What Jobs Come to Mind?
Stereotypes about Fields of Study

John J. Conlon* Dev Patel†

January 9, 2025

Abstract

Using both large-scale nationally representative data and surveys administered among undergraduates at the Ohio State University, we measure how US freshmen perceive the relationship between college majors and occupations. We show that students stereotype fields of study, greatly exaggerating the likelihood that majors lead to their distinctive jobs (e.g., counselor for psychology, journalist for journalism). Estimates from a structural model suggest that such stereotyping distorts decisions because students have strong preferences for future occupations when choosing their major. In a field experiment, we find that reducing stereotyping has significant effects on students' intentions about what to study as well as the classes and majors in which they enroll. Misperceptions also skew towards the careers and majors of people students know personally, consistent with a recall-based model of belief formation.

*Carnegie Mellon University, jconlon@andrew.cmu.edu. †Harvard University, devpatel@fas.harvard.edu.
We thank the Exploration Program at Ohio State for their support in implementing our surveys, and Jenna Anders, Zach Bleemer, Pedro Bordalo, Emily Breza, Katie Coffman, Lucas Coffman, Meg Edison, Ben Enke, Nicola Gennaioli, Ed Glaeser, Claudia Goldin, Larry Katz, Gabriel Kreindler, Ross Mattheis, Muriel Niederle, Matthew Rabin, Gautam Rao, Josh Schwartzstein, Andrei Shleifer, Evan Soltas, and Basit Zafar for helpful feedback. This study underwent human subjects review at Harvard University under IRB20-1581. We are grateful for financial support from the Lab for Economic Applications and Policy, the Mind Brain Behavior Initiative, the National Science Foundation Graduate Research Fellowship under grant DGE1745303, and the Alfred P. Sloan Foundation Predoctoral Fellowship in Behavioral Macroeconomics, awarded through the NBER. The experiment described in this paper was pre-registered at the AEA RCT Registry (ID AEARCTR-0008339).

1 Introduction

Stereotypes (in)famously bias perceptions of nationalities, genders, and racial and ethnic groups, leading to inaccurate statistical discrimination, distorted policy preferences, and adverse economic outcomes (e.g., Chan 2022, Bohren et al. 2023, Bursztyn et al. 2023, Haaland & Roth 2023). To explain such stereotypes, psychologists have advocated for a cognitive approach based on a "kernel of truth": mental representations of groups focused on their distinctive characteristics—those more common among them than among others—leading to oversimplified and exaggerated stereotypes based on these traits (Schneider et al. 1979, Judd & Park 1993, Hilton & Von Hippel 1996, Schneider 2005). Existing work in economics has largely either applied this framework to shed light on well-known stereotypes (e.g., about race and gender) or tested underlying cognitive mechanisms using lab experiments (Bordalo et al. 2016, Coffman et al. 2023, Esponda et al. 2023). But this approach also seems to provide a recipe for *discovering* novel stereotypes: wherever there are distinctive traits, we might expect stereotypes about their prevalence to develop, distorting choices that depend on such beliefs.

In this paper, we apply this logic to uncover stereotypes in a new and high-stakes economic context: human capital investment. We focus in particular on students' choice of college major—one of the most consequential economic decisions many people ever make—and how that choice affects the occupation that graduates later attain. Crucially, college majors often have distinctive associated jobs. For example, 16% of political science majors in the US are lawyers, but only 2% of other majors are; 4% of communications and journalism majors are journalists or writers, yet that rate falls to less than 1% among other majors. If stereotypes indeed form by exaggerating kernels of truth, then students may overestimate the likelihood of attaining jobs that are distinctive of the field they choose, distorting their beliefs about the expected returns (both pecuniary and non-pecuniary) to different majors. Despite its potential importance for human capital investment, the extensive literature on beliefs about the returns to education (e.g., Dominitz & Manski 1996, Jensen 2010, Wiswall & Zafar 2015a, Sequeira et al. 2016, Dizon-Ross 2019, de Koning et al. 2022, Alfonsi et al. 2023) has not examined biases about how college majors map onto occupations.

Does stereotyping distort students' beliefs about the occupations they will attain depending on their choice of major? If so, where do these beliefs come from, and how might they distort students' decisions about what major to pursue? We begin by presenting motivating evidence from nationally representative surveys of millions of US college freshmen spanning over 40 years. We define a major M's distinctive occupation as the career c that is most common in M compared to other majors: i.e., that maximizes $P(c|M)/P(c|\text{ not } M)$. Comparing students' expectations about their future major and career with government data on the same cohorts, we find that dramatically more US freshmen expect to attain their major's distinctive career than actually end up working in that job: 63% of biology majors expect to be doctors (in reality, 23% are), 62% of psychology majors expect to be counselors (21% are), 65% of prospective art majors expect to be artists (17% are), 42% of communications/journalism majors expect to be writers or journalists (4% are), and so on. Pooling across majors, these facts combine to produce large gaps between the expected and actual careers that college graduates pursue. For professions that are primarily the rare-but-distinctive outcome of particular majors—e.g., doctor, counselor, journalist—two to four times more college freshmen expect to work in these jobs than actually do. In contrast, many fewer expect to be teachers, working in business, or non-employed than ultimately are because these are the common alternatives to the distinctive career of many majors. These overall gaps—amounting to between 40,000 and 200,000 students a year—appear largely unchanged since at least the 1970s.

By themselves, of course, these patterns are consistent with explanations other than stereotyping of distinctive features. For instance, they could be driven by students exaggerating their own abilities, overestimating the demand for certain jobs, holding motivated beliefs, or selecting fields of study about which they hold particularly extreme beliefs. The nationally representative data are also qualitative in that they ask students to select one of many jobs as their "probable" career, raising the possibility these patterns are an artifact of survey elicitation.

To distinguish between these possibilities and to isolate the role of stereotyping, we designed and administered surveys among first-year students at The Ohio State University (OSU). We find that these students exaggerate distinctive careers even when answering quantitative probabilistic questions about careers conditional on major (allowing them to express uncertainty precisely), about people other than themselves (shutting down overconfidence), and about majors other than their own (shutting down selection, motivated reasoning, and other concerns). In regression analyses, we additionally control for individual-by-career fixed effects to show that these biases are not about salient careers *per se* but about the relationship *between* majors and careers. In other words, our results are not driven by the general salience of certain jobs but rather reflect a within-major phenomenon. These differences are strikingly large and mirror those in the nationally representative survey: OSU students exaggerate the share of artists among art majors by 36 percentage points (p.p.) or 211%, of doctors among biology majors by 11 p.p. (48%), of journalists among journalism majors by 43 p.p. (1,100%), of counselors among psychology majors by 22 p.p. (105%), and so on. We show that the magnitude of biases in the OSU data appear sufficient to predict the majority of the aggregate biases in the nationally representative sample. Using a Shapley-style decomposition, we furthermore show that stereotyping appears *more important* than any of the other channels we consider, in the sense that it explains a greater share of the variance in students' beliefs. We also document substantial stereotyping in an incentivized online survey of US adults (for both college and non-college educated as well as younger and older Americans), suggesting that this misperception is widespread and persists through college and beyond.

We then turn to the implications of stereotyping using two complementary approaches, both of which examine the role of students' career beliefs for their choice of major: a structural model of field of study choice and a field experiment. We first describe a stylized model in which preferences over three dimensions contribute to students' decisions about their major: non-career factors (e.g., difficulty/enjoyability of classes), expected salary, and an individual-specific non-monetary preference for one "preferred" career. We estimate the model using survey data from our OSU sample and document two main findings. First, the jobs students expect to hold play a substantial role in their choice of major, evidence that validates our focus on this set of beliefs. This relationship is not driven by concerns about earnings: students' choices are barely sensitive to (their beliefs about) expected salary across majors. Instead, non-pecuniary preferences over jobs play a large role, and these exhibit a close link to stereotyping: we estimate that the preferred career of over half of students coincides with their most likely major's distinctive job. Second, we find evidence that misperceptions about the mapping between majors and occupations significantly shape their study decisions. Counterfactually replacing students' beliefs with the true relationships based on Census data generates a 15.9 p.p. average decline in students' likelihood of enrolling in their top-ranked major. This same simulation yields increases in the likelihood of majoring in students' second-ranked major because its distinctive job tends not to be their most preferred career.

We complement this structural analysis with a field experiment estimating the effect of a light-touch intervention providing students statistics on the actual past distribution of careers by major. This exercise of course lacks the control and interpretability of the structural model, but it provides evidence on the impact of reducing stereotyping with fewer assumptions. We estimate the impact of information provision on students' self-reported intentions as well as their actual course enrollments and college major declarations up to three years later, measured using administrative data. Consistent with the choice model described above (where students had especially strong preferences for their most likely major's distinctive career), we find that the effect of information depends critically on students' heterogeneous preferences over their future jobs. They react negatively to news that the distinctive career of their *ex ante* most preferred major is unlikely, whereas if anything they react *positively* to news that less preferred majors' distinctive careers are unlikely. We estimate that changing beliefs about the likelihood of attaining the distinctive career of students' top-ranked or second-ranked major by 10 percentage points (p.p.) would change their beliefs about the likelihood of pursuing that major by -3.5 p.p. ($p < 0.05$) or +2.1 p.p. ($p = 0.17$), respectively, and we can reject that these two effects are equal ($p < 0.05$). The same shift in stereotypical beliefs has immediate effects on course enrollment in each field in the following semester, changing class-taking by -0.22 courses ($p < 0.05$) or +0.20 courses ($p < 0.10$, p-value for difference < 0.01) and the probability of having declared a major in that subject a year later by -6.1 p.p. ($p = 0.23$) or +9.9 p.p. ($p < 0.01$, p-value for difference < 0.05). Though the effects of our survey experiment appear to partly fade out over the following years, the treatment also had a large and significant effect on *when* students declare their major: treated students spend on average 0.21 more semesters undecided before declaring a major ($p < 0.05$).[1]

Having documented that stereotypes distort students' beliefs about jobs as well as influence their choices between majors, we conclude by investigating the origins of these belief biases. Recent work explores how stereotyping can arise endogenously when beliefs depend

[1]If anything, treated students are slightly more likely to still be taking classes two or three years later, so the effects on major declaration do not appear driven by students dropping out or becoming discouraged with college.

on what comes to mind (Bordalo et al. 2023). We therefore explore a natural source of heterogeneity across students in the careers and majors that are likely to be top of mind for them: those of people personally close to students (i.e., their parents and other role models). We find that students' beliefs are strongly predicted by the careers and majors of people they happen to know personally. These role-model "effects" appear even when looking at students' beliefs about the joint distribution of careers and majors nationwide. For example, students with a parent who has career-major combination (c, M) believe the share of M majors who have career c is 3.11 p.p. higher ($p < 0.01$) than students without such a role model. For distinctive careers—which students overestimate at baseline—this implies that knowing someone with such a career (which further boosts their beliefs about $P(c|M)$) makes students' beliefs *less* accurate. These results are broadly in line with theories of belief-formation from memory.

This paper contributes to a growing literature on stereotyping. Many studies explore the origin and consequences of stereotypes about race, immigrants, political parties, and gender (e.g., Bordalo et al. 2019, Coffman et al. 2020a, Coffman et al. 2020b, Alesina et al. 2022, **?**). To our knowledge, we are novel in applying the logic of stereotyping based on a "kernal of truth" to *discover* new stereotypes in a high-stakes economic context: namely, human capital investment. Further, our evidence connecting stereotyping and role model effects relates to a growing literature on beliefs arising from memory and what comes easily to mind (e.g., Enke et al. 2021, Malmendier & Wachter 2022, Graeber et al. 2022, Bordalo et al. 2023, Bordalo et al. 2024, Augenblick et al. 2023).

Our study also adds to a rich literature on beliefs and human capital investment (see Giustinelli 2022 for a review). The importance of field of study for economic outcomes has led many studies to investigate whether biased beliefs distort students' college major choices,[2] but these papers tend to focus on beliefs about average salary conditional on major (e.g., Betts 1996, Arcidiacono et al. 2012, Wiswall & Zafar 2015b, Baker et al. 2018, Conlon 2021).[3]

[2]College major choice plays a large and increasing role in shaping the economic prospects of college graduates (Altonji et al., 2014). Differences in, for example, earnings across majors often rival or exceed the wage premium from attending college at all, and they appear to primarily reflect causal effects rather than selection (Hastings et al., 2013; Kirkeboen et al., 2016; Bleemer & Mehta, 2020). These effects appear driven at least in part by the actual classes that students take, rather than (just) the official major they graduate with (Arteaga, 2018).

[3]A notable exception is Arcidiacono et al. (2020), who decompose beliefs about the salary returns to majors into their effects on salaries within occupations and on the likelihood of attaining certain occupations.

To our knowledge, we are the first to document biases in beliefs about the distribution of occupations by major and, therefore, the first to document stereotyping in this domain. Our results also echo a small but growing literature showing that students in vocational programs in developing countries greatly overestimate their own employment prospects post-graduation (e.g., Bandiera et al. 2023) and that reducing these misperceptions can have positive effects (e.g., Alfonsi et al. 2023)..

2 Do Students Stereotype Majors?

In this section, we provide evidence that students stereotype majors, exaggerating the prevalence of jobs that are distinctive of particular fields of study. We start by documenting motivating qualitative evidence from a large-scale and nationally representative survey of US freshmen. We then introduce and show results from our main surveys, which allow us to isolate the role of stereotyping from other forces, including uncertainty, overconfidence, selection, motivated beliefs, and salience. We conclude by discussing external validity, including by showing that similar belief patterns hold in a separate incentivized survey of US adults (including college graduates and respondents with substantial labor market experience).

2.1 Motivating Evidence

We begin by providing motivating evidence from the CIRP Freshman Survey administered by the Higher Education Research Institute (henceforth, the "Freshman Survey"), which surveys incoming first-year students typically during the first weeks of the school year. We pool survey data between 1976 and 2015.[4] We restrict the data to students younger than 24 years with non-missing location (home zip code), race, gender, expected career, and expected major, which leaves 9,068,064 students from 1,587 schools (95.9% of students are at 4-year institutions). Column 1 of Table A.I shows self-reported demographic information about students in the Freshman Survey. Throughout the analysis, we use census data to weight

Their study does not attempt to test whether students' beliefs are consistent with rational expectations. See also Wiswall & Zafar (2021) and Ersoy & Speer (2022) for beliefs about non-labor-market consequences of major choice.

[4] We use data from all years in this range except 1977 and 1978. We choose these years because they include information on students' home zip code which we use for weighting. See https://heri.ucla.edu/instruments/ for a list of participating schools and survey instruments by year.

the Freshman Survey data to match US residents of the same birth cohorts with at least some college education on race, gender, and census division of birth.[5] In that sense, we call this sample nationally representative of incoming college freshmen.

We focus on two questions from the Freshman Survey. First, students are asked to mark their "probable field of study" from a list of around 80 options, including "Other" and "Undecided."[6] We group these fields into 10 major groups (plus "other" and "undecided"), as shown in Table A.II. Similarly, students are asked to report their "probable career occupation" from a list of approximately 45 options, which we group into nine occupation categories (plus "other," "non-employment," and "undecided") as shown in Table A.III. The qualitative nature of the Freshman Survey—i.e., asking students to pick which job is their "probable career" rather than eliciting probabilistic beliefs about a well-defined event—of course raises questions about how to interpret students' responses. We address this and other issues in Section 2.2 and use the Freshman Survey merely as motivating evidence.

Do students in the Freshman Survey appear to stereotype majors? As described above, we focus on stereotyping based on distinctiveness: we define a career c as the most distinctive of major M if it maximizes $p_{c|M}/p_{c|-M}$. The careers that are most distinctive of each major by this definition are intuitive: doctors for biology/chemistry, lawyers for government, counselor for psychology, teachers for education, etc (see Table A.IV for the complete list).

The dark blue bars in Figure 1 show the fraction of students in the Freshman Survey who list their expected major's most distinctive career as their expected career occupation. The dotted lines show the true fraction of college graduates with each major who are working in its distinctive career, which we calculate using the 2017-2019 ACS (Ruggles et al., 2022). We restrict to college-graduate respondents born between 1958 and 1997 who are between 30 and 50 years old when answering the ACS. We see a clear pattern: students in every major are significantly more likely to expect to work in that major's most distinctive career than in fact do. For example, 65% of prospective art majors expect to be artists (only 17% are), 60% of biology majors expect to be doctors (23% are), 42% of communications/journalism majors expect to be writers or journalists (4% are), 62% of psychology majors expect to be counselors (21% are), and so on. All of these differences are statistically significant at the

[5] For people born outside the U.S., we use current location as a proxy for birthplace. We include students in the Freshman Survey data that are non-citizens so long as they self-report a U.S. zip code.

[6] The exact list of majors varies from year to year.

$p < 0.001$ level.

We next examine how the gaps between expected and actual careers have changed over time. Because the US census data that include respondents' college major only began in 2012, we instead compare students' expected careers (unconditional on their major) to the actual distribution of occupations from the Annual Social and Economic Supplement (ASEC) to the Current Population Survey (CPS) (Flood et al., 2021). We restrict the data to those aged 33 to 37, because by this time the vast majority of people are no longer students and have started their career.[7] This also matches well with the age of 35 that we ask about in the 2021 OSU survey, described further below. We match occupation codes from the CPS to the same nine occupation groups (see Table A.V).

Figure 2 shows large, systematic, and persistent differences between the careers that freshmen expect to attain and the actual occupations they go on to have. The blue lines show the share of first-year students each year who expect to have each career. The gray lines show the share of college graduates in the same cohort that are working in that occupation in the CPS. Panel A of Table A.VI shows the corresponding share expecting and actually working in each career, pooling across cohorts.[8] We see that twice as many students expect to become artists, counselors, and lawyers (about 5% each) than actually do (2-3% each). Four times as many students expect to become writers and doctors (2.7% and 11.1%) than do (0.7% and 2.8%).[9] Note that these gaps between expected and actual careers are exactly what we would expect from students stereotyping majors based on their distinctive careers. The occupations with the largest positive gaps—e.g., doctors, writers, lawyers, artists—are those that are primarily the (rare) distinctive outcome of a particular major. In contrast, those with negative gaps—teaching, business, non-employment—are those that tend to be the most common alternatives to other majors' distinctive jobs. Further, these patterns appear quite stable over time, suggesting that the apparent stereotyping in Figure 1 reflects

[7]Our results are not sensitive to this specific age range.

[8]If students' beliefs were on average correct, we might expect the outcomes data to lag the expectations data. For simplicity, we do not attempt to correct for this, and the two series are plotted contemporaneously in Figure 2. Note however that there is no lag in outcomes for which the patterns would match.

[9]Panel B of Table A.VI shows that there do not appear to be similarly large differences between the fraction of students who expect to pursue each major and the fraction of students who actually attain such majors, calculated from the American Community Survey (ACS) (Table A.VII shows how we categorize majors from the ACS into our 10 major groups). Thus, differences between expected and actual careers are unlikely to be driven by systematic biases in the majors with which students expect to graduate.

a long-standing pattern in students' expectations.

2.2 Isolating Stereotyping

The patterns described in Section 2.1, while consistent with stereotyping, by themselves could reflect several other potential mechanisms. In this section, we describe these alternative explanations and our primary surveys, which we designed to isolate the role of stereotyping. To administer these surveys, we partnered with the "Exploration" program at the Ohio State University (OSU). Entering OSU students are automatically enrolled in this program if they have not yet officially declared a major. Although these students may be selected relative to college students in general, we show later that our results are externally valid to other samples. The Exploration program includes a mini-course that enrolled students take their first semester at Ohio State. This course is typically taught by students' academic advisor and includes self-assessments (meant to point students toward majors that suit them), information about degree requirements, and help planning course schedules. At the time we ran our surveys, the program did not provide students quantitative information about employment outcomes by major. Students received extra credit in this course for completing our survey, ensuring a high response rate of around 80%.

We ran three surveys among students in the Exploration program. Our main survey occurred during Fall semester of 2020. The other two surveys both occurred during Fall semester of 2021 among a new cohort of Exploration students. The first 2021 survey served as a replication of our 2020 results, while the second included an information-provision experiment to evaluate the effect of reducing stereotyping on students' intentions and choices (see Section 3.2). See Appendix B for more details about the surveys and their implementation.

Columns 2 and 3 of Table A.I give self-reported demographic information about the students in our OSU samples, which are broadly similar demographically to the overall student body at OSU, though with a somewhat higher share of first-generation college students. They are also similar to the Freshman Survey along gender, ethnicity, first-generation status, and self-reported family income.

The 2020 survey began by displaying the ten groups of college majors (henceforth, just "majors") and asking students to rank them by how likely they thought they were to graduate from OSU with a degree in each. It then asked them detailed questions about a subset of these majors. In the next paragraphs, we discuss how the design of this survey isolates stereotyping from alternative mechanisms that might underlie students' beliefs.

Qualitative vs Quantitative Expectations

One difficulty in interpreting the Freshman Survey stems from the fact that it asks student to mark one job as their "probable career occupation." One might reasonably worry that this way of eliciting expectations makes it difficult to interpret the patterns we have documented as biased beliefs. For example, if students tend to mark an occupation as their "probable" career when in reality they think they only have a relatively small chance of working in that job, then we could be overstating the extent of bias in students' true beliefs. To avoid this issue, we ask students in the OSU sample quantitative probabilistic questions about a well-defined event, allowing them to express uncertainty precisely. Specifically, for each student's top-ranked major, we asked: "Imagine that you successfully graduate from OSU with a major in X. What is your best guess about the percent chance that, when you are 30 years old, you would be..." It then listed the nine careers in a random order, plus "working in any other job" and "not working for pay." We required students' answers to sum to 100%.

The light blue bars in Figure 1 show the average answer that students who ranked each major highest gave about their likelihood of working in that major's most distinctive career. We see a striking pattern: OSU students in every major believe that they have a higher chance of working in that major's most distinctive career than the true fraction who in fact work in that career. For every major, average beliefs in the OSU sample are very close to the fraction of students in the Freshman Survey who said they would "probably" have that career. Note that these similarities between the OSU and Freshman Survey samples appear not only despite the difference in elicitation method (qualitative vs quantitative expectations) but also despite differences in time period (1970s-2010s vs 2020) and sample (students around the country vs only Ohio State). We take these results as evidence both that the qualitative nature of the questions in the Freshman Survey does not explain the patterns documented above and of the external validity of our OSU sample.

Ruling Out Overconfidence, Selection, and Motivated Beliefs

So far, we have focused on students' beliefs about their own future career. However, if students believe their outcomes will be systematically different from population outcomes—for example, due to overconfidence—this could lead to exaggerated beliefs about their own likelihood of attaining distinctive careers. We address this issue in the OSU survey by asking students not only about their own future outcomes conditional on major but also about the US population as a whole. More precisely, *before* asking students about their own future jobs, the survey asked them to give their "best guess about the percent of Americans aged 30-50 (note, not just from Exploration or OSU) who graduated with a major in X that are..." It then listed the same 11 outcomes. We call these students' "population beliefs," in contrast to their "self beliefs" about their own outcomes. These questions were designed such that we could compare students' answers to an objective benchmark using the ACS data.

We have also so far restricted attention to students' beliefs about the major they themselves intend to pursue. If students' beliefs are correlated with their intentions about their own future major, then biased beliefs among people pursuing a particular major could reflect this correlation structure rather than a more general underlying feature of students' beliefs. Most obviously, students might *select* into majors on the basis of their beliefs about employment outcomes. For example, students who especially think a journalism major leads to a career in journalism might select into that major, leading to a bias in beliefs *conditional on pursuing journalism* despite no underlying bias in the population at large. Other channels that could lead to such a correlation include wishful thinking or motivated beliefs: students might hold mistaken beliefs in order to *ex post* justify their chosen major, spur them to study harder, or because it makes them feel better about their choices or future outcomes. Another such factor is persuasion: one could imagine that academic departments may try to convince students taking introductory classes that their field's distinctive outcomes are more likely than they are in an attempt to maintain enrollment.

To address such issues, we asked students in the OSU survey about not only their top-ranked major but also about their second-ranked major and two additional majors chosen randomly from the remaining eight. We then use inverse probability weights to estimate average beliefs *unconditional* on major ranking.[10] This approach of estimating average beliefs among the whole group of students, rather than only those considering particular majors, rules out any proposed explanation for biased beliefs that rests on students with different majors holding systematically different beliefs. Throughout the analyses to follow, we employ such weighting whenever we pool beliefs about students' top-ranked majors with beliefs about their lower ranked majors. In practice these weights have little impact on our results,

[10] In particular, a student's two top majors receive a weight of one, while the two other majors receive a weight of four (because there were eight other majors, and thus a one in four chance that each was selected).

indicating a minor role for this alternative set of explanations.

The gray bars in Figure 1 show the 2020 OSU sample's average population belief, including all four majors that each student was asked about, rather than restricting it to their top-ranked major. We see that for nine of the ten majors—all except nursing—students believe that a major's most dinstinctive career is substantially (and statistically significantly) more common among graduates with that major than it actually is.[11] These differences are again quite large and comparable to both the OSU self beliefs and Freshman Survey expectations: students exaggerate the share of artists among art majors by 36 p.p. or 211%, of doctors among biology and chemistry majors by 11 p.p. (48%), of counselors among psychology majors by 22 p.p. (105%), of writers and journalists among communications majors by 43 p.p. (1,075%), and so on. All of these differences are statistically significant at the $p < 0.01$ level. These results suggest that the earlier patterns in self-beliefs were not primarily driven by confidence, selection, or motivated beliefs. They also suggest that students believe the causal effect that majoring in different fields would have on them (expressed by their self beliefs) are quite similar to the (perceived) cross-sectional differences in occupations across majors.

Ruling out Salience and Other Career-Specific Explanations

Next, students could simply tend to overestimate more common careers or more salient careers, unrelated to considerations about majors. For example, if distinctive jobs tend to be common and students simply exaggerate or latch onto more common outcomes by major, this could superficially look like stereotyping. Additionally, the role of stereotyping could be confounded if students underestimate the unconditional likelihood of careers that tend not to be distinctive of most majors. For example, perhaps many non-distinctive occupations are less salient (independent of major), and some students might fail to correct for the fact that there are many jobs they have not heard of or that are less top-of-mind.

We address these issues—and any other explanation that applies only at the individual-by-career-level—with a regression analysis. Table 1 shows OLS estimates of the regression specification in equation 1, where $\pi^i_{c|M}$ is student i's population belief about career c conditional on M, $p_{c|M}$ is the true fraction of those with that major who are working in that

[11] Even the exception to this pattern is instructive. Though students underestimate the share of nursing majors working as nurses, this is in large part because they dramatically overstate the share of such majors who eventually become doctors, which is nursing's second most distinctive outcome. In fact, 4% of nursing majors work as doctors, but the average belief is 23%.

career, and μ_c^i are career-by-individual fixed effects.

$$\pi_{c|M}^i = \gamma p_{c|M} + \theta \mathbb{1}\left(c = \text{argmax } \frac{p_{c,M}}{p_{c,-M}}\right) + \mu_c^i + \epsilon_{c,M}^i \tag{1}$$

The coefficient θ is our measure of stereotyping. Controlling for true frequencies $p_{c|M}$ lets us account for the possibility that students may simply exaggerate more likely careers (which would manifest itself as estimating γ to be larger than one). Controlling for career-by-individual fixed effects allows us to separate stereotyping from biases unrelated to major (e.g., underweighting less salient careers irrespective of major). For example, if students simply neglected non-employment or "other" jobs but otherwise responded only to true frequency, this would be captured by μ_c^i and γ, and our estimate of θ would be zero.

Column 1 of Table 1 shows OLS estimates of equation 1 using the full 2020 OSU sample. We see a large and statistically significant estimate of 0.29 for θ, the coefficient measuring stereotyping. This estimate can be interpreted as saying that the average student's belief about the fraction of graduates with a major's most distinctive career is 29 percentage points higher ($p < 0.01$) than similarly frequent but less distinctive outcomes. Finally, the estimate of 0.51 for γ, the coefficient on $p_{c|M}$, shows that after accounting for stereotyping, students' beliefs are undersensitive to true frequencies.

We can visualize these coefficients by looking at Figure 3, which plots the average population belief for each career-major pair against its true conditional likelihood. Panel A shows that students overestimate nine of the ten distinctive pairs (these points correspond to the gray bars in Figure 1), but within these distinctive outcomes the relationship between beliefs and true frequencies is attenuated (i.e., the line of best fit is flatter than 45 degrees). Panel B shows a similar attenuation between beliefs and true frequencies among non-distinctive outcomes, such that sufficiently rare non-distinctive outcomes are on average overestimated while sufficiently common ones are neglected. These trends explain why the estimate of γ from equation 1 is less than one. The role of stereotyping is also obvious from Figure 3; comparing objectively similarly frequent careers that are distinctive (Panel A) versus not (Panel B), we see that distinctive careers are perceived to be dramatically more common.

Importantly, note that students substantially overestimate business and teaching jobs for business and education majors (respectively) but underestimate these outcomes for almost every other major. This result shows that these biases are not just about careers *per se* (e.g., students simply having not heard of certain professions) but rather about the relationship *between* careers and majors. It also explains why our estimate of θ is so large and significant despite the inclusion of career-fixed effects.

Figure 3 also shows that students greatly underestimate the share of college graduates who are not working for pay. Across majors, students believe 1% to 4% of graduates are not employed. Note that these numbers are lower even than the non-employment rate for *male* college graduates (6.3%). Further, in our 2021 replication (more details below), we updated the wording of these questions to explicity ask about "not working for pay (e.g., unemployed or a full-time parent)", indicating these results are not driven by a failure to consider leaving the workforce for reasons related to child-care.

Stereotyping thus remains a powerful determinant of beliefs even after isolating it from the other potential mechanisms described above. To quantify this importance, Table A.VIII presents the results of a Shapley-Sharrocks decomposition. This analysis effectively conducts a horse-race, comparing the extent to which stereotyping versus other potential mechanisms explain the variance in students' beliefs (see notes of Table A.VIII for details). We find that stereotyping distinctive careers appears to be the dominant explanation for the patterns of expectations in our survey data, accounting for a greater share of the variation in beliefs than any of the other mechanisms we explore.

Is Stereotyping Unique to Undecided Freshman?

One might wonder whether these mistaken beliefs are unique to the sample of Ohio State freshmen we happen to survey. For example, perhaps stereotyping vanishes over time as students learn more about the labor market, or those undecided about their majors upon entering college may be differentially selected. To explore this possibility, we conducted an online survey of US adults, recruited through CloudResearch, where respondents were incentivized to provide accurate responses to the same population-beliefs questions as our OSU survey (see Appendix B for details on this survey). Columns 8-11 of Table 1 shows somewhat smaller but still large and statistically significant stereotyping even among these respondents, both when restricting to college- vs non-college-educated and younger vs older adults. Thus, even people with substantial labor market experience and who have attended college beyond their freshman year exhibit similar patterns of stereotyping as in our main OSU sample. Note that these results also rule out inattention stemming from lack of incentives being the driving factor behind stereotyping in our main surveys; the same patterns appear when respondents are paid for accuracy.

Next, perhaps stereotyping is primarily prevalent only in a demographic group that happens to be overrepresented in our Ohio State sample. To test this, columns 2-7 of Table 1 show a similarly large magnitude of stereotyping across a range of demographic cuts: male vs female students, underrepresented-minority vs non-minority students, and first-generation vs non-first-generation students. In no case do we find significant differences in the level of stereotyping across these groups ($p > 0.10$ for all comparisons). Thus stereotyping appears to be widespread across demographic groups.

Figure 1 suggests a close parallel between mistaken beliefs in the Ohio State surveys and the patterns of career expectations in the nationally representative Freshman Survey. To formalize that comparison, we ask whether the differences between actual and expected careers in the qualitative Freshman Survey (shown in Figure 2) are consistent with being driven by biases in population beliefs of the same magnitude as in our Ohio State sample. To do so, we conduct the following back-of-the-envelope calculation. Let $\pi_{c|M}$ be the average OSU population belief about the fraction of people with major M who are working in career c. Let κ_M be the fraction of the Freshman Survey respondents that say they expect to graduate with major M.[12] We calculate what we call the "implied error" about the probability of working in career c as shown in equation 2, where p_c is the true fraction working in c.

$$\text{ImpliedError}_c = \sum_M \kappa_M \pi^i_{c|M} - p_c \tag{2}$$

We compare these implied errors to a corresponding notion of "error" from the Freshman Survey data: the difference between the fraction of students who expect to have a career in each occupation minus the true proportion of college graduates with that occupation. Intuitively, this analysis asks whether the Freshman Survey respondents' expectations match what we would expect if they held the same (population) beliefs as the OSU sample. As Figure A.I shows, there is a robust positive relationship between actual "errors" in the Freshman Survey data and ImpliedError_c. The correlation between between implied and actual error is 0.81 and is highly statistically significant ($p < 0.01$). An OLS regression of the error in the Freshman Survey data on the implied error yields an R^2 of 0.71 with a coefficient of 0.84 ($p < 0.01$), which is not statistically distinguishable from one ($p = 0.39$). We conclude

[12] For this analysis, we drop students who list their probable major as "Undecided."

from this exercise that the pattern of overestimation of careers in the Freshman Survey is quite close to what we would expect if they were driven by the same errors as in the OSU students' beliefs.

Finally, the first 2021 OSU survey served as a replication exercise for the results from the 2020 sample. For brevity, this survey asked students only about their two top-ranked majors, but otherwise asked the same self- and population beliefs about career likelihoods as the 2020 survey did. Table A.IX shows very similar average population beliefs in this sample compared to the 2020 sample; in particular, students again exaggerate the share of graduates working in distinctive jobs for nine of the ten majors (all but nursing).

3 Implications of Stereotyping

Having documented stereotypical thinking in how students perceive the link between college majors and occupations, we next turn to the economic consequences of these beliefs. We use two complementary approaches: a structural model, which allows us to estimate student preferences and simulate choice behavior with counterfactual beliefs, and a field experiment, which while limiting our ability to perfectly control beliefs provides a policy-relevant counterfactual with fewer assumptions.

3.1 A Stylized Model of Major Choice

Set-Up We begin by describing and estimating a stylized model of major choice. Assume that student i is choosing their major $M \in \{A, B, ...\}$. If they choose M, the probability that they will have career $c \in \{a, b, ...\}$ is $p^i_{c|M}$. Their *belief* about this probability is $\pi^i_{c|M}$. We assume their perceived expected utility (i.e., given their potentially incorrect beliefs) from choosing M is then given by equation 3:

$$\widehat{EU}^i[M] = \sum_c \pi^i_{c|M} \left(\alpha w^i_{c,M} + \beta^i_c \right) + \mu^i_M + \nu^i_M \tag{3}$$

In equation 3, $w^i_{c,M}$ is the salary i believes they would earn conditional on c and M, and α is the (homogeneous and constant) marginal utility of income. We allow for i to have idiosyncratic non-pecuniary preferences over jobs, which are denoted by β^i_c. Next, μ^i_M

indicates the known non-labor-market benefits the student would derive from majoring in M (enjoyment of classes, parental approval, etc.). Finally, ν_M^i indicates an unrealized preference shock, whose distribution i knows but whose realized value they do not.

We make a series of simplifying assumptions to facilitate estimating the model. First, we assume ν_M^i is a type 1 extreme value random variable that is i.i.d. across majors and students. We also assume that the student has a non-monetary preference for working in one career, which we denote by $c^*(i)$: that is, $\beta_c^i = \beta\mathbb{1}(c = c^*(i))$. From these assumptions, equation 4 follows, where π_M^i is i's belief about the probability they will graduate with M:

$$\log \frac{\pi_M^i}{\pi_{M'}^i} = \alpha \sum_c \left(\pi_{c|M}^i w_{c,M}^i - \pi_{c|M'}^i w_{c,M'}^i\right) + \beta\left(\pi_{c^*(i)|M}^i - \pi_{c^*(i)|M'}^i\right) + \mu_M^i - \mu_{M'}^i \qquad (4)$$

Estimation To estimate the model, we use data collected in our 2020 OSU survey. For four majors, we directly elicited $\pi_{c|M}^i$, each student's self beliefs about their likelihood of working in c conditional on majoring in M. We also elicited students' beliefs about their own expected salary at age 30 for the same four majors (see Appendix B for details), which we use as a proxy for $\sum_c \pi_{c|M}^i w_{c,M}^i$. We asked students the percent chance they thought they would graduate from OSU with each of the four majors, which we employ as our measure of π_M^i. Finally, we assume μ_M^i is normally distributed and i.i.d. with mean μ_M and variance σ^2.

To summarize, the parameters to be estimated are α (salary preferences), β (non-pecuniary preference for favorite careers), $c^*(i)$ (each student's favorite career), μ_M (the mean non-labor market preference for each major), and σ (the variance of non-labor-market preferences for majors). We collect these parameters into a vector that we denote by $\xi = (\alpha, \beta, c^*, \mu, \sigma)$.

The survey asked students questions about four majors, meaning that for each student we have data on three independent pairs of majors: M_1 vs M_2, M_2 vs M_3, and M_3 vs M_4. Let π_M be the student's reported probability of graduating with a major in M, and $\widehat{\pi_M}$ be the model's prediction, given ξ, of that probability. Note that this is a random variable given the distribution of non-labor-market preferences. Then, let $L_{i,j}(\xi)$ be the likelihood given the model and parameters ξ that $\log(\widehat{\pi}_{M_j}/\widehat{\pi}_{M_4}) = \log(\pi_{M_j}/\pi_{M_4})$.

The maximum likelihood estimate for ξ is then given by equation 5:

$$\hat{\xi} = \underset{\xi}{\operatorname{argmax}} \sum_{i} \sum_{j=1}^{3} \log(L_{i,j}(\xi)) \tag{5}$$

We then construct confidence intervals and standard errors using a Bayesian bootstrap, clustered at the individual level.

Results Our structural model yields two main results. First, we show that students have strong preferences about the specific job they will hold—above and beyond its salary—when choosing college majors. Second, we simulate decisions under alternative beliefs about the mapping between field of study and occupation and show that eliminating stereotyping has large effects on students' choices.

To examine the premium placed by students on their future occupations, Column 1 of Table A.X shows estimates from this baseline model. We see a positive coefficient of 0.066 ($p < 0.01$) for α, students' preferences for expected salary. To facilitate interpretation, consider a student who believes there is a 50% chance each that they will major in A and in B (i.e., they are only considering those two majors but are indifferent between them). Our estimate of α implies that if the expected salary of major A increased by \$10,000, they would only increase their perceived probability of majoring in A by 1.6 percentage points. This result is reminiscent of previous work that finds a surprisingly small elasiticity of major choice with respect to earnings using both survey and observational evidence (e.g., Arcidiacono 2004, Beffy et al. 2012, Wiswall & Zafar 2015a, and Long et al. 2015).

In contrast, column 1 of Table A.X shows substantial non-monetary preferences for working in preferred careers. Returning to our hypothetical student who is on the fence between majors A and B, the estimate of 4.56 ($p < 0.01$) for β implies that increasing the chance that i could work in their preferred career by 10 percentage points if they majored in A would increase their chance of graduating with that major from 50% to 61.2%. This change—more than six times larger than that of increasing salaries by \$10,000 a year—implies a very large willingness-to-pay to work in preferred careers: our estimates suggest a student would give up almost \$6,900 a year in expectation (95% confidence interval = $[\$4,500, \$18,500]$) to increase their chances of working in their preferred career by one percentage point. Note however that this extremely large WTP is driven by the low estimates for salary preferences rather than by preferences for careers being unrealistically large. To make this claim more precise, note that the variance of non-career preferences for majors (μ^i_M) is estimated to be 1.04. This implies that a one standard deviation increase in non-career preferences toward a major is equivalent to increasing i's belief about their chances of attaining their preferred career by 23 p.p., which is about 70% of the standard deviation in self-beliefs about the distinctive career of students' top-ranked major. In this sense, non-career preferences and non-monetary career preferences are roughly comparable in magnitude.

Our estimates suggest not only that students have strong non-monetary preferences for careers, but also that these preferences are (endogenously) focused in particular on distinctive careers. Over half (50.5%) of students' estimated preferred careers are the distinctive job of the major they had initially ranked highest. In contrast, only 10.1% percent of students' preferred careers are the distinctive job of their second-ranked major.

The remaining columns of Table A.X show estimates of modifications to this baseline model. Columns 2 and 3 show that these results are not sensitive to how the beliefs data are winsorized (i.e., how 0's and 1's are treated). The model in column 4 does not use students' beliefs about their expected salary but instead the average actual realized earnings by career and major from the ACS (column 4 of Table A.X). Column 5 allows for homogeneous non-pecuniary preferences for each career (retaining the individual-specific additional preference for a single career). Column 6 allows students' expected GPA in each major to influence their perceived utility from pursuing it (where unsurprisingly we find students prefer majors they will succeed in). None of these modifications makes a qualitative difference in our estimates of students pecuniary or non-pecuniary preferences over careers: for example, the WTP to increase a student's chance of working in their most preferred career by 1p.p. is never estimated to be below $3,000 per year.

Equipped with our baseline estimates, we can examine student behavior under alternative beliefs via simulation. We replace respondents' reported beliefs with accurate population beliefs (about both career shares and salary) from the Census. Relative to the field experiment in the following section, this in effect analyzes a scenario in which we completely shift students' perceptions about all majors towards the information we provide in the experiment. Doing so reduces the likelihood of students' majoring in their ex ante top-ranked majors by 15.9 p.p. and increases their likelihood of pursuing lower-ranked majors. This result is driven by students' preferences over careers: because the distinctive job of students' top-ranked major (endogenously) tends to be their preferred career, reducing beliefs about this outcome pushes students away from this major. In contrast, reducing stereotyping regarding less preferred majors if anything makes those majors look *more* attractive, because those majors' distinctive jobs tend not to be students' preferred career. See Appendix B for more details on this counterfactual analysis.

These analyses suggest that students perceive their eventual career to be quite important, over and above the income it may generate, when deciding what to study in college. We conclude this subsection by presenting two corroborating pieces of evidence for this result. First, in the second 2021 survey, we asked the control group of the experiment (described below) to rate on a scale from 0 to 100 how important various factors were to them as they decided what to major in. The "effect on what job I will get after college" received the highest average ranking (83 out of 100), above enjoyability of classes (70), difficulty of classes (58), and the opinions/choices of family, friends, and peers (all below 30).

Second, in Appendix B we analyze data from the 2013 National Survey of College Graduates, which asks employed respondents about the salary of their job, whether it is related to their highest degree (which we use as a proxy for its distinctiveness), and their overall satisfaction with this job. We regress this satisfaction variable on salary, degree-relatedness, and demographic controls. We find that having a job unrelated to one's highest degree (conditional on salary) is associated with an 11.2 percentage point (93.9%, $p < 0.01$) increase in the likelihood of being dissatisfied with that job. In contrast, a \$10,000 increase in salary is only associated with a 0.4 percentage points decrease ($p < 0.01$) in likelihood of job dissatisfaction. Thus, at least in the cross-section, the relationship between salary, degree-occupation match, and satisfaction qualitatively mirrors students' revealed preferences as estimated by our choice model above. Furthermore, 38% of graduates whose job does not match their major report that this is because no such job is available to them, and such graduates are 80% more likely to be dissatisfied with their current job ($p < 0.01$). Finally, of graduates who work in an occupation outside their degree field, only 17.1% say the reason they do so is because of a change in career or professional interests. In contrast, 59.1% list factors related to aptitude (e.g., promotion opportunities, whether a job is available) or worse-than-expected job amenities (e.g., pay, location, working conditions). These results suggest that much of the gap between expected and actual careers relates to information frictions, rather than changing preferences.

3.2 A Field Experiment Providing Information

The analysis in the previous section suggested that beliefs about distinctive jobs—and therefore stereotyping—may have a large effect on what students choose to study. We now turn to complementary evidence from a field experiment in which we tested a low-cost, light-touch information intervention embedded in the second 2021 OSU survey. The survey began by asking students the percent chance that they would graduate with the two majors they selected as being most likely to pursue (henceforth, their "top-ranked" and "second-ranked" majors). It then asked their self and population beliefs about the likelihood of each career group conditional on these two majors. Students were then randomly sorted into a control group and a treatment group. Those in the control arm answered questions about their classes so far that semester and how they had (or had not) contributed to their major and career plans. These questions were designed to be similar in overall length and broadly about the same topic as the information module in the treatment arm but without providing students any new objective information.

In the treatment arm, information modules provided students with the actual distribution of careers conditional on each of their top two majors according to data from the ACS. For each major, it told them several headline numbers about the frequency of the careers they had listed as their most likely jobs if they graduated with that major.[13] We then provided interactive infographics depicting the share of graduates with each major that were working in each career group (plus "other" and non-employed). A further graphic broke down these groups into more detailed occupation titles. After showing this information for each major, we re-asked students how likely they thought they would be to have each job if they graduated with that major. The information module (filled in with fictitious previous answers) can be accessed at this link.

Because the treatment group saw information about the two majors they thought they were most likely to pursue, in practice each student was assigned to either zero (in the control group) or two (in the treatment group) of ten possible information modules. A natural first question is the extent to which each of these treatments succeeded in changing students' beliefs about their own chances of attaining each major's distinctive career. Figure A.II

[13]These headline numbers were always about the two careers they said they would be most likely to be working in if they graduated with that major, plus (if not already included) that major's most distinctive career.

shows the average revision, among those in the treatment group who saw information about each major, in students' beliefs about their own chances of attaining that major's distinctive career. It plots this statistic against the average error in students' population beliefs: i.e., the average belief about the share of Americans with that major working in its distinctive career minus the true share. Two facts stand out. First, while students on average revise in a sensible direction (reducing self-beliefs when they overestimate population outcomes), this updating is far from one-for-one: a regression of average revisions on average errors yields a coefficient of only -0.28. Second, average revisions are *heterogeneous* across majors: for some majors, average revisions are well above or well below the trend predicted by the -0.28 coefficient, and we can reject the null hypothesis that each major's average revision is on this trend line ($p < 0.01$).[14] This result is perhaps not surprising: each module in the treatment group necessarily provided many distinct pieces of information (about each career group as well as more fine-grained occupations) that differed across majors, and we intentionally did not provide students guidance on how this information should impact their self-beliefs.

Our question of interest is whether reducing students' beliefs about their likelihood of attaining distinctive careers influences their intentions and behavior. Following the literature on information interventions, we account for heterogeneity in treatment-induced belief changes (see, e.g., Haaland et al. 2023 and Stantcheva 2023). This is particularly important in our setting because students received differing (and high-dimensional) information, depending on their major preferences, as discussed above. We adopt a data-driven approach to test the causal effect of interest by constructing a simple measure of treatment intensity: the leave-out mean reduction by major in treated students' self-beliefs about their chances of having that major's distinctive career. Note that using the leave-out mean reduction ensures that we are not conditioning on any post-treatment data for student i. We estimate equation 6 by OLS:

$$Y_{i,M} = \beta \cdot T_i \cdot AvgReduction_{i,M} + \alpha_1 Intentions_{i,M,pre} + \alpha_2 Classes_{i,M,pre} + \epsilon_{i,M} \tag{6}$$

In the above equation, $Y_{i,M}$ is an outcome variable of interest for a given major M for student

[14]This p-value comes from regressing revisions in self-beliefs about a major's distinctive career on the average population error for that career as well as major fixed effects. If average errors simply reflected the average error across majors, we would expect these fixed effects to all be zero, and the p-value is for the null hypothesis that the major-fixed effects are jointly zero.

i. T_i is a dummy variable for treatment status. $AvgReduction_{i,M}$ is the measure of treatment intensity described above: the leave-out mean reduction in self-beliefs about the distinctive career for major M. Finally, we include two controls: $Intentions_{i,M,pre}$ is i's pre-treatment belief about their likelihood of graduating with major M, and $Classes_{i,M,pre}$ is the number of courses in M in their (pre-treatment) Fall 2021 class schedule.

Figure 4 shows the coefficient on the interaction term from equation 6 for different outcome variables (Tables A.XI and A.XII show the full estimates). The dark and light blue dots correspond to regressions where the dependent variable involves students' top- and second-ranked majors respectively. The leftmost pair of dots show that reducing stereotyping has significantly different effects on students' intentions (the post-intervention percent chance they assigned to graduating with each major) for their top- and second-ranked majors. Consistent with intuition, reducing stereotyping decreases intentions toward students' top-ranked majors: the coefficient of -1.1 implies that reducing students' self-beliefs about their chance of having their top-ranked major's distinctive job by 10 percentage points (p.p.) decreases intentions toward that major by 0.11 standard deviations (about 3.5 p.p., $p < 0.05$). For students' second-ranked major (light blue), we see if anything the opposite pattern: treatments that reduce beliefs about attaining the distinctive job of students' second ranked major directionally *increase* intentions toward it (coefficient $= 0.66$, $p = 0.17$). This difference between first- and second-ranked majors is statistically significant ($p = 0.01$). The direction of this difference is also what we would expect from Section 3.1, where we estimated that students tend to have strong non-monetary preferences for their first-ranked major's distinctive career. That result would suggest that learning one's top-ranked major's distinctive job is unlikely should count as bad news, whereas learning that one's second-ranked major's distinctive job is unlikely would not (and could even count as good news if students do not particularly value that job).

The next six pairs of dots in Figure 4 show analogous results but where the dependent variable is the number of classes students took in each major every semester between Spring 2022 (immediately post treatment) and Fall 2024. Initially, in the Spring 2022 semester immediately following the intervention, we find very similar effects as with intentions: reducing stereotyping regarding students' top-ranked major by 10 p.p. decreases class-taking in that major by 0.22 standard deviations (about 0.21 classes, $p < 0.05$). Reducing stereotyping about their second-ranked major again has if anything a positive effect, 0.21 standard deviations (about 0.20 classes, $p < 0.10$), with the difference between top- and second-ranked majors significant at the $p < 0.01$ level. Looking at the subsequent semesters, two patterns emerge. First, both treatment effects eventually fade, becoming smaller and statistically insignificant by Fall 2024, three years after treatment. Second, this fading is much more rapid for students' top-ranked major, disappearing entirely by Fall 2022. While caution is warranted in interpreting these coefficients *ex post*, one possibility is that students manage to selectively forget the "bad news" about their top-ranked major more readily than the on-average good news about their second-ranked majors, consistent with "asymmetric updating" or motivated memory (e.g., Möbius et al. 2022, Amelio & Zimmermann 2023, Conlon 2024).

The final five pairs of dots show effects on whether students' have declared a major in their top- and second-ranked fields between Spring of 2022 and Spring of 2024, the last semester for which these data are available. We see no effects on major declarations in the semester immediately after treatment (Spring 2022), which is unsurprising given that only 7% of students have declared a major by then, whereas 45% have done so by the first semester of their sophomore year (Fall 2022). Here we see that reducing stereotyping by 10 p.p. directionally reduces major declarations in students' top-ranked major by 6 p.p. in Fall 2022 ($p = 0.23$) and increases declarations in students' second-ranked major by 10 p.p. ($p < 0.01$, difference significant at $p = 0.01$). These estimates persist through Spring and Fall of 2023, but appear to narrow by Spring 2024, where neither interaction is significant, nor is the difference between them ($p = 0.16$).

The results described above investigate the effect of changing students' beliefs about the likelihood of attaining distinctive jobs by leveraging variation, across information modules, in how much stereotypical beliefs changed in response to the treatment. Tables A.XIII and A.XIV instead regress various outcomes on controls and an indicator for treatment assignment, without interacting treatment status with this measure of treatment intensity. This specification tests not the extent to which changes in students' beliefs about their chances of distinctive careers affects their choices but rather the average effect of being assigned to the treatment group. These average effects are of course generally smaller (and usually not statistically significant), reflecting the fact that our light-touch intervention providing population statistics for some majors often had small (and for some majors null) effects on what jobs students expected themselves to have. However, column 1 of Table A.XV shows that the treatment did have a significant average effect on the timing of students' major decisions:

among students who have declared a major by Spring 2024, those in the treatment group spent an additional 0.21 semesters ($p = 0.01$) undecided before doing so. This appears to be because they are somewhat less likely to have declared a major during their sophomore year but have done so at equal or even somewhat higher rates by the end of their junior year (a 6 p.p. increase, $p < 0.10$). This latter effect appears driven by the fact by those in the treatment group being somewhat less likely to have dropped out by that time (5 p.p. effect, $p < 0.10$), which we measure by whether they are still enrolling in classes (see columns 7-12 of Table A.XV).

4 How Do Students Form Beliefs?

What drives stereotyping in our context? More broadly, how do students form their beliefs about careers and majors? Bordalo et al. (2023) argue that exaggeration of distinctive types arises when agents form beliefs from memory. The crucial mechanism is cued recall, whereby when considering a particular hypothesis, items consistent with that hypothesis disproportionately come to mind. For example, consider a student forming beliefs about what jobs psychology majors go on to have. When they consider the hypothesis that they become counselors, cueing prompts them to think of people with this job, who overwhelmingly tend to have majored in psychology. In contrast, when they consider the hypothesis that they become teachers and cueing prompts them to think of people with that job, many *non*-psychology majors come to mind (e.g., education majors who become teachers). Thus, the student can think of many psychology majors who are counselors but few who are teachers. Beliefs thus increase in distinctiveness, yielding stereotyping. Appendix C formalizes this argument by adapting Bordalo et al. (2023)'s model to our setting.

A basic implication of such a framework is that differences in beliefs across students should systematically correlate with who is likely to be top-of-mind for them (and therefore who are easily retrieved when forming beliefs). To provide some correlational evidence in this direction, we collected data on the careers and majors of adults who are personally close to students and therefore plausibly more likely to come to mind for them. In particular, the first 2021 OSU survey asked students to think of "three people in your life whom you might consider role models. These should be people whom you might turn to for advice about choosing your college major or other aspects of planning for your schooling and eventual career." The survey then asked the student's relationship to this person (84% of students answered about at least one parent, and 50% answered about two), their level of education, college major (if applicable), and occupation. The options for their role models' major and occupation were the same groups of careers and majors that we focus on throughout the paper.[15] All questions about role models were asked after eliciting students' beliefs in order to avoid appearing to suggest that they should base their beliefs on the careers/majors of the people they know personally.

Table 2 shows OLS estimates of equations 7 and 8.

$$\pi^i_{c|M} = \alpha + \beta_1 RM^i_{c,M} + \beta_2 RM^i_{c,-M} + \mu_{c,M} + \epsilon^i_{c,M} \quad (7)$$

$$\pi^i_c = \alpha + \beta RM^i_c + \mu_c + \epsilon^i_c \quad (8)$$

In equation 7, $\pi^i_{c|M}$ is the student's population belief (in columns 1-3) or self belief (columns 5-7) about the likelihood of career c conditional on major M, $RM^i_{c,M}$ indicates the number of role models they listed with c and M, and $RM^i_{c,-M}$ indicates the number with c but a major other than M. Finally, $\mu_{c,M}$ are career-by-major fixed effects, indicating that all estimates are identified off variation across individuals in the career/major of their role models.

The first 2021 OSU survey also elicited students beliefs about share of college graduates working in each career group *unconditional* on major. We use these data to explore whether the careers of students' role models correlate with their beliefs about the unconditional frequency of careers. In particular, in equation 8, π^i_c is the student's unconditional belief about career c, RM^i_c is the number of role models they list with that career, and μ_c are career fixed effects.

Columns 1 and 5 of Table 2 show that, pooling across all majors and careers, knowing someone with a particular career-major pair (c, M) boosts beliefs about the frequency of c conditional on M by 3.1 p.p. ($p < 0.01$) for population beliefs and by 3.7 p.p. ($p < 0.01$) for self beliefs. Column 4 shows that having a role model with a particular career boosts students' population beliefs about the unconditional frequency of that career among college graduates by 1.8 percentage points ($p < 0.01$). Column 8 shows similar—and indeed larger—effects on their self beliefs about their own future career. Column 9 shows analogous estimates

[15] In addition to the ten groups of majors and "other," students could also mark that they "have no idea" what their role model's major was. In practice, we have major data for 93% of college graduate role models, suggesting that students are relatively well informed about their role models' majors.

from the Freshman Survey (which asks the career, but not the major, of students' parents): students are 4.2 p.p. ($p < 0.01$) more likely to list an occupation as their probable career if it is one of their parents' careers.

By themselves, the results on self beliefs are consistent with many mechanisms (e.g., students may intrinsically prefer or have greater access to their role models' careers). In contrast, we view such large "effects" on population beliefs (that is, about the current distribution of careers and majors nationwide) as most naturally interpreted through the lens of what comes to mind: i.e., students think these career paths are especially likely because they can easily think of someone who followed them. Consistent with this interpretation, these effects even appear (and indeed, are larger) when we restrict the conditional beliefs data to students' population beliefs about each major's most distinctive career (column 2). Because students already exaggerate these careers, these role models therefore make students' beliefs *less* accurate, which we would not naturally expect if greater access to people with these careers only entailed better information about them (i.e., how unlikely such careers often are).

So far we have considered the effect on beliefs about $P(c|M)$ of knowing someone with both c and M. However, we can also ask how knowing someone with the correct career c but the "wrong" major M' correlates with beliefs about $P(c|M)$. Column 1 of Table 2 shows that knowing someone with the correct career but the wrong major boosts beliefs by 0.34 percentage points ($p = 0.04$). Columns 2 and 3 show that this positive effect is entirely driven by non-distinctive outcomes; restricting the data to distinctive careers, we see a much larger but negative effect of -4.4 percentage points ($p < 0.01$). What might explain these ambiguous effects? In Appendix C, we show that the simplest model we consider predicts a negative effect of knowing someone with c but not M on beliefs about $P(c|M)$. However, a simple extension of the model adding a role for extrapolation (Kahneman & Tversky, 1981; Gilboa & Schmeidler, 1995; Bordalo et al., 2024) can predict that these negative effects flip to being positive for sufficiently rare or implausible outcomes, as we see in the data.

5 Discussion

Across multiple survey samples, time periods, and elicitation methods, we find that U.S. undergraduate students greatly oversimplify the college-to-career process. Students appear to stereotype majors ("Art majors become artists," "Political science majors become lawyers"), exaggerating the share of college graduates who are working in their major's most distinctive job. We estimate that students have strong preferences over the job in which they end up, suggesting that stereotyping could have large welfare implications. In a field experiment testing a light-touch intervention, we find that information changing these beliefs can have significant effects on students' intentions and later choices. Finally, we find evidence consistent with students forming their career expectations by relying on who comes to mind, producing systematic differences across students in their beliefs.

One natural follow-up question to our results is why stereotyping persists despite the apparently large incentives that students have to make informed decisions about their education. For example, one could imagine students seeking out information (online, from better-informed friends, etc.) to correct their biased initial perceptions, or learning from better informed peers Enke et al. (2022). We find, however (see Appendix B for more details), that the students in our data who are most confident that their beliefs about careers are correct are also the ones who stereotype most. These results, though only suggestive, point toward the possibility that biased students may fail to correct their beliefs because they are confident in their misperceptions.

More speculatively, our results may help to partly explain several striking and perhaps puzzling facts about students' human capital decisions. For example, more American undergraduates are currently pursuing a bachelor's degree in journalism than there are journalists in the entire country. Psychology majors outnumber accounting majors in the United States, and yet there are eight times as many accountants as psychologists. Students take on considerable debt to fund Master's programs with appealing but unlikely associated careers (e.g., film studies).[16] *Ex ante*, of course, rational mechanisms could have fully explained these patterns: e.g., students with correct beliefs might rationally pursue certain career paths which, though very unlikely to pan out, they feel are worth the risk (e.g., journalism or film), or students may realize that certain majors (e.g., psychology) provide a general education not intended for use in any particular sector. Our findings suggest that mistaken beliefs may also contribute to these patterns: certain fields of study may appear especially appealing because

[16]Shares of majors come from the American Community Survey (authors' calculation), and the number of college graduates comes from the National Center for Education Statistics. Counts of occupations come from the Bureau of Labor Statistics' Occupational Outlook Handbook. See Korn & Fuller (2021) for the article on film studies Master's programs.

students believe they lead with exaggerated likelihoods to attractive distinctive jobs. These human capital investments carry substantial monetary and opportunity costs, and therefore it may be beneficial to find ways to help students make better informed decisions or to nudge them toward less risky academic paths.

References

Alesina, A., Miano, A., & Stantcheva, S. (2022). Immigration and Redistribution. *NBER Working Paper 24733*, 5–24.

Alfonsi, L., Namubiru, M., & Sara, S. (2023). Meet Your Future: Experimental Evidence on the Labor Market Effects of Mentors. *Working Paper*.

Altonji, J. G., Kahn, L. B., & Speer, J. D. (2014). Trends in Earnings Differentials across College Majors and the Changing Task Composition of Jobs. *American Economic Review P&P*, *104*(5), 387–393.

Amelio, A., & Zimmermann, F. (2023). Motivated memory in economics-a review. *Games*.

Arcidiacono, P. (2004). Ability sorting and the returns to college major. *Journal of Econometrics*, *121*, 343-375.

Arcidiacono, P., Hotz, V. J., & Kang, S. (2012). Modeling college major choices using elicited measures of expectations and counterfactuals. *Journal of Econometrics*, *166*(1), 3–16.

Arcidiacono, P., Hotz, V. J., Maurel, A., & Romano, T. (2020). Ex Ante Returns and Occupational Choice. *Journal of Political Economy*(March).

Arteaga, C. (2018, 1). The effect of human capital on earnings: Evidence from a reform at colombia's top university. *Journal of Public Economics*, *157*, 212-225.

Augenblick, N., Jack, K., Kaur, S., Masiy, F., & Swanson, N. (2023). Retrieval failures and consumption smoothing: A field experiment on seasonal poverty. *Working paper*.

Baker, R., Bettinger, E., Jacob, B., & Marinescu, I. (2018). The Effect of Labor Market Information on Community College Students' Major Choice. *Economics of Education Review*, *65*, 18–30.

Bandiera, O., Bassi, V., Burgess, R., Rasul, I., Sulaiman, M., & Vitali, A. (2023). The Search for Good Jobs: Evidence from a Six-year Field Experiment in Uganda. *Working Paper*.

Beffy, M., Fougère, D., & Maurel, A. (2012). Choosing the field of study in postsecondary education: Do expected earnings matter? *Review of Economics and Statistics*, *94*(1), 334–347.

Betts, J. R. (1996). What do students know about wages? Evidence from a survey of undergraduates. *Journal of Human Resources*, 27–56.

Bleemer, Z., & Mehta, A. S. (2020). Will Studying Economics Make You Rich? A Regression Discontinuity Analysis of the Returns to College Major. *American Economic Journal: Applied Economics*, 1–26.

Bohren, J. A., Haggag, K., Imas, A., & Pope, D. G. (2023). Inaccurate statistical discrimination: An identification problem. *Review of Economics and Statistics*.

Bordalo, P., Burro, G., Coffman, K., Gennaioli, N., & Shleifer, A. (2024). Imagining the future: Memory, simulation, and beliefs about covid. *Review of Economic Studies*.

Bordalo, P., Coffman, K., Gennaioli, N., & Shleifer, A. (2016). Stereotypes. *Quarterly Journal of Economics*, 1753–1794.

Bordalo, P., Coffman, K., Gennaioli, N., & Shleifer, A. (2019, 3). *Beliefs about gender* (Vol. 109). American Economic Association.

Bordalo, P., Conlon, J. J., Gennaioli, N., Kwon, S. Y., & Shleifer, A. (2023, 12). Memory and probability. *The Quarterly Journal of Economics*, *138*, 265-311.

Bursztyn, L., Cappelen, A. W., Tungodden, B., Voena, A., & Yanagizawa-Drott, D. (2023). *How are gender norms perceived?*

Chan, A. (2022). Discrimination against doctors: A field experiment. *Unpublished manuscript*.

Coffman, K., Collis, M., & Kulkarni, L. (2023). Whether to apply. *Management Science*.

Coffman, K., Exley, C. L., & Niederle, M. (2020a). The role of beliefs in driving gender discrimination. *Management Science*.

Coffman, K., Exley, C. L., & Niederle, M. (2020b). The role of beliefs in driving gender discrimination. *Management Science*.

Conlon, J. J. (2021). Major Malfunction: A Field Experiment Correcting Undergraduates' Beliefs about Salaries. *Journal of Human Resources*.

Conlon, J. J. (2024). Memory rehearsal and belief biases. *Working paper*.

de Koning, B. K., Dur, R., & Fouarge, D. (2022). Correcting beliefs about job opportunities and wages: A field experiment on education choices.

Dizon-Ross, R. (2019). Parents' Beliefs About Their Children's Academic Ability: Implications for Educational Investments. *American Economic Review*(8), 2728-65.

Dominitz, J., & Manski, C. F. (1996). *Eliciting student expectations of the returns to schooling* (Vol. 31). Winter.

Enke, B., Schwerter, F., & Zimmerman, F. (2021). *Associative memory and belief formation.*

Enke, B., Thomas, G., & Oprea, R. (2022). Confidence, self-selection and bias in the aggregate. *Working paper.*.

Ersoy, F., & Speer, J. D. (2022). Opening the black box of college major choice: Evidence from an information intervention. *Working paper*.

Esponda, I., Oprea, R., & Yuksel, S. (2023, 11). Seeing what is representative. *Quarterly Journal of Economics*, *138*, 2607-2657.

Flood, S., King, M., Rodgers, R., Ruggles, S., Warren, R. J., & Westberry, M. (2021). *Integrated Public Use Microdata Series, Curren Population Survey: Version 9.0 [dataset]* (Tech. Rep.). Minneapolis, MN: IPUMS.

Gilboa, I., & Schmeidler, D. (1995). Case-based decision theory. *The Quarterly Journal of Economics*.

Giustinelli, P. (2022). Expectations in education. *Working paper.*.

Graeber, T., Roth, C., & Zimmermann, F. (2022). *Stories, statistics, and memory.*

Haaland, I., & Roth, C. (2023). Beliefs about racial discrimination and support for pro-black policies. *Review of Economics and Statistics*, *105*(1), 40–53.

Haaland, I., Roth, C., & Wohlfart, J. (2023). Designing information provision experiments. *Journal of economic literature*, *61*(1), 3–40.

Hastings, J. S., Neilson, C. A., & Zimmerman, S. D. (2013, jul). Are Some Degrees Worth More Than Others? Evidence from College Admission Cutoffs in Chile. *National Bureau of Economic Research*(w19241).

Hilton, J. L., & Von Hippel, W. (1996). Stereotypes. *Annual review of psychology*, *47*(1), 237–271.

Jensen, R. (2010). The (perceived) returns to education and the demand for schooling. *The Quarterly Journal of Economics*, *125*(2), 515–548.

Judd, C. M., & Park, B. (1993). Definition and assessment of accuracy in social stereotypes. *Psychological review*, *100*(1), 109.

Kahneman, D., & Tversky, A. (1981). The simulation heuristic. *Working paper.1*.

Kirkeboen, L. J., Leuven, E., & Mogstad, M. (2016). Field of study, earnings, and self-selection. *Quarterly Journal of Economics*, *131*(3), 1057–1111.

Korn, M., & Fuller, A. (2021). 'financially hobbled for life': The elite master's degrees that don't pay off. *The Wall Street Journal*.

Long, M. C., Goldhaber, D., & Huntington-Klein, N. (2015). Do completed college majors respond to changes in wages? *Economics of Education Review*, *49*, 1–14.

Malmendier, U., & Wachter, J. A. (2022). Memory of past experiences and economic decisions. *Handbook of Human Memory*.

Möbius, M. M., Niederle, M., & Niehaus, P. (2022). Managing self-confidence. *Management Science*.

Ruggles, S., Flood, S., Goeken, R., Schouweiler, M., & Sobek, M. (2022). *Integrated Public Use Microdata Series, USA: Version 12.0 [dataset]* (Tech. Rep.). Minneapolis, MN: IPUMS.

Schneider, D. J. (2005). *The psychology of stereotyping.* Guilford Press.

Schneider, D. J., Hastorf, A. H., & Ellsworth, P. (1979). *Person perception.* Addison-Wesley.

Sequeira, S., Spinnewijn, J., & Xu, G. (2016). Rewarding schooling success and perceived returns to education: Evidence from india. *Journal of Economic Behavior & Organization*, *131*, 373–392.

Shorrocks, A. (1982). Inequality decomposition by factor components. *Econometrica*, *50*(1), 193-211.

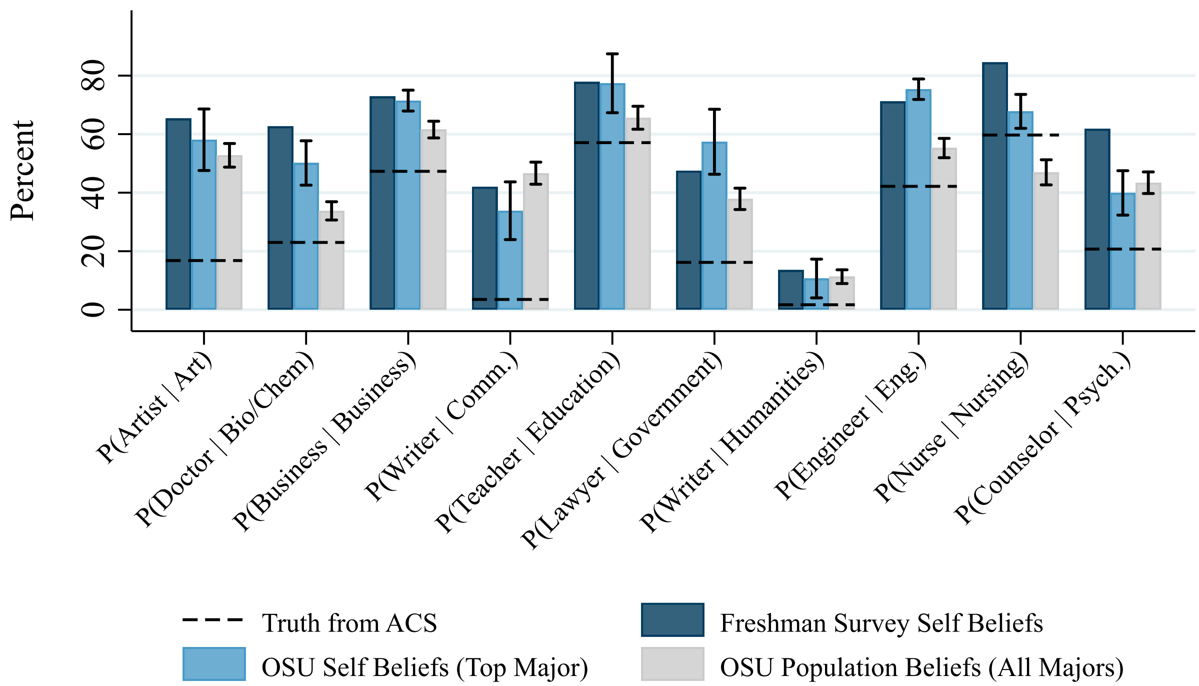
Stantcheva, S. (2023). How to run surveys: A guide to creating your own identifying variation and revealing the invisible. *Annual Review of Economics*, *15*(1), 205–234.

Wiswall, M., & Zafar, B. (2015a). Determinants of college major choice: Identification using an information experiment. *The Review of Economic Studies*, *82*(2), 791–824.

Wiswall, M., & Zafar, B. (2015b). How Do College Students Respond to Public Information about Earnings? *Journal of Human Capital*, *9*(2), 117–169.

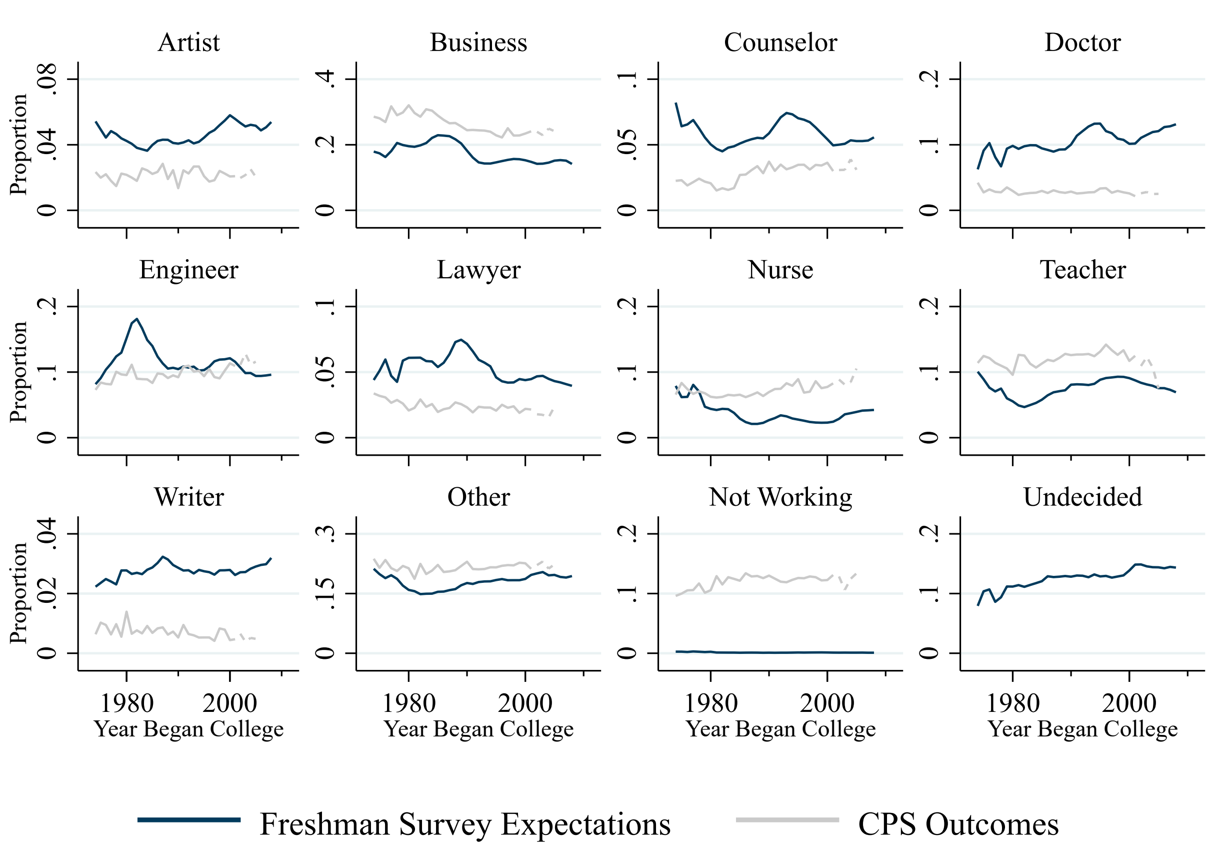
Wiswall, M., & Zafar, B. (2021). Human capital investments and expectations about career and family. *Journal of Political Economy*, *129*(5).

Figure 1: Stereotyping Distinctive Careers



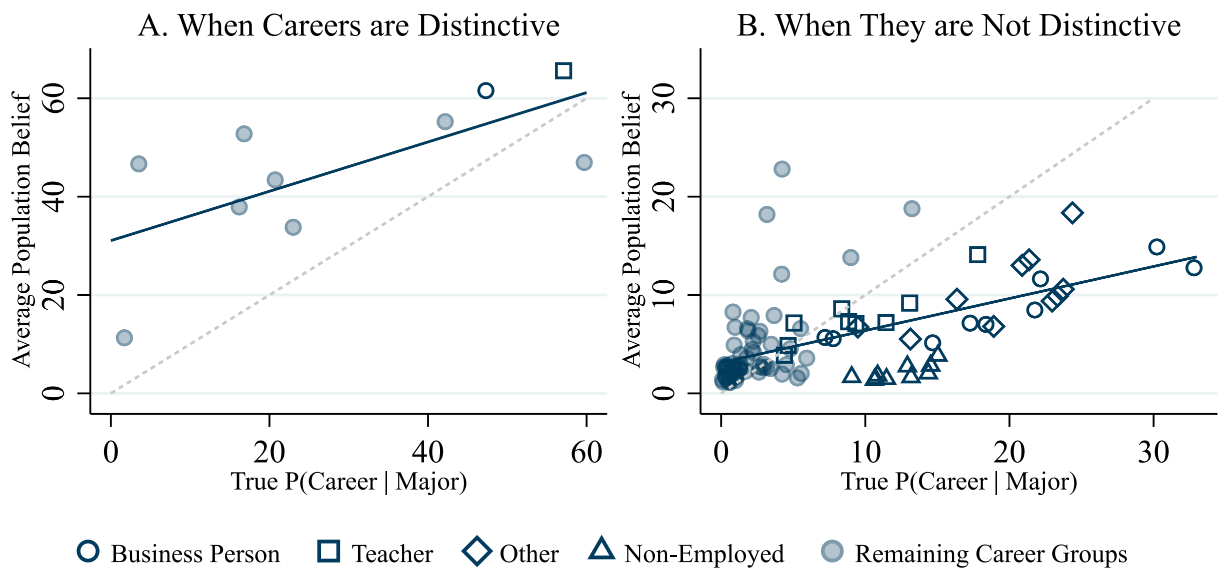
Notes: Figure 1 presents average statistics regarding the most distinctive career (as defined in section 2) for each major. The dashed horizontal lines denote the actual proportion of college graduates with each major between the ages of 30 and 50 that are working in that major's most distinctive career, based on data from the 2017-2019 American Community Survey. The dark blue bar shows, among students in the Freshman Survey who expect to pursue each major, what fraction list that major's distinctive outcome as their probable career occupation. The light blue bar plots the average belief for the 2020 OSU sample about the probability that they would be working in each career at age 30 if they graduated from Ohio State with their top-ranked (i.e., most likely) major. The gray bars show the average belief among our 2020 OSU sample about the fraction of Americans between the ages of 30 and 50 who graduated college with each major (not only their top-ranked major) that are working in each occupation. Error bars show 95% confidence intervals for the mean of the OSU beliefs.

Figure 2: Career Expectations vs. Outcomes Over Time



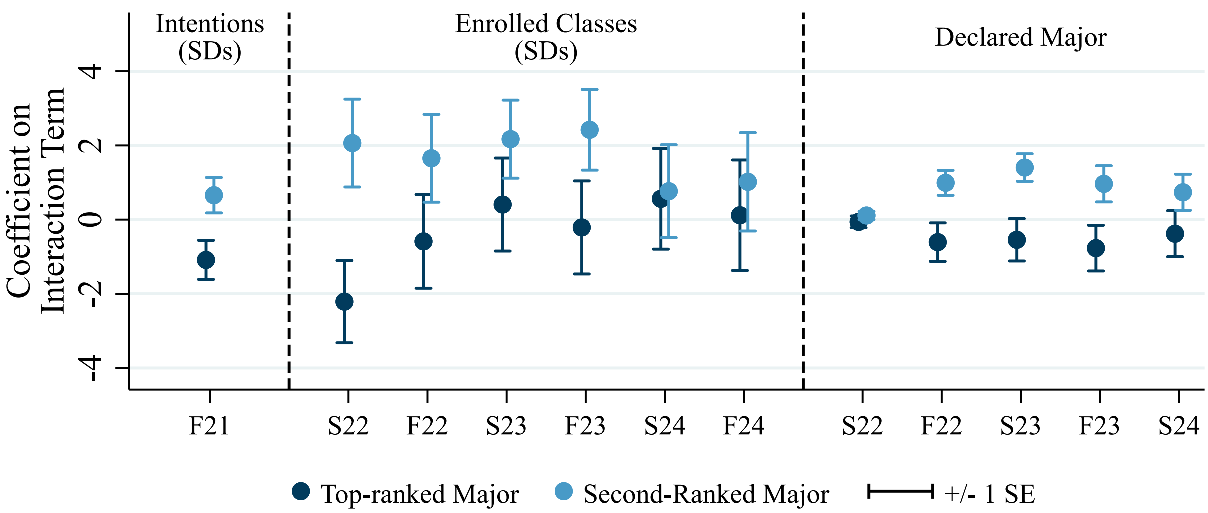
Notes: Figure 2 compares the share of first-year undergraduates by the year they began college (birth year plus 18) who expect to have a career in each occupation (blue line) according to the Freshman Survey data along with the share of college graduates (gray line) aged 33 to 37 who work in that occupation in the same cohort, according to the Current Population Survey. The gray line becomes dashed when CPS outcomes begin to only include graduates younger than 37.

Figure 3: Beliefs about Careers Conditional on Majors



Notes: Each dot in Figure 3 represents a career-major pair (where non-employment is also one the "careers"). Panel A restricts these pairs to when the career is most distinctive of the major, and Panel B restricts them to when the career is not distinctive of the major. The x-axis of both panels is the share of graduates with that major who are working in that career in the ACS. The y-axis is the average population belief, among the 2020 OSU sample, about the fraction of graduates with that major who are working in that career. Lines show OLS regressions including all career-major pairs within each panel.

Figure 4: Effect of Information Intervention



Notes: Each dot in Figure 4 shows the estimate for the interaction term in equation 6 in a separate OLS regression. Dark blue dots indicate regressions where M in equation 6 is participants' top-ranked major, while for light blue dots it is their second-ranked major. The leftmost dots indicate regressions where the dependent variable is students' post-intervention updated belief about their likelihood of graduating with M. The middle dots indicate regressions where the dependent variable is the number of classes in M that students took during Spring semester 2022 (S22), Fall 2022 (F22), etc. The dots on the right indicate similar regressions where the dependent variable is an indicator for whether students' had declared a major in M during those semesters.

Table 1: Testing for Stereotyping

	2020 Ohio State Sample							Online Sample of US Adults			
	All (1)	Men (2)	Women (3)	Non-URM (4)	URM (5)	FG (6)	Non-FG (7)	Non-College (8)	College (9)	Younger (10)	Older (11)
P(Career | Major)	0.51***	0.45***	0.56***	0.46***	0.52***	0.47***	0.53***	0.49***	0.48***	0.46***	0.51***
	(0.12)	(0.14)	(0.11)	(0.11)	(0.12)	(0.13)	(0.12)	(0.08)	(0.07)	(0.08)	(0.08)
$\mathbb{1}$(Most Distinctive)	0.29***	0.29***	0.29***	0.23***	0.30***	0.26***	0.30***	0.23***	0.20***	0.19***	0.23***
	(0.04)	(0.05)	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
Constant	0.02**	0.02**	0.01*	0.03***	0.02*	0.02**	0.02*	0.03***	0.03***	0.03***	0.02***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Observations	33,220	16,016	17,204	6,908	26,312	11,660	21,560	4,664	6,072	4,928	5,808
Individuals	755	364	391	157	598	265	490	212	276	224	264
R^2	0.67	0.64	0.70	0.61	0.69	0.63	0.69	0.76	0.76	0.75	0.77

Notes: Table 1 presents OLS estimates of equation 1 using the 2020 OSU sample (columns 1-7) and a sample of US adults recruited online through CloudResearch (columns 8-119). The dependent variable is respondents' population beliefs about the fraction of graduates with each major working in each career. All regressions include all majors that respondents were asked about and, for each of these majors, all eleven careers (where non-employment is one of the "careers"). "P(Career | Major)" is the true fraction of graduates with a major that are working in that career, calculated from the 2017-2019 American Community Survey. $\mathbb{1}$(Most Distinctive) is a dummy variable indicating whether an occupation is the most distinctive outcome for a major. All regressions cluster standard errors at the individual level and at the career-by-major level. Column 1 includes all the 2020 OSU sample. Columns 2 and 3 split the 2020 OSU sample by gender, columns 4 and 5 by underrepresented-minority status, and columns 6 and 7 by first-generation student status. Columns 8 and 9 split the CloudResearch sample by whether the respondent has a four-year college degree, and columns 10 and 11 split the same sample by whether the respondent is above median age (39 years). *, **, and *** indicate significance at the 10%, 5%, and 1% levels.

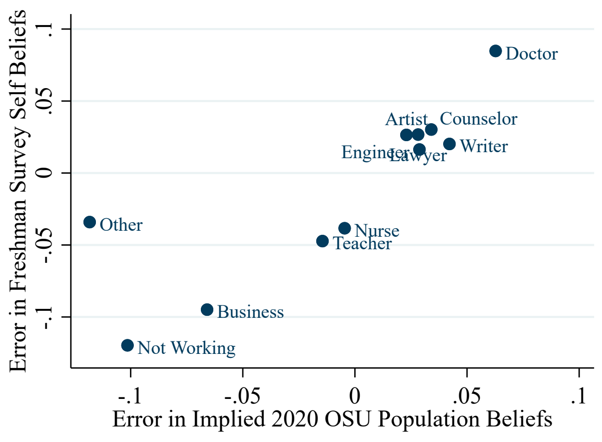
Table 2: Role Models and What Comes to Mind

	Population Beliefs				Self Beliefs				
	P(c | M)			P(c)	P(c | M)			P(c)	
	All	D	ND		All	D	ND	OSU	TFS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$RM_{c,M}$	3.11***	6.26***	1.16**		3.73***	7.68***	1.26**		
	(0.59)	(1.21)	(0.45)		(0.68)	(1.32)	(0.61)		
$RM_{c,-M}$	0.34**	-4.44***	0.63***		1.62***	-3.95**	1.97***		
	(0.16)	(1.57)	(0.15)		(0.26)	(1.87)	(0.25)		
RM_c				1.78***				7.43***	4.19***
				(0.30)				(0.62)	(0.01)
Constant	8.85***	46.86***	5.01***	8.80***	8.53***	50.66***	4.27***	7.87***	7.64***
	(0.05)	(0.78)	(0.08)	(0.05)	(0.07)	(0.85)	(0.09)	(0.11)	(0.00)
Individuals	894	894	894	894	894	894	894	894	8,979,362
Career-by-Major Fixed Effects	Yes	Yes	Yes	No	Yes	Yes	Yes	No	No
Career Fixed Effects	No	No	No	Yes	No	No	No	Yes	Yes

Notes: Table 2 presents OLS estimates of equation 7 (columns 1-3 and 5-7) and equation 8 (columns 4, 8, and 9). The dependent variable in columns 1 to 3 are the population beliefs of students in the 2021 OSU data of the fraction of college graduates working in each occupation conditional on each major. The dependent variable in columns 5 to 7 are the corresponding self beliefs: i.e., students' beliefs about their own chance of working in each career if they gradauted with each major. Columns 2 and 6 restrict the sample to career-major pairs in which the career is that major's most distinctive career (D). Columns 3 and 7 restrict the sample to all career-major pairs where the career is not the most distinctive (ND) of the major. The dependent variable in column 4 is population belief in the 2021 OSU data about the fraction of college graduates working in each occupation unconditional on major. The dependent variable in column 8 is the corresponding self belief: i.e., students' beliefs about their own chance of working in each career (not conditioning on their major). The dependent variable in column 9 is whether a student in the Freshman Survey listed each career as their probable career occupation. $RM_{c,M}$ is the number of role models that the student listed who have that career c and that major M. $RM_{c,-M}$ is the number of role models that the student listed who have career c but do not have major M. All regressions cluster standard errors at the individual level. *, **, and *** indicate significance at the 10%, 5%, and 1% levels.

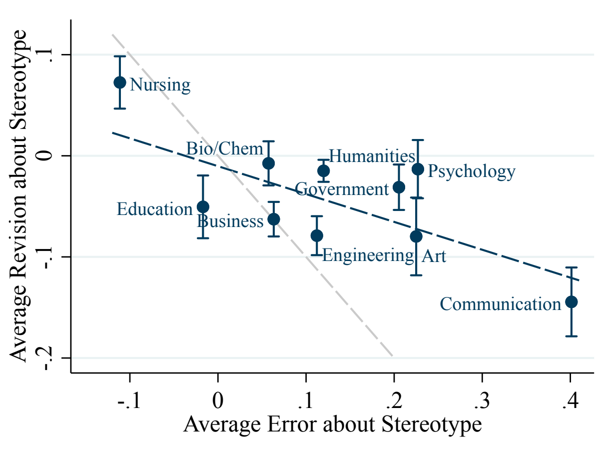
A Online Appendix: Supplementary Figures and Tables

Figure A.I: OSU Beliefs Predict Aggregate Biases in the Freshman Survey



Notes: The y-axis in Figure A.I is the difference between the fraction of students in the Freshman Survey who list each occupation as their probable career and the fraction of 33-37 year old college graduates in the CPS (of the same cohorts, up to birth year 1987) who are working in each occupation. The x-axis is the difference between the 2020 OSU students' "implied" beliefs about the frequency of each career and the true frequency. To construct these implied beliefs, we first take the average population belief of the fraction working in each occupation conditional on each major. We then take a weighted average of these values, where the weights are the fraction of students in the Freshman Survey who expect to pursue each major. See Section 2.2 for further details on the construction of this statistic.

Figure A.II: Revision in Self Beliefs after Information Intervention



Notes: The x-axis of Figure A.II is the true share of college graduates working in the most distinctive career of each major minus the 2021 OSU sample's average beliefs about that share. The y-axis is the revision in treatment-group students' beliefs about their own chance of working in the most distinctive career conditional on graduating with each major, from before to after the information intervention. The blue line shows an OLS line of best fit, while the gray line shows the line y=-x.

Table A.I: Summary Statistics

	Freshman Survey	Ohio State 2020	Ohio State 2021
Female (%)	54.0	51.8	52.9
Non-Hispanic White (%)	72.0	74.2	74.5
First Generation (%)	36.8	35.1	33.0
Mean Family Income ($1,000s)	124.5	118.7	127.0
	(95.7)	(74.8)	(76.7)
Year Began College	1976-2015	2020	2021
N	9,068,064	755	894

Notes: Table A.I presents summary statics for the Freshman Survey (column 1), students in our 2020 Ohio State survey (column 2), and students in our 2021 Ohio State surveys. We use the CPI-U to convert family income in the Freshman Survey into September 2020 dollars. Freshman Survey results are weighted by gender, race, and US census division to be nationally representative. See `http://enrollmentservices.osu.edu/report.pdf` for data on the overall Ohio State student body.

Table A.II: Majors Groups in the Freshman Survey

Full Group Name	Short Name	Nationally Representative Survey Major Names
Art or Entertainment	Art	Art, fine and applied, Drafting or Design, Media/Film Studies, Music
Biology or Chemistry	Bio/Chem	Animal Biology, Biochemistry/Biophysics, Biology (general), Botany, Chemistry, Ecology and Evolutionary Biology, Environmental Science, Marine (life) Science, Marine Biology, Medical, Dental, Veterinary, Microbiology, Microbiology or Bacteriology, Molecular, Cellular & Developmental Biology, Neurobiology/Neuroscience, Other Biological Science, Pharmacy, Plant Biology, Zoology
Business or Economics	Business	Accounting, Business Administration (general), Computer/Management Information Systems, Economics, Entreprenuership, Finance, Hospitaliy/Tourism, Human Resource Management, International Business, Management, Marketing, Other Business, Real Estate, Secretarial Studies, Speech, Speech or Theater, Theater/Drama
Communication or Journalism	Communication	Communications (radio, TV, etc.), Journalism, Journalism/Communication
Education	Education	Business Education, Elementary Education, Music/Art Education, Other Education, Physical Education/Recreation, Secondary Education, Special Education
Government or Political Science	Government	Law, Political Science (gov't., international
Humanities	Humanities	Classical and Modern Language and Litera, English (language & literature), Ethnic Studies, Ethnic/Cultural Studies, History, Language and Literature (except English), Other Arts and Humanities, Philosophy, Sociology, Theology/Religion, Women's Studies, Women's/Gender Studies
Math, Engineering, or Computer Science	Engineering	Aeronautical or Astronautical Eng, Aerospace/Aeronautical/Astronautical Engineering, Biological/Agricultural Engineering, Biomedical Engineering, Chemical Engineering, Civil Engineering, Clinical Laboratory Science, Computer Engineering, Computer Science, Data Processing or Computer Programming, Electrical or Electronic Engineering, Electrical/Electronic Communications Engineering, Electronics, Engineering Science/Engineering Physics, Environmental/Enviromental Health Engineering, Health Technology, Industrial Engineering, Industrial/Manufacturing Engineering, Materials Engineering, Mathematics, Mathematics/Statistics, Mechanical Engineering, Other Engineering, Other Math and Computer Science, Statistics
Nursing or Non-Doctor Health Professions	Nursing	Health Care Administration/Studies, Kinesiology, Nursing, Other Health Profession
Psychology or Social Work	Psychology	Psychology, Social Work, Therapy (occupational, physical, speech)
Other	Other	Agriculture, Agriculture/Natural Resources, Anthropology, Architecture/Urban Planning, Astronomy, Astronomy & Astrophysics, Atmospheric Sciences, Building Trades, Criminal Justice, Earth & Planetary Sciences, Earth Science, Forestry, Geography, Home Economics, Law Enforcement, Library Science, Library or Archival Science, Marine Sciences, Mechanics, Military Science, Military Sciences/Technology/operations, Other, Other Physical Science, Other Professional, Other Social Sciences, Other Technical, Physics, Security & Protective Services
Undecided	Undecided	Undecided

Notes: Table A.II presents the groupings of majors we use to aggregate the options in the Freshman Survey.

Table A.III: Career Groups in the Freshman Survey

Full Group Name	Short Name	Nationally Representative Survey Career Names
Artist or Entertainer	Artist	Actor or Entertainer, Artist, Graphic Designer, Musician, Writer/Producer/Director
Business Person	Business	Accountant, Accountant or Actuary, Advertising, Business (clerical), Business Manager/Executive, Business Owner/Entrepreneur, Business Salesperson or Buyer, Finance, Human Resources, Management Consultant, Public/Media Relations, Real Estate, Sales/Marketing, Sports Management
Social Worker or Counselor	Counselor	Clinical Psychologist, School Counselor, Social, Welfare, or Recreation Worker, Social/Non-profit Services, Therapist (e.g., Physical, Occupational,
Doctor	Doctor	Dentist/Orthodontist, Medical Doctor/Surgeon, Optometrist, Pharmacist, Physician, Veterinarian
Engineer or Computer Scientist	Engineer	Computer Programmer or Analyst, Computer Programmer/Developer, Computer/Systems Analyst, Engineer, Web Designer
Lawyer or Judge	Lawyer	Lawyer/Judge
Health Care Worker (non-doctor)	Nurse	Home Health Worker, Medical/Dental Assistant (e.g. Hygienist, Registered Nurse
Teacher	Teacher	Elementary School Teacher, K-12 Administrator, Other K-12 Professional, School Principal or Superintendent, Secondary School Teacher, Secondary School Teacher in Science, Technology, Engineering, or Math (STEM), Secondary School Teacher in a non-STEM subject, Teacher or Administrator (elementary), Teacher or Administrator (secondary), Teacher's Assistant/Paraprofessionsal
Journalist or Writer	Writer	Journalist, Writer or journalist
Other	Other	Administrative Assistant, Architect, Cleargy, Clergy (minister, priest), Clergy (other religious), College Administrator/Staff, College Faculty, Conservationist or forester, Custodian/Janitor/Housekeeper, Dietitcian/Nutritionist, Dietitian or Home Economist, Early Childcare Provider, Farmer or Forester, Farmer or Rancher, Food Service, Foreign Service Worker (including diplom, Government Official, Hair Stylist, Interior Designer, Interpreter (translator), Law Enforcement Officer, Librarian, Military, Natural Resource Specialist/Environmentalist, Other, Paralegal, Policymaker/Government, Postal Worker, Protective Services, Research Scientist, Retail Sales, Scientific Researcher, Skilled Trades (e.g., Plumber, Electrici, Statistician, Unemployed, Urban Planner/Architect
Not Working for Pay	Not Working	Homemaker (full-time), Homemaker/Stay at Home Parent

Notes: Table A.III presents the groupings of careers we use to aggregate the options in the Freshman Survey.

Table A.IV: Beliefs about Careers Conditional on Major

	Artist	Business	Counselor	Doctor	Engineer	Lawyer	Nurse	Teacher	Writer	Other	Not Working
Panel A, Freshman Survey: P(Expected Career | Expected Major)											
Art	**0.65**	0.03	0.01	0.01	0.01	0.01	0.00	0.04	0.01	0.17	0.00
Bio/Chem	0.00	0.01	0.02	**0.63**	0.01	0.01	0.02	0.01	0.00	0.22	0.00
Business	0.05	**0.73**	0.00	0.00	0.01	0.05	0.00	0.01	0.00	0.08	0.00
Communication	0.07	0.07	0.01	0.00	0.00	0.03	0.00	0.01	**0.42**	0.25	0.00
Education	0.02	0.02	0.03	0.00	0.00	0.00	0.01	**0.78**	0.00	0.08	0.00
Government	0.00	0.03	0.01	0.01	0.00	**0.48**	0.00	0.01	0.01	0.36	0.00
Humanities	0.06	0.04	0.04	0.02	0.00	0.13	0.00	0.10	**0.14**	0.29	0.00
Engineering	0.00	0.04	0.00	0.04	**0.71**	0.01	0.01	0.01	0.00	0.11	0.00
Nursing	0.00	0.01	0.03	0.04	0.00	0.00	**0.85**	0.00	0.00	0.05	0.00
Psychology	0.01	0.02	**0.62**	0.04	0.00	0.03	0.00	0.01	0.00	0.14	0.01
Panel B, OSU: Average Beliefs about Self (Restricting to Top-Ranked Major)											
Art	**0.58**	0.04	0.01	0.03	0.03	0.04	0.00	0.04	0.05	0.16	0.03
Bio/Chem	0.01	0.03	0.02	**0.50**	0.06	0.02	0.16	0.05	0.01	0.13	0.03
Business	0.03	**0.71**	0.01	0.02	0.03	0.04	0.02	0.03	0.02	0.08	0.01
Communication	0.06	0.21	0.06	0.02	0.04	0.03	0.02	0.04	**0.34**	0.19	0.00
Education	0.01	0.01	0.03	0.00	0.00	0.04	0.03	**0.77**	0.00	0.08	0.01
Government	0.01	0.15	0.03	0.00	0.00	**0.57**	0.00	0.07	0.07	0.08	0.00
Humanities	0.07	0.09	0.09	0.00	0.01	0.06	0.07	0.11	**0.11**	0.38	0.02
Engineering	0.03	0.07	0.00	0.02	**0.75**	0.01	0.01	0.03	0.01	0.06	0.00
Nursing	0.00	0.03	0.02	0.15	0.01	0.01	**0.68**	0.02	0.00	0.07	0.01
Psychology	0.02	0.08	**0.40**	0.05	0.00	0.04	0.12	0.07	0.02	0.18	0.01
Panel C, OSU: Average Beliefs about Population (Full Sample)											
Art	**0.53**	0.07	0.02	0.01	0.02	0.02	0.03	0.09	0.07	0.11	0.04
Bio/Chem	0.01	0.05	0.03	**0.34**	0.12	0.02	0.19	0.07	0.02	0.14	0.01
Business	0.02	**0.62**	0.02	0.03	0.04	0.07	0.03	0.05	0.03	0.09	0.01
Communication	0.05	0.13	0.06	0.02	0.02	0.04	0.03	0.07	**0.47**	0.10	0.03
Education	0.02	0.06	0.06	0.02	0.03	0.03	0.03	**0.66**	0.03	0.06	0.02
Government	0.02	0.15	0.05	0.03	0.03	**0.38**	0.03	0.09	0.08	0.13	0.02
Humanities	0.06	0.08	0.18	0.04	0.02	0.07	0.08	0.14	**0.11**	0.18	0.03
Engineering	0.01	0.12	0.01	0.05	**0.55**	0.03	0.04	0.07	0.01	0.10	0.02
Nursing	0.01	0.06	0.04	0.23	0.03	0.03	**0.47**	0.04	0.01	0.07	0.01
Psychology	0.02	0.07	**0.43**	0.08	0.02	0.05	0.14	0.07	0.03	0.07	0.02
Panel D, ACS: True P(Career | Major)											
Art	**0.17**	0.18	0.02	0.01	0.06	0.01	0.03	0.13	0.01	0.24	0.15
Bio/Chem	0.01	0.15	0.01	**0.23**	0.04	0.01	0.13	0.09	0.00	0.21	0.11
Business	0.01	**0.47**	0.01	0.00	0.06	0.02	0.03	0.05	0.00	0.23	0.11
Communication	0.05	0.33	0.03	0.00	0.05	0.02	0.03	0.09	**0.04**	0.23	0.13
Education	0.01	0.08	0.03	0.00	0.01	0.00	0.03	**0.57**	0.00	0.13	0.13
Government	0.01	0.30	0.03	0.01	0.05	**0.16**	0.03	0.08	0.01	0.21	0.11
Humanities	0.02	0.22	0.03	0.01	0.04	0.06	0.04	0.18	**0.02**	0.24	0.15
Engineering	0.01	0.22	0.01	0.01	**0.42**	0.01	0.02	0.05	0.00	0.16	0.09
Nursing	0.00	0.07	0.02	0.04	0.01	0.00	**0.60**	0.04	0.00	0.09	0.11
Psychology	0.01	0.17	**0.21**	0.02	0.03	0.02	0.09	0.11	0.00	0.19	0.14
Panel E, ACS: True P(Career | All Other Majors)											
Art	**0.01**	0.26	0.04	0.03	0.10	0.02	0.09	0.13	0.01	0.21	0.12
Bio/Chem	0.02	0.26	0.04	**0.01**	0.10	0.02	0.08	0.13	0.01	0.21	0.12
Business	0.02	**0.19**	0.04	0.03	0.11	0.02	0.10	0.15	0.01	0.21	0.12
Communication	0.02	0.25	0.04	0.03	0.10	0.02	0.09	0.13	**0.00**	0.21	0.12
Education	0.02	0.27	0.04	0.03	0.10	0.02	0.09	**0.08**	0.01	0.22	0.12
Government	0.02	0.25	0.03	0.03	0.10	**0.02**	0.09	0.13	0.01	0.21	0.12
Humanities	0.02	0.26	0.04	0.03	0.10	0.02	0.09	0.12	**0.00**	0.21	0.12
Engineering	0.02	0.26	0.04	0.03	**0.04**	0.02	0.10	0.14	0.01	0.22	0.12
Nursing	0.02	0.27	0.04	0.02	0.10	0.02	**0.04**	0.13	0.01	0.22	0.12
Psychology	0.02	0.26	**0.02**	0.03	0.10	0.02	0.08	0.13	0.01	0.21	0.12

Notes: Panel A of Table A.IV presents the fraction of students in the Freshman Survey sample that expect to have a career in each of the careers listed in the column headings, conditional on expecting to major in the field listed in the rows. Panel B shows the average self-beliefs of students in the 2020 OSU sample about the probability that they will be working in each career if they graduate with that major, restricting the data to students' first-ranked major field. Panel C shows the average population belief in the 2020 OSU sample about the fraction of graduates with each major that is working in each career. Panel D shows the true fraction of college graduates aged 30-50 working in each career conditional on their major, calculated from the 2017-2019 American Community Survey. Panel E shows the fraction working in each career conditional on having a major *other* than the one listed in the row. This is the denominator in our definition of distinctiveness: $p_{c,m}/p_{c,-m}$. The most distinctive career for each major by this metric is bolded.

Table A.V: Careers in the American Community Survey

Full Group Name	Short Name	ACS Career Names
Artist or Entertainer	Artist	Actors, Producers, And Directors, Announcers, Artists And Related Workers, Athletes, Coaches, Umpires, And Related Workers, Dancers And Choreographers, Designers, Entertainers And Performers, Sports And Related Workers, All Other, Musicians, Singers, And Related Workers, Photographers
Business Person	Business	Accountants And Auditors, Actuaries, Administrative Services Managers, Advertising Sales Agents, Agents And Business Managers Of Artists, Performers, And Athletes, Appraisers And Assessors Of Real Estate, Budget Analysts, Chief Executives And Legislators/Public Administration, Constructions Managers, Credit Analysts, Credit Counselors And Loan Officers, Financial Analysts, Financial Examiners, Financial Managers, Financial Specialists, Nec, First-Line Supervisors Of Sales Workers, Food Service And Lodging Managers, Gaming Managers, General And Operations Managers, Human Resources Managers, Human Resources, Training, And Labor Relations Specialists, Industrial Production Managers, Insurance Sales Agents, Insurance Underwriters, Management Analysts, Managers In Marketing, Advertising, And Public Relations, Managers, Nec (Including Postmasters), Natural Science Managers, Operations Research Analysts, Other Business Operations And Management Specialists, Parts Salespersons, Personal Financial Advisors, Property, Real Estate, And Community Association Managers, Public Relations Specialists, Purchasing Managers, Real Estate Brokers And Sales Agents, Sales And Related Workers, All Other, Sales Representatives, Services, All Other, Sales Representatives, Wholesale And Manufacturing, Securities, Commodities, And Financial Services Sales Agents, Tax Examiners And Collectors, And Revenue Agents, Tax Preparers, Transportation, Storage, And Distribution Managers, Travel Agents
Social Worker or Counselor	Counselor	Community And Social Service Specialists, Nec, Counselors, Psychologists, Social And Community Service Managers, Social Workers
Doctor	Doctor	Audiologists, Dentists, Optometrists, Pharmacists, Physicians And Surgeons, Podiatrists, Veterinarians
Engineer or Computer Scientist	Engineer	Aerospace Engineers, Architectural And Engineering Managers, Broadcast And Sound Engineering Technicians And Radio Operators, And Media And Communication Equipment Workers, All Other, Chemical Engineers, Civil Engineers, Computer And Information Systems Managers, Computer Hardware Engineers, Computer Programmers, Computer Scientists And Systems Analysts/Network Systems Analysts/Web Developers, Computer Support Specialists, Database Administrators, Electrical And Electronics Engineers, Engineering Technicians, Except Drafters, Engineers, Nec, Environmental Engineers, Industrial Engineers, Including Health And Safety, Marine Engineers And Naval Architects, Materials Engineers, Mechanical Engineers, Network And Computer Systems Administrators, Petroleum, Mining And Geological Engineers, Including Mining Safety Engineers, Sales Engineers, Software Developers, Applications And Systems Software, Surveying And Mapping Technicians
Lawyer or Judge	Lawyer	Lawyers, And Judges, Magistrates, And Other Judicial Workers, Legal Support Workers, Nec, Paralegals And Legal Assistants
Health Care Worker (non-doctor)	Nurse	Chiropractors, Clinical Laboratory Technologists And Technicians, Dental Assistants, Dental Hygienists, Diagnostic Related Technologists And Technicians, Dieticians And Nutritionists, Emergency Medical Technicians And Paramedics, Health Diagnosing And Treating Practitioner Support Technicians, Health Diagnosing And Treating Practitioners, Nec, Health Technologists And Technicians, Nec, Healthcare Practitioners And Technical Occupations, Nec, Licensed Practical And Licensed Vocational Nurses, Medical And Health Services Managers, Medical Assistants And Other Healthcare Support Occupations, Nec, Medical Records And Health Information Technicians, Medical, Dental, And Ophthalmic Laboratory Technicians, Nursing, Psychiatric, And Home Health Aides, Occupational Therapists, Occupational Therapy Assistants And Aides, Opticians, Dispensing, Personal Care Aides, Physical Therapist Assistants And Aides, Physical Therapists, Physician Assistants, Radiation Therapists, Recreational Therapists, Registered Nurses, Respiratory Therapists, Speech Language Pathologists, Therapists, Nec
Teacher	Teacher	Education Administrators, Education, Training, And Library Workers, Nec, Elementary And Middle School Teachers, Other Teachers And Instructors, Postsecondary Teachers, Preschool And Kindergarten Teachers, Secondary School Teachers, Special Education Teachers, Teacher Assistants
Journalist or Writer	Writer	Editors, News Analysts, Reporters, And Correspondents, Writers And Authors
Other	Other	All other occupation titles
Not Working for Pay	Not Working	All non-employed people

Notes: Table A.V presents the groupings of careers we use to aggregate the occupation titles in the American Community Survey.

Table A.VI: Career and Major Expectations Among College First-Years in the U.S.

Panel A: Career Expectations

Career	Outcomes	Expectations	p-value
Artist	0.022	0.048	0.000
Business	0.260	0.165	0.000
Counselor	0.029	0.059	0.000
Doctor	0.028	0.113	0.000
Engineer	0.098	0.114	0.000
Lawyer	0.024	0.050	0.000
Nurse	0.073	0.035	0.000
Teacher	0.121	0.074	0.000
Writer	0.007	0.027	0.000
Other	0.217	0.183	0.000
Not Working	0.121	0.002	0.000
Undecided	0.000	0.130	0.000

Panel B: Major Expectations

Major	Outcomes	Expectations	p-value
Art	0.042	0.042	0.188
Bio/Chem	0.063	0.148	0.000
Business	0.235	0.193	0.000
Communication	0.045	0.037	0.000
Education	0.089	0.075	0.000
Government	0.030	0.036	0.000
Humanities	0.092	0.061	0.000
Engineering	0.141	0.151	0.000
Nursing	0.072	0.034	0.000
Psychology	0.066	0.073	0.000
Other	0.126	0.080	0.000
Undecided	0.000	0.071	0.000

Notes: Table A.VI shows the distribution of career and major expectations and outcomes in the United States. "Expectations" indicates the fraction of college first-years in the Freshman Survey, spanning the years 1976-2015, that report that their "probable" career (Panel A) or "probable field of study" (Panel B) would fall into each group. "Outcomes" in Panel A indicates the true distribution of occupations of Americans aged 33 to 37 between 1976 and 2020 among the same cohorts (up to birth year 1987), according to data from the Current Population Survey. "Outcomes" in Panel B indicates the true distribution of college majors according to data from the 2017-2019 American Community Survey, using the 1958 to 1997 birth cohorts. p-value is from a t-test for whether the shares are equal across columns.

Table A.VII: Majors in the American Community Survey

Full Group Name	Short Name	ACS Major Names
Art or Entertainment	Art	Commercial Art And Graphic Design, Drama And Theater Arts, Film, Video And Photographic Arts, Fine Arts, Miscellaneous Fine Arts, Music, Studio Arts, Visual And Performing Arts
Biology or Chemistry	Bio/Chem	Biochemical Sciences, Biology, Chemistry, Genetics, Microbiology, Miscellaneous Biology, Molecular Biology, Neuroscience, Nutrition Sciences, Pharmacology, Pharmacy, Pharmaceutical Sciences, And Administration, Physiology
Business or Economics	Business	Accounting, Actuarial Science, Advertising And Public Relations, Business Economics, Business Management And Administration, Economics, Finance, General Business, Hospitality Management, Human Resources And Personnel Management, International Business, Management Information Systems And Statistics, Marketing And Marketing Research, Miscellaneous Business And Medical Administration, Operations, Logistics And E-Commerce
Communication or Journalism	Communication	Communication Technologies, Communications, Composition And Speech, Journalism, Mass Media
Education	Education	Art And Music Education, Early Childhood Education, Educational Administration And Supervision, Elementary Education, General Education, Language And Drama Education, Mathematics Teacher Education, Miscellaneous Education, Physical And Health Education Teaching, Science And Computer Teacher Education, Secondary Teacher Education, Social Science Or History Teacher Education, Special Needs Education, Teacher Education: Multiple Levels
Government or Political Science	Government	International Relations, Political Science And Government, Pre-Law And Legal Studies, Public Administration, Public Policy
Humanities	Humanities	Area, Ethnic, And Civilization Studies, Art History And Criticism, English Language And Literature, French, German, Latin And Other Common Foreign Language Studies, History, Humanities, Intercultural And International Studies, Liberal Arts, Linguistics And Comparative Language And Literature, Other Foreign Languages, Philosophy And Religious Studies, Theology And Religious Vocations, United States History
Math, Engineering, or Computer Science	Engineering	Aerospace Engineering, Applied Mathematics, Architectural Engineering, Biological Engineering, Biomedical Engineering, Chemical Engineering, Civil Engineering, Computer And Information Systems, Computer Engineering, Computer Information Management And Security, Computer Networking And Telecommunications, Computer Programming And Data Processing, Computer Science, Electrical Engineering, Electrical Engineering Technology, Engineering And Industrial Management, Engineering Mechanics, Physics, And Science, Engineering Technologies, Environmental Engineering, General Engineering, Geological And Geophysical Engineering, Industrial And Manufacturing Engineering, Industrial Production Technologies, Information Sciences, Materials Engineering And Materials Science, Materials Science, Mathematics, Mathematics And Computer Science, Mechanical Engineering, Mechanical Engineering Related Technologies, Metallurgical Engineering, Mining And Mineral Engineering, Miscellaneous Engineering, Miscellaneous Engineering Technologies, Naval Architecture And Marine Engineering, Nuclear Engineering, Nuclear, Industrial Radiology, And Biological Technologies, Petroleum Engineering, Statistics And Decision Science
Nursing or Non-Doctor Health Professions	Nursing	Communication Disorders Sciences And Services, Community And Public Health, General Medical And Health Services, Health And Medical Administrative Services, Health And Medical Preparatory Programs, Medical Assisting Services, Medical Technologies Technicians, Miscellaneous Health Medical Professions, Nursing, Treatment Therapy Professions
Psychology or Social Work	Psychology	Clinical Psychology, Cognitive Science And Biopsychology, Counseling Psychology, Educational Psychology, Human Services And Community Organization, Industrial And Organizational Psychology, Miscellaneous Psychology, Psychology, School Student Counseling, Social Psychology, Social Work
Other	Other	Agricultural Economics, Agriculture Production And Management, Animal Sciences, Anthropology And Archeology, Architecture, Astronomy And Astrophysics, Atmospheric Sciences And Meteorology, Botany, Construction Services, Cosmetology Services And Culinary Arts, Court Reporting, Criminal Justice And Fire Protection, Criminology, Ecology, Electrical And Mechanic Repairs And Technologies, Environmental Science, Family And Consumer Sciences, Food Science, Forestry, General Agriculture, General Social Sciences, Geography, Geology And Earth Science, Geosciences, Interdisciplinary And Multi-Disciplinary Studies (General), Interdisciplinary Social Sciences, Library Science, Military Technologies, Miscellaneous Agriculture, Miscellaneous Social Sciences, Multi-Disciplinary Or General Science, Natural Resources Management, Oceanography, Physical Fitness, Parks, Recreation, And Leisure, Physical Sciences, Physics, Plant Science And Agronomy, Precision Production And Industrial Arts, Sociology, Soil Science, Transportation Sciences And Technologies, Zoology
Undecided	Undecided	

Notes: Table A.VII presents the groupings of majors we use to aggregate the options in the American Community Survey.

Table A.VIII: Decomposing Belief Errors: A Shapley Approach

Variable	Shapley Value 2020	Shapley Value 2021
$\mathbb{1}$(Most Distinctive)	35.1 %	33.6 %
Career FEs	34.1 %	8.9 %
$\mathbb{1}$(Most Distinctive)*$\mathbb{1}$(Self Beliefs)	10.2 %	12.1 %
$\mathbb{1}$(Most Distinctive)*$\mathbb{1}$(Top Major)	7.5 %	
$\mathbb{1}$(Most Distinctive)*$\mathbb{1}$(Top Major)*$\mathbb{1}$(Self Beliefs)	4.2 %	
Truth	10.8 %	28.8 %
Truth*$\mathbb{1}$(Self Beliefs)	3.5 %	10.1 %
Truth*$\mathbb{1}$(Top Major)	1.8 %	
Truth*$\mathbb{1}$(Top Major)*$\mathbb{1}$(Self Beliefs)	1.2 %	
Role Model Variables		4.3 %
Role Model Variables*$\mathbb{1}$(Self Beliefs)		2.3 %

Notes: Table A.VIII presents a Shorrocks-Shapley decomposition of the R^2 of an OLS regression. Let $Y_{i,c,m,p}$ denote the belief of individual i about the probability of entering career c conditional on major m from perspective p, where p is either that student's own belief (self) or belief about others (population). Let $T_{c,m}$ denote the true probability from the American Community Survey of someone entering career c conditional on majoring in m. We estimate equations 9 and 10 by OLS. $\psi_{c,s,Top(i,m)}$ are career-by-perspective-by-top fixed effects and $\psi_{c,s}$ are career-by-perspective fixed effects, where top $Top(i,m)$ indicates whether student i listed m as their most likely major. $Self_{i,p}$ indicates whether the belief was about the student's own outcomes or others. $Dist_{m,c}$ indicates whether c is the most distinctive career of major m. Let $RM_{i,c,m}$ be a vector of variables indicating the number of role models i listed with c and m, with m but a career other than c, and with c but a major other than m. We only include the 2020 OSU sample for equation 9 and only the 2021 OSU sample for equation 10.

$$\begin{aligned} Y_{i,c,m,p} - T_{c,m} = \psi_{c,p,Top(i,m)} &+ \beta_1 Self_{i,p} + \beta_2 t_{i,m} + \beta_3 Self_{i,p} \times Top_{i,m} + \beta_4 Dist_{m,c} + \beta_5 Dist_{m,c} \times Self_{i,p} + \\ &\beta_6 Dist_{m,c} \times Top_{i,m} + \beta_7 Dist_{m,c} \times Self_{i,p} \times Top_{i,m} + \beta_8 T_{c,m} + \beta_9 T_{c,m} \times Self_{i,p} + \beta_{10} T_{c,m} \times Top_{i,m} \\ &+ \beta_{11} T_{c,m} \times Self_{i,p} \times Top_{i,m} + \varepsilon_{i,c,m,p} \end{aligned} \tag{9}$$

$$\begin{aligned} Y_{i,c,m,p} - T_{c,m} = \psi_{c,p} &+ \beta_1 Self_{i,p} + \beta_2 t_{i,m} + \beta_4 Dist_{m,c} + \beta_5 Dist_{m,c} \times Self_{i,p} + \beta_8 T_{c,m} + \beta_9 T_{c,m} \times Self_{i,p} + \\ &\beta_{1,2} RM_{i,c,m} + \beta_{1,3} RM_{i,c,m} \times Self_{i,p} + \varepsilon_{i,c,m,p} \end{aligned} \tag{10}$$

After running estimating these regressions, we decompose the R^2 of each model following the Shapley-style method of Shorrocks (1982). In the table above, we show the results of this exercise, where "Career FEs" includes $\{\psi_{c,p,Top(i,m)}, Self_{i,p}, t_{i,m}, Self_{i,p} \times Top_{i,m}\}$.

Table A.IX: OSU 2021: Beliefs about P(Career | Major)

	Artist	Business	Counselor	Doctor	Engineer	Lawyer	Nurse	Teacher	Writer	Other	Not Working
Art	**0.46**	0.10	0.02	0.01	0.02	0.01	0.02	0.07	0.08	0.17	0.05
Bio/Chem	0.01	0.04	0.03	**0.32**	0.11	0.02	0.22	0.11	0.02	0.11	0.02
Business	0.03	**0.56**	0.03	0.02	0.05	0.05	0.03	0.05	0.04	0.10	0.03
Communication	0.06	0.15	0.08	0.01	0.01	0.03	0.02	0.05	**0.44**	0.10	0.04
Education	0.02	0.04	0.06	0.02	0.02	0.01	0.03	**0.65**	0.04	0.09	0.03
Government	0.02	0.16	0.08	0.01	0.02	**0.37**	0.01	0.07	0.10	0.14	0.03
Humanities	0.09	0.09	0.15	0.02	0.01	0.09	0.04	0.19	**0.13**	0.14	0.05
Engineering	0.02	0.12	0.02	0.03	**0.58**	0.02	0.03	0.07	0.02	0.09	0.02
Nursing	0.02	0.05	0.05	0.16	0.03	0.02	**0.51**	0.04	0.02	0.07	0.03
Psychology	0.02	0.06	**0.44**	0.06	0.02	0.04	0.12	0.08	0.04	0.09	0.03

Notes: Table A.IX presents average population beliefs in the 2021 OSU sample about the fraction of graduates with each major that is working in each career. The most distinctive career for each major (where we define distinctiveness by $p_{c,m}/p_{c,-m}$) is bolded.

Table A.X: Choice Model Estimates

	(1)	(2)	(3)	(4)	(5)	(6)
α: E[Salary | M]	0.066***	0.054***	0.039***	0.150***	0.063***	0.068***
	(0.021)	(0.018)	(0.017)	(0.038)	(0.020)	(0.021)
β: P(Favorite Career | M)	4.563***	4.225***	4.115***	4.556***	4.603***	4.468***
	(0.099)	(0.098)	(0.091)	(0.100)	(0.098)	(0.109)
α_2: E[GPA | M]						0.947***
						(0.153)
σ^2: Variance of μ_M^i	1.073***	0.756***	0.596***	1.068***	1.059***	1.042***
	(0.039)	(0.028)	(0.023)	(0.038)	(0.038)	(0.038)
Implied WTP for 1pp Increase in Favorite Career ($10ks)	0.689***	0.781***	1.058***	0.304***	0.731***	0.655***
	(0.380)	(1.130)	(3.245)	(0.091)	(0.501)	(0.326)

Notes: Table A.X shows parameter estimates of the baseline model described in Section 3.1. The model is estimated by maximum likelihood, and standard errors/confidence intervals are constructed by Bayesian bootstrap. *, **, and *** indicate significance at the 10%, 5%, and 1% levels. We omit estimates of μ_M and $c^*(i)$ for readability. Each column shows estimates of a variant of the model, as described below.

Column 1. Baseline model described in Section 3.1.

Column 2. In the baseline model, we winsorize beliefs at 1% and 99% to allow us to take logs. Model 2 is identical to the baseline model, except we instead winsorize beliefs at 2% and 98%.

Column 3. Identical to the baseline model, except we instead winsorize beliefs at 3% and 97%.

Column 4. Identical to the baseline model, except we use do not use the directly elicited salary beliefs. Instead, we construct a weighted average of actual expected salaries by career and major (from the ACS), where the weights are each student's self beliefs about their likelihood of having each career conditional on each major.

Column 5. Identical to the baseline model, except we allow non-pecuniary preferences for each career to be non-zero (but homogeneous across people, up to β). That is, we add $\sum_c \zeta_c \pi^i_{c|M}$ to equation 3. We omit estimates of ζ_c from Table A.X for readability, and instead simply note that their inclusion in the model does not substantially alter the estimates of the main parameters of interest.

Column 6. Identical to the baseline model, except we allow preferences over the difficulty of a major to affect student's choice. To proxy for this, we use answers students gave to the following question: "imagine that you decided to pursue a major in m. What is your best guess about the percent chance that you would earn a sophomore year GPA of..." Students then entered percents into five bins labeled "less than 2.3 (that is, less than a C+)," "from 2.3 to 2.7 (from a C+ to B-)," "from 2.7 to 3.3 (from a B- to a B+)," "from 3.3 to 3.7 (from a B+ to A-)," and "more than 3.7 (more than A-)." We then compute the expected GPA from these answers as our proxy for difficulty. That is, we add α_2E[GPA | M] to equation 4.

Table A.XI: Effects of an Information Intervention on Major Intentions and Class Enrollment

	Intentions (SDs)	Enrolled Classes (SDs)					
	F21	S22	F22	S23	F23	S24	F24
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Top-Ranked Major							
Treatment x Average Reduction in $\pi_{c\|M}$	-1.09**	-2.21**	-0.59	0.41	-0.21	0.56	0.12
	(0.53)	(1.11)	(1.26)	(1.25)	(1.25)	(1.36)	(1.49)
Treatment	-0.06	0.07	-0.12	-0.12	-0.06	-0.04	-0.08
	(0.04)	(0.08)	(0.09)	(0.08)	(0.08)	(0.09)	(0.10)
Average Reduction in $\pi_{c\|M}$	0.02	3.20***	2.58***	3.00***	2.39***	0.73	0.87
	(0.37)	(0.76)	(0.85)	(0.89)	(0.88)	(0.99)	(1.08)
Pre-Treatment Belief P(M)	3.02***	0.84***	1.60***	1.69***	1.83***	1.79***	1.83***
	(0.06)	(0.16)	(0.17)	(0.16)	(0.16)	(0.16)	(0.19)
Pre-Treatment Classes	0.02	0.37***	0.16***	0.07	0.01	0.05	0.06
	(0.02)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Constant	-1.25***	-0.67***	-0.89***	-0.86***	-0.86***	-0.85***	-0.83***
	(0.04)	(0.11)	(0.11)	(0.11)	(0.11)	(0.12)	(0.14)
Panel B: Second-Ranked Major							
Treatment x Average Reduction in $\pi_{c\|M}$	0.66	2.06*	1.65	2.17**	2.42**	0.77	1.02
	(0.48)	(1.18)	(1.18)	(1.05)	(1.09)	(1.25)	(1.33)
Treatment	-0.03	-0.08	-0.10	-0.07	-0.04	-0.00	0.04
	(0.03)	(0.07)	(0.07)	(0.06)	(0.06)	(0.07)	(0.08)
Average Reduction in $\pi_{c\|M}$	-0.15	-0.56	-0.54	-0.32	-0.58	-0.65	-0.67
	(0.32)	(0.77)	(0.85)	(0.63)	(0.65)	(0.89)	(0.86)
Pre-Treatment Belief P(M)	2.61***	0.82***	0.86***	1.02***	1.04***	1.15***	1.08***
	(0.10)	(0.22)	(0.21)	(0.20)	(0.20)	(0.24)	(0.27)
Pre-Treatment Classes	0.01	0.35***	0.27***	0.14***	0.12***	0.03	0.01
	(0.02)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.05)
Constant	-1.20***	-0.58***	-0.59***	-0.65***	-0.68***	-0.61***	-0.62***
	(0.03)	(0.07)	(0.06)	(0.06)	(0.05)	(0.07)	(0.07)
Observations	783	752	726	690	668	642	516
p-value: 1st vs 2nd Major Interaction Equal	0.014	0.009	0.195	0.282	0.113	0.912	0.651

Notes: Table A.XI presents OLS regressions of equation 6 including data from the 2021 OSU sample. The dependent variable in column 1 is students' post-intervention belief about their likelihood of majoring in M (their top-ranked major in Panel A and second-ranked major in Panel B). The dependent variable in columns 2-7 is the number of classes they enrolled in M suring Spring semester 2022 (S22), Fall 2022 (F22), etc."Treatment" is an indicator for treatment status. "Average reduction in $\pi_{c|M}$" is our measure of treatment intensity: the leave-out mean revision in self-beliefs about the likelihood of working in the distinctive career of M conditional on graduating with it."Pre-Treatment Belief P(M)" is students' pre-treatment belief about their likelihood of graduating with a major in their first-ranked major group (odd columns) or second-ranked major group (even columns). "Pre-Treatment Classes" is the number of courses that participants enrolled in in M during Fall 2021. Robust standard errors in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels.

Table A.XII: Effects of an Information Intervention on Major Declarations

	Declared Major				
	S22	F22	S23	F23	S24
	(1)	(2)	(3)	(4)	(5)
Panel A: Top-Ranked Major					
Treatment x Average Reduction in $\pi_{c\|M}$	-0.06	-0.61	-0.54	-0.77	-0.38
	(0.16)	(0.52)	(0.57)	(0.62)	(0.62)
Treatment	-0.01	-0.01	-0.01	0.01	0.02
	(0.01)	(0.03)	(0.04)	(0.04)	(0.04)
Average Reduction in $\pi_{c\|M}$	0.02	1.27***	1.36***	0.65	0.28
	(0.08)	(0.36)	(0.40)	(0.44)	(0.44)
Pre-Treatment Belief P(M)	0.01	0.57***	0.63***	0.68***	0.59***
	(0.03)	(0.06)	(0.06)	(0.07)	(0.07)
Pre-Treatment Classes	-0.00	0.02	0.04*	0.02	0.03
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Constant	0.03	-0.12***	-0.09**	-0.02	0.02
	(0.02)	(0.04)	(0.04)	(0.05)	(0.05)
Panel B: Second-Ranked Major					
Treatment x Average Reduction in $\pi_{c\|M}$	0.11	0.99***	1.40***	0.96**	0.74
	(0.11)	(0.34)	(0.37)	(0.49)	(0.49)
Treatment	-0.00	-0.04**	-0.05**	-0.00	0.01
	(0.01)	(0.02)	(0.02)	(0.03)	(0.03)
Average Reduction in $\pi_{c\|M}$	-0.06	-0.17	-0.29	-0.14	-0.25
	(0.08)	(0.19)	(0.23)	(0.30)	(0.32)
Pre-Treatment Belief P(M)	0.07*	0.20***	0.38***	0.45***	0.48***
	(0.04)	(0.07)	(0.08)	(0.09)	(0.09)
Pre-Treatment Classes	0.02**	0.04**	0.04**	0.03**	0.01
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
Constant	-0.01	0.00	0.00	-0.01	0.01
	(0.01)	(0.02)	(0.02)	(0.03)	(0.03)
Observations	783	783	783	783	783
p-value: 1st vs 2nd Major Interaction Equal	0.374	0.010	0.004	0.028	0.157

Notes: Table A.XII presents OLS regressions of equation 6 including data from the 2021 OSU sample. The dependent variable is an indicated for whether students had declared a major in M (their top-ranked major in Panel A and second-ranked major in Panel B) during Spring semester 2022 (S22), Fall 2022 (F22), etc."Treatment" is an indicator for treatment status. "Average reduction in $\pi_{c|M}$" is our measure of treatment intensity: the leave-out mean revision in self-beliefs about the likelihood of working in the distinctive career of M conditional on graduating with it."Pre-Treatment Belief P(M)" is students' pre-treatment belief about their likelihood of graduating with a major in their first-ranked major group (odd columns) or second-ranked major group (even columns). "Pre-Treatment Classes" is the number of courses that participants enrolled in in M during Fall 2021. Robust standard errors in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels.

Table A.XIII: Effects of an Information Intervention on Major Intentions and Class Enrollment

	Intentions (SDs)	Enrolled Classes (SDs)					
	F21	S22	F22	S23	F23	S24	F24
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Top-Ranked Major							
Treatment	-0.10***	-0.01	-0.14*	-0.09	-0.07	-0.02	-0.08
	(0.03)	(0.07)	(0.08)	(0.08)	(0.08)	(0.08)	(0.09)
Pre-Treatment Belief P(M)	3.02***	0.86***	1.62***	1.70***	1.85***	1.79***	1.82***
	(0.06)	(0.16)	(0.17)	(0.16)	(0.16)	(0.16)	(0.19)
Pre-Treatment Classes	0.02	0.40***	0.19***	0.12***	0.05	0.06	0.08
	(0.02)	(0.05)	(0.05)	(0.04)	(0.04)	(0.05)	(0.05)
Constant	-1.25***	-0.59***	-0.82***	-0.79***	-0.81***	-0.83***	-0.81***
	(0.04)	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)	(0.13)
Panel B: Second-Ranked Major							
Treatment	-0.01	0.00	-0.03	0.02	0.06	0.03	0.08
	(0.03)	(0.06)	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)
Pre-Treatment Belief P(M)	2.60***	0.81***	0.86***	1.00***	1.02***	1.16***	1.08***
	(0.10)	(0.22)	(0.21)	(0.20)	(0.20)	(0.23)	(0.26)
Pre-Treatment Classes	0.01	0.34***	0.27***	0.14***	0.11***	0.03	0.01
	(0.02)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.05)
Constant	-1.21***	-0.60***	-0.61***	-0.66***	-0.70***	-0.64***	-0.65***
	(0.03)	(0.06)	(0.05)	(0.05)	(0.04)	(0.05)	(0.06)
Observations	783	752	726	690	668	642	516
p-value: 1st vs 2nd Major Interaction Equal	0.021	0.884	0.207	0.236	0.162	0.604	0.157

Notes: Table A.XIII presents OLS regressions of equation 6, omitting $AvgReduction_{i,M}$, including data from the 2021 OSU sample. The dependent variable in column 1 is students' post-intervention belief about their likelihood of majoring in M (their top-ranked major in Panel A and second-ranked major in Panel B). The dependent variable in columns 2-7 is the number of classes they enrolled in M suring Spring semester 2022 (S22), Fall 2022 (F22), etc."Treatment" is an indicator for treatment status. "Pre-Treatment Belief P(M)" is students' pre-treatment belief about their likelihood of graduating with a major in their first-ranked major group (odd columns) or second-ranked major group (even columns). "Pre-Treatment Classes" is the number of courses that participants enrolled in in M during Fall 2021. Robust standard errors in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels.

Table A.XIV: Effects of an Information Intervention on Major Declarations

	Declared Major				
	S22	F22	S23	F23	S24
	(1)	(2)	(3)	(4)	(5)
Panel A: Top-Ranked Major					
Treatment	-0.05	-0.10	-0.07	-0.04	0.01
	(0.08)	(0.08)	(0.08)	(0.08)	(0.03)
Pre-Treatment Belief P(M)	0.07	1.53***	1.52***	1.55***	0.60***
	(0.18)	(0.16)	(0.15)	(0.15)	(0.07)
Pre-Treatment Classes	-0.00	0.09*	0.13***	0.05	0.03
	(0.06)	(0.05)	(0.05)	(0.04)	(0.02)
Constant	0.04	-0.68***	-0.68***	-0.61***	0.03
	(0.14)	(0.11)	(0.11)	(0.11)	(0.05)
Panel B: Second-Ranked Major					
Treatment	0.02	-0.00	0.01	0.08	0.04*
	(0.06)	(0.04)	(0.04)	(0.05)	(0.02)
Pre-Treatment Belief P(M)	0.48*	0.51***	0.87***	0.98***	0.47***
	(0.27)	(0.18)	(0.20)	(0.20)	(0.09)
Pre-Treatment Classes	0.14**	0.10**	0.08**	0.08*	0.01
	(0.07)	(0.04)	(0.04)	(0.04)	(0.02)
Constant	-0.22***	-0.46***	-0.56***	-0.63***	-0.00
	(0.07)	(0.04)	(0.05)	(0.05)	(0.02)
Observations	783	783	783	783	783
p-value: 1st vs 2nd Major Interaction Equal	0.454	0.321	0.375	0.170	0.511

Notes: Table A.XII presents OLS regressions of equation 6, omitting $AvgReduction_{i,M}$, including data from the 2021 OSU sample. The dependent variable is an indicated for whether students had declared a major in M (their top-ranked major in Panel A and second-ranked major in Panel B) during Spring semester 2022 (S22), Fall 2022 (F22), etc."Treatment" is an indicator for treatment status. "Average reduction in $\pi_{c|M}$" is our measure of treatment intensity: the leave-out mean revision in self-beliefs about the likelihood of working in the distinctive career of M conditional on graduating with it."Pre-Treatment Belief P(M)" is students' pre-treatment belief about their likelihood of graduating with a major in their first-ranked major group (odd columns) or second-ranked major group (even columns). "Pre-Treatment Classes" is the number of courses that participants enrolled in in M during Fall 2021. Robust standard errors in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels.

Table A.XV: Effects of an Information Intervention on Timing of Major Declarations

	Semesters Undecided	Any Major Declared					Still Taking Classes					
		S22	F22	S23	F23	S24	S22	F22	S23	F23	S24	F24
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Treatment	0.21**	-0.00	-0.05	-0.04	0.02	0.06*	0.01	0.03	0.03	0.03	0.05*	0.03
	(0.08)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
Constant	2.57***	0.07***	0.48***	0.61***	0.74***	0.74***	0.92***	0.88***	0.83***	0.81***	0.77***	0.62***
	(0.05)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Observations	681	814	814	814	814	814	814	814	814	814	814	814

Notes: Table A.XV presents OLS regressions including data from the 2021 OSU sample. The dependent variable in column 1 is the number of semesters the student spend undecided before officially declaring any major. The dependent variable in columns 2-6 is an indicator for whether had declared any major by Spring semester 2022 (S22), Fall 2022 (F22), etc. The dependent variable in columns 7-12 is an indicator for they enrolled in any classes at Ohio State during these semesters. “Treatment” is an indicator for treatment status. Robust standard errors in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels.

B Online Appendix: Data and Additional Analyses

The section describes the data sources used in this paper in greater detail, along with additional analyses.

The text and order of the OSU survey questions can be found at the following link: https://johnjconlon17.github.io/website/survey_instruments_conlon_patel.pdf

B.1 Fall 2020 Ohio State Survey

We embedded the 2020 OSU survey into the Fall semester course associated with the Exploration program. Students accessed the survey through the official course website. They took the survey between October and December and earned extra credit in their Exploration course for doing so. The median student took 27 minutes to complete the survey. Our main study sample includes 755 completed responses, amounting to a roughly 80% response rate.[17]

These surveys focused on the 10 major groups described in Section 2.1. Whenever the surveys mentioned a group of majors, the name of the group appeared in blue font to indicate that students could click it to see which particular majors were included in the group.[18] The surveys also focused on the nine career groups mentioned in Section 2.1. As with majors, the names of our nine groups of careers also always appeared in blue to students, indicating that they could click on the name to see what occupations titles (from the ACS) were included in that group.

One may worry that the quantitative nature of the questions in this survey makes them more difficult and time-consuming to answer than simple multiple choice questions, and that this could be driving our main results. For example, if some respondents found entering percentages tedious and therefore just put salient focal answers (e.g., 0%, 50%, and 100%) to all or many questions, that could bias our results if they did so in a way that disproportionately increased measured beliefs about distinctive careers. While some students do give such answers (about 5% of students' reported beliefs for career distributions by major include an answer of 100% or two answers of 50%), our main results are

[17]Due to a coding error, an additional 44 responses were not usable.

[18]While the list of majors in each group came from the American Community Survey (ACS), in most cases they match very closely with majors that OSU actually offers.

nearly identical if we exclude such responses or such respondents. At the end of the survey, we also asked students how difficult they found it to answer the percent chance questions in the survey. The majority (55%) responded that they found them "moderately difficult".[19] However, in open-ended feedback the overwhelming reason given was that they took longer to fill out than multiple choice questions would have.[20] In addition, all of the main results described in Section 2.2 are nearly identical for students who did and did not report finding these questions difficult to answer.

B.1.1 Eliciting Salary Beliefs

The 2020 OSU survey elicited students' salary beliefs in addition to beliefs about the likelihood of different occupations. The self-beliefs version of this question asked students to imagine they graduated from Ohio State with each of the four majors they were asked about. It then asled their "best guess about the percent chance that, when you are 30 years old, you would earn an annual salary of..." It then listed six ranges of salaries, starting with "less than $30,000" and ending with "more than $150,000" with intervals of $30,000 between. We asked a similar question eliciting students' population salary beliefs.

These questions give us a measure of students' beliefs about the distribution of salaries conditional on majors or careers. We then calculate expected values from these distributions to ease interpretation and compare them to the ACS data. To do so, we assume that salaries are uniformly distributed within the ranges that the survey asked about. We apply a similar assumption to the actual distribution of salaries using ACS data. Namely, we first calculate the share of people with salaries in the ranges listed in the OSU survey. We then calculate the average salary assuming that salaries are uniformly distributed within these ranges.[21]

[19] In our 2021 OSU survey (described below), we added a question about whether students found the percent chance questions annoying to answer directly before a question asking if they found them difficult or confusing. This framing dramatically reduced the fraction of students who rated them as difficult. The mean answer, on a scale from 0 to 100, for the "annoying" question was 67, compared to 33 for the "difficult or confusing" question.

[20] Indeed, one students' reason for finding them difficult was "I find it more efficient to just click an answer that comes first to my mind," which we take to be indication that our questions, at least for this student, induced more careful answers than quicker multiple-choice questions would have.

[21] For the highest bin ("greater than $150,000"), we simply assume a maximum salary of $180,000.

B.2 Fall 2021 Ohio State Surveys

In Fall 2021, we partnered again with the Exploration program to administer two surveys to its incoming cohort, the first between August and September and the second between October and November. The median respondent took 30 minutes to complete the first survey and 25 minutes to complete the second survey. A total of 894 students completed the first survey, and 814 completed the second survey, amounting to approximately 80-90% response rates. Students received a small amount of class credit for their Exploration course for completing the survey.

B.3 CloudResearch Survey

In November of 2021, we recruited 706 US respondents through CloudResearch's mTurk Toolkit to take a short survey. Each participant was asked population beliefs questions about the frequency of careers conditional on a randomly selected two majors (we used the same career and major groups as we focus on throughout the paper). In addition to a $1 completion payment, participants received a $1 bonus if they answered a randomly chosen question in the survey correctly (within 5 percentage points). When scoring the beliefs questions, we chose a random career from among the careers the question asked about and paid participants if their answer about that career was close enough to the correct answer.

Three-fourths of respondents were asked the same population beliefs questions asking for the likelihood of careers conditional on majors as the 2021 OSU sample was asked. The remaining 25% were asked similarly worded questions except the only three options were the distinctive career of that major, "other," and non-employment. We find that in the latter case participants assign a significantly higher probability to distinctive outcomes (analysis available upon request). In the main text, we restrict the data to those who are asked about all nine career groups (plus other and non-employment), to facilitate comparison with the OSU surveys.

Respondents were asked demographics questions about themselves at the end of the survey, including their highest level of education (from which the college vs non-college education split in Table 1 are derived).

B.4 2013 National Survey of College Graduates

In Section 3.1, we mentioned a regression involving data from the 2013 National Survey of College Graduates. Here we give more details about those data and that regression. The dependent variable we use is an indicator variable for whether they express that they are very or somewhat dissatisfied with their primary job. The independent variables are a dummy variable indicating whether the respondent said their principal job was not related to their highest degree and their salary (in $10,000s). We restrict the data to college graduates between the ages of 30 and 50. We additionally includes fixed effects for college major, age, race, and gender. This regression yields a coefficeint of 0.117 ($p < 0.01$) for the unrelated job dummy and a coefficient of -0.004 ($p < 0.01$) for the salary variable (full regression results available upon request).

The survey also asks respondents who work in an area outside their degree field the reason they do so. Only 17.1% list a "Change in career or professional interests" as a reason. The other reasons given include "Pay, promotion opportunities" (26.9%), "Working conditions (e.g., hours, equipment, working environment)" (10.5%), "Job location" (4.4%), "Family-related reasons (e.g., children, spouse's job moved)" (16.0%), "Job in highest degree field not available" (17.3%), and "Some other reason" (7.8%).

B.5 Counterfactual Analysis from Structural Model

Here we provide more details about the counterfactual analysis described in Section 3.1. For this analysis, because students provided beliefs only about four majors (their two most likely plus a randomly selected other two), we first scale up beliefs such that the total probability of graduating with any of the four majors adds up to one. The average student assigns a probability of 85% (median = 95%) to doing so, so this adjustment does not greatly affect the results.

We then compare students stated probabilities of graduating with each major to the model-implied probabilities after changing students' beliefs about salaries and about careers. We assign the true average salary and true population career shares for each major. Note that these are data on the population (since of course we cannot observe the truth for each individual student). We then use the preference parameters estimated from the model to update the students' probability of graduating with each major.

Table B.I: Counterfactual Simulations

	Rank 1 (1)	Rank 2 (2)	Rank 3 (3)	Rank 4 (4)
Reduction in Stereotypical Beliefs	-0.17***	0.07***	0.03***	0.03***
	(0.03)	(0.02)	(0.01)	(0.01)
Constant	-0.12***	0.05***	0.04***	0.03***
	(0.01)	(0.01)	(0.00)	(0.00)
N	747	747	747	747
R^2	0.047	0.017	0.020	0.050
Mean of Dependent Variable	-0.156	0.068	0.043	0.034

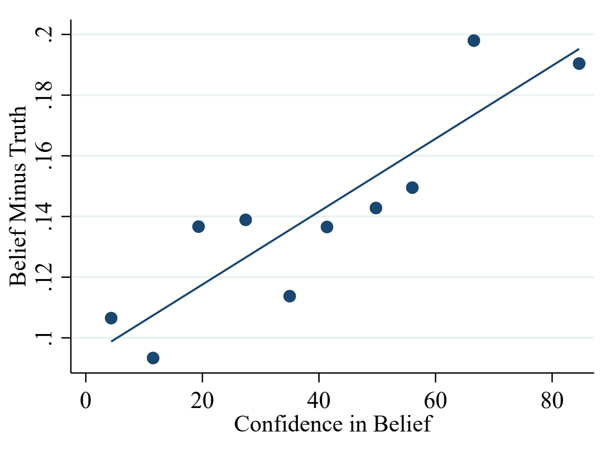
Notes: Table B.I presents OLS regressions of the simulated reduction in students' beliefs about their likelihood of graduating with each major on the counterfactual reduction in their beliefs about their likelihood of attaining that major's distinctive career. Columns 1 through 4 include data on students first- through fourth-ranked majors, respectively. *, **, and *** indicate significance at the 10%, 5%, and 1% levels.

Table B.I shows the average revision in probability for each of the four majors. We see that students move away from their top-ranked major (by 15.6 p.p.) and correspondingly toward their lower ranked majors. Table B.I also shows OLS regressions of this revision on the reduction in students' beliefs about the distinctive career of each major. We see that such reductions reduce students' propensity to pursue their top-ranked major. This is because for most students (50.5%) the distinctive career of this major is their preferred job. In contrast, greater reductions in stereotyping move students more *toward* their lower ranked majors. This is intuitive: few students' preferred career is the distinctive outcome of their lower ranked majors, so these reductions do not count as bad news for the large majority of students.

B.6 Correlation between Stereotypes and Confidence

The first 2021 OSU survey, in addition to asking students' beliefs about the frequency of careers conditional on majors, also asked how confident they were in their in their guesses. In particular, immediately after each question about careers conditional on major, students were asked "And on a scale between 0 (completely uncertain) and 100 (completely certain), how confident are you that the answers above are close to correct?" We find that, controlling for major-by-career and individual fixed effects, a one standard deviation increase in this confidence variable (24 points on the 100 point scale) predicts a 7.4 percentage point greater exaggeration in population beliefs about distinctive careers ($p < 0.01$). Figure B.I shows this regression visually using a binscatter plot.

Figure B.I: Stereotyping and Confidence are Positively Correlated



Notes: Figure B.I shows a binscatter plot correlating students' confidence that their population beliefs are correct with their error in the share of graduates working in each major's distinctive career.

C Online Appendix: Model and Proofs

Here we present a model of belief formation from "what comes to mind," building off Bordalo et al. (2023) and Bordalo et al. (2024). Assume students form beliefs about the likelihood of careers either conditional on a particular major or unconditional on major. We use lower-case letters to denote careers (i.e., $c \in \{a, b, ...\}$) and upper-case letters to denote majors (i.e., $M \in \{A, B\}$). For simplicity, we assume there are only two majors.

The student first separately assesses the "plausibility" $F(H)$ of each relevant "hypothesis" H. When assessing unconditional probabilities, these hypotheses H_c simply correspond to the set of people with each career c. When assessing probabilities of careers conditional on major M, these hypotheses $H_{c,M}$ correspond to the set of people with *both* career c and major M. The student's probabilistic beliefs about H, shown in equation 11, are then the plausibility of H normalized such that their beliefs about all relevant hypotheses sum to one.

$$\pi_c = \frac{F(H_c)}{\sum_{z \in \{a,b,...\}} F(H_z)} \qquad \pi_{c|M} = \frac{F(H_{c,M})}{\sum_{z \in \{a,b,...\}} F(H_{z,M})} \tag{11}$$

Beliefs from Recall To assess plausibilities, we assume the student repeatedly follows a two-stage process for each hypothesis separately. First, they recall an person e from their memory "database" D. To highlight that belief biases arise even without biased data, we begin by assuming that D is representative of the population, in the sense that the people in it reflect the true joint distribution of careers and majors. Later, we explore how differences across individuals in whom students know systematically predict their beliefs. We also assume D is sufficiently large that we can take derivatives with respect to the true fraction $p_{c,M}$ of people with major M and career c. We can think of the database as comprised of people the student knows personally like friends or family, those they have met or seen only a few times, as well as people they have merely heard about, e.g., from the media or second-hand from others.

Let $r(e, H)$ be the probability that a person e is recalled when assessing hypothesis H. Critically, because the student recalls people one at a time, memories must compete for recall. To capture this, we assume in equation 12 that $r(e, H)$ depends not only on how memorable (or "available" in memory) e is, which we denote by $a(e, H)$, but also the availability of others.

$$r(e, H) = \frac{a(e, H)}{\sum_u a(u, H)} \tag{12}$$

The numerator of equation 12 implies that more available people are more likely to come to mind. The denominator captures competition for recall: factors that make one person come to mind more easily do so at the expense of others (that is, they "interfere" with each other).

Associative and Frequency-based Recall We assume two factors matter for the availability of a given item e: associativity (i.e., similarity-based) and frequency (i.e., items people have more experience with come to mind more easily). For assosicativity, we assume that the availability $a(e, H)$ of person e when assessing hypothesis H depends on the similarity $S(e, H)$ between e and H. We define the similarity between e and H as simply the average pair-wise similarity $s(e, u)$ between e and everyone u who is consistent with H.

Second, items are more likely to come to mind the more experiences the agent has with them. Let $N(e)$ denote this measure of quantity of experiences with e. For simplicity, we assume that the student has one role model, whom we call x, and a fraction ϕ of the student's experiences are with this person, whereas they have only one experience with everyone else. We can then evaluate how beliefs change as we increase ϕ. This comparative static can be thought of as asking about the effect of increasing the student's exposure to their role model x. Taking these assumptions together, total availability is given by 13.

$$a(e,H) = N(e)S(e,H) \quad \text{where} \quad N(e) = \begin{cases} \frac{\phi}{D} & \text{if } e = x \\ 1 & \text{if } e \neq x \end{cases} \tag{13}$$

We assume that the similarity between two people e and u decreases by a factor of $\delta_c \leq 1$ if they have different careers and by $\delta_M \leq 1$ if they have different majors, as in equation 14.

$$s(e,u) = \delta_c^{\mathbb{1}(c(e) \neq c(u))} \times \delta_M^{\mathbb{1}(M(e) \neq M(u))} \tag{14}$$

Simulation given Retrieval Second, following Bordalo et al. (2024), the student tries to "simulate" the hypothesis H_c (i.e., imagine someone having the career/major she is assessing) using the person e that they have recalled. Let $\sigma(e,H)$ be how easy it is to simulate H after recalling e. The plausibility of H is then the average ease of simulation among the people that the student recalled.

We assume the functional form in equation 15, whereby ease-of-simulation decreases by a factor of $\eta_c \leq 1$ if e lacks the relevant career and by $\eta_M \leq 1$ if e lacks the relevant major.

$$\sigma(e, H_c) = \eta_c^{\mathbb{1}(c(e) \neq c)} \qquad \sigma(e, H_{c,M}) = \eta_c^{\mathbb{1}(c(e) \neq c)} \times \eta_M^{\mathbb{1}(M(e) \neq M)} \tag{15}$$

First, let T be the number of times the student samples an item from their database and uses it to simulate the hypothesis H. Let e_t be the tth item that they sample. Then $\sigma(e_t, H)$ is the ease of simulating H given e_t. The expected value of $\sigma(e_t, H)$ can be written as follows:

$$E[\sigma(e_t, H)] = \sum_{e \in \mathcal{D}} P(e_t = e) E[\sigma(e, H)] = \sum_{e \in \mathcal{D}} r(e,H)\sigma(e,H)$$

The plausibility of H is the average ease of simulation of the items the student samples. The law of large numbers then implies the following as the number of samples T goes to infinity:

$$\frac{1}{T}\sum_{t=1}^{T}\sigma(e_t, H) \xrightarrow{p} \sum_{e\in\mathcal{D}} r(e,H)\sigma(e,H)$$

This model naturally nests the rational-expectations benchmark. In particular, if $a\big(e,H\big)$ is constant and $\sigma(e,H) = \mathbb{1}(e \in H)$ (i.e., $\eta_c = \eta_M = 0$), then the student's beliefs will be correct. This corresponds to the case where the student simply takes an unbiased random sample of people in their database and counts the number consistent with each hypothesis.

We derive predictions regarding beliefs about careers conditional on major. The plausibility of two hypotheses $H_{a,A}$ and $H_{b,A}$ can be written as follows:

$$F(H_{a,A}) = \frac{p_{a,A} + \delta_c\eta_c p_{b,A} + \delta_c\eta_c \sum_{z\notin\{a,b\}} p_{z,A} + \delta_M\eta_M p_{a,B} + \delta_c\delta_M\eta_c\eta_M p_{b,B} + \delta_c\delta_M\eta_c\eta_M \sum_{z\notin\{a,b\}} p_{z,B} + \frac{1}{D}a(x|a,A)\sigma(x|a,A)}{p_{a,A} + \delta_c p_{b,A} + \delta_c \sum_{z\notin\{a,b\}} p_{z,A} + \delta_M p_{a,B} + \delta_c\delta_M p_{b,B} + \delta_c\delta_M \sum_{z\notin\{a,b\}} p_{z,B} + \frac{1}{D}a(x|a,A)}$$

$$F(H_{b,A}) = \frac{\delta_c\eta_c p_{a,A} + p_{b,A} + \delta_c\eta_c \sum_{z\notin\{a,b\}} p_{z,A} + \delta_c\delta_M\eta_c\eta_M p_{a,B} + \delta_c\eta_c p_{b,B} + \delta_c\delta_M\eta_c\eta_M \sum_{z\notin\{a,b\}} p_{z,B} + \frac{1}{D}a(x|b,A)\sigma(x|b,A)}{\delta_c p_{a,A} + p_{b,A} + \delta_c \sum_{z\notin\{a,b\}} p_{z,A} + \delta_c\delta_M p_{a,B} + \delta_c p_{b,B} + \delta_c\delta_M \sum_{z\notin\{a,b\}} p_{z,B} + \frac{1}{D}a(x|b,A)}$$

To see how this model endogenously generates distinctiveness-based stereotyping, let p_B be the fraction of people with major B. We can consider how the agent's beliefs $\pi_{a|A}$ change as we increase the fraction of people with major B who have career a. More precisely, let $p_{a,B} = \alpha p_B$ and $p_{c,B} = (\beta - \alpha)p_B$ for some other career c. We can then ask how beliefs change as we increase α: that is, as we shift a fraction of B majors from having career c to career a. Additionally, let $\phi = 0$ so that we can ignore role models for now. Then,

$$\frac{\partial}{\partial \alpha}\left[\log \frac{\pi_{a|A}}{\pi_{b|A}}\right] = p_B \frac{\delta_M \eta_M - \delta_c \delta_M \eta_c \eta_M}{p_{a,A} + \delta_c \eta_c p_{b,A} + \delta_c \eta_c \sum_{z \notin \{a,b\}} p_{z,A} + \delta_M \eta_M p_{a,B} + \delta_c \delta_M \eta_c \eta_M p_{b,B} + \delta_c \delta_M \eta_c \eta_M \sum_{z \notin \{a,b\}} p_{z,B}}$$
$$- p_B \frac{\delta_M - \delta_c \delta_M}{p_{a,A} + \delta_c p_{b,A} + \delta_c \sum_{z \notin \{a,b\}} p_{z,A} + \delta_M p_{a,B} + \delta_c \delta_M p_{b,B}}$$

Note that when $\eta_c = \eta_M = 0$ this derivative is unambiguously negative whenever $\delta_c < 1$. This reflects the fact that when we increase the number of B majors with career a, these people increasingly are cued when the agent tries to think of A majors with this career. They interfere with thinking about people with (a, A), reducing beliefs.

To see how role models affect beliefs in this setup, let $\Big(c(x), m(x)\Big) = (a, A)$. That is, the student's role model has career a and major A. How does increasing their exposure to this role model (i.e., increasing ϕ) impact beliefs about careers conditional on A?

$$\begin{aligned}
\log \frac{\pi_{a|A}}{\pi_{b|A}} &= \log \frac{F(H_{1,1})}{F(H_{2,1})} \\
&= \log\Big(p_{a,A} + \delta_c \eta_c p_{b,A} + \delta_c \eta_c \sum_{z \notin \{a,b\}} p_{z,A} + \delta_M \eta_M p_{a,B} + \delta_c \delta_M \eta_c \eta_M p_{b,B} + \delta_c \delta_M \eta_c \eta_M \sum_{z \notin \{a,b\}} p_{z,B} + \phi\Big) \\
&\quad - \log\Big(p_{a,A} + \delta_c p_{b,A} + \delta_c \sum_{z \notin \{a,b\}} p_{z,A} + \delta_M p_{a,B} + \delta_c \delta_M p_{b,B} + \delta_c \delta_M \sum_{z \notin \{a,b\}} p_{z,B} + \phi\Big) \\
&\quad - \log\Big(\delta_c \eta_c p_{a,A} + p_{b,A} + \delta_c \eta_c \sum_{z \notin \{a,b\}} p_{z,A} + \delta_c \delta_M \eta_c \eta_M p_{a,B} + \delta_M \eta_M p_{b,B} + \delta_c \delta_M \eta_c \eta_M \sum_{z \notin \{a,b\}} p_{z,B} + \delta_c \eta_c \phi \\
&\quad + \log\Big(\delta_c p_{a,A} + p_{b,A} + \delta_c \sum_{z \notin \{a,b\}} p_{z,A} + \delta_c \delta_M p_{a,B} + \delta_M p_{b,B} + \delta_c \delta_M \sum_{z \notin \{a,b\}} p_{z,B} + \delta_c \phi\Big)
\end{aligned}$$

We can then take the derivative of the agent's subjective log odds with respect to ϕ:

$$
\begin{aligned}
\frac{\partial}{\partial \phi} \log \frac{\pi_{a|A}}{\pi_{b|A}} =& \frac{1}{p_{a,A} + \delta_c \eta_c p_{b,A} + \delta_c \eta_c \sum_{z \notin \{a,b\}} p_{z,A} + \delta_M \eta_M p_{a,B} + \delta_c \delta_M \eta_c \eta_M p_{b,B} + \delta_c \delta_M \eta_c \eta_M \sum_{z \notin \{a,b\}} p_{z,B} + \phi} \\
&- \frac{1}{p_{a,A} + \delta_c p_{b,A} + \delta_c \sum_{z \notin \{a,b\}} p_{z,A} + \delta_M p_{a,B} + \delta_c \delta_M p_{b,B} + \delta_c \delta_M \sum_{z \notin \{a,b\}} p_{z,B} + \phi} \\
&- \frac{\delta_c \eta_c}{\delta_c \eta_c p_{a,A} + p_{b,A} + \delta_c \eta_c \sum_{z \notin \{a,b\}} p_{z,A} + \delta_c \delta_M \eta_c \eta_M p_{a,B} + \delta_M \eta_M p_{b,B} + \delta_c \delta_M \eta_c \eta_M \sum_{z \notin \{a,b\}} p_{z,B} + \delta_c \eta_c \phi} \\
&+ \frac{\delta_c}{\delta_c p_{a,A} + p_{b,A} + \delta_c \sum_{z \notin \{a,b\}} p_{z,A} + \delta_c \delta_M p_{a,B} + \delta_M p_{b,B} + \delta_c \delta_M \sum_{z \notin \{a,b\}} p_{z,B} + \delta_c \phi}
\end{aligned}
$$

Note that the first term is larger in magnitude than the second term. The third term is smaller than the fourth term (guaranteeing that the whole derivative is positive) whenever

$$
(1 - \eta_c) p_{b,A} > \delta_c \delta_M \eta_c (1 - \eta_M) [p_{a,B} + \sum_{z \notin \{a,b\}} p_{z,B}] + \delta_M (\eta_c - \eta_M) p_{b,B}
$$

Thus, a sufficient condition for the derivative to be positive is for either η_c to be close to zero or $p_{b,A}$ to be large. These conditions require the "spillover" effect of simulation (retrieving someone with the same career but different major) is not larger than the distraction effect (knowing someone with a and A makes it harder to think of someone with a and B). Note that the model of Bordalo et al. (2023) corresponds to the case where $\eta_M = \eta_c = 0$, so in that case this prediction is unambiguous.

Note also that if we remove major B (in essence changing the belief to be unconditional on major rather than conditional on major A vs B), so that $p_{c,B} = 0$ for all c, then this condition is also satisfied. Hence we should expect positive role model effects for the unconditional beliefs about careers, as we saw in Section 4.

To analyze the effect of having a role model with the "correct" career a but "wrong" major B, let $\Big(c(x), m(x)\Big) = (a, B)$. Then

$$
\begin{aligned}
\log \frac{\pi_{a|A}}{\pi_{b|A}} &= \log \frac{F(H_{1,1})}{F(H_{2,1})} \\
&= \log\Big(p_{a,A} + \delta_c \eta_c p_{b,A} + \delta_c \eta_c \sum_{z \notin \{a,b\}} p_{z,A} + \delta_M \eta_M p_{a,B} + \delta_c \delta_M \eta_c \eta_M p_{b,B} + \delta_c \delta_M \eta_c \eta_M \sum_{z \notin \{a,b\}} p_{z,B} + \delta_M \eta_M \phi\Big) \\
&\quad - \log\Big(p_{a,A} + \delta_c p_{b,A} + \delta_c \sum_{z \notin \{a,b\}} p_{z,A} + \delta_M p_{a,B} + \delta_c \delta_M p_{b,B} + \delta_c \delta_M \sum_{z \notin \{a,b\}} p_{z,B} + \delta_M \phi\Big) \\
&\quad - \log\Big(\delta_c \eta_c p_{a,A} + p_{b,A} + \delta_c \eta_c \sum_{z \notin \{a,b\}} p_{z,A} + \delta_c \delta_M \eta_c \eta_M p_{a,B} + \delta_M \eta_M p_{b,B} + \delta_c \delta_M \eta_c \eta_M \sum_{z \notin \{a,b\}} p_{z,B} + \delta_c \delta_M \eta_c \eta_M \phi\Big) \\
&\quad + \log\Big(\delta_c p_{a,A} + p_{b,A} + \delta_c \sum_{z \notin \{a,b\}} p_{z,A} + \delta_c \delta_M p_{a,B} + \delta_M p_{b,B} + \delta_c \delta_M \sum_{z \notin \{a,b\}} p_{z,B} + \delta_c \delta_M \phi\Big)
\end{aligned}
$$

Then,

$$
\begin{aligned}
\frac{\partial}{\partial \phi} \log \frac{\pi_{a|A}}{\pi_{b|A}} &= \frac{\delta_M \eta_M}{p_{a,A} + \delta_c \eta_c p_{b,A} + \delta_c \eta_c \sum_{z \notin \{a,b\}} p_{z,A} + \delta_M \eta_M p_{a,B} + \delta_c \delta_M \eta_c \eta_M p_{b,B} + \delta_c \delta_M \eta_c \eta_M \sum_{z \notin \{a,b\}} p_{z,B} + \delta_M \eta_M \phi} \\
&\quad - \frac{\delta_M}{p_{a,A} + \delta_c p_{b,A} + \delta_c \sum_{z \notin \{a,b\}} p_{z,A} + \delta_M p_{a,B} + \delta_c \delta_M p_{b,B} + \delta_c \delta_M \sum_{z \notin \{a,b\}} p_{z,B} + \delta_M \phi} \\
&\quad - \frac{\delta_c \delta_M \eta_c \eta_M}{\delta_c \eta_c p_{a,A} + p_{b,A} + \delta_c \eta_c \sum_{z \notin \{a,b\}} p_{z,A} + \delta_c \delta_M \eta_c \eta_M p_{a,B} + \delta_M \eta_M p_{b,B} + \delta_c \delta_M \eta_c \eta_M \sum_{z \notin \{a,b\}} p_{z,B} + \delta_c \delta_M \eta_c \eta_M \phi} \\
&\quad + \frac{\delta_c \delta_M}{\delta_c p_{a,A} + p_{b,A} + \delta_c \sum_{z \notin \{a,b\}} p_{z,A} + \delta_c \delta_M p_{a,B} + \delta_M p_{b,B} + \delta_c \delta_M \sum_{z \notin \{a,b\}} p_{z,B} + \delta_c \delta_M \phi}
\end{aligned}
$$

Note that when $\eta_c = \eta_M = 0$, this derivative is unambiguously negative, as alluded to in the main text. To investigate this prediction allowing for simulation/extrapolation, we compute a first-order Taylor approximation. First,

$$
\begin{aligned}
\frac{\partial^2}{\partial \phi \partial \delta_c} \log \frac{\pi_{a|A}}{\pi_{b|A}} &= \frac{-\delta_M \eta_M (\eta_c p_{b,A} + \eta_c \sum_{z \notin \{a,b\}} p_{z,A} + \delta_M \eta_M p_{a,B} + \delta_M \eta_c \eta_M p_{b,B} + \delta_M \eta_c \eta_M \sum_{z \notin \{a,b\}} p_{z,B})}{(p_{a,A} + \delta_c \eta_c p_{b,A} + \delta_c \eta_c \sum_{z \notin \{a,b\}} p_{z,A} + \delta_M \eta_M p_{a,B} + \delta_c \delta_M \eta_c \eta_M p_{b,B} + \delta_c \delta_M \eta_c \eta_M \sum_{z \notin \{a,b\}} p_{z,B} + \delta_M \eta_M \phi)^2} \\
&\quad + \frac{\delta_M (p_{b,A} + \sum_{z \notin \{a,b\}} p_{z,A} + \delta_M p_{b,B} + \delta_M \sum_{z \notin \{a,b\}} p_{z,B})}{(p_{a,A} + \delta_c p_{b,A} + \delta_c \sum_{z \notin \{a,b\}} p_{z,A} + \delta_M p_{a,B} + \delta_c \delta_M p_{b,B} + \delta_c \delta_M \sum_{z \notin \{a,b\}} p_{z,B} + \delta_M \phi)^2} \\
&\quad - \frac{\delta_M \eta_c \eta_M}{\delta_c \eta_c p_{a,A} + p_{b,A} + \delta_c \eta_c \sum_{z \notin \{a,b\}} p_{z,A} + \delta_c \delta_M \eta_c \eta_M p_{a,B} + \delta_M \eta_M p_{b,B} + \delta_c \delta_M \eta_c \eta_M \sum_{z \notin \{a,b\}} p_{z,B} + \delta_c \delta_M \eta_c \eta_M \phi} \\
&\quad + \frac{\delta_c \delta_M \eta_c \eta_M (\delta_c \eta_c p_{a,A} + p_{b,A} + \delta_c \eta_c \sum_{z \notin \{a,b\}} p_{z,A} + \delta_c \delta_M \eta_c \eta_M p_{a,B} + \delta_M \eta_M p_{b,B} + \delta_c \delta_M \eta_c \eta_M \sum_{z \notin \{a,b\}} p_{z,B} + \delta_c \delta_M \eta_c \eta_M \phi)}{(\eta_c p_{a,A} + \eta_c \sum_{z \notin \{a,b\}} p_{z,A} + \delta_M \eta_c \eta_M p_{a,B} + \delta_M \eta_c \eta_M \sum_{z \notin \{a,b\}} p_{z,B} + \delta_M \eta_c \eta_M \phi)^2} \\
&\quad + \frac{\delta_M}{\delta_c p_{a,A} + p_{b,A} + \delta_c \sum_{z \notin \{a,b\}} p_{z,A} + \delta_c \delta_M p_{a,B} + \delta_M p_{b,B} + \delta_c \delta_M \sum_{z \notin \{a,b\}} p_{z,B} + \delta_c \delta_M \phi} \\
&\quad - \frac{\delta_M (\delta_c p_{a,A} + \sum_{z \notin \{a,b\}} p_{z,A} + \delta_M p_{a,B} + \delta_M p_{b,B} + \delta_M \sum_{z \notin \{a,b\}} p_{z,B} + \delta_M \phi)}{(\delta_c p_{a,A} + p_{b,A} + \delta_c \sum_{z \notin \{a,b\}} p_{z,A} + \delta_c \delta_M p_{a,B} + \delta_M p_{b,B} + \delta_c \delta_M \sum_{z \notin \{a,b\}} p_{z,B} + \delta_c \delta_M \phi)^2}
\end{aligned}
$$

We can evaluate this derivative at the rational benchmark:

$$\frac{\partial^2}{\partial\phi\partial\delta_c}\left[\log\frac{\pi_{a|A}}{\pi_{b|A}}\right]\Bigg|_{\delta_c=\delta_M=1,\eta_c=\eta_M=0} = \frac{p_{b,A}+\sum_{z\notin\{a,b\}}p_{z,A}+p_{b,B}+\sum_{z\notin\{a,b\}}p_{z,B}}{(1+\phi)^2}$$
$$= \frac{1-p_{b,A}-p_{a,A}+p_{b,B}-p_{a,B}}{(1+\phi)^2}$$

Similar derivations show that at the rational benchmark,

$$\frac{\partial^2}{\partial\phi\partial\delta_M}\left[\log\frac{\pi_{a|A}}{\pi_{b|A}}\right]\Bigg|_{\delta_c=\delta_M=1,\eta_c=\eta_M=0} = \frac{\partial^2}{\partial\phi\partial\eta_c}\left[\log\frac{\pi_{a|A}}{\pi_{b|A}}\right]\Bigg|_{\delta_c=\delta_M=1,\eta_c=\eta_M=0} = 0$$
$$\frac{\partial^2}{\partial\phi\partial\eta_M}\left[\log\frac{\pi_{a|A}}{\pi_{b|A}}\right]\Bigg|_{\delta_c=\delta_M=1,\eta_c=\eta_M=0} = \frac{1}{p_{a,A}}$$

Combining these, we can approximate $\frac{\partial}{\partial\phi}\left[\log\frac{\pi_{a|A}}{\pi_{b|A}}\right]$:

$$\frac{\partial}{\partial\phi}\left[\log\frac{\pi_{a|A}}{\pi_{b|A}}\right] \approx \eta_M\frac{1}{p_{a,A}} - (1-\delta_c)\frac{1-p_{b,A}-p_{a,A}+p_{b,B}-p_{a,B}}{(1+\phi)^2}$$

Thus, the effect of the student's role model is now ambiguous. If $p_{a,A}$ is small enough, the first term will dominate and the effect will be negative. If $p_{a,A}$ is large (and η_M is not too large), then the second term will dominate and the effect will be negative. These results reflect the countervailing roles of extrapolation and interference. On the one hand, knowing someone with the right career but wrong major makes it harder to recall people with both the right major and career. This is interference, captured by the second term. In contrast, knowing someone with the right career but wrong major partially helps, by extrapolation, to simulate the hypothesis of an A major working as a. If $p_{a,A}$ is small, then there are few relevant people to distract from, so in this case the simulation term dominates. In contrast, if there are many people whom x might distract from ($p_{a,A}$ is large), then interference will dominate and the overall effect will be negative.