What Jobs Come to Mind? Stereotypes about Fields of Study

John J. Conlon* Dev Patel†

October 26, 2023

Abstract

Using both large-scale nationally representative data and surveys administered among undergraduates at the Ohio State University, we document that US freshmen hold systematic misperceptions about the relationship between college majors and occupations. Students stereotype fields of study, greatly exaggerating the likelihood that majors lead to their distinctive jobs (e.g., counselor for psychology, journalist for journalism). In a field experiment, we find that reducing stereotyping has significant effects on students' intentions about what to study as well as the classes and majors they actually choose. Finally, we present a model of belief formation in which stereotyping arises as a product of associative recall. The model makes additional predictions—which new survey evidence broadly confirms—both about average beliefs and how heterogeneity in beliefs should systematically depend on the careers and majors of people students know personally.

*Stanford University, jjconlon@stanford.edu. †Harvard University, devpatel@g.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, Gautam Rao, Matthew Rabin, Josh Schwartzstein, Andrei Shleifer, Evan Soltas, and Basit Zafar for helpful feedback. This study underwent human subjects review at Harvard University under IRB20-1581. We are grateful for financial support from the Lab for Economic Applications and Policy, the Mind Brain Behavior Initiative, the National Science Foundation Graduate Research Fellowship under grant DGE1745303, and the Alfred P. Sloan Foundation Predoctoral Fellowship in Behavioral Macroeconomics, awarded through the NBER. The experiment described in this paper was pre-registered at the AEA RCT Registry (ID AEARCTR-0008339). All errors are ours.

1 Introduction

A young person deciding on their education must first form beliefs about the consequences of their choices. Despite the importance of these decisions, a growing body of evidence across many educational contexts suggests that students' expectations often stray far from the truth. However, the origins of these mistaken beliefs often remain unclear, and perhaps as a result biases can appear unstable across samples or contexts. Similar puzzles—pervasive errors in expectations that defy simple characterization—occur across a wide range of economic domains. Are there systematic patterns in these errors? If so, what mechanisms underpin them, and what implications might they have for human capital investments?

We explore these questions in the context of one of the most important economic decisions many people ever make: their choice of what to study in college.[1] We focus on students' beliefs about the link between their college major and the job they will hold in the years to come. We begin by providing motivating evidence documenting large and persistent differences between the careers that undergraduates expect to attain and the actual occupations they go on to have. To do so, we compare nationally representative survey data from millions of US college first-years with government data on the same cohorts. Two to four times more college freshmen expect to work in certain professions—e.g., doctor, counselor, journalist—than actually do. In contrast, many fewer expect to be teachers, working in business, or non-employed than ultimately are. These gaps—amounting to between 40,000 and 200,000 students a year, depending on the profession—appear largely unchanged since at least the 1970s.

In this paper, we test the hypothesis that these patterns are driven by stereotyping: that is, when considering a major, students may form an oversimplified picture—a stereotype—of the career they would have if they pursued that major. Following Bordalo et al. (2016), we define a stereotype by *distinctiveness*: i.e., a career c is the stereotype of major M if it maximizes $P(c|M)/P(c|\text{ not } M)$. We then show that many more students expect to attain their major's stereotypical career than actually work in that job: 65% of prospective art

[1] 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).

majors expect to be artists (in reality, 17% are), 63% 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. In turn, few students expect to work in business or to teach (unless they are pursuing business or education majors), and almost none expects to be non-employed. By themselves, of course, these results are consistent with mechanisms other than stereotyping. For instance, they could be driven by students exaggerating their own abilities, overestimating the demand for certain jobs, or selecting fields of study about which they hold particularly extreme beliefs.

To distinguish between these possibilities and to test for 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 stereotypical 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 and related confounds). In regression analyses, we additionally control for individual-by-career fixed effects to show that these biases are not about salient careers *per se* but about the relationship *between* majors and careers. 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. Furthermore, we also find substantial stereotyping in an online survey of US adults (for both college and non-college educated as well as younger and older Americans), suggesting that this misperception is widespread and persists through college and beyond.

We then turn to the implications of stereotyping for students' choice of major. We first describe a stylized model of major choice in which three factors contribute to students' decisions: non-career factors (e.g., difficulty/enjoyability of classes), expected salary, and an individual-specific non-monetary preference for one "preferred" career. We estimate the model using survey data from our OSU sample and find that, among career factors, non-monetary considerations dominate: students' choices are barely sensitive to (their beliefs about) expected salary across majors. In contrast, the chance to attain careers that students value for non-pecuniary reasons plays a large role, comparable in magnitude to that of non-career factors. Further, these preferences appear closely linked to stereotyping: we estimate that the preferred career of over half of students is their most likely major's stereotypical job. Errors in beliefs about occupations—and in particular stereotyping—therefore potentially carry substantial welfare implications.

Do students' beliefs about stereotypical jobs matter for their schooling decisions, as the above discussion suggests? We address this question by conducting a field experiment testing a light-touch intervention providing information meant to combat stereotyping: namely, statistics on the actual past distribution of careers by major. We investigate the impact of such information on students' self-reported intentions as well as administrative data on their course enrollments and college major declarations up to two years later. Consistent with the choice model described above (where students had especially strong preferences for their most likely major's stereotypical career), we find that the effect of this information depends critically on students' heterogeneous preferences. They react negatively to news that the stereotypical career of their *ex ante* most preferred major is unlikely, whereas if anything they react *positively* to news that less preferred majors' stereotypical careers are unlikely. More precisely, we estimate that changing beliefs about the likelihood of attaining the stereotypical career of students' top-ranked or second-ranked major by 10 percentage points (p.p.) would change their beliefs about the likelihood of pursuing that major by -3.5 p.p. ($p < 0.05$) or +2.1 p.p. ($p = 0.17$), respectively, and we can reject that these two effects are equal ($p < 0.05$). The same shift in stereotypical beliefs would change later course enrollment in that major's subject by -2.8 credit hours ($p < 0.10$) or +4.2 credit hours ($p < 0.05$, p-value for difference < 0.01) and the probability of having declared a major in that subject a year and a half later by -6.6 p.p. ($p = 0.25$) or +10.4 p.p. ($p < 0.01$, p-value for difference < 0.01). The treatment also had a large and significant effect on *when* students declare a major: treated students are 9.2 p.p. less likely ($p < 0.01$) to have officially declared a major three semesters after the intervention.[2]

Finally, we turn to the origins of these beliefs: why do students stereotype majors? We

[2]If anything, treated students are slightly more likely (though not significantly so) to still be taking classes two years later, so the effects on major declaration do not appear driven by students dropping out or becoming discouraged with college. Given the ongoing effects on the timing of major declaration, however, effects of *which* major students declare are of course preliminary.

explore one hypothesis: that stereotypes simply come to mind easily. To formalize this idea, we present a simple model of belief formation. In our framework, students assess the likelihood of career c after majoring in M by trying to think of people they know or have heard of who have c and M. The more such people they can think of, the higher their belief about $P(c|M)$. Because beliefs depend on who comes to mind in this way, two well-established principles of recall play a key role (Kahana, 2012; Bordalo, Conlon, et al., 2023). First, recall is limited, meaning that memories must compete for recall. Thus, factors that make one person easier to recall thereby make others harder to recall. Second, recall is associative, which the model captures by assuming that people come to mind more easily when they are similar to (i.e., share a career or major with) the hypothesis the student is considering. These features of memory drive biases in beliefs.

We first show that the model predicts stereotyping endogenously. When thinking of one major, people with its stereotypical career face little competition for recall because their distinctive job makes them dissimilar to those with other majors. People with stereotypical jobs thus come to mind easily. In contrast, those with non-stereotypical jobs are difficult to recall because their careers—which are common across majors—make them similar to people with other majors. Intuitively, when trying to think of biology majors, it is easy to think of doctors but hard to think of business people: there are few doctors who studied something else, but many business people who did.

To further test this explanation for why students stereotype majors, we examine other predictions from this same model. First, among the non-stereotypical jobs for a given major, it predicts that students will on average underestimate more common careers while overestimating rare ones. Second, unconditional on major, students should exaggerate the likelihood of both rare jobs and also those that are concentrated within particular majors. Third, people with whom students have more experiences—e.g., their parents and other close role models—should come to mind more easily, and thus those people's careers and majors should systematically predict heterogeneity in students' beliefs about the distribution of careers by major. Note the model predicts such effects not only on students' expectations about the jobs the they themselves will have but also their beliefs about the general population. We find broad empirical support for all of these predictions in our data.

Our study adds to a rich literature on beliefs and human capital investment.[3] In the

[3]See, for instance, studies linking beliefs to effort in secondary school (Jensen, 2010) and in college

domain of college major choice, many papers have shown that subjective expectations measured in surveys robustly predict later life outcomes and help to explain students' human capital decisions (e.g., Arcidiacono et al. 2012, Wiswall & Zafar 2021). These studies tend to focus on beliefs about average salary conditional on major (e.g., Betts 1996, Baker et al. 2018, Conlon 2021).[4] To our knowledge, we are the first to systematically investigate biases in beliefs about the distribution of occupations by major and, therefore, the first to document stereotyping in this domain. Our results also echo a small but growing literature showing that students in vocational programs in developing countries greatly overestimate their own employment prospects post-graduation (e.g., Bandiera et al. 2023), and that reducing these misperceptions can have positive effects (e.g., Alfonsi et al. 2023).

This paper is also linked to a growing body of evidence on belief formation across economic contexts. Many studies document how stereotypes can distort beliefs about race, immigrants, political parties, and gender.[5] One notable recent example is Bursztyn et al. (2023), who find that beliefs about social attitudes exhibit both stereotyping—exaggerating the differences between men's and women's attitudes—but also overestimation of rare attitudes. Both these findings have direct parallels in our data. We are, to our knowledge, novel in applying the logic of stereotyping—and systematically testing its implications—in a high-stakes economic context outside the domain of demographic groups. Further, our conception of stereotyping as a byproduct of associative recall also speaks to the broader literature studying the origins of beliefs and their biases. Recent work in economics increasingly studies the role of memory in belief formation and choices, though ours is among the first papers we are aware of to test the predictions of a model of associative recall in a field context.[6]

(Delavande et al., 2020), to the choice of which university to attend (Delavande & Zafar, 2019), to the choice of whether to attend university at all (Boneva & Rauh, 2019), and to parents' investment in their children's schooling (Dizon-Ross, 2019). See Giustinelli (2022) for a review.

[4] One 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.

[5] See, for example, Alesina et al. (2022), Bordalo et al. (2019), Coffman et al. (2020), Bohren et al. (2019), Exley et al. (2022), and Carlana (2019).

[6] For recent work in economics studying the role of memory, see Wachter & Kahana (2019), Bordalo et al. (2020), Koszegi et al. (2021), Malmendier & Wachter (2022), Fudenberg et al. (2022), Exley et al. (2022), and Sial et al. (2023). Enke et al. (2021) and Graeber et al. (2022) use lab experiments to study how associative memory distorts belief updating. In contemporaneous work, Bordalo, Burro, et al. (2023) study how associative recall can affect beliefs about the lethality of Covid-19, and Charles (2022b,a) shows that companies plausibly associated in memory—by alphabetical proximity or earnings announcement dates—cue

2 Motivating Evidence

To provide motivating evidence on the beliefs of first-year college students in the US, we use 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.[7] 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.[8] In that sense, we call this sample nationally representative of incoming college freshmen.

We focus on two questions from the Freshman Survey. First, students are asked to mark their "probable field of study" from a list of around 80 options, including "Other" and "Undecided."[9] We group these fields into 10 major groups (plus "other" and "undecided"), as shown in Table A.II. Similarly, students are asked to report their "probable career occupation" from a list of approximately 45 options, which we group into nine occupation categories (plus "other," "non-employment," and "undecided") as shown in Table A.III. The qualitative nature of the Freshman Survey—i.e., asking students to pick which job is their "probable career" rather than eliciting probabilistic beliefs about a well-defined event—of course raises questions about how to interpret students' responses. We address this and other issues in the Section 3.2 and use the Freshman Survey merely as motivating and suggestive evidence.

To compare students' expected careers to the actual distribution of occupations, we use the Annual Social and Economic Supplement (ASEC) to the Current Population Survey (CPS) data from 1976 onward (Flood et al., 2021). We restrict the data to those aged 33 to

each other, inducing correlated trading behavior.

[7]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.

[8]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.

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

37, because by this time the vast majority of people are no longer students and have started their career.[10] 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.IV).

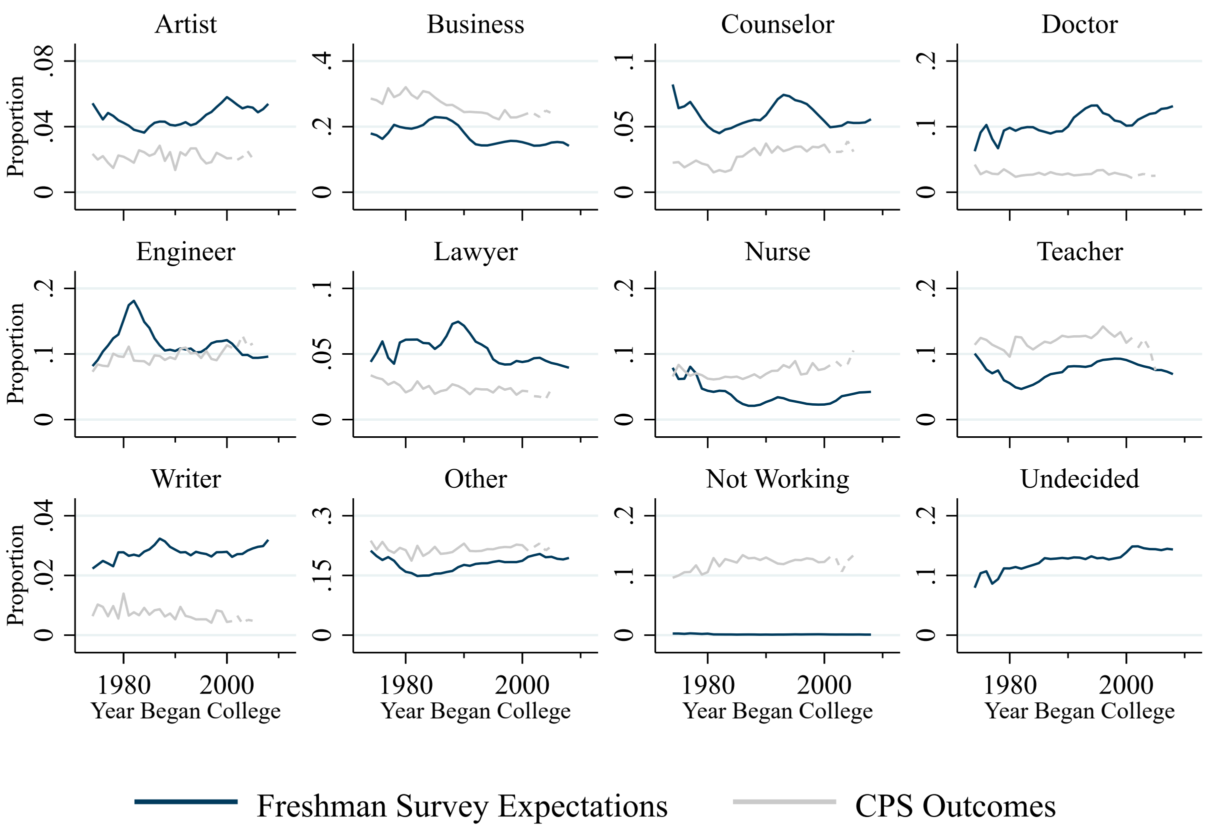
Figure 1 shows large, systematic, and persistent differences between the careers that freshmen expect to attain and the actual occupations they go on to have. The blue lines show the share of first-year students each year who expect to have each career. The gray lines show the share of college graduates in the same cohort that are working in that occupation in the CPS. Panel A of Table A.V shows the corresponding share expecting and actually working in each career, pooling across cohorts. Around 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%).[11] Focusing in on doctors, these rates imply that at least 8% of college freshmen in the U.S.—about 150,000 students every year in recent cohorts—expect to become doctors but will not. Note that the fact that students could list their probable career as "Undecided" makes these findings of overestimation of certain careers all the more extreme, since of course there is no corresponding category in the CPS. Next, though 12.1% of college graduates are not working for pay, only 0.2% of students report their probable career as "Homemaker," "Stay-at-Home Parent," or "Unemployed." Of course, a student who expects to drop out of the labor force temporarily (e.g., to take care of a child) may still reasonably consider their career to be something other than "Homemaker," so we take this result to be merely suggestive.[12]

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

[11]Panel B of Table A.V 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.VI 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.

[12]Note that if we restrict the CPS data to employed college graduates, this does not substantially change the conclusion of overestimation of the careers previously mentioned.

Figure 1: Career Expectations vs. Outcomes Over Time



Notes: Figure 1 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 dotted when CPS outcomes begin to only include graduates younger than 37.

3 Do Students Stereotype Majors?

We hypothesized that these patterns in the Freshman Survey were driven by stereotyping: students may exaggerate the likelihood that majors lead to their distinctive jobs. Following Bordalo et al. (2016), we define stereotypical-ness using *distinctiveness*: i.e., a career c is the stereotype of major M if it maximizes $p_{c|M}/p_{c|-M}$. The careers that are most stereotypical 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.VII). Such a bias would lead students to underestimate their chances of having outcomes that are common alternatives to many major's stereotypical job. Empirically (see Panel D of Table A.VII) these tend to be teaching, business, and non-employment, exactly the outcomes that students in the Freshman Survey appear to neglect.

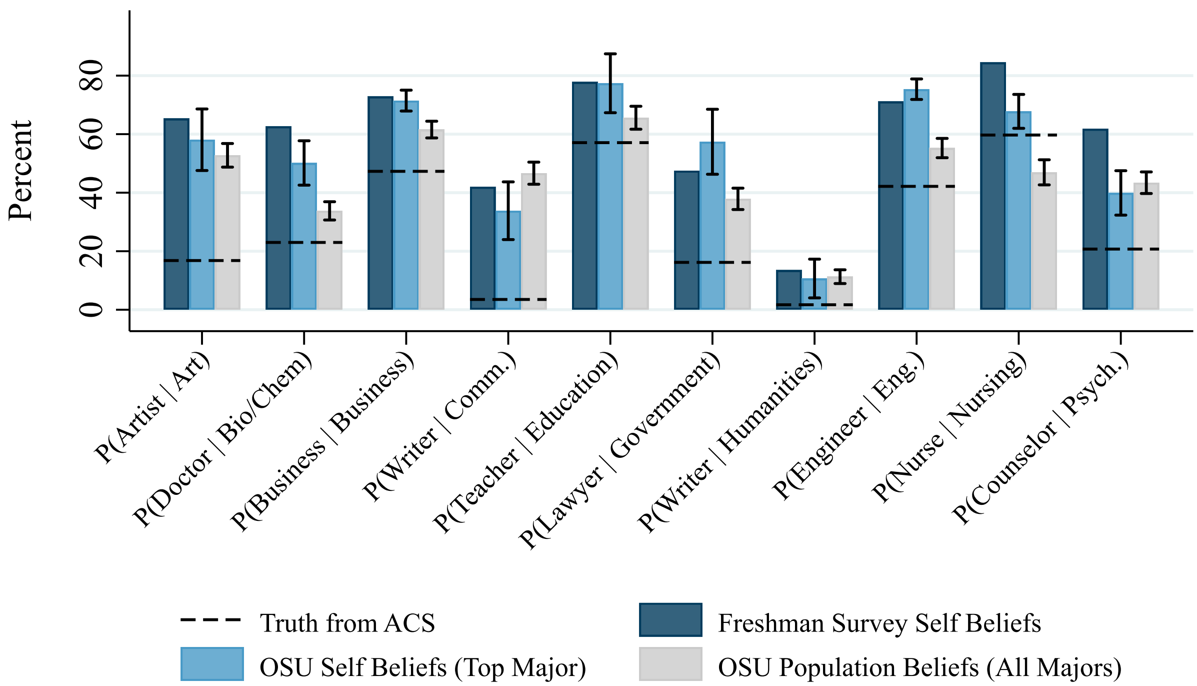
3.1 Freshman Survey Respondents Overwhelmingly Expect Their Major's Stereotypical Career

To provide a first piece of suggestive evidence in favor of our stereotyping hypothesis, the dark blue bars in Figure 2 show the fraction of students in the Freshman Survey who list their "probable" major's most stereotypical career as their "probable" career occupation. The dotted lines show the true fraction of college graduates with each major who are working in its stereotypical 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 stereotypical 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.

3.2 Isolating Stereotyping from Other Mechanisms

The patterns described in Sections 2 and 3.1, while consistent with stereotyping, by themselves could reflect several over potential mechanisms. In this section, we describe these alternative explanations and a survey we designed to isolate the role of stereotyping from them. To administer these surveys, we partnered with the "Exploration" program at the Ohio State University (OSU) in Fall Semester 2020. Entering OSU students are automatically enrolled in this program if they have not yet officially declared a major. The 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,

Figure 2: Exaggerating Stereotypical Careers



Notes: Figure 2 presents average statistics regarding the most stereotypical career (as defined in section 3) 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 stereotypical 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 stereotypical 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.

and help planning course schedules. The program does not, however, provide students' quantitative information about employment outcomes by major. Students received extra credit in this course for completing our survey, ensuring a very high response rate of around 80%.

Column 2 of Table A.I gives self-reported demographic information about the 755 respondents in this sample, which we call our "2020 OSU data." The sample is broadly similar demographically to the overall student body at OSU, though with a somewhat higher share of first-generation college students. It is also similar to the Freshman Survey along gender, ethnicity, first-generation status, and self-reported family income. See Appendix B for more details about the survey and implementation. The survey began by displaying the ten groups of college majors (henceforth, just "majors") and asking students to rank them by how likely they thought they were to graduate from OSU with a degree in each. It then asked them detailed questions about a subset of these majors.

Qualitative vs Quantitative Expectations

One difficulty in interpreting the Freshman Survey stems from the fact that it asks student to mark one job as their "probable career occupation." One might reasonably worry that this way of eliciting expectations makes it difficult to interpret the patterns we have documented as biased beliefs. For example, if students tend to mark an occupation as their "probable" career when in reality they think they only have a relatively small chance of working in that job, then we could be overstating the extent of bias in students' true beliefs. To avoid this issue, we ask students in the OSU sample quantitative probabilistic questions about a well-defined event, allowing them to express uncertainty precisely. Specifically, for each 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." Students' answers had to add up to 100%.

The light blue bars in Figure 2 show the average answer that students who ranked each major highest gave about their likelihood of working in that major's stereotypical 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 stereotypical 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 that the qualitative nature of the questions in the Freshman Survey does not explain the patterns documented above.

Confidence and Selection

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 likelihood of attaining stereotypical careers. We address this issue in the OSU survey by asking students not only about their own future outcomes conditional on major but also about the US population as a whole. More precisely, *before* asking students about their own future jobs, the survey asked them to give their "best guess about the percent of Americans aged 30-50 (note, not just from Exploration or OSU) who graduated with a major in X that are..." It then listed the same 11 outcomes. We call these students' "population beliefs," in contrast to their "self beliefs" about their own outcomes. These questions were designed such that we could compare students' beliefs 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 systematically select into majors depending on their beliefs, then biased beliefs conditional on pursuing a particular major could reflect this selection process rather than a more general underlying feature of students' beliefs. 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. Note that, though we describe this mechanism as "selection," it additionally encompasses any proposed explanation for biased beliefs that rests on students with different majors holding systematically different beliefs. For example, one might worry that our results are driven by a form of wishful thinking in which students hold mistaken beliefs in order to *ex post* justify their chosen major. Or, one could imagine that academic departments may try to convince students taking introductory classes that stereotypical outcomes are more likely than they are in an attempt to increase enrollment.

To address this selection issue, the OSU survey asked students, not only about their top-ranked major, but also 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.[13] Throughout the analyses to follow, we employ such weighting whenever we pool beliefs about students' top-ranked majors with

[13] 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).

beliefs about their lower ranked majors, though in practice these weights have little impact on our main results.

The gray bars in Figure 2 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 stereotypical career is substantially (and statistically significantly) more common among graduates with that major than it actually is.[14] 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 or selection. They also suggest that students' believe the causal effect of majors (expressed by their self beliefs) are quite similar to the (perceived) cross-sectional differences in occupations across majors.

Biases toward Common or Salient Careers

Next, students could systematically overestimate more common careers. For example, if stereotypical jobs tend to be common and students simply exaggerate or latch onto more common outcomes by major, this could superficially look like stereotyping. And finally, the role of stereotyping could be confounded if students underestimate the unconditional likelihood of careers that tend not to be stereotypical of most majors. For example, perhaps many non-stereotypical occupations are less salient (independent of major) and therefore students have not heard of many of them.

We address these final two issues 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

[14] 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 stereotypical outcome. In fact, 4% of nursing majors work as doctors, but the average belief is 23%.

in that career, and μ_c^i 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 less salient careers irrespective of major). For example, if students simply neglected non-employment or "other" jobs but otherwise responded only to true frequency, this would be captured by μ_c 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 stereotypical career is 29 percentage points higher ($p < 0.01$) than similarly frequent but non-stereotypical 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 stereotypical pairs (these points correspond to the gray bars in Figure 2), but within stereotypical 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-stereotypical outcomes, such that sufficiently rare non-stereotypical outcomes are on average overestimated while sufficiently common ones are neglected (we return to these facts in Section 5). These trends explain why the estimate of γ from equation 1 is less than one.

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 Stereotypical)	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 Stereotypical) is a dummy variable indicating whether an occupation is the most stereotypical 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.

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 reveals why our estimate of θ is so large and significant despite the inclusion of career-fixed effects.

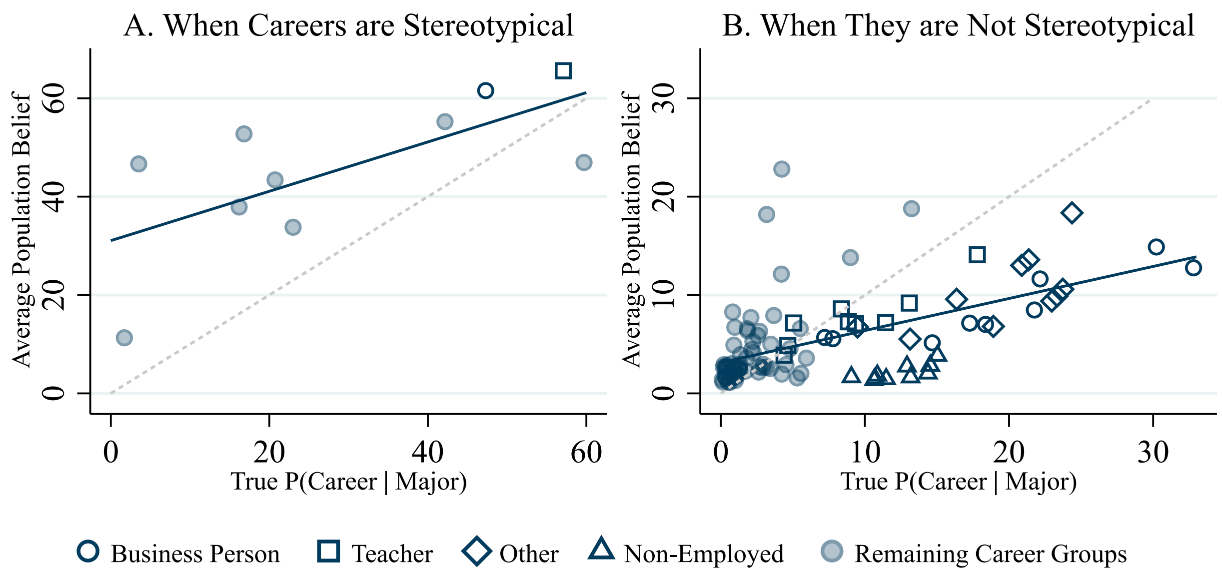
Stereotyping thus remains apparent after isolating it from the other potential mechanisms described above. Moreover, Table A.VIII describes a Shapley-Sharrocks decomposition, which shows that stereotyping is also *more important* than these other mechanisms, in the sense that it explains a greater proportion of the variance of students' beliefs (see notes of Table A.VIII for details).

Our Survey Sample and External Validity

Finally, one might wonder whether these mistaken beliefs are unique to the sample of Ohio State freshmen we happen to survey. For example, perhaps stereotyping is primarily prevalent only in a demographic group that happens to be overrepresented in our Ohio State sample. Or, one might worry that stereotyping may be unique to undecided college freshmen or to unincentivized surveys, either because students quickly learn more about career outcomes or because stereotyping is primarily a result of inattentive survey responses. To address the first possibility, 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). To address the second possibility, we conducted an online survey of US adults, recruited through CloudResearch, where respondents were incentivized to provide accurate responses to the same population-beliefs questions as our OSU survey (see Appendix B for details on this survey). Columns 8-11 of Table 1 shows somewhat smaller, but still large and statistically significant stereotyping even among these respondents, both when restricting to college- vs non-college-educated and younger vs older adults. We conclude that stereotyping appears to be a broad-based phenomenon not restricted to particular demographic groups, to college freshmen, or to unincentivized surveys.

Finally, Figure 2 suggests a close parallel between mistaken beliefs in the Ohio State surveys and the patterns of career expectations in the nationally representative Freshman

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 stereotypical of the major, and Panel B restricts them to when the career is not stereotypical 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.

Survey. We now attempt to formalize that comparison, asking whether the differences between actual and expected careers in the qualitative Freshman Survey (shown in Figure 1) are consistent with being driven by biases in population beliefs of the same magnitude as in our Ohio State sample. To do so, we conduct the following back-of-the-envelope calculation. Let $\pi_{c|M}$ be the average OSU population belief about the fraction of people with major M who are working in career c. Let κ_M be the fraction of the Freshman Survey respondents that say they expect to graduate with major M.[15] We can then calculate what we call the "implied error" about the probability of working in career c as shown in equation 2, where p_c is the true fraction working in c.

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

We can then 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 is asking whether the Freshman Survey respondents' expectations are what we would expect if they held the same (population) beliefs as the OSU sample. As Figure A.I shows, there is a robust positive relationship between actual "errors" in the Freshman Survey data and ImpliedError_c. The correlation between between implied and actual error is 0.81 and is highly statistically significant ($p < 0.01$). An OLS regression of the error in the Freshman Survey data on the implied error has a coefficient of 0.84 ($p < 0.01$), with an R^2 of 0.71. 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.

3.3 More Confident Students Stereotype More

Why might stereotyping persist 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. This question echoes similar issues surrounding whether and

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

when other behavioral biases are likely to persist and affect aggregate outcomes. Enke et al. (2022) suggest that biases are more likely to persist when they are positively correlated with decisions-makers' confidence that they are making the correct decision.

In 2021, we administered two similar surveys among a new cohort of the same Exploration program at Ohio State (see Appendix B for more details). The first of these also asked students' beliefs about the frequency of careers conditional on majors. In addition, immediately after each such question, 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 stereotypical careers ($p < 0.01$, regression not shown). 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.

4 Implications of Stereotyping

4.1 A Stylized Model of Major Choice

What are the implications of stereotyping for students' choices and for their welfare? The answer, of course, depends on students' preferences and whether stereotyping distorts their educational decisions. In this section, we begin by describing and estimating a stylized model of major choice. We then analyze data from a field experiment attempting to reduce stereotyping.

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

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

In equation 3, $w^i_{c,M}$ is the salary i believes they would earn conditional on c and M, and α is the (homogeneous and constant) marginal utility of income. We allow for i to have idiosyncratic non-pecuniary preferences over jobs, which are denoted by β_c^i. Next, μ_M^i indicates the known non-labor-market benefits the student would derive from majoring in M (enjoyment of classes, parental approval, etc.). Finally, ν_M^i indicates an unrealized preference shock, whose distribution i knows but whose realized value they do not.

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

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

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

To summarize, the paramaters 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 students reported probability of graduating with a major in M, and $\widehat{\pi_M}$ be the model's prediction, given ξ, of that probability. Note that this is a random variable given the distribution of non-labor-market preferences. Then, let $L_{i,j}(\xi)$ be the likelihood given the model and parameters ξ that $\log(\widehat{\pi}_{M_j}/\widehat{\pi}_{M_4}) = \log(\pi_{M_j}/\pi_{M_4})$.

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

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

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

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

In contrast, column 1 of Table A.IX 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 = $[\$4500, \$18500]$) to increase their chances of working in their preferred career by one percentage point.[16] Note however that this extremely large WTP is driven by the low estimates for salary preferences rather than (arguably) by preferences for careers being unrealistically large. To make this claim more precise, note that the variance of non-career preferences for majors (μ^i_M) is estimated

[16] The remaining columns of Table A.IX show estimates of modifications to this baseline model (see table notes for details). Across all of them we find large preferences for working in preferred careers: 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.

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 stereotypical 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 particularly about stereotypical careers. Over half (50.5%) of students' estimated preferred careers are the stereotypical job of the major they had initially ranked highest. In contrast, only 10.1% percent of students' preferred careers are the stereotype of their second-ranked major.

These analyses suggest that students perceive their eventual career to be quite important, over and above the income it may generate, when deciding what to study in college. We conclude this section with two corroborating pieces of evidence. First, in our 2021 survey, we asked the control group of the experiment (described below) to rate on a scale from 0 to 100 how important various factors were to them as they decided what to major in. The "effect on what job I will get after college" received the highest average ranking (83 out of 100), above enjoyability of classes (70), difficulty of classes (58), and the opinions/choices of family, friends, and peers (all below 30). Second, in Appendix B we analyze data from the 2013 National Survey of College Graduates, which asks employed respondents about the salary of their job, whether it is related to their highest degree (which we use as a proxy for stereotypical-ness), and their overall satisfaction with this job. We regress this satisfaction variable on salary, degree-relatedness, and demographic controls. We find that having a job unrelated to one's highest degree (conditional on salary) is associated with an 11.2 percentage point increase in the likelihood of being dissatisfied with that job. In contrast, a $10,000 increase in salary is only associated with a 0.4 percentage points decrease in likelihood of job dissatisfaction. Thus, at least in the cross-section, the relationship between salary, degree-occupation match, and satisfaction qualitatively mirrors students' revealed preferences as estimated by our choice model above.

4.2 A Field Experiment Providing Information

The analysis in the previous section suggested that beliefs about stereotypical jobs may have a large effect on what students choose to study, but of course such results are only correlational. We now turn to causal evidence from a field experiment. We tested a low-cost, light-touch information intervention embedded in the second 2021 OSU survey. The survey began by asking students the percent chance that they would graduate with the two majors they selected as being most likely to pursue (henceforth, their "top-ranked" 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. Students 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.[17] 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 stereotypical career. Figure A.II shows the average revision, among those in the treatment group who saw information about

[17]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 stereotypical career.

each major, in students' beliefs about their own chances of attaining that major's stereotypical 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 stereotypical 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 ($p < 0.01$).[18] This result is perhaps not surprising: each module in the treatment group necessarily provided many distinct pieces of information (about each career group and 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 stereotyping changes students' intentions and behavior. We therefore construct a simple measure of treatment intensity: the leave-out mean reduction by major in treated students' self-beliefs about their chances of having that major's stereotypical career. We view this specification as providing a data-driven approach to test the causal effect we hope to estimate. More precisely, Table 2 shows OLS estimates of equation 6:

$$Y_{i,M} = \beta \cdot T_i \cdot AvgReduction_{i,M} + \alpha_1 Intentions_{i,M,pre} + \alpha_2 Credits_{i,M,pre} + \epsilon_{i,M} \qquad (6)$$

In the above equation, $Y_{i,M}$ is an outcome variable of interest for a given major M. T_i is a dummy variable for treatment status. $AvgReduction_{i,M}$ is the measure of treatment intensity described above: the leave-out mean reduction in self-beliefs about the stereotypical career for major M. Finally, we always include two controls: $Intentions_{i,M,pre}$ is i's pre-treatment belief about their likelihood of graduating with major M, and $Credits_{i,M,pre}$ is the number of credit hours their (pre-treatment) Fall 2021 class scheduled contained.

Column 1 of Table 2 looks at participants' first-ranked major. The dependent variable

[18] This p-value comes from regressing revisions in self-beliefs about a major's stereotypical 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.

is students' post-treatment belief about their likelihood of graduating with that major. We see a statistically significant coefficient of -0.345 ($p < 0.05$) on the interaction of treatment intensity and treatment status. The negative coefficient is the expected direction: treatments that reduce students' perceived chances of attaining their first-ranked major's stereotypical career reduce students' intentions toward that major. Furthermore, the magnitude of this effect is substantial: it implies that information reducing beliefs about stereotypical careers by 10 percentage points (p.p.) reduces intentions toward student's first-ranked majors by 3.45 p.p.

Column 2 of Table 2 looks at each student's second-ranked major, where if anything we see the opposite pattern: treatments that reduce beliefs about attaining the stereotypical job of students' second ranked major directionally *increase* intentions toward it (coefficient = 0.208, $p = 0.17$). This difference between first- and second-ranked majors is statistically significant ($p < 0.05$). The direction of this difference is also what we would expect from Section 4.1, where we estimated that students tend to have strong non-monetary preferences for their first-ranked major's stereotypical career. That result would suggest that learning one's first major's stereotype is unlikely should count as bad news, whereas learning that one's second-ranked major's stereotype is unlikely would not (and could even count as good news if students do not particularly value that job).

Columns 3 and 4 of Table 2 repeat this analysis but changing the dependent variable to the number of course credit hours the student enrolled in post intervention (i.e., between Spring 2022 and Fall 2023) as measured by administrative data from Ohio State Registrar's Office. We see a similar pattern as with intentions. Information that reduces stereotypical beliefs about a student's first-ranked major reduces course-taking in that major (interaction coefficient = -28.0, $p < 0.10$) whereas the reverse pattern holds for second-ranked majors (coefficient = 41.7, $p < 0.05$). Again the difference in these coefficients is statistically significant ($p < 0.01$). The magnitude of these estimated effects are also large: they suggest that changing beliefs about the likelihood of a stereotypical career by 10 percentage points changes course taking by -2.8 or +4.2 credit hours, which is roughly the number of credit hours in a single semester-long class (between 3 and 5 for typical classes).

Finally, columns 5 and 6 of Table 2 show similar regressions but where the dependent variable is a indicator for having declared a major by Spring 2023 (the latest semester for which these data are available) in the students' first-ranked (column 5) or second-ranked

Table 2: Effects of Reducing Stereotyping on Major Intentions and Later Choices

	Post-Treatment Belief P(M)		Post-Treatment Course Credits		Major in Spring 23	
	1st	2nd	1st	2nd	1st	2nd
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment x Average Reduction in $\pi_{c\|M}$	-0.35**	0.21	-28.05*	41.65**	-0.66	1.38***
	(0.17)	(0.15)	(16.86)	(16.57)	(0.57)	(0.37)
Treatment	-0.02	-0.01	0.13	-1.63*	-0.03	-0.06***
	(0.01)	(0.01)	(1.11)	(0.93)	(0.04)	(0.02)
Average Reduction in $\pi_{c\|M}$	0.01	-0.05	45.81***	-19.56*	1.52***	-0.26
	(0.12)	(0.10)	(11.33)	(10.19)	(0.40)	(0.23)
Pre-Treatment Belief P(M)	0.96***	0.83***	22.74***	16.32***	0.66***	0.39***
	(0.02)	(0.03)	(1.96)	(2.69)	(0.06)	(0.08)
Pre-Treatment Course Credits	0.00	0.00	0.93***	0.91***	0.01	0.01**
	(0.00)	(0.00)	(0.18)	(0.16)	(0.01)	(0.00)
Constant	0.02	0.03***	-0.30	3.37***	-0.09*	0.01
	(0.01)	(0.01)	(1.37)	(0.87)	(0.05)	(0.02)
Observations	783	783	783	783	783	783
p-value: 1st vs 2nd Major Interaction Equal		0.014		0.003		0.003

Notes: Table 2 presents OLS regressions including data from the 2021 OSU sample. The dependent variable in columns 1 and 2 is students' post-intervention belief about the percent chance that they will graduate with their first-ranked major (column 1) or second-ranked major (column 2) (recall that these rankings are according to how likely the student thought they were to pursue each major). The dependent variable in columns 3 and 4 is the number of course credit hours that students enrolled in in their first-ranked major (column 3) or second-ranked major (column 4) post-intervention (between Spring 2022 and Fall 2023). The dependent variable in columns 5 and 4 is an indicator for whether students' declared major in Spring 2023 is in their first-ranked major group (column 5) or second-ranked major group (column 6). "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 stereotypical career for students' first-ranked major (odd columns) or second-ranked major (even columns). "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 Course Credits" is the number of course credit hours that participants enrolled in in their first-ranked major (odd columns) or second-ranked major (even columns) during Fall 2021. Robust standard errors in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels.

major (column 6). We see directionally similar effects as with intentions and course taking: a coefficient of -0.66 for first-ranked majors ($p = 0.25$) and 1.39 for second-ranked majors ($p < 0.01$), with the difference between first- and second-ranked majors again being statistically significant ($p < 0.01$). These estimates suggest that changing beliefs about the likelihood of a stereotypical career by 10 percentage points changes the likelihood of declaring a major by -6.6 or +13.9 percentage points. We note that these coefficients should be interpreted cautiously, however. Column 1 of Table A.X shows that the treatment had a large negative effect (-9.2 p.p., $p < 0.01$) on whether participants have officially declared a major by Spring 2023. The effects on *which* major they've declared (in Table 2) are therefore preliminary. We do not see corresponding effects on whether participants are still taking courses in Fall 2023 (column 2 of Table A.X, coefficient = +1.6 p.p., $p = 0.48$), so this effect does not appear driven by dropping out, but rather by waiting longer before declaring a major.

Table 2 treats students' first- and second-ranked majors separately, but of course we might expect spillovers of learning about one major on intentions/choices regarding another. Table A.XI regresses outcomes regarding students' first-ranked major (odd columns) or second-ranked major (even columns) on treatment intensity and controls regarding both majors. As before, learning that the stereotypical job for one's first-ranked major is less likely is "bad news" (i.e., reduces intentions/choices) about that major, but in addition it appears to be "good news" (increasing intentions/choices) about one's second ranked major. Conversely, learning the stereotypical job for one's second-ranked major is less likely is "good news" for one's second ranked major and "bad news" for one's first-ranked major. These results suggest that the average student particularly values her first-ranked major's stereotypical job but not that of her second-ranked major.

The results described above investigate the effect of changing students' beliefs about the likelihood of attaining stereotypical jobs by leveraging variation, across modules, in how much stereotypical beliefs changed in response to the information. Columns 3 to 8 of Table A.X 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 whether changing students' beliefs about their chances of stereotypical careers affects choices, but rather the average effect of being assigned to the treatment group. These average effects are of course generally smaller (and often not statistically significant), reflecting the fact that our light-touch intervention providing population statistics for some majors had only small (or null) effects on what jobs students expected themselves to have.

5 Why do Students Stereotype Majors?

What drives stereotyping in our context? More broadly, how do students form their career expectations? To explore these questions, in this section we describe a simple model of belief formation. The core idea that the model formalizes is that the more easily a student can think of or recall people with a given career path (i.e., the more easily they come to mind), the more likely the student will think that career path is. Within this framework, well-established regularities in human recall then have scope to distort students' beliefs. In particular, we apply the basic "similarity and interference" framework of Bordalo, Conlon, et al. (2023), which draws on memory research in psychology (Kahana, 2012). We first show that this model generates stereotyping endogenenously. We then derive additional predictions that we test using survey data from our OSU samples. We note at the outset that this model is not the only plausible explanation one could give for why students stereotype majors.[19] As we show, however, our framework provides a parsimonious way of organizing and interpreting a fairly broad range of patterns in the data, over and above stereotyping.

5.1 Setup

We 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

[19]Other potential mechanisms that might contribute to stereotyping include forms of base-rate neglect (Benjamin et al., 2019), the representativeness heuristic (Tenenbaum & Griffiths, 2001), a desire for or expectation of "sense making" (Chater & Loewenstein, 2016), and versions of projection bias (Loewenstein et al., 2003). We note, however, that none of these explanations predict the other empirical patterns that we document in this section.

major M. The student's probabilistic beliefs about H, shown in equation 7, are then the plausibility of H normalized such that their beliefs about all relevant hypotheses sum to one.

$$\pi_c = \frac{F(H_c)}{\sum_{z\in\{a,b,\ldots\}} F(H_z)} \qquad \pi_{c|M} = \frac{F(H_{c,M})}{\sum_{z\in\{a,b,\ldots\}} F(H_{z,M})} \tag{7}$$

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. Second, the student simply checks whether the person they recalled matches the career and/or major associated with the hypothesis they are assessing. The more such "successes" they have, the more likely a hypothesis will seem.

This setup assumes that, for any person e that the student might recall when assessing hypothesis H, e either counts fully in favor of H (when $e \in H$) or fully against H (when $e \notin H$). Two potential relaxations of this assumption bear mentioning. First, students may not know with certainty whether $e \in H$. In such a case, it may seem more plausible to assume that e counts as evidence for H according to how likely the student thinks it is that $e \in H$. For example, if they do not know for sure what their doctor's major was, they may add partial evidence in favor of the hypothesis that biology majors become doctors according to how *likely* they think it is that their doctor had that major. We show in Appendix C that if these beliefs are unbiased, then this alternative is equivalent to the model described in the main text.

Second, the model treats all failures to recall someone consistent with a given hypothesis equally, which may be unrealistic. For example, suppose the student is assessing how many petroleum engineering majors work as engineers in the oil industry. Perhaps they cannot think of anyone with exactly that major or job, but they can think of other types of engineering majors and the sorts of jobs they went into. Taken literally, equation 9 would not allow the student to count these people as evidence for the hypothesis they are considering. This is clearly an oversimplification. In Appendix C, we describe a more general version of the model, which allows for this sort of extrapolation (Kahneman & Tversky 1981, Gilboa & Schmeidler 1995, Bordalo, Burro, et al. 2023). We return to this point below when discussing the effect role models have on students' beliefs.

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 8 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.

$$r(e, H) = \frac{a(e, H)}{\sum_u a(u, H)} \tag{8}$$

The numerator of equation 8 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).

After a person e comes to mind when the student is assessing H, the student checks whether e has the career/major associated with H. We assume the student repeats this process of recall-and-checking many times, counting up the number of "successes" they recall for each hypothesis. Their assessed plausibility then converges in probability to equation 9.

$$F(H) = \sum_{e \in \mathcal{D}} r(e, H) \mathbb{1}(e \in H) = \frac{E\Big[a\big(e, H\big) \mathbb{1}(e \in H)\Big]}{E\Big[a(e, H)\Big]} \tag{9}$$

This model naturally nests the rational-expectations benchmark. In particular, if $a\big(e, H\big)$ is constant, 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. We explore two factors that allow availability to deviate from this benchmark: the facts that human recall is associative and frequency-based (see Kahana 2012 for a review of the related memory research in psychology). For both of these forces, we first derive predictions and then test them in our OSU survey data.

5.2 Associative Recall

We first incorporate into the model the fact that human recall is associative (i.e., similarity-based). To do so, 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. These assumptions are reflected in equation 10.

$$a(e, H) = S(e, H) = \frac{1}{|H|} \sum_{u \in H} s(e, u) \tag{10}$$

Equation 10 implies that when people consistent with H are mostly similar to each other, examples consistent with that hypothesis will come to mind easily. For example, one might struggle to think of college dropouts but easily think of college dropouts who became billionaire tech company founders—despite the latter being a subset of the former—because college dropouts are much more heterogeneous. However, these assumptions also imply that even people *in*consistent with H can be highly available if they are similar in some dimensions to those who are consistent. For instance, when thinking of college dropouts who became billionaire tech company founders, CEOs who actually finished college might come to mind "erroneously" and need to be discarded.

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 11.

$$s(e, u) = \delta_c^{\mathbb{1}(c(e) \neq c(u))} \times \delta_M^{\mathbb{1}(M(e) \neq M(u))} \tag{11}$$

This feature-based approach to modeling similarity is standard in psychological work (Tversky, 1977), and the exponential form in equation 11 is standard in models with discrete features (Mack & Palmeri, 2020; Evers et al., 2021; Bordalo, Conlon, et al., 2023). We intentionally focus on the simplest possible formulation, where the only features are career and major; an enriched similarity function could yield more nuanced predictions, but such an analysis is beyond the scope of this paper.

Associative recall implies that groups will seem more common when their members are dissimilar from others outside the group (because, if so, they face little interference from similar-but-inconsistent people). In beliefs about careers conditional on majors, this force produces stereotyping. Recall that a career c is stereotypical of major M when it is more common conditional on M than conditional on other majors. Let $p_{a|B}$ be the true frequency of career a conditional on major B. How do beliefs about major A change as we decrease $p_{a|B}$? Despite keeping constant the share of people with major A and career a, this makes such people more dissimilar from those with other majors (because they less frequently share a career). Therefore people with both a and A come to mind more easily, boosting beliefs about $P(a|A)$. Intuitively, when thinking about psychology majors, it is easy to think of counselors and difficult to think of teachers. Interpreted through the mechanism of associative memory, this is because (essentially) all counselors are psychology majors, whereas many teachers majored in something else. Prediction 2 summarizes this result.

Prediction 2: *Conditional beliefs increase in stereotypical-ness: decreasing $P(a|B)$ increases beliefs about $P(a|A)$. More precisely, $\frac{\partial}{\partial p_{a|B}}\pi_{a|A} < 0$ whenever $\delta_c < 1$.*

This same force operates on beliefs about careers *un*conditional on major. Here, people with career c will be less similar to those with other careers if they have a distinctive major as opposed to a major that is common across careers. Thus, careers that are relatively concentrated within a particular major will seem more common. Intuitively, it easy to think of how one becomes a doctor because the path to medicine is so dissimilar from other career paths. In contrast, it is difficult to think of the many heterogeneous ways people become schoolteachers. Prediction 3 summarizes this result.

Prediction 3: *Unconditional beliefs about the frequency of a career increase in the extent to which it is concentrated within particular majors.*

More precisely, let $p_{A|c} = p_{B|c}$ for all careers $c \neq a$. Then equation 12 follows if $\delta_M < 1$:

$$\frac{\partial}{\partial p_{B|a}}\pi_a > 0 \iff p_{B|a} > p_{A|a} \tag{12}$$

Associative recall has one additional implication. Any hypothesis the student considers picks out people on the basis of features that distinguish them from others (e.g., people with a given career and/or major), meaning that they are to that extent mechanically dissimilar from people inconsistent with the hypothesis. This means that, when the student considers each hypothesis, examples consistent with it disproportionately come to mind. For example, when thinking about artists the student tends to succeed in recalling artists, and when thinking about businesspeople they tend to recall people with that profession. This disproportionately benefits rarer hypotheses which, under unbiased random sampling, would more often fail to come to mind. Thus, for both conditional and unconditional beliefs, students *ceteris paribus* tend to overestimate rare hypotheses and underestimate common ones. Prediction 4 states this result.

Prediction 4: *For both conditional and unconditional beliefs, the student,* ceteris paribus, *exaggerates rare careers and underestimates common careers.*

More precisely, for conditional beliefs, let $p_{c,B} = 0$ for all careers c, shutting down stereotyping effects. For unconditional beliefs, let $p_{A|c} = p_{B|c}$ for all careers c, shutting down interference due to differing concentration of careers within majors. Then equation 13 follows if $\delta_c < 1$.

$$\frac{\pi_{a|A}}{\pi_{b|A}} > \frac{p_{a|A}}{p_{b|A}} \iff p_{a|A} < p_{b|A} \quad \text{and} \quad \frac{\pi_a}{\pi_b} > \frac{p_a}{p_b} \iff p_b > p_a \tag{13}$$

5.3 Empirical Tests of Associative Recall Predictions

We test these predictions using the survey data from OSU. Figure 3, which tests Predictions 2 and 4 regarding conditional beliefs, plots for each career-major pair the true share of graduates with that major who are working in that career (x-axis) and the average population belief in the 2020 OSU survey about this likelihood (y-axis). Panel A restricts the data to each major's stereotypical outcome, where we see (consistent with the gray bars in Figure 2), that students significantly exaggerate the stereotype of nine out of the ten majors. Panel B shows all non-stereotypical combinations of careers and majors. In line with Prediction 4, we see that students tend to exaggerate rare careers and underestimate common careers. For example, students exaggerate the share of communication majors who become counselors (6% belief in the OSU sample vs 4% true rate), engineering majors who become doctors (5% vs 1%), or government majors who become writers (8% vs 1%). They underestimate relatively common non-stereotypical outcomes, like working in business, as a teacher, in an "other" job, and not working for pay. A particularly striking case is non-employment (triangles in Figure 3). Students underestimate non-employment for every major: for no major does the average student believe more than 4% of graduates are not working for pay, while in every major the true rate of non-employment is 9% or more.[20]

To test the model's predictions regarding unconditional beliefs, we asked students about the distribution of careers of US college graduates unconditional on major in the first of the 2021 OSU surveys. This question read "What is your best guess about the percent of 35 year-old Americans (note, not just from Exploration or OSU) who have have graduated from a 4-year college that are..." and then listed the nine career groups, "working in any other job," and "not working for pay (e.g., unemployed or a full-time parent)." These questions occurred before those that conditioned on major in order to avoid confusion.

To test Predictions 3 and 4, we estimate equation 14 by OLS.

$$\overline{BeliefError}_c = \alpha + \beta_1 TrueShare_c + \beta_2 HHI_c + \epsilon_c \tag{14}$$

In equation 14, $\overline{BeliefError}_c$ is the difference between the average population belief about that career and the true fraction of college graduates working in that career. $TrueShare_c$ is then the true share working in that career. Finally, HHI_c is a Herfindahl–Hirschman index measuring how concentrated a career is within particular majors: $HHI_c = \sum p^2_{m|c}$.

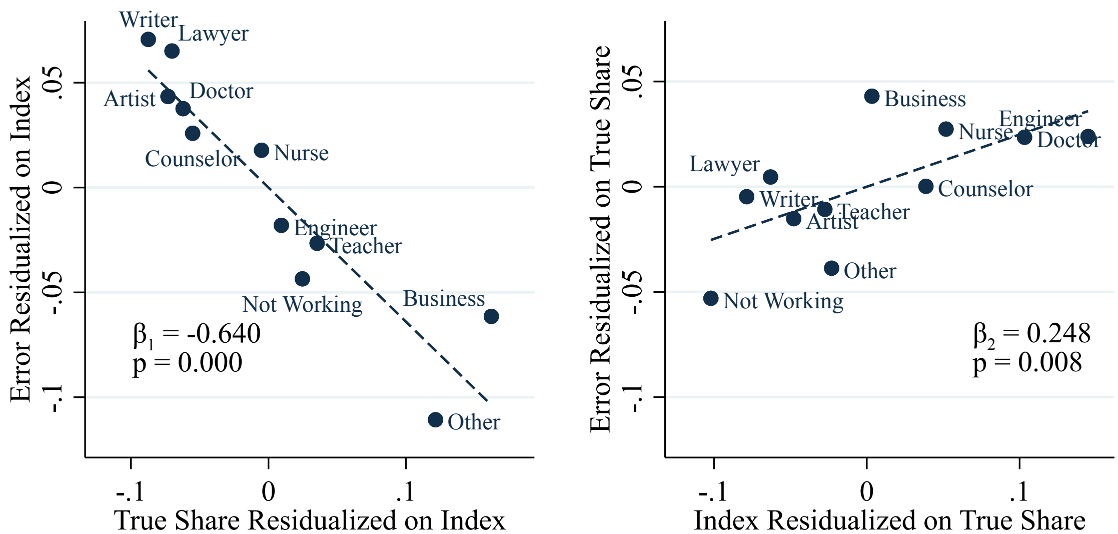
Figure 4 plots the partial correlations between $\overline{BeliefError}_c$ and $TrueShare_c$ as well as between $\overline{BeliefError}_c$ and HHI_c. We see clear evidence in favor of both Predictions 3 and 4: errors in beliefs decrease in the true share of people with each career ($p < 0.01$) and increase in how concentrated the career is within majors ($p < 0.01$).

5.4 Frequency-Based Recall and Role-Model Effects

To further explore the model's primary mechanism, that beliefs depend on who comes to mind, we now analyze a second feature of recall beyond associativeness. Namely, recall

[20] In the first of our 2021 OSU surveys, we asked similar questions about the frequency of careers (plus non-employment) by major. The only differences between these 2021 question and the 2020 questions were 1) they asked students about both themselves and others at age 35, 2) they only asked students about their two top majors (i.e., not also about two other randomly selected majors), and 3) the "not working for pay" category instead read "Not working for pay (e.g., unemployed or a full-time parent)". We made the latter change to explicitly prompt students to consider both involuntary and voluntary non-employment. Table A.XII shows, however, that these changes appear to have had little impact on students' beliefs, and in particular the underestimation of non-employment clearly persists despite these changes.

Figure 4: Beliefs about Careers Unconditional on Majors



Notes: Figure 4 shows partial correlations between errors in unconditional beliefs and the true share of college graduates working in each career (left panel), as well as between errors in unconditional beliefs and a Herfindahl–Hirschman Index (HHI) measuring how concentrated each career is within majors (right panel). Errors in beliefs are the average difference between the 2021 OSU students' unconditional population beliefs about the frequency of each career minus the true fraction of college graduates aged 30-50 working in that career (calculated using the 2017-2019 ACS).

is *frequency-based*, meaning that things come to mind more easily the more experiences that we have with them. To capture this idea, we replace equation 10 above with equation 15.

$$a(e,H) = N(e)S(e,H) \quad \text{where} \quad N(e) = \begin{cases} \frac{\phi}{D} & \text{if } e = x \\ 1 & \text{if } e \neq x \end{cases} \tag{15}$$

The new term $N(e)$ in equation 15 represents the number of experiences that the student has with e. This new force implies that people the student is personally close to—e.g., their parents or other role models, whom they have many experiences with—will have an outsized impact on their beliefs compared to people they have only met or heard of a few times. 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.

Prediction 5: *Personally knowing someone with career a and major A increases beliefs about career a's unconditional frequency and its frequency conditional on major A.*

More precisely, let $\Big(c(x), M(x)\Big) = (a, A)$*. Then* $\frac{\partial}{\partial \phi}\pi_a > 0$ *and* $\frac{\partial}{\partial \phi}\pi_{a|A} > 0$.

The intuition behind Prediction 5 is straightforward: the student's role model is very likely to come to mind, adding evidence in favor of the career path they followed.

For conditional beliefs, we can ask a further question about the effect of role models: how the student's beliefs about one major are affected by knowing someone with a *different* major. Prediction 6 summarizes such effects.

Prediction 6: *Personally knowing someone with career a but major B should reduce the student's beliefs about the likelihood of career a conditional on major A:*

More precisely, let $\Big(c(x), M(x)\Big) = (a, B)$*. Then* $\frac{\partial}{\partial \phi}\pi_{a|A} < 0$ *whenever* $\delta_c < 1$.

The intuition behind Prediction 6 is the same as for stereotyping and follows from the same associative-recall mechanism. The student's role model x comes to mind easily when they try to think about people with career a and major A, because they are similar to (albeit inconsistent with) that hypothesis. This distracts from recalling people who have both career a and major A, reducing the student's beliefs about that hypothesis.

5.5 Empirical Tests of Role-Model Effects

To test Predictions 5 and 6, 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.[21] 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 3 shows OLS estimates of equations 16 and 17.

$$\pi^i_{c|M} = \alpha + \beta_1 RM^i_{c,M} + \beta_2 RM^i_{c,-M} + \mu_{c,M} + \epsilon^i_{c,M} \tag{16}$$

$$\pi^i_c = \alpha + \beta RM^i_c + \mu_c + \epsilon^i_c \tag{17}$$

In equation 16, $\pi^i_{c|M}$ 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. In equation 17, 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.

[21] 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.

Table 3: Role Models and What Comes to Mind

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

Notes: Table 3 presents OLS estimates of equation 16 (columns 1-3 and 5-7) and equation 17 (columns 4, 8, and 9). The dependent variable in columns 1 to 3 are the population beliefs of students in the 2021 OSU data of the fraction of college graduates working in each occupation conditional on each major. The dependent variable in columns 5 to 7 are the corresponding self beliefs: i.e., students' beliefs about their own chance of working in each career if they gradauted with each major. Columns 2 and 6 restrict the sample to career-major pairs in which the career is that major's most stereotypical career (S). Columns 3 and 7 restrict the sample to all career-major pairs where the career is not the most stereotypical (NS) 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.

Columns 1 and 5 of Table 3 show that, pooling across all majors and careers, knowing someone with a particular career-major pair (c, M) boosts beliefs about the frequency of c conditional on M by 3.1 p.p. ($p < 0.01$) for population beliefs and by 3.7 p.p. ($p < 0.01$) for self beliefs. Column 4 shows that having a role model with a particular career boosts students' population beliefs about the unconditional frequency of that career among college graduates by 1.8 percentage points ($p < 0.01$). Column 8 shows similar, and indeed larger, effects on their self beliefs about their own future career. Column 9 shows analogous estimates from the Freshman Survey (which asks the career, but not the major, of students' parents): students are 4.2 p.p. ($p < 0.01$) more likely to list an occupation as their probable career if it is one of their parents' careers.

By themselves, the results on self beliefs could be consistent with many mechanisms (e.g., students may intrinsically prefer or have greater access to their role model's 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 stereotypical career (column 2). Because students already exaggerate stereotypes, these role models therefore make students' beliefs *less* accurate, which we would not naturally expect if greater access to people with those careers only entailed better information about them (i.e., how unlikely such careers often are).

Prediction 6 said that we should expect students' beliefs about $P(a|A)$ to be lower if they have a role model with career a but the other major B. Column 1 of Table 3 shows, instead, a small but *positive* and statistically significant effect; 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-stereotypical outcomes; restricting the data to stereotypical careers, we see a larger and negative effect of -4.4 percentage points ($p < 0.01$). Thus Prediction 6 only appears true for stereotypical outcomes.

What might explain these ambiguous effects? In Appendix C, we show that a simple extension of the model adding a role for extrapolation (Kahneman & Tversky, 1981; Gilboa & Schmeidler, 1995; Bordalo, Burro, et al., 2023) can predict positive effects for sufficiently rare or implausible outcomes. The intuition behind this result is that when a career is very implausible (e.g., becoming an artist after majoring in business), the student may struggle to think of anyone with the right career and major. Therefore, their role model cannot distract much from such people. Instead, the student may extrapolate from the fact that their role model has a similar career path (e.g., an artist who was an art major) to conclude that the hypothesized career path is slightly less implausible.

6 Conclusion

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 stereotypical job. We estimate that students have strong preferences over which job they end up in, suggesting that stereotyping could have large welfare implications. In a field experiment testing a light-touch intervention, we find that information changing these beliefs can have significant effects on students' intentions and later choices. Finally, we use a simple model of belief formation to show that stereotyping can arise as a natural consequence of associative recall. The model makes additional predictions—which new survey evidence broadly confirms—both about average beliefs and how heterogeneity in beliefs should systematically depend on the careers and majors of people students know personally.

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).[22] *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,

[22]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 stereotypical jobs. These human capital investments carry substantial monetary and opportunity costs, and therefore it may be beneficial to find ways to help students make better informed decisions or to nudge them toward less risky academic paths.

References

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

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

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

Arcidiacono, P. (2004). Ability sorting and the returns to college major. *Journal of Econometrics*, *121*, 343-375. Retrieved from www.elsevier.com/locate/econbase doi: 10.1016/j.jeconom.2003.10.010

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. doi: 10.1016/j.jpubeco.2017.10.007

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

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

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

Benjamin, D. J., Bodoh-Creed, A., & Rabin, M. (2019). *Base-rate neglect: Foundations and implications **.

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. (2019). Inaccurate Statistical Discrimination: An Identification Problem. *Ssrn*.

Boneva, T., & Rauh, C. (2019). Socio-economic gaps in university enrollment: The role of perceived pecuniary and non-pecuniary returns.

Bordalo, P., Burro, G., Coffman, K., Gennaioli, N., & Shleifer, A. (2023). Imagining the future: Memory, simulation, and beliefs about covid. *Working paper.*.

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). Beliefs about gender. *American Economic Review*, *109*(3), 739–773.

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. doi: 10.1093/qje/qjac031

Bordalo, P., Gennaioli, N., & Shleifer, A. (2020). Memory, attention, and choice. *Quarterly Journal of Economics*.

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

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

Charles, C. (2022a). *Memory and trading.*

Charles, C. (2022b). *Memory moves markets.*

Chater, N., & Loewenstein, G. (2016, 6). The under-appreciated drive for sense-making. *Journal of Economic Behavior and Organization*, *126*, 137-154. doi: 10.1016/j.jebo.2015.10.016

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*.

Delavande, A., Bono, E. D., & Holford, A. (2020). Academic and non-academic investments at university : The role of expectations , preferences and constraints Angus Holford. *Journal of Econometrics*, 1–68.

Delavande, A., & Zafar, B. (2019). University choice: The role of expected earnings, non-pecuniary outcomes, and financial constraints. *Journal of Political Economy*, *127*(5), 2343–2393.

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

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

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

Evers, E. R. K., Imas, A., & Kang, C. (2021). On the role of similarity in mental accounting and hedonic editing. *Psychological Review*.

Exley, C. L., Hauser, O. P., Moore, M., & Pezzuto, J.-H. (2022). *Beliefs about gender differences in social preferences.*

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.

Fudenberg, D., Lanzani, G., & Strack, P. (2022). *Selective memory equilibrium.* Retrieved from https://ssrn.com/abstract=4015313

Gilboa, I., & Schmeidler, D. (1995). Case-based decision theory. *The Quarterly Journal of Economics*. Retrieved from https://academic.oup.com/qje/article/110/3/605/1859208

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

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

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).

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

Kahana, M. (2012). *Foundations of human memory.* OUP, USA.

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

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

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

Koszegi, B., Loewenstein, G., & Murooka, T. (2021). Fragile self-esteem. *Working paper.*.

Loewenstein, G., O'Donoghue, T., & Rabin, M. (2003). Projection bias in predicting future utility. *Quarterly Journal of Economics*, *118*, 1209-1248. doi: 10.1162/003355303322552784

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.

Mack, M. L., & Palmeri, T. J. (2020). *Discrimination, recognition, and classification* (Vol. 1).

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

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.

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

Sial, A. Y., Sydnor, J. R., & Taubinsky, D. (2023). Biased memory and perceptions of self-control. *NBER Working Paper 30825*. Retrieved from http://www.nber.org/papers/w30825

Tenenbaum, J., & Griffiths, T. (2001). The rational basis of representativeness. *... of the 23Rd Annual Conference of the ...*.

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*.

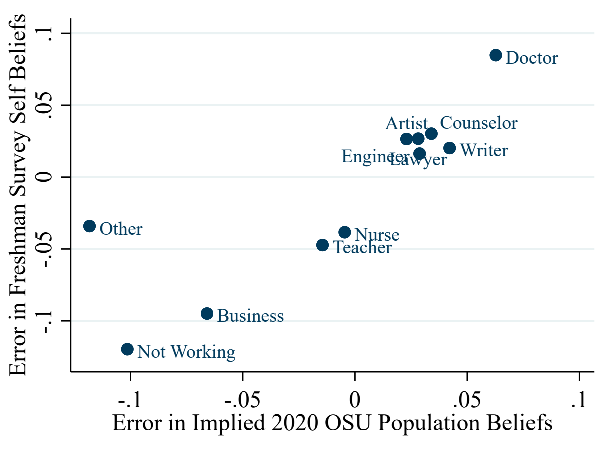
Wachter, J. A., & Kahana, M. J. (2019). A retrieved-context theory of financial decisions. *NBER Working Ppaer 26200*.

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

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

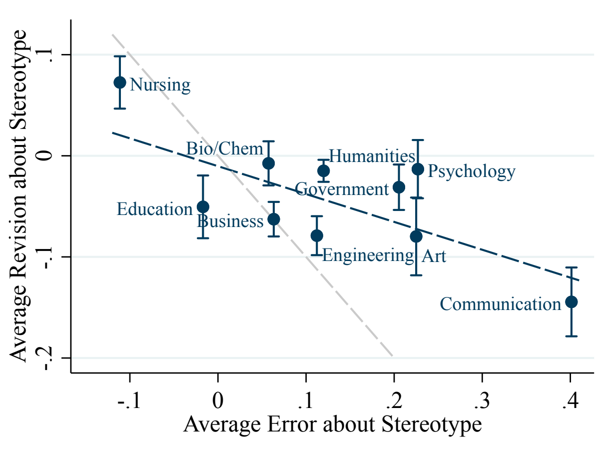
A Online Appendix: Supplementary Figures and Tables

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



Notes: The y-axis in Figure A.I is the difference between the fraction of students in the Freshman Survey who list each occupation as their probable career and the fraction of 33-37 year old college graduates in the CPS (of the same cohorts, up to birth year 1987) who are working in each occupation. The x-axis is the difference between the 2020 OSU students' "implied" beliefs about the frequency of each career and the true frequency. To construct these implied beliefs, we first take the average population belief of the fraction working in each occupation conditional on each major. We then take a weighted average of these values, where the weights are the fraction of students in the Freshman Survey who expect to pursue each major. See Section 3.2 for further details on the construction of this statistic.

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



Notes: The x-axis of Figure A.II is the true share of college graduates working in the stereotypical 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 stereotypical career conditional on graduating with each major, from before to after the information intervention. The solid line shows an OLS line of best fit, while the dashed 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: 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.IV presents the groupings of careers we use to aggregate the occupation titles in the American Community Survey.

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

Table A.VII: 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.VII 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 stereotypicalness: $p_{c,m}/p_{c,-m}$. The most stereotypical career for each major by this metric is bolded.

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

Variable	Shapley Value	
	2020	2021
$\mathbb{1}$(Most Stereotypical)	35.1 %	34.2 %
Career FEs	34.1 %	8.8 %
$\mathbb{1}$(Most Stereotypical)*$\mathbb{1}$(Self Beliefs)	10.2 %	12.3 %
$\mathbb{1}$(Most Stereotypical)*$\mathbb{1}$(Top Major)	7.5 %	
$\mathbb{1}$(Most Stereotypical)*$\mathbb{1}$(Top Major)*$\mathbb{1}$(Self Beliefs)	4.2 %	
Truth	10.8 %	28.5 %
Truth*$\mathbb{1}$(Self Beliefs)	3.5 %	10.0 %
Truth*$\mathbb{1}$(Top Major)	1.8 %	
Truth*$\mathbb{1}$(Top Major)*$\mathbb{1}$(Self Beliefs)	1.2 %	
Role Model Variables		4.1 %
Role Model Variables*$\mathbb{1}$(Self Beliefs)		2.2 %

Notes: Table A.VIII presents a Shorrocks-Shapley decomposition of the R^2 of an OLS regression. Let $Y_{i,c,m,p}$ denote the belief of individual i about the probability of entering career c conditional on major m from perspective p, where p is either that student's own belief (self) or belief about others (population). Let $T_{c,m}$ denote the true probability from the American Community Survey of someone entering career c conditional on majoring in m. We estimate equations 18 and 19 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. $Ster_{m,c}$ indicates whether c is the most stereotypical 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 18 and only the 2021 OSU sample for equation 19.

$$
\begin{aligned}
Y_{i,c,m,p} - T_{c,m} = \psi_{c,p,Top(i,m)} &+ \beta_1 Self_{i,p} + \beta_2 t_{i,m} + \beta_3 Self_{i,p} \times Top_{i,m} + \beta_4 Ster_{m,c} + \beta_5 Ster_{m,c} \times Self_{i,p} + \\
&\beta_6 Ster_{m,c} \times Top_{i,m} + \beta_7 Ster_{m,c} \times Self_{i,p} \times Top_{i,m} + \beta_8 T_{c,m} + \beta_9 T_{c,m} \times Self_{i,p} + \beta_{10} T_{c,m} \times Top_{i,m} \\
&+ \beta_{11} T_{c,m} \times Self_{i,p} \times Top_{i,m} + \varepsilon_{i,c,m,p} \qquad (18)
\end{aligned}
$$

$$
\begin{aligned}
Y_{i,c,m,p} - T_{c,m} = \psi_{c,p} &+ \beta_1 Self_{i,p} + \beta_2 t_{i,m} + \beta_4 Ster_{m,c} + \beta_5 Ster_{m,c} \times Self_{i,p} + \beta_8 T_{c,m} + \beta_9 T_{c,m} \times Self_{i,p} + \\
&\beta_{1,2} RM_{i,c,m} + \beta_{1,3} RM_{i,c,m} \times Self_{i,p} + \varepsilon_{i,c,m,p} \qquad (19)
\end{aligned}
$$

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.IX: Choice Model Estimates

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

Notes: Table A.IX shows parameter estimates of the baseline model described in Section 4.1. The model is estimated by maximum likelihood, and standard errors/confidence intervals are constructed by Bayesian bootstrap. *, **, and *** indicate significance at the 10%, 5%, and 1% levels. We omit estimates of μ_M and $c^*(i)$ for readability. Each column shows estimates of a variant of the model, as described below.

Column 1. Baseline model described in Section 4.1.

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

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

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

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

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

Table A.X: Effects of an Information Intervention on Major Intentions and Later Choices

	Any S23 Major	Not Missing	Belief P(M)		Course Credits		Major in S23	
			1st	2nd	1st	2nd	1st	2nd
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	-0.09***	0.02	-0.03***	-0.00	-0.91	0.15	-0.06*	-0.00
	(0.03)	(0.02)	(0.01)	(0.01)	(0.98)	(0.68)	(0.03)	(0.02)
Pre-Treatment Belief P(M)			0.96***	0.83***	23.08***	16.29***	0.67***	0.38***
			(0.02)	(0.03)	(1.98)	(2.70)	(0.06)	(0.08)
Pre-Treatment Course Credits			0.00	0.00	1.08***	0.89***	0.01**	0.01**
			(0.00)	(0.00)	(0.17)	(0.16)	(0.01)	(0.00)
Constant	0.69***	0.88***	0.02	0.03***	0.88	2.51***	-0.05	-0.00
	(0.02)	(0.02)	(0.01)	(0.01)	(1.34)	(0.70)	(0.05)	(0.02)
Observations	783	783	783	783	783	783	783	783
p-value: 1st vs 2nd Effect Equal				0.020		0.370		0.167

Notes: Table A.X presents OLS regressions including data from the 2021 OSU sample. The dependent variable in column 1 is an indicator for whether students have declared a major by Spring 2023. The dependent variable in column 2 is an indicator for whether they enrolled in any classes during Fall 2023. The dependent variable in columns 3 and 4 is students' post-intervention belief about the percent chance that they will graduate with their first-ranked major (column 3) or second-ranked major (column 4). The dependent variable in columns 5 and 6 is the number of course credit hours that students enrolled in in first-ranked major (column 5) or second-ranked major (column 6) post-intervention (between Spring 2022 and Fall 2023). The dependent variable in columns 7 and 8 is an indicator for whether students' declared major in Spring 2023 is in their first-ranked major group (column 7) or second-ranked major group (column 8). "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 Course Credits" is the number of course credit hours that participants enrolled in in their first-ranked major (odd columns) or second-ranked major (even columns) during Fall 2021. 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 Later Choices

	Belief P(M)		Course Credits		Major in S23	
	1st	2nd	1st	2nd	1st	2nd
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment x Average Reduction in $\pi_{\text{Stereotype}\|M1}$	-0.29*	0.09	-25.90	22.59*	-0.58	-0.07
	(0.17)	(0.14)	(17.04)	(11.94)	(0.58)	(0.34)
Treatment x Average Reduction in $\pi_{\text{Stereotype}\|M2}$	-0.44**	0.22	-13.55	36.88**	-0.41	1.41***
	(0.20)	(0.15)	(19.95)	(16.50)	(0.68)	(0.38)
Treatment	-0.00	-0.01	0.69	-2.27**	-0.01	-0.06**
	(0.01)	(0.01)	(1.33)	(1.05)	(0.05)	(0.03)
Average Reduction in $\pi_{S1\|M1}$	-0.02	0.04	42.91***	-20.39**	1.42***	0.09
	(0.11)	(0.11)	(11.63)	(9.50)	(0.41)	(0.28)
Average Reduction in $\pi_{S2\|M2}$	0.15	-0.06	20.26	-14.29	0.69	-0.27
	(0.11)	(0.10)	(13.65)	(10.13)	(0.46)	(0.24)
Pre-Treatment Belief P($M1$)	0.98***	-0.09**	21.17***	-0.76	0.68***	-0.03
	(0.02)	(0.04)	(2.07)	(1.61)	(0.07)	(0.05)
Pre-Treatment Belief P($M2$)	0.04	0.73***	-3.81	15.72***	0.09	0.36***
	(0.04)	(0.06)	(3.69)	(3.33)	(0.12)	(0.10)
Pre-Treatment Course Credits in $M1$	0.00	-0.00*	0.97***	0.17	0.01	0.00
	(0.00)	(0.00)	(0.18)	(0.11)	(0.01)	(0.00)
Pre-Treatment Course Credits in $M2$	-0.00	0.00	0.14	0.92***	0.00	0.01**
	(0.00)	(0.00)	(0.20)	(0.16)	(0.01)	(0.00)
Constant	-0.01	0.11***	0.32	4.08**	-0.16**	0.03
	(0.02)	(0.04)	(1.91)	(1.76)	(0.06)	(0.05)
N	783	783	783	783	783	783
p-value: M1 vs M2 Interactions Equal	0.583	0.556	0.655	0.493	0.865	0.005
p-value: 1st vs 2nd $M1$ Interactions Equal		0.089		0.020		0.454
p-value: 1st vs 2nd $M2$ Interactions Equal		0.010		0.052		0.020

Notes: Table A.XI presents OLS regressions including data from the 2021 OSU sample. The dependent variable in columns 1 and 2 is students' post-intervention belief about the percent chance that they will graduate with their first-ranked major (column 1) or second-ranked major (column 2). The dependent variable in columns 3 and 4 is the number of course credit hours that students enrolled in in first-ranked major (column 3) or second-ranked major (column 4) post-intervention (between Spring 2022 and Fall 2023). The dependent variable in columns 5 and 6 is an indicator for whether students' declared major in Spring 2023 is in their first-ranked major group (column 5) or second-ranked major group (column 6). "Treatment" is an indicator for treatment status. "Average reduction in $\pi_{c|M1}$" (or $\pi_{c|M2}$) is our measure of treatment intensity: the leave-out mean revision in self-beliefs about the likelihood of working in the stereotypical career for students first-ranked major ($M1$) or second-ranked major ($M2$). "Pre-Treatment Belief P($M1$)" (or $M2$) is students' pre-treatment belief about their likelihood of graduating with a major in their first-ranked major group ($M1$) or second-ranked major group ($M2$). "Pre-Treatment Course Credits in $M1$" (or $M2$) is the number of course credit hours that participants enrolled in in their first-ranked major ($M1$) or second-ranked major ($M2$) during Fall 2021. Robust standard errors in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels.

Table A.XII: 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.XII 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 stereotypical career for each major (where we define stereotypicalness by $p_{c,m}/p_{c,-m}$) is bolded.

B Online Appendix: Data and Additional Analyses

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

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

B.1 Fall 2020 Ohio State Survey

We embedded the 2020 OSU survey into the Fall semester course associated with the Exploration program. Students accessed the survey through the official course website. They took the survey between October and December and earned extra credit in their Exploration course for doing so. The median student took 27 minutes to complete the survey. Our main study sample includes 755 completed responses, amounting to a roughly 80% response rate.[23]

These surveys focused on the 10 major groups described in Section 2. Whenever the surveys mentioned a group of majors, the name of the group appeared in blue font to indicate that students could click it to see which particular majors were included in the group.[24] The surveys also focused on the nine career groups mentioned in Section 2. 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 stereotypical 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".[25] However, in open-ended feedback the overwhelming reason given was that they

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

[24]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.

[25]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 diffi-

took longer to fill out than multiple choice questions would have.[26] In addition, all of the main results described in Section 3.2 are nearly identical for students who did and did not report finding these questions difficult to answer.

B.1.1 Eliciting Salary Beliefs

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

These questions give us a measure of students' beliefs about the distribution of salaries conditional on majors or careers. We then calculate expected values from these distributions to ease interpretation and compare them to the ACS data. To do so, we assume that salaries are uniformly distributed within the ranges that the survey asked about. We apply a similar assumption to the actual distribution of salaries using ACS data. Namely, we first calculate the share of people with salaries in the ranges listed in the OSU survey. We then calculate the average salary assuming that salaries are uniformly distributed within these ranges.[27]

B.2 Fall 2021 Ohio State Surveys

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

cult 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.

[26]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.

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

B.3 CloudResearch Survey

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

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

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

B.4 2013 National Survey of College Graduates

In Section 4.1, we mentioned a regression involving data from the 2013 National Survey of College Graduates. Here we give more details about those data and that regression. The dependent variable we use is an indicator variable for whether they express that they are very or somewhat dissatisfied with their primary job. The independent variables are a dummy variable indicating whether the respondent said their principal job was not related to their highest degree and their salary (in \$10,000s). We restrict the data to college graduates between the ages of 30 and 50. We additionally includes fixed effects for college major, age, race, and gender. This regression yields a coefficeint of 0.117 ($p < 0.01$) for the unrelated job dummy and a coefficient of -0.004 ($p < 0.01$) for the salary variable (full regression results available upon request).

C Online Appendix: Proofs

Here we provide derivations of the results in Section 5. In the main text, we first analyze the effects of associative recall without frequency effects and then later introduce frequency effects. Here we start with the full model (i.e, with equation 15), setting $\phi = 0$ (the no role-models parameterization) when appropriate. Note also that we take the sum in equation 10 to be over *experiences*, not individuals (i.e., if the student has multiple experiences with her role model, they get extra weight in the similarity function $S(e, H)$. This latter assumption does not affect any of the results.

Following the discussion in Section 5.5, we first generalize the model in the following way. Each time the student samples someone e, we no longer assume that she simply checks whether e is consistent with the hypothesis H that she is considering. Instead, we now assume she attempts to "simulate" the hypothesis (i.e., imagine someone having the career/major she is assessing) using e. 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. Note that if $\sigma(e, H) = \mathbb{1}(e \in H)$, this is equivalent to the model described in the main text.

Following Kahneman & Tversky (1981) and Bordalo, Burro, et al. (2023), we assume ease of simulation depends on similarity: more specifically, we assume $\sigma(e, H)$ is weakly increasing in the similarity between e and H. We assume the functional form in equation 20, 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.

$$\sigma(e, H_c) = \eta_c^{\mathbb{1}(c(e) \neq c)} \qquad \sigma(e, H_{c,M}) = \eta_c^{\mathbb{1}(c(e) \neq c)} \times \eta_M^{\mathbb{1}(M(e) \neq M)} \tag{20}$$

Note that the model in the main text corresponds to the case where $\eta_c = \eta_M = 0$.

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:

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

In Section 5.1, we discussed a modification to the model whereby the student may not know with certainty what e's career and major are. We claimed that if her beliefs about e

were unbiased, then this modification is equivalent to the model in the main text. We now make this claim precise. Suppose $\{G_1, G_2, ...,\}$ are mutually exclusive groups of people that partition the student's database D. Let $p_{c,M,g} = P(e \in H_{c,M} | e \in G_g)$ be the known fraction of people in G_g who have c and M. We can then assume that the student simulates H using e's *group*, along with her (correct) beliefs about the likelihood that someone in that group has each major and career.

Then,

$$\begin{aligned}
\widetilde{\sigma}(e_t, H) &\equiv \sum_c \sum_M p_{c,M,g(e_t)} \sigma(c, M, H) \\
\implies E[\widetilde{\sigma}(e_t, H)] &= \sum_e P(e_t = e) \sum_c \sum_M p_{c,M,g(e)} \sigma(c, M, H) \\
&= \sum_g P(e_t \in G_g) \sum_c \sum_M p_{c,M,g} \sigma(c, M, H) \\
&= \sum_g P(e_t \in G_g) \sum_c \sum_M P(e_t \in H_{c,M} | e_t \in G_g) \sigma(c, M, H) \\
&= \sum_g P(e_t \in G_g) \sum_c \sum_M P(e_t \in H_{c,M} | e_t \in G_g) E[\sigma(e_t, H) | e_t \in H_{c,M} \cap G_g] \\
&= \sum_g P(e_t \in G_g) E[\sigma(e_t, H) | e_t \in G_g] \\
&= E[\sigma(e_t, H)]
\end{aligned}$$

Thus, because plausibility in the two models converge in probability to $E[\sigma(e_t, H)]$ and $E[\widetilde{\sigma}(e_t, H)]$, and these are equal to each other, these two models are equivalent (in the limit where $T \to \infty$).

Next, we derive the 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 yield Prediction 2, let p_B be the fraction of people with major B. We can consider how the agent's beliefs 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 as in the main text. 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 $\eta_c = \eta_M = 0$, as in main specification, this derivative is unambiguously negative whenever $\delta_c < 1$.

Next, to derive Prediction 4, let $p_{z,B} = 0$ for all careers z and set $\phi = 0$. Then, (with some algebra) it is simple to show that $\frac{\pi_{a|A}}{\pi_{b|A}} > \frac{p_{a,A}}{p_{b,A}}$ whenever $p_{b,A} > p_{a,A}$, meaning that the agent relatively overestimates rarer careers (conditional on major).

To derive Prediction 5 for conditional beliefs, let $\Big(c(x), m(x)\Big) = (a, A)$. Then

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

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

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

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. Note that this is satisfied in the main specification where $\eta_M = \eta_c = 0$.

To yield Prediction 6, let $\Big(c(x), m(x)\Big) = (a, B)$. Then

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

Then,

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

Note that when $\eta_c = \eta_M = 0$, this derivative is unambiguously negative. 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 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.

Next, we turn to results concerning beliefs about careers unconditional on major. To do so, we first need to derive the average similarity $S(e_{c,m}, H_{c'})$ of experiences with career-major (c, m) to each hypothesis $H_{c'}$. Let $\Big(c(x), m(x)\Big) = (a, A)$. Then,

$$S(e_{a,A}, H_a) = \frac{p_{a,A} + \delta_M p_{a,B} + \phi}{p_{a,A} + p_{a,B} + \phi} \text{ and } S(e_{z,A}, H_a) = \frac{\delta_c p_{a,A} + \delta_c \delta_M p_{a,B} + \delta_c \phi}{p_{a,A} + p_{a,B} + \phi} \text{ for } z \neq a$$
$$S(e_{a,B}, H_a) = \frac{\delta_M p_{a,A} + p_{a,B} + \delta_M \phi}{p_{a,A} + p_{a,B} + \phi} \text{ and } S(e_{z,B}, H_a) = \frac{\delta_c \delta_M p_{a,A} + \delta_c p_{a,B} + \delta_c \delta_M \phi}{p_{a,A} + p_{a,B} + \phi} \text{ for } z \neq a$$
$$S(e_{b,A}, H_b) = \frac{p_{b,A} + \delta_M p_{b,B}}{p_{b,A} + p_{b,B}} \text{ and } S(e_{z,A}, H_B) = \frac{\delta_c p_{b,A} + \delta_c \delta_M p_{b,B}}{p_{b,A} + p_{b,B}} \text{ for } z \neq a$$
$$S(e_{b,B}, H_b) = \frac{\delta_M p_{b,A} + p_{b,B}}{p_{b,A} + p_{b,B}} \text{ and } S(e_{z,B}, H_b) = \frac{\delta_c \delta_M p_{b,A} + \delta_c p_{b,B}}{p_{b,A} + p_{b,B}} \text{ for } z \neq a$$

Then the plausibility of each hypothesis can be written as follows:

$$F(H_a) = 1/\Bigg((p_{a,A}+\phi)(p_{a,A}+\delta_M p_{a,B}+\phi) + p_{b,A}(\delta_c p_{a,A}+\delta_c\delta_M p_{a,B}+\delta_c\phi) + \sum_{z\notin\{a,b\}} p_{z,A}(\delta_c p_{a,A}+\delta_c\delta_M p_{a,B}+\delta_c\phi)$$

$$+ p_{a,B}(\delta_M p_{a,A}+p_{a,B}+\delta_M\phi) + p_{b,B}(\delta_c\delta_M p_{a,A}+\delta_c p_{a,B}+\delta_c\delta_M\phi) + \sum_{z\notin\{a,b\}} p_{z,B}(\delta_c\delta_M p_{a,A}+\delta_c p_{a,B}+\delta_c\delta_M\phi)\Bigg)$$

$$\times\Bigg((p_{a,A}+\phi)(p_{a,A}+\delta_M p_{a,B}+\phi) + \eta_c p_{b,A}(\delta_c p_{a,A}+\delta_c\delta_M p_{a,B}+\delta_c\phi) + \eta_c \sum_{z\notin\{a,b\}} p_{z,A}(\delta_c p_{a,A}+\delta_c\delta_M p_{a,B}+\delta_c\phi)$$

$$+ p_{a,B}(\delta_M p_{a,A}+p_{a,B}+\delta_M\phi) + \eta_c p_{b,B}(\delta_c\delta_M p_{a,A}+\delta_c p_{a,B}+\delta_c\delta_M\phi) + \eta_c \sum_{z\notin\{a,b\}} p_{z,B}(\delta_c\delta_M p_{a,A}+\delta_c p_{a,B}+\delta_c\delta_M\phi)\Bigg)$$

$$F(H_b) = 1/\Bigg((p_{a,A}+\phi)(\delta_c p_{b,A}+\delta_c\delta_M p_{b,B}) + p_{b,A}(p_{b,A}+\delta_M p_{b,B}) + \sum_{z\notin\{a,b\}} p_{z,A}(\delta_c p_{b,A}+\delta_c\delta_M p_{b,B})$$

$$+ p_{a,B}(\delta_c\delta_M p_{b,A}+\delta_c p_{b,B}) + p_{b,B}(\delta_c p_{b,A}+p_{b,B}) + \sum_{z\neq b} p_{z,B}(\delta_c\delta_M p_{b,A}+\delta_c p_{b,B})\Bigg)$$

$$\times\Bigg(\eta_c(p_{a,A}+\phi)(\delta_c p_{b,A}+\delta_c\delta_M p_{b,B}) + p_{b,A}(p_{b,A}+\delta_M p_{b,B}) + \eta_c \sum_{z\notin\{a,b\}} p_{z,A}(\delta_c p_{b,A}+\delta_c\delta_M p_{b,B})$$

$$+ \eta_c p_{a,B}(\delta_c\delta_M p_{b,A}+\delta_c p_{b,B}) + p_{b,B}(\delta_c p_{b,A}+p_{b,B}) + \eta_c \sum_{z\neq b} p_{z,B}(\delta_c\delta_M p_{b,A}+\delta_c p_{b,B})\Bigg)$$

Some algebra then shows that $\frac{\partial}{\partial\phi}\log F(H_a) > 0$ and $\frac{\partial}{\partial\phi}\log F(H_b) < 0$. Analogous results for the case where $m(x) = B$ are also simple to derive. Thus, no matter what the role model's major is, beliefs about a increase in ϕ.

We next derive Prediction 3. Let $p_{a,A} = p_a - p_{a,B}$ and let $\phi = 0$. Then

$$\frac{\partial}{\partial p_{a,B}}\log F(H_a) =$$

$$= \frac{2(p_{a,B}-p_{a,A})(1-\delta_M) + \eta_c\delta_c(1-\delta_M)\sum_{z\notin\{a\}}(p_{z,B}-p_{z,A})}{\begin{array}{l} p_{a,A}(p_{a,A}+\delta_M p_{a,B}) + \eta_c p_{b,A}(\delta_c p_{a,A}+\delta_c\delta_M p_{a,B}) + \eta_c\sum_{z\notin\{a,b\}} p_{z,A}(\delta_c p_{a,A}+\delta_c\delta_M p_{a,B}) \\ \quad + p_{a,B}(\delta_M p_{a,A}+p_{a,B}) + \eta_c p_{b,B}(\delta_c\delta_M p_{a,A}+\delta_c p_{a,B}) + \eta_c\sum_{z\notin\{a,b\}} p_{z,B}(\delta_c\delta_M p_{a,A}+\delta_c p_{a,B}) \end{array}}$$

$$- \frac{2(p_{a,B}-p_{a,A})(1-\delta_M) + \delta_c(1-\delta_M)\sum_{z\notin\{a\}}(p_{z,B}-p_{z,A})}{\begin{array}{l} p_{a,A}(p_{a,A}+\delta_M p_{a,B}) + p_{b,A}(\delta_c p_{a,A}+\delta_c\delta_M p_{a,B}) + \sum_{z\notin\{a,b\}} p_{z,A}(\delta_c p_{a,A}+\delta_c\delta_M p_{a,B}) \\ \quad + p_{a,B}(\delta_M p_{a,A}+p_{a,B}) + p_{b,B}(\delta_c\delta_M p_{a,A}+\delta_c p_{a,B}) + \sum_{z\notin\{a,b\}} p_{z,B}(\delta_c\delta_M p_{a,A}+\delta_c p_{a,B}) \end{array}}$$

which, if $p_{B|z} = p_{A|z}$ for other careers z, is positive when $p_{B|a} > p_{A|a}$. However, note that when most *other* careers are mostly in major B that can make this effect go negative. This makes sense, as it would then be making career a more similar to other careers, increasing interference.

Now let's do the same for H_b.

$$\frac{\partial}{\partial p_{a,B}} \log F(H_b) =$$

$$= \frac{\eta_c \delta_c (1-\delta_M)(p_{b,B} - p_{b,A})}{\begin{array}{r}\eta_c(p_{a,A}+\phi)(\delta_c p_{b,A} + \delta_c\delta_M p_{b,B}) + p_{b,A}(p_{b,A}+\delta_M p_{b,B}) + \eta_c \sum_{z\notin\{a,b\}} p_{z,A}(\delta_c p_{b,A} + \delta_c \delta_M p_{b,B}) \\ + \eta_c p_{a,B}(\delta_c\delta_M p_{b,A} + \delta_c p_{b,B}) + p_{b,B}(\delta_c p_{b,A} + p_{b,B}) + \eta_c \sum_{z\neq b} p_{z,B}(\delta_c \delta_M p_{b,A} + \delta_c p_{b,B})\end{array}}$$

$$- \frac{\delta_c (1-\delta_M)(p_{b,B} - p_{b,A})}{\begin{array}{r}\eta_c(p_{a,A}+\phi)(\delta_c p_{b,A} + \delta_c\delta_M p_{b,B}) + p_{b,A}(p_{b,A}+\delta_M p_{b,B}) + \eta_c \sum_{z\notin\{a,b\}} p_{z,A}(\delta_c p_{b,A} + \delta_c \delta_M p_{b,B}) \\ + \eta_c p_{a,B}(\delta_c\delta_M p_{b,A} + \delta_c p_{b,B}) + p_{b,B}(\delta_c p_{b,A} + p_{b,B}) + \eta_c \sum_{z\neq b} p_{z,B}(\delta_c \delta_M p_{b,A} + \delta_c p_{b,B})\end{array}}$$

which has the same sign as $p_{b,A} - p_{b,B}$.

Thus, when $p_{z,A} = p_{z,B}$ for all $z \neq a$, $\frac{\partial}{\partial p_{a,B}} \log \frac{\pi_a}{\pi_b} > 0 \iff p_{B|a} > p_{A|a}$.

To investigate Prediction 4 for unconditional beliefs, let $\phi = 0$ (shutting down role model effects) and let $p_{A|z} = K$ for all z. If so, then (with some algebra) it is simple to show that $\frac{\pi_a}{\pi_b} > \frac{p_a}{p_b}$ when $p_b > p_a$.