tropp 2006). The model provides a new lens to study policies like desegregation, internships, worker subsidies, cluster hiring, and affirmative action, which may not only increase diversity but also efficiency by inducing employers to learn (Miller 2017; Aizer et al. 2020). It generates the clear prediction that group information, in particular from own experience, leads employers to hold more accurate beliefs *on average*. This prediction contrasts with previous models: preferences should not respond to information about productivity (and it is unclear how they would respond to exposure more generally), information on groups should not affect average outcomes if they already reflect true group productivity, and it is unclear how group information would mitigate biased beliefs if these arise from a static bias or failure to update from new information.

Biased beliefs are conceptually straightforward to distinguish from classical theories: agents act on their imperfect information rather than objective group differences (statistical discrimination) or nonproductivity-related preferences. Empirical advances have identified discrimination from biased beliefs by leveraging dynamic patterns, marginal outcomes tests, and the provision of group-level information (Sarsons 2019; Bohren et al. 2023; Hull 2021; Benson and Lepage 2023). Beyond documenting biased beliefs, data on individual decisions across time and experience can also be used to uncover their specific source and derive targeted policy implications, as demonstrated in the experiment below.

Lastly, hiring algorithms provide another potential tool to mitigate discrimination. First, algorithms could provide recommendations based on pooled information sets across decision-makers and over time, reducing the extent to which employers rely on their own experiences. Second, algorithms are a natural way to implement numerical solutions, which approximate optimal strategies in bandit problems. Since solving these problems analytically is typically challenging, a large body of work especially in computer science has studied how to design tractable algorithms to maximize expected payoffs. Namely, some strategies including the "epsilon-greedy" strategy and some versions of minimax strategies balance exploration and extraction by always including a small positive probability of exploring in any given period (Watkins 1989; Kuleshov and Precup 2014). In the context of my model, these strategies could ensure that employers never fully stop hiring from group B, mitigating biased beliefs.[24] In contrast, algorithms, which add a bonus to

[23] Sarsons (2019) provides evidence that negative experiences with minority workers affect hiring, but because employers exhibit a static bias in the way they weight signals: biased beliefs are not shaped by experiences, but employers interpret experiences in a biased way. Although the two mechanisms are largely complementary, the predicted impact of policy interventions differs between the two.

[24] Theoretically, abstracting from market exit, if employers used such strategies rather than trying to solve the full dynamic programming problem as in my model, then there would be no long-run discrimination since every employer would eventually have approximately correct beliefs.

exploration as in Li et al. (2020), are conceptually not enough to eliminate discrimination in my model if the exploration bonus decreases with additional exploration, because my model already considers employers who fully internalize the value of learning through Equation (2).

III. Labor Market Experiment

The model rests on one fundamental idea: employers learn about groups from their hiring experiences. Learning is therefore endogenous because it is shaped by previous hiring decisions, which themselves were shaped by previous learning. For the mechanism to operate in a hiring context, employers must first recognize that learning about group B is valuable and hire them even though their productivity is uncertain. Second, they must extrapolate from their experiences with individuals to update their perception of the group. Third, this updated perception must affect subsequent hiring of group B and therefore learning about its productivity. I design an experimental market to test whether decision-making and learning are consistent with this mechanism, how features of the hiring context can exacerbate or mitigate bias formation, and how discrimination can be mitigated through policy interventions.

Experiments have frequently been used to study discrimination, particularly belief-based discrimination, because they provide an environment in which beliefs can be observed and mapped into behavior (Charness and Kuhn 2011). In contrast to most existing work, I go beyond documenting bias to focus on its endogenous formation through hiring experiences (Fershtman and Gneezy 2001; Bordalo et al. 2019; Bohren et al. 2023). The experiment focuses on individual hiring decisions, making two simplifications informed by the model: wages are exogenous constants, and every employer is matched with one worker from each group each period. These assumptions simplify the experimental design into a one-armed bandit framework without impacting qualitative predictions.

Bandit problems have been implemented in experiments studying whether participants follow optimal strategies, showing that participants value learning but switch between arms too often and experiment less than optimal (Meyer and Shi 1995; Banks, Olson, and Porter 1997). Rather than studying whether participants play optimally, I frame the problem within a hiring framework and focus on the belief distribution resulting from the strategies participants use in practice. The experiment illustrates how the mechanics of belief updating and sampling behavior observed robustly across different applications in the bandit literature can generate persistent labor market discrimination. That is, combining insights from bandit problems with hiring generates important implications that have been missing from much of the discrimination literature.

I first test the following hypotheses regarding employers learning from experience:

HYPOTHESIS 1: *Positive hiring experiences lead to a higher estimate of group B's mean productivity and more hiring from the group.*

HYPOTHESIS 2: *Negative hiring experiences lead to a lower estimate of group B's mean productivity and less hiring from the group.*

HYPOTHESIS 3: *Through increased hiring, positive experiences increase learning and lead to more accurate beliefs about group B's productivity.*

HYPOTHESIS 4: *Through decreased hiring, negative experiences decrease learning and lead to less accurate beliefs about group B's productivity.*

HYPOTHESIS 5: *Since negative biases are more persistent than positive ones, the final belief distribution about group B's productivity is negatively skewed across employers.*

I then test additional hypotheses along two dimensions, which helps provide a deeper understanding of bias formation. First, since discrimination fundamentally arises from the individual experience of employers, I test the effectiveness of some policy interventions that decrease the extent to which employers should and must rely on their own previous experiences to hire workers. Namely, I investigate whether incentivizing employers to experiment hiring from group *B*—approximating policies like subsidies, quotas, and affirmative action—or directly providing them with additional information about group *B*—approximating policies like hiring centralization and hiring algorithms—mitigate bias formation. Second, since belief updating and associations between the productivity of individual workers and that of their group could differ based on the hiring context and the framing of worker groups, I investigate how bias formation is influenced by the minority status of the uncertain group or by the use of salient worker group labels (gender).

A. *Experimental Design*

A group of 200 workers and 1,299 employers was recruited through Amazon's Mechanical Turk (MTurk) using a subject pool restricted to US adults. Data gathered through MTurk have been found to be reliable and consistent with data obtained from a traditional laboratory environment or other survey methods (Buhrmester, Kwang, and Gosling 2011).[25]

Summary statistics on workers and employers are presented in Table 1, and additional details on recruitment and sample restrictions are presented in online Appendix 3.

Workers.—To construct a hiring pool for employers, workers were assigned the real-effort cognitive task of solving character puzzles under a piece rate. An example puzzle is shown in online Appendix Figure A3-1. Workers were given one practice puzzle followed by four minutes to solve as many puzzles as they could, which corresponds to their productivity. Of the workers, 25 percent were randomly assigned to group *B* and the rest to group *A* so that both groups would have equal productivity

[25] The subject pool is likely younger, more educated, and more liberal in their views than the US average (Berinsky, Huber, and Lenz 2012). Theoretically, the qualitative predictions of the framework do not depend on these characteristics, but to the extent that prior beliefs and willingness to hire minority workers may be higher in the MTurk subject pool, then my results likely underestimate the extent of negatively biased beliefs that experience-based discrimination would generate in a representative sample of the population.

Table 1—Summary Statistics

	Group A (1)	Group B (2)	Male (3)	Female (4)
Panel A. Puzzles solved by workers				
Mean	9.23	9.12	9.38	8.91
Standard deviation	(3.44)	(3.68)	(3.63)	(3.27)
Median	9	9	9	9
Min	1	1	1	1
Max	18	18	18	18
Number of observations	150	50	123	77
p-value				
H0: $\mu_O = \mu_G$	0.85			
H0: $\mu_M = \mu_F$			0.35	

	Mean (1)	Standard deviation (2)
Panel B. Employer demographics		
Age	36.14	10.45
Male	0.63	0.48
White	0.75	0.43
Black	0.09	0.28
Asian	0.07	0.26
Hispanic	0.06	0.24
At least some college	0.85	0.36
Employment beyond MTurk	0.72	0.45
Number of observations	1,299	

Notes: p-values are from t-tests for the equality of means. In the experiment, color labels *Gray* and *Orange* were used rather than labels B and A to avoid indicating an ordering. Employment Beyond MTurk is an indicator variable for the participant being employed outside of the MTurk platform.

distributions.[26] In practice, color labels *Gray* and *Orange* were used rather than labels B and A, which could indicate an ordering.[27] There is no interaction between employers and workers, nor do employers belong to either group, so there is little room for taste-based discrimination or group attachment to arise, especially since those mechanisms would not interact with the random nature of hiring experiences with group B. Workers solved nine puzzles on average with a minimum of 1, a maximum of 18, and no statistically significant average difference across groups.

Employers.—The experiment was designed to create the simplest setting to study how biased beliefs arise endogenously from experience in a setting abstract enough to study the primal mechanism through which discrimination arises. Employers were incentivized with hiring the most productive workers over 15 periods $t = 1, \ldots, 15$, which required hiring from the group with higher expected

[26] Given 15 periods of hiring and 200 workers in total, the 25 percent fraction was chosen so that group B would be large enough that employers could not expect to hire all or most of it as part of their task but small enough that it would appear as a clear minority. While the framework predictions do not depend on this choice, results from an additional experimental treatment presented below show that the minority status of group B does affect belief updating of employers.

[27] To control for preferences, the colors green and purple were also used, and the color order was varied such that some employers saw green or orange as the uncertain group and others purple or gray. The different color variations are pooled together in the analysis.

productivity.[28] Each period, they observed their hire's productivity and received credits at a flat rate for each puzzle that their worker solved. There is no ground for statistical discrimination to arise since groups are equally productive, but because this is initially unknown to employers, group information is relevant.

Before hiring, employers were shown an example puzzle and given the size of both groups, revealing that group *B* is a minority, as well as the mean productivity of group *A*, μ. Group *B* is never explicitly referred to or labeled as a minority group to avoid potential connotations with the term. The group is framed in a neutral way—its workers could just as well be more rather than less productive on average—to make clear that uncertainty drives discrimination.

When hiring from group *B*, they were given a randomly drawn worker from the group (drawn without replacement). When hiring from group *A*, they were given a worker with productivity equal to the group average of nine. Theoretically, this simplification has no impact on behavior based on expected productivity. In practice, it simplifies the instructions substantially and should be of little consequence because I focus on the impact of hiring experiences on subsequent hiring and beliefs rather than baseline hiring differentials across groups.[29] To investigate the role of ambiguity aversion, employers completed a separate task after hiring to obtain an individual measure of ambiguity aversion following Gneezy et al. (2015), but I show that neither risk nor ambiguity aversion are plausible alternative explanations for my results.

Beliefs about group *B*'s mean productivity were elicited using a binarized scoring rule as proposed in Hossain and Okui (2013)—incentivized for a random sample of two periods as detailed in online Appendix 3. Beliefs were elicited before an employer's first hire and after each period they hired from group *B*. When an employer hired from group *A*, their beliefs about group *B* carried over from the last period. Belief elicitation incentives were chosen to be small compared to hiring payoffs to minimize distortions in hiring incentives. Still, to investigate whether the timing of belief elicitation affects bias formation, an additional group of employers only had their beliefs elicited at the end of the hiring task, providing little evidence that it substantively affected belief formation (Columns 6 and 7 of panel B of Table 3).

Treatments and Empirical Strategy.—To test Hypotheses 1–5, employers were assigned to one of two treatments:

- Treatment *Baseline*: Each period, employers choose between hiring from group *A* or *B*. Group *A* is the majority with 75 percent of workers.
- Treatment *Control*: As in Treatment *Baseline*, but employers can only hire from group *B* each period.

[28] Benson and Lepage (2023) report that the median number of hires by managers at a large US retailer is 10 over a six-year period, suggesting that 15 hires can correspond to a substantial time frame.

[29] Risk-averse employers have an incentive to hire group *A*, but it does not interact with the random nature of hiring experiences with group *B*. That is, the goal is not to document baseline hiring differentials between groups, but how better or worse hiring experiences impact hiring of group *B* relative to group *A* and learning. Further, evidence below shows a clear link between beliefs about the *mean* productivity of group *B* and hiring, supporting an interpretation based on expected productivity.

Treatment *Baseline* allows me to test Hypotheses 1 and 2 by observing how hiring experiences impact subsequent hiring and beliefs. Testing Hypotheses 3 and 4 is complicated by the fact that hiring experiences affect posterior beliefs in two distinct ways: they mechanically lead to belief updating and they impact hiring, indirectly affecting belief updating in future periods. The second corresponds to the mechanism of interest. The *Control* treatment allows me to separately identify the two by providing variation in belief updating that is independent of hiring choices. For those employers, hires influence beliefs about group B, but the mechanism of interest is shut down because they cannot stop hiring from the group. Contrasting the final belief distributions across the two treatments allows me to test Hypothesis 5.

I use nine puzzles as the cutoff for a good experience given the implicit comparison to hiring from group A, which is known to yield nine puzzles. While the productivity of group B hires is randomly drawn irrespective of an employer's hiring history, the decision to hire group B and observe a productivity draw beyond the first is endogenously determined by previous experiences. Accordingly, I present results isolating the impact of a first group B hire's productivity or holding constant the previous number of group B hires across employers. The first experience is exogenous and captures the total impact on hiring and beliefs over subsequent periods. Considering the productivity of later hires allows me to test how experiences affect hiring and belief updating more generally as well as how impacts vary with an employer's previous number of hires.

B. *Evidence on Experience-Based Discrimination*

I first characterize how the hiring and learning of employers in the *Baseline* treatment is shaped by their experiences hiring from group B. I then isolate the impact of endogenous learning on hiring and bias formation by comparing employers across the *Baseline* and *Control* treatments.

Previous Experiences, *Hiring*, *and Beliefs*.—I provide evidence for Hypotheses 1 and 2 in Figure 1 and Table 2, focusing on the impact of hiring experiences with group B on subsequent hiring of the group and final beliefs about its productivity. The impact of experiences on subsequent hiring can be seen as a first stage since the mechanism posits that experiences impact beliefs specifically though changes in hiring. Throughout the analysis, I use a 1 percent statistical significance level unless specified otherwise.

Panel A of Figure 1 shows a clear relationship between the productivity of an employer's hires from group B and their subsequent hiring of the group over the remaining periods. The figure plots estimates from linear regressions of total future hires from group B on the number of puzzles solved by a given group B hire, estimated separately for each group B hire. I plot up to the first nine hires from group B, which corresponds to the average number of total hires from the group over the 15 periods. The figure shows that subsequent B hiring increases if B hires have higher productivity, especially if productivity is above nine (positive experience), and decreases if productivity is below nine (negative experience). The relationship appears particularly strong for early hires, consistent with employers responding

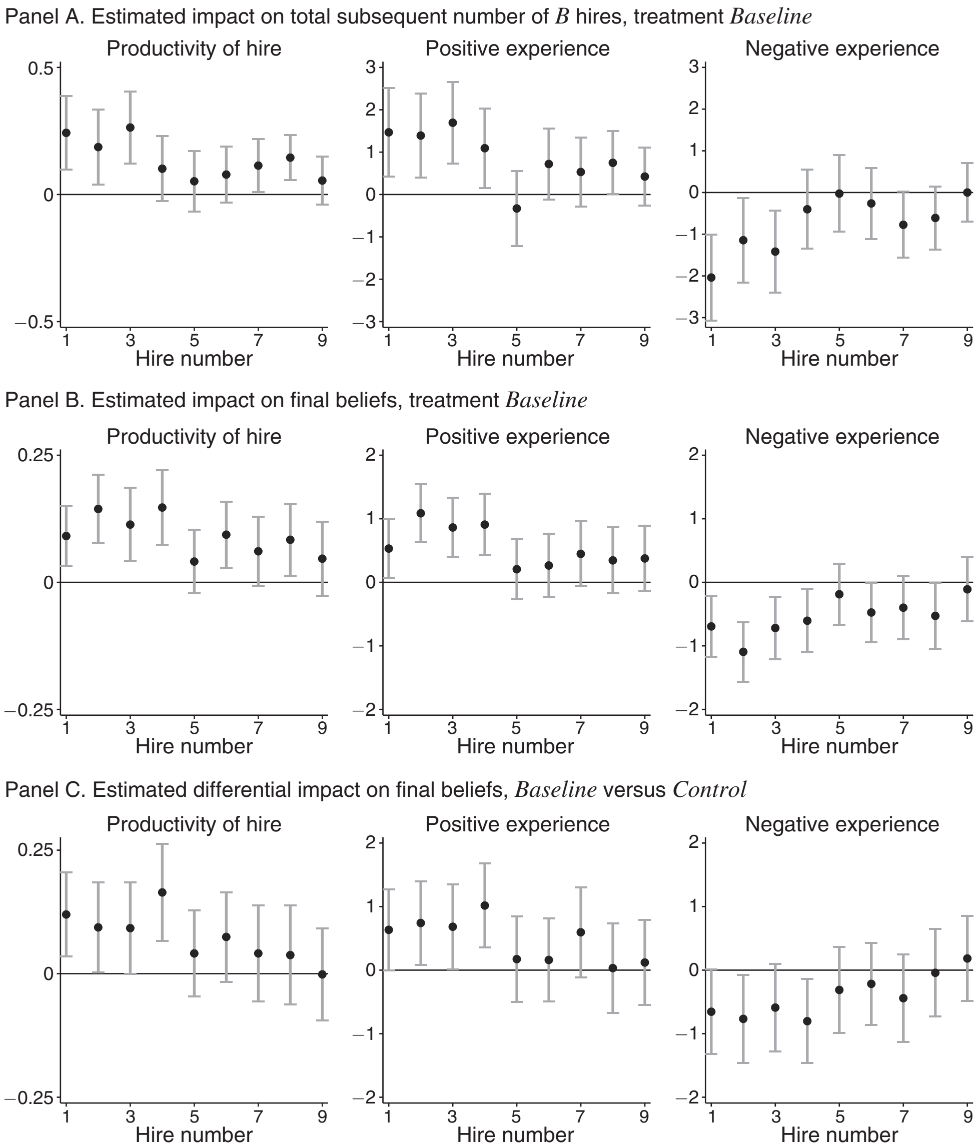

Figure 1. Estimates of the Impact of the Productivity of Group *B* Hires on Hiring and Beliefs

Notes. Panels A and B show the estimated impact of the productivity of group *B* hires on subsequent hiring of the group and final beliefs about its productivity for the *Baseline* treatment. Estimates are obtained from linear regressions of total subsequent group *B* hires or final beliefs about group *B* productivity on the number of puzzles solved by a group *B* hire. Panel C shows the estimated differential impact of the productivity of group *B* hires on final beliefs for the *Baseline* versus the *Control* treatment. Estimates are obtained from linear regressions of final beliefs about group *B* productivity on the number of puzzles solved by a group *B* hire, an indicator variable for an employer being assigned to the *Baseline* treatment, and an interaction term between the two. Effects are estimated separately for each group *B* hire. Treatment *Baseline*: each period, employers choose between hiring from group *A* or *B*. Group *A* is the majority with 75 percent of workers. Beliefs about the mean productivity of group *B* are elicited before the first hire and after every hire from the group. Treatment *Control*: as in treatment *Baseline*, but employers can only hire from group *B* each period. A negative (positive) experience is defined as a hire from group *B* having productivity < 9 (> 9), the mean productivity of group *A*.

TABLE 2—IMPACT OF THE PRODUCTIVITY OF GROUP *B* HIRES ON HIRING AND BELIEFS

	Total subsequent number of group *B* hires					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Baseline treatment						
Productivity of *B* hire	0.107 (0.021)					
Productivity of *B* hire × # of previous *B* hires		−0.018 (0.004)				
Positive experience			0.727 (0.153)			
Positive experience × # of previous *B* hires				−0.124 (0.030)		
Negative experience					−0.623 (0.154)	
Negative experience × # of previous *B* hires						0.136 (0.029)
Outcome mean	5.115	5.115	5.115	5.115	5.115	5.115
Number of observations	2,389	2,389	2,389	2,389	2,389	2,389
	Final beliefs about group *B* productivity					
Panel B. Baseline treatment						
Productivity of *B* hire	0.091 (0.011)					
Productivity of *B* hire × # of previous *B* hires		−0.005 (0.002)				
Positive experience			0.555 (0.078)			
Positive experience × # of previous *B* hires				−0.038 (0.017)		
Negative experience					−0.565 (0.076)	
Negative experience × # of previous *B* hires						0.036 (0.018)
Outcome mean	8.975	8.975	8.975	8.975	8.975	8.975
Number of observations	2,389	2,389	2,389	2,389	2,389	2,389
	Final beliefs about group *B* productivity					
Panel C. Differential impact, baseline versus control						
Baseline × prod. of *B* hire	0.070 (0.014)					
Baseline × productivity of *B* hire × # of previous *B* hires		−0.007 (0.003)				
Baseline × positive experience			0.440 (0.096)			
Baseline × positive experience × # of previous *B* hires				−0.052 (0.022)		
Baseline × negative experience					−0.407 (0.096)	
Baseline × negative experience × # of previous *B* hires						0.055 (0.023)
Outcome mean	8.913	8.913	8.913	8.913	8.913	8.913
Number of observations	4,414	4,414	4,414	4,414	4,414	4,414

Notes. Panels A and B show estimates of the impact of a group *B* hire's productivity on subsequent hiring of the group and beliefs about its productivity for the *Baseline* treatment. Estimates are obtained from linear regressions of the total number of subsequent group *B* hires by an employer or their final beliefs about group *B* productivity on the number of puzzles solved by a given group *B* hire. In Columns 2, 4, and 6, the regression model is augmented with a variable indicating the number of previous hires from group *B* and an interaction term between this variable and the number of puzzles solved by a given group *B* hire. Panel C shows estimates of the differential impact of a group *B* hire's productivity on beliefs about the group's productivity for the *Baseline* versus the *Control* treatment. Estimates are obtained from linear regressions of an employer's final beliefs about group *B* productivity on the number of puzzles solved by a given group *B* hire, an indicator variable for whether the employer was assigned to the *Baseline* treatment, and an interaction term between the two. In Columns 2, 4, and 6, the regression model is augmented with a variable indicating the number of previous hires from group *B* as of a given period, an interaction term between this variable and the number of puzzles solved by a given group *B* hire, and interaction terms between these two variables and whether the employer was assigned to the *Baseline* treatment. Robust standard errors are presented in parentheses. Treatment *Baseline*: each period, employers choose between hiring from group *A* or *B*. Treatment *Control*: as in treatment *Baseline*, but employers can only hire from group *B* each period. Group *A* is the majority with 75 percent of workers. Beliefs about the mean productivity of group *B* are elicited before the first hire and after every hire from the group. Regressions include the employer's prior beliefs about group *B*'s average productivity elicited before the hiring task. Regressions in panels A and B also include an individual measure of ambiguity aversion calculated as in Gneezy et al. (2015). A negative (positive) experience is defined as a hire from group *B* having productivity < 9 (> 9), the mean productivity of group *A*.

strongly to their first experiences because they are particularly uncertain about group *B*'s productivity.[30] While neither the theoretical results nor predictions of the experiment depend on the timing of negative experiences for any given employer, if negative experiences lead an employer to reduce their hiring of group *B* early on, then their impact may be particularly large and persistent by reducing subsequent information acquisition.

Panel A of Table 2 displays a similar pattern using the same specification, but averaging over all *B* hires for each employer. In Column 2, 4, and 6, I also present estimates from linear regressions interacting the productivity of a group *B* hire with the number of previous hires from group *B* while also controlling for the main effects of both of these variables. The productivity of *B* hires statistically significantly affects subsequent hiring, with 0.1 additional hire for each additional puzzle solved by a *B* hire, but the relationship weakens with each additional hire from the group. Similarly, a hire with productivity above nine increases hiring by 0.73 workers, while a hire with productivity below decreases it by 0.62. Throughout the analysis, controlling for employer priors and individual measures of ambiguity aversion has negligible impact on the results (online Appendix Table A3-7).[31]

Panel B of Figure 1 and Panel B of Table 2 show a similar relationship between the productivity of *B* hires and final beliefs of an employer about the group's mean productivity as the outcome variable. Estimates in Table 2 are statistically significant at the 1–5 percent level, namely showing that an increase of 1 puzzle solved by a *B* hire increases final beliefs about the group's productivity by 0.1 puzzles, while a hire with productivity above (below) 9 increases (decreases) final beliefs by 0.64 puzzles. Once again, interacting these impacts with the previous number of group *B* hires indicates that the impacts decrease with additional hiring experience. These impacts are consistent with bias formation from experience but conflate the impact of experiences on beliefs through hiring and mechanical belief updating based on the productivity of hires. Since initial experiences play an important role in shaping hiring and beliefs, the timing of experiences matters, relating intuitively to work on the lasting consequences of first impressions (Agnew et al. 2018).

The results highlight that hiring experiences, particularly early ones, set employers on persistently different paths of hiring and learning, directly supporting Hypotheses 1 and 2.

Endogenous Learning and Bias Formation.—Next, I provide evidence relating to Hypotheses 3 and 4 in panel C of Figure 1, Figure 2, and panel C of Table 2 by comparing the effect of hiring experiences across the *Baseline* and the *Control*

[30] Employers switch between groups more times than is optimal, on average 3.46 times with a standard deviation of 3.23. Still, nearly 40 percent switch at most once and nearly half switch at most twice. Increased switching could mitigate bias formation because employers do not completely stop hiring from group *B* after switching to group *A* once. Yet if employers are quicker to switch away from group *B* in the first place, this may decrease hiring and learning if early experiences are particularly important. Evidence that agents switch more than optimal in bandit problems is a common finding (e.g., Meyer and Shi 1995; Banks, Olson, and Porter 1997). It is unlikely to represent a lack of comprehension by employers given that they had to complete comprehension questions targeting important aspects of the task and setting.

[31] There is little relationship between ambiguity aversion and total *B* hiring in general and little interaction with the productivity of the first hire (online Appendix Table A3-8).

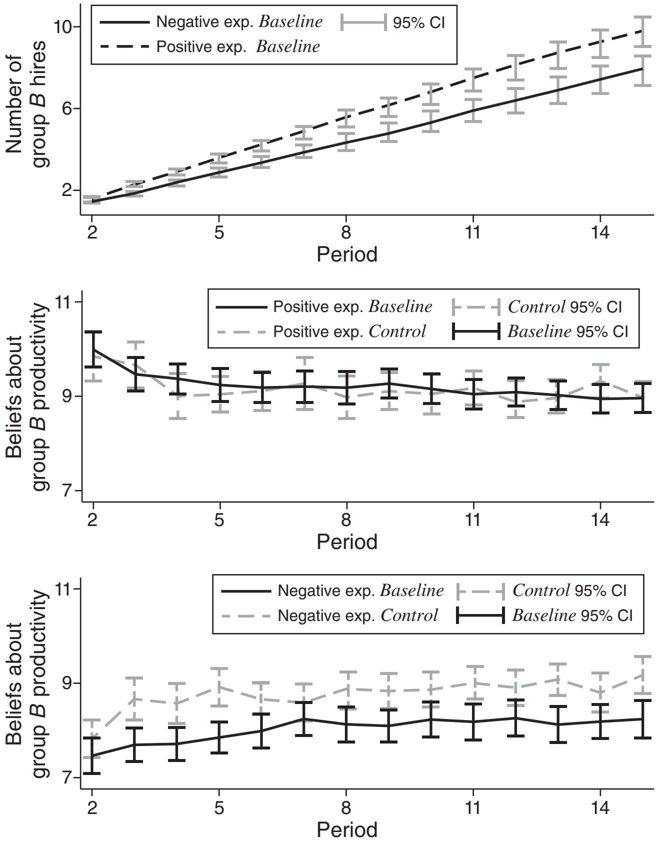

Figure 2. Impact of First Experience with Group *B* on Hiring and Beliefs

Notes. Panel A shows the impact of an employer's first experience hiring from group *B* on hiring of the group for the *Baseline* treatment, separated by whether the first experience was positive or negative. Estimates are obtained from linear regressions of the number of group *B* hires by an employer on an indicator variable for whether their first experience hiring group *B* was positive or negative. Panel B (C) shows the impact of an employer's first experience hiring from group *B* being positive (negative) on beliefs about the group's productivity. Estimates are obtained from linear regressions of the beliefs of an employer about the productivity of group *B* on an indicator variable for whether their first experience hiring group *B* was positive or negative, estimated separately for the *Baseline* and *Control* treatment. A negative (positive) experience is defined as the first hire from group *B* having productivity < 9 (> 9), the mean productivity of group *A*.

treatments. I isolate the impact of endogenous learning, showing that previous experiences have larger impacts on final beliefs when employers selectively decide whether to hire more from the group based on these experiences and that negative biases in particular are persistent.

Panel C of Figure 1 and panel C of Table 2 show that experiences hiring group B, both positive and negative, have larger impacts on final beliefs of employers in the *Baseline* treatment compared to the *Control* treatment, consistent with experiences shaping hiring and therefore subsequent learning. The estimates come from linear regressions using data from both the *Baseline* and *Control* treatments, which control for an indicator variable for an employer being assigned to the *Baseline* treatment, the productivity of a given group *B* hire, and the interaction between the two, which isolates the impact of endogenous learning. In Figure 1, I show estimates from separate regressions for each group *B* hire, while Table 2 pools estimates across all group *B* hires from a given employer. Columns 2, 4, and 6 of Table 2 present results of the triple interaction between being assigned to the *Baseline* treatment, the productivity of a given group *B* hire, and the number of group *B* hires by the employer prior to that group *B* hire—obtained from a regression also including the full set of main effects and interactions—again to investigate the relative importance of early hiring experiences. The impacts in Table 2 are statistically significant at the 1–5 percent level and once again subside with each additional hire from group *B*. The latter result is unsurprising because employers from the *Baseline* treatment who hire more *B* workers end up with a more similar number of signals from which to update their beliefs to employers from the *Control* treatment and because the scope for experiences to affect beliefs through subsequent hiring decreases with each additional hire.

Figure 2 traces out the impact of a first negative versus positive experience with group *B* on hiring and belief updating, providing an intuitive way to visualize the experiment's main results. Each estimate comes from a linear regression of the number of group *B* hires or beliefs about the group's productivity as of a given period of the hiring task on an indicator variable for whether an employer's first experience hiring group *B* was positive or negative. In the top panel, only employers in the *Baseline* treatment are included in the estimation. In the middle and bottom panels, coefficients are estimated and presented separately for the *Baseline* and *Control* treatments.

The top panel shows that a first negative experience persistently lowers group *B* hiring compared to a positive one, with a difference of two hires or 20 percent after 15 periods. The middle panel shows the evolution of beliefs following a first positive experience. A first positive experience leads to positively biased beliefs, which dissipate over subsequent periods in both treatments. In later periods, beliefs across the two treatments converge to the true mean productivity of group *B*.

As shown in the bottom panel, following a first negative experience, employers in both treatments have negatively biased beliefs. In contrast to positive bias, negative bias only dissipates in the control treatment, in which employers keep hiring from group *B* even when they believe it to be less productive and therefore keep learning. Negative bias corrects much more slowly in the *Baseline* treatment because it decreases hiring and therefore learning.[32] After 15 periods, those employers hold average beliefs substantially below nine and corresponding to a decrease in bias of less than 50 percent from period 2.[33]

[32] Consistent with the trade-off that employers face between the two worker groups, the observed impact of a first negative experience on final beliefs is not solely driven by particularly negative first experiences but rather any first hire with productivity below that expected of group *A* (online Appendix Figure A3-4).

[33] *t*-tests indicate no statistically significant differences across treatments in periods 2 and 15 conditional on a first positive experience (*p*-values 0.678, 0.959) or in period 15 for the control treatment regardless of the first experience (*p*-value 0.475). In contrast, while beliefs are not statistically significantly different across treatments in period 2 conditional on a first negative experience (*p*-value 0.230), they are in period 15 both across treatments and whether the first experience was negative or positive within the *Baseline* treatment (0.003, 0.010).

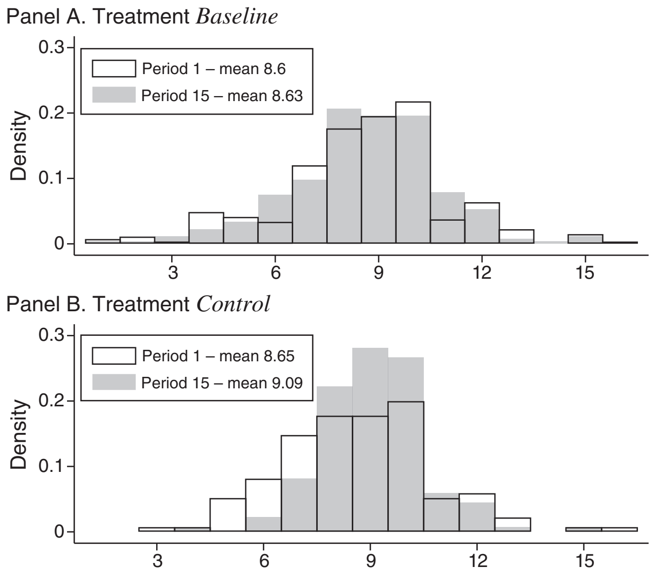

Figure 3. Distribution of Employer Beliefs, Treatment Baseline versus Control

Notes: Treatment *Control*: as in treatment *Baseline*, but employers can only hire from group *B* each period. A small fraction of employers from treatment *Baseline* who did not hire from group *B* are excluded from the sample. See Figure 1 for more details.

I present evidence for Hypothesis 5 in Figure 3, contrasting the change in the belief distribution between treatments from the first period to the last.[34] Both treatments have similar negatively biased initial beliefs about the mean productivity of group *B* of around 8.6. A mass of initial beliefs distributed around nine likely results from anchoring in the instructions, which mentioned that the average productivity of group *A* was nine but said nothing of group *B*'s.[35] The *Control* treatment generally corrected their biases, with increased mass around 9 and average beliefs of 9.09 after 15 periods. In contrast, *Baseline* treatment employers had essentially the same average beliefs as in period 1, and proportionally little changed in the left tail. Both Wilcoxon rank-sum and Kolmogorov-Smirnov tests reject the null hypothesis of equal period 15 belief distributions across treatments (at the 5 and 7 percent level, respectively).[36] Online Appendix A3-2 plots the difference in the final belief distributions across treatments, highlighting that much of the difference lies in the *Baseline* treatment being more likely to have beliefs below 8 and less likely to have beliefs of 9 and 10.

[34] To focus on bias formation, I restrict the sample to the vast majority of employers from the *Baseline* treatment who hired at least one group *B* worker.

[35] This design choice approximates the case of unbiased prior distribution considered in the model and should yield malleable beliefs given how little information is provided to form beliefs. Still, average initial beliefs below nine indicates that employers seemingly incorrectly expected group *B* to be less productive, for example, because of their minority status. Consistent with this possibility, initial beliefs in the treatment described below presenting both groups as equally sized were closer to nine.

[36] In contrast, the tests fail to reject the null that the period 1 belief distributions are equal across treatments, with p-values of 0.77 and 0.92.

Persistent priors don't appear to be the cause of these biased beliefs. Even with a relatively accurate prior distribution, over 80 percent of employers finished the experiment with beliefs that differed from those they held in period 1 by more than 1 puzzle. There is also clear heterogeneity in beliefs based on the number of group *B* hires, with *Baseline* treatment employers who hired at least one but less than nine *B* workers having average final beliefs of 7.7 versus 9.4 for those who hired more than 9.[37] This pattern highlights how beliefs may not converge or converge slowly with experience when experience itself is endogenous. This is particularly striking given that, as shown in the previous subsection, employers responded roughly symmetrically to positive and negative experiences in terms of hiring and belief updating, but negative bias persists by decreasing subsequent hiring and learning.

Risk or ambiguity aversion also don't provide a plausible alternative interpretation for these findings. Negative experiences cause employers to hire fewer group *B* workers rather than a fundamental unwillingness to hire the group. Even though employers are not forced to hire group *B*, approximately 90 percent of them do at least once. The dynamics of hiring and belief updating, for example, evidence that beliefs converge with experience, are also consistent with a learning interpretation. Moreover, evidence below shows that keeping uncertainty and risk constant in a way that does not affect the propensity of employers to hire from group *B* still affects bias formation by affecting how employers update their beliefs from experience. Alternatively, changing incentives of employers to trade off hiring and learning in a way that decreases the scope for risk aversion does not simply decrease bias by increasing hiring—but by weakening the relationship between previous experiences and hiring. Lastly, providing information on average group productivity in a way that leaves hiring incentives constant increases group *B* hiring, highlighting that correcting negative biases has a direct impact on hiring and learning.

C. *Evidence on Policy Implications and Changing the Hiring Context*

I now consider additional experimental treatments to gain a deeper understanding of how experience-based discrimination operates. First, I test some empirical predictions of the model related to commonly used policies to mitigate discrimination. Second, I vary the hiring context to investigate how it affects bias formation.

Policy Tools.—First, I consider an *Exploration* treatment varying the cost of exploration by giving employers a bonus equivalent to two puzzles solved (440 credits or about 22 percent of average productivity) for each group *B* hire. Incentivizing hiring should decrease the extent to which it is shaped by previous experience and therefore the extent to which learning is endogenous.

Second, I consider an *Information* treatment providing employers with information on group *B* from outside their hiring. In periods 10–15, regardless of hiring, employers were given the mean productivity of five randomly selected group *B* workers previously hired by other employers in that period, and their beliefs were elicited.

[37] Online Appendix Figure A3-3 displays a strong positive relationship between average final employer beliefs and the total number of *B* hires.

Table 3—Differences in Hiring and Beliefs across Treatments Compared to Treatment Baseline

	Total number of *B* hires (1)	Final bias in beliefs (2)	Probability of hiring *B* (3)	Bias in beliefs (4)
Panel A. Policy interventions				
Exploration treatment	1.814	−0.340		
	(0.519)	(0.139)		
Period 10–15			−0.024	−0.053
			(0.016)	(0.057)
Information treatment × period 10–15			0.079	−0.551
			(0.032)	(0.138)
Outcome mean	8.65	1.55	0.49	1.71
Number of observations	445	445	6,517	6,517

	Total number of *B* hires (1)	Final bias in beliefs (2)	Final bias in beliefs (3)	Total number of *B* hires (4)	Final beliefs (5)	Total number of *B* hires (6)	Final bias in beliefs (7)
Panel B. Changes in hiring context							
Equal treatment	0.660	−0.259					
	(0.496)	(0.138)					
Gender treatment			0.325	0.236	−0.380		
			(0.154)	(0.732)	(0.356)		
Male employer				0.015	0.072		
				(0.575)	(0.289)		
Gender treatment × male employer				−0.375	0.795		
				(0.923)	(0.457)		
Elicitation treatment						0.572	0.159
						(0.435)	(0.139)
Outcome mean	8.27	1.58	1.82	8.97	8.67	8.27	1.73
Number of observations	449	449	534	468	468	487	487

Notes. The table displays differences in hiring and beliefs across experimental treatments compared to the *Baseline* treatment. Estimates in columns 1 and 2 of panel A and 1, 2, 6, and 7 of panel B were obtained from a linear regression of the total hiring of group *B* by an employer or their final bias in beliefs about the group's productivity on an indicator variable for an employer being assigned to the *Exploration*, *Equal*, or *Elicitation* treatment. Estimates in columns 3 and 4 of panel A were obtained from linear regressions of the probability of an employer hiring from group *B* in a given period or their current bias in beliefs about the group's productivity on an indicator variable for the employer being assigned to the *Information* treatment, an indicator for being in periods 10–15 of the hiring task when additional information was provided to employers, and an interaction term between the two. Estimates in Column 3 of panel B were estimated from a regression of final bias in beliefs about group *B*'s productivity on an indicator variable for an employer being assigned to the *Gender* treatment. Columns 4 and 5 also include an indicator variable for an employer being male and its interaction with the employer being assigned to the *Gender* treatment. Robust standard errors are presented in parentheses for columns 1 and 2 of panel A and panel B, and clustered standard errors at the employer level are presented in parentheses for columns 3 and 4 of panel A. The reference treatment in all columns is treatment *Baseline*. *Exploration*: as in treatment *Baseline*, but employers are given a 440-credit bonus each period they hire from group *B*. Treatment *Information*: as in treatment *Baseline*, but employers are given additional information on group *B* in periods 10–15, and their beliefs in those periods are elicited regardless of hiring. Treatment *Equal*: as in treatment *Baseline*, but groups are equally sized with 50 workers each. Treatment *Gender*: as in treatment *Baseline*, but groups correspond to male (123) and female (77) workers. *Elicitation* is an indicator variable for the employer having been assigned to a group of 190 employers who only had their beliefs elicited once at the end of the hiring task. The specification for columns 3 and 4 of panel A include employer fixed effects to capture time-invariant tendencies across employers to hire from a group and update their belief.

Columns 1 and 2 of panel A of Table 3 show that the *Exploration* treatment hired two more *B* workers on average and had 22 percent lower final bias. The estimates shown were obtained from a linear regression of the total hiring of group *B* by an employer or their final bias in beliefs about the group's productivity on an indicator variable for an employer being assigned to the *Exploration* treatment. These employers finished with average beliefs of 8.97, correcting a substantial fraction of negatively biased beliefs. Moreover, online Appendix Table A3-3 presents evidence that lowering the cost of exploration specifically weakened the relationship between the productivity of previous *B* hires and subsequent hiring of the group, decreasing selection in hiring and therefore endogeneity in learning.

Columns 3 and 4 of panel A of Table 3 present estimates from linear regressions of the probability of an employer hiring from group *B* in a given period or their current bias in beliefs about the group's productivity on an indicator variable for the employer being assigned to the *Information* treatment—an indicator for being in periods 10–15 of the hiring task when additional information was provided to employers—and an interaction term between the two. Column 3 shows that the *Information* treatment was statistically significantly 16 percent more likely to hire from group *B* in periods 10–15. This increase is consistent with inducing employers with negative biases to hire from group *B*, even when the information across periods indicated that the two groups were equally productive on average, leaving hiring incentives unchanged. Column 4 indicates a decrease of 32 percent in bias for periods 10–15. Employers finished with average beliefs of 8.79, a roughly 50 percent reduction in negatively biased beliefs in particular.

These findings highlight that employers internalize the exploration-extraction tradeoff and that decreasing the cost of exploration mitigates bias formation. Moreover, while the *Information* treatment provides a lot of information, equivalent to 25 hires, and it may be harder to convey to employers in other settings, the findings stress that the issue driving discrimination is a lack of information. For policy, these results highlight how interventions that increase minority hiring or information about minority workers can not only improve their short term outcomes but also shape a different path of hiring and employer perceptions for the future. Several policies could approximate the *Exploration* treatment, including subsidies as well as affirmative action. Others could serve a similar role as the *Information* treatment, namely information aggregation through centralized hiring. Lastly, hiring algorithms could serve a similar role as both the *Exploration* and the *Information* treatments if they explicitly assign value to learning about worker groups and aggregate information across managers or establishments. Li et al. (2020) shows that an algorithm that explicitly puts weight on exploration about worker characteristics that are more uncertain—rather than simply extraction from previous experiences—can improve hiring diversity at no efficiency cost. In my setting, the exploration bonus directly affects the employer's payoff rather than affecting a hiring recommendation given to the employer, but the two measures essentially target the same issue and yield the same benefits.

Hiring Context.—First, I consider an *Equal* treatment investigating the impact of framing group *B* as a minority. The mechanism should operate regardless because it fundamentally arises from relative uncertainty about the productivity of groups, but minority status itself may affect group perceptions, for example, through stereotyping. In both the *Baseline* and *Equal* treatments, group *B* has 50 workers. In the latter, group *A* also has 50 rather than 150 workers, framing groups as equally sized but presenting group *B* identically across treatments.

Second, I consider a *Gender* treatment with a salient group characteristic, gender, using self-reported male (123) and female (77) workers.[38] In practice, employers enter the labor market with beliefs that may affect their hiring and learning from the start. While gender labels could make that characteristic particularly salient, gender is generally easily observable, an established literature documents its salience, and this simple framing allows for a direct comparison with other treatments. Moreover, while the task was chosen to be reasonably gender neutral, some employers evoked common stereotypes to motivate their beliefs, such as women having "less computer training," but being "more detail-oriented," suggesting a meaningful relationship with views held more broadly. Employers were also more likely to report that intelligence and experience were important or very important to explain differences in group productivity compared to the *Baseline* treatment, at 50 percent and 60 percent versus 25 percent and 38 percent.

Estimates in columns 1 and 2 of panel B of Table 3 were obtained from linear regressions of the total hiring of group *B* by an employer or their final bias in beliefs about the group's productivity on an indicator variable for an employer being assigned to the *Equal* treatment. Column 2 presents evidence of decreased bias by around 16 percent when groups are equally sized. The *Equal* treatment has average final beliefs of 8.8 versus 8.6 for the *Baseline* treatment. Column 1 indicates that these impacts are not driven by increased hiring. Online Appendix Table A3-4 suggests that they appear due to less updating following early negative experiences with group *B*. This finding is striking given how similar the two treatments are and the neutral framing of the uncertain group, suggesting that the minority label plays an important role in shaping perceptions. One potential explanation is that, when group *B* is framed as a minority, negative experiences trigger a negative association employers have between productivity and minority status.

Column 3 of panel B of Table 3 shows a 18 percent increase in final bias about female workers estimated from a regression of final bias in beliefs about group *B*'s productivity on an indicator variable for an employer being assigned to the *Gender* treatment. Employers assigned to the *Gender* treatment have lower final beliefs of 8.48, with strong heterogeneity across employer gender (8.67 and 8.16 for male and female employers). I then augment the specification in column 3 with an indicator for an employer being male and its interaction with being assigned to the *Gender* treatment. As shown in column 5, when the uncertain group corresponds to female workers, male employers report 9 percent higher beliefs about its productivity, although the estimate is only statistically significant at the 10 percent level. In contrast, there is little evidence of relationships between bias formation and employer characteristics or prejudice measures with artificial worker groups, as shown in online Appendix Table A3-5. Much of this difference across employer gender arises from the hiring task because initial beliefs are more similar (8.22 for female employers and 8.34 for male employers). Combined with the absence of a large difference in female

[38] Table 1 shows that both groups solved approximately nine puzzles on average with no statistically significant difference. Tests for differences in distribution also yield p-values above 0.39.

hiring across employer gender shown in column 4, this suggests that the difference lies in belief updating. Evidence that female employers form more negative biases is consistent with evidence that they evaluate female workers comparatively harshly (Ellemers et al. 2004; Bagues and Esteve-Volart 2010).

Lastly, online Appendix 3 investigates how employer behavior departs from Bayesian updating. I find that employers appear quicker to develop group associations than a Bayesian benchmark, consistent with stereotype formation amplifying experience-based discrimination.

IV. Conclusion

Evidence from surveys and recent studies supports the notion that discrimination can arise from employers developing inaccurate group perceptions, but this feature is absent from classical models of discrimination. I present a new model in which persistent, heterogeneous employer-biased beliefs about the productivity of a worker group arise from employers' individual hiring experiences. These biased beliefs can create discrimination against worker groups whose productivity is initially more uncertain to employers, like minority groups, even with expected profit-maximizing employers and equally productive worker groups, with no prior bias or prejudice, and without endogenous worker investments.

I then present the results of an online experiment finding support for the model's key predictions. Namely, negative experiences of an employer hiring a group lead to persistent negative biases specifically by decreasing subsequent hiring of that group and therefore learning about its productivity. I also show that the hiring context matters: whether a group represents a minority or whether the group label is arbitrary (color) or has real world saliency (gender) affects the extent to which biased beliefs arise. Further, by explicitly providing an origin for biased beliefs, the model generates clear predictions regarding the effectiveness of some policies to mitigate discrimination. I test some of these predictions in the experiment, showing that incentivizing employers to hire more from the uncertain group or providing them with additional information on its productivity mitigates biased beliefs. These findings provide a new lens to analyze the impact of diversity, equity, and inclusion (DEI) policies. Not only can they increase representation but promote employer learning about the productivity of minority groups, potentially leading to longer-term increases in both diversity and efficiency. Similarly, my findings suggest that the use of tools like hiring algorithms can mitigate discrimination if they decrease the extent to which employers rely on their own personal experiences when making hiring decisions.

This paper studies an intuitive feature of hiring that provides a new way to understand prejudice in the labor market as the result of interactions between groups distorting beliefs and behavior. Biased beliefs arise because employers learn from a selected sample of observations about worker productivity, selected by their own hiring, with implications for our theoretical understanding of labor market discrimination, empirical studies on the source of discrimination, and policy. Key insights from this paper likely carry over to other scenarios in which individuals make consequential decisions from experience. The increasing availability of data on decision-makers making repeated decisions in the labor market, criminal justice system, medicine, education, and financial or credit services suggests opportunities to investigate how experience-based discrimination arises in other important settings.

Appendix A. Proofs and Equilibrium Definition

A. Proofs

PROOF OF PROPOSITION 1:
Market clearing implies $\lambda_{\mathrm{mt}} = \mathrm{w}_{\mathrm{Bt}}(\Psi_t) - \mathrm{w}_{\mathrm{A}}$. Define $\lambda_{\mathrm{mt}} := \lambda_{\mathrm{t}}^{\mathrm{c}}$. From (3), employers with $\lambda_{jt} > \lambda_{\mathrm{t}}^{\mathrm{c}}$ strictly prefer to hire group B while those with $\lambda_{\mathrm{jt}} < \lambda_{\mathrm{t}}^{\mathrm{c}}$ strictly prefer to hire group A. Thus, $\lambda_{\mathrm{t}}^{\mathrm{c}}$ is the cutoff λ_{jt} for a B worker in period t. ∎

PROOF OF PROPOSITION 2:
The Bayesian central limit theorem implies that $\mu_B \to_d \mu$ as $K \to \infty$ with $K \leq t$ for employers with $\lambda_{jt} > \lambda_t^c$ as $t \to \infty$ under standard regulatory conditions on $G(\cdot)$ and $h(\cdot)$. For almost all of these employers, $\lambda_{jt} \to 0$ as $K \to \infty$ and $\lambda_{jt} \geq w_{Bt}(\Psi_t) - w_A$, implying $w_A \geq w_{Bt}(\Psi_t)$ asymptotically. From (3) and (4), fraction $1 - F_B$ of employers hire group A asymptotically, implying $\lambda_{jt} \leq \lambda_t^c$. Since their value of information $E_t\Big[V\Big(\psi'_{\mathcal{S}_{jt+1}}, \cdot\Big)\Big] - E_t\Big[V\Big(\psi_{\mathcal{S}_{jt+1}}, \cdot\Big)\Big]$ is weakly positive and $w_A \geq w_{Bt}(\Psi_t)$, then $E\big[\mu_B | \mathcal{S}_{jt}\big] < \mu$. ∎

PROOF OF PROPOSITION 3:
Define $\mathcal{E}_{Bt}$ as the set of employers with $\lambda_{jt} \geq \lambda_t^C$ in period t, with mass equal to F_B. Given a continuum of employers, there exists $\mathcal{Z}_{Bt+1} \subset \mathcal{E}_{Bt}$ with $\lambda_{jt+1} < w_{Bt} - w_A \leq \lambda_{jt}$. Suppose $w_{Bt+1} \geq w_B$, then $\mathcal{E}_{Bt+1} \subset E_{Bt}$, and the market doesn't clear. Thus, $w_{Bt+1} < w_{Bt}$ for all t.

Since w_{Bt} is strictly decreasing in t, showing $w_{Bt} \to c \in \mathbb{R}$ as $t \to \infty$ is equivalent to showing that w_{Bt} cannot fall below an arbitrarily low limit $\underline{c} > -\infty$. Employers with $\lambda_{jt} \leq \lambda_t^c$ have observed a finite number of signals (if any), have a strictly positive value of learning, and $E\big[\mu_B|\ _{jt}\big] > -\infty$. Define $\lambda_{\underline{j}}$ as the supremum λ_j for employers with $\lambda_{jt} \leq \lambda_t^c$ as $t \to \infty$, with $\lambda_{\underline{j}} = \underline{c}$. Then $w_{Bt} \geq \underline{c}$ for any t.

For any $\varepsilon > 0$, there exists t large enough such that fraction $F_B - \varepsilon$ of employers with $\lambda_{jt} \geq \lambda_t^c$ have value of learning smaller than ε and will hire group B in the limit. There also exists $t' > t$ arbitrarily large such that beliefs of employers hiring from group B at t' are almost entirely driven by signals observed between t and t': $\mu_B|\mathcal{S}_{t'j}$ follows approximately the same distribution as $\mu_B|\{\mathcal{S}_{t'j} \setminus \mathcal{S}_{jt}\}$. Given that $E\big[\mu_B|\{\mathcal{S}_{t'j} \setminus \mathcal{S}_{jt}\}\big] \to \mu$ almost surely, some employers with $\lambda_{jt} \geq \lambda_t^c$ have $E\big[\mu_B|\mathcal{S}_{jt}\big] < \mu$ and a value of learning smaller than ε, such that their λ_{jt} is below 0.[39] By market clearing, λ_{mt} is no greater than the infimum λ_{jt} of employers

[39] The probability that employer beliefs all converge in distribution to μ from above is 0 given a large number of employers.

with $\lambda_{jt} \geq \lambda_t^c$, implying $\lambda_{mt} = w_{Bt} - w_A < 0$ and $w_{Bt} < w_A$ for $t > t'$. Since w_{Bt} is strictly decreasing in t, then $c < w_A$. ∎

B. *Equilibrium Definition*

An equilibrium is a stochastic process over beliefs and a mapping from beliefs to wages. Given a continuum of agents on each side of the market, this corresponds to a deterministic Markov process with transition functions characterized by the following definition.

DEFINITION 1: *An equilibrium is a Markov process with a distribution over beliefs $\boldsymbol{\Psi}_t$ evolving according to a transition function $T : \Delta(\mathbb{R} \times \mathbb{R}_+) \rightarrow \Delta(\mathbb{R} \times \mathbb{R}_+)$, a wage function $w_{Bt} : \Delta(\mathbb{R} \times \mathbb{R}_+) \rightarrow \Delta\mathbb{R}$, and an initial state $\boldsymbol{\Psi}_0 \in \Delta(\mathbb{R} \times \mathbb{R}_+)$ such that every period:*

(*i*) *Employers make expected profit maximizing hiring decisions following Equation* (2) *and Proposition* 1 *for all* $\left(\psi_{s_{jt}}, w_{Bt}(\boldsymbol{\Psi}_t)\right)$.

(*ii*) *The labor market clears according to* (4).

(*iii*) *Employers update their beliefs:*

(*a*) *Those with* $\psi_{s_{jt}}$ *such that* $\lambda_{jt} < \lambda_t^c\left(w_{Bt}(\boldsymbol{\Psi}_t)\right)$ *hold posterior beliefs* $\psi_{_{jt+1}} = \psi_{_{jt}}$.

(*b*) *Those with* $\psi_{s_{jt}}$ *such that* $\lambda_{jt} \geq \lambda_t^c\left(w_{Bt}(\boldsymbol{\Psi}_t)\right)$ *hold posterior beliefs* $\psi'_{\mathcal{S}_{jt+1}}$ *derived according to Equation* (1).

REFERENCES

Agnew, Julie R., Hazel Bateman, Christine Eckert, Fedor Iskhakov, Jordan Louviere, and Susan Thorp. 2018. "First Impressions Matter: An Experimental Investigation of Online Financial Advice." *Management Science* 64 (1): 288–307.

Aizer, Anna, Ryan Boone, Adriana Lleras-Muney, and Jonathan Vogel. 2020. "Discrimination and Racial Disparities in Labor Market Outcomes: Evidence from WWII." NBER Working Paper 27689.

Aguirregabiria, Victor, and Jihye Jeon. 2019. "Firms' Beliefs and Learning: Models, Identification, and Empirical Evidence." *Review of Industrial Organization* 56: 203–35.

Aigner, Dennis J., and Glen G. Cain. 1977. "Statistical Theories of Discrimination in Labor Markets." *ILR Review* 30 (2): 175–87.

Altonji, Joseph G., and Charles R. Pierret. 2001. "Employer Learning and Statistical Discrimination." *Quarterly Journal of Economics* 116 (1): 313–50.

Arnold, David, Will Dobbie, and Crystal S. Yang. 2018. "Racial Bias in Bail Decisions." *Quarterly Journal of Economics* 133 (4): 1885–1932.

Arrow, Kenneth J. 1973. "The Theory of Discrimination." In *Discrimination in Labor Markets*, edited by Orley Ashenfelter and Albert Rees, 3–33. Princeton, NJ: Princeton University Press.

Ashenfelter, Orley, and Timothy Hannan. 1986. "Sex Discrimination and Product Market Competition: The Case of the Banking Industry." *Quarterly Journal of Economics* 101 (1): 149–73.

Audretsch, David B. 1991. "New-Firm Survival and the Technological Regime." *Review of Economics and Statistics* 73 (3): 441–50.

Bagues, Manuel F., and Berta Esteve-Volart. 2010. "Can Gender Parity Break the Glass Ceiling? Evidence from a Repeated Randomized Experiment." *Review of Economic Studies* 77 (4): 1301–28.

Banks, Jeffrey, Mark Olson, and David Porter. 1997. "An Experimental Analysis of the Bandit Problem." *Economic Theory* 10 (1): 55–77.

Bardhi, Arjada, Yingni Guo, and Bruno Strulovici. 2020. "Early-Career Discrimination: Spiraling or Self-Correcting?" Northwestern University Working Paper.

Bartoš, Vojtěch, Michal Bauer, Julie Chytilová, and Filip Matějka. 2016. "Attention Discrimination: Theory and Field Experiments with Monitoring Information Acquisition." *American Economic Review* 106 (6): 1437–75.

Becker, Gary S. 1957. *The Economics of Discrimination*. Chicago, IL: University of Chicago Press.

Bergemann, Dirk, and Juuso Välimäki. 2008. "Bandit Problems." In *The New Palgrave Dictionary of Economics*, Vol. 1–8, 336–40. London: Palgrave Macmillan.

Berinsky, Adam J., Gregory A. Huber, and Gabriel S. Lenz. 2012. "Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk." *Political Analysis* 20 (3): 351–68.

Bertrand, M., and E. Duflo. 2017. "Field Experiments on Discrimination." In *Handbook of Economic Field Experiments*, Vol. 1, edited by Abhijit Vinayak Banerjee and Esther Duflo, 309–93. Amsterdam: Elsevier.

Bertrand, Marianne, Dolly Chugh, and Sendhil Mullainathan. 2005. "Implicit Discrimination." *American Economic Review* 95 (2): 94–98.

Black, Sandra E., and Philip E. Strahan. 2001. "The Division of Spoils: Rent-Sharing and Discrimination in a Regulated Industry." *American Economic Review* 91 (4): 814–31.

Bohren, J. Aislinn, Alex Imas, and Michael Rosenberg. 2019. "The Dynamics of Discrimination: Theory and Evidence." *American Economic Review* 109 (10): 3395–3436.

Bohren, J. Aislinn, Kareem Haggag, Alex Imas, and Devin G. Pope. 2023. "Inaccurate Statistical Discrimination: An Identification Problem." Review of Economics and Statistics.

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer. 2016. "Stereotypes." *Quarterly Journal of Economics* 131 (4): 1753–94.

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer. 2019. "Beliefs about Gender." *American Economic Review* 109 (3): 739–73.

Buhrmester, Michael, Tracy Kwang, and Samuel D. Gosling. 2011. "Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data?" *Perspectives on Psychological Science* 6 (1): 3–5.

Campos-Mercade, Pol, and Friederike Mengel. 2024. "Non-Bayesian Statistical Discrimination." Management Science 70 (4): 2549–67.

Charness, Gary, and Peter Kuhn. 2011. "Lab Labor: What Can Labor Economists Learn from the Lab?" In *Handbook of Labor Economics*, Vol. 4A, edited by Orley Ashenfelter and David Card, 229–330. Amsterdam: Elsevier.

Charness, Gary, Luca Rigotti, and Aldo Rustichini. 2007. "Individual Behavior and Group Membership." *American Economic Review* 97 (4): 1340–52.

Chen, Yan, and Sherry Xin Li. 2009. "Group Identity and Social Preferences." *American Economic Review* 99 (1): 431–57.

Coate, Stephen, and Glenn C. Loury. 1993. "Will Affirmative-Action Policies Eliminate Negative Stereotypes?" *American Economic Review* 83 (5): 1220–40.

Cornell, Bradford, and Ivo Welch. 1996. "Culture, Information, and Screening Discrimination." *Journal of Political Economy* 104 (3): 542–71.

Denrell, Jerker. 2005. "Why Most People Disapprove of Me: Experience Sampling in Impression Formation." Psychological Review 112 (4): 951–78.

Ellemers, Naomi, Henriette Van den Heuvel, Dick de Gilder, Anne Maass, and Alessandra Bonvini. 2004. "The Underrepresentation of Women in Science: Differential Commitment or the Queen Bee Syndrome?" *British Journal of Social Psychology* 43 (3): 315–38.

Eyster, Erik, Matthew Rabin, and Dmitri Vayanos. 2019. "Financial Markets Where Traders Neglect the Informational Content of Prices." *Journal of Finance* 74 (1): 371–99.

Farber, Henry S., and Robert Gibbons. 1996. "Learning and Wage Dynamics." *Quarterly Journal of Economics* 111 (4): 1007–47.

Fershtman, Chaim, and Uri Gneezy. 2001. "Discrimination in a Segmented Society: An Experimental Approach." *Quarterly Journal of Economics* 116 (1): 351–77.

Fershtman, Daniel, and Alessandro Pavan. 2021. "'Soft' Affirmative Action and Minority Recruitment." American Economic Review: Insights 3 (1): 1–18.

Fudenberg, Drew, and David K. Levine. 1993. "Self-Confirming Equilibrium." *Econometrica* 61 (3): 523–45.

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): 1219–60.

Gneezy, Uri, Alex Imas, and John List. 2015. "Estimating Individual Ambiguity Aversion: A Simple Approach." NBER Working Paper 20982.

Hossain, Tanjim, and Ryo Okui. 2013. "The Binarized Scoring Rule." *Review of Economic Studies* 80 (3): 984–1001.

Hull, Peter. 2021. "What Marginal Outcome Tests Can Tell Us about Racially Biased Decision-Making." NBER Working Paper 28503.

Kahn, Lisa B. 2013. "Asymmetric Information between Employers." *American Economic Journal: Applied Economics* 5 (4): 165–205.

Keller, Godfrey, Sven Rady, and Martin Cripps. 2005. "Strategic Experimentation with Exponential Bandits." *Econometrica* 73 (1): 39–68.

Kirschenman, Joleen, and Kathryn M. Neckerman. 1991. "'We'd Love to Hire Them, but…': The Meaning of Race for Employers." In *The Urban Underclass*, edited by Christopher Jencks, and Paul E. Peterson, 203–32. Washington DC: Brookings Institution Press.

Komiyama, Junpei, and Shunya Noda. 2024. "On Statistical Discrimination as a Failure of Social Learning: A Multiarmed Bandit Approach." Management Science.

Kuleshov, Volodymyr, and Doina Precup. 2014. "Algorithms for Multi-armed Bandit Problems." arXiv: 1402.6028.

Lang, Kevin. 1986. "A Language Theory of Discrimination." *Quarterly Journal of Economics* 101 (2): 363–82.

Lang, Kevin, and Ariella Kahn-Lang Spitzer. 2020. "Racial Discrimination: An Economic Perspective." *Journal of Economic Perspective* 34 (2): 68–89.

Lang, Kevin, and Jee-Yeon K. Lehmann. 2012. "Racial Discrimination in the Labor Market: Theory and Empirics." *Journal of Economic Literature* 50 (4): 959–1006.

Le Mens, G., and J. Denrell. 2011. "Rational Learning and Information Sampling: On the 'Naivety' Assumption in Sampling Explanations of Judgment Biases." Psychological Review 118 (2): 379–92.

Lepage, Louis-Pierre. 2024. *Data and Code for: "Experience-Based Discrimination."* Nashville, TN: American Economic Association; distributed by Inter-university Consortium for Political and Social Research, Ann Arbor, MI. https://doi.org/10.3886/E192292V1.

Leung, Ming D. 2017. "Learning to Hire? Hiring as a Dynamic Experiential Learning Process in an Online Market for Contract Labor.' *Management Science* 64 (12): 5651–68.

Li, Danielle, Lindsey R. Raymond, and Peter Bergman. 2020. "Hiring as Exploration." NBER Working Paper 27736.

Lundberg, Shelly J., and Richard Startz. 1983. "Private Discrimination and Social Intervention in Competitive Labor Market." *American Economic Review* 73 (3): 340–47.

Malmendier, Ulrike. 2021. "FBBVA Lecture 2020: Exposure, Experience, and Expertise: Why Personal Histories Matter in Economics." *Journal of the European Economic Association* 19 (6): 2857–94.

March, James G. 1991. "Exploration and Exploitation in Organizational Learning." *Organization Science* 2 (1): 71–87.

Meyer, Robert J., and Yong Shi. 1995. "Sequential Choice under Ambiguity: Intuitive Solutions to the Armed-Bandit Problem." *Management Science* 41 (5): 817–34.

Miller, Conrad. 2017. "The Persistent Effect of Temporary Affirmative Action." *American Economic Journal: Applied Economics* 9 (3): 152–90.

Moore, Don A., Elizabeth R. Tenney, and Uriel Haran. 2015. "Overprecision in Judgment." In *The Wiley Blackwell Handbook of Judgment and Decision Making*, edited by Gideon Keren and George Wu, 182–209. New York, NY: Wiley.

Morgan, John, and Felix Várdy. 2009. "Diversity in the Workplace." *American Economic Review* 99 (1): 472–85.

Pager, Devah, and Diana Karafin. 2009. "Bayesian Bigot? Statistical Discrimination, Stereotypes, and Employer Decision Making." *Annals of the American Academy of Political and Social Science* 621 (1): 70–93.

Pettigrew, T. F., and L. R. Tropp. 2006. "A Meta-analytic Test of Intergroup Contact Theory." *Journal of Personality and Social Psychology* 90 (5): 751–83.

Phelps, Edmund S. 1972. "The Statistical Theory of Racism and Sexism." *American Economic Review* 62 (4): 659–61.

Pinkston, Joshua C. 2009. "A Model of Asymmetric Employer Learning with Testable Implications." *Review of Economic Studies* 76 (1): 367–94.

Reuben, Ernesto, Paola Sapienza, and Luigi Zingales. 2014. "How Stereotypes Impair Women's Careers in Science." *PNAS* 111 (12): 4403–08.

Sarsons, Heather. 2019. "Interpreting Signals in the Labor Market: Evidence from Medical Referrals." Unpublished.

Schary, Martha A. 1991. "The Probability of Exit." *RAND Journal of Economics* 22 (3): 339–53.

Schönberg, Uta. 2007. "Testing for Asymmetric Employer Learning." *Journal of Labor Economics* 25 (4): 651–91.

Watkins, Christopher John Cornish Hellaby. 1989. "Learning from Delayed Rewards." PhD thesis. University of Cambridge.

Wilson, William J. 1996. *When Work Disappears: The World of the New Urban Poor*. New York, NY: Alfred A. Knopf.