Statistical Discrimination and Optimal Mismatch in College Major Selection

Mary Kate Batistich Timothy N. Bond Sebastian Linde
Kevin J. Mumford

May 8, 2025

Abstract

We develop a model of college major selection in an environment where firms and students have incomplete information about the students' aptitude. Students must choose from a continuum of majors which differ in their human capital production function and can act as a signal to the market. Whether black students choose more or less difficult majors than similar white students, and whether they receive a higher or lower return to major difficulty, depends on the extent to which employers statistically discriminate. We find strong evidence that statistical discrimination influences major choice using administrative data from several large universities and two nationally representative surveys.

JEL Codes: J71, J15, I26

Batistich: Department of Economics, University of Notre Dame, 3060-I Jenkins Nanovic Hall, Notre Dame, IN 46556, USA; email: mbatisti@nd.edu. Bond: Department of Economics, Daniels School of Business, Purdue University, 403 Mitch Daniels Blvd., West Lafayette, IN 47907, USA; email: tnbond@purdue.edu. Linde: School of Public Health, Texas A&M, 212 Adriance Lab Rd. 1266 TAMU, College Station, TX 77843, USA; email: slinde@tamu.edu. Mumford: Department of Economics, Daniels School of Business, Purdue University, 403 Mitch Daniels Blvd., West Lafayette, IN 47907, USA; email: mumford@purdue.edu. We are grateful to Tyler Ransom, Jeff Smith, and seminar and conference participants at Georgia State University, the University of Oklahoma, the European Economics Association Annual Congress, the Midwest Economics Association Annual Meeting, the Mid-Midwest Applied Microeconomics Workshop, the Minorities in Economics Best Paper Award Virtual Session, the Notre Dame/Midwest Economics of Education Conference, the Royal Economic Society and Scottish Economic Society Annual Conference, the Society of Labor Economics Annual Meeting, the Southern Economic Association Annual Meetings, and the Western Economic Association International Annual Conference for their helpful comments. All remaining errors are our own.

1 Introduction

The wage difference between college graduates with high paying and low paying degrees is nearly as large as that between high school and college graduates (Altonji et al., 2012). It is natural then that researchers and policymakers are concerned about major choice, especially for underrepresented minority (URM) students. In fact, a central argument against affirmative action in university admissions is that it causes URM students to graduate in lower paying majors than they would have if they had instead attended a less prestigious institution (e.g., Arcidiacono et al., 2012, 2016). This "mismatch" hypothesis was recently cited by Supreme Court Justice Clarence Thomas in his concurring opinion on the *Students for Fair Admissions v. Harvard* decision which curtailed the use of racial admissions preferences in the United States.

In this paper, we show that mismatch in human capital investment can occur even in the absence of affirmative action when information is incomplete, with a specific application to college major choice. Drawing inspiration from both the mismatch literature and the statistical discrimination literature, we build a theoretical model that allows for racial differences in the precision of information about aptitude both by students and employers. We use this model to develop several novel empirical predictions which we test and confirm using three different datasets.

In our model, students choose from a continuum of college majors which augment their initial level of aptitude to create human capital. Students make choices based on noisy beliefs about their true ability, and black students have noisier beliefs than white students. Employers do not observe students' accumulated human capital, but do observe their major choice and two other unbiased pieces of information: course grades and a labor market signal. While grades are equally informative across race, the labor market signal is less precise for black workers, as in the statistical discrimination literature.

We show that student and employer information frictions have opposing effects on equilibrium mismatch.[1] Statistical discrimination raises employer reliance on observable information, like major choice, when evaluating black applicants. Just as in Lang and Manove (2011), this higher return to signaling pushes black students to "mismatch" by selecting majors that are more difficult than those chosen by white students with similar academic backgrounds. However, when black students make their major choices with less precise beliefs about their aptitude, it reduces the reliability of major choice as a productivity indicator. This in turn reduces the reliance of firms on major choice in their evaluations, reducing the

[1] Effects of incomplete information on both supply and demand sides in a labor market have also been studied empirically by Carranza et al. (2022) and others.

signaling value of major for black students and pushing black students to choose less difficult majors than similarly prepared white students. Whether black students take more or less difficult majors than white students will depend on which of these two forces is stronger.

We test this using administrative data from twelve large public universities and two nationally representative surveys on labor market outcomes. The evidence strongly suggests that statistical discrimination dominates student information frictions in major choice. First, we find that black students select and graduate in higher paying majors than white students with similar academic preparation. Consistent with the outcome of the signaling game when statistical discrimination is stronger, we find that the black-white difference in major difficulty is increasing in academic preparation. In addition, we find that the within-major racial wage gap is increasing in major difficulty. This is consistent with our model but stands in contrast to conventional wisdom that the black-white wage gap is lowest among the highest skilled workers (e.g., Lang and Lehmann, 2012). Our results hold for measures of difficulty based on labor market returns as well as course content, are true for both early career and prime age workers, and are robust to controls for institution quality.

While our results indicate that statistical discrimination is a more important factor, we nonetheless find evidence that black students have worse information than white students when making their human capital decisions. We show theoretically that black students will have a higher observed labor market return to grades only if they have chosen their major with less precise beliefs about their aptitude, and confirm this using data on early career wages of recent university graduates.

Despite the dramatic differences in labor market returns across college major, racial differences in major choice have received surprisingly little attention.[2] Arcidiacono et al. (2012) show that black students at Duke University are more likely to begin schooling in a science major than white students, but have lower rates of finishing a major in science. Arcidiacono et al. (2016) similarly find substantial gaps in preparation between URM students who finish a STEM degree and those who do not within the University of California system. Sovero et al. (2021) show minority students at the University of California, Los Angeles actually have higher rates of STEM persistence after controlling for preparation. Bleemer and Mehta (2023) document a trend toward lower paying degrees for URM students since the 1990s, which they attribute to an increase in major enrollment restrictions.

Our study differs from these papers in several important ways. First, we show that racial disparities in major choice are reversed after controlling for college preparation across a large set of universities of varying selectivity. Second, we document this finding across a fuller set of majors than typically studied in this literature. Third, we provide a theoretical foundation,

[2]See Altonji et al. (2016) for a recent review on the returns across major.

grounded in the statistical discrimination literature, for understanding why URM students would enroll in higher paying majors. Fourth, our model generates additional predictions on labor market outcomes which we confirm using two nationally representative data sets.

Our paper contributes more broadly to the literature on race in higher education. It has been widely observed that black students attend higher quality universities than white students with similar academic backgrounds, possibly due to race conscious admissions policies (e.g., Arcidiacono et al., 2011b). Whether this leads to better or worse outcomes for black students is a question of much debate.[3] Hinrichs (2012) finds that statewide bans on affirmative action in admissions cause a shift of black students away from highly selective institutions, but do not decrease the share of black degree holders in the population.[4] Hinrichs (2014) finds similarly that such bans raise the graduation rate for black students at selective institutions while lowering the overall number of black graduates from these same selective institutions. Arcidiacano et al. (2014) find that a specific ban introduced in California led to improvements in black college graduation rates, in part by causing black students to attend less selective institutions. However, Bleemer (2022) finds this same ban led to decreased wages for URMs as adults, driven by Hispanics. Looking across a wider set of states, Antman et al. (2024) find that affirmative action bans reduce long-run earnings and employment for Hispanic women. Yet, they find ambiguous results for other underrepresented groups, with possible positive effects for black men and notable heterogeneity across states. Mountjoy and Hickman (2021) instead compare students with identical application and admissions portfolios in Texas, and find no evidence that black students who attend more selective institutions perform worse in the labor market than those who attend less selective institutions.[5]

We also apply the insights of our model to reconcile the seemingly disparate findings in the university admissions literature. If black students anticipate facing statistical discrimination in the labor market, those who are admitted (with or without affirmative action) will optimally choose to enroll in more selective universities than similarly prepared white students because they are disproportionately rewarded for observable information. Affirmative action increases the choice set of colleges that black students can use to signal, which could potentially lower overall welfare for black students. However, all college decisions are correct

[3]The theoretical ambiguity is due to the "quality-fit tradeoff." Higher quality institutions have better resources, faculty, and peers, but less prepared students may struggle to learn if the teaching material is targeted towards a more advanced audience. For evidence that student quality and institution quality are complements see Light and Strayer (2000), Sallee et al. (2008) and Dillon and Smith (2020).

[4]Using a similar approach, Backes (2012) finds affirmative action bans led to modest decreases in the number of black graduates of public institutions.

[5]Mountjoy and Hickman (2021) note one exception: Black students who attend the two historically black universities (HBCUs) in Texas earn higher wages than those who attend more selective institutions.

in equilibrium. Moving a black student from a highly selective to a less selective institution will make that student worse off, due to a decrease in the market's beliefs about their aptitude. Thus, empirical strategies that compare marginal students across institutions will be unable to detect negative effects of mismatch when information about workers is incomplete. In contrast, affirmative action bans change all student investment choices, and thus market beliefs. If mismatch is detrimental, empirical strategies that use bans would be better able to detect any negative effects in an incomplete information environment.

The weakness of using affirmative action bans for identification however is that they can only identify mismatch relative to a policy of no affirmative action; they are not able to estimate the effect on mismatch by raising or lowering the amount of admission preferences. Thus even in an environment with complete information, it is difficult for such an approach to definitively *reject* mismatch, at least on the margin. Further, the results will always be local to the context, depending on the amount of admissions preferences given in that particular state and to each individual URM group.[6]

Our empirical results contribute to the growing body of evidence that student major selection responds to labor market incentives.[7] Previous studies have found that students switched majors in response to cyclical fluctuations in energy prices, the dot-com bust, the fracking boom, and the 2007-2008 financial crisis (Ersoy, 2020; Han and Winters, 2020; Weinstein, 2022). Similarly, Aalto et al. (2022) find the COVID-19 pandemic caused a decrease in applications to hospitality vocational programs by high school students in Sweden, while Ganguli et al. (2024) find the pandemic increased the demand for online courses promoting telework skills in Saudi Arabia. Blom et al. (2021) show that students enroll in majors with better labor market prospects during recessions. Our paper's empirical results add a key labor market characteristic that affects student major choice: statistical discrimination.

We also contribute important evidence on the role of anticipated discrimination in influencing premarket factors. Lang and Manove (2011) use a model closely related to our own to show that statistical discrimination will cause black workers to overinvest in education, an observable measure of productivity. Consistent with this, they find that black students obtain nearly a year more of education on average than white students with the same AFQT score.[8] Conversely, Coate and Loury (1993) show theoretically that statistical discrimination

[6]Others have noted the importance of such context when comparing undergraduate and law school admissions, see for example Dillon and Smith (2020).

[7]Labor market incentives are just one factor among many that influence college major choices. Other important influencers include student risk preferences and heterogeneous tastes (e.g., Wiswall and Zafar, 2015; Patnaik et al., 2022).

[8]Building on work by Arcidiacono et al. (2010), Lang and Manove (2011) also allow education to increase the precision of the labor market signal received by employers. Thus their model predicts that racial educational attainment will converge moving up the AFQT distribution, which they confirm empirically. While

will cause black workers to underinvest in unobservable measures of productivity. Fryer and Loury (2005) use a tournament model to show that affirmative action can increase effort provision by a disadvantaged group. Akhtari et al. (2024) find empirical evidence for this theory using data on student SAT scores before and after racial preference bans in university admissions. Our paper shows that major selection also responds to anticipated labor market discrimination.

Finally, we provide a contribution to understanding the mechanisms by which affirmative action may or may not cause mismatch. In a complete information environment, it is difficult to rationalize how racial admissions preferences, which simply increase the choice set of black students, could make black students worse off. Arcidiacono et al. (2011a) argue that because students have imprecise beliefs about their aptitude, and universities have private information of student qualifications, such preferences may induce some black students to attend universities they are not sufficiently qualified for. This is because students who are admitted primarily to achieve university diversity goals incorrectly interpret their admission as a positive signal of aptitude. Our study provides mixed support for this idea. We find strong empirical evidence that black students especially have noisy beliefs about their aptitude when making their major choice. However, we also show that these noisy beliefs reduce the signaling value of investment choice, which would cause black students to be *less* likely to attempt highly challenging investments. This is true regardless of the distribution of underlying aptitude. Whether the positive effect on individual beliefs from the admissions signal would dominate the equilibrium effects of the lower signaling value of investments would likely depend on the parameters of the model.

The rest of the paper is organized as follows. In Section 2, we introduce our model where students select a college major based on imprecise beliefs while taking into account the statistical discrimination behavior of future employers. In Section 3, we describe our three data sources. In Section 4, we empirically test our model's predictions on major selection and labor market outcomes. In Section 5, we discuss the implication of our results for empirical strategies designed to test the mismatch hypothesis. Section 6 concludes.

2 A Model Of Major Selection with Statistical Discrimination

We extend the insights of Lang and Manove (2011) to college major choice. Beyond institutional factors, we make two key departures. First, we allow students to have uncertainty about their aptitude, which affects the return on their investment, and allow the degree of

we find some suggestive evidence for this convergence effect for college majors, we are unable to formally test this theory, as we lack sufficient samples of black students who overlap with the highest SAT score white students.

that uncertainty to vary by race. Second, we augment the standard statistical discrimination framework by including an additional signal of aptitude (grades) which is observable to both employers and the researcher, and is equally precise across race.

There exists a large number of students who are either (b)lack or (w)hite and differ in a_i. We will interchangeably refer to a_i as "aptitude" or "college preparation" for ease of exposition, but it more accurately measures the stock of skills a student possesses when making postsecondary educational decisions. It reflects both innate ability as well as the impact of early childhood investments, primary and secondary school quality, etc. Students do not observe their aptitude. Based on high school performance, entrance exam scores, university admissions decisions, and other factors, student i of race r has normally distributed beliefs about their own ability with mean ρ_i and variance ς_r^2, where $\varsigma_b^2 \geq \varsigma_w^2$. This allows the possibility for black students to have less precise beliefs about their aptitude than white students, which could reflect differences in the quality of their high school preparation (e.g., less access to challenging advanced placement high school courses, worse guidance counseling) or home environment (e.g., parents with less college experience to help guide postsecondary decisions). Following Arcidiacono et al. (2011a), it may also reflect that racial preferences in admissions could make acceptance decisions less reliable signals of aptitude to minority students.[9]

The twice differentiable cdf $F_r(\rho)$ governs the distribution of ρ across students over the bounded interval $[0, 1]$.[10] We impose $f_r(\rho) > 0, \forall \rho \in [0, 1]$ but otherwise make no assumption on this distribution, or on differences in the distribution of ρ by race. Thus we allow that racial admissions preferences or population-level differences in preparation may lead to potentially large racial differences in the distribution of ρ among college students.

In period 1, students select from a continuum of investments m which differ in their human capital production function. In our empirical section m will represent college major choice, but our arguments would follow for any observable investment, including university quality. A student who selects m will produce p_i when they enter the labor market, where

$$p_i = ma_i - m^2. \tag{1}$$

[9]While Arcidiacono et al. (2011a) do not explicitly model the impact admissions decisions can have on the variance of student beliefs, racial differences in said variance would be a natural consequence when students know that white students are evaluated only based on their perceived match with the university, while black students are also evaluated on their ability to help the university reach its diversity goals.

[10]While the lower bound in particular could reflect that college students must pass some minimum observable aptitude criteria to gain admission to secondary education, we use a bounded distribution primarily for tractability.

Thus m is indexed by its complementarity with a. Further, it is straightforward to see that

$$\arg\max_{m} ma_i - m^2 = \frac{a_i}{2}. \tag{2}$$

and thus choosing m above this point will lead to lower levels of human capital accumulation. The linear in a functional form is necessary for a tractable characterization of the equilibrium.[11] However, the intuition should follow for any single-peaked human capital production function where a and m are complementary.[12] We will refer to higher levels of m as being "more difficult", again for ease of exposition.

Students enter the labor market in period 2. Employers do not observe p, a, or ρ, but they do observe m. In addition they observe an unbiased signal of a student's aptitude s:

$$s_i = a_i + \epsilon_i, \tag{3}$$

where ϵ_i is normally distributed with mean zero and variance σ_r^2.[13] This reflects information that is learned, for example, from an interview, and is unobservable to the researcher. As is standard in the statistical discrimination literature, employers are better able to interpret this information for whites, so that $\sigma_w^2 \leq \sigma_b^2$. This reflects, among other things, communication differences between white employers and black potential employees (Lang, 1986).

Employers also observe a student's grades, g_i, a second unbiased signal of their aptitude:

$$g_i = a_i + \zeta_i, \tag{4}$$

where ζ_i is normally distributed with mean zero and variance ϱ^2, and is independent of ϵ_i. There are two key differences between g and s. First, g is equally precise for black and white students. Second, g is observable to the researcher.

Equilibrium requires that both students and employers make optimal choices. Denote

[11] Specifically it ensures that we can use conjugate priors for the normal distribution to calculate employer beliefs.

[12] In an earlier version of this paper, we derived our main predictions in an environment where there were no differences in uncertainty about aptitude by race for a general human capital production function with those properties.

[13] We choose to have the market observe signals of a rather than p because our choice of human capital production function causes the equilibrium variance of p to be increasing in m. Thus, holding the variance of the signal fixed, any signal of p would be relatively more informative for higher levels of m. This would generate additional testable predictions and is consistent with assumptions made on the relationship between signal quality and education choice in, for example, Arcidiacono et al. (2010) and Lang and Manove (2011). However these predictions would *only* be due to our functional form choice, and could easily be *reversed* by choosing a functional form that caused the variance of productivity to be decreasing in m. By having the market instead observe signals of a, the functional form assumption does not influence the relative informativeness of the signals. Note also that in equilibrium the expected value of a could be easily computed by an employer with knowledge of m and the expected value of p.

π_r as the race-specific function which maps from observable information to employer beliefs, w_r as the race-specific wage function, and M_r as the race-specific function which maps from aptitude to educational investment. We define an equilibrium as follows:

Definition. *An equilibrium is a set of functions π_r, w_r, and M_r such that*

1. *w_r generates zero expected profit for firms given π_r.*
2. *M_r maximizes expected utility given w_r.*
3. *π_r is defined by Bayes' rule whenever possible.*

Denote $P_r(m)$ as the inverse of $M_r(\rho)$. Following Lang and Manove (2011) we restrict attention to separating equilibria which are "well-behaved" as defined below.

Definition. *A well-behaved equilibrium is an equilibrium with the following properties:*

1. *M_r is smooth, continuous, differentiable, and monotonically increasing in aptitude, a.*
2. *For any major, m, which is not utilized by any students of race r in equilibrium, $P_r = 0$.*

2.1 Employer Beliefs and Wages

Note that in a well-behaved equilibrium, college major selection reveals a student's aptitude beliefs, ρ, to the market. Since student beliefs are unbiased, the distribution of aptitude for students of race r with major m is normally distributed with mean $P_r(m)$ and variance ς_r^2. We can then apply Bayes' rule to find employer beliefs conditional on m, s, and g for any m that is used in equilibrium,

$$\pi_r(m, s, g) = \tau_r^{-1} \left[\varsigma_r^{-2} P_r(m) + \sigma_r^{-2} s + \varrho^{-2} g \right], \tag{5}$$

where $\tau_r \equiv \varsigma_r^{-2} + \sigma_r^{-2} + \varrho^{-2}$ is the precision of the market's posterior beliefs and follows from Bayesian updating with a normally distributed prior and signals. From the zero profit condition wages are

$$w_r(\pi_r) = m\pi_r(m, s, g) - m^2. \tag{6}$$

2.2 Optimal Major Selection and Educational Outcomes

Now consider a student's optimal investment problem:

$$\max_m E_r(w|m, \rho), \tag{7}$$

where $E_r(w|m,\rho)$ is the expected wage for a student of race r with beliefs ρ who attempts investment m,

$$E_r(w|m,\rho) = m\tau_r^{-1}\left[\varsigma_r^{-2}P_r(m) + \left(\sigma_r^{-2} + \varrho^{-2}\right)\rho\right] - m^2. \tag{8}$$

This follows from taking the expectation of (6), recognizing that both s and g are equal to ρ in expectation. In choosing a more difficult major students benefit from a "sheepskin" effect $[\varsigma^{-2}P_r(m)]$, but beyond a certain point, students bear a cost of lower actual human capital from being in a major that is more difficult than optimal for their expected aptitude. The benefit of the sheepskin effect depends on the weight employers place on a student's major when forming their posterior, which is in turn determined by the informativeness of major relative to grades and the signal. It is larger when grades and the signal are less precise (ϱ, σ_r), but smaller when students choose their major with less precise information about their aptitude (ς_r).

Proposition 1. *Denote $M^*(\rho)$ as $\arg\max_m E[p]$. In any well-behaved equilibrium, $M_r(0) = M^*(0) = 0$, and $M_r(a) \geq M^*(a), \forall a > 0$.*[14]

Proposition 1 follows from employer belief structures in well-behaved equilibria and says that students with the lowest beliefs about their aptitude will select the least difficult major. These students do not receive a benefit from choosing a higher m than the full-information optimum because they receive no sheepskin effect. In equilibrium, employers believe the least difficult major that is utilized must indicate workers of the lowest belief type, and therefore the lowest type workers will want to choose their full-information optimal major.

Proposition 2. *In equilibrium, $M_r(\rho)$ can be characterized by the differential equation*

$$\frac{\partial M_r(\rho)}{\partial \rho} = \frac{M_r(\rho)}{2M_r(\rho) - \rho}\frac{\varsigma_r^{-2}}{\tau_r}.$$

Propositions (1) and (2) characterize the full set of equilibrium major choices. The next propositions summarize how incomplete information and statistical discrimination can lead to racial differences in these choices.

Proposition 3. *If $\sigma_w^2 = \sigma_b^2$ and $\varsigma_w^2 < \varsigma_b^2$, black students attempt less difficult majors than white students conditional on ρ for all $\rho > 0$.*

Proposition 4. *Provided σ_b^2 is sufficiently larger than σ_w^2, black students attempt more difficult majors than white students conditional on ρ for all $\rho > 0$.*

[14] Proofs of this and all other results can be found in Appendix A.

Propositions 3 and 4 illustrate the tension between racial differences in student information and racial differences in employer information. When black students have less precise beliefs about their aptitude, their major choice becomes a less reliable indicator of their productivity to employers. Thus employers will put less weight on major choice and more weight on the labor market signal and college grades when making wage offers. This in turn reduces the sheepskin effect earned for black students when choosing a more difficult major, which incentivizes black students to take less difficult majors than white students. Put another way, if black students have worse information about their aptitude than white students, it will lead black students to be *less* overmatched than white students.[15]

Statistical discrimination has the opposite effect. When the labor market signal is less precise for black students, employers will put relatively more weight on observable information, including major choice, in their evaluations. This increases the sheepskin effect, which in turn induces black students to take more difficult majors. When statistical discrimination is sufficiently strong, the sheepskin incentives will dominate, and black students will optimally choose majors that are more mismatched than white students.

Our model predicts no differences in major choice for the least academically prepared students, and that a racial difference will emerge as we move up the preparation distribution. The direction of this difference will depend on which information force is more important for black students' investment choices. We will test this in our empirical section.

2.3 Labor Market Returns to Major and Grades

We will now explore the labor market implications of our model of major choice with specific focus on the equilibrium returns to majors and grades.

Proposition 5. *If $\sigma_w^2 = \sigma_b^2$ and $\varsigma_w^2 < \varsigma_b^2$, black students earn higher wages than white students conditional on m for all $\rho > 0$. Provided σ_b^2 is sufficiently larger than σ_w^2, black students earn lower wages than white students conditional on m for all $\rho > 0$.*

Proposition 5 follows directly from Propositions 1, 3, and 4. When black students have worse information about their aptitude than white students, and the market signals across race are similarly precise, black students select less difficult majors than white students with the same ρ. Since productivity is decreasing on the margin in m, black students are thus more productive conditional on major at graduation. As the labor market signal becomes relatively less precise for black students, the reverse becomes true. The value of the sheepskin

[15]Note that Proposition 1 still holds, so this will not induce black students to "undermatch" by choosing majors that are less than productivity maximizing.

effect increases, causing black students to select more difficult majors than white students with the same ρ, and thus become less productive at graduation conditional on major.

Proposition 6. *If $\sigma_w^2 = \sigma_b^2$ and $\varsigma_w^2 < \varsigma_b^2$, the observed labor market return to major difficulty for black college graduates is higher than for white college graduates. Provided σ_b^2 is sufficiently larger than σ_w^2, the observed labor market return to major difficulty for black college graduates is lower than for white college graduates.*

The equilibrium of the signaling game is non-distortionary for the lowest abilities and majors, but both imprecise information about aptitude and statistical discrimination induce racial differences for higher m. The former induces a productivity gap in favor of black students for more difficult majors, while the latter pushes towards a productivity gap in favor of whites. Whichever force is stronger will determine the direction of the observed return to major.

Note that this prediction is on the cross-sectional *correlation* between majors and wages, *not* the causal effect of major choice. Changing major choice causes both a change in market perceptions of aptitude and a change in accumulated human capital, with the net effect depending on the relative importance of each in wage determination. While increasing m has a smaller effect on market beliefs about black workers due to equilibrium sorting patterns, black workers see a larger marginal gain from increasing those beliefs due to a higher σ_b^2 (when statistical discrimination dominates). With cross-sectional data, we observe only the equilibrium wages, which are equal to market beliefs of average productivity by major and race.

The previous propositions will allow us to test whether black students have less precise information about their aptitude or whether statistical discrimination is strong. However, they do not allow us to test independently for whether there are racial differences in the quality of information about aptitude. The next proposition provides an independent test using the observed return to college grades.

Proposition 7. *If $\varsigma_w^2 < \varsigma_b^2$, the observed return to g will be higher for black college graduates than for white college graduates.*

To understand the proposition, one must think about how the estimated regression results are influenced by the incomplete information possessed by the researcher. Following standard intuition from statistical discrimination, when σ_b^2 is relatively large, grades have a stronger *causal* effect on black students' wages because employers put more weight on observable information and grades are observable. However, this will not come through in regression estimates when only g and m are observed. We cannot observe s, so our regressions will be estimates of $E[[w|m,s,g]|m,g]$. This is equal to $E[w|m,g]$ by the law of iterated expectations, and σ_b^2 and σ_w^2 have no impact on this expectation. However, if black students have worse information about their aptitude when choosing their major, g becomes a more reliable predictor of productivity and wages relative to m. Thus we will find a stronger relationship between g and wages if and only if $\varsigma_b^2 > \varsigma_w^2$.

Finally, we close by emphasizing that our results do not in any way depend on the distribution of ρ within race. Even for very skewed distributions, which might result from diversity preferences in admissions or differences in pre-market factors, all of our results hold.

3 Administrative and Survey Data

In this section we describe our three main data sources as well as how we construct our major difficulty measures.

3.1 Major Difficulty Measures

In our model, we classify the "difficulty" of educational investments on an index, m, related to their complementarity with aptitude in the production of human capital. We construct three measures to translate this idea to our empirical analysis of college majors.

The first two are constructed from labor market outcomes in the American Community Survey (ACS) from 2011 to 2021 excluding the year 2020.[16] We aggregate field of degree (college major) to 173 categories for all individuals holding a bachelor's degree or above. We adopt a similar approach to Bleemer and Mehta (2023), and compute the residuals from a regression of real log earnings on indicators for major as well as age and year fixed effects on a sample of white, prime age (25-54 year old), native-born, full-time, year-round, employed workers with at least a bachelor's degree. Our results are robust to instead using all prime age workers, or using only white men.[17] Our "Wage Return" measure is simply the average value of these residuals for each major, while "Percentile Return" is the major's percentile ranking in the distribution of average residuals. This approach follows naturally from our model where, in equilibrium, the highest m educational investments are chosen by the highest a students, and produce the highest p workers, who receive the highest wages.

We also construct a separate measure of major difficulty that relies on course content rather than adult outcomes. We use administrative student transcript records from 12 large public universities, which we refer to as the "state schools sample," and calculate the fraction of course credits in STEM for the average graduate of each major. We describe these data

[16]We exclude 2020 due to potential sampling difficulties related to the COVID-19 pandemic.

[17]See Appendix Tables C.2, C.3, C.4, and C.5.

more fully in Appendix B. We report the values for all three of our measures for each major in Appendix Table C.1.

We again emphasize the use of the word "difficulty" is for ease of exposition. It is immaterial whether our difficulty variables measure "true" difficulty in some absolute sense. Instead what we require is that students with higher academic preparation select majors with higher values on our metrics, and that majors with higher values on our metrics yield higher wages post-graduation.[18] Both of these are easily verified empirically for all three of our measures. Nonetheless, we also note that our measures are highly correlated with efforts to measure true major difficulty. Novik (2023) uses online ratings of college professors to produce an index of college major difficulty for 144 majors, a subset of the majors used in our analysis. His measure has a correlation of 0.82 with our wage return measure, 0.81 with our wage percentile measure, and 0.65 with our STEM measure. Further, we find similar results when using his measure on our subsample of students who choose majors for which he measures difficulty.[19]

3.2 State Schools Sample

The state schools sample is constructed from administrative student transcript records from 12 large public universities. See Appendix B for details. Table 1 reports summary statistics for this sample. We restrict attention to students who identify as black or white, and exclude students without a reported SAT or ACT test score.[20] The primary advantage of these data is the large sample size, with over 900,000 student records. Black students have lower graduation rates, are disproportionately female, and have lower high school and college GPAs. Strikingly, despite having on average 111 point lower SAT scores, black students initially enroll in majors that have a higher wage return on both the return and percentile measures during their first year of college. Consistent with Arcidiacono et al. (2012)'s findings for Duke, these differences are reversed by graduation, with whites graduating in more difficult majors by all three of our measures. We also see black students are more likely to be enrolled in chemistry, biology, or business while they are less likely to be enrolled in history, English, or agriculture.

[18]The latter reflects the wage equilibrium if employers believe that these majors contain on average more productive workers.

[19]These results are available upon request.

[20]To convert ACT scores to SAT scores, we use the ACT to SAT Concordance published by the ACT Education Corp. at www.act.org/content/act/en/products-and-services/the-act/scores/act-sat-concordance.html

3.3 American Community Survey Wages

To test our model's predictions on wages, we return to our ACS sample. For these analyses, we restrict our sample to working age (25 to 54) native non-Hispanic white and black workers with at least a bachelor's degree who were employed full-time year-round in the previous year. Table 2 displays descriptive statistics for this sample. We find that the black workforce has a higher fraction of female workers than the white workforce, in line with well-known racial differences in labor force participation (Neal, 2004). We also observe a substantial racial earnings gap of $16,250 annually (a 0.26 difference in log earnings). As in the state schools sample, black workers have degrees in less difficult majors than white workers.

3.4 The Baccalaureate and Beyond

The biggest weakness for our purposes in the ACS is a lack of information on college quality. One possibility then is that any racial differences we find in the labor market returns to major choice would be due to differences in university enrollment patterns between black and white students. A central concern of the affirmative action and mismatch literature is that affirmative action in admissions leads black students to graduate in lower-return majors than they would have had they attended a less selective college. (e.g., Arcidiacono et al., 2016).[21] We therefore provide additional evidence from the Baccalaureate and Beyond 2008/18 (B&B), a nationally representative longitudinal study of 2007-2008 college graduates. See Appendix B for details on data and sample construction.

We show descriptive statistics in Table 3. Similar to what we observe in the state school data, black students are more likely to be female and graduate with a lower GPA than white students. There is a 150 point racial gap in average SAT scores. In raw terms, the racial wage gap in each year is much smaller than the unconditional racial gap in the United States. This reflects both the youth of the sample, as well as the fact that the racial wage gap is generally thought to be lower in more educated individuals (Lang and Lehmann, 2012). Black students graduate in majors that are more difficult according to our wage percentile measure, but less difficult in terms of STEM courses. The largest disadvantage of using the B&B data is the much smaller sample size with only about 11,500 individuals in the sample.[22]

[21] Note that our results would not be biased if black students attended worse colleges than white students, so long as black major selection is uncorrelated with college quality. The college quality effect would be accounted for by the black indicator.

[22] The NCES requires that the number of observations in each table is rounded to the nearest 10.

4 Testing for Optimal Mismatch

We now test our model's main predictions in the environment of college major selection. The major itself is our measure of investment, while we use SAT scores, a measure of college preparation, as the stand-in for our model's aptitude parameter.

4.1 Academic Preparation and Major Selection

Our model makes a specific prediction on the relationship between major choice and academic preparation. There should be no racial differences in major choice for the least prepared students (see Proposition 1). However, if statistical discrimination is sufficiently strong, black students will select more difficult majors than white students as we move up the academic preparation distribution (see Proposition 4). If instead student information frictions are stronger, we would expect white students to select more difficult majors than black students for higher levels of academic preparation.

We first analyze this in raw means for our state schools sample by plotting the relationship between SAT scores (in 25 equal sized bins for each race) and the major percentile return in Figure 1. In panel A, we show first-year major selection among all students. Consistent with our model, black and white students at the very bottom of the SAT distribution initially select similar majors. However, as we move up the SAT distribution, black students rapidly overtake white students in percentile return. There is possibly some convergence at the top of the SAT distribution but we caution that the upper SAT bins are much wider for black students.[23] Many students change majors during college, and less than half in our sample ultimately graduate. In panel B, we instead consider the graduation major among graduates. We saw earlier that white students graduate in more difficult majors than black students (Table 1). However, once we account for student SAT scores, the racial gap reverses.

The state schools sample is nearly ideal for testing our model's predictions regarding major choice as it is a large administrative data set spanning several different universities across a long time period. One shortcoming is its lack of national representation. We therefore also test our predictions on the B&B sample, which is a nationally representative survey of a single graduating cohort. Figure 2 reports the raw relationship between SAT score (in 20 equal sized bins) and the major percentile return in the B&B data for graduates.[24] The figure is much noisier than Figure 1 because of the smaller sample size, but the relationship

[23]One explanation for this convergence is provided by Lang and Manove (2011), who show in their model of statistical discrimination with education choice that if higher levels of education also make the productivity signal more precise, as proposed by Arcidiacono et al. (2010), then racial differences in education choice should converge for high levels of aptitude.

[24]The B&B does not include first-year major.

has a similar pattern. Students with the lowest SAT scores appear to select similar majors regardless of race. As we move up the SAT distribution, black students tend to graduate in higher return majors than white students.

To formally test our prediction, we estimate:

$$\text{Major}_{ijt} = \beta_1 \text{Black}_i + \beta_2 \left(\text{Black}_i \times \text{SAT}_i\right) + \boldsymbol{\theta X_i} + \gamma_{jt} + \epsilon_{ijt} \tag{9}$$

where the subscript i indicates the individual student, j the educational institution, and t is the year of enrollment. Black_i is an indicator for the student having identified as black. X_i is a set of individual characteristics. γ_{jt} is a vector of institution-by-enrollment year fixed effects. Major_{ijt} is one of our three measures of major difficulty for the primary major selected by student i in their first year (first-year major) or for the student's primary major at graduation (graduation major).

We report estimates of Equation 9 using both the state schools and B&B data in Table 4. Due to differences in data availability, the two samples offer slightly different control variables. For the state schools sample, student characteristics include age at matriculation, a female indicator, and a transfer student indicator. For the B&B sample, student characteristics only include the student's age at matriculation and a female indicator. All specifications across both samples include student SAT score fixed effects. The state schools sample controls for institution-by-start-year fixed effects. As the B&B sample is drawn from a single cohort and has few observations within any single institution, we include the 25th and 75th math and verbal SAT scores at the institution as controls for institution quality.[25]

Panel A of Table 4 uses our wage return measure as the outcome variable. Consistent with Table 1, we find in column (1) that black students enroll in first-year majors with a 3.1 log point higher residual wage than white students with the same SAT score. In column (2) we add to the SAT fixed effects an interaction between SAT and race. The racial gap in major selection is increasing in SAT score, which is consistent with our model when statistical discrimination is sufficiently strong. Specifically, our estimates in panel B for the major percentile return imply that the average black student enrolls in a major that is 4.8 percentile points higher in the earnings distribution, and that this gap grows by 0.6 percentile points for every 100 point increase in SAT score. We find similar results in columns (3) and (4) when we look at graduation major for the sample of college graduates. In columns (5) and (6) we turn to graduates from the B&B sample and again find similar results.

In panel C of Table 4 we instead use our STEM courses major difficulty measure. Similar

[25] We find similar point estimates when controlling for institution fixed effects in the B&B sample, but with larger standard errors due to the small number of student observations per institution. See Appendix Table C.6.

to our findings in panels A and B, we find that black students select first-year majors with 4.4 percent more STEM course credits on average than white students with the same SAT score, and this gap is increasing in SAT score (column 6). These results also hold for graduation major selection across both samples. In sum, we find strong evidence in Table 4 consistent with statistical discrimination as the driving force in student major choice.[26]

One advantage of focusing on majors rather than university enrollment is that major, as a student choice, will be less sensitive to racial admissions preferences. Nonetheless, our results may still be confounded by major-specific recruiting efforts to increase diversity, particularly among the most competitive degree offerings. The common link for universities in our state school sample is that all are engineering-focused. We therefore test for this by excluding engineering majors and re-estimating Equation 9 in Appendix Table C.8. Our results are similar.

4.2 Major Selection and Career Outcomes

Our model makes two further predictions on racial differences in career outcomes. If statistical discrimination is sufficiently strong, as suggested by the results in Section 4.1, black workers will earn less than white workers who graduated with the same college major (see Proposition 5), and this racial wage gap will grow in major difficulty (see Proposition 6). To test this, we first use the ACS to estimate:

$$Y_{irst} = \alpha_1 \text{Black}_i + \alpha_2 \text{Major}_i + \alpha_3 \left(\text{Black}_i \times \text{Major}_i\right) + \boldsymbol{\theta X_i} + \gamma_{rs} + \delta_t + \epsilon_{irst} \tag{10}$$

where subscript i is for the individual, r indicates race, s indicates state of residence, and t indicates time. Y_{irst} is individual i's log earnings in year t. X_i is a set of individual controls. γ_{rs} is a set of possibly race-specific state fixed effects. δ_t is a set of time fixed effects. Black_i is an indicator for the student having identified as black. Major_{ijt} is one of our three measures of major difficulty for the individual's primary major at graduation. If statistical discrimination is sufficiently large, our model predicts $\alpha_1 < 0$ and $\alpha_3 < 0$. Black college graduates should have lower wages than white graduates in the same major, and the measured return to major difficulty should be higher for white graduates.

In columns (1) through (3) of Table 5, we estimate equation (10) using our ACS sample and cluster the standard errors by graduation major. In panel A we use the wage return as the measure of major difficulty. With only a basic set of controls (gender, age, and age-squared) we find strong evidence for both predictions our model. Black graduates earn 22% lower wages than white graduates in the same major, and have an observed return to

[26]Controlling for high school GPA leaves the results unchanged, as shown in Appendix Table C.7.

major difficulty that is 33.1% lower than white graduates. This result is unchanged with the addition of state and year fixed effects in column (2) and race-specific state fixed effects in column (3). Panels B and C repeat this analysis using the percentile return and STEM courses measures of major difficulty, respectively. Our results are similar.

While the ACS offers a large and nationally representative sample, it does not contain information on the university the worker attended. To ensure our estimates are not driven by differences in the quality of the degree-granting institution, we turn to the B&B data in columns (4) and (5) of Table 5. We control for the 25th and 75th percentile math and verbal SAT scores at the institution to control for institution quality.[27] The cost of using the B&B data is a much smaller sample which is limited to early career outcomes for a specific cohort. We use up to three earnings observations (at 1, 4, and 10 years after graduation) for each individual and again cluster the standard errors by graduation major. In column (4), we find similar though somewhat smaller point estimates, which perhaps reflects the youth of the sample, and we lose statistical significance on the interaction term when using the STEM courses measure. However, our estimates remain consistent with our model.[28]

Our model's predictions on wages are driven by the interaction between student major choice and market beliefs. Black students choose more difficult majors than similarly prepared white students, for which they are a worse match. This leads black students to graduate with less human capital than white students on average in the same major. As the market cannot observe human capital, but can observe major and race, firms pay lower wages to all black workers. One concern then is that our empirical results are entirely driven by the within-major differences in preparation rather than the equilibrium effects of incomplete information. We test this in column (5) of Table 5 by including SAT fixed effects. If anything, this strengthens the results, consistent with the importance of statistical discrimination.

Our results so far are consistent with an environment where labor market statistical discrimination plays a larger role in determining black major choices than incomplete information about aptitude. However, they do not necessarily mean that there are no racial differences in initial information quality. Proposition 7 provides a direct test of this hypothesis, which we implement in column (6) by adding college GPA and college GPA interacted with black. Unsurprisingly we find that higher GPA students earn higher wages as adults. Consistent with Proposition 7, we also find a larger GPA effect for black students, confirming that black students have less precise beliefs about their aptitude when making their major choices. Including college GPA in column (6) only strengthens the evidence for statistical

[27] We find similar point estimates when controlling for institution fixed effects in specifications using the B&B sample, but with larger standard errors due to the small number of student observations per institution. These results are included in Appendix Table C.6.

[28] These results are robust to controlling for high school GPA. See Appendix Table C.7

discrimination. Black graduates earn lower wages than white graduates in the same major and have a lower observed return to major difficulty than white graduates.

4.3 Testing Alternative Explanations: Race or SES?

One reason a student may pursue a more difficult, higher paying major is financial need. For example, low socioeconomic status (SES) students may be less likely to select "risky" majors with lower expected payoffs, or that require graduate school (Monaghan and Jang, 2017). As black students come from lower average SES backgrounds than whites, this provides a potential alternative mechanism for our empirical results. It is however difficult to argue that low SES whites face statistical discrimination, at least to the extent faced by blacks.[29] Thus, we can test our theory both by analyzing whether our racial effects hold after accounting for measures of SES (and thus whether they hold for both high and low SES blacks), as well as by comparing the outcomes of low SES whites to that of blacks.

Unfortunately none of our data have direct measures of childhood conditions or SES background. However both the state schools sample and the B&B include students' home zip codes. We therefore test this alternative hypothesis by including controls for three zip code SES characteristics from the student's childhood years: median household income, median education, and income mobility. Median income and education measures are taken from the Decennial Censuses and the ACS, while the income mobility measure is from Opportunity Insights (Chetty et al., 2018).[30]

We begin with graduation majors in the state schools sample. As not all students have zip code data we have fewer observations for this exercise. In column (1) of Table 6 we reproduce column (4) of Table 4 for this sample, using wage return as our measure of difficulty. Reassuringly, the change in sample has little impact on our results. In column (2) we

[29] The most common argument for why blacks would face stronger statistical discrimination is rooted in differences in language usage, which may be difficult to interpret by white employers (e.g., Lang, 1986). While there may be differences in English dialects between low and high SES whites, it seems unlikely they would be as severe as differences between Standard American English and African-American English. Bond and Salisbury (2018) argue that those outside of a region are unable to ascertain the information content of within-region variation in white dialects.

[30] We connect home zip code to ZCTA using the UDS mapper (https://udsmapper.org/zip-code-to-zcta-crosswalk/), and then merge it with IPUMS NHGIS Data (Manson et al., 2023) at the ZCTA level. We use the 1990 and 2000 Censuses and the 2008-2012 (as 2010) 5-year ACS data. The 1990 Census is available at the zipcode level so we convert it to ZCTA using the crosswalk. We use 1990 variables for college start years up to 1999, 2000 variables for college start years up to 2009, and 2010 variables for college start years up to 2019. We also connect home zip code to county using the HUD crosswalk from first quarter 2010 (https://www.huduser.gov/apps/public/uspscrosswalk/home), and then merge it with county-level data from Opportunity Insights (https://www.opportunityatlas.org/). Income mobility is measured by county and is the share of individuals whose parents' incomes were at the 25th percentile that are in the top 20% of household incomes at age 35.

include the median household income control as well as its interaction with that student's SAT score. In contrast to the alternative hypothesis, we find that students from wealthier zip codes graduate in *more* difficult majors than those from less wealthy zip codes, though this difference is decreasing in SAT scores. Including these controls has no impact on our point estimates for the black indicator or its interaction with SAT. We find similar results when we instead measure SES status through median education (column 3) or income mobility (column 4). Columns (5) through (8) provide analogous results to columns (1) through (4) using the B&B data. While we find some evidence that students from low SES zip codes are more likely to graduate in more difficult majors, our results on race are unaffected.

We produce results from several related exercises in Appendix C. In Appendix Table C.9 we find evidence that low SES students choose more difficult first-year majors in the state schools sample. However, our main race effects remain robust. We reproduce our results for the state schools sample with the SES controls using the Percentile Return measure in Appendix Table C.10 and the STEM courses measure in Appendix Table C.11. Our results are again robust. We repeat this exercise using for the B&B data in Appendix Table C.12 and find similar results.

In Table 7 we compare the race and SES effects on the earnings of B&B graduates. Using our sample of graduates with observable home zip code, we estimate a model similar to column (4) of Table 5. Our results again provide evidence for both predictions our model. Black graduates earn lower wages than white graduates in the same major, and have an observed return to major difficulty that is 32.8% lower. In column (2) we include the median income measure. Students from high SES zip codes earn higher wages than students from low SES zip codes, but including this measure has little effect on the other estimates. We find similar results when we instead use median education (column 3) or income mobility (column 4). Our results are further unchanged when we include SAT fixed effects along with the median income measure in column (5). We repeat this exercise for our percentile return and STEM courses difficulty measures in Appendix Table C.13. We find little qualitative difference across these difficulty measures.

In summary, the results of Section 4.3 strongly support labor market statistical discrimination as the mechanism for our empirical findings. The predicted racial effects hold across all specifications that include SES controls. In one dataset, we find that low SES students graduate in *less* difficult majors than high SES students with similar academic preparation. Finally, we find no evidence that that observed labor market return to major difficulty varies with SES status.

4.4 Heterogeneous Effects

We explore gender differences in major selection by estimating Equation 9 separately for female and male students in both the state schools and B&B data. The results are reported in Appendix Table C.14. We find a larger racial gap in major selection for female students and a smaller, though still statistically significant, racial gap for male students. Consistent with Proposition 4, we find that both male and female black students select more difficult majors than white students as we move up the academic preparation distribution, though the racial gap in major selection increases more rapidly in SAT score for male students than for female students.

We examine wage outcomes by gender and age group by estimating Equation (10) using the ACS sample. We show our results in Appendix Table C.15. We find some evidence that the black-white within major earnings gap is larger for men than women, but the differences in the observed return to major difficulty are similar by sex. We also find that the within major racial earnings gap is smallest among workers under the age of 30. This provides a potential explanation for why we generally find smaller effect sizes on the recent college graduates of the B&B. We find little evidence that racial differences in the observed return to major difficulty vary across age.

5 Empirical Strategies for Affirmative Action and Mismatch

Thus far we have derived a theoretical model of educational investment choice where individuals face uncertainty about their own aptitude and anticipate statistical discrimination in the labor market. We showed that these two forces have opposing effects. Student uncertainty reduces the reliability of their investment choices as an indicator of productivity, which reduces the signaling value of these credentials and leads to lower equilibrium investment. Statistical discrimination increases firm reliance on observable measures of productivity, which raises the signaling value of credentials and leads to higher equilibrium investment. When statistical discrimination against black workers is sufficiently strong, black students optimally mismatch by choosing more difficult investments than white students with similar academic preparation. This causes black workers to have a lower observed return to investment. Our empirical results were consistent with this in an environment, college major choice, that is likely to be less confounded by other factors that may lead to mismatch, such as affirmative action in college admissions.

In Appendix D we extend our model to an environment where black students face additional costs to acquiring human capital, reflecting potential discriminatory barriers. We then analyze how affirmative action policy will change equilibrium investment choices when black students anticipate statistical discrimination, before finally assessing the validity of two different common empirical strategies for estimating the consequences of affirmative action and mismatch. We summarize the key results here.

First, introducing costly barriers will unambiguously lower the investment choices for black students. If these barriers are sufficiently large, black students will in equilibrium undermatch, by choosing $M_b(\rho) < M^*(\rho)$. This provides room for policy to improve outcomes through an "affirmative action" subsidy to counteract that barrier. Such a subsidy could be designed to raise $M_b(\rho)$ to the human capital optimum or raise it to $M_w(\rho)$, which will be above the human capital optimum as in Proposition 1. If the subsidy is set exactly equal to the additional costs black students face, and assuming statistical discrimination is sufficiently strong, black students will overmatch relative to whites as seen in Section 2.

Whether an affirmative action policy is well-designed is an empirical question. To test this, we must first specify the policy objective. We differentiate between two definitions of mismatch. "Weak mismatch" is defined as any state of the world where $M_b(\rho) > M^*(\rho)$. If black students are weakly mismatched, then a marginal decrease in the amount of affirmative action will raise their human capital. "Strong mismatch" is in contrast defined as when abolishing affirmative action (setting the subsidy to 0) leads black students to acquire more human capital than under the current affirmative action regime.

Previous research that tested for mismatch has adopted two primary strategies. The first relies on a natural experiment that leads to as-good-as-random assignment of a set of students to highly selective and unselective universities (e.g Hoekstra, 2009; Zimmerman, 2019; Mountjoy and Hickman, 2021). This approach will identify the causal effect of university selectivity on the population studied. However, this does not necessarily identify mismatch if information is incomplete. This is because the causal estimate will include both the effect of university selectivity on human capital, but also the sheepskin effect on market beliefs. Even if the human capital effect is negative, we should expect a positive wage effect in equilibrium when mismatch is an optimal response to anticipated discrimination. In contrast, if information is complete this approach will identify whether or not there is weak mismatch. Thus this strategy would be ideal on a set of mid-career or older workers where signaling should be less important (Lange, 2007; Aryal et al., 2022).

The alternative approach uses state-wide racial admissions preferences bans for identification (e.g., Hinrichs, 2012; Arcidiacono et al., 2016; Bleemer, 2022; Antman et al., 2024). By estimating the causal effect of the affirmative action ban, this approach directly tests for strong mismatch. The approach can also confirm weak mismatch whenever an affirmative action ban causes a positive effect on black wages. However, it can never *reject* weak mismatch. Even if black students were overmatched under the affirmative action regime, they may still be worse off after the policy change if it results in undermatch that is more severe. Note that since the policy change affects all student choices, it also affects equilibrium market beliefs, and thus it does not matter for interpretation whether the sample is from a complete or incomplete information environment.

6 Conclusion

In this paper we study the impact of two sources of information frictions on college major choice and subsequently the labor market. When firms have imprecise information on workers, they rely more heavily on college major as a screening device, which increases the incentive for students to overinvest by choosing a more difficult major than the human capital maximizing choice. Conversely, when students choose their major with imprecise information on their aptitude, it reduces the reliability of college major as a screening device, reducing the incentives to overinvest. Following the statistical discrimination and student mismatch literatures, we postulate that these information frictions are more severe for black students and workers. Whichever of these forces is stronger will determine whether black students attempt more or less difficult majors than white students, and whether they see a higher or lower return to major difficulty in the labor market.

We test our theory using administrative data from 12 large public universities, the ACS, and the B&B, and find evidence consistent with the view that statistical discrimination has a larger impact on black student choices than incomplete information about their own aptitude. Specifically, we find that black students attempt more difficult majors than white students with similar academic preparation, that this difference is increasing in SAT scores, and that black students have a lower observed return to major difficulty in the labor market. Our model also provides a test for whether black students have less precise beliefs about their aptitude than white students when making their major choice, which we confirm by estimating the labor market return to college grades.

Our paper provides a novel contribution to the literature on academic mismatch and affirmative action. The empirical results indicate black students are "overmatched" in their major choices, but not due to information deficiencies or affirmative action. Instead, it is the rational response to anticipated statistical discrimination in the labor market. It also provides a valuable lesson on the interpretation of as-good-as-random assignment estimators when the measured outcome is determined in a market with incomplete information. In fact, we should expect a discontinuity in wage outcomes between individuals just below and just above a university admissions cutoff, independent of any human capital effect of that university itself, because there is a sharp change in employer beliefs at this cutoff. Depending on the context, it may be appropriate to focus on older samples of workers where signaling is less important, or to use alternative identification strategies relying on large shocks that change market beliefs.

References

Aalto, Aino-Maija, Dagmar Müller, and J. Lucas Tilley (2022) "From epidemic to pandemic: Effects of the COVID-19 outbreak ou high school program choices in Sweden," *Labour Economics*, Vol. 82.

Akhtari, Mitra, Natalie Bau, and Jean-William P. Laliberté (2024) "Affirmative action and pre-college human capital," *American Economic Journal: Applied Economics*, Vol. 16, pp. 1–32.

Altonji, Joseph G., Peter Arcidiacono, and Arnaud Maurel (2016) "The analysis of field choice in college and graduate school: Determinants and wage effects," in Eric A. Hanushek, Stephen Manchin, and Ludger Wößmann eds. *Handbook of the Economics of Education*, Vol. 5: North Holland, pp. 305–396.

Altonji, Joseph G., Erica Blom, and Costas Meghir (2012) "Heterogeneity in human capital investments: High school curriculum, college major, and careers," *Annual Review of Economics*, Vol. 4, pp. 185–223.

Antman, Francisca M., Brian Duncan, and Michael F. Lovenheim (2024) "The Long-Run Impacts of Banning Affirmative Action in US Higher Education," NBER Working Paper 32778.

Arcidiacano, Peter, Esteban Aucejo, Patrick Coate, and V. Joseph Hotz (2014) "Affirmative action and university fit: Evidence from Proposition 209," *IZA Journal of Labor Economics*, Vol. 3.

Arcidiacono, Peter, Esteban M. Acuejo, and Ken Spenner (2012) "What happens after enrollment? An analysis of the time path of racial differences in GPA and major choice," *IZA Journal of Labor Economics*, Vol. 1.

Arcidiacono, Peter, Esteban M. Aucejo, Hanmin Fang, and Kenneth I. Spenner (2011a) "Does affirmative action lead to mismatch? A new test and evidence?" *Quantitative Economics*, Vol. 2, pp. 303–333.

Arcidiacono, Peter, Esteban M. Aucejo, and V. Joseph Hotz (2016) "University differences in the graduation of minorities in STEM fields: Evidence from California," *American Economic Review*, Vol. 106, pp. 525–562.

Arcidiacono, Peter, Patrick Bayer, and Aurel Hizmo (2010) “Beyond signaling and human capital: Education and the revelation of ability,” *American Economic Journal: Applied Economics*, Vol. 2, pp. 76–104.

Arcidiacono, Peter, Shakeeb Khan, and Jacob L. Vigdor (2011b) “Representation versus assimilation: How do preferences in college admissions affect social interactions?” *Journal of Public Economics*, Vol. 95, pp. 1–15.

Aryal, Gaurab, Manudeep Bhuller, and Fabian Lange (2022) “Signaling and employer learning with instruments,” *American Economic Review*, Vol. 112, pp. 1669–1702.

Backes, Ben (2012) “Do affirmative action bans lower minority college enrollment and attainment? Evidence from statewide bans,” *Journal of Human Resources*, Vol. 47, pp. 435–455.

Bleemer, Zachary (2022) “Affirmative action, mimsatch, and economic mobility after California’s Proposition 209,” *Quarterly Journal of Economics*, Vol. 137, pp. 115–160.

Bleemer, Zachary and Aashish Mehta (2023) “College major restrictions and student stratification,” Unpublished.

Blom, Erica, Brian C. Cadena, and Benjamin K. Keys (2021) “Investment over the business cycle: Insights from college major choice,” *Journal of Labor Economics*, Vol. 39, pp. 1043–1082.

Bond, Timothy N. and Laura Salisbury (2018) “Local information, income segregation, and geographic mobility,” *B.E. Journal of Economic Analysis and Policy*, Vol. 18.

Carranza, Eliana, Robert Garlick, Kate Orkin, and Neil Rankin (2022) “Job search and hiring with limited information about workseekers’ skills,” *American Economic Review*, Vol. 112, pp. 3547–3583.

Chetty, Raj, John N Friedman, Nathaniel Hendren, Maggie R Jones, and Sonya R Porter (2018) “The opportunity atlas: Mapping the childhood roots of social mobility,”Technical report, National Bureau of Economic Research.

Coate, Stephen and Glenn C. Loury (1993) “Will affirmative-action policies eliminate negative stereotypes,” *American Economic Review*, Vol. 83, pp. 1220–1240.

Denning, Jeffrey T., Eric R. Eide, Kevin J. Mumford, Richard W. Patterson, and Merril Warnick (2022) “Why have college completion rates increaseed?” *American Economic Journal: Applied Economics*, Vol. 14, pp. 1–29.

Dillon, Eleanor Wiske and Jeffrey Andrew Smith (2020) "The consequences of academic match between students and colleges," *Journal of Human Resources*, Vol. 55, pp. 767–808.

Ersoy, Fulya Y. (2020) "The effects of the Great Recession on college majors," *Economics of Education Review*, Vol. 77.

Fryer, Roland G. Jr. and Glenn C. Loury (2005) "Affirmative action in winner-take-all markets," *Journal of Economic Inequality*, Vol. 3, pp. 263–280.

Ganguli, Ina, Jamal I. Haidar, Asim Ijaz Khwaja, Samuel W. Stemper, and Basit Zafar (2024) "Economic shocks and skill acquisition: Evidence from a a national online learning program at the onset of COVID-19," *Labour Economics*, Vol. 90.

Han, Luyi and John V. Winters (2020) "Industry fluctations and college major choices: Evidence from an energy boom and bust," *Economics of Education Review*, Vol. 77.

Hinrichs, Peter (2012) "The effects of affirmative action bans on college enrollment, educational attianment, and the demographic composition of universities," *Review of Economics and Statistics*, Vol. 94, pp. 712–722.

——— (2014) "Affirmative action bans and college graduation rates," *Economics of Education Review*, Vol. 42, pp. 43–52.

Hoekstra, Mark (2009) "The effect of attending the flagship state university on earnings: A discontinuity-based approach," *Review of Economics and Statistics*, Vol. 91, pp. 717–724.

Lang, Kevin (1986) "A language theory of discrimination," *Quarterly Journal of Economics*, Vol. 101, pp. 363–382.

Lang, Kevin and Jee-Yeon K. Lehmann (2012) "Racial discrimination in the labor market: Theory and empirics," *Journal of Economic Litarature*, Vol. 50, pp. 959–1006.

Lang, Kevin and Michael Manove (2011) "Education and labor market discrimination," *American Economic Review*, Vol. 101, pp. 1467–1496.

Lange, Fabian (2007) "The speed of employer learning," *Journal of Labor Economics*, Vol. 25, pp. 1–35.

Light, Audrey and Wayne Strayer (2000) "Determinants of college completion: School quality or student ability?" *Journal of Human Resources*, Vol. 35, pp. 299–332.

Manson, Steven, Jonathan Schroeder, David Van Riper, Katherine Knowles, Tracy Kugler, Finn Roberts, and Steven Ruggles (2023) "IPUMS National Historical Geographic Information System: Version 18.0 [dataset]," University of Minnesota.

Monaghan, David and Sou Hyun Jang (2017) "Major payoffs: Postcollege income, graduate school, and the choice of "risky" undergraduate majors," *Sociological Perspectives*, Vol. 60, pp. 722–746.

Mountjoy, Jack and Brent R. Hickman (2021) "The returns to college(s): Value-added and match effects in higher education," NBER Working Paper No. 29276.

Neal, Derek (2004) "The measured black-white wage gap among women is too small," *Journal of Political Economy*, Vol. 112, pp. S1–S28.

Novik, Vitaliy (2023) "The Role of Learning in Returns to College Major: Evidence from 2.8 million reviews of 150,000 Professors," Unpublished.

Patnaik, Arpita, Joanna Venator, Matthew Wiswall, and Basit Zafar (2022) "The role of heterogeneous risk preferences, discount rates, and earnings expectations in college major choice," *Journal of Econometrics*, Vol. 231, pp. 98–122.

Sallee, James M., Alexandra M. Resch, and Paul N. Courant (2008) "On the optimal allocation of students and resources in a system of higher education," *B.E. Journal of Economic Analysis & Policy*, Vol. 8.

Sovero, Veronica, Moshe Buchinsky, and Matthew D. Baird (2021) "Playing catch up: A term-level investigation of the racial gap in STEM retention," *Economics of Education Review*, Vol. 83.

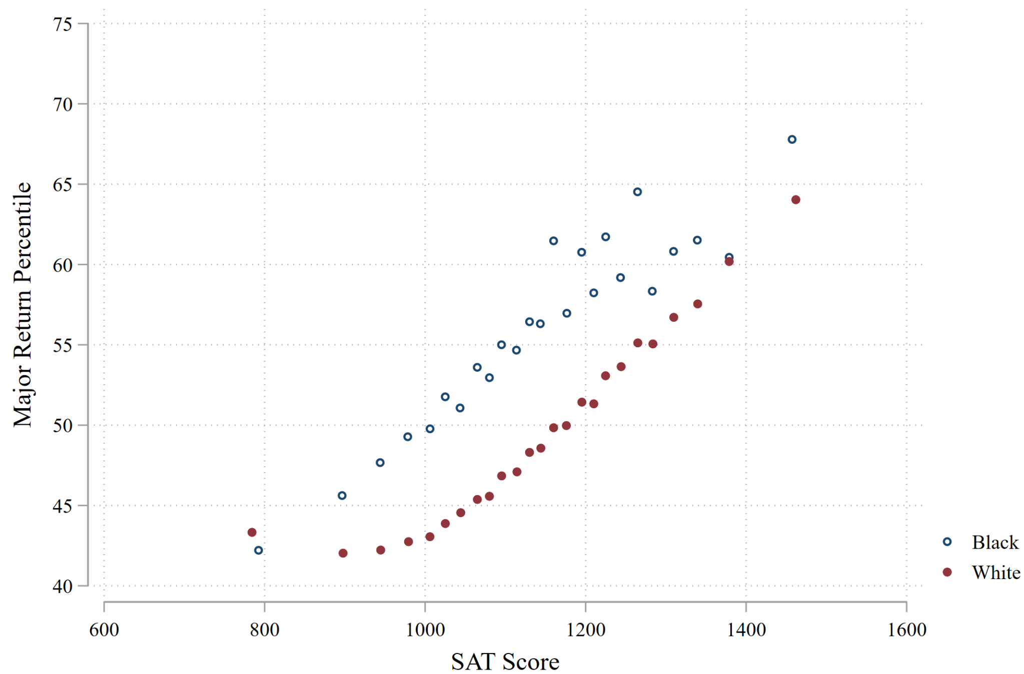
Weinstein, Russell (2022) "Local labor markets and human capital investments," *Journal of Human Resources*, Vol. 57, pp. 1498–1525.

Wiswall, Matthew and Basit Zafar (2015) "Determinants of college major choice: Identification using an information experiment," *The Review of Economic Studies*, Vol. 82, pp. 791–824.

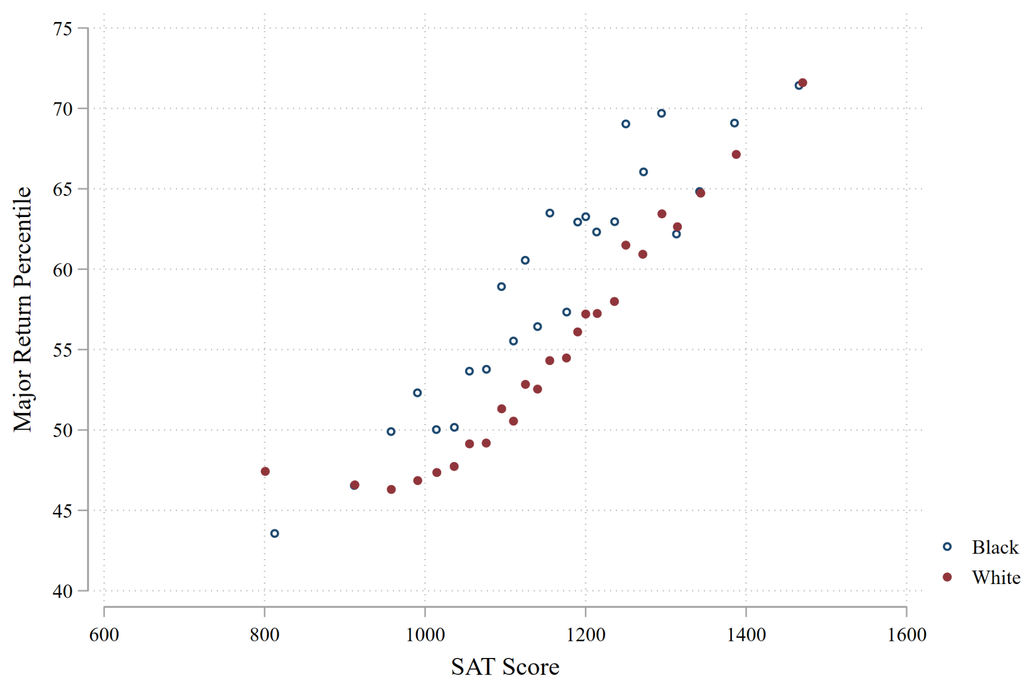
Zimmerman, Seth D. (2019) "Elite colleges and upward mobility to top jobs and top incomes," *American Economic Review*, Vol. 109, pp. 1–47.

Figure 1: SAT Scores and Major Percentile Return by Race: State Schools Sample

Panel A. First Year Major

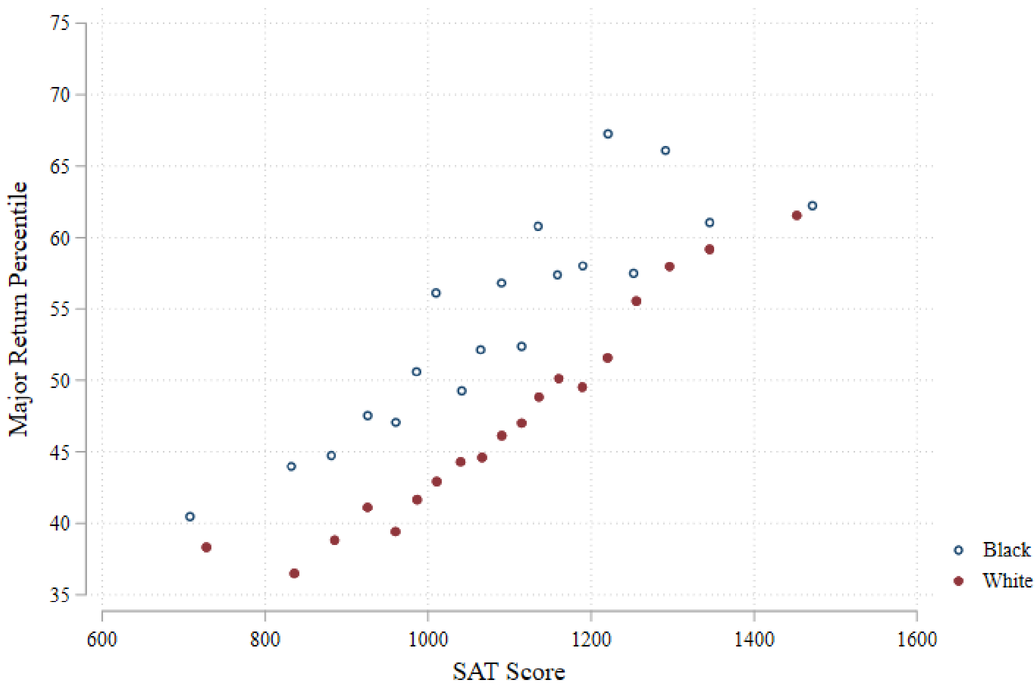


Panel B. Graduation Major



Source – Multiple-Institution Database for Investigating Engineering Longitudinal Development (MIDFIELD)
Notes – The state schools sample includes all black and white students with observed SAT scores at Clemson, Colorado, Colorado State, Florida, Florida State, Georgia Tech, North Carolina State, North Carolina – Charlotte, Oklahoma, Purdue, Utah State, and Virginia Tech. The sample includes students who entered college between 1987 and 2018, with incomplete time coverage for some institutions. Students over age 30 and those not identified as either Black or White are excluded from the sample.

Figure 2: SAT Scores and Major Percentile Return by Race: B&B Sample



Source – U.S. Department of Education, National Center for Education Statistics, 2008/18 Baccalaureate and Beyond Longitudinal Study (B&B:08/18)
Notes – The B&B sample is a nationally representative survey of 2007-2008 college graduates. Students over age 30 and those not identified as either Black or White are excluded from the sample.

Table 1: Descriptive Statistics, State Schools Sample

	White (1)	Black (2)	T-Test P-Value (3)
Female	0.475	0.562	0.000
Transfer Student	0.151	0.123	0.000
Year Entered College	2001.7	2000.2	0.000
High School GPA	3.032	2.854	0.000
SAT Score	1144.6	1033.4	0.000
First-Year College GPA	2.848	2.459	0.000
College GPA at Graduation	2.899	2.441	0.000
First Major Wage Return	-0.0029	0.0117	0.000
First Major Percentile Return	0.497	0.516	0.000
First Major STEM Courses	0.354	0.352	0.006
Graduation Major Wage Return	0.0364	0.0250	0.000
Graduation Major Percentile Return	0.551	0.535	0.000
Graduation Major STEM Courses	0.333	0.302	0.000
Graduated College	0.489	0.394	0.000
Chemistry Major	0.010	0.011	0.001
Biology Major	0.073	0.092	0.000
Social Science Major	0.042	0.053	0.000
Communications Major	0.040	0.042	0.010
Business/Econ Major	0.126	0.130	0.010
Liberal Arts Major	0.168	0.171	0.044
Engineering Major	0.178	0.163	0.000
History Major	0.012	0.007	0.000
English Major	0.019	0.015	0.000
Education Major	0.041	0.037	0.000
Agriculture Major	0.031	0.012	0.000
Observations	873,662	60,786	

Source – Multiple-Institution Database for Investigating Engineering Longitudinal Development (MIDFIELD)
Notes – The state schools sample includes all black and white students with observed SAT scores at Clemson, Colorado, Colorado State, Florida, Florida State, Georgia Tech, North Carolina State, North Carolina – Charlotte, Oklahoma, Purdue, Utah State, and Virginia Tech. The sample includes students who entered college between 1987 and 2018, with incomplete time coverage for some institutions. Students over age 30 and those not identified as either Black or White are excluded from the sample.

Table 2: Descriptive Statistics, American Community Survey Sample

	White (1)	Black (2)	T-Test P-Value (3)
Female	0.465	0.628	0.000
Age	43.43	42.99	0.000
Log Earnings	11.17	10.91	0.000
Major Wage Return	-0.0022	-0.0308	0.000
Major Percentile Return	0.504	0.470	0.000
Major STEM Courses	0.278	0.246	0.000
Observations	2,585,094	200,428	

Source – U.S. Census Bureau, 2011-2021 American Community Survey, Public Use Microdata
Notes – The ACS sample includes working age (16 to 64) native non-Hispanic black and white college graduates who were employed full time in the previous year. Survey years 2011 through 2021 are included with the year 2020 excluded. Log earnings is the log of the sum of wage income and salary income in 2020 dollars.

Table 3: Descriptive Statistics, Baccalaureate and Beyond Sample

	White (1)	Black (2)	T-Test P-Value (3)
Female	0.577	0.673	0.000
SAT Score	1099.3	949.0	0.000
GPA at Graduation	3.347	3.089	0.000
Age	18.19	18.19	0.914
Major Wage Return	-0.020	-0.012	0.196
Major Return Percentile	0.475	0.490	0.061
Major STEM Percentage	0.338	0.325	0.057
Log Salary 2009	10.16	10.10	0.023
Log Salary 2012	10.58	10.50	0.000
Log Salary 2018	11.08	10.94	0.000
Observations	10,420	1,210	

Source – U.S. Department of Education, National Center for Education Statistics, 2008/18 Baccalaureate and Beyond Longitudinal Study (B&B:08/18)
Notes – The B&B sample includes 2007-2008 college graduates with follow-up surveys at 1, 4, and 10 years after graduation. Students over age 30 and those not identified as either Black or White are excluded from the sample.

Table 4: Major Selection by Race and SAT Score

	State Schools				B&B	
	1st-Yr. Major		Grad. Major		Grad. Major	
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Major Wage Return					
Black	0.031***	0.037***	0.029***	0.037***	0.058***	0.070***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.007)	(0.009)
Black × SAT		0.005***		0.007***		0.007**
		(0.001)		(0.001)		(0.003)
	Panel B: Major Percentile Return					
Black	0.042***	0.048***	0.037***	0.046***	0.082***	0.098***
	(0.002)	(0.002)	(0.004)	(0.004)	(0.009)	(0.012)
Black × SAT		0.006***		0.010***		0.010**
		(0.001)		(0.001)		(0.004)
	Panel C: Major STEM Courses					
Black	0.030***	0.036***	0.021***	0.032***	0.044***	0.061***
	(0.002)	(0.002)	(0.004)	(0.004)	(0.008)	(0.012)
Black × SAT		0.006***		0.011***		0.010**
		(0.001)		(0.001)		(0.004)
Student Characteristics	X	X	X	X	X	X
Student SAT FE	X	X	X	X	X	X
Institution × Start Year FE	X	X	X	X		
Institution SAT Percentiles					X	X
Observations	934,456	934,456	450,994	450,994	11,550	11,550

Source – Multiple-Institution Database for Investigating Engineering Longitudinal Development (MIDFIELD) and U.S. Department of Education, National Center for Education Statistics, 2008/18 Baccalaureate and Beyond Longitudinal Study (B&B:08/18)
Notes – The panel A outcome variable is the average wage return from the ACS for white graduates by major. The panel B outcome variable is the percentile ranking of the average wage return from the ACS for white graduates by major. The panel C outcome variable is the fraction of course credits in STEM courses by major in the state school sample. Student characteristics include student age at matriculation, a female indicator, SAT fixed effects, and in the state school sample a transfer student indicator is also included. In the state schools sample, institution by start year fixed effects are included. In the B&B sample, the 25th and 75th percentile math and verbal SAT scores for the institution (4 variables) are included to control for institution quality. Students not identified as either black or white are excluded from the analysis. Standard errors clustered by institution (both samples) and year of college entry (state schools sample) are reported in parenthesis: $*p < .1, **p < .05, ***p < .01$

Table 5: Adult Log Earnings by Graduation Major Selection and Race

	ACS			B&B		
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Major = Wage Return					
Black	-0.220***	-0.229***		-0.068***	-0.068***	-0.039**
	(0.016)	(0.016)		(0.017)	(0.015)	(0.019)
Major	0.866***	0.832***	0.833***	0.564***	0.568***	0.559***
	(0.024)	(0.033)	(0.033)	(0.131)	(0.132)	(0.131)
Major × Black	-0.325***	-0.321***	-0.325***	-0.131*	-0.137*	-0.140*
	(0.052)	(0.051)	(0.052)	(0.073)	(0.075)	(0.071)
College GPA						0.061***
						(0.018)
College GPA × Black						0.113***
						(0.035)
	Panel B. Major = Percentile Return					
Black	-0.099***	-0.109***		-0.066***	-0.065***	-0.037*
	(0.013)	(0.013)		(0.017)	(0.015)	(0.019)
Major	0.649***	0.625***	0.625***	0.402***	0.403***	0.398***
	(0.023)	(0.029)	(0.029)	(0.104)	(0.105)	(0.104)
Major × Black	-0.246***	-0.242***	-0.245***	-0.094*	-0.099*	-0.101*
	(0.038)	(0.037)	(0.038)	(0.054)	(0.056)	(0.053)
College GPA						0.062***
						(0.018)
College GPA × Black						0.113***
						(0.035)
	Panel C: Major = STEM Courses					
Black	-0.182***	-0.193***		-0.056***	-0.053***	-0.025
	(0.040)	(0.041)		(0.018)	(0.017)	(0.020)
Major	0.460***	0.452***	0.452***	0.352***	0.350***	0.350***
	(0.080)	(0.084)	(0.084)	(0.115)	(0.116)	(0.114)
Major × Black	-0.121**	-0.118**	-0.118**	-0.074	-0.073	-0.079
	(0.057)	(0.057)	(0.058)	(0.079)	(0.081)	(0.076)
College GPA						0.067***
						(0.018)
College GPA × Black						0.115***
						(0.036)
State FE		X		X	X	X
State × Race FE			X			
Student SAT FE					X	
Institution SAT Percentiles				X	X	X
Observations	2,650,399	2,650,399	2,650,399	26,400	26,400	26,400

Source – U.S. Census Bureau, 2011-2021 American Community Survey, Public Use Microdata and U.S. Department of Education, National Center for Education Statistics, 2008/18 Baccalaureate and Beyond Longitudinal Study (B&B:08/18)
Notes – The outcome variable is log earnings (wage and salary income). Major is defined as the graduation major average wage return from the ACS for white graduates in panel A, as the percentile wage return from the ACS for white graduates in panel B, and as the fraction of course credits in STEM courses by major in panel C. Worker characteristics (gender and age) and year fixed effects are included in all specifications. In the B&B sample, log earnings are measured at 1, 4 and 10 years after graduation and the 25th and 75th percentile math and verbal SAT scores for the institution (4 variables) are included to control for institution quality. Standard errors clustered by the graduation major are reported in parenthesis: $*p < .1, **p < .05, ***p < .01$

Table 6: Graduation Major Selection by Race, SAT Score, and Neighborhood Characteristics, State Schools Sample and Baccalaureate and Beyond Sample

	State Schools				B&B			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Black	0.039***	0.040***	0.040***	0.040***	0.073***	0.072***	0.073***	0.073***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.010)	(0.010)	(0.010)	(0.010)
Black × SAT	0.009***	0.009***	0.009***	0.008***	0.007*	0.006*	0.006*	0.007*
	(0.002)	(0.002)	(0.002)	(0.002)	(0.004)	(0.004)	(0.004)	(0.004)
Median Income (10,000s)		0.001*				-0.002*		
		(0.000)				(0.001)		
Median Income × SAT		-0.001*				-0.000		
		(0.000)				(0.001)		
Median Education			0.001				-0.002	
			(0.001)				(0.002)	
Median Education × SAT			-0.001*				-0.000	
			(0.000)				(0.001)	
Income Mobility				0.077***				-0.058
				(0.019)				(0.053)
Income Mobility × SAT				-0.087***				-0.001
				(0.012)				(0.025)
Student Characteristics	X	X	X	X	X	X	X	X
Student SAT FE	X	X	X	X	X	X	X	X
Institution × Start Year FE	X	X	X	X				
Institution SAT Percentiles					X	X	X	X
Observations	312,538	312,538	312,538	312,538	8,370	8,370	8,370	8,370

Source - Multiple-Institution Database for Investigating Engineering Longitudinal Development (MIDFIELD) and U.S. Department of Education, National Center for Education Statistics, 2008/18 Baccalaureate and Beyond Longitudinal Study (B&B:08/18)

Notes - The outcome is graduation major average wage return from the ACS for white graduates. Only students who identified as either black or white and have an observable home ZCTA and county are included in the analysis. Median household income and median education are measured at the ZCTA level in the census year (1990, 2000, 2010) prior to the student entering college. Income mobility is measured at the county level. Student characteristics include student age at matriculation, a female indicator, SAT fixed effects, and in the state school sample a transfer student indicator is also included. In the state schools sample, institution by start year fixed effects are included. In the B&B sample, the 25th and 75th percentile math and verbal SAT scores for the institution (4 variables) are included to control for institution quality. Standard errors clustered by institution (both samples) and year of college entry (state schools sample) are reported in parenthesis: $*p < .1, **p < .05, ***p < .01$

Table 7: Log Earnings by Graduation Major Selection, Race, and Neighborhood Characteristics

	B&B				
	(1)	(2)	(3)	(4)	(5)
Black	-0.059***	-0.044**	-0.056***	-0.058***	-0.059***
	(0.021)	(0.021)	(0.021)	(0.020)	(0.021)
Major	0.589***	0.498***	0.528***	0.553***	0.532***
	(0.127)	(0.152)	(0.143)	(0.193)	(0.156)
Black × Major	-0.328***	-0.338***	-0.323***	-0.321***	-0.263***
	(0.092)	(0.090)	(0.093)	(0.087)	(0.094)
Median Income (10,000s)		0.028***			0.027***
		(0.003)			(0.004)
Median Income × Major		0.018			0.010
		(0.013)			(0.013)
Median Education			0.018***		
			(0.005)		
Median Education × Major			0.012		
			(0.017)		
Income Mobility				0.902***	
				(0.149)	
Income Mobility × Major				0.267	
				(0.777)	
Student Characteristics	X	X	X	X	X
Student SAT FE					X
Year FE	X	X	X	X	X
Institution SAT Percentiles	X	X	X	X	X
Observations	22,670	22,670	22,670	22,670	20,500

Source – U.S. Department of Education, National Center for Education Statistics, 2008/18 Baccalaureate and Beyond Longitudinal Study (B&B:08/18)
Notes – The outcome variable is log earnings measured at either 1, 4 or 10 years after graduation. Only students who identified as either black or white and have an observable home ZCTA and county are included in the analysis. Median household income and median education are measured at the ZCTA level and income mobility is measured at the county level. Student characteristics include student age at matriculation and a female indicator. The 25th and 75th percentile math and verbal SAT scores for the institution (4 variables) are included to control for institution quality. Standard errors clustered by institution (both samples) and year of college entry (state schools sample) are reported in parenthesis: $*p < .1, **p < .05, ***p < .01$

Appendix

A Proofs of Main Results

A.1 Proposition 1

Proof. For the first part of the proposition, suppose not and that $M_r(0) > 0$ for some r. Since beliefs are correct in equilibrium, it is then the case that $P_r(M_r(0)) = 0$. As employers believe off-equilibrium major choices are associated with the lowest type, it must be that $M_r(0)$ provides higher expected wages for the 0 type than $m = 0$. From equation (8), this implies

$$-M_r(0)^2 \geq 0, \tag{11}$$

which is a contradiction. Lang and Manove (2011) prove a similar proposition in a more general environment.

For the second part of the proposition, suppose that for some ρ, $M_r(\rho) < M^*(\rho)$. Note that in equilibrium, $E[\pi_r(m, s, g)|M_r(\rho)] = \rho$. By definition of $M^*(\rho)$,

$$M^*(\rho)\rho - M^*(\rho)^2 > M_r(\rho)\rho - M_r(\rho)^2. \tag{12}$$

Further, as $P_r(m)$ is monotonically increasing in m, it must be the case that $P_r(M^*(\rho)) > P_r(M_r(\rho))$. From equation (8), the expected wages from deviating to $M^*(\rho)$ are

$$M^*(\rho)\tau_r^{-1}\left[\varsigma_r^{-2}P_rM^*(\rho) + \left(\sigma_r^{-2} + \varrho^{-2}\right)\rho\right] - M^*(\rho)^2 > M^*(\rho)\rho - M^*(\rho)^2 \tag{13}$$

This proves the proposition. □

A.2 Proposition 2

Proof. By taking the derivative of (8) with respect to m and recognizing that in equilibrium $P_r(m) = \rho$, we arrive at a first order condition of

$$m\tau_r^{-1}\varsigma_r^{-2}\frac{\partial P_r(m)}{\partial m} + \rho - 2m = 0.$$

Note that since $P_r(m) = M_r^{-1}(\rho)$, $\frac{\partial P_r(m)}{\partial m} = \frac{\partial M_r^{-1}(\rho)}{\partial \rho}$. Rearranging terms then proves the proposition. □

A.3 Proposition 3

Proof. First note that if $\sigma_b^2 = \sigma_w^2$, then $\frac{\varsigma_w^{-2}}{\tau_w} > \frac{\varsigma_b^{-2}}{\tau_b}$. Next note that per Proposition 1 we know that $M_b(0) = M_w(0)$. Now suppose that there was some $\rho > 0$ such that $M_b(\rho) > M_w(\rho)$. As M_r is continuous and monotonically increasing in a well-behaved equilibrium, it must then be the case that there exists some $\rho' < \rho$ such that $M_b(\rho') = M_w(\rho')$ and $\frac{\partial M_b(\rho')}{\partial \rho} > \frac{\partial M_w(\rho')}{\partial \rho}$. But, as $\frac{\varsigma_w^{-2}}{\tau_w} > \frac{\varsigma_b^{-2}}{\tau_b}$, whenever $M_b(\rho) = M_w(\rho)$, $\frac{\partial M_b(\rho')}{\partial \rho} < \frac{\partial M_w(\rho')}{\partial \rho}$, which is a contradiction. □

A.4 Proposition 4

Proof. Note that as σ_b^2 increases relative to σ_w^2, $\frac{\varsigma_w^{-2}}{\tau_w} - \frac{\varsigma_b^{-2}}{\tau_b}$ decreases, and at some point $\frac{\varsigma_w^{-2}}{\tau_w} < \frac{\varsigma_b^{-2}}{\tau_b}$. When $\frac{\varsigma_w^{-2}}{\tau_w} < \frac{\varsigma_b^{-2}}{\tau_b}$, a *mutatis mutandi* proof from Proposition 3 proves the proposition. □

A.5 Proposition 5

Proof. For the first part of the proposition, from Proposition 3 we know that when $\sigma_b^2 = \sigma_w^2$ and $\varsigma_b > \varsigma_w$, black workers will have higher on average ρ within m, and thus $P_b(m) > P_w(m)$ for all $m > 0$. Noting that in expectation $\rho = P_r(m)$, we can see from equation (8) that

$$E_r(w|m) = mP_r(m) - m^2, \tag{14}$$

which is strictly increasing in $P_r(m)$.

For the second part of this proposition, we know from Proposition 4 that with sufficiently high σ_b^2, black workers will have lower on average ρ within m. The same arguments from above then apply. □

A.6 Proposition 6

Proof. A similar proof is provided in Lang and Manove (2011). First note from Proposition 1, that $M_b(0) = M_w(0) = 0$. Now consider the equilibrium observed return to human capital from major $m > 0$, given by

$$\frac{mP_r(m) - m^2}{m}. \tag{15}$$

The numerator is the difference in productivity between the equilibrium worker who chose major m and the equilibrium worker who chose major 0, while the denominator is simply the difference between m and 0. It is straightforward to see that this expressions is increasing in $P_r(m)$.

From Proposition 3 we know that when $\sigma_b^2 = \sigma_w^2$ and $\varsigma_b > \varsigma_w$, $P_b(m) > P_w(m)$ for all $m > 0$, while from Proposition 4, we know that with sufficiently high σ_b^2, $P_b(m) < P_w(m)$ for all $m > 0$. This proves the proposition. $\square$

A.7 Proposition 7

Proof. First we can find the wage offer for any worker given full information on m, s, and g by taking the expectation of (8),

$$w_r(\pi_r) = mE[p|m, s, g] - m^2. \tag{16}$$

However the researcher does not observe s, she instead only observes m and g. Thus the regression will be an estimate of

$$\begin{aligned} E[w_r(\pi_r)|m, g] &= mE_r[[p|m, s, g]|m, g] - m^2 \\ &= mE[p|m, g] - m^2, \end{aligned} \tag{17}$$

where the latter expression follows from the law of iterated expectations. Since both m and g are normally distributed, we can apply Bayes' rule to find $E[p|m, g]$,

$$E_r[p|m, g] = \frac{\varrho^2}{\varrho^2 + \varsigma_r^2} P_r(m) + \frac{\varsigma_r^2}{\varrho^2 + \varsigma_r^2} g. \tag{18}$$

The derivative of (18) with respect to g is clearly increasing in ς_r^2. $\square$

B Details on Data Sources and Construction

B.1 State Schools Sample

The state schools sample is constructed from administrative student transcript records from 12 large public universities: Clemson, Colorado, Colorado State, Florida, Florida State, Georgia Tech, North Carolina State, North Carolina – Charlotte, Oklahoma, Purdue, Utah State, and Virginia Tech.[31] While these universities are not nationally representative, Denning et al. (2022) show that these students are quite similar to those from the nationally representative NELS:88 and ELS:2002 sample of top-50 public universities in race, gender, and the distribution of SAT scores. The data were obtained from school registrars through the MIDFIELD partnership. Institutions that participate in the MIDFIELD partnership share de-identified longitudinal student records for all degree-seeking undergraduate students. These records include demographic characteristics and admissions data as well as course grades, major, and degree earned. They cover the years 1987 through 2018, though not all universities are included in all years. The records contain no information on post-graduation outcomes.

B.2 The Baccalaureate and Beyond

The B&B is a nationally representative longitudinal study of 2007-2008 college graduates collected by the National Center for Education Statistics (NCES). It combines demographic characteristics, college admissions measures, and college academic records with follow-up surveys focused on employment, post-baccalaureate education, and other outcomes. Follow-up surveys were conducted one, four, and ten years after graduation (in 2009, 2012, and 2018). We restrict the sample to students who identify as either black or white and exclude those who are age 30 or older when they graduate from college.

[31]The MIDFIELD partnership does not allow us to report any results separately by institution that would enable readers to identify the institution.

Online Appendix

C Supplementary Tables

Table C.1: Wage Return, Wage Percentile, and STEM Courses by Major

Major Code and Name		Wage Return	Wage Percentile	STEM
1100	General Agriculture	-0.192	20	0.261
1101	Agriculture Production and Management	-0.072	42	0.260
1102	Agricultural Economics	0.054	59	0.233
1103	Animal Sciences	-0.148	27	0.322
1104	Food Science	0.119	66	0.378
1105	Plant Science and Agronomy	-0.190	21	0.369
1106	Soil Science	-0.162	25	0.373
1199	Miscellaneous Agriculture	-0.218	18	0.184
1301	Environmental Science	-0.098	37	0.410
1302	Forestry	-0.147	28	0.322
1303	Natural Resources Management	-0.180	24	0.355
1401	Architecture	0.032	57	0.147
1501	Area, Ethnic, and Civilization Studies	0.000	55	0.132
1901	Communications	-0.043	47	0.122
1902	Journalism	-0.065	44	0.108
1903	Mass Media	-0.130	32	0.141
1904	Advertising and Public Relations	-0.003	54	0.126
2001	Communication Technologies	-0.151	27	0.297
2100	Computer and Information Systems	0.068	64	0.647
2101	Computer Programming and Data Processing	0.029	57	0.201
2102	Computer Science	0.260	84	0.714
2105	Information Sciences	0.173	73	0.294
2106	Computer Information Management and Security	0.039	58	-
2107	Computer Networking and Telecommunications	-0.032	49	-
2201	Cosmetology Services and Culinary Arts	-0.324	6	-
2300	General Education	-0.308	7	0.205
2301	Educational Administration and Supervision	-0.130	33	-
2303	School Student Counseling	-0.284	9	-
2304	Elementary Education	-0.373	3	0.164
2305	Mathematics Teacher Education	-0.190	21	0.551
2306	Physical and Health Education Teaching	-0.233	15	0.157
2307	Early Childhood Education	-0.418	1	0.145
2308	Science and Computer Teacher Education	-0.237	14	0.530
2309	Secondary Teacher Education	-0.254	11	0.219
2310	Special Needs Education	-0.291	8	0.113
2311	Social Science or History Teacher Education	-0.221	18	0.155
2312	Teacher Education: Multiple Levels	-0.358	4	0.196
2313	Language and Drama Education	-0.271	10	0.104
2314	Art and Music Education	-0.312	6	0.065
2399	Miscellaneous Education	-0.232	16	0.183
2400	General Engineering	0.233	82	0.713
2401	Aerospace Engineering	0.393	98	0.738
2402	Biological Engineering	0.211	78	0.676
2403	Architectural Engineering	0.242	83	0.763
2404	Biomedical Engineering	0.375	96	0.771
2405	Chemical Engineering	0.408	99	0.797
2406	Civil Engineering	0.259	83	0.794
2407	Computer Engineering	0.385	96	0.508
2408	Electrical Engineering	0.355	94	0.530
2409	Engineering Mechanics, Physics, and Science	0.301	90	0.821
2410	Environmental Engineering	0.185	75	0.650
2411	Geological and Geophysical Engineering	0.271	85	-
2412	Industrial and Manufacturing Engineering	0.315	91	0.742
2413	Materials Engineering and Materials Science	0.271	85	0.513
2414	Mechanical Engineering	0.338	92	0.786
2415	Metallurgical Engineering	0.346	93	0.309
2416	Mining and Mineral Engineering	0.339	92	0.837
2417	Naval Architecture and Marine Engineering	0.350	94	-
2418	Nuclear Engineering	0.361	95	0.693
2419	Petroleum Engineering	0.696	100	0.848
2499	Miscellaneous Engineering	0.221	79	0.712
2500	Engineering Technologies	-0.013	51	0.233
2501	Engineering and Industrial Management	0.206	77	0.420
2502	Electrical Engineering Technology	0.054	60	0.198
2503	Industrial Production Technologies	0.063	62	0.293
2504	Mechanical Engineering Related Technologies	0.076	65	0.448
2599	Miscellaneous Engineering Technologies	0.065	64	0.195
2601	Linguistics and Comparative Language and Literature	-0.064	44	0.139
2602	French, German, and Other Common Languages	-0.078	42	0.169
2603	Other Foreign Languages	-0.013	51	0.199
2901	Family and Consumer Sciences	-0.293	7	0.192
3201	Court Reporting	-0.240	14	-
3202	Pre-Law and Legal Studies	-0.097	39	0.123
3301	English Language and Literature	-0.096	40	0.168
3302	Composition and Speech	-0.226	17	0.150
3401	Liberal Arts	-0.161	26	0.166
3402	Humanities	-0.183	23	0.119
3501	Library Science	-0.385	2	0.182
3600	Biology	0.153	72	0.675
3601	Biochemical Sciences	0.261	84	0.669
3602	Botany	-0.120	36	0.612
3603	Molecular Biology	0.230	81	0.673
3604	Ecology	-0.203	20	0.615
3605	Genetics	0.135	68	0.661
3606	Microbiology	0.182	74	0.630
3607	Pharmacology	0.275	87	-
3608	Physiology	0.123	68	0.574
3609	Zoology	0.147	70	0.604
3611	Neuroscience	0.197	77	0.602
3699	Miscellaneous Biology	-0.081	41	0.488
3700	Mathematics	0.143	69	0.590
3701	Applied Mathematics	0.291	88	0.756
3702	Statistics and Decision Science	0.233	81	0.677
3801	Military Technologies	0.180	74	-
4000	Interdisciplinary Studies (General)	-0.244	13	0.212
4001	Intercultural and International Studies	-0.005	53	0.182
4002	Nutrition Sciences	-0.086	40	0.438
4005	Mathematics and Computer Science	0.276	87	0.649
4006	Cognitive Science and Biopsychology	0.273	86	0.182
4007	Interdisciplinary Social Sciences	-0.147	29	0.331
4101	Physical Fitness, Parks, Recreation, and Leisure	-0.139	32	0.188
4801	Philosophy and Religious Studies	-0.034	48	0.156
4901	Theology and Religious Vocations	-0.384	2	-
5000	Physical Sciences	0.046	58	0.750
5001	Astronomy and Astrophysics	0.154	72	0.693
5002	Atmospheric Sciences and Meteorology	0.064	63	0.726
5003	Chemistry	0.229	80	0.675
5004	Geology and Earth Science	-0.011	52	0.697
5005	Geosciences	0.096	66	0.771
5006	Oceanography	-0.060	45	0.746
5007	Physics	0.223	80	0.704
5008	Materials Science	0.332	91	-
5098	Multi-disciplinary or General Science	-0.016	50	0.370
5102	Nuclear, Industrial, and Biological Technologies	-0.033	48	0.363
5200	Psychology	-0.142	30	0.214
5201	Educational Psychology	-0.244	12	-
5202	Clinical Psychology	-0.180	24	-
5203	Counseling Psychology	-0.345	5	-
5205	Industrial and Organizational Psychology	0.047	59	0.403
5206	Social Psychology	-0.231	17	0.116
5299	Miscellaneous Psychology	-0.123	35	0.252
5301	Criminal Justice and Fire Protection	-0.127	34	0.099
5401	Public Administration	0.006	56	0.130
5402	Public Policy	0.299	90	0.299
5403	Human Services and Community Organization	-0.414	1	-
5404	Social Work	-0.348	5	0.113
5500	General Social Sciences	-0.184	22	0.199
5501	Economics	0.369	95	0.488
5502	Anthropology and Archeology	-0.145	29	0.157
5503	Criminology	-0.114	36	0.183
5504	Geography	-0.096	39	0.310
5505	International Relations	0.223	79	0.211
5506	Political Science and Government	0.189	76	0.163
5507	Sociology	-0.141	31	0.145
5599	Miscellaneous Social Sciences	0.064	62	-
5601	Construction Services	0.152	71	0.202
5701	Electrical and Mechanic Repairs and Technologies	-0.147	28	0.153
5901	Transportation Sciences and Technologies	0.122	67	0.165
6000	Fine Arts	-0.271	10	0.095
6001	Drama and Theater Arts	-0.269	11	0.088
6002	Music	-0.231	16	0.086
6003	Visual and Performing Arts	-0.284	9	0.092
6004	Commercial Art and Graphic Design	-0.184	22	0.129
6005	Film, Video and Photographic Arts	-0.165	25	0.105
6006	Art History and Criticism	-0.127	33	0.091
6007	Studio Arts	-0.365	3	0.107
6099	Miscellaneous Fine Arts	-0.216	19	0.186
6100	General Medical and Health Services	-0.047	47	0.264
6102	Communication Disorders Sciences and Services	-0.097	38	0.092
6103	Health and Medical Administrative Services	-0.126	35	0.000
6104	Medical Assisting Services	-0.057	46	-
6105	Medical Technologies Technicians	-0.066	43	0.218
6106	Health and Medical Preparatory Programs	0.389	97	0.366
6107	Nursing	-0.001	54	0.124
6108	Pharmacy Sciences, and Administration	0.400	98	0.112
6109	Treatment Therapy Professions	-0.059	46	0.119
6110	Community and Public Health	-0.098	37	0.281
6199	Miscellaneous Health Medical Professions	-0.243	13	0.296
6200	General Business	0.059	61	0.193
6201	Accounting	0.148	70	0.174
6202	Actuarial Science	0.465	99	0.616
6203	Business Management and Administration	0.005	55	0.208
6204	Operations, Logistics and E-Commerce	0.161	73	0.355
6205	Business Economics	0.283	88	0.449
6206	Marketing and Marketing Research	0.056	61	0.191
6207	Finance	0.294	89	0.219
6209	Human Resources and Personnel Management	-0.068	43	0.171
6210	International Business	0.140	69	0.190
6211	Hospitality Management	-0.141	31	0.158
6212	Management Information Systems and Statistics	0.193	76	0.273
6299	Misc Business and Medical Administration	0.071	65	0.199
6402	History	-0.006	53	0.144
6403	United States History	-0.030	50	-

Notes – The wage return and the percentile wage return are calculated for white, prime age (25-54), native-born, full-time, year-round, employed workers with at least a bachelor's degree. Survey years 2011 through 2021 are included with the year 2020 excluded. The major STEM content is calculated from the state schools sample and is the fraction of course credits in STEM courses by major.

Table C.2: Major Selection by Race and SAT Score: Alternative Measures of Major Return, State Schools Sample

	All		White male	
	Wage	Pctl	Wage	Pctl
	(1)	(2)	(3)	(4)
Black	0.034***	0.043***	0.033***	0.048***
	(0.003)	(0.004)	(0.003)	(0.004)
Black × SAT	0.008***	0.011***	0.005***	0.007***
	(0.001)	(0.002)	(0.001)	(0.001)
Student Characteristics	X	X	X	X
Student SAT FE	X	X	X	X
Institution x Start Year FE	X	X	X	X
Observations	450,994	450,994	450,994	450,994

Source – Multiple-Institution Database for Investigating Engineering Longitudinal Development (MIDFIELD)
Notes – This table reports estimates similar to those in Table 4, but uses alternative measures of the return to college major for the dependent variable. In Columns (1) and (2) the major wage return / percentile return are computed using all prime age workers. In Columns (3) and (4) the major wage return / percentile return are computed using white male prime age workers. Student characteristics include student age at matriculation, a female indicator, SAT fixed effects, and a transfer student indicator. Institution by start year fixed effects are also included. Students not identified as either black or white are excluded from the analysis. Standard errors clustered by institution and year of college entry are reported in parenthesis: $*p < .1, **p < .05, ***p < .01$

Table C.3: Major Selection by Race and SAT Score: Alternative Measures of Major Return, Baccalaureate and Beyond Sample

	All		White male	
	Wage	Pctl	Wage	Pctl
	(1)	(2)	(3)	(4)
Black	0.067***	0.095***	0.071***	0.104***
	(0.009)	(0.012)	(0.008)	(0.012)
Black × SAT	0.007**	0.010**	0.005*	0.007
	(0.003)	(0.004)	(0.003)	(0.005)
Student Characteristics	X	X	X	X
Student SAT FE	X	X	X	X
Institution SAT Percentiles	X	X	X	X
Observations	11,620	11,620	11,620	11,620

Source – U.S. Department of Education, National Center for Education Statistics, 2008/18 Baccalaureate and Beyond Longitudinal Study (B&B:08/18)
Notes – This table reports estimates similar to those in Table 4, but uses alternative measures of the return to college major for the dependent variable. In Columns (1) and (2) the major wage return / percentile return are computed using all prime age workers. In Columns (3) and (4) the major wage return / percentile return are computed using white male prime age workers. Student characteristics include student age at matriculation, a female indicator, and SAT fixed effects. The 25th and 75th percentile math and verbal SAT scores for the institution (4 variables) are included to control for institution quality. Students not identified as either black or white are excluded from the analysis. Standard errors clustered by institution are reported in parenthesis: $*p < .1, **p < .05, ***p < .01$

Table C.4: Adult Log Earnings by Graduation Major and Race: Alternative Measures of Major Return, American Community Survey Sample

	All		White male	
	Wage (1)	Pctl (2)	Wage (3)	Pctl (4)
Major	0.861***	0.630***	0.895***	0.614***
	(0.031)	(0.026)	(0.044)	(0.034)
Major × Black	-0.329***	-0.252***	-0.357***	-0.249***
	(0.050)	(0.035)	(0.049)	(0.034)
Worker Characteristics	X	X	X	X
Year FE	X	X	X	X
State × Race FE	X	X	X	X
Observations	2,701,293	2,701,293	2,701,293	2,701,293

Source – U.S. Census Bureau, 2011-2021 American Community Survey, Public Use Microdata
Notes – Robust standard errors clustered at the major level in parenthesis. Worker characteristics include age fixed effects and a gender indicator. Column (1) uses the Wage Return difficulty measure computed using all prime age workers. Column (2) uses the Percentile Return difficulty measure computed using all prime age workers. Column (3) uses the Wage Return difficulty measure computed using white male prime age workers. Column (4) uses the Percentile Return difficulty measure computed using white male prime age workers. Standard errors clustered by graduation major are reported in parenthesis: $*p < .1, **p < .05, ***p < .01$

Table C.5: Adult Log Earnings by Graduation Major and Race: Alternative Measures of Major Return, Baccalaureate and Beyond Sample

	All		White male	
	Wage (1)	Pctl (2)	Wage (3)	Pctl (4)
Major	0.551***	0.392***	0.497***	0.337***
	(0.137)	(0.104)	(0.146)	(0.103)
Black × Major	-0.134*	-0.095*	-0.145*	-0.101*
	(0.077)	(0.055)	(0.077)	(0.053)
Student Characteristics	X	X	X	X
State & Year FE	X	X	X	X
Institution SAT Percentiles	X	X	X	X
Observations	26,400	26,400	26,400	26,400

Source – U.S. Department of Education, National Center for Education Statistics, 2008/18 Baccalaureate and Beyond Longitudinal Study (B&B:08/18)
Notes – This table reports estimates similar to those in Table 5, but uses alternative measures of the return to college major for the independent variable *Major*. In Columns (1) and (2) the major wage return / percentile return are computed using all prime age workers. In Columns (3) and (4) the major wage return / percentile return are computed using white male prime age workers. Student characteristics (age and gender) as well as year and state fixed effects are included in all specifications. The 25th and 75th percentile math and verbal SAT scores for the institution (4 variables) are included to control for institution quality. Students not identified as either black or white are excluded from the analysis. Standard errors clustered by graduation major are reported in parenthesis: $*p < .1, **p < .05, ***p < .01$

Table C.6: Major Selection and Log Earnings Regressions with Institution Fixed Effects, Baccalaureate and Beyond Sample

	Grad Major		Log Earnings	
	(1)	(2)	(3)	(4)
	Panel A: Major Wage Return			
Black	0.046***	0.055***	-0.071***	-0.075***
	(0.007)	(0.009)	(0.020)	(0.019)
Black × SAT		0.005		
		(0.003)		
Major			0.543***	0.553***
			(0.108)	(0.108)
Black × Major			-0.134*	-0.134
			(0.080)	(0.085)
	Panel B: Major Percentile Return			
Black	0.067***	0.078***	-0.069***	-0.073***
	(0.009)	(0.012)	(0.020)	(0.019)
Black × SAT		0.007		
		(0.005)		
Major			0.387***	0.392***
			(0.085)	(0.085)
Black × Major			-0.093	-0.093
			(0.059)	(0.063)
	Panel C: Major STEM Courses			
Black	0.040***	0.050***	-0.061***	-0.062***
	(0.008)	(0.013)	(0.020)	(0.020)
Black × SAT		0.006		
		(0.005)		
Major			0.342***	0.346***
			(0.096)	(0.098)
Black × Major			-0.084	-0.087
			(0.085)	(0.086)
Student Characteristics	X	X	X	X
Student SAT FE	X	X		X
State & Year FE			X	X
Institution FE	X	X	X	X
Observations	11,480	11,480	26,390	26,390

Source – U.S. Department of Education, National Center for Education Statistics, 2008/18 Baccalaureate and Beyond Longitudinal Study (B&B:08/18)
Notes – For columns (1) and (2), the outcome variable for panel A is the major average wage return from the ACS for white graduates, for panel B is the percentile ranking of the average wage return from the ACS for white graduates by major, and for panel C is the fraction of course credits in STEM courses by major. For columns (3) and (4), the outcome variable is log earnings and the definition of the major variable is given in the panel title. Student characteristics include an indicator for gender and student age at matriculation. Students not identified as either black or white are excluded from the analysis. Standard errors clustered by institution in columns (1) and (2) and clustered by graduation major in columns (3) and (4) are reported in parenthesis: $*p < .1, **p < .05, ***p < .01$

Table C.7: Major Selection and Log Earnings Regressions Controlling for High School GPA

	State Schools			B&B
	1st-Yr Major	Grad Major		Log Earnings
	(1)	(2)	(3)	(4)
	Panel A: Major Wage Return			
Black	0.035***	0.036***	0.071***	-0.068***
	(0.002)	(0.003)	(0.009)	(0.015)
Black × SAT	0.005***	0.009***	0.008**	
	(0.001)	(0.001)	(0.003)	
Major				0.564***
				(0.132)
Black × Major				-0.137*
				(0.076)
	Panel B: Major Percentile Return			
Black	0.046***	0.046***	0.099***	-0.065***
	(0.003)	(0.004)	(0.012)	(0.015)
Black × SAT	0.006***	0.011***	0.010**	
	(0.001)	(0.002)	(0.004)	
Major				0.400***
				(0.105)
Black × Major				-0.098*
				(0.056)
	Panel C: Major STEM Courses			
Black	0.032***	0.031***	0.063***	-0.053***
	(0.002)	(0.004)	(0.012)	(0.016)
Black × SAT	0.005***	0.013***	0.011**	
	(0.001)	(0.002)	(0.004)	
Major				0.346***
				(0.117)
Black × Major				-0.075
				(0.082)
Student Characteristics	X	X	X	X
Student SAT FE	X	X	X	
State & Year FE				X
High School GPA	X	X	X	X
Observations	805,728	381,275	11,550	26,400

Source – Multiple-Institution Database for Investigating Engineering Longitudinal Development (MIDFIELD) and U.S. Department of Education, National Center for Education Statistics, 2008/18 Baccalaureate and Beyond Longitudinal Study (B&B:08/18)
Notes – For columns (1), (2), and (3), the outcome variable for panel A is the major average wage return from the ACS for white graduates, for panel B is the percentile ranking of the average wage return from the ACS for white graduates, and for panel C is the fraction of course credits in STEM courses by major. For column (4), the outcome variable is log earnings and the definition of the major variable is given in the panel title. High school GPA is included as a control variable in columns (1) and (2) and high school GPA categories as reported in the B&B are included as fixed effects in columns (3) and (4). Students not identified as either black or white are excluded from the analysis. Standard errors clustered by institution in columns (1), (2), and (3) and clustered by graduation major in column (4) are reported in parenthesis: $*p < .1, **p < .05, ***p < .01$

Table C.8: Major Selection by Race and SAT Score, Engineering Majors Excluded

	State Schools				B&B	
	1st-Yr. Major		Grad. Major		Grad. Major	
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Major Wage Return					
Black	0.029***	0.031***	0.017***	0.021***	0.056***	0.065***
	(0.002)	(0.001)	(0.002)	(0.002)	(0.006)	(0.009)
Black × SAT		0.001**		0.003***		0.005
		(0.001)		(0.001)		(0.003)
	Panel B: Major Percentile Return					
Black	0.039***	0.042***	0.021***	0.027***	0.080***	0.093***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.009)	(0.012)
Black × SAT		0.002**		0.005***		0.007
		(0.001)		(0.001)		(0.005)
	Panel C: Major STEM Courses					
Black	0.030***	0.034***	0.007**	0.015***	0.042***	0.052***
	(0.002)	(0.002)	(0.003)	(0.004)	(0.008)	(0.012)
Black × SAT		0.003***		0.007***		0.006
		(0.001)		(0.001)		(0.004)
Student Characteristics	X	X	X	X	X	X
Student SAT FE	X	X	X	X	X	X
Institution × Start Year FE	X	X	X	X		
Institution SAT Percentiles					X	X
Observations	769,267	769,267	362,334	362,334	10,750	10,750

Source – Multiple-Institution Database for Investigating Engineering Longitudinal Development (MIDFIELD) and U.S. Department of Education, National Center for Education Statistics, 2008/18 Baccalaureate and Beyond Longitudinal Study (B&B:08/18)
Notes – Students not identified as either black or white and student in an Engineering major are excluded from this analysis. The panel A outcome variable is the average wage return from the ACS for white graduates by major. The panel B outcome variable is the percentile ranking of the average wage return from the ACS for white graduates by major. The panel C outcome variable is the fraction of course credits in STEM courses by major in the state school sample. Student characteristics include student age at matriculation, a female indicator, SAT fixed effects, and in the state school sample a transfer student indicator is also included. In the state schools sample, institution by start year fixed effects are included. In the B&B sample, the 25th and 75th percentile math and verbal SAT scores for the institution (4 variables) are included to control for institution quality. Standard errors clustered by institution (both samples) and year of college entry (state schools sample) are reported in parenthesis: $*p < .1, **p < .05, ***p < .01$

Table C.9: First-Year Major Selection (wage return) by Race, SAT, and Neighborhood Characteristics

	State Schools			
	(1)	(2)	(3)	(4)
Black	0.031***	0.029***	0.029***	0.031***
	(0.003)	(0.003)	(0.003)	(0.003)
Black × SAT	0.005***	0.005***	0.005***	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)
Median Income (10,000s)		-0.002***		
		(0.000)		
Median Income × SAT		-0.000**		
		(0.000)		
Median Education			-0.003***	
			(0.000)	
Median Education × SAT			-0.000***	
			(0.000)	
Income Mobility				-0.037***
				(0.012)
Income Mobility × SAT				-0.034***
				(0.011)
Student Characteristics	X	X	X	X
Student SAT FE	X	X	X	X
Institution × Start Year FE	X	X	X	X
Observations	626,180	626,180	626,180	626,180

Source – Multiple-Institution Database for Investigating Engineering Longitudinal Development (MIDFIELD)
Notes – The outcome is first-year major wage return from the ACS for white graduates by major. This table is similar to Table 6 where the outcome is the graduation major wage return. Only students who identified as either black or white and have an observable home ZCTA and county are included in the analysis. Median household income and median education are measured at the ZCTA level in the census year (1990, 2000, 2010) prior to the student entering college. Income mobility is measured at the county level. Student characteristics include student age at matriculation, a female indicator, SAT fixed effects, and a transfer student indicator. Institution by start year fixed effects are included. Standard errors clustered by institution and year of college entry are reported in parenthesis: $*p < .1, **p < .05, ***p < .01$

Table C.10: Major Selection (Wage Percentile) by Race, SAT, and Neighborhood Characteristics

	State Schools					
	First Major			Graduation Major		
	(1)	(2)	(3)	(4)	(5)	(6)
Black	0.039***	0.039***	0.041***	0.051***	0.050***	0.051***
	(0.004)	(0.004)	(0.004)	(0.006)	(0.006)	(0.006)
Black × SAT	0.006***	0.006***	0.006***	0.012***	0.012***	0.010***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
Median Income (10,000s)	-0.002***			0.002***		
	(0.000)			(0.001)		
Median Income × SAT	-0.000*			-0.001*		
	(0.000)			(0.000)		
Median Education		-0.003***			0.002*	
		(0.001)			(0.001)	
Median Education × SAT		-0.000**			-0.001	
		(0.000)			(0.000)	
Income Mobility			-0.038**			0.116***
			(0.017)			(0.024)
Income Mobility × SAT			-0.052***			-0.107***
			(0.014)			(0.016)
Student Characteristics	X	X	X	X	X	X
Student SAT FE	X	X	X	X	X	X
Institution × Start Year FE	X	X	X	X	X	X
Observations	626,180	626,180	626,180	316,259	316,259	316,259

Source – Multiple-Institution Database for Investigating Engineering Longitudinal Development (MIDFIELD)
Notes – The outcome is major wage percentile from the ACS for white graduates. This table is similar to Table 6 where the outcome is the graduation major wage return. Only students who identified as either black or white are included in the analysis. Student characteristics include student age at matriculation, a female indicator, SAT fixed effects, and a transfer student indicator. Median household income and median education, are measured at the ZCTA level and income mobility is measured at the county level. Standard errors clustered by institution and year of college entry are reported in parenthesis: $*p < .1, **p < .05, ***p < .01$

Table C.11: Major Selection (STEM Courses) by Race, SAT, and Neighborhood Characteristics

	State Schools					
	First Major			Graduation Major		
	(1)	(2)	(3)	(4)	(5)	(6)
Black	0.027***	0.027***	0.029***	0.025***	0.026***	0.028***
	(0.003)	(0.003)	(0.003)	(0.005)	(0.005)	(0.005)
Black × SAT	0.008***	0.008***	0.006***	0.012***	0.012***	0.011***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
Median Income (10,000s)	-0.004***			-0.004***		
	(0.000)			(0.000)		
Median Income × SAT	0.000			-0.000		
	(0.000)			(0.000)		
Median Education		-0.005***			-0.006***	
		(0.001)			(0.001)	
Median Education × SAT		0.000			-0.000	
		(0.000)			(0.000)	
Income Mobility			-0.111***			-0.122***
			(0.016)			(0.023)
Income Mobility × SAT			-0.085***			-0.065***
			(0.011)			(0.013)
Student Characteristics	X	X	X	X	X	X
Student SAT FE	X	X	X	X	X	X
Institution × Start Year FE	X	X	X	X	X	X
Observations	626,180	626,180	626,180	316,259	316,259	316,259

Source – Multiple-Institution Database for Investigating Engineering Longitudinal Development (MIDFIELD)
Notes – The outcome is the fraction of course credits in STEM courses in the major. This table is similar to Table 6 where the outcome is the graduation major wage return. Only students who identified as either black or white are included in the analysis. Student characteristics include student age at matriculation, a female indicator, SAT fixed effects, and a transfer student indicator. Median household income and median education, are measured at the ZCTA level and income mobility is measured at the county level. Standard errors clustered by institution and year of college entry are reported in parenthesis: $*p < .1, **p < .05, ***p < .01$

Table C.12: Major Selection (Other Measures) by Race, SAT, and Neighborhood Characteristics

	B&B					
	Percentile Return			STEM Courses		
	(1)	(2)	(3)	(4)	(5)	(6)
Black	0.100***	0.101***	0.102***	0.061***	0.064***	0.066***
	(0.013)	(0.013)	(0.013)	(0.014)	(0.014)	(0.014)
Black x SAT	0.009*	0.009*	0.009*	0.007	0.008	0.008
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Median Income (10,000s)	-0.003			-0.010***		
	(0.002)			(0.001)		
Median Income x SAT	-0.000			-0.001		
	(0.001)			(0.001)		
Median Education		-0.003			-0.013***	
		(0.002)			(0.002)	
Median Education x SAT		-0.000			-0.001	
		(0.001)			(0.001)	
Income Mobility			-0.062			-0.306***
			(0.072)			(0.067)
Income Mobility x SAT			-0.001			-0.038
			(0.034)			(0.032)
Student Characteristics	X	X	X	X	X	X
Student SAT FE	X	X	X	X	X	X
Institution SAT Percentiles	X	X	X	X	X	X
Observations	8,367	8,367	8,367	8,323	8,323	8,323

Source – U.S. Department of Education, National Center for Education Statistics, 2008/18 Baccalaureate and Beyond Longitudinal Study (B&B:08/18)
Notes – The outcome in columns (1), (2), and (3) is major wage percentile from the ACS for white graduates. The outcome in columns (4), (5), and (6) is the fraction of course credits in STEM courses in the major. This table is similar to Table 6 where the outcome is the graduation major wage return. Only students who identified as either black or white are included in the analysis. Student characteristics include student age at matriculation, a female indicator, and SAT fixed effects. Median household income and median education, are measured at the ZCTA level and income mobility is measured at the county level. Standard errors clustered by the institution are reported in parenthesis: $*p < .1, **p < .05, ***p < .01$

Table C.13: Log Earnings by Graduation Major Selection, Race, and Neighborhood Characteristics

	B&B					
	Major = Percentile Return			Major = STEM Courses		
	(1)	(2)	(3)	(4)	(5)	(6)
Black	-0.051**	-0.062***	-0.064***	-0.038	-0.050**	-0.052**
	(0.022)	(0.022)	(0.021)	(0.024)	(0.023)	(0.023)
Major	0.371***	0.349***	0.399**	0.355***	0.349***	0.466***
	(0.120)	(0.114)	(0.156)	(0.102)	(0.117)	(0.105)
Black × Major	-0.193***	-0.175**	-0.183***	-0.128	-0.116	-0.123
	(0.069)	(0.069)	(0.066)	(0.114)	(0.114)	(0.108)
Median Income (10,000s)	0.027***			0.030***		
	(0.003)			(0.004)		
Median Income × Major	0.010			0.007		
	(0.010)			(0.015)		
Median Education		0.019***			0.024***	
		(0.005)			(0.005)	
Median Education × Major		0.014			0.006	
		(0.013)			(0.016)	
Income Mobility			0.911***			1.004***
			(0.153)			(0.137)
Income Mobility × Major			0.168			-0.672
			(0.617)			(0.412)
Student Characteristics	X	X	X	X	X	X
Year FE	X	X	X	X	X	X
Institution SAT Percentiles	X	X	X	X	X	X
Observations	20,394	20,394	20,394	20,394	20,394	20,394

Source – U.S. Department of Education, National Center for Education Statistics, 2008/18 Baccalaureate and Beyond Longitudinal Study (B&B:08/18)
Notes – The outcome variable is log earnings measured at either 1, 4 or 10 years after graduation. Major is defined in columns (1) - (3) as the major wage return percentile from the ACS for white graduates and in columns (4) - (6) as the fraction of credits from STEM courses by major. This table is similar to Table 7 where major is defined as the wage return for the major. Median household income and median education, are measured at the ZCTA level and income mobility is measured at the county level. Standard errors clustered by the graduation major are reported in parenthesis: $*p < .1, **p < .05, ***p < .01$

Table C.14: Major Selection by Race and SAT Score by Gender

	State Schools				B&B	
	1st-Yr. Major		Grad. Major		Grad. Major	
	Female (1)	Male (2)	Female (3)	Male (4)	Female (5)	Male (6)
	Panel A: Major Wage Return					
Black	0.056***	0.017***	0.055***	0.009**	0.085***	0.045***
	(0.002)	(0.003)	(0.004)	(0.004)	(0.011)	(0.013)
Black × SAT	0.001	0.012***	0.003**	0.015***	0.006	0.012**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.004)	(0.005)
	Panel B: Major Percentile Return					
Black	0.075***	0.020***	0.071***	0.009*	0.120***	0.063***
	(0.003)	(0.003)	(0.005)	(0.005)	(0.014)	(0.017)
Black × SAT	0.001	0.015***	0.003*	0.019***	0.008	0.016**
	(0.001)	(0.001)	(0.002)	(0.002)	(0.005)	(0.007)
	Panel C: Major STEM Courses					
Black	0.057***	0.013***	0.053***	0.002	0.079***	0.035*
	(0.003)	(0.003)	(0.005)	(0.005)	(0.015)	(0.019)
Black × SAT	0.001	0.013***	0.005***	0.019***	0.011**	0.012
	(0.001)	(0.001)	(0.002)	(0.002)	(0.005)	(0.007)
Student Characteristics	X	X	X	X	X	X
Student SAT FE	X	X	X	X	X	X
Institution × Start Year FE	X	X	X	X		
Institution SAT Percentiles					X	X
Observations	449,226	485,224	217,364	233,624	6,784	4,764

Source – Multiple-Institution Database for Investigating Engineering Longitudinal Development (MIDFIELD) and U.S. Department of Education, National Center for Education Statistics, 2008/18 Baccalaureate and Beyond Longitudinal Study (B&B:08/18)
Notes – The panel A outcome variable is the average wage return from the ACS for white graduates by major. The panel B outcome variable is the percentile ranking of the average wage return from the ACS for white graduates by major. The panel C outcome variable is the fraction of course credits in STEM courses by major in the state school sample. Student characteristics include student age at matriculation, SAT fixed effects, and in the state school sample a transfer student indicator is also included. In the state schools sample, institution by start year fixed effects are included. In the B&B sample, the 25th and 75th percentile math and verbal SAT scores for the institution (4 variables) are included to control for institution quality. Students not identified as either black or white are excluded from the analysis. Standard errors clustered by institution (both samples) and year of college entry (state schools sample) are reported in parenthesis: $*p < .1, **p < .05, ***p < .01$

Table C.15: Effect of Major Choice on Log Earnings, Heterogeneous Effects, American Community Survey Sample

	Gender		Age		
	Male	Female	< 30 Yrs.	31-50 Yrs.	51+ Yrs.
	(1)	(2)	(3)	(4)	(5)
	Panel A: Major Wage Return				
Major	0.873***	0.760***	0.727***	0.864***	0.840***
	(0.037)	(0.053)	(0.083)	(0.030)	(0.050)
Major × Black	-0.280***	-0.232***	-0.285***	-0.337***	-0.332***
	(0.063)	(0.051)	(0.047)	(0.070)	(0.058)
	Panel B: Major Percentile Return				
Major	0.668***	0.560***	0.547***	0.648***	0.630***
	(0.024)	(0.043)	(0.068)	(0.023)	(0.042)
Major × Black	-0.222***	-0.170***	-0.211***	-0.253***	-0.253***
	(0.042)	(0.041)	(0.036)	(0.050)	(0.042)
	Panel C: Major STEM Courses				
Major	0.452***	0.419***	0.360**	0.461***	0.491***
	(0.084)	(0.116)	(0.150)	(0.083)	(0.088)
Major × Black	-0.044	-0.056	-0.132**	-0.095	-0.150**
	(0.052)	(0.052)	(0.050)	(0.066)	(0.059)
Female Indicator			X	X	X
Year Fixed Effect	X	X	X	X	X
Age Fixed Effects	X	X	X	X	X
State X Black Fixed Effects	X	X	X	X	X
Observations	1,371,887	1,278,512	465,175	1,352,932	832,292

Source – U.S. Census Bureau, 2011-2021 American Community Survey, Public Use Microdata
Notes – Data from the American Community Survey. Outcome is log wage and salary income in 2020 real dollars. Robust standard errors clustered at the major level reported in parentheses. Column (1) includes only male workers. Column (2) includes only female workers. Column (3) includes only workers less than thirty years old. Column (4) includes only workers 31 to 50 years old. Column (5) includes only workers over fifty-one years old. Standard errors clustered by the graduation major are reported in parenthesis: $*p < .1, **p < .05, ***p < .01$

D Extended Theory: Empirical Strategies for Mismatch

Consider a modified version of our model. First, suppose that black students face an investment cost $c(m_i)$, with $\frac{\partial c}{\partial m_i} > 0$ and $\frac{\partial^2 c}{\partial m_i} > 0$. This could represent structural barriers caused by discrimination in higher educational institutions. It could also represent a mitigable (with cost) preparation disadvantage due to inequalities in secondary and primary education. For example, black students may have less access to AP credits than white students, and must instead make relatively costlier choices, like entering student government, in order to have a strong enough application to be accepted to top institutions.[32]

Proposition 8. *When black students face an additional investment cost, they may choose investments that are more difficult, less difficult, or equal to those of white students. They may choose investments that are more difficult, less difficult, or equal to* $M^*(\rho)$

Proof. The proposition follows because $c(m_i)$ directly influences black student investment choice. Thus, for sufficiently low levels of $c(m_i)$, we can obtain qualitatively identical results to Proposition 4, and for sufficiently high levels of $c(m_i)$ we can obtain the opposite results.

For black students with beliefs 0, note that if $c(0) = 0$, $M_b(0) = M^*(0) = 0$ as in Proposition 1. If $c(M^*(0)) > 0$ it then follows by the same arguments as in the proof of Proposition 1 that $M_b(0) = \arg\max E_r(w|m, 0) - c(m)$ which is lower than $M^*(0)$ as $c(m) > 0$.

Now, consider the modified first order condition for black students in this environment:

$$m\tau_r^{-1}\varsigma_b^{-2}\frac{\partial P_b(m)}{\partial m} + \rho - 2m = \frac{\partial c}{\partial m}. \tag{19}$$

Rearranging terms,

$$\frac{\partial M_b(\rho)}{\partial \rho} = \frac{M_b(\rho)}{\frac{\partial c}{\partial m} + 2M_b(\rho) - \rho}\frac{\varsigma_r^{-2}}{\tau_b}. \tag{20}$$

$$\tag{21}$$

Compared to the differential equation in Proposition 2, it is clear that an increase in $\frac{\partial c}{\partial m}$ reduces $\frac{\partial M_b(\rho)}{\partial \rho}$. Thus, for a sufficiently large $\frac{\partial c}{\partial m}$, black students may choose the same investments as whites, or choose less difficult investments. $\square$

The proposition shows that adding costs may cause black students to choose less difficult investments than white students. However, this need not be the case, and depends on the

[32]Here we mean costlier in the sense that student government may require similar time commitments while generating less human capital than advanced course work, but the general point is that a student with a smaller investment choice set must select a weakly worse investment portfolio to generate the same quality application.

nature of the cost function. The incentives created by statistical discrimination still motivate black students to take on more difficult investments that partially counteract their higher costs.

Perversely, some amount of barriers may actually *improve* black labor market outcomes. When these costs are relatively low, they push black student investments closer to $M^*(\rho)$, which will raise their accumulated human capital. It will also raise black wages conditional on ρ, as firm beliefs are correct in equilibrium. However as the costs continue to increase, black students undertake investments that are *below* $M^*(\rho)$, creating a human capital gap with white students due to *underinvestment*, and a larger resulting wage gap as well.

Policymakers concerned about this loss of equity may seek remedy through an "affirmative action" subsidy, $b(m)$. What should this subsidy be? We can imagine several different aims. A policymaker may choose to set $b(m) = c(m)$, so that black and white students face an identical investment choice problem. Under this regime, the model reduces to the one we analyzed in Section 2. If black students face sufficiently strong statistical discrimination, they will choose more difficult investments than white students, and receive lower wages in the market. Alternatively, a policymaker may choose a $b(m)$ so that $M_b(\rho) = M^*(\rho)$. That is, choose an affirmative action policy which incentivizes black students to select their in expectation human capital maximizing investment. Following our analysis in Section 2, under this regime, white students will overmatch, choosing more difficult investments than black students. Black students will then outearn white students with the same ρ, as black students' equilibrium investment choices generate higher levels of human capital than whites'. Finally, a policymaker may choose a $b(m)$ so that $M_b(\rho) = M_w(\rho)$. Under this regime, any aggregate racial wage gap will be due only to differences in ρ (and in turn a).

Critics of affirmative action often raise concerns that racial admissions preferences lead black students to enroll in universities that are too difficult given their academic preparation, leading to worse outcomes than had they enrolled in a less selective institutions. Denote $b_1(m)$ as the affirmative action policy being evaluated. To be more precise, we will differentiate between two different versions of this mismatch hypothesis.

Definition. *Black students are weakly mismatched if, at* $b_1(m)$, $M_b(\rho) > M^*(\rho)$.

Definition. *Black students are strongly mismatched if, relative to* $b_1(m)$, $b_2(m) = 0$ *leads to higher equilibrium black wages.*

Weak mismatch does not necessarily imply that optimal policy should eliminate racial admissions preferences. Instead, it simply states that current levels of affirmative action are too high, and that reducing them will lead to better outcomes for black students.[33] In

[33]Note that even in our signaling framework, any policy which induced $M_b(\rho) = M^*(\rho)$ would maximize

contrast, strong mismatch occurs only if racial admissions preferences are so inefficiently high that abolishing these preferences would improve black student outcomes. Even under strong mismatch, it is not obvious that abolishing affirmative action would be optimal. When $c(m) > 0$ the policy that maximizes black labor outcomes may, and arguably likely will, be some $b_3(m)$ with $0 < b_3(m) < b_1(m)$. Strong mismatch implies weak mismatch, but not the reverse.

Two different approaches have been proposed to test the mismatch hypothesis. The first relies on some students being as-good-as-randomly assigned to universities, while the second relies on state-level policy changes to affirmative action. We will now discuss the effectiveness of these strategies in light of our findings. We will consider both using data on younger workers, where information is incomplete and firms engage in statistical discrimination as in our model, as well as older workers where learning causes information to be complete, and wages are equal to productivity.

The as-good-as-random assignment strategy uses a natural experiment which shifts a small number of black students from a high-quality institution to a lower quality institution and then compares the wages of these students. In principle, this could come through a regression discontinuity design at an admissions test threshold, as is common in the returns to school quality literature (e.g., Hoekstra, 2009; Zimmerman, 2019). Mountjoy and Hickman (2021) instead compare students who have applied and were admitted to the same set of universities, but made different matriculation choices. In their data from Texas, they find large disparities in preparation between black and white students at top public universities. Yet, once excluding historically black colleges and universities (HBCUs), black students who attend better universities perform better in the labor market 8-10 years post graduation. They thus conclude that mismatch does not harm black students. The following proposition shows that the validity of such approaches depends crucially on whether firms possess complete or incomplete information about workers.

Proposition 9. *The as-good-as-random assignment approach can confirm but cannot reject weak mismatch for young workers. The as-good-as-random assignment approach can confirm or reject weak mismatch for older workers.*

Proof. For the first part of the proposition, note that the estimator compares workers with $M^*(\rho)$ to those who are as-good-as-randomly assigned to $m' < M^*(\rho)$. As the assignment mechanism does not change market beliefs the differences in wages between these two groups

black wages, because it would maximize the average black worker's human capital.

is

$$M_b(\rho)\rho - M_b(\rho)^2 - \left(m'\tau_r^{-1}\left[\varsigma_r^{-2}P_r(m') + \left(\sigma_r^{-2} + \varrho^{-2}\right)\rho\right] - [m']^2\right). \tag{22}$$

Now instead consider an alternative $b(m)$ such that $M_b(\rho) = M^*(\rho)$ for all ρ. Because $M^*(\rho)$ maximizes productivity, it must be the case that

$$M^*(\rho)\rho - M^*(\rho)^2 > m'\tau_r^{-1}\left[\varsigma_r^{-2}P_r(m') + \left(\sigma_r^{-2} + \varrho^{-2}\right)\rho\right] - [m']^2. \tag{23}$$

It thus follows from transitivity that if equation (22) is negative then black workers are weakly mismatched under the current $b(m)$. By definition of a well-behaved equilibrium $P_r(m') < \rho$. Thus a positive estimate for equation (22) does not imply $M_b(\rho) < M^*(\rho)$.

For the second part of the proposition, note that when workers are old their ability is known as a which is equal to ρ in expectation. Thus, the estimate is simply

$$M_b(\rho)\rho - M_b(\rho)^2 - \left[m'\rho - (m')^2\right], \tag{24}$$

which is the causal effect of investment on human capital. □

While this approach identifies the causal return to university selectivity on labor market outcomes at any age, it is not effective at evaluating whether $M_b(\rho) > M^*(\rho)$ when investments act as signals and information is incomplete.[34] To see this, assume that $M_b(\rho) > M^*(\rho)$, and consider taking a small number of students and instead assigning them to to $M^*(\rho)$. Crucially, such a policy will not change the market beliefs on the relationship between m and ρ. From equation (8), such a student will receive expected wages

$$E_r(w|M^*(\rho), \rho) = M^*(\rho)\tau_r^{-1}\left[\varsigma_r^{-2}P_r(M^*(\rho)) + \left(\sigma_r^{-2} + \varrho^{-2}\right)\rho\right] - M^*(\rho)^2. \tag{25}$$

While this change in assignment raises accumulated human capital, it lowers the signaling value of investment, $\varsigma_r^{-2}P_r(M^*(\rho))$, as the market now believes the worker has the lower level of aptitude typical associated with $M^*(\rho)$. This wage is thus lower than an alternative $b(m)$ which maximized human capital, where $P_r(M^*(\rho)) = \rho)$. Finding empirically that students who attend lower ranked institutions earn higher wages would clearly indicate weak mismatch; it could only be because $b(m)$ is so high that black students are choosing to forego a higher wage in exchange for the subsidy of attending a high ranked institution. But finding that students who attend lower ranked institutions earn less could still occur under weak mismatch, because the gains from signaling at higher levels of m offset the losses to human capital.

[34] Hoekstra (2009) also notes that such estimators will include both the causal effect of education quality on human capital, and the signaling value of institution quality.

In contrast, with older workers for whom the market has full information, wage is equal to productivity, and thus the estimator gives the causal effect of university choice on labor market capital. A positive causal estimate of university selectivity would correctly be interpreted as indicating that at $b(m)$ increasing racial admissions preferences would increase black workers' human capital, while a negative would instead indicate that reducing $b(m)$ would improve black outcomes.

Now consider the second identification strategy. Several states have banned racial preferences in admissions, which allows researchers to compare the outcomes of cohorts who differed in their exposure to affirmative action (e.g., Hinrichs, 2012, 2014; Arcidiacono et al., 2016). More recently Bleemer (2022) finds that a ban on racial admissions preferences in California led to reduced wages as young adults for URMs, with effects concentrated on Hispanics. Across a wider set of states, Antman et al. (2024) find that bans had a negative impact on Hispanic women, but evidence for other URMs is mixed. In addition, they find substantial heterogeneity in estimates across states. Because these bans change the investment decisions of all black students, they also change employer beliefs.

Proposition 10. *Affirmative action bans can confirm or reject strong mismatch for young or older workers. Affirmative action bans can confirm but cannot reject weak mismatch for young or older workers.*

Proof. For the first part of the proposition, note that the equilibrium average wage for a worker with beliefs ρ is

$$E_r(w|\rho) = M_r(\rho)\rho - M_r(\rho)^2. \tag{26}$$

Denote M_{r1} as the mapping from ρ to m under $b_1(m)$ and similarly M_{r2} for $b_2(m) = 0$. The approach is thus estimating

$$M_{b2}(\rho)\rho - M_{b2}(\rho)^2 - M_{b1}(\rho)\rho - M_{b1}(\rho)^2, \tag{27}$$

which is a direct test of strong mismatch.

For the second part of the proposition, suppose that black wages are higher under $b_2(m) = 0$ than $b_1(m)$. Since $M_{b2}(\rho) < M_{b1}(\rho)$ this can only be the case when lowering $M_b(\rho)$ raises black wages, which can only be the case if $M_{b1}(\rho) > M^*(\rho)$ and thus $b_1(m)$ induces weak mismatch.

However, suppose that black wages are lower under $b_2(m) = 0$ than $b_1(m)$. From Proposition 8 we know that at $b_2(m) = 0$, $M_{b2}(\rho)$ may be greater than, less than, or equal to $M^*(b)$. If $M_{b2}(\rho) < M^*(\rho)$, there will be other $M_b(\rho) > M^*(\rho)$ that can lead to higher wages than M_{b2}, but these will still have weak mismatch.

□

The affirmative action ban identification strategy directly tests current $b_1(m)$ versus $b_2(m) = 0$; that is, it tests for strong mismatch. Such bans change all students' behavior and thus also change employer beliefs. Since employer beliefs are correct on average in equilibrium, it does not matter whether the workers studied are young or older.

However, this strategy can only confirm weak mismatch but cannot reject it. To see this, note that under large barriers (high $c(m)$) without affirmative action ($b(m) = 0$), black investments will be too low relative to the human capital maximizing optimum. Thus, *even if* affirmative action induces weak mismatch, black students may see higher wages because this mismatch is less severe than the *undermatch* they experience in its absence. If we observe that black wages decrease after a ban in racial admissions preferences, it only tells us that black students were undermatched without these preferences. It does not tell us whether some other policy that lessened, but did not eliminate, racial admissions preferences would lead to higher black wages by reducing the amount of overmatch in black human capital investments. In contrast, since a ban on racial admissions preferences can only reduce the difficulty of black human capital investments, an increase in black wages would provide strong evidence that these preferences induced mismatch. Systematically reducing $M_b(\rho)$ will raise wages only if $M_b(\rho) > M^*(\rho)$.

The above analysis shows that both methods have their strengths and weaknesses when evaluating the mismatch hypothesis. The affirmative action bans approach can test and reject strong mismatch even with young workers. This provides an advantage when we are interested in evaluating more recent policy changes where full career histories are not yet available. However, it can only confirm weak mismatch and cannot reject it, even in a full information environment. The as-good-as-random assignment method is capable of rejecting weak mismatch, but only in a full information environment. This is plausible with older workers, because as the market learns about worker productivity the signaling value of any investment heads to zero. In practice, the market appears to learn about worker productivity relatively quickly (Lange, 2007; Aryal et al., 2022). A regression discontinuity or similar approach using the wages of even mid-career workers may be able to credibly reject weak mismatch.