What Jobs Come to Mind? Stereotypes about Fields of Study

John J. Conlon* Dev Patel†

August 31, 2025

Abstract

We test for stereotyping—the exaggeration of distinctive traits—in a high-stakes economic environment. 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 overestimating the likelihood that majors lead to their distinctive jobs (e.g., counselor for psychology, journalist for journalism). Using an implicit association test, we show that students associate majors with their distinctive careers and that these associations strongly predict belief biases, in line with a stereotyping mechanism. A simple equilibrium model of the labor market predicts that stereotyping reduces welfare costs by increasing misallocation, which suggestive evidence on job/major mismatch corroborates. In a field experiment, we test a light-touch policy to reduce stereotyping and find significant effects on students' intentions about what to study as well as the classes and majors in which they enroll.

*Carnegie Mellon University, jconlon@andrew.cmu.edu. †Harvard and MIT, 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 experiments described in this paper were pre-registered at the AEA RCT Registry (IDs AEARCTR-0008339 and AEARCTR-0015937).

1 Introduction

Stereotypes (in)famously bias perceptions of social groups, leading to inaccurate statistical discrimination, distorted policy preferences, and adverse economic outcomes (e.g., Bohren et al. 2023, Bursztyn et al. 2023, Haaland & Roth 2023, Chan 2024). To explain such stereotypes, psychologists have posited a cognitive mechanism 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 theory also makes a prediction for any setting that features highly distinctive traits: even among groups not typically thought of as being subject to stereotypes, we might expect biased beliefs to develop via the same psychological mechanism, distorting choices that hinge on these expectations.

In this paper, we test this prediction in a novel and high-stakes economic environment: the link between college major choice and student's future occupations. We first define for each college major its "distinctive" career: the one that is most common among that major's graduates *compared to* other majors (e.g., lawyers for government majors, journalists for communication majors). We show that for such careers, absolute frequencies of distinctive careers conditional on major are often modest—between 2% and 60% depending on field—but relative frequencies are nonetheless very extreme: graduates are between 155% and 1,751% more likely to have their major's distinctive career than graduates with other majors. If such distinctiveness generates stereotypes that exaggerate these kernels of truth, we should expect students' beliefs to systemically overweight such careers.

For four reasons, this context represents an ideal field setting to test distinctiveness-based theories of stereotyping. First, as described above, the key ingredient proposed as leading to biased beliefs—a disconnect between absolute and relative frequencies—is present to a very large degree. Second, our context lacks many features that plausibly serve as confounds to well-known stereotypes surrounding social groups: e.g., animus, historical or contemporary discrimination, and political or cultural overtones. Third, it is an extremely high-stakes economic choice where students have strong incentive to learn the truth, but where several facts point toward potentially consequential mistakes in need of explanation: for example, almost half of employed US college graduates report that their job is either not related or only somewhat related to their field of study, and a full 36% report regretting their choice of field of study.[1] Finally, and perhaps most importantly, to our knowledge there is no evidence yet on whether students have biased beliefs about the distribution of occupations by major.[2] This context therefore provides an "out-of-sample" test of distinctiveness-based theories of stereotyping in the field.

We begin by presenting motivating evidence of stereotyping in this setting from nationally representative surveys of millions of US college freshmen spanning over 40 years. 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 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 irrespective of major,

[1] These numbers come, respectively, from the 2013 National Survey of College Graduates and the 2021 Survey of Household Economics and Decisionmaking.

[2] 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, Arcidiacono et al. 2020, de Koning et al. 2025, Alfonsi et al. 2025) has not examined biases about how college majors map onto occupations.

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 particular careers *per se* but about the relationship *between* majors and careers. 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 incentivized online surveys 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.

To further test whether stereotyping is the correct interpretation of these results, we adapt to our setting the most prevelant non-belief measure of stereotyping from the psychology literature—the implicit association test or IAT (Greenwald et al. 1998)—which is increasingly used even in economics (Bertrand et al. 2005, Lowes et al. 2015, Carlana 2019, Chetty et al. 2020, Avitzour et al. 2020, Kline et al. 2022, Alesina et al. 2024, Owen & Rury 2025). In the IAT, participants must sort stimuli into categories using two response keys. Their speed is taken as evidence of underlying associations: if participants sort more quickly when two categories share a key than when they do not, this is evidence of an implicit link or association in memory. In our adaptation, the categories are major and career groups, and the stimuli are individual occupation and degree fields. For each round, participants sort words as either belonging vs not to a focal major group (e.g., "Humanities") or as belonging vs not to a focal career group (e.g., "Writers and Journalists"). The key comparison is between "matched" blocks, where the focal major and career share a response key, and "unmatched" blocks, where each is instead paired with the other's alternative category. The difference in completion times across these two blocks provides a measure of an implicit association between the focal major and career.

Using this design, we find that participants strongly and systematically associate majors with their distinctive careers, as predicted by the psychology of stereotyping. Implicit associations were 0.30–0.36 standard deviations higher for distinctive major–career pairs than for non-distinctive pairs ($p < 0.01$). This effect remains (0.24–0.28 SDs, $p < 0.01$) even controlling for the true conditional frequency of careers. Furthermore, a one standard deviation increase in the associations between a major and its distinctive career predicts 4.1 p.p. greater stereotyping as measured by beliefs ($p < 0.01$). This relationship remains large and statistically significant even controlling for career-by-major fixed effects (2.8 p.p., $p < 0.01$); that is, for a given major, those who reveal a greater association between that field of study and its distinctive career also have more biased beliefs. These results strongly suggest that the same psychological mechanisms that drive stereotyping in other settings also underpin the biases about college majors that we uncover. Additionally, psychologists (and increasingly behavioral economists) argue that such associations and the belief biases they lead to are intimately linked to memory processes (e.g., Greenwald & Banaji 1995, Gawronski & Bodenhausen 2011, Bordalo et al. 2016, 2023). Consistent with this idea, we find that heterogeneity across students in their experiences—the careers and majors of people they know personally—strongly predicts biases in their beliefs at the individual level.

Having provided evidence on the severity and origins of stereotyping in our context, we next turn to its implications for economic outcomes and policy. We describe a stylized model of major choice in which students decide their major with imperfect information about the quality of jobs that will be available to them depending on their field of study. The model makes two intuitive points. First, stereotyping leads students to believe that distinctive jobs must be more attractive than they truly are: e.g., that the jobs require less search to find, pay more than they do, or provide better amenities than in reality. Second, stereotyped beliefs lead to greater *misallocation*—given the occupations graduates ultimately end up pursuing, they would have been better off had they majored in a different field. Intuitively, this occurs through two forces. The first is through wages: stereotyping leads more graduates to compete for the distinctive job, which depresses salaries in it. The second is through selection: stereotyping draws in students on the margin who are less well-suited to a major's distinctive job. Larger belief biases therefore cause more graduates to end up in jobs other than those for which their major prepared them.

We test whether stereotyping is correlated with misallocation by compiling data from multiple survey datasets of college graduates. We find that alumni whose majors (in our OSU beliefs data) are the subject of greater stereotyping are significantly more likely to be dissatisfied with their jobs, feel their job does not fit their skills and experience, report their job is unrelated to their field of study, and report that they regret the field of study they chose ($p < 0.05$ for all variables). These results are of course only correlational (and limited to the coarse major groups that we define), but they nonetheless provide suggestive evidence of the potential importance of stereotyped beliefs in distorting post-graduation labor market outcomes.

What policies can help students to make better informed choices? We conclude the paper by analyzing the impact of a light-touch information intervention 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. We find that the effect of information depends critically on students' heterogeneous preferences over their future jobs. In response to news that the distinctive career of their *ex ante* most preferred major is 10 p.p. less likely, students reduce their intentions toward that major by 3.5 p.p. ($p < 0.01$), enroll in fewer courses in that major in the next semester by -0.22 courses ($p < 0.05$), and lower their probability of having declared that major a year later by -6.1 p.p. ($p = 0.23$). Although the effects of our light-touch intervention appear to partially fade out over the following years, it 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 ($p < 0.05$).[3] In contrast to the top-ranked major impacts, reducing stereotyping

[3] 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

if anything boosts the attractiveness of less preferred fields, consistent with heterogeneous preferences over careers: every 10 p.p. reduction in stereotyping about students' second-ranked major *boosts* intentions toward it by +2.1 p.p. ($p = 0.17$), immediate class-taking by +0.20 courses ($p < 0.10$), and major declaration within a year by +9.9 p.p. ($p < 0.01$). All these differences between first and second majors are significant at the $p < 0.05$ level.

This paper contributes to a growing literature on stereotyping by providing field evidence on its importance in a high-stakes setting. Many studies explore stereotypes surrounding race, immigration, political affiliation, and gender (e.g., Bordalo et al. 2019, Coffman et al. 2020, Alesina et al. 2023, Bohren et al. 2023). To our knowledge, we are novel in applying the logic of stereotyping based on a "kernel of truth" to test for (and therefore discover) new stereotypes in a high-stakes economic context. Further, our evidence connecting stereotyping, implicit associations, and role model effects relates to a growing literature on beliefs arising from associations, memory, and what comes easily to mind (e.g., Malmendier & Wachter 2022, Graeber et al. 2022, Bordalo et al. 2023, Augenblick et al. 2023, Enke et al. 2024, Bordalo et al. 2024, Conlon & Kwon 2025).

Our study also adds to a rich literature on beliefs and human capital investment (see Giustinelli 2023 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,[4] 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).[5] 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

with college.

[4]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).

[5]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. Their study does not attempt to test whether students' beliefs are consistent with rational expectations. See also Wiswall & Zafar (2021) and Ersoy & Speer (2025) for beliefs about non-labor-market consequences of major choice.

programs in developing countries greatly overestimate their own employment prospects post-graduation (e.g., Bandiera et al. 2025) and that reducing these misperceptions can have positive effects (e.g., Alfonsi et al. 2025).

2 Do Students Stereotype Majors?

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.[6] 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 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.[7] 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, which vary from year to year, including "Other" and "Undecided." 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.

[6] 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.

[7] 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.

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? We focus on stereotyping based on distinctiveness and 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 $p < 0.001$ level.[8] In Appendix B.1, we additionally bring in data from the CIRP Senior Survey, which allows us to link a subset of our Freshman Survey data to the same students' major and career expectations when they are seniors. This yields 258,134 senior students spanning 1994 to 2015. There we find that, while there is substantial heterogeneity across majors, overall rates of expecting distinctive jobs are largely unchanged from freshman to senior year, falling by only 1.1 percentage points.

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 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

[8]None of these results is driven by ambiguity in the mapping between occupations and career groups. See Appendix B.2 where we explore reclassifying occupations across career groups.

started their career, though our results are not sensitive to this specific age range. 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.[9] 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%).[10] 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 particular majors. 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 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

[9] 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.

[10] 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.

University (OSU). Entering OSU students are automatically enrolled in this program if they have not yet officially declared a major. 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 4). See Appendix B.3 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 our 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 OSU students 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%.[11]

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

[11] This wording, which conditions the belief on students' graduating with each major, avoids another issue in the Freshman Survey Qualitative data: namely, if a student is uncertain about her eventual major, this may add to any uncertainty about her future career. While this presents difficulties in interpreting the Freshman Survey, note that it will naturally tend to push *against* finding stereotyping, as a student who is uncertain whether she will major in (say) biology should be even less likely to report that her probable career is being a doctor.

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.[12] 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, indicating at best 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 distinctive career is substantially (and statistically significantly)

[12]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).

more common among graduates with that major than it actually is.[13] 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 Career-Specific Explanations

Next, students could simply tend to overestimate some careers *unconditional* on major. For example, perhaps many non-distinctive occupations are less well-known (independent of major), and some students might fail to correct for the fact that there are some jobs they have not heard of. 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 career, and μ^i_c are career-by-individual fixed effects.

$$\pi^i_{c|M} = \gamma p_{c|M} + \theta \mathbb{1}\left(c = \text{argmax } \frac{p_{c|M}}{p_{c|-M}}\right) + \mu^i_c + \epsilon^i_{c,M} \tag{1}$$

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 some careers irrespective of major). For example, if students simply neglected non-employment or "other" jobs but otherwise responded only to true frequency,

[13]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%.

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 effect 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 only by a failure to consider leaving the workforce for reasons related to child-care.

These results so far pertain to the career *categories* we have defined, but one might reasonably wonder whether biases specific to more fine-grained occupations drive our results. For example, if students are unaware of many non-distinctive jobs within some categories, and fail to correct for this when forming their beliefs about majors, this could plausibly lead to biases like those we find. In Appendix B.12, we describe a pre-registered survey in which we explicitly test for—and fail to find evidence of—this mechanism. In particular, we ask current college students and recent graduates whether they are aware of the 100 most common non-distinctive occupations. We find that participants are aware of the overwhelming majority of such jobs and that the small amount of unawareness we find is actually *negatively* correlated with stereotyping, the opposite of what we would expect if it were driving our main results. This speaks against unawareness of specific occupations as a driving force behind stereotyping.

Additionally, 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.VIII 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).

Explanatory Power of Stereotyping

Having isolated stereotyping from alternative mechanisms, we next investigate its importance in explaining students' beliefs compared to the other mechanisms we have considered. To do so, we conduct two exercises. First, Table A.IX 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.IX for details). We find that stereotyping distinctive careers appears to be the dominant explanation for the patterns of expectation biases in our survey data, accounting for a greater share of the variation in beliefs than any of the other mechanisms we explore. Second, we examine the extent to which stereotyping appears to underlie the patterns of career expectations we document in the nationally representative Freshman Survey. Specifically, 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 (see Appendix B.7 for more details on this exercise). We find that the implied error based on stereotyping alone from the OSU survey strongly correlates with the actual error in the Freshman Survey (.81 with a p-value < 0.01). We conclude 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.

Persistence of Stereotyping

We have thus far documented stereotyping among freshman students, and particularly among those actively choosing their major. This population captures a policy-relevant group for whom stereotyping poses the largest economic stakes. One might nevertheless wonder whether these mistaken beliefs persist as students age. A natural question is whether belief biases driven by stereotyping disappear over time as students and graduates are exposed to additional information (such as our RCT in Section 4) and experiences. While such learning if it appeared would be consistent with our stereotyping mechanism—which is a theory of students *priors*, not of how they update over time—it might have important independent consequences (e.g., for the sorts of advice students receive from older adults).

Overall, we find strong persistence of stereotyping and little evidence of systematic learning. 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.[14] Further more, 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.

[14] Note that these results also rule out inattention stemming from lack of incentives being the driving factor behind stereotyping in our main surveys, as the same patterns appear when respondents are paid for accuracy.

These persistent biases exist even for graduates' own major, where we might have expected learning to be greatest. Our CloudResearch sample did not include data on participants' own career and major, but in Appendix B.12 we describe a survey of recent college graduates (aged 30 or younger) recruited from Prolific in which we do collect such data. There we replicate our main stereotyping result (19.0 p.p., $p < 0.01$, column 1 of Table B.XI) and show that stereotyping appears even when focusing on graduates' own major (21.4 p.p., $p < 0.01$, column 2). We also find some imprecise evidence that those whose jobs are not distinctive of their major stereotype it somewhat less (-5.2 p.p., $p = 0.24$), despite stereotyping other majors more (+4.4 p.p., $p < 0.05$ for difference between own and other majors). We also find no systematic age gradient: graduates aged 26 and older (the median age among this sample) do not stereotype more or less than younger respondents (Table B.XII). These results provide further evidence that stereotyping remains substantial even after college and even with labor market experience.

2.3 Evidence on Stereotyping from Implicit Associations

We next provide additional data exploring whether stereotyping is the right interpretation of the belief biases we document. We do so by adapting the implicit association test (IAT)—an extermely prevelant measure of stereotyping from the psychology literature (Greenwald et al. 1998)—to our setting. In two pre-registered experiments, we design and conduct IATs to measure whether people implicitly associate majors with their distinctive careers and whether these associations predict belief biases. Experiment 1 asks whether participants exhibit especially strong implicit associations between majors and their distinctive careers. Experiment 2 then asks whether implicit associations between majors and distinctive careers correlate with belief biases at the individual level. See Appendix B.11 for details on this data collection, including incentives, full experimental design, and pre-registration.

Both Experiments 1 and 2 included IATs following a similar design modeled closely off the design from Project Implicit. A typical IAT requires participants to sort words into two pairs of mutually exclusive categories. For example, a test might ask participants to sort distinctively black vs distinctively white names (e.g., "Latonya" vs "Heather") as well as to sort pleasant vs unpleasant words ("lucky" vs "poison"). Different rounds of the test differ according to which categories share a response key. For example, in one round, both distinctively black names and pleasant words need to be sorted to the left (by pressing the "Q" key) while distinctively white names and unpleasant words need to be sorted to the right (by pressing the "P" key). In a second round, these pairings are reversed. A participant then might reveal an implicit association between black names and unpleasant words if she sorts words faster when these two categories are paired together than when black names are paired with pleasant words.

Our IAT measures whether participants implicitly associate certain career groups with certain groups of college majors. To do so, for each of our 11 major groups (including "Other"), we select the individual majors in which at least 0.5% of graduates with a degree in that group had that major. Similarly, for each of our ten career groups (including "Other"), we select the occupations in which at least 0.5% of college graduates with in that career group had that occupation. These word lists (after some cleaning of occupation/major titles, such as removing "Other", "Except", etc) comprise the stimuli that participants needed to sort between groups of majors and/or careers.

Participants underwent multiple rounds of the IAT. Each round had a focal major group (e.g., "Humanities") and a focal career group (e.g., "Writers and Journalists"). The alternative career group and major group in all rounds were "Other Careers" and "Other Majors" (i.e., majors/careers that did not fall into any of the ten major groups/nine career groups). Participants always sorted words to the left or the right using the "Q" and "P" keys. Whenever participants correctly sorted a word in the correct direction, a green check mark appeared and the screen progressed to the next word. Whenever they incorrectly sorted a word, a red X appeared, and participants had to wait one second before trying again. Thus though all participants had to correctly sort all words, incorrect responses slowed them down. Participants were incentivized to complete all rounds as quickly as possible, as the fastest 10% of participants earned an additional $5 bonus.

Each round began with three practice sorting tasks to familiarize participants with the careers and majors in each group (see Appendix B.11 for full details). In the final two sorting tasks—which occurred in a random order—the focal career and major groups either shared a response key (the "matched" task) or they did not (the "unmatched" task). The crucial variable we construct is $\text{Difference}_{i,c,M}$, the difference between these the time taken in unmatched compared to matched task. Participant i is said to have revealed an implicit association between c and M to the extent that $\text{Difference}_{i,c,M}$ is positive. To avoid outliers driving our results, we winsorize this variable at the 95th percentile, and then we standardize it to have mean zero and standard deviation one.

Implicit Associations between Majors and Distinctive Careers. Experiment 1 tests whether participants implicitly associate majors with their distinctive careers. To this end, participants completed four rounds of the IAT. Each round included one of our ten major groups as the focal major, and the focal career was chosen from that major's distinctive career group plus "Teachers" and "Businesspeople" (unless this group is already included as the distinctive career for that major). We chose these focal career/major pairs to ensure that we have measures of associations between each major and its most common careers, allowing us to measure whether participants' have stronger associations between majors and their distinctive careers than other common careers. This process results in 28 pairs of focal careers/majors, and each participant underwent rounds for a randomly chosen four pairs. We recruited 100 participants on Prolific to complete Experiment 1.

Panel A of Table 2 shows OLS regressions where the dependent variable is $\text{Difference}_{i,c,M}$, our measure of i's implicit association between the focal career c and major M. Column 1 regresses this variable on an indicator for whether the round's focal career is the most distinctive career of its focal major. We see that for rounds with distinctive career-major pairs, $\text{Difference}_{i,c,M}$ is 0.36 standard deviations (6.5 seconds) larger than for non-distinctive career-major pairs ($p < 0.01$). Column 2 adds individual-fixed effects, which do not qualitatively change the main result (0.30 SDs, or 5.4 seconds, $p < 0.01$). Columns 3 and 4 add to this regression the true population frequency of the focal career among graduates with the focal major. These frequencies also predict implicit associations between the focal career and major, but the magnitude of this difference is much smaller than for distinctiveness: the coefficient on the distinctiveness dummy is 45 to 55 times as large as that on true frequency, indicating that the boost in implicit associations from distinctiveness is equivalent to a 45 to 55 percentage point boost in true frequency. Finally, columns 5 and 6 add as controls the (winsorized) time participants took in the earlier tasks within each practice sorting task, which have no impact on our main conclusions.[15]

[15]Note that one might have worried that the names of some majors partially overlap with their distinctive careers (e.g., "psychology" and "psychologist"). This worry prompted us to include a sorting task (Part 3, see the Appendix B.11 for details) in which participants must sort just between the focal major and focal career. If it is simply difficult to tell apart members of the focal major and career, and this were driving any

We conclude from these results that participants appear to implicitly associate majors with their distinctive careers, and that the size of these associations are much greater than the true conditional frequency of those careers alone would predict. These results appear to mirror our main results concerning students'/participants' beliefs about the frequency of careers. We now turn to Experiment 2 to explore this connection more directly.

Implicit Associations Predict Beliefs Biases. Experiment 2 tests whether heterogeneity across participants in associations between majors and their distinctive career predicts stereotyping in their beliefs (i.e., overestimation of the distinctive career's conditional frequency). The IAT portion of this experiment was identical to Experiment 1 except that participants did five rounds (instead of four) and the focal career group was always the distinctive career of a randomly chosen major. In addition, participants in Experiment 2 answered population beliefs questions about all ten major groups. The IAT and belief modules of the survey occurred in a random order. We recruited 396 participants on Prolific to complete Experiment 2, which we conducted separately from Experiment 1 to keep the length of each survey manageable.

Panel B of Table 2 shows OLS regressions asking whether participants' population beliefs are predicted by their implicit associations between majors and their distinctive careers, as measured by our IAT. In column 1, we regress population beliefs about $P(c|M)$ on the true conditional frequency, an indicator for whether c is M's most distinctive career, and an interaction between this indicator and participants' (winsorized and normalized) implicit association between M and the distinctive career. The coefficient on this interaction is positive and statistically significant ($p < 0.01$), indicating that participants who reveal a stronger implicit association between a major and its distinctive career have a more exaggerated belief about that career's conditional frequency. The coefficient of 4.06 can be interpreted as saying that every standard deviation increase in implicit associations predicts an increase of 4.06 p.p. in participants' stereotyped beliefs.

A natural question is whether these results are driven by heterogeneity across *majors*—some majors being stereotyped to a greater degree and participants holding stronger implicit associations between those majors/careers—or heterogeneity across *people*. To investigate

of our results, we should expect inclusion of time to complete part 3 to significantly predict $\text{Difference}_{i,c,M}$ and to reduce the coefficient on distinctiveness. In fact, neither of these turn out to be true, suggesting that this is not an important confound.

this, column 4 adds career-by-major fixed effects to these regressions, meaning that all variation in implicit associations and beliefs stems from differences across participants within career/major pairs. We see that the correlation between implicit associations and stereotyping is still large and statistically significant even controlling for these fixed effects: a standard deviation increase in associations predicts an increase of 2.80 p.p. in beliefs ($p < 0.01$).

A second natural question is whether the results presented so far reflect differences across participants in an overall tendency to stereotype majors—and to implicitly associate them with their distinctive careers—or instead reflect within-participant differences across majors. To answer this question, we look at whether stereotyping of a major M is predicted by participants' average implicit association between *other* majors and their distinctive careers. Column 2 shows that these leave-out-mean (LoM) implicit associations are strongly correlated with participants' beliefs ($p < 0.01$). This remains true controlling for career-by-major fixed effects (column 5, $p < 0.01$). Finally, columns 3 and 6 regress beliefs about $\mathrm{P}(c|M)$ on both participants' implicit association between M and its distinctive career as well as the leave-out-mean constructed from that participants' associations for other majors. We see that each variable continues to predict stereotyping conditional on the other ($p < 0.01$ for the this-major association and $p < 0.10$ for the leave-out mean, both with and without career-by-major-fixed effects). We conclude from this analysis that the correlation across participants between associations and stereotypes is driven in part by there being types who stereotype more and have stronger associations between majors and their distinctive careers.

Explicit Associations and Stereotypical Images. The above results show that participants appear to have implicit associations between majors and their distinctive careers and that these associations in part appear to drive stereotyped beliefs. Note that such implicit measures need not reflect *unconscious* attitudes or views (Gawronski et al. 2006), in the sense of people believing they do not hold such associations. Indeed, we show in Appendix B.12 that directly asking participants to select which careers they associate with each major, and to describe in their own words their stereotypical image of graduates with each major, yield very similar findings as the IAT. In particular, participants tend to explicitly associate majors with their distinctive careers, and their stereotypical image of someone with a major tends to include that career. Furthermore, both of these measures are correlated at the individual level with beliefs even controlling for major-by-career fixed effects. Thus the implicit associations we uncover go hand-in-hand with more direct measures of participants' stereotypes.

Heterogeneity by Students' Experiences. Implicit associations like those described above are thought to be tightly linked to memory (e.g., Greenwald & Banaji 1995, Phelps et al. 2000), and recent work in economics has begun to formalize the connections between associations, memory, and belief biases (e.g., Bordalo et al. 2023, Enke et al. 2024, Conlon & Kwon 2025). A basic implication of these ideas is that heterogeneity across students in their experiences should have outsized impacts on their beliefs. In Appendix B.10, we show that beliefs in the Freshman Survey and in our OSU data about the likelihood of careers—both self-beliefs but also (more tellingly) population beliefs—are strongly correlated with the careers and majors of students' parents and other role models. Further, these results appear inconsistent with some natural rational-learning stories: for example, students with a role model in a major's distinctive career—who we might think should have better information about it—stereotype it *more* (6.13 p.p., $p < 0.01$). These results instead look consistent with a simple memory framework in which beliefs depend on who comes readily to mind. See Appendix B.10 for full details.

3 Implications of Stereotyping

What are the implications of stereotyping for labor-market outcomes and for welfare? In this section, we describe and test the predictions of a simple model of college major choice to explore how biases about occupations can distort education choices.

3.1 Setup.

Assume there are two majors $M \in \{A, B\}$ and two careers $c \in \{a, b\}$. For simplicity, we assume that students can only enter job a if they major in A, but students of either major can choose to enter job b. Intuitively, job a requires special training—e.g., being a graphic designer might require having majored in art (A)—while the other job b (e.g., working in sales) is open to people from multiple educational backgrounds but may nonetheless benefit from having majored in an aligned field. Student j's utility *ex post* (i.e., after graduating) from choosing career c is:

$$u_{j,c,M} = \begin{cases} w_{a,M} + \epsilon_j + \psi_j & \text{if } c = a \\ w_{b,M} & \text{if } c = b \end{cases}$$

That is, they earn wage $w_{c,M}$ depending on their major/career and additionally have individual-specific preferences ϵ_j and ψ_j toward job a (we normalize preferences toward job b to be zero). We assume that the timing of the model is as follows: while choosing between majors, the students observes wages and ψ_j but must rely on their beliefs about the distribution of ϵ_j. For simplicity, we assume that $\epsilon_j \sim U(-k+s_k, k+s_k)$, $\psi_j \sim U(-h+s_h, h+s_h)$ and that ϵ_j and ψ_j are independent of each other.

We refer to both ϵ_j and ψ_j as relating to the "amenities" that career a delivers, though other interpretations would deliver identical results. For example, either/both ψ_j and ϵ_j could also be interpreted as also capturing the students' idiosyncratic productivity (and therefore individual-specific wages) in job a or as capturing the difficulty of finding a job in a (conditional on its wage). The crucial distinction between these two objects is that ψ_j captures what the agent knows in advance about a, whereas ϵ_j captures aspects of the job that they learn only after graduating.

This environment implies that students will employ a nested threshold strategy. Working backward, student j will realize that after majoring in A, they will choose career a if $\epsilon_j > T(\psi_j) \equiv w_{b,A} - w_{a,A} - \psi_j$. They will then choose to major in A only if ψ_j—which they observe *ex ante*—is larger than some threshold ψ^*. The value of ψ^* will depend on what they know about the relative attractiveness of a vs b.

While we assume that students are price takers, we allow wages within each occupation to be a decreasing function of the fraction of graduates working in it (i.e., downward sloping labor demand):

$$w_{c,M} = w_{c,M,0} - \gamma p_c,$$

where p_c is the share of graduates working in job c. Finally, we assume that $w_{b,B,0} > w_{b,A,0}+\gamma$, which ensures that students earn more in career b if they have its associated major B than if they had majored in A. Thus, the share $p_{b|A}$ of A majors with career b represents the extent of *ex post misallocation* of human capital investments. Notice however that even students without any *ex ante* biases will have some probability of *ex post* misallocation: those A majors with a sufficiently negative ϵ shock will choose career b.

Discussion of Assumptions. Besides introducing necessary simplifications, the model incorporates two crucial assumptions. The first is that college majors matter for productivity and therefore wages, which is consistent with many studies finding that college majors have large effects on earnings (e.g., Hastings et al. 2013; Kirkeboen et al. 2016; Bleemer & Mehta 2020).

The second crucial assumption is that students care about their future occupation when choosing their major and are at least partially uncertain over their future preferences between occupations. Besides being intuitive, several pieces of suggestive evidence corroborate this assumption. 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.8, we use our 2020 survey data to structurally estimate students' preferences when choosing their major. There, we find that students care both about the expected salary their major will earn them but also have strong (indeed, much stronger) non-pecuniary preferences over the occupations they might pursue depending on their choice of major.

Finally, Table A.X that in two datasets—the National Survey of College Graduates (NSCG) and Survey of Consumer Expectations (SCE)—shows having a job that does not match one's field of study or skills/experience is strongly positively correlated with college graduates' job dissatisfaction ($p < 0.01$ in both samples). This kind of "mismatch" is also negatively correlated with income, but the correlation between it and dissatisfaction appears even controlling for income ($p < 0.01$ in both samples). Note that all these regressions control for major-fixed effects, so these correlations reflect differences within major across respondents who do vs do not experience mismatch. These correlational results are consistent with both pecuniary and non-pecuniary career concerns being important to graduates' welfare.[16]

[16] In addition, the NSCG data also ask respondents who report their job is unrelated to their field of study

3.2 Stereotyping

We allow students to have potentially mis-specified beliefs about two objects. First, they might stereotype major A: that is, they might have an exaggerated belief $\pi_{a|A}$ about the fraction of A majors who choose career a (where we denote the true fraction by $p_{a|A}$). Second, they might have an incorrect belief $\widehat{s_k}$ about the mean s_k of the distribution of ϵ. We assume that students have correct beliefs about the salaries that different careers/majors pay. This assumption is motivated by survey data we collect showing that OSU students' salary beliefs are largely unbiased (see Appendix Section B.3.1). However, this assumption is not crucial, since (as described above) we can also interpret ϵ_j as capturing any unknown part of wages.

Our first result is that any biases in students' beliefs—about the share of graduates' choosing a given occupation on the one hand and about the distribution of amenities in that occupation on the other hand—must be tightly linked (see Section C for all proofs):

Proposition 1 *Students who stereotype major A thereby overestimate the average amenity value of job a. Let $\theta \equiv \pi_{a|A} - p_{a|A}$ be the degree of stereotyping. Then $\widehat{s_k} - s_k = 2k\theta$.*

Proposition 1 follows from the fact that stereotyping is a bias in beliefs about an endogenous object: the share of A majors who *choose* job a. Therefore, a student cannot "only" stereotype: students who believe distinctive jobs are more common than they are must *thereby* believe that some feature of distinctive jobs is more attractive than it is, else they would realize that people do not choose job a so often. In our model, this means that students who stereotype must also have overly optimistic beliefs the amenities of job a.

3.3 Misallocation

We can now ask how choices change as stereotyping grows more severe, as described in Proposition 2:

why they do not: 72.5% report a factor related to pay, amenities, location, or the availability of such a job as being the most important reason, while only 18.5% report a "change in career or professional interest". This suggests that the major driver of such mismatch is related jobs being worse than anticipated, rather than changing preferences (where welfare analyses would be less straightforward).

Proposition 2 *Misallocation increases in stereotyping:*

$$\frac{dp_{b|A}}{d\theta} = \frac{\gamma}{2k}\frac{dp_a}{d\theta} + \frac{1}{2}, \qquad \textit{where} \qquad \frac{dp_a}{d\theta} > 0$$

Misallocation increases in stereotyping for two reasons. First, more people major in A because stereotyping implies overestimating the average amenity shock s_k associated with career a. This increases p_a, the share of graduates choosing career a, which pushes down wages in that job compared to b and therefore persuades some A majors to choose b. Second, the marginal student induced to major in A is negatively selected on ψ and therefore is less likely to end up choosing career a than inframarginal A majors are. Note that *all three* of p_a, $p_{b,A}$, and $p_{b|A}$ increase: stereotyping draws more students into major A, which mechanically increases p_a, but a disproportionate share of them end up (unexpectedly) choosing job b, which increases both $p_{b,A}$ and $p_{b|A}$.

To test Proposition 2, we use multiple data sources to construct six different proxies for misallocation at the major level. First, we use the share of graduates not working in their major's distinctive career in the ACS. Second, we use average responses in the 2013 National Survey of College Graduates (NSCG) to the question "To what extent was your work on your principal job related to your highest degree?", where answers are on a scale from 1 (meaning closely related) to 3 (meaning not related). Third, we use average responses in the New York Fed's Survey of Consumer Expectations (SCE) to the question "On a scale from 1 to 7, how well do you think this job fits your experience and skills?" (we reverse the scale so higher numbers imply poorer fit).

Our next three proxies reflect the idea that misallocation implies a form of regret or dissatisfaction with one's career path. Our fourth proxy is average answers in the 2013 NSCG to the question "How would you rate your overall satisfaction with [your] principal job" on a scale from 1 (meaning very satisfied) to 4 (meaning very dissatisfied). Fifth, we use average answers in the SCE to the question "Taking everything into consideration, how satisfied would you say you are, overall, in your [current/main] job?" on a scale from 1, meaning very dissatisfied, to 5, meaning very satisfied (we reverse the scale so that higher numbers indicate more dissatisfaction). Finally, we use the share of respondents in 2021 Survey of Household Economics and Decision Making who reported that they would have chosen a different field of study if they "could go back and make your education decisions again".

Figure 4 shows that all six of these proxies for misallocation are positively and significantly ($p < 0.05$) correlated at the major level by average stereotyping in our 2020 OSU data (i.e., average population beliefs about distinctive careers minus the truth).[17] In majors where there is greater stereotyping, people are more likely to work in non-distinctive careers (Panel A), report their jobs are more unrelated to their degree (Panel B), report their jobs are a poor fit with their skills and experience (Panel C), report being more dissatisfied with their jobs (Panels D and E), and report wishing they could go back a choose a different field of study (Panel F). All of these correlations are significant at the $p < 0.05$ level. We conclude that these results, though of course only correlational, appear to corroborate the intuitive prediction from the model that stereotyping produces misallocation (and also thereby dissatisfaction and regret about education choices) by pushing people to choose majors whose associated jobs they end up not pursuing.

3.4 Welfare

Finally, we can also ask how overall *ex ante* welfare changes as stereotyping becomes more severe.

Proposition 3 *Stereotyping especially reduces welfare when misallocation leads to larger ex post wage losses. Let $W(\theta)$ be average ex ante welfare:*

$$\frac{dW(\theta)}{d\theta} = \underbrace{-\frac{dp_A}{d\theta}\left[2\theta\sqrt{k(w_{b,B} - w_{b,A})} - k\theta^2\right]}_{\textit{Loss from misallocation}} + \underbrace{\gamma\frac{dp_a}{d\theta}\Big(1 - 2p_a\Big)}_{\textit{Wage Externality}}$$

As is intuitive, Proposition 3 makes clear that stereotyping reduces welfare especially when the productivity stakes of major choice are high. That is, if B majors are much more productive in career b than A majors are, then errors in major choice driven by stereotyping have larger negative consequences.[18]

[17]Note that each dataset has somewhat different categories of majors, and sometimes the mapping between our major groups in the ACS to those in the other datasets is not always perfect or one-for-one.

[18]The wage externality term in Proposition 3 captures our assumption that a student entering career a reduces wages among those who already were choosing it (and boosts wages for career b). Whether this

Proposition 3 suggests that the welfare losses from stereotyping—and therefore the majors for which combating its negative effects may be more urgent—are those where belief biases (θ) are larger and ones where the wage loss from misallocation ($w_{b,B} - w_{a,A}$) is larger. While we can compute average stereotyping across majors from our survey data, note that the wage loss term is counterfactual (how much less is student j earning in her job than she *would have* if she majored in something else) and therefore unobservable. We can nonetheless construct a plausible proxy for this term by looking at wages among graduates not working in their major's distinctive career.

Figure A.I shows for each of our ten major groups average stereotyping in our OSU sample along the x-axis. The y-axis shows average earnings among those without their major's most distinctive career (net of career-fixed effects to account for the fact that alternative careers differ across majors).[19] Proposition 3 suggests that majors to the bottom-right of the figure, with more severe belief biases and less lucrative alternative careers, are ones where stereotyping may be especially harmful on the margin. As is perhaps intuitive, the majors with the largest marginal welfare costs are fields such as communication, art, and psychology, each of which (as we've seen) have very large average biases and also tend to result in lower paying non-distinctive jobs. In contrast, majors like nursing and engineering have lower implied costs from stereotyping, either because the average belief bias is small or because graduates with non-distinctive careers nonetheless earn high salaries.

Note that all of these welfare analyses concern the effect on the *marginal* student: i.e., one who perceives herself to be indifferent between the major she is stereotyping and a less risky alternative major. This analysis therefore suggests that there may be particularly large benefits from nudging such students toward less risky majors.

4 Testing a Light-Touch Policy Intervention

The analysis in the previous section suggested that stereotyping may have substantial long-run labor-market consequences. We now turn to evidence on a light-touch policy in-

improves welfare (ignoring the misallocation term) depends on whether there are more people with a vs b already.

[19]More precisely, we estimate by OLS a regression of earnings among those without their major's most distinctive career on fixed effects for major group and for occupation (note this is more fine-grained than our career groups). Figure A.I then plots the major-group fixed effects on the y-axis.

tervention: a field experiment in which we tested an 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.[20] 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 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

[20]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.

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 ($p < 0.01$).[21] 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. 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 (i.e., the average excluding the student herself) reduction by major in treated students' self-beliefs about their chances of having that major's distinctive career.[22] 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 2 by OLS:

$$Y_{i,M} = \beta \cdot T_i \cdot AvgReduction_{i,M} + \alpha_1 Intentions_{i,M,pre} + \alpha_2 Classes_{i,M,pre} + \epsilon_{i,M} \tag{2}$$

In the above equation, $Y_{i,M}$ is an outcome variable of interest for a given major M for student i. T_i is a dummy variable for treatment status. $AvgReduction_{i,M}$ is the measure of treatment

[21] 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.

[22] More precisely, we compute, for every student i, the average reduction in self-beliefs about $P(c|M)$ among all students other than i. For students in the control group, this is simply the average reduction, since control-group students were not asked to update these self-beliefs. For treated students, we exclude i herself when computing these leave-out means for them.

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 5 shows the coefficient on the interaction term from equation 2 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 dark blue dot shows that 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., $p < 0.05$). The next six dark blue dots in Figure 5 show analogous results but where the dependent variable is the number of classes students took in their top major every semester between Spring 2022 (immediately post treatment) and Fall 2024. We find very similar initial 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, $p < 0.05$) in the semester immediately following treatment.[23]

In contrast to these effects on top-ranked majors, the light blue dots in Figure 5 show that, if anything, reducing beliefs about attaining the distinctive job of students' second ranked major directionally *increases* intentions toward it (coefficient $= 0.66$ SDs, $p = 0.17$) and immediate class-taking in it (coefficient $= 0.21$ SDs, or about 0.20 classes, $p < 0.10$). Both of these coefficients are statistically significantly different from the analogous results regarding top-ranked majors ($p < 0.01$ for intentions and $p < 0.05$ for classes). While these results may appear surprising, in Appendix Section B.8 we show that they are consistent with a simple counterfactual simulation given estimates of students' heterogeneous preferences over careers: in particular, this is the pattern we would expect if students have much stronger preferences toward their top-ranked major's distinctive career than that of their second-ranked major.[24]

[23]We are underpowered to investigate which if any majors are driving this result, but we do not see large differences in this coefficient if we drop any individual major. We can also investigate major switching in the administrative data. We find that the magnitudes of this phenomenon are far too small to meaningfully impact our results, however: by the eighth semester of college, only 2.46 percent of students have ever switched their declared major across our major groups.

[24]Of course, while these patterns are consistent with this story, there may be other explanations as well.

The rightmost dots in Figure 5 show that both treatment effects eventually fade, becoming smaller and statistically insignificant by Fall 2024, three years after treatment. 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 2025).[25]

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.XIV and A.XV 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 intervention often had small (and for some majors null) effects on what jobs students expected themselves to have. However, column 1 of Table A.XVI 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.XVI).

[25]Note that all these results exclude from each semester's data students who are no longer taking classes at OSU. Table A.XIII shows very similar results if we instead impute zeros for such students.

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. These stereotypes in students' beliefs are strongly mirrored in their implicit associations between majors and careers, and appear partly driven by past experiences. A stylized model predicts—and empirically we confirm—that greater stereotyping produces misallocation, with more graduates ended up in jobs unrelated to their field of study. 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.

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. 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 (Enke et al. 2023).

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).[26] *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,

[26] 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.

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 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., Carlana, M., La Ferrara, E., & Pinotti, P. (2024). Revealing stereotypes: Evidence from immigrants in schools. *American Economic Review*, *114*(7), 1916–1948.

Alesina, A., Miano, A., & Stantcheva, S. (2023). Immigration and redistribution. *The Review of Economic Studies*, *90*(1), 1–39.

Alfonsi, L., Namubiru, M., & Sara, S. (2025). 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*.

Avitzour, E., Choen, A., Joel, D., & Lavy, V. (2020). *On the origins of gender-biased behavior: The role of explicit and implicit stereotypes* (Tech. Rep.). National Bureau of Economic Research.

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. (2025). The search for good jobs: evidence from a six-year field experiment in uganda. *Journal of labor economics*, *43*(3), 000–000.

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.

Bertrand, M., Chugh, D., & Mullainathan, S. (2005). Implicit discrimination. *American Economic Review*, *95*(2), 94–98.

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*, 1–45.

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. H. (2023). *How are gender norms perceived?* (Tech. Rep.). National Bureau of Economic Research.

Carlana, M. (2019). Implicit Stereotypes: Evience from Teachers' Gender Bias. *The Quarterly Journal of Economics*, 1163–1224.

Chan, A. (2024). Discrimination against doctors: A field experiment.

Chetty, R., Hendren, N., Jones, M. R., & Porter, S. R. (2020). Race and economic opportunity in the united states: An intergenerational perspective. *The Quarterly Journal of Economics*, *135*(2), 711–783.

Coffman, K., Collis, M. R., & Kulkarni, L. (2023). Stereotypes and belief updating. *Journal of the European Economic Association*, jvad063.

Coffman, K., Exley, C. L., & Niederle, M. (2020). 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. (2025). Memory rehearsal and belief biases. *Working paper*.

Conlon, J. J., & Kwon, S. Y. (2025). Beliefs from cues.

de Koning, B. K., Dur, R., & Fouarge, D. (2025). 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., Graeber, T., & Oprea, R. (2023). Confidence, self-selection, and bias in the aggregate. *American Economic Review*, *113*(7), 1933–1966.

Enke, B., Schwerter, F., & Zimmermann, F. (2024). Associative memory, beliefs and market interactions. *Journal of Financial Economics*, *157*, 103853.

Ersoy, F., & Speer, J. D. (2025). Opening the black box of college major choice: Evidence from an information intervention. *Journal of Economic Behavior & Organization*, *231*, 106800.

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.

Gawronski, B., & Bodenhausen, G. V. (2011). The associative–propositional evaluation model: Theory, evidence, and open questions. *Advances in experimental social psychology*, *44*, 59–127.

Gawronski, B., Hofmann, W., & Wilbur, C. J. (2006). Are "implicit" attitudes unconscious? *Consciousness and cognition*, *15*(3), 485–499.

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

Giustinelli, P. (2023). Expectations in education. *Handbook of economic expectations*, 193–224.

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

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychological review*, *102*(1), 4.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of personality and social psychology*, *74*(6), 1464.

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*.

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

Kline, P., Rose, E. K., & Walters, C. R. (2022). Systemic discrimination among large us employers. *The Quarterly Journal of Economics*, *137*(4), 1963–2036.

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.

Lowes, S., Nunn, N., Robinson, J. A., & Weigel, J. (2015). Understanding ethnic identity in africa: Evidence from the implicit association test (iat). *American Economic Review*, *105*(5), 340–345.

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*.

Owen, S., & Rury, D. (2025). Implicit gender-stem stereotypes and college major choice.

Phelps, E. A., O'Connor, K. J., Cunningham, W. A., Funayama, E. S., Gatenby, J. C., Gore, J. C., & Banaji, M. R. (2000). Performance on indirect measures of race evaluation predicts amygdala activation. *Journal of cognitive neuroscience*, *12*(5), 729–738.

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.

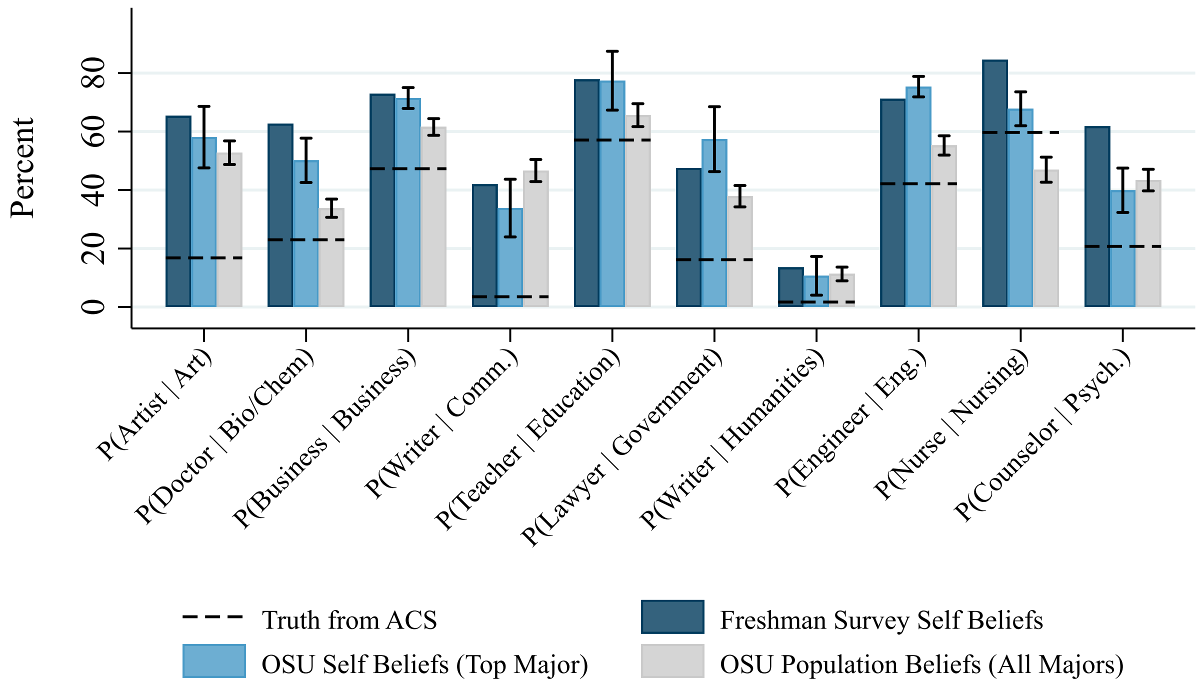
Stantcheva, S. (2023). How to run surveys: A guide to creating your own identifying variation and revealing the invisible. *Annual Review of Economics*, *15*(1), 205–234.

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.

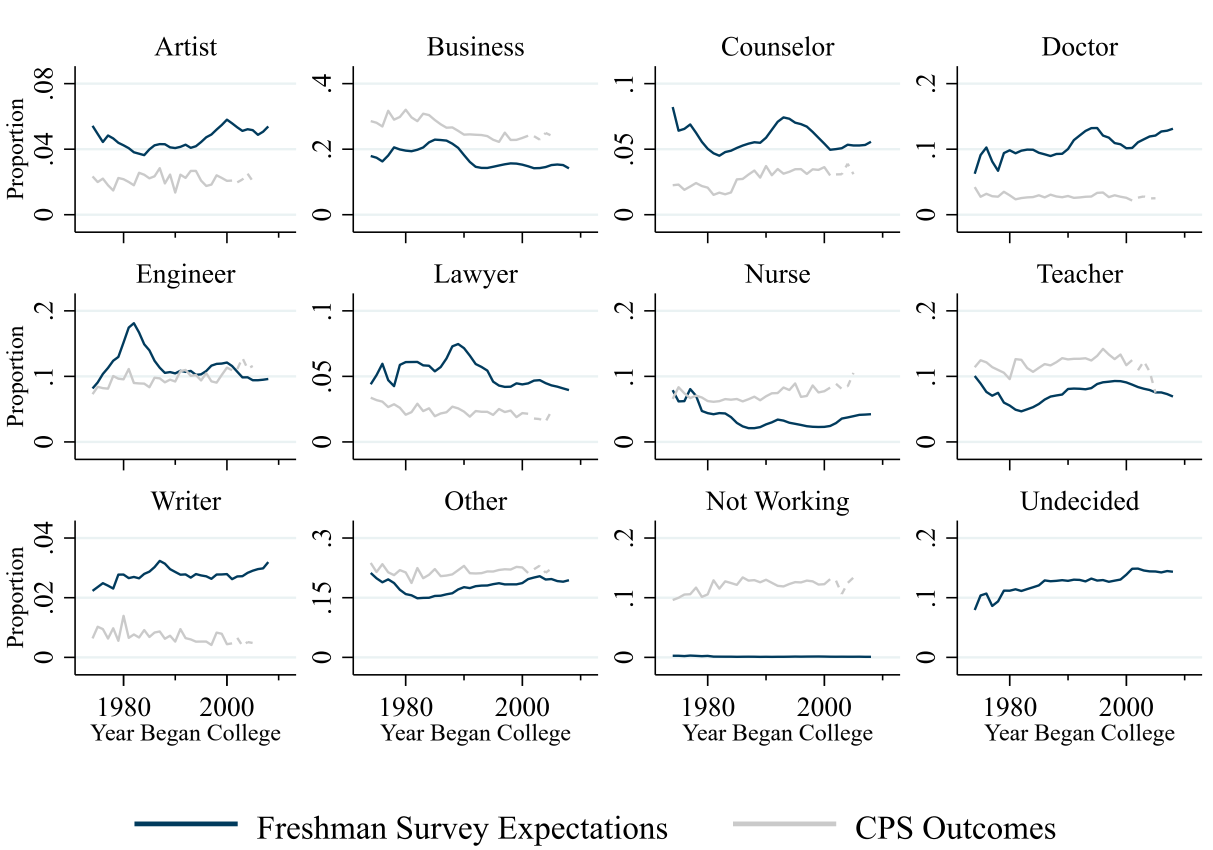
Wiswall, M., & Zafar, B. (2021). Human capital investments and expectations about career and family. *Journal of Political Economy*, *129*(5).

Figure 1: Stereotyping Distinctive Careers



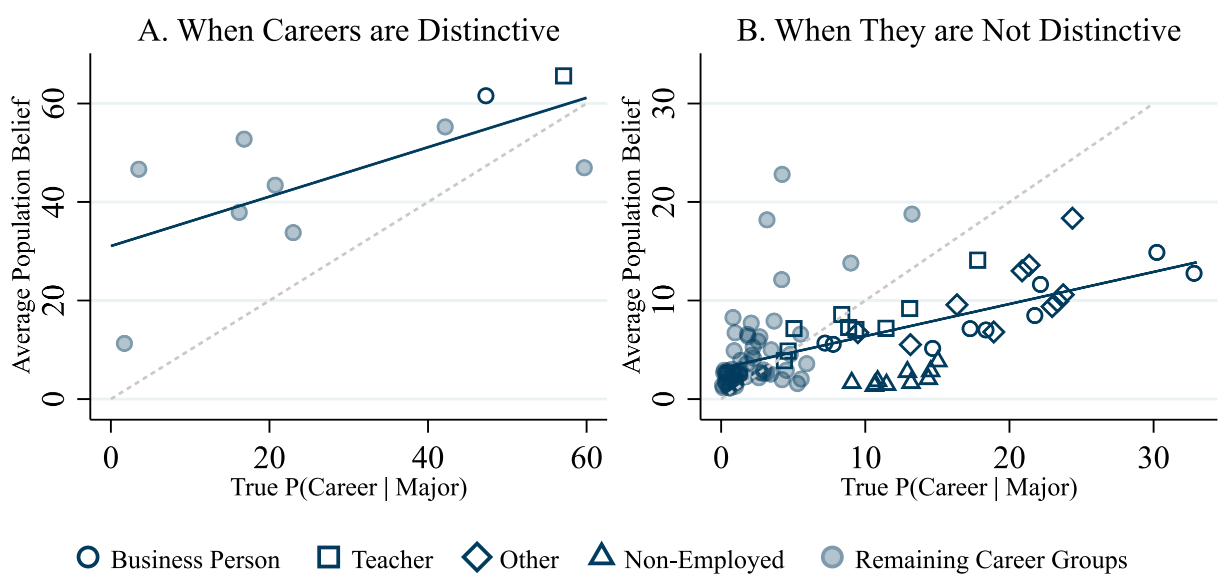
Notes: Figure 1 presents average statistics regarding the most distinctive career (as defined in section 2) for each major. The dashed horizontal lines denote the actual proportion of college graduates with each major between the ages of 30 and 50 that are working in that major's most distinctive career, 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



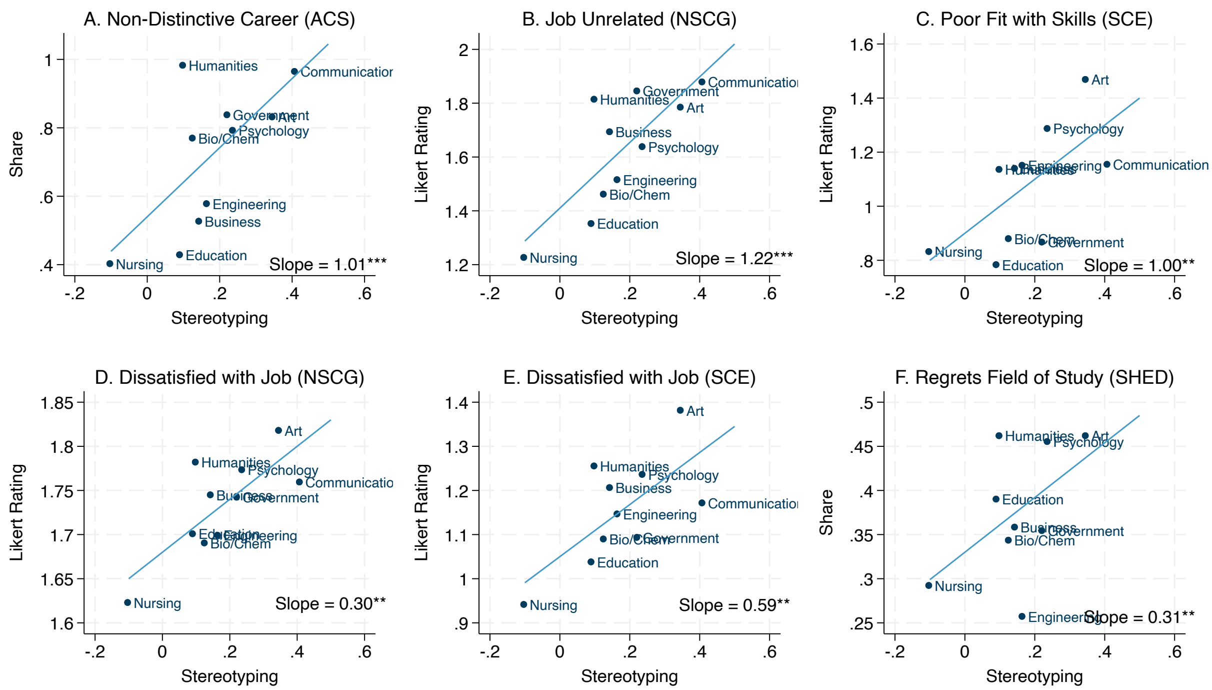
Notes: Figure 2 compares the share of first-year undergraduates by the year they began college (birth year plus 18) who expect to have a career in each occupation (blue line) according to the Freshman Survey data along with the share of college graduates (gray line) aged 33 to 37 who work in that occupation 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



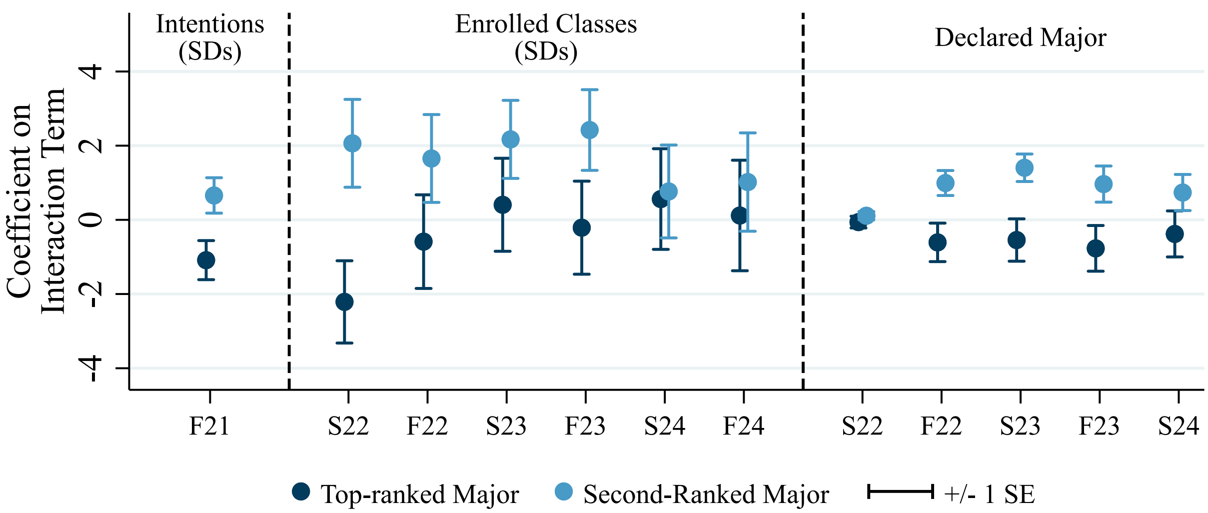
Notes: Each dot in Figure 3 represents a career-major pair (where non-employment is also one the "careers"). Panel A restricts these pairs to when the career is most distinctive of the major, and Panel B restricts them to when the career is not distinctive of the major. The x-axis of both panels is the 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: Stereotyping Predicts Misallocation



Notes: The x-axis in each panel of Figure 4 shows average stereotyping of each major (population beliefs minus truth from the 2020 OSU data). The y-axes in each panel are: the share of ACS respondents who are not working in that major's distinctive career (Panel A); the average rating (on a scale from 1 to 3) of how unrelated employed NSCG respondents' job is to their highest degree (Panel B); the average rating (on a scale from 1 to 7) of how well employed SCE respondents say their job fits their skills and experience, reverse coded so high ratings indicate a poor fit (Panel C); the average rating (on a scale from 1 to 4) employed NSCG respondents give to how satisfied their are with their job, coded so that high ratings indicate dissatisfaction (Panel D); the average rating (on a scale from 1 to 5) employed SCE respondents give to how satisfied their are overall with their job, coded so that high ratings indicate dissatisfaction (Panel E); and the share of SHED respondents who report that they would have chosen a different field of study if they could go back and make their education decisions again. The slope reported is a the coefficient on stereotyping in a univariate OLS regression. *, **, and *** indicate significance at the 10%, 5%, and 1% levels.

Figure 5: Effect of Information



Notes: Each dot in Figure 5 shows the estimate for the interaction term in equation 2 in a separate OLS regression. Dark blue dots indicate regressions where M in equation 2 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	Men	Women	Non-URM	URM	FG	Non-FG	Non-College	College	Younger	Older
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(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.

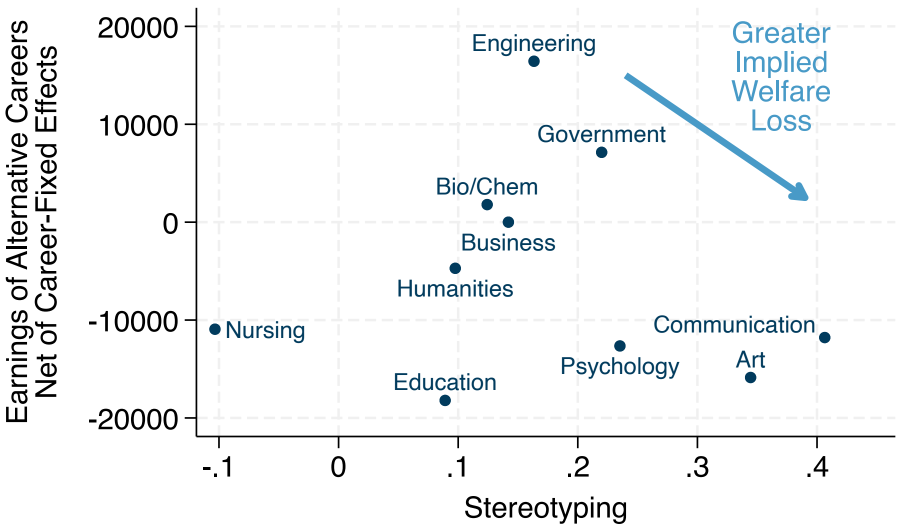
Table 2: Implicit Associations

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Dependent Variable = Difference$_{i,c,M}$						
$\mathbb{1}$(Most Distinctive)	0.36***	0.30***	0.28***	0.24***	0.30***	0.23***
	(0.09)	(0.07)	(0.08)	(0.06)	(0.09)	(0.08)
P(Career | Major)			0.01**	0.00*	0.01**	0.00
			(0.00)	(0.00)	(0.00)	(0.00)
Part 1 Time					0.01	0.06
					(0.05)	(0.12)
Part 2 Time					0.08*	0.12
					(0.04)	(0.08)
Part 3 Time					-0.05	0.06
					(0.07)	(0.11)
Constant	-0.12*	-0.10***	-0.20**	-0.16***	-0.22**	-0.13**
	(0.06)	(0.02)	(0.08)	(0.04)	(0.10)	(0.06)
Observations	400	400	400	400	400	400
Individuals	100	100	100	100	100	100
Individual Fixed Effects	No	Yes	No	Yes	No	Yes
Panel B: Dependent Variable = Population Belief of P(Career | Major)						
$\mathbb{1}$(Most Distinctive) x Difference$_{i,M}$	4.06***		2.69***	2.80***		0.82***
	(0.71)		(0.66)	(0.68)		(0.26)
$\mathbb{1}$(Most Distinctive) x LoM Difference$_{i,M}$		3.25***	1.93*		3.23***	2.71***
		(0.94)	(1.10)		(0.95)	(1.01)
$\mathbb{1}$(Most Distinctive)	16.42***	15.74***	15.85***			
	(3.30)	(4.13)	(3.27)			
P(Career | Major)	0.35***	0.38***	0.36***			
	(0.07)	(0.09)	(0.07)			
Observations	21,780	43,560	21,780	21,780	43,560	21,780
Individuals	396	396	396	396	396	396
Individual-by-Career Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Major-by-Career Fixed Effects	No	No	No	Yes	Yes	Yes

Notes: Panel A of Table 2 presents OLS estimates where the dependent variable is the difference between the time participants' took to complete Parts 5 and 4 (Unmatched Pairing minus Matched Pairing) of the implicit association test in Experiment 1. We winsorize this variable at the 95th percentile and then normalize it to have mean zero and a standard deviation of one. "P(Career | Major)" is the true percent of graduates with the the IAT round's focal major that are working in its focal career, calculated from the 2017-2019 American Community Survey. $\mathbb{1}$(Most Distinctive) is a dummy variable indicating whether the focal career is the most distinctive outcome for the focal major. The time control variables are the times that participants took to complete parts 1 to 3 of the same IAT round (winsorized at the 5th and 95th percentiles and then normalized). Panel B presents OLS estimates where the dependent variable is Experiment-2 participants' population beliefs about the fraction of graduates with a certain major with a certain career. Difference$_{i,M}$ is (winsorized and normalized) difference between the time participants took to complete Parts 5 and 4 (Unmatched Pairing minus Matched Pairing) of the implicit association test for that major and its distinctive career. The leave-out-mean (LoM) difference is calculated by taking the average of Difference$_{i,M'}$ for all other majors M' that participant took the IAT for. All regressions cluster standard errors at the individual level and at the career-by-major level. *, **, and *** indicate significance at the 10%, 5%, and 1% levels.

A Online Appendix: Supplementary Figures and Tables

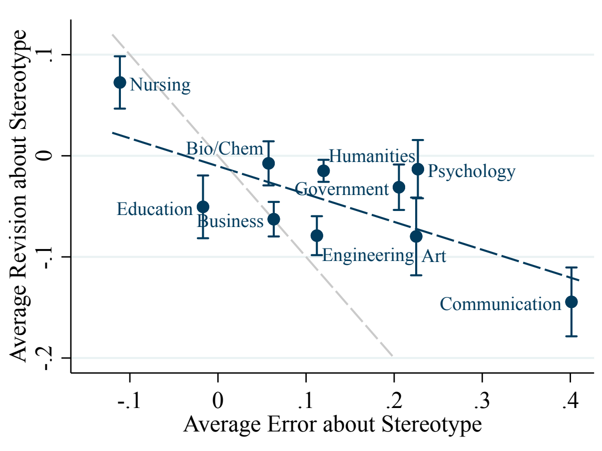
Figure A.I: Implied Welfare Costs of Stereoptying



Notes: The x-axis of Figure A.I shows average stereotyping of each major (population beliefs minus truth from the 2020 OSU data). The y-axis shows average earnings in the ACS among graduates with that major but without that major's distinctive career, net of career-fixed effects. The arrow indicates the direction in which the model from Section 3 predicts that the marginal welfare costs of stereotyping are larger.

Notes:

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: 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.VIII 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.IX: 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.IX 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 3 and 4 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 3 and only the 2021 OSU sample for equation 4.

$$\begin{aligned}
Y_{i,c,m,p} - T_{c,m} = \psi_{c,p,Top(i,m)} &+ \beta_1 Self_{i,p} + \beta_2 t_{i,m} + \beta_3 Self_{i,p} \times Top_{i,m} + \beta_4 Dist_{m,c} + \beta_5 Dist_{m,c} \times Self_{i,p} + \\
&\beta_6 Dist_{m,c} \times Top_{i,m} + \beta_7 Dist_{m,c} \times Self_{i,p} \times Top_{i,m} + \beta_8 T_{c,m} + \beta_9 T_{c,m} \times Self_{i,p} + \beta_{10} T_{c,m} \times Top_{i,m} \\
&+ \beta_{11} T_{c,m} \times Self_{i,p} \times Top_{i,m} + \varepsilon_{i,c,m,p} \qquad (3)
\end{aligned}$$

$$\begin{aligned}
Y_{i,c,m,p} - T_{c,m} = \psi_{c,p} &+ \beta_1 Self_{i,p} + \beta_2 t_{i,m} + \beta_4 Dist_{m,c} + \beta_5 Dist_{m,c} \times Self_{i,p} + \beta_8 T_{c,m} + \beta_9 T_{c,m} \times Self_{i,p} + \\
&\beta_{1,2} RM_{i,c,m} + \beta_{1,3} RM_{i,c,m} \times Self_{i,p} + \varepsilon_{i,c,m,p} \qquad (4)
\end{aligned}$$

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.X: Correlation between Job Satisfaction, Salary, and Job Mismatch

	NSCG			**SCE**		
	(1) Dissatisfied	(2) Salary	(3) Dissatisfied	(4) Dissatisfied	(5) Salary	(6) Dissatisfied
Mismatch	0.20***	-0.86***	0.16***	0.54***	-1.12***	0.53***
	(0.01)	(0.03)	(0.01)	(0.02)	(0.08)	(0.02)
Salary ($10ks)			-0.04***			-0.01***
			(0.00)			(0.00)
Constant	-0.02**	6.72***	0.26***	0.00	7.97***	0.10***
	(0.01)	(0.03)	(0.02)	(0.02)	(0.09)	(0.03)
N	84,045	83,896	83,896	2,832	2,794	2,794

Notes: Table A.X presents OLS regressions using the 2013 round of the National Survey of College Graduates (NSCG, columns 1-3) and the 2017-2020 Survey of Consumer Expectations (SCE, columns 4-6). The dependent variable in columns 1, 3, 4, and 6 is a likert scale (normalized to have mean zero and standard deviation one) indicating dissatisfaction with respondents' job. The dependent variable in columns 2 and 5 is respondents' annual salary ($10ks). "Mismatch" is a normalized likert indicating the extent to which the respondent said their job was not related to their highest degree (NSCG) or did not fit their skills/experience (SCE). We restrict the data to employed college graduates. All regressions control for major-fixed effects. Robust standard errors in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels.

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 2 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 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.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 2 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 (with Dropouts)

	Enrolled Classes (SDs)					
	S22	F22	S23	F23	S24	F24
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Top-Ranked Major						
Treatment x Average Reduction in $\pi_{c\|M}$	-2.32**	-1.48	-0.03	-1.51	-0.66	-0.50
	(1.10)	(1.24)	(1.22)	(1.22)	(1.33)	(1.39)
Treatment	0.07	-0.05	-0.06	0.03	0.07	0.00
	(0.07)	(0.08)	(0.08)	(0.08)	(0.09)	(0.10)
Average Reduction in $\pi_{c\|M}$	3.27***	3.22***	2.85***	2.76***	0.80	-0.56
	(0.75)	(0.83)	(0.86)	(0.85)	(0.94)	(0.97)
Pre-Treatment Belief P(M)	0.84***	1.54***	1.52***	1.59***	1.42***	0.90***
	(0.15)	(0.16)	(0.15)	(0.16)	(0.16)	(0.16)
Pre-Treatment Classes	0.38***	0.16***	0.09**	0.02	0.09*	0.14***
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Constant	-0.65***	-0.86***	-0.78***	-0.76***	-0.69***	-0.36***
	(0.11)	(0.10)	(0.10)	(0.11)	(0.11)	(0.12)
Panel B: Second-Ranked Major						
Treatment x Average Reduction in $\pi_{c\|M}$	2.01*	1.69	2.34**	2.40**	0.87	0.76
	(1.17)	(1.16)	(1.02)	(1.03)	(1.12)	(1.04)
Treatment	-0.08	-0.09	-0.08	-0.04	0.00	0.04
	(0.07)	(0.07)	(0.06)	(0.06)	(0.07)	(0.07)
Average Reduction in $\pi_{c\|M}$	-0.64	-0.58	-0.40	-0.61	-0.73	-0.67
	(0.75)	(0.81)	(0.59)	(0.60)	(0.77)	(0.65)
Pre-Treatment Belief P(M)	0.84***	0.90***	1.00***	0.95***	1.04***	0.98***
	(0.22)	(0.21)	(0.19)	(0.18)	(0.21)	(0.21)
Pre-Treatment Classes	0.34***	0.25***	0.12***	0.11***	0.03	-0.00
	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)
Constant	-0.53***	-0.54***	-0.57***	-0.58***	-0.51***	-0.46***
	(0.07)	(0.06)	(0.06)	(0.05)	(0.06)	(0.06)
Observations	783	783	783	783	783	783
p-value: 1st vs 2nd Major Interaction Equal	0.007	0.062	0.134	0.014	0.379	0.470

Notes: Table A.XIII presents OLS regressions of equation 2 including data from the 2021 OSU sample. The dependent variable is the number of classes they enrolled in M during Spring semester 2022 (S22), Fall 2022 (F22), etc., where students who dropped out are assigned 0. "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.XIV: 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.XIV presents OLS regressions of equation 2, 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.XV: 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 2, 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.XVI: 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.XVI 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.

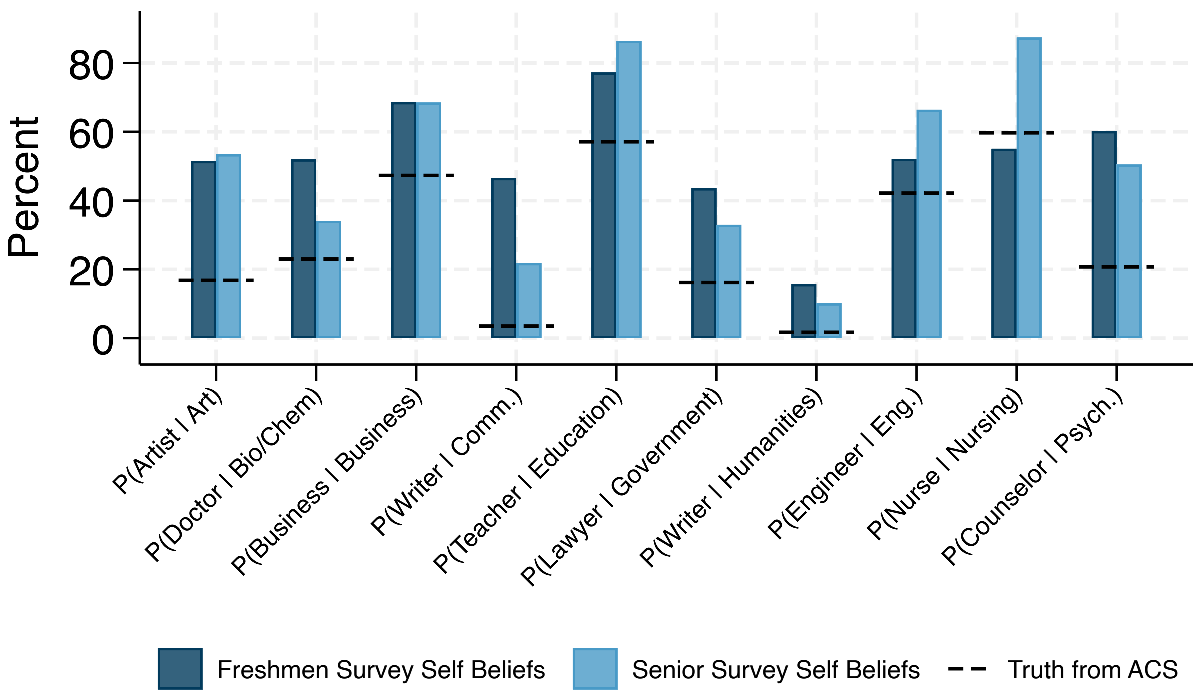
B Online Appendix: Data and Additional Analyses

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

B.1 Matched Freshman and Senior CIRP Data

We match data from the CIRP Senior Data to the Freshman Survey. The matched sample includes 258,134 respondents with non-missing major and career expectation data in both surveys. These data cover the 1994 to 2015 Senior Surveys. We define senior's major as their primary major (they are allowed to mark a secondary major as well). Figure B.I shows the share of respondents in each survey who report their probable career to be their major's distinctive career. While there is some variation across majors, the overall share expecting their major's distinctive career remains largely unchanged: 47.4% in the Freshman Survey and 46.3% in the Senior Survey (though this difference is of course statistically significant given the large sample size, $p < 0.01$). Note that these are self-beliefs (students' beliefs) about what their own future career, which are the only expectations data the Freshman and Senior Surveys collect. See Section B.12 for evidence on how population beliefs change over time.

Figure B.I: Comparing the Freshman and Senior Surveys



Notes: Figure B.I 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. These data are restricted to students who also respond to the Senior Survey. The light blue bars show, among students in the Senior Survey who expect to pursue each major, what fraction list that major's distinctive outcome as their probable career occupation.

B.2 Robustness of Career Classifications

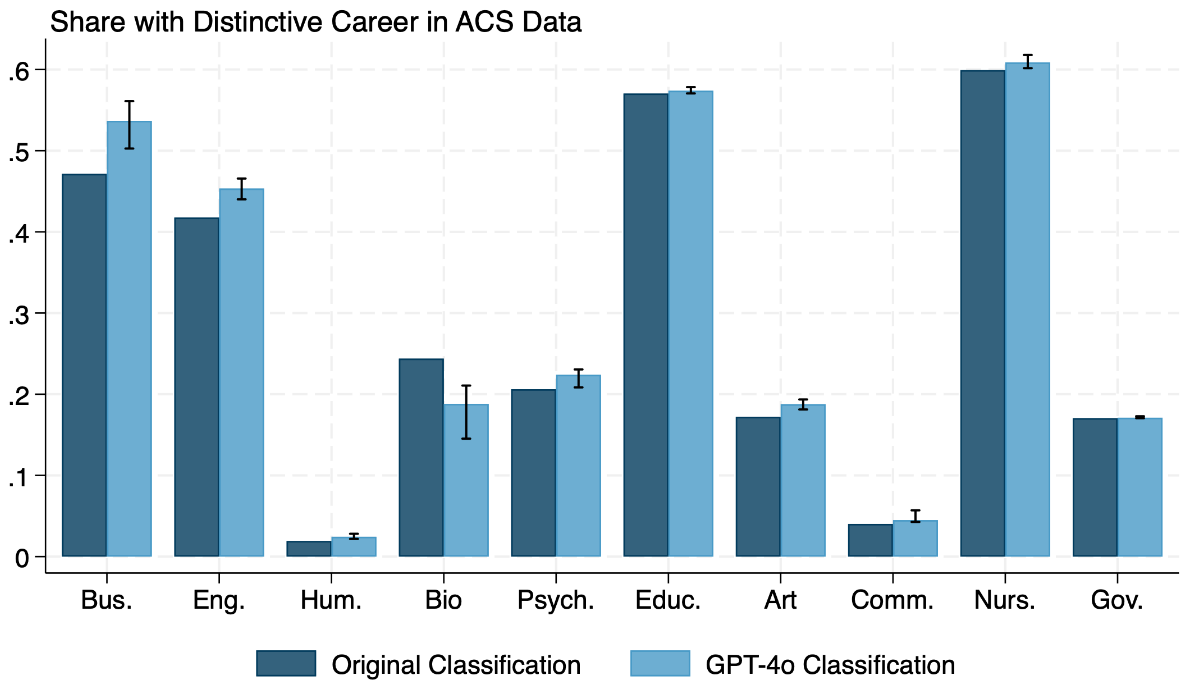
A natural question is how sensitive our main stereotyping result is to the way we classify occupation codes into career groups. To explore this, we use the OpenAI API to have gpt-4o (a large language model) sort each occupation from the ACS into our career groups. More specifically, for each occupation title, we give it the following prompt:

```
System message: You are classifying occupation titles into exactly one career group.
Output strict JSON with keys 'career_group_name' and 'confidence' (0..100).

Allowed career groups:
[List of our 11 career groups, including "Other"]

Return ONLY strict JSON: {"career_group_name":"<one of the allowed groups>",
"confidence":0..100}

User message: Title:
{occupation title}
Respond with JSON only.
```

We then return 50 responses for each occupation with temperature set to 2.0 (to measure the extent to which the model returns variable answers).

Figure B.II shows the share of each major's graduates whom we code as working in its distinctive career compared to the share gpt-4o codes as doing so. Our first result is that these two shares are extremely correlated: the correlation coefficient across majors is 0.988. Note that this high correlation appears despite the model often disagreeing with our classification: for only 38% of occupations does the model always sort it into the same career group as our original classification. Rather, these disagreements tend to be over rare occupations. The average occupation frequency over which our and the model's classifications always agree is 0.38%, compared to only 0.12% for occupations with at least one disagreement.

Figure B.II: Robustness of Career Categorization



Notes: Figure B.II compares the fraction 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, according to our career classifications (dark blue bars) and gpt-4o's (light blue bars). Whiskers show the central 90% of estimates from the multiple queries of gpt-4o.

B.3 Details on 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.[27] 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

These surveys focused on the 10 major groups described in Section 2.1. Whenever the

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

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.[28] 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 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".[29] However, in open-ended feedback the overwhelming reason given was that they took longer to fill out than multiple choice questions would have.[30] 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.

[28]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.

[29]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.

[30]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.

B.3.1 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.[31]

Table B.I shows the average beliefs about expected salary by major and by career among the 2020 OSU sample. While there are some differences between average perceived and actual salaries, they are generally much smaller than the errors in beliefs about the frequency of careers that we primarily focus on. For example, the largest difference (in absolute terms) is that the average student underestimates the expected salary of humanities majors by about 15%.

Nonetheless, students' beliefs about salaries are are tightly linked to their beliefs about the frequency of careers conditional on major. Denote the expected value of the elicited distribution of salaries students' "direct" salary beliefs. Column 1 of Table B.II regresses direct self beliefs on direct population beliefs. We see a robust positive relationship, which persists in Column 2 after adding major- and individual-fixed effects. These results show that students' earnings expectations, depending on major, are tightly linked to the salaries that they think others earn. Column 3 of Table B.II replaces direct population beliefs with what we call students' "implied" population beliefs. To construct these, we first take the

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

actual mean salary from the ACS for each major-career pair. We then take a weighted average of these using students' beliefs about their likelihood of having each career conditional on major. Thus, their implied population beliefs are the average salaries that follow from their beliefs about the likelihood of careers, assuming they know the average salary of each career. We see in column 3 that implied population beliefs are very predictive of direct self beliefs, and this relationship persists in Column 4 after we add major- and individual-fixed effects. Intuitively, these results show that students expect to earn a high salary with a major when they think higher-paying careers are more common among that major's graduates.

Table B.I: Beliefs about Salaries in 2020 OSU Survey

Majors				Careers			
	ACS	PB OSU	SB OSU		ACS	PB OSU	SB OSU
Art	69,681	60,511	69,266	Artist	74,651	64,065	67,854
		(1,604)	(2,214)			(1,908)	(2,287)
Bio/Chem	97,057	94,524	96,771	Business	98,464	92,031	95,266
		(1,632)	(1,733)			(1,314)	(1,517)
Business	89,083	89,266	93,564	Counselor	61,516	66,613	70,724
		(1,319)	(1,491)			(1,368)	(1,600)
Communication	79,123	69,120	74,867	Doctor	129,837	125,054	123,295
		(1,592)	(1,805)			(1,598)	(1,774)
Education	61,501	58,961	64,066	Engineer	104,205	104,861	103,190
		(1,372)	(1,680)			(1,397)	(1,603)
Government	93,965	88,321	90,317	Lawyer	110,042	110,737	110,521
		(1,843)	(2,068)			(1,621)	(1,699)
Humanities	78,365	66,651	73,633	Nurse	77,607	82,677	87,541
		(1,548)	(2,110)			(1,416)	(1,725)
Engineering	103,011	97,445	97,490	Teacher	62,779	58,059	62,583
		(1,457)	(1,686)			(1,248)	(1,449)
Nursing	80,182	90,558	94,243	Writer	77,487	66,869	69,982
		(1,683)	(1,773)			(1,496)	(1,993)
Psychology	71,549	72,321	80,695	Other	66,258		
		(1,470)	(1,837)				
Other	77,383						

Notes: Table B.I compares salary beliefs of students in our 2020 OSU sample to average salaries calculated from the 2017-2019 ACS. The left panel shows (beliefs about) average salary conditional on major, while the left panel shows (beliefs about) average salary by career. “SB” indicates that the beliefs are OSU students’ “self-beliefs,”, i.e., the expected value of their beliefs of what they would earn conditional on having that major or career. “PB” indicates their “population beliefs”, i.e., the expected value of their beliefs about the distribution of earnings of those with that major or occupation. Standard errors in parentheses.

Table B.II: Connecting Career and Salary Beliefs

	Dependent Variable: Direct Self Beliefs			
	(1)	(2)	(3)	(4)
Direct Population Belief	0.82***	0.66***		
	(0.03)	(0.03)		
Implied Population Belief			0.81***	0.54***
			(0.04)	(0.06)
Constant	19.06***	31.75***	15.36***	37.85***
	(2.46)	(2.55)	(3.40)	(5.32)
Observations	3,020	3,020	3,016	3,015
Individuals	755	755	755	754
R^2	0.55	0.85	0.19	0.76
Major Fixed Effects	No	Yes	No	Yes
Individual Effects	No	Yes	No	Yes

Notes: Table B.II presents OLS regressions, where the dependent variable is the 2020 OSU sample's beliefs about their expected salary conditional on graduating with each major. "Direct Population Belief" is students' belief about the average salary of graduates with that major aged 30-50. "Implied Population Belief" is constructed by taking a weighted average of actual average salaries (calculated in the ACS) for each occupation (conditional on major), where the weights are each student's beliefs about the fraction of graduates with that major who have that occupation. All regressions cluster standard errors at the individual level. *, **, and *** indicate significance at the 10%, 5%, and 1% levels.

B.4 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.5 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.6 2013 National Survey of College Graduates

In Section B.8, 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.7 Implied Error Analysis

To formalize the comparison between the Ohio State surveys and the Freshman Survey, 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.[32] We calculate what we call the "implied error" about the probability of working in career c as shown in equation 5, where p_c is the true fraction working in c.

$$\text{ImpliedError}_c = \sum_M \kappa_M \pi^i_{c|M} - p_c \tag{5}$$

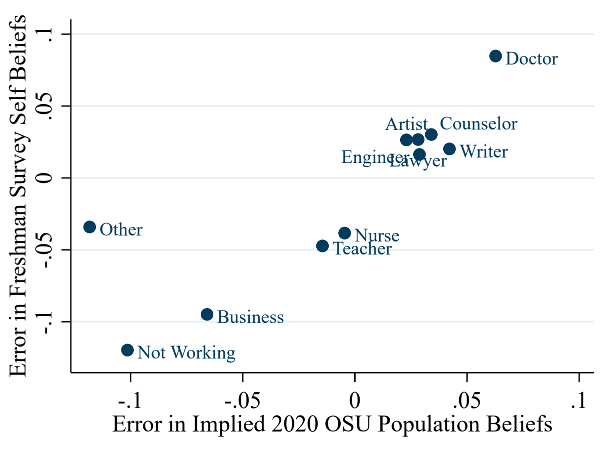
We compare these implied errors to a corresponding notion of "error" from the Freshman Survey data: the difference between the fraction of students who expect to have a career in each occupation minus the true proportion of college graduates with that occupation. Intuitively, this analysis asks whether the Freshman Survey respondents' expectations match what we would expect if they held the same (population) beliefs as the OSU sample. As Figure B.III shows, there is a robust positive relationship between actual "errors" in the Freshman Survey data and ImpliedError_c. The correlation 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 ($p = 0.39$).

Note that the above analysis assumes, among other things, no major switching and no bias in expectations about majors. Table A.VI shows that students' expectations about their future major are much closer to true major shares than for career expectations. Note in addition that any expected major switching would push *against* our finding that the

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

OSU beliefs data predict career expectations in the Freshman Survey. Intuitively, a student who expects that she might switch her major should reduce her likelihood of expecting that major's distinctive career.

Figure B.III: OSU Beliefs Predict Aggregate Biases in the Freshman Survey



Notes: The y-axis in Figure B.III 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.

B.8 Estimating Students' Preferences

Set-Up Here we structurally estimate the preferences that determine student's choice of major. 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 6:

$$\widehat{EU}^i[M] = \sum_c \pi^i_{c|M}\left(\alpha w^i_{c,M} + \beta^i_c\right) + \mu^i_M + \nu^i_M \tag{6}$$

In equation 6, $w^i_{c,M}$ 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^i_M 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 estimation of the relevant preference parameters. First, we assume ν^i_M 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, $\beta^i_c = \beta\mathbb{1}(c = c^*(i))$. From these assumptions, equation 7 follows, where π^i_M is i's belief about the probability they will graduate with M:

$$\log \frac{\pi^i_M}{\pi^i_{M'}} = \alpha \sum_c \left(\pi^i_{c|M} w^i_{c,M} - \pi^i_{c|M'} w^i_{c,M'}\right) + \beta\Big(\pi^i_{c^*(i)|M} - \pi^i_{c^*(i)|M'}\Big) + \mu^i_M - \mu^i_{M'} \tag{7}$$

Estimation To estimate this model, we use data collected in our 2020 OSU survey. For four majors, we directly elicited $\pi^i_{c|M}$, 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 $\sum_c \pi^i_{c|M} w^i_{c,M}$. 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^i_M. Finally, we assume μ^i_M 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 8:

$$\hat{\xi} = \underset{\xi}{\operatorname{argmax}} \sum_i \sum_{j=1}^{3} \log(L_{i,j}(\xi)) \tag{8}$$

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

Results This structural exercise 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 would have large effects on students' choices.

To examine the premium placed by students on their future occupations, Column 1 of Table B.III shows estimates from our baseline specification. 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 elasticity 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 B.III 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_M^i) 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% of students' preferred careers are the distinctive job of their second-ranked major.

The remaining columns of Table B.III 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 B.III). 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). Column 7 allows students to derive non-pecuniary utility from a second-favorite career. Column 8 allows richer non-pecuniary preferences over careers—different preferences for a full ranking of careers from most to least preferred—but uses students' stated ranking of how *likely* they are to have each career as their ranking of their preferences toward them. Column 9 is the same as the baseline model but estimated using students' population beliefs about career likelihoods and salaries. 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.

Table B.III: Choice Model Estimates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
α: E[Salary | M]	0.066***	0.054***	0.039***	0.150***	0.063***	0.068***	0.059***	0.049***	0.055***
	(0.021)	(0.018)	(0.017)	(0.038)	(0.020)	(0.021)	(0.020)	(0.017)	(0.021)
β: P(Favorite Career | M)	4.563***	4.225***	4.115***	4.556***	4.603***	4.468***	6.134***	3.797***	5.384***
	(0.099)	(0.098)	(0.091)	(0.100)	(0.098)	(0.109)	(0.381)	(0.414)	(1.511)
σ^2: Variance of μ_M^i	1.073***	0.756***	0.596***	1.068***	1.059***	1.042***	0.668***	1.482***	1.002***
	(0.039)	(0.028)	(0.023)	(0.038)	(0.038)	(0.038)	(0.031)	(0.055)	(0.052)
α_2: E[GPA | M]						0.947***			
						(0.153)			
β_2: P(2nd Favorite Career | M)							3.851***	2.179***	
							(0.933)	(0.424)	
β_3: P(3rd Favorite Career | M)								1.064**	
								(0.432)	
β_4: P(4th Favorite Career | M)								0.630**	
								(0.439)	
β_5: P(5th Favorite Career | M)								0.436**	
								(0.456)	
β_6: P(6th Favorite Career | M)								0.113**	
								(0.442)	
β_7: P(7th Favorite Career | M)								-0.174**	
								(0.435)	
β_8: P(8th Favorite Career | M)								-0.039**	
								(0.416)	
β_9: P(9th Favorite Career | M)								0.164**	
								(0.485)	
Implied WTP for 1pp Increase in Favorite Career ($10ks)	0.689***	0.781***	1.058***	0.304***	0.731***	0.655***	1.032***	0.776***	0.986***
	(0.380)	(1.130)	(3.245)	(0.091)	(0.501)	(0.326)	(0.527)	(0.748)	(0.731)

Notes: Table B.III shows parameter estimates of the structural estimation exercise. Each 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.

Equipped with these 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 we conduct, this in effect analyzes a scenario in which we completely shift students' perceptions about all majors towards the information we provide in the experiment.

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. For the average student, the total probability she assigns to graduating with one of these major is 85% (median = 95%), 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.

The dependent variable in Table B.IV is revision in probability for each of the four majors. We also show the mean of this variable for each specification. We see that students move away from their top-ranked major (by 14 p.p. in our baseline Model 1) and correspondingly toward their lower ranked majors. The independent variable in the regressions in Table B.IV is 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 (e.g., 50.5% in Model 1) 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.9 Correlation between Stereotypes and Confidence

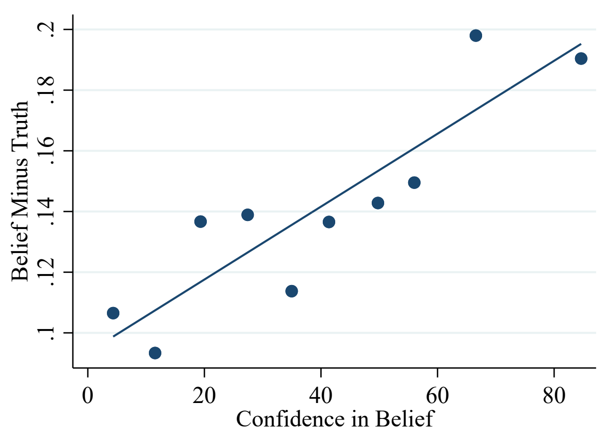
The first 2021 OSU survey, in addition to asking students' beliefs about the frequency of careers conditional on majors, also asked how confident they were in their in their guesses. In particular, immediately after each question about careers conditional on major, students were asked "And on a scale 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.IV shows this regression visually using a binscatter plot.

Table B.IV: Counterfactual Simulations

	Rank 1	Rank 2	Rank 3	Rank 4
	(1)	(2)	(3)	(4)
Model 1				
β: Reduction in Stereotypical Beliefs	-0.13***	0.07***	0.04***	0.02***
	(0.03)	(0.02)	(0.01)	(0.01)
Mean of Dependent Variable	-0.14*	0.07*+	0.04*+	0.03*+
Model 2				
β: Reduction in Stereotypical Beliefs	-0.13***	0.08***	0.02***	0.01***
	(0.02)	(0.02)	(0.01)	(0.01)
Mean of Dependent Variable	-0.13*	0.07*+	0.03*+	0.03*+
Model 3				
β: Reduction in Stereotypical Beliefs	-0.17***	0.07***	0.05***	0.05***
	(0.03)	(0.03)	(0.01)	(0.01)
Mean of Dependent Variable	-0.23*	0.08*+	0.08*+	0.07*+
Model 4				
β: Reduction in Stereotypical Beliefs	-0.13***	0.07***	0.02***	0.02***
	(0.02)	(0.02)	(0.01)	(0.01)
Mean of Dependent Variable	-0.13*	0.07*+	0.04*+	0.03*+
Model 5				
β: Reduction in Stereotypical Beliefs	-0.15***	0.10***	0.02***	0.02***
	(0.02)	(0.02)	(0.01)	(0.01)
Mean of Dependent Variable	-0.14*	0.07*+	0.04*+	0.03*+
Model 6				
β: Reduction in Stereotypical Beliefs	-0.13***	0.09***	0.02***	0.03***
	(0.03)	(0.02)	(0.01)	(0.01)
Mean of Dependent Variable	-0.14*	0.07*+	0.04*+	0.03*+
Model 7				
β: Reduction in Stereotypical Beliefs	-0.13***	0.08***	0.02**	0.02***
	(0.03)	(0.02)	(0.01)	(0.01)
Mean of Dependent Variable	-0.14*	0.07*+	0.04*+	0.03*+
Model 8				
β: Reduction in Stereotypical Beliefs	-0.12***	0.04**	0.06***	0.03***
	(0.02)	(0.02)	(0.01)	(0.01)
Mean of Dependent Variable	-0.06*	-0.01*+	0.04*+	0.03*+
Model 9				
β: Reduction in Stereotypical Beliefs	-0.23***	0.10***	0.04***	0.04***
	(0.03)	(0.02)	(0.01)	(0.01)
Mean of Dependent Variable	-0.14*	0.08*+	0.03*+	0.03*+

Notes: Table B.IV 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. The models correspond to the columns in Table B.III. *, **, and *** indicate significance of the regression coefficients at the 10%, 5%, and 1% levels.

Figure B.IV: Stereotyping and Confidence are Positively Correlated



Notes: Figure B.IV 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.

B.10 Heterogeneity by Role Models

Implicit associations like those described in section 2.3 are thought to be tightly linked to memory (e.g., Greenwald & Banaji 1995, Phelps et al. 2000). Recent work in economics has begun to formalize these connections between associations, memory, and belief biases like stereotyping. For example, Bordalo et al. (2023) argue that exaggeration of distinctive types arises when agents form beliefs by relying on associative memory. The crucial mechanism is cued recall, whereby when considering a particular hypothesis, items associated with that hypothesis disproportionately come to mind. For example, consider a student forming beliefs about what jobs psychology majors go on to have. When she considers the hypothesis that they become counselors, cueing prompts her to think of people with this job, who overwhelmingly tend to have majored in psychology. In contrast, when she considers the hypothesis that they become teachers and cueing prompts her 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 D 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.[33] 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 B.V shows OLS estimates of equations 9 and 10.

$$\pi^i_{c|M} = \alpha + \beta_1 RM^i_{c,M} + \beta_2 RM^i_{c,-M} + \mu_{c,M} + \epsilon^i_{c,M} \tag{9}$$

$$\pi^i_c = \alpha + \beta RM^i_c + \mu_c + \epsilon^i_c \tag{10}$$

In equation 9, $\pi^i_{c|M}$ 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 10, 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 B.V 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 3.25 p.p. ($p < 0.01$) for population beliefs and by 3.68 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 1.67 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 from the Freshman Survey (which asks the career, but not the major, of students' parents): students are 4.19 p.p. ($p < 0.01$) more likely to list an occu-

[33] 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.

pation 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 B.V 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 D, 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.

Table B.V: 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)
$RM_{c,M}$	3.25***	6.13***	1.38***		3.68***	7.19***	1.32*		
	(0.63)	(1.27)	(0.50)		(0.75)	(1.40)	(0.68)		
$RM_{c,-M}$	0.31*	-4.79***	0.61***		1.67***	-3.56*	2.00***		
	(0.17)	(1.70)	(0.16)		(0.29)	(2.08)	(0.27)		
RM_c				1.63***				7.91***	4.19***
				(0.30)				(0.67)	(0.01)
Constant	8.84***	46.77***	5.01***	8.82***	8.52***	50.51***	4.28***	7.76***	7.64***
	(0.05)	(0.84)	(0.08)	(0.05)	(0.08)	(0.91)	(0.09)	(0.12)	(0.00)
Individuals	772	772	772	772	772	772	772	772	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 B.V presents OLS estimates of equation 9 (columns 1-3 and 5-7) and equation 10 (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 graduated 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.

B.11 Details on Implicit Association Test

Here we provide more details on Experiments 1 and 2, which implemented our implicit association tests. Both experiments were run in May of 2025. Participants in Experiment 1 were paid $3.00 for completing the survey. In addition, they earned a $5.00 bonus if they were among the 10% fastest (among those in Experiment 1) in the IAT task. The median participant in Experiment 1 took 19.8 minutes to complete the experiment. Participants in Experiment 2 were paid a $7.50 completion payment. In addition, they earned a $1.00 bonus if their answer to a randomly selected belief question was within five percentage points of the correct value and a $5.00 bonus if they were among the 10% fastest (among those in Experiment 2) in the IAT task. The median participant in Experiment 2 took 45.9 minutes to complete the experiment. For both experiments, we recruited participants between the ages of 18 to 30.[34]

This online document includes screenshots of Experiments 1 and 2 (as well as Experiment 3, described below). The Qualtrics survey for Experiment 1 can be taken **by following this link**, while Experiment 2 can be taken **by following this link**.

Participants underwent multiple rounds of the IAT, each of which comprised five parts. Rounds differed by the major/career groups that they were about, whereas parts differed by the sorting task participants were faced with. Each round had a focal major group (e.g., "Humanities") and a focal career group (e.g., "Writers and Journalists"). The alternative career group and major group in all rounds were "Other Careers" and "Other Majors" (i.e., majors/careers that did not fall into any of the ten major groups/nine career groups).

The five parts of each round are described below:

1. Only Majors. In this part, participants sorted majors into either the focal major group or into "Other Majors". For each group, five individual majors were drawn randomly (with replacement) from the group in proportion to their true frequency in the population. These ten majors (five from the focal group and five from "Other Majors") were then presented in a random order, and participants needed to sort them into their respective groups.

2. Only Careers. In this part, participants sorted occupations into either the focal career group or into "Other Careers". For each group, five individual careers were drawn randomly (with replacement) from the group in proportion to their true frequency in the population. These ten occupations (five from the focal group and five from "Other Careers") were then presented in a random order, and participants needed to sort them into their respective groups.

3. Focal Groups. In this part, participants sorted majors and occupations into either the focal career group or the focal major group. For each group, five individual ca-

[34] For those under age 22 (about 10% of the sample), we restricted recruitment to current college students (self-reported to Prolific), while for those 22 and older, we restricted the sample to college graduates (self-reported to Prolific).

reers/majors were drawn randomly (with replacement) from the group in proportion to their true frequency in the population. These ten words (five occupations from the focal career group and five majors from the focal major group) were then presented in a random order, and participants needed to sort them into their respective groups.

4. Matched Pairing. This part involved twenty occupations/majors, five each from the focal major group, the focal career group, "Other Majors," and "Other Careers". The individual occupations/majors were drawn randomly from their respective groups like in previous rounds. In "Matched Pairing", the focal career and focal major groups shared a response key (e.g., pressing "Q"), and the "Other Majors" and "Other Careers" groups shared a response key (e.g., pressing "P").

5. Unmatched Pairing. The "Unmatched Pairing" part is identical to the "Matched Pairing" round, except that the focal career and "Other Majors" groups shared a response key (e.g., pressing "Q"), and the focal major and "Other Careers" groups shared a response key (e.g., pressing "P").

Across all parts, participants always sorted words to the left or the right using the "Q" and "P" keys. The left/right assignment was randomly assigned across parts but held constant within parts. Whenever participants correctly sorted a word in the correct direction, a green check mark appeared and the screen progressed to the next word. Whenever they incorrectly sorted a word, a red X appeared, and participants had to wait one second before trying again. Thus though all participants had to correctly sort all words, incorrect responses slowed them down. Participants were incentivized to complete all rounds as quickly as possible, as the fastest 10% of participants earned an additional $5 bonus.

The crucial parts within each round are parts 4 and 5. The key variable we construct is $\text{Difference}_{i,c,M}$, which is the difference between the time it took participant i to complete part 5 (Unmatched Pairing) minus the time it took her to complete part 4 (Matched Pairing) for focal career and major c and M. Participant i is said to have revealed an implicit association between c and M to the extent that $\text{Difference}_{i,c,M}$ is positive. Parts 1-3 within each round are meant to acquaint participants with the task and the categories being employed within the round. We can also use performance (time to completion) within these parts as a control variable, though as we will see doing so makes little difference to the results. Parts 1 and 2 occurred first but in a random order. Part 3 was always third, and then parts 4 and 5 occurred last in a random order.

Because our measure of implicit associations—the difference between time to complete parts 4 and 5 of the IAT—is potentially subject to outliers, we preregistered that we would look at both a winsorized version of this variable (Panel A) and an indicator for whether it is above the median (Panel B). We showed results using the winsorized measure in the main text in Table 2. Table B.VI shows analogous regressions instead using an indicator for above-median $\text{Difference}_{i,M}$. We see broadly similar patterns as with the continuous measure, though the leave-out-mean results are less precise.

Table B.VI: Implicit Associations Predict Stereotyping

	(1)	(2)	(3)	(4)	(5)	(6)
$\mathbb{1}$(Most Distinctive) x $\mathbb{1}$(Difference$_{i,r}$ > Med.)	7.23***		5.69***	4.84***		3.12*
	(1.63)		(1.97)	(1.44)		(1.66)
$\mathbb{1}$(Most Distinctive) x $\mathbb{1}$(Leave-out-Mean Difference$_{i,r}$ > Med.)		3.50**	3.43		3.48**	3.80
		(1.60)	(2.44)		(1.61)	(2.39)
$\mathbb{1}$(Most Distinctive)	12.74***	13.99***	11.43***			
	(3.24)	(4.04)	(3.39)			
P(Career | Major)	0.35***	0.38***	0.35***			
	(0.08)	(0.09)	(0.08)			
Observations	21,780	43,560	21,780	21,780	43,560	21,780
Individuals	396	396	396	396	396	396
Individual-by-Career Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Major-by-Career Fixed Effects	No	No	No	Yes	Yes	Yes

Notes: Table B.VI presents OLS estimates where the dependent variable is Experiment-2 participants' population beliefs about the fraction of graduates with a certain major with a certain career. Difference$_{i,M}$ is (winsorized and normalized) difference between the time participants took to complete Parts 5 and 4 (Unmatched Pairing minus Matched Pairing) of the implicit association test for that major and its distinctive career. The leave-out-mean (LoM) difference is calculated by taking the average of Difference$_{i,M'}$ for all other majors M' that participant took the IAT for. The indicators are dummies for these variables being above the median. "P(Career | Major)" is the true percent of graduates with the the IAT round's focal major that are working in its focal career, calculated from the 2017-2019 American Community Survey. $\mathbb{1}$(Most Distinctive) is a dummy variable indicating whether the focal career is the most distinctive outcome for the focal major. All regressions cluster standard errors at the individual level and at the career-by-major level. *, **, and *** indicate significance at the 10%, 5%, and 1% levels.

B.12 Experiment 3: Other Measures

In addition to Experiments 1 and 2 that investigated implicit associations, we also simultaneously conducted a third experiment eliciting other measures potentially related to stereotyping. See **this online document** for screenshots from this experiment and **this link** to take the associated Qualtrics survey.

B.12.1 Design

Experiment 3 had four modules, which occurred in a random order:

1. Population Beliefs. This module was identical to the population beliefs module in Experiment 2. If any of these questions was chosen to determine participants' bonus, they earned an additional $1.00 payment if a randomly selected answer for that major was within five percentage points of the correct answer.

2. Explicit Associations. In this module, participants repeatedly were presented with a major and two careers, and they had to indicated which of the two careers they more "associated" with the major. For each major, we selected its most distinctive career plus (when not already included) teaching and businessperson, and included all pairwise comparisons, resulting in 26 combinations of majors and career-pairs. Participants answered all 26 questions one-by-one in a random order. This elicitation was unincentivized.

3. Expressed Stereotypes. In this module, which consisted only a single question, participants were asked of a randomly selected major: "We'd like to ask you to describe your stereotype of what sort of job a 35-year old American college graduate who majored in M has. What job comes to mind when you think of a person like that?" We then use GPT-4o-mini, a large language model, to analyze which career(s) the participants' description of her stereotype appears to mention. This elicitation was unincentivized.

4. Awareness. A reasonable question is whether participants are aware in the first place of the many non-distinctive jobs that college graduates end up working in, and whether any such unawareness might be driving our main stereotyping result. To assess this possibility, this module of Experiment 3 sequentially presented participants with job titles and asked of each, "As far as you are aware, are there people in the United States with this job?" We asked about 40 jobs, drawn randomly (without replacement) from a list that included the 100 occupation titles (drawn from ACS) most common among college graduates with our 10 majors who are not working in their major's most distinctive career group. To incentivize this measure, we additionally included a group of 15 plausible-but-fictitious job titles (e.g., "Output Specialist", "Kinetic Resource Planners", etc.), and—if this module was chosen to determine their bonus—participants earned an additional \$1.00 payment if they correctly expressed awareness of all real jobs and none of the fictitious jobs they were asked about.

B.12.2 Logistics

We recruited 252 participants from Prolific to take Experiment 3 in May 2025. The median participant too 25.3 minutes to complete the experiment and was paid a \$4.50 completion payment for doing so. One question was randomly chosen—either from the population beliefs questions or the awareness module—to determine their bonus payment.

B.12.3 Results

Explicit Associations. We first analyze which careers participants explicitly associate with majors and whether such associations are correlated with participant beliefs. Table B.VII shows OLS regressions where the dependent variable is $\text{Association}_{i,M,c_1>c_2}$, an indicator for whether individual i more associates major M with c_1 than with c_2. We include all 26 M-c_1-c_2 combinations for each participant and cluster standard standard errors at both the individual and combination levels. We regress $\text{Association}_{i,M,c_1>c_2}$ on the difference between indicators for whether c_1 or c_2 is the most distinctive career for M, as well as the difference in their true conditional frequency among M majors. Columns 1 and 2 of Panel A show that both of these variables are individually predictive of explicit associations: participants are more likely to associate c_1 with M (compared to c_2) if it is M's distinctive career and if it is more common among M majors. Column 3 shows that both covariates retain predictive power even controlling for the other, though distinctiveness is much more predictive of explicit associations than true frequencies. Holding frequency fixed, a major's distinctive career is 35 p.p. more likely to be explicitly associated with it than non-distinctive careers ($p < 0.01$). Holding (non) distinctiveness fixed however, a 10 p.p. increase in the true frequency of a career within a major only increases its chance of being explicitly associated with it by 0.7 percentage points ($p < 0.01$).

Panel B of B.VII shows similar regressions but where 1) the dependent variable is the (winsorized) time that it took participant i to decide whether she associated M more with c_1 than with c_2, and 2) the independent variables include the *absolute value* of the differences in distinctiveness and raw frequency of c_1 and c_2. Thus, these regressions let us see whether participants *more quickly* express explicit associations when careers differ in their distinctiveness or differ more in their frequencies. We see that participants decide more quickly (negative coefficients) when careers differ along these dimensions (columns 1 and 2), but that only distinctiveness significantly predicts speed when both are included in the regression.

Do these explicit associations predict participants' beliefs? Column 1 of Table B.VIII

Table B.VII: Explicit Associations between Majors and Distinctive Careers

	(1)	(2)	(3)
Panel A: Associate M with Career 1 more than Career 2			
$\mathbb{1}$(Career 1 is Most Distinctive) - $\mathbb{1}$(Career 2 is Most Distinctive)	0.36***		0.35***
	(0.01)		(0.01)
P(Career 1 | Major) − P(Career 2 | Major)		0.57***	0.07***
		(0.02)	(0.02)
Constant	0.50***	0.50***	0.50***
	(0.01)	(0.01)	(0.01)
Observations	6,552	6,552	6,552
Individuals	252	252	252
Panel B: Time Spent Deciding Association			
Abs($\mathbb{1}$(Career 1 is Most Distinctive) −$\mathbb{1}$(Career 2 is Most Distinctive))	-0.54***		-0.53***
	(0.04)		(0.04)
Abs(P(Career 1 | Major) − P(Career 2 | Major))		-0.57***	-0.10
		(0.10)	(0.10)
Constant	3.09***	2.82***	3.10***
	(0.07)	(0.06)	(0.07)
Observations	6,552	6,552	6,552
Individuals	252	252	252

Notes: Table B.VII shows OLS regressions where the dependent variable in Panel A is an indicator for whether a participant said they more associate the given major with career 1 than with career 2. In Panel B, the dependent variable is the time (winsorized at the 5th and 95th percentiles) that participants took to answer this association question. Standard errors are clustered at the individual and major-by-career-pair level. "P(Career 1/2| Major)" is the true percent of graduates with that major that are working in career 1/2, calculated from the 2017-2019 American Community Survey. $\mathbb{1}$(Career 1/2 Most Distinctive) is a dummy variable indicating whether the career is the most distinctive outcome for the major. *, **, and *** indicate significance at the 10%, 5%, and 1% levels.

regresses participant i's population beliefs about $\mathrm{P}(c_1|M)$ minus her beliefs about $\mathrm{P}(c_2|M)$ on 1) the difference in their distinctiveness, 2) the difference in their true conditional frequency, and 3) an indicator for whether i explicitly associated M more with c_1 than with c_2. We see that participants on average believe that careers they associate with a major are 10.3 p.p. ($p < 0.01$) more common, holding fixed distinctiveness and frequency differences. Column 2 shows that this correlation remains large and significant (6.7 p.p., $p < 0.01$) even controlling for major-by-career-pair fixed effects, indicating that these results reflect differences across participants, not only fixed differences across careers/majors.

Columns 3-6 of Table B.VIII investigate whether the speed with which participants decide on their explicit associations has additional predictive power in explaining beliefs. They interact participants' association of c_1 with c_2 with the time it took to answer with that association (winsorized in columns 3-4 and an indicator for above-median time in columns 5-6). We see that slower decisions reduce the correlation between associations and beliefs or, equivalently, that faster associations are more predictive of beliefs: a participant who expressed one second faster that she associates M more with c_1 that c_2 has 1.3 p.p. higher beliefs about its relative frequency (column 1, $p < 0.10$). Those whose decision took a below-median amount of time have 4.9 p.p. higher beliefs (column 5, $p < 0.05$). These patterns also appear controlling for major-by-career-pair fixed effects (columns 4 and 6, $p < 0.10$ for all comparisons).

We conclude from these results that participants tend to explicitly associate majors with their distinctive careers over and above their true conditional frequency, and that differences across participants in these explicit associations help to explain their beliefs. Further, decision times seem to serve as a proxy for the strength of these associations, in that they moderate the correlation between associations and beliefs.

Table B.VIII: Explicit Associations Predict Beliefs

	(1)	(2)	(3)	(4)	(5)	(6)
$\text{Association}_{M,C1>C2}$	10.34***	6.70***	14.04***	9.47***	13.07***	8.89***
	(2.19)	(1.21)	(3.03)	(2.05)	(2.63)	(1.67)
P(Career 1 | Major) − P(Career 2 | Major)	0.39***		0.39***		0.39***	
	(0.08)		(0.07)		(0.07)	
$\mathbb{1}$(Career 1 Most Distinctive) $-\mathbb{1}$(Career 2 Most Distinctive)	16.29***		16.16***		16.13***	
	(3.20)		(3.15)		(3.14)	
$\text{Association}_{M,C1>C2} \times$ Time			-1.29*	-0.98*		
			(0.67)	(0.56)		
Time			1.25**	0.80*		
			(0.51)	(0.42)		
$\text{Association}_{M,C1>C2} \times \mathbb{1}(\text{Time}>\text{Med.})$					-4.93**	-4.13**
					(2.32)	(1.98)
$\mathbb{1}$(Time>Med.)					4.72***	3.42**
					(1.64)	(1.35)
Constant	-9.22***	-5.77***	-12.69***	-8.00***	-11.67***	-7.54***
	(1.30)	(0.60)	(1.94)	(1.35)	(1.58)	(0.97)
Observations	6,552	6,552	6,552	6,552	6,552	6,552
Individuals	252	252	252	252	252	252
Major by Career-Pair Fixed Effects	No	Yes	No	Yes	No	Yes

Notes: Table B.VIII presents OLS estimates where the dependent variable is Experiment-3 participants' population beliefs about the fraction of graduates with a certain major with career 1 minus their corresponding belief about career 2. "$\text{Association}_{M,c_1,c_2}$ is an indicator for whether a participant said they more associate the given major with career 1 than with career 2. "P(Career 1/2\| Major)" is the true percent of graduates with that major that are working in career 1/2, calculated from the 2017-2019 American Community Survey. $\mathbb{1}$(Career 1/2 Most Distinctive) is a dummy variable indicating whether the career is the most distinctive outcome for the major. "Time" is the number of seconds (winsorized at the 5th and 95th percentiles) it took the participant to choose which career she more associated with the major. Standard errors are clustered at the individual and major-by-career-pair level. *, **, and *** indicate significance at the 10%, 5%, and 1% levels.

Expressed Stereotypes Next, recall participants were asked to describe in their own words their stereotype of the job(s) that people with a randomly selected major have. Column 1 of Table B.IX regresses an indicator for whether a participant's response for major M mentioned career c (according to GPT-4o-mini) on c's true conditional frequency and an indicator for whether c is M's most distinctive career. We see that, holding frequency fixed, participants are 32 p.p. ($p < 0.01$) more likely to mention a job if it is M's most distinctive career. This number increases to 42 p.p. ($p < 0.01$) if we include career fixed effects (column 2). Column 3 regresses participants' population beliefs about $\mathrm{P}(c|M)$ on c's true frequency, an indicator for distinctiveness, and an indicator for whether the participant's expressed stereotype mentions c. We see that even conditional on objective frequency and distinctiveness, participants' beliefs about $\mathrm{P}(c|M)$ are 5.2 p.p. ($p < 0.01$) higher when they express that c is the stereotypical career of M. Column 4 shows that this pattern holds even controlling for major-by-career-fixed effects (3.0 p.p., $p < 0.01$), indicating that these results at least in part reflect differences across participants in their expressed stereotypes.

Awareness of Non-Distinctive Jobs. Next, recall that participants were asked which jobs they were aware of. These job were chosen from the most common non-distinctive occupations as well as a list of fictional jobs (to enable incentivization). Our first result is that participants express awareness of the vast majority of jobs. The average participant expresses that she is aware of 94% of real jobs. In part this reflects the fact that many of the most non-distinctive jobs are quite familiar (teachers, secretaries, lawyers, managers, salespeople, etc). In contrast, participants report being "aware" of only 41% of the fictitious jobs we included, indicating a high level of engagement and attention from participants.

We can next ask how unawareness of non-distinctive jobs correlates with participants' beliefs. Column 1 of Table B.X regresses population beliefs about $\mathrm{P}(c|M)$ on c's true frequency among M majors and an indicator for c being M's most distinctive job, where we replicate our original stereotyping result: participants believe distinctive jobs are 19.3 ($p < 0.01$) more common holding objective frequency fixed.

If this stereotyping is largely a result of participants being unaware of non-distinctive jobs (and failing to correct for this in their beliefs), then we should expect that participants who are unaware of more jobs should stereotype majors to a greater degree. Columns 2-7 show that, if anything, the opposite is true. In columns 2 and 6, we interact the distinctiveness indicator with participants' error rates in the awareness module: that is, the fraction of real occupations they express unawareness of, and the fraction of fictitious jobs they express awareness of. We see that participants who express relatively more unawareness of non-distinctive jobs stereotype *less.* This is true whether we use for the error rate itself (column 2, $p < 0.10$), if we include for an indicator for above-median error

Table B.IX: Expressed Stereotypes Mention Distinctive Careers and Predict Beliefs

	Expressed Stereotype Mentions Career		Population Belief: P(Career | Major)	
	(1)	(2)	(3)	(4)
P(Career | Major)	0.01***	0.01***	0.31***	
	(0.00)	(0.00)	(0.04)	
$\mathbb{1}$(Most Distinctive)	0.32***	0.42***	18.79***	
	(0.04)	(0.05)	(1.84)	
$\mathbb{1}$(Expressed Stereotype Mentions Career)			5.19***	3.04**
			(1.20)	(1.45)
Constant	-0.00	0.01	3.85***	8.69***
	(0.01)	(0.01)	(0.33)	(0.19)
Observations	2,772	2,772	2,772	2,772
Individuals	252	252	252	252
Career Fixed Effects	No	Yes		
Major-by-Career Fixed Effects			No	Yes

Notes: Table B.IX presents OLS estimates where the dependent variable in columns 1 and 2 is $\mathbb{1}$(Expressed Stereotype Mentions Career), an indicator for whether Experiment-3 participants mention a career when asked to describe their stereotype of the job that people with that major have. In columns 3 and 4, the dependent variable is participants' population beliefs about the fraction of graduates with a certain major with a certain career. "P(Career | Major)" is the true percent of graduates with that 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 the career is the most distinctive outcome for the major. Standard errors are clustered at the individual and major-by-career level. *, **, and *** indicate significance at the 10%, 5%, and 1% levels.

(column 4, $p < 0.10$), or directionally if we additionally include interactions with actual $\mathrm{P}(c|M)$ (columns 6 and 7, $p > 0.10$). We therefore find no evidence that stereotyping is driven by participants being simply unaware of the jobs that tend to be non-distinctive.

Table B.X: Correlation between Awareness of Occupations and Stereotyping

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
P(Career | Major)	0.44***	0.44***	0.62***	0.44***	0.62***	0.56***	0.57***
	(0.10)	(0.11)	(0.12)	(0.11)	(0.12)	(0.11)	(0.11)
$\mathbb{1}$(Most Distinctive)	19.28***	24.88***	19.28***	24.64***	19.28***	22.00***	21.51***
	(4.59)	(5.02)	(4.59)	(4.99)	(4.59)	(4.95)	(4.97)
$\mathbb{1}$(Most Distinctive) × Fraction Real Occupations Unaware of		-27.50*				-5.51	
		(15.69)				(10.89)	
$\mathbb{1}$(Most Distinctive) × Fraction Fake Occupations "Aware" of		-9.94***				-5.91***	
		(3.02)				(2.22)	
P(Career | Major) × Fraction Real Occupations Unaware of			-1.01**			-0.88***	
			(0.44)			(0.24)	
P(Career | Major) × Fraction Fake Occupations "Aware" of			-0.30***			-0.16***	
			(0.09)			(0.05)	
$\mathbb{1}$(Most Distinctive) × Fraction Real Occupations Unaware of > Median				-4.35*			-1.48
				(2.26)			(1.95)
$\mathbb{1}$(Most Distinctive) × Fraction Fake Occupations "Aware" of > Median				-7.39***			-3.43*
				(2.39)			(1.90)
P(Career | Major) × Fraction Real Occupations Unaware of > Median					-0.15**		-0.12***
					(0.06)		(0.03)
P(Career | Major) × Fraction Fake Occupations "Aware" of > Median					-0.24***		-0.16***
					(0.06)		(0.03)
Constant	3.32***	3.32***	3.32***	3.32***	3.32***	3.32***	3.32***
	(0.75)	(0.75)	(0.75)	(0.75)	(0.75)	(0.75)	(0.75)
Observations	27,720	27,720	27,720	27,720	27,720	27,720	27,720
Individuals	252	252	252	252	252	252	252
Career by Individual Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Table B.X presents OLS estimates where the dependent variable is Experiment-3 participants' population beliefs about the fraction of graduates with a certain major with a certain career. "P(Career | Major)" is the true percent of graduates with that 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 the career is the most distinctive outcome for the major. "Fraction Real Occupations Unaware of" is the share of actual jobs that participants said they were not aware of. "Fraction Fake Occupations "Aware" of" is the share of fictitious jobs that participants nonetheless said they were aware of. Standard errors are clustered at the individual and major-by-career level. *, **, and *** indicate significance at the 10%, 5%, and 1% levels.

Correlation between own outcomes and beliefs In both Experiments 2 and 3, which asked population beliefs questions, we additionally asked participants what their college major and current occupation are. We can thus look at any correlations between own outcomes and population beliefs. Column 1 of Table B.XI first replicates our main stereotyping result among this sample, combining college graduate participants from Experiments 2 and 3: participants believe that distinctive jobs are 19.0 p.p. more common ($p < 0.01$) than non-distincitve jobs, holding fixed true conditional frequency. Participants stereotype their own major by 2.6 p.p. more than they do other majors (columns 2 vs 5), though this difference is not statistically significant ($p = 0.26$). They also stereotype their own major 5.2 p.p. more if they themselves are working in their major's most distinctive career (columns 3 vs 4), though again this difference is not significant ($p = 0.23$). In contrast, participants with their own major's distinctive career stereotype other majors less (-4.4 p.p., $p < 0.01$), and this difference between own and other majors is statistically significant (-4.4 vs $+5.2$ p.p., $p < 0.05$). We conclude from these results that there is some suggestive evidence in favor of own experiences affecting levels of stereotyping, though our estimates are not always precise enough to draw sharp conclusions.

Another way of investigating the role of experience in explaining participants' beliefs is to see whether stereotyping appears related to participants' age. Experiments 2 and 3 recruited participants between the ages of 18 and 30, restricting to college graduates among those older than 22. Table B.XII splitting the sample by median age. We see no significant differences in stereotyping across age groups. Neither do we see any systematic pattern of participants becoming aware of more jobs as they get older (bottom row of table).

Table B.XI: Correlation between Own Career/Major and Stereotyping

	All College Grads	Own Major			Other Majors		
	Pooled	Pooled	ND	D	Pooled	ND	D
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
P(Career | Major)	0.33***	0.29***	0.34***	0.22***	0.33***	0.35***	0.31***
	(0.05)	(0.06)	(0.08)	(0.07)	(0.05)	(0.06)	(0.05)
$\mathbb{1}$(Most Distinctive)	18.99***	21.40***	19.53***	24.77***	18.84***	20.96***	16.57***
	(3.42)	(2.99)	(3.67)	(3.16)	(3.49)	(3.68)	(3.39)
Constant	4.36***	4.48***	4.27***	4.80***	4.36***	4.02***	4.77***
	(0.46)	(0.48)	(0.62)	(0.49)	(0.47)	(0.50)	(0.47)
Observations	62,370	5,621	2,607	3,014	56,749	29,623	27,126
Individuals	567	511	237	274	567	293	274

Notes: Table B.XI presents OLS estimates where the dependent variable is Experiment-3 participants' population beliefs about the fraction of graduates with a certain major with a certain career. "P(Career | Major)" is the true percent of graduates with that 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 the career is the most distinctive outcome for the major. All columns restrict data to college graduates, columns 2-4 further restrict the data to participants' own major, while columns 5-7 restrict to majors other than participants' own. Columns 4 and 7 further restrict the data to participants who report having the distinctive career of their own major, while columns 3 and 6 restrict to those who do not. *, **, and *** indicate significance at the 10%, 5%, and 1% levels.

Table B.XII: Correlation between Age and Stereotyping

	All Majors		Own Major		Other Majors	
	18-25	26+	18-25	26+	18-25	26+
	(1)	(2)	(3)	(4)	(5)	(6)
P(Career | Major)	0.42***	0.39***	0.29***	0.30***	0.43***	0.39***
	(0.10)	(0.10)	(0.06)	(0.07)	(0.10)	(0.09)
$\mathbb{1}$(Most Distinctive)	16.55***	17.52***	20.91***	21.68***	16.38***	17.20***
	(4.46)	(4.20)	(2.72)	(3.44)	(4.51)	(4.18)
Constant	3.73***	3.99***	4.59***	4.42***	3.68***	3.98***
	(0.73)	(0.69)	(0.51)	(0.53)	(0.73)	(0.67)
Observations	30,470	40,810	2,013	3,608	28,457	37,202
Individuals	277	371	183	328	277	371
E3: Share Real Jobs Unaware of	0.06	0.06				
E3: Share Fake Jobs "Aware" Of	0.43	0.39				
Individual-by-Career-Fixed Effects	Yes	Yes	No	No	Yes	Yes

Notes: Table B.XII presents OLS estimates where the dependent variable is Experiment-2 and Experiment-3 participants' population beliefs about the fraction of graduates with a certain major with a certain career. "P(Career | Major)" is the true percent of graduates with that 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 the career is the most distinctive outcome for the major. Columns 1-2 include all majors, columns 3-4 restrict the data to participants' own major, while columns 5-6 restrict to majors other than participants' own (including college-age participants who do report any major). Each column restricts the data to participants whose age falls into the range listed in the column heading (split by median). *, **, and *** indicate significance at the 10%, 5%, and 1% levels.

C Proofs for Section 3

Here we prove the propositions from Section 3. Recall that there must be a threshold strategy for choosing major A: student j majors in A if $\psi_j > \psi^*$.

$$
\begin{aligned}
w_{b,B} &= \int_{-k+\widehat{s_k}}^{T(\psi^*)} w_{b,A}\frac{1}{2k}d\epsilon + \int_{T(\psi^*)}^{k+\widehat{s_k}} (w_{a,A}+\epsilon+\psi_j)\frac{1}{2k}d\epsilon \\
\iff 0 &= 2w_{b,A}\Big(w_{b,A}-w_{a,A}-\psi^*+k-\widehat{s_k}\Big) \\
&+ \Big(k+\widehat{s_k}-w_{b,A}+w_{a,A}+\psi_j\Big)(w_{a,A}+\psi^*+k+\widehat{s_k}+w_{b,A}) - 4kw_{b,B} \\
&= \psi^{*2} \\
&+ \psi^*\cdot 2\Bigg[w_{a,A}-w_{b,A}+k+\widehat{s_k}\Bigg] \\
&+ \Bigg[\Big(k+\widehat{s_k}+w_{a,A}-w_{b,A}\Big)^2 - 4k(w_{b,B}-w_{b,A})\Bigg]
\end{aligned}
$$

We can then apply the quadratic formula:

$$
\psi^* = -\Bigg[w_{a,A}-w_{b,A}+k+\widehat{s_k}\Bigg] \pm 2\sqrt{k(w_{b,B}-w_{b,A})}
$$

The larger root is the relevant one here, so:

$$
\psi^* = 2\sqrt{k(w_{b,B}-w_{b,A})} - \Bigg[w_{a,A}-w_{b,A}+k+\widehat{s_k}\Bigg]
$$

Note that, to avoid edge cases where the student always chooses b with a positive probability, we assume that $-k+s_k < T(\psi^*) < k+s_k$: in other words, the marginal student would have a non-zero chance of choosing career a were she to enter major A.

We can now derive the share of people choosing major A:

$$
p_A = 1 - \frac{1}{2h}\Big(\psi^*+h-s_h\Big) = \frac{1}{2h}(h-\psi^*+s_h)
$$

Of these, we can now derive what fraction students *think* choose a:

$$\pi_{a|A} = \frac{1}{p_A} p_{a,A} = \frac{1}{4hkp_A} \int_{\psi^*}^{h+s_h} \int_{T(\psi)}^{k+\widehat{s_k}} d\epsilon d\psi$$
$$= \frac{1}{2k} \left[k + \widehat{s_k} + w_{a,A} - w_{b,A} + \frac{1}{2}(h + s_h + \psi^*) \right]$$

The *actual* fraction $p_{a|A}$ is then given by

$$p_{a|A} = \frac{1}{2k} \left[k + s_k + w_{a,A} - w_{b,A} + \frac{1}{2}(h + s_h + \psi^*) \right]$$

Note that $p_{a|A}$ uses the true s_k rather than beliefs $\widehat{k}$ but that ψ^* is still defined in terms of beliefs. This incorporates the fact that beliefs about s_k drive major choices, whereas the actual distribution of ϵ drives actual career choices conditional on major.

Next, define $\theta \equiv \pi_{a|A} - p_{a|A}$, which, by substituting in the above expressions, is easily shown to be equal to $\frac{1}{2k}(\widehat{s_k} - s_k)$. This proves Proposition 1.

Next, we can derive comparative statics with respect to θ.

$$\frac{dp_{a|A}}{d\theta} = -\frac{\gamma}{k}\frac{dp_a}{d\theta} + \frac{1}{4k}\frac{d\psi^*}{d\theta}$$
$$= -\frac{\gamma}{k}\frac{dp_a}{d\theta} + \frac{1}{4k}\Big(2\gamma\frac{dp_a}{d\theta} - 2k\Big)$$
$$= -\frac{\gamma}{2k}\frac{dp_a}{d\theta} - \frac{1}{2}$$
$$\frac{dp_A}{d\theta} = -\frac{1}{2h}\frac{d\psi^*}{d\theta} = -\frac{1}{2h}\left[2\gamma\frac{dp_a}{d\theta} - 2k\right] = \frac{1}{h}\Big(k - \gamma\frac{dp_a}{d\theta}\Big)$$

$$
\begin{aligned}
\frac{dp_a}{d\theta} &= p_A \frac{dp_{a|A}}{d\theta} + \frac{dp_A}{d\theta} p_{a|A} \\
&= -\frac{1}{2k} p_A \Big(\gamma \frac{dp_a}{d\theta} + k\Big) + \frac{1}{h}\Big(k - \gamma \frac{dp_a}{d\theta}\Big) p_{a|A} \\
&= \frac{\frac{k}{h} p_{a|A} - \frac{1}{2} p_A}{1 + \frac{\gamma}{2k} p_A + \frac{\gamma}{h} p_{a|A}} \\
&= \frac{k + s_k + w_{a,A} - w_{b,A} + \psi^*}{2h + \frac{\gamma h}{k} p_A + 2\gamma p_{a|A}} \\
&= \frac{\sqrt{k(w_{b,B} - w_{b,A})} - k\theta}{h + \frac{\gamma h}{2k} p_A + \gamma p_{a|A}}
\end{aligned}
$$

Note that our earlier assumption that the marginal student has a non-zero chance of choosing a (i.e., $\psi^* < k + s_k$) ensures that this derivative above is positive:

$$
\begin{aligned}
w_{b,A} - w_{a,A} - 2\sqrt{k(w_{b,B} - w_{b,A})} + \Big(w_{a,A} - w_{b,A} + k + s_k + 2k\theta\Big) < k + s_k \\
\iff k\theta < \sqrt{k(w_{b,B} - w_{b,A})}
\end{aligned}
$$

Thus, $\frac{dp_a}{d\theta}$ is positive. Recognizing that $\frac{dp_{b|A}}{d\theta} = -\frac{dp_{a|A}}{d\theta}$ then completes the proof of Proposition 2.

Finally we can derive Proposition 3.

$$
\begin{aligned}
W(\theta) &= \int_{-h+s_h}^{\psi^*} w_{b,B}\frac{1}{2h}d\psi \\
&+ \int_{\psi^*}^{k+s_k}\left[\int_{-k+s_k}^{T(\psi)} w_{b,A}\frac{1}{2k}d\epsilon + \int_{T(\psi)}^{k+s_k}(w_{a,A}+\epsilon+\psi)\frac{1}{2k}d\epsilon\right]\frac{1}{2h}d\psi \\
&= \int_{-h+s_h}^{\psi^*} w_{b,B}\frac{1}{2h}d\psi \\
&+ \int_{\psi^*}^{h+s_h}\frac{1}{2k}\left[2w_{b,A}k + \frac{1}{2}\Big(w_{a,A}-w_{b,A}+k+s_k+\psi\Big)^2\right]\frac{1}{2h}d\psi
\end{aligned}
$$

Finally we can differentiate this with respect to θ:

$$
\begin{aligned}
\frac{dW(\theta)}{d\theta} &= \gamma\frac{dp_a}{d\theta}\frac{1}{2h}\int_{-h+s_h}^{\psi^*} d\psi \\
&+ \frac{1}{2hk}\frac{d\psi^*}{d\theta}\left[2(w_{b,A}-w_{b,B})k + \frac{1}{2}\Big(w_{a,A}-w_{b,A}+k+s_k+\psi^*\Big)^2\right] \\
&- \frac{1}{4hk}2\gamma\frac{dp_a}{d\theta}\int_{\psi^*}^{h+s_h}\Big[w_{a,A}-w_{b,A}+s_k+\psi\Big]d\psi
\end{aligned}
$$

After some algebra, this simplifies to:

$$
\frac{dW(\theta)}{d\theta} = \gamma\frac{dp_a}{d\theta}\Big(1-2p_a\Big) + \theta\frac{dp_A}{d\theta}\left[k\theta - 2\sqrt{k(w_{b,B}-w_{b,A})}\right]
$$

which is Proposition 3. Note that the same condition from before ($\psi^* < k + s_k$) guarantees that the second term in the above equation is negative.

D Online Appendix: Model of "What Comes to Mind" 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.

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

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.

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

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?

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

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 B.10.

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,

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

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,

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

We can evaluate this derivative at the rational benchmark:

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

Similar derivations show that at the rational benchmark,

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

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.