What Jobs Come to Mind? Stereotypes about Fields of Study

Please click here for the latest version.

John J. Conlon Dev Patel

October 10, 2022*

Abstract

What do college students know about how their choice of what to study will affect their future job? Using both nationally representative survey data and surveys that we administered among undergraduates at the Ohio State University, we document that U.S. freshmen hold systematically incorrect beliefs about the relationship between college majors and occupations. Students appear to stereotype fields of study, greatly exaggerating the likelihood that majors lead to their most distinctive jobs (e.g., counselor for psychology, journalist for journalism, teacher for education). We estimate a stylized model of college major choice, which suggests that stereotyping boosts demand for "risky" majors: ones with rare stereotypical careers and low-paying alternative jobs. In a field experiment among the same Ohio State sample, providing statistical information on career frequencies to first-year college students has significant effects on their intended fields of study (and, less precisely, on their choices of which classes to enroll in), with larger effects on students considering risky majors. To shed light on why students might stereotype majors, we present a model in which regularities in recall shape how individuals draw on past experiences to form expectations, generating predictions about both average biases in beliefs (including stereotyping) and heterogeneity across people. We confirm—among other predictions—that the careers and majors of people close to students (their parents and other role models) have systematic, and at times counterintuitive, effects on their beliefs.

*Department of Economics, Harvard University. Conlon: johnconlon@g.harvard.edu. Patel: 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 to the Lab for Economic Applications and Policy, the Mind Brain Behavior Initiative, and the National Science Foundation Graduate Research Fellowship under grant DGE1745303 for financial support. All errors are ours.

1 Introduction

A young person deciding on their education must first form beliefs about the consequences of their choices. A growing body of evidence suggests that across many educational contexts—and despite the importance of these choices—such expectations often stray far from the truth.[1] However, the *origin* of these mistaken beliefs is often unclear, and biases can seem unstable across people and contexts. Similar puzzles—pervasive errors in expectations that defy simple characterization—appear across a wide range of economic domains.[2] 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.[3] We focus on students' beliefs about the job they will have depending on their college major. We begin by 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 are. These differences—between 40,000 and 200,000 students a year, depending on the profession—appear largely unchanged since at least the 1970s.

What explains these patterns? We hypothesize that they may be the result of stereotypical thinking: 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, indeed, many more students expect to attain their major's stereotypical career than actually work in that job: 65% of prospective art majors expect to be artists (only 17% are), 63% of biology

[1] For instance, see Wiswall & Zafar (2015b), Hastings et al. (2016), Betts (1996), Jensen (2010), Dominitz & Manski (1996), and Baker et al. (2018).

[2] See, among many other examples: Giglio et al. (2021) in personal finance; Conlon et al. (2018), Mueller et al. (2021), and Jäger et al. (2021) in job search; Weber et al. (2022) in inflation expectations; Alesina et al. (2021) and Alesina et al. (2022) in perceptions of race and immigration; and Benjamin (2018) in lab experiments.

[3] College major choice plays a large and increasing role in shaping the economic prospects of college graduates (Altonji et al., 2012, 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; Blom et al., 2021).

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 have particularly extreme beliefs.

To distinguish between these possibilities and to test for the role of stereotypical thinking, 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 students to precisely express uncertainty), about people other than themselves (shutting down overconfidence), and about majors other than their own (shutting down selection and other confounds). In regression analyses, we additionally control for individual-by-career fixed effects to show that these biases are not about focal careers *per se* but about the relationship *between* majors and careers. These differences are strikingly large and mirror those in the Freshman 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 respondents' career expectations.

We then turn to the implications of stereotypical thinking for students' choice of major and for welfare. We estimate a stylized model of major choice and find that, while student's have a significant preference for higher salaries, non-pecuniary amenities of jobs loom large in their decision of what to study. A simple welfare analysis shows that for students on the margin between majors, two factors then matter for the welfare consequences of biased beliefs: the likelihood of stereotypical outcomes and the salaries of non-stereotypical alternative careers. We show that there is wide variation across majors in the US along these two dimensions. Some fields of study—e.g., fine arts, humanities, communications, psychology—have both rare stereotypical careers and low-paying alternatives. The model predicts that welfare costs of stereotypical thinking are largest for students whose biased beliefs induce them to choose "risky" majors like these. In contrast, welfare costs are lower for fields like STEM, business, education, and nursing, either because these majors' stereotypical careers are objectively quite common or because wages are high even for those who end up in non-stereotypical careers. Consistent with this analysis, we show correlational evidence that graduates with risky majors earn less, are less satisfied with their jobs, have greater student debt, and are more likely to regret their field of study. Together, these results suggest that, for students who are on the margin between majors, there may be benefits to encouraging them to pursue less risky academic paths.

Next, in a field experiment, we test the effect of a light-touch information intervention providing statistics on the joint distribution of majors and careers. We find statistically significant effects on students' intended major, with students on average moving away from the major they initially listed as their most likely field of study. This effect is largest—a shift of about seven percentage points ($p < 0.01$)—for students considering a risky major who exaggerate the likelihood of its stereotypical outcome. We also find some corroborating evidence that the information affected students' choices of classes in subsequent semesters. For example, the information induces students who overestimate the frequency of their major's stereotypical career to take about 0.5 fewer classes in that subject, compared to students who underestimate that career ($p < 0.10$). We find these effects despite the fact that students appear to substantially discount the information we provided about the careers of others when updating their beliefs about their own future career. This latter fact may explain the sometimes modest size of the estimated treatment effects.

Finally, we turn to the origins of these beliefs: why do students stereotype majors? To explore this question, we present a simple cognitive model of belief formation. In the model, 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, well-established principles of in human recall play a key role (Kahana, 2012; Bordalo, Conlon, et al., 2022). First, recall is limited, meaning that memories must compete for retrieval. Thus, factors that make one person easier to recall thereby make others harder to recall. Second, and crucially, recall is associative—that is, based on similarity—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 as an endogenous consequence of associative memory. 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. They thus come to mind easily. In contrast, those with non-stereotypical jobs are difficult to retrieve 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.

While other mechanisms could contribute to our headline stereotyping results, associative memory makes further predictions, which we then test, for how beliefs should depart from the truth.[4] For instance, which outcomes do students underestimate when they are exaggerating stereotypical careers? One plausible *ex ante* possibility was that students would neglect all other outcomes, but the model makes a different prediction. Associative recall implies that when assessing a particular career-major pair, the student is disproportionately likely to think of people with exactly that outcome, because they are maximally similar to the hypothesis under consideration. This force especially boosts beliefs about rare careers, which otherwise would less frequently come to mind. The model thus predicts exaggeration of very rare careers and neglect of more common careers (unless they are stereotypical). This is precisely the pattern we find, with students underestimating the likelihood of common-but-non-stereotypical outcomes, such as working in business (except for business majors), as a teacher (except for education majors), and not working at all. In contrast, they overestimate very rare outcomes, like becoming a doctor after majoring in engineering or a counselor after majoring in communications.

Next, the model makes predictions for beliefs about careers *unconditional on major*, also based on similarity. Like with conditional beliefs, it predicts that students should overweight rare careers. A more novel prediction says that beliefs about a career should increase with the extent to which it is *concentrated* within particular majors. To see this, note that when a career (e.g., doctor) is attained primarily after majoring in a small number of fields, people with that career tend to be similar to each other and so come to mind easily, compared to outcomes (e.g., teaching, non-employment) that include people from many different backgrounds. We find statistically significant evidence for both predictions, with students exaggerating careers unconditional on major more when they are rare and when they are concentrated within major, which we proxy for using a Herfindahl–Hirschman index.

To provide further evidence for the model's primary mechanism—that beliefs depend on how easily examples of people with various majors/careers come to mind—we leverage the fact that, besides being associative, memory is frequency-based: it is easier to recall things we have more experience with. We show that the careers and majors of people students know particularly well—their parents and other role models—systematically affect their beliefs in

[4] Other potential mechanisms that could 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).

directions predicted by the model. Personally knowing someone with a particular career or career-major pair boosts students' beliefs about the likelihood of that outcome, not just for themselves (which could be consistent with many other explanations) but for others as well. For instance, having a parent who became a lawyer after majoring in political science boosts beliefs about the number of lawyers nationwide and about the share of political science majors who become lawyers. Note that for stereotypical careers, which students already tend to overestimate, this means that having a role model with that career and major makes students' beliefs *less* accurate. In some ways, these results echo the outsized role of personal experiences on beliefs and economic choices in other contexts and may shed light on the potential mechanisms behind the impacts of role models.[5]

Our study is related to a rich literature on beliefs and human capital investment.[6] In the domain of college major choice, many papers have shown that subjective-expectations survey data robustly predict later life outcomes and help to explain students' human capital decisions (e.g., Arcidiacono et al. 2012, Arcidiacono et al. 2020, Wiswall & Zafar 2021). These studies tend to focus on beliefs about average salary conditional on major (e.g., Conlon 2021), though recent notable exceptions include Wiswall & Zafar (2018) and Arcidiacono et al. (2020). Our finding of systematic biases in expectations about the distribution of occupations may shed light on the origins of students' salary beliefs. More broadly, stereotypical thinking may lead many graduates to end up in jobs that they did not expect or prepare for while in college, with implications for wages and career dynamics (Robst 2007, Nordin et al. 2010).

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, immigration, and gender, including in education settings.[7] Our conception of stereotyping as a byproduct of associative memory also speaks to the broader literature studying the origins of beliefs and their biases.[8] The underlying mechanics of memory in our model build off literatures in psychology, and recent work in economics increasingly incorporates many

[5] See, for instance, Georganas et al. (2014), Kuchler & Zafar (2019), Malmendier (2021), and Conlon et al. (2022) for evidence on overweighting personal experiences and Bleemer (2016), Chung et al. (2018), Bell et al. (2019), Porter & Serra (2020), Altmejd et al. (2021), and Riise et al. (2022) on role models.

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

[7] See Agan & Starr (2017), Arnold et al. (2018), Alesina et al. (2021), Alesina et al. (2022), Bordalo et al. (2019), K. Coffman et al. (2020), Bohren et al. (2019), Exley et al. (2022), Coate & Loury (1993), Shih et al. (1999), and Carlana (2019).

[8] See, among many other examples, Rabin (2002), Bénabou & Tirole (2016), Gagnon-Bartsch et al. (2017), Benjamin et al. (2017), Enke & Zimmermann (2019), Enke (2020), Golman et al. (2021), Augenblick & Rabin (2021), and Gagnon-Bartsch & Bushong (2022).

of these assumptions.[9]

2 Motivating Evidence: Nationally Representative Expectations

To investigate beliefs among a nationally representative sample of first-year college students, 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 to 2015.[10] 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 1 shows self-reported demographic information about students in the Freshman Survey. Throughout the analysis, we use census data to weight the 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.[11] 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."[12] We group these fields into 10 major groups (plus "other" and "undecided"), as shown in Table A.I. 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.II. 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

[9]See Kahneman & Tversky (1981), Dougherty et al. (1997), Hassabis et al. (2007), Kahana (2012), Schacter et al. (2012), Biderman et al. (2020), Wachter & Kahana (2019), Enke et al. (2021), and Koszegi et al. (2021).

[10]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.

[11]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.

[12]The exact list of majors varies from year to year. In the 2007 wave of the Freshman Survey, for example, there are 84 options.

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 37, because by this time the vast majority of people are no longer students and have started their career.[13] 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.III).

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. Table A.IV 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%).[14] 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.[15] 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.[16]

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

[14]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.

[15]For this statistic, we take the number of first-year undergraduates per year is taken from the National Center for Education Statistics: `https://nces.ed.gov/programs/digest/d18/tables/dt18_302.10.asp`. Even this number is surely an underestimate for several reasons. First, we restrict the CPS statistic to college graduates only; many students in the Freshman Survey sample, in contrast, will end up dropping out of college (and thus cannot become physicians). This calculation also assumes that only students who say when they are freshmen that they expect to become doctors actually do. In fact, some students who expect to enter different professions in fact become physicians, making the 150,000 number even more of an underestimate.

[16]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.

3 Stereotypical Thinking

We hypothesized that these patterns in the Freshman Survey were driven by stereotypical thinking: students may exaggerate the likelihood that majors lead to their distinctive jobs. Following Bordalo et al. (2016), we can identify stereotypical-ness with *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 stereotypical thinking 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 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 have a career in that major's most stereotypical career than in fact work in that career. 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 Stereotypical Thinking

The patterns described in Sections 2 and 3.1, while consistent with stereotypical thinking, 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 stereotypical thinking 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. Students received extra credit in the course associated with the program for completing our survey.

Column 2 of Table 1 gives self-reported demographic information about the 755 respondents in this sample, which we call our "2020 OSU data." The sample is broadly comparable to the overall student body at OSU, though with a somewhat higher share of first-generation college students.[17] It is also similar to the Freshman survey along gender, ethnicity, first-generation status, and self-reported family income.[18] 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

An obvious 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 wording makes it difficult to interpret the patterns of expectations that 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, the OSU survey asked students quantitative probabilistic questions about a well-defined event, allowing them to express uncertainty precisely. Namely, we asked each student, about their top-ranked major, "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 just despite the difference in elicitation method (qualitative vs quantitative expectations), but also despite differences in time period (1970s-2010s vs 2020) and sample

[17]For further details, see http://enrollmentservices.osu.edu/report.pdf.

[18]We use the CPI-U to convert family income in the Freshman Survey into September 2020 dollars.

(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 about the outcomes of others as well. More precisely, *before* asking students about their own future jobs, the survey asked students 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.

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 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.[19] Throughout the analyses to follow, we employ such weighting whenever we pool beliefs about students' top-ranked majors with beliefs about their lower ranked majors, though in practice these weights have little impact on our main

[19]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).

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.[20] 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.

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 stereotypical thinking. And finally, the role of stereotyping could be confounded if students exaggerate the unconditional likelihood of careers that happen to be stereotypical of majors, perhaps because certain occupations are simply more or less salient. For example, if students simply underestimate non-employment and "other" jobs (which are not stereotypical of any career), this could spuriously be interpreted as stereotyping.

We address these final two issues with a regression analysis. Table 2 shows OLS estimates of the regression specification in equation 1, where $\pi^i_{c|M}$ is student i's population belief about career c conditional on M, $p_{c|M}$ is the true fraction of those with that major who are working in that career, and μ^i_c are career-by-individual fixed effects.

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

The coefficient θ is our measure of stereotypical thinking. 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 fixed effects allows us to separate stereotyping from biases unrelated to major (e.g.,

[20]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%.

overweighting salient careers irrespective of major). For example, if students simply neglected non-employment but otherwise responded only to frequency, this would be captured by μ_c and γ.

Column 1 of Table 2 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 stereotypical thinking. 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. 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).

The remaining columns of Table 2 show that we see similarly large effects for male vs female students, underrepresented-minority vs non-minority students, and first-generation vs non-first-generation students. Columns 8 and 9 show estimates using survey data we collected on Amazon Mechanical Turk, where participants were incentivized to provide accurate responses to the same questions (see Appendix B for details on this survey). We see somewhat smaller but broadly similar estimates of θ for these respondents, both when restricting to college-educated and non-college-educated adults. We conclude that stereotyping appears to be a broad-based phenomenon, not restricted to particular groups or to college freshmen.

3.3 Connecting OSU and Freshman Survey Beliefs

In this section, we ask to what extent the biases in beliefs from the OSU data are predictive of the gaps between expected and actual careers in the Freshman Survey data.[21] 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.[22] 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

[21] In Appendix Section B.1.2, we additionally provide evidence that errors in OSU students' population beliefs do not appear to be predicted by ways in which the outcomes of OSU graduates differ from the outcomes of college graduates in the United States more generally. We view this as providing further evidence that biases in beliefs in the OSU data and in the Freshman Survey are likely to reflect common mechanisms, rather than aspects of expectations that are specific to students from certain schools.

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

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 from the OSU students' beliefs.

3.4 More Confident Students Stereotype More

Why might stereotypical beliefs 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 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?" Figure A.II shows that more confident students exaggerate stereotypical careers *more* than less confident students ($p < 0.01$, from regressing error on confidence). 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 for Choice, Welfare, and Policy

4.1 Estimating Preferences for Careers and Salary

To investigate potential consequences of stereotypical thinking for students' choices and welfare, in this section we describe a stylized model of major choice. Assume that student i is choosing their major $M \in \{A, B, ...\}$. If they choose M, the probability that they will have career $c \in \{a, b, ...\}$ is $p^i_{c|M}$. Their *belief* about this probability is $\pi^i_{c|M}$. We assume their perceived expected utility (i.e., given their potentially incorrect beliefs) from choosing M are 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 would earn conditional on c and M, which we assume to be known, and α is the (homogeneous and constant) marginal utility of income. We allow for i to have idiosyncratic non-pecuniary preferences over jobs, which are denoted by β^i_c. Next, μ^i_M indicates the known non-labor-market benefits the student would derive from majoring in M (enjoyment of classes, parental approval, etc.). Finally, ν^i_M indicates an unrealized preference shock, whose distribution i knows but whose realized value they do not.

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

$$\log\frac{\pi^i_M}{\pi^i_{M'}} = \alpha\sum_c\left(\pi^i_{c|M}w^i_{c,M} - \pi^i_{c|M'}w^i_{c,M'}\right) + \beta\left(\pi^i_{c^*(i)|M} - \pi^i_{c^*(i)|M'}\right) + \mu^i_M - \mu^i_{M'} \tag{4}$$

To estimate the model, we use data collected in our 2020 OSU survey. We directly elicited $\pi^i_{c|M}$, 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 $\sum_c \pi^i_{c|M}w^i_{c,M}$. We asked students the percent chance they thought they would graduate from OSU with each of the four majors, which we employ as our measure of π^i_M. Finally, we assume μ^i_M is normally distributed and i.i.d. with mean μ_M and variance σ^2. We then estimate the model by maximum likelihood, and compute standard errors and confidence intervals using the Bayesian bootstrap. See Appendix B for more details on estimation and alternative specifications.

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

In contrast, column 1 of Table A.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. We conclude from this analysis 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.

4.2 Which Majors are Worse for Stereotypical Thinkers?

Given students' apparent preferences for careers and salary, how might stereotypical thinking distort students' choices? We focus our analysis on a simple intuitive case: a student on the margin between two majors A and B (i.e., $\widehat{EU}^i[A] = \widehat{EU}^i[B]$), where the student has a positive preference β^i_a for the non-monetary amenities from A's stereotypical career a.[23] To simplify expressions, we also assume this career is (known to be) impossible with major B (i.e., $p^i_{a|B} = \pi^i_{a|B} = 0$). Let $\widehat{EW}^i[M]$ be i's perceived expected salary conditional on majoring in M: i.e., $\widehat{EW}^i[M] = \sum_c \pi^i_{c|M} w^i_{c,M}$. The following expression must then hold:

$$\beta^i_a = \frac{1}{\pi^i_{a|A}} \left[\alpha \Big(\widehat{EW}^i[B] - \widehat{EW}^i[A] \Big) + \mu^i_B - \mu^i_B \right] \tag{5}$$

[23]We assume, as before, that they are indifferent between the non-pecuniary amenities of other jobs.

Equation 5 simply states that, given i's non-monetary preference for A's stereotypical career, their indifference between A and B implies that B must have a higher expected salary or better non-labor-market attributes, or both.

When will stereotypical thinking lead the student to choose a major that leads to an especially lower (objective) expected utility? To provide an answer to this question, we make one final assumption. Namely, we assume that i's beliefs about salaries are unbiased, in the sense that, though their beliefs about likely careers may be incorrect, she does not systematically exaggerate more or less lucrative careers. In Appendix B, we show that this assumption is approximately true on average in our 2020 OSU sample: though students' beliefs about the likelihood of careers are highly predictive of their beliefs about average salaries, beliefs about average salaries by major are nonetheless roughly unbiased.

Given this assumption, equation 6 follows:

$$\text{Bias}^i = EU^i[B] - EU^i[A] = \frac{\pi^i_{a|A} - p^i_{a|A}}{\pi^i_{a|A}} \left[\alpha \Big(EW^i[B] - EW^i[A|c \neq a] \Big) + \mu^i_A - \mu^i_B \right] \tag{6}$$

Equation 6 shows that, unsurprisingly, the student makes a larger mistake (i.e., the expected utility loss is higher) when their beliefs are more distorted (i.e., when $\pi^i_{a|A} - p^i_{a|A}$ is large). Bias is also larger the "riskier" major A is. This is true when the wage penalty between A and B conditional on not working in the stereotypical career ($EW[B] - EW[A|c \neq a]$) is large, and when A's stereotypical job is unlikely ($\pi^i_{a|A}$ is small).[24] The intuition behind this latter result is that, when the stereotypical job is rare, i is accepting a high probability of a wage reduction in exchange for a small chance of achieving the stereotypical job. The fact that i is nonetheless marginal between the two majors implies that they place a high non-pecuniary value on the stereotypical job (i.e., β^i_a is large). This in turn means that errors in the likelihood of attaining that job have especially large consequences for the perceived returns to choosing major A. We summarize these results in Prediction 1.

Prediction 1: *Bias in the returns to major are larger for riskier majors. Majors are "risky" to the extent that the following are both true:*

- *Their stereotypical career is unlikely.*
- *Wages conditional on not having a stereotypical career are low.*

To investigate which majors are "risky" according to the definition in Prediction 1, we look at the 20 most common majors in the US from the American Community Survey. We also separate out economics from business. Motivated by equation 6, we calculate two

[24] Bias would also larger when major A is less attractive for non-labor-market reasons than major B.

statistics for each such major. First, we calculate the share of graduates with that major who are working in a 4-digit occupation that is sufficiently stereotypical of that major. This statistic is our proxy for $\pi^i_{a|A}$ in equation 6. As always, we use a likelihood ratio to define how stereotypical a career is of a major: $Stereotype_{c,M} = p_{c|M}/p_{c|-M}$. We define a career as stereotypical for a major if this ratio is at least 10.[25] Tables A.X and A.XI give examples of the most common stereotypical and non-stereotypical jobs for each major. Second, we calculate the average salary of people with that major who are *not* working in a stereotypical job, which is our measure of $EW^i[A|c \neq a]$.

Figure 3 plots these two statistics for each major, where we see wide variation across both dimensions. First, some majors rarely result in a stereotypical job but have high-paying alternative jobs (the top-left region of Figure 3). For example, only 1.3% of economics majors are "economists and market researchers," but economics has the highest paying alternative jobs of any major. Math, physics, and business are also broadly "general" in this sense. Second, moving to the top-right region of Figure 3, STEM majors such as engineering, computer science, and biology all have relatively more common stereotypical jobs and have higher-paying non-stereotypical jobs. At the bottom-right are health services and education, both of which are lower-paying (conditional on not working in their associated occupations) but which have very common stereotypical jobs. Finally, at the bottom-left of Figure 3 we see which majors are "risky" in the sense described above: majors with rare stereotypical jobs and lower-paying alternatives. Examples of such majors are fine arts, English, communications, and psychology. Note that these majors also tended to be the ones in the Ohio State surveys where the bias from stereotypical thinking was especially severe (that is, where overestimation of stereotypical outcomes was particularly large).

One might worry that for a variety of potential reasons, students may be limited in the number of majors they could successfully pursue. If, for instance, students in risky majors only have the academic preparation to pursue other risky majors, then the welfare implications of stereotypical thinking may be smaller. However, there is growing evidence that the returns to college majors are very large even for academically marginal students (Hastings et al., 2013; Kirkeboen et al., 2016). For example, Bleemer & Mehta (2020) show that American undergraduates who barely pass the academic cutoff to major in economics (a GPA of 2.8 in their setting) reap average salary returns that are almost identical to the cross-sectional differences in average earnings between economics and the major they would otherwise have studied.

[25]The broad conclusions that we draw in this section are not sensitive to this choice of cutoff. We include those without a college degree when calculating $p_{c|-M}$ to reduce noise, which also does not affect any of the main conclusions.

A second worry is that these biased beliefs might motivate students to work harder in school than they otherwise would. If students are studying too little (e.g., because they are present biased), then stereotypical thinking could be welfare-improving through its effect on study effort. To provide suggestive evidence on this possibility, the first 2021 OSU survey asked students how many hours a week they anticipated studying the following (Spring) semester. Later in the survey, it asked students to imagine that they had to enter a career other than the one they listed as their most likely job conditional on their top ranked major. It then asked, if this were true, how many hours a week they would expect to study next semester. If stereotyped beliefs contributed to high effort, we might expect a large dropoff in effort across these two questions. In fact, however, students report that they would study slightly *more* if they had to choose an alternative career (12.6 vs 12.4 hours, $p = 0.02$), with a slightly larger difference (12.1 vs 11.5 hours, $p < 0.01$) if we restrict to students intending to pursue risky majors.[26] We conclude from these analyses that, at least according to students themselves, high beliefs about the likelihood of preferred jobs are not maintaining high effort in school.

4.3 Suggestive Evidence on Long-Term Implications

How do long-run outcomes vary across these types of majors? To shed light on this question, we examine data from the 2013 National Survey of College Graduates (National Science Foundation, 2013) and the 2021 Survey of Household Economics and Decisionmaking to provide suggestive correlational evidence on post-college outcomes depending on students' field of study. Table 3 regresses various outcomes on a dummy variable indicating whether a graduate's major is risky (which in these datasets we define as humanities, psychology, art, communications, and social/behavioral sciences). We see that graduates with risky majors are 39% more likely not to be employed; 17% more likely to be dissatisfied with their job; 17% less likely to have a job related to their highest degree; 65% more likely to report that the reason they do not have such a job is that one was not available; earn 22% lower salaries; have 28% more outstanding student debt; are 78% more likely to say the costs of their Bachelor's degree were somewhat or much larger than the lifetime financial benefits; and are 35% more likely to report that they would choose a different field of study if they could go back and make decisions regarding their education again (all comparisons significant at the $p < 0.05$ level). This correlational evidence—while merely suggestive—is consistent with Prediction

[26]These results are not changed if we restrict the sample to students whose most likely career is also the stereotypical career of their top major. We did not automatically ask about stereotypical careers in the survey because we did not want to alert students that this was our focus. Instead, the survey informs them (truthfully) that we chose that career because they listed it as their most likely job. In practice, for 84% of students, their reported most likely career is also their top major's most stereotypical career.

1's claim that stereotypical thinking may lead some students to make worse human capital investments, especially when it pushes them to pursue risky subjects.

4.4 The Impact of Providing Statistical Information: Evidence from a Field Experiment

Given the potential welfare consequences of stereotypical thinking, what types of policies might help students to make more informed decisions? We tested one such intervention during the second 2021 OSU survey: a low-cost, light-touch information intervention. In a randomized controlled trial, we provided half of students with truthful statistical information about the joint distribution of majors and careers to test for impacts on students' beliefs and choices. The survey began by asking students the percent chance that they would graduate with their top two 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, an information module 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, we told them several headline numbers about the frequency of the careers they had listed as their most likely if they graduated with that major. 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.[27] Figure A.III shows that students do update their beliefs about their own careers in a sensible direction, reducing beliefs about their likelihood of achieving their top major's stereotypical career when they overestimated its frequency. However, this updating is far from one-for-one (an OLS regression yields a coefficient of 0.30), indicating that students appear to substantially discount this population information when updating their self beliefs.

We look at treatment effects on two main outcomes. First, we use a within-survey measure of change in intentions about what to major in. At the end of the survey, both

[27]The information module (filled in with fictitious previous answers) can be accessed at this link.

treatment and control groups were re-asked the earlier question about the percent chance that they would graduate with their two top majors.[28] Second, we use administrative data, linked to students' survey responses, on the classes they took in each major during the Fall 2021 semester (pre-treatment) as well as Spring and Fall 2022 semesters (post-treatment).

Column 1 of Table 4 show OLS estimates where the dependent variable is students' updated belief about their likelihood of graduating with their top-ranked major. We regress this outcome on a treatment dummy interacted with a dummy indicating whether the student over or underestimated their top major's stereotypical job (a proxy for whether the information was "good news" or not). We also control for their original belief about their chance of graduating with that major. We see that students on average reduce their stated likelihood of graduating with their top major in response to the information, with slightly (but not significantly) larger effects for the majority of students who overestimated the frequency of that major's stereotype (-3.6 percentage points, $p < 0.01$) than those underestimated it (-2.9 p.p., $p < 0.10$). Columns 3 and 5 show similar regressions restricting to students whose top-ranked major is "risky," according to the definition from Section 4.2, or not. Among our 10 groups of college majors, risky majors include humanities, psychology, art, and communications. For students considering a risky major, we see negative effects concentrated among those who overestimated its stereotypical career (-7.5 p.p., $p < 0.01$) but no effects for those who underestimate it (0.6 p.p, $p = 0.91$). For those considering less risky majors, the pattern is harder to interpret: slightly (but not significantly) larger effects for those who *under*estimate their majors stereotype (-3.2 p.p., $p < 0.10$, vs -2.1 p.p., $p = 0.12$).

Columns 2, 4, and 6 of Table 4 show similar regressions but where the dependent variable is the number of classes they took in their top-ranked major post-intervention (adding up their Spring and Fall 2022 classes) and controlling for the number of such classes they took in Fall 2021 (pre-treatment). We see that, for the full sample, students who overestimated their top major's stereotype take slightly fewer classes (-0.29 classes, $p = 0.15$) and those who underestimated it take slightly more (0.22 classes, $p = 0.35$) in response to the information. This difference in estimated treatment effects is marginally statistically significant ($p < 0.10$). We see larger differences for students considering a risky major (-.29 vs 1.52 classes, difference-in-difference significant at $p < 0.05$) than for those considering a non-risky major (-.23 vs 0.08, $p = 0.27$).

Recall that risky majors, by our definition, are ones that have both a rare stereotypical

[28] This question was phrased so that it would make sense to both the treatment and control group. In particular, it read, "After thinking about things throughout the course of the survey, we'd like to ask you again about your likely major. It's perfectly fine if your answers don't change compared to before. We just want your honest opinion. What do you think is the percent chance that you will... graduate with a major in M1? graduate with a major in M2? Graduate from OSU with any other major? Not graduate from OSU?"

career and low-paying alternatives to that career. Tables A.XII and A.XIII show analogous regressions to those in Table 4 but splitting the data, respectively, just by whether student's top major has a rare stereotypical outcome (i.e., the risky majors plus Biology/Chemistry and Government/Political Science) and just by whether the alternatives to that career are lower-paying (i.e., the risky majors plus Education and Nursing). We see less pronounced hetereogeneity when we split only according to salary and more pronounced heterogeneity when we split only according to how rare stereotypical outcomes are. These differences from the main specification are small, so caution is warranted when interpreting them, but they are consistent with the finding in Section 4.1 that students had stronger preferences for the non-pecuniary aspects of jobs than for expected salaries.

5 A Memory Model of Belief Formation

Given the potential welfare implications of stereotypical thinking, a natural question is why this bias arises. More broadly, how do students form their career expectations? To explore these questions, in this section we describe a simple cognitive model of belief formation in which students form their expectations by drawing on their experiences—the people they know or have heard about with certain combinations of jobs and fields of study. Our structure applies the basic "similarity and interference" framework of Bordalo, Conlon, et al. (2022) to model retrieval from memory. We first show that the model predicts stereotyping as an endogenous consequence of assoicative recall. We then derive additional predictions, which we test using survey data from our OSU samples.

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 in this section that there are only two majors. As our running example, we will often refer to major A as art and major B as business, and to career a as being an artist and career b as being a businessperson.

The student separately assesses the "plausibility" $F(H)$ of each relevant "hypothesis" H. When assessing unconditional probabilities, these hypotheses H_c just correspond to the set of people with each career c. When assessing probabilities of careers conditional on major M, these hypotheses $H_{c,M}$ correspond to the set of people with *both* career c and major M. The student's probabilistic beliefs about H, shown in equation 7, are then just the plausibility of

H normalized such that their beliefs about all relevant hypotheses sum to one.

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

To assess plausibilities, we assume the student repeatedly follows a two-stage process for each hypothesis separately. First, they retrieve an experience from their memory "database" D. We can think of the database as comprised of people the student knows personally like friends or family, those she has 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 retrieved has the career/major associated with the hypothesis they are assessing. The more such "successes" they have, the more likely a hypothesis will seem.

More precisely, we assume that the likelihood of retrieving a person e when assessing the plausibility of hypothesis H depends on e's *availability*, denoted by $a(e, H)$. We assume retrieval is both associative—that is, similarity-based—and frequency-based, meaning that people she has encountered more times will come to mind more easily (Kahana, 2012). These two forces correspond to the two components of equation 8, which we describe in turn.

$$a(e, H) = N(e)S(e, H) \tag{8}$$

Frequency-Based Memory. First, someone is more likely to come to mind the more times the student has encountered them. This force, captured by $N(e)$ in equation 8, will mean that people the student is personally close—e.g., their parents or other role models, whom they have encountered many times—will have an outsized impact on their beliefs compared to people they have only met or heard of a few times.[29]

To tractably analyze these frequency effects, we assume one person, whom we call x, is the student's role model and therefore more likely to come to mind than their similarity to H would otherwise suggest. We also assume $N(e)$ follows the simple functional form in equation 9, where $\phi \geq \frac{1}{D}$ represents the fraction of the student's experiences that are with x (and, abusing notation somewhat, D denotes both the memory database itself and the number of people in it).

$$N(e) = \Big(\phi D\Big)^{\mathbb{1}(e=x)} \tag{9}$$

[29] We assume D is representative of the population, in the sense that the people in it reflects the true joint distribution of careers and majors. If there are people the student has never heard of or encountered, this would be captured by $N(e) = 0$. 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.

Associative Memory. Second, we assume someone is more likely to come to mind the more similar they are to the hypothesis at hand, which is captured by a similarity function $S(e, H)$. More precisely, let $s(e, u)$ be the similarity between two people e and u. We then define the similarity between a person and a hypothesis as simply the average pair-wise similarity between that person and all experiences consistent with the hypothesis, as shown in equation 10.

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

These assumptions imply 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 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 and need to be discarded.

We assume the functional form for the similarity function $s(e, u)$ between people e and u given by equation 11.

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

Similarity thus decreases by a factor of $\delta_c \leq 1$ if e and u have different careers and by $\delta_M \leq 1$ if they have different majors. 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., 2022). We 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 the present paper.

Equipped with this definition of availability, we can define the probability $r(e, H)$ that a person e is retrieved when assessing hypothesis H by equation 12.

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

The numerator of equation 12 captures the notion that more available people are more likely to come to mind. The denominator illustrates the idea that experiences compete for retrieval, or "interfere" with each other: thus, factors that make one person come to mind more easily (i.e., similarity or frequency) do so at the expense of others.

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 retrieval-and-checking many times, counting up the number of "successes" they retrieve for each hypothesis.[30] Their assessed plausibility then converges in probability to equation 13 where the expectation operator is with respect to frequencies in the student's database D. All proofs are in Appendix C.

$$F(H) = \sum_{e \in \mathcal{D}} r(e, H)\mathbb{1}(e \in H) = \frac{E\Big[a\big(e, H\big)\mathbb{1}(e \in H)\Big]}{E\Big[a\big(e, H\big)\Big]} \tag{13}$$

This model naturally nests the rational-expectations benchmark. In particular, if $a\big(e, H\big)$ is constant, meaning $\delta_c = \delta_M = 1$ and $N(e) = 1$ for all e, 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, without overweighting people they have encountered more times.

5.2 Discussion of Assumptions

The above 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 what e's career and major are. 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 i does not know 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. However, we show in Appendix C that, if the student's beliefs about e's career and major are unbiased, then this alternative model is equivalent to the model described above.[31]

Second, the model described above does not allow the student to extrapolate from one major to another or from one career to another, which may not be realistic. For example,

[30] Note that this formulation allows "double-counting"; i.e., if someone comes to mind twice, they get double the weight.

[31] Of course, there could be interesting extensions of the model incorporating potentially biased beliefs about individual's careers and majors. For example, perhaps students stereotype people's *major* on the basis of their career.

suppose i's is assessing how many humanities majors become artists. According to equation 13, thinking of artists with the wrong major (e.g., art) counts equally against this hypothesis as thinking of people with completely different careers and majors. Intuitively, however, we might think that being able to think of many artists—albeit with the wrong major—might by extrapolation make that career seem less implausible. In Appendix C, we describe a more general version of the model, allowing for "simulation" from one experience to similar-but-distinct possibilities, that accommodates this intuition (Kahneman & Tversky 1981, Bordalo, Burro, et al. 2022). We return to this point below when discussing the effect role models have on students' beliefs.

5.3 Beliefs about Careers Conditional on Major

Stereotyping. We start by showing that the model predicts the pattern of stereotypical thinking that we documented in Section 3. 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. To investigate stereotypical thinking, we can ask how beliefs about major A change as we increase $p_{a|B}$. With rational expectations, changing the distribution of careers within major B should not change beliefs about A. In our model, however, the distribution of careers in the alternative major can affect retrieval despite being normatively irrelevant for the question at hand. These effects are summarized in Prediction 2.

Prediction 2: *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 $\delta_c < 1$.*

To intuitively interpret Prediction 2, consider a student forming beliefs about how many art majors go on to become artists. Prediction 2 says their belief will be higher if fewer people with the other major (business) have that job. Intuitively, the more business majors are artists, the harder it is for the student to think of art majors with that job. This reflects the role of associative recall in the student's belief formation process. In contrast, if *no* business majors are artists, then the student will easily retrieve art majors who are, since there are few other similar people in their database that compete much for retrieval. In the latter case, becoming an artist after majoring in art will seem more plausible, even if the true share of art majors with that job is held constant.

Figure 4 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 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.

Neglecting common alternatives to stereotypes. When students exaggerate the likelihood that a major leads to its stereotypical career, which outcomes do they underestimate? One plausible *ex ante* possibility was that students might underestimate all non-stereotypical outcomes. However, the model predicts that, besides stereotyping, the student's beliefs should be *undersensitive* to true frequencies, meaning that they should underestimate common careers and overestimate rare careers. We summarize this result in Prediction 3.

Prediction 3: *Absent distortions from stereotyping or role models, the student exaggerates rare careers and underestimates common careers.*

More precisely, let $p_{c,B} = 0$ for all careers c (shutting down stereotyping effects) and $\phi = 0$ (shutting down the effect of role models). Then equation 14 follows whenever $\delta_c < 1$.

$$\frac{\pi_{a|A}}{\pi_{b|A}} > \frac{p_{a|A}}{p_{b|A}} \iff p_{a|A} < p_{b|A} \tag{14}$$

The intuition behind this result is straightforward. When $\delta_c < 1$, meaning that retrieval is based on similarity, the student disproportionately retrieves people consistent with the career that they are assessing. For example, when thinking about artists they tend to retrieve artists, and when thinking about businesspeople they tend to retrieve people with that profession. This disproportionately benefits rarer hypotheses which, under unbiased random sampling, would more often fail to come to mind.

Panel B of Figure 4 plots the actual and perceived conditional likelihood for all non-stereotypical combinations of careers and majors. We see that students tend to exaggerate very rare careers such as 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%). Instead, students tend to underestimate relatively common but non-stereotypical outcomes. Figure 4 highlights four such outcomes: working in business, as a teacher, in an "other" job, and not working for pay. These four outcomes tend to be common alternatives to many majors' stereotypical career. A particularly striking case is non-employment (triangles in Figure 4). 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.[32,33]

Role models. We next analyze how the increased availability of role models can systematically affect beliefs. To do so, we 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. As we will see, the effect depends on the career and major of x. First, suppose that the student is forming beliefs about major A and that their role model graduated with that major. Prediction 4 summarizes the effect this has on their beliefs.

Prediction 4: *The student's beliefs $\pi_{a|A}$ about the probability of career a conditional on major A increase if they personally know someone with that career-major pair.*

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

This result is intuitive; knowing someone personally with a career-major pair boosts the chances that the student retrieves someone with that exact outcome, which counts as evidence in favor of that hypothesis.

Second, we can ask how the student's beliefs about one major are affected by knowing someone with a *different* major. Prediction 5 summarizes such effects.

Prediction 5: *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 5 is the same as for stereotyping. The student's role model x comes to mind easily when they try to think about people with career a. This distracts from recalling people who have both career a and major A, reducing the student's beliefs about that hypothesis.

To test Predictions 4 and 5, 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

[32]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.XIV 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.

[33]Figure A.IV in the Appendix breaks up survey responses and actual employment probabilities by gender. The broad patterns are very similar to those in Figure 4. In particular, both men and women underestimate the share of graduates (of each major) who is non-employed, and this holds even when comparing their beliefs (about people of any gender) to data just on graduates of the student's gender.

might turn to for advice about choosing your college major or other aspects of planning for your schooling and eventual career." We chose the "role models" framing to allow students to name influential individuals other than a mother or father, though 84% of students answered about at least one parent, and 50% answered about two. The survey then asked the student's relation to this person, their level of education, gender, race, college major (if applicable), and occupation.[34] The options for their role models' major and occupation were the same groups of careers and majors that we focus on throughout the paper.[35] All questions about role models were asked after we elicited students' beliefs about careers, in order to avoid appearing to suggest that they should base their beliefs on the careers/majors of the people they know personally.

Table 5 shows OLS estimates of the following regression specification using these 2021 OSU data:

$$\pi^i_{c|M} = \alpha + \beta_1 RM^i_{c,M} + \beta_2 RM^i_{c,-M} + \mu_{c,M} + \epsilon^i_{c,M} \tag{15}$$

In equation 15, $\pi^i_{c|M}$ is student i'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 not M. Finally, $\mu_{c,M}$ are career-by-major fixed effects, indicating that all effects are driven by variation across individuals in the career/major of their role models. We cluster standard errors at the student level.

Columns 1 and 5 of Table 5 shows 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. These results are consistent with Prediction 3. Note that these effects persist (and indeed, are larger) when we restrict the data to students' population beliefs about each major's stereotypical career (column 2). Knowing someone with these careers, which student's already exaggerate, therefore makes students' beliefs *less* accurate.

Prediction 5 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 5 shows, instead,

[34] One might worry that because, almost by definition, role models are more influential than others, our results are for that reason unsurprising. However, note that our model predicts, not just that role models will have *any* effect on students' beliefs, but that beliefs respond to role models' careers and majors *in a particular direction.* These signed predictions, in our view, are not always what one would expect absent the model. For example, we might have expected students who personally know a journalist to be better informed (i.e., have lower beliefs about P(journalist | journalism)) than students without such a role model.

[35] 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.

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 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 5 only appears true for stereotypical outcomes.[36]

What might explain these ambiguous effects? In Appendix C, we show that a simple extension of the model adding a role for extrapolation or simulation can predict positive effects for sufficiently rare or implausible outcomes (Kahneman & Tversky, 1981; Bordalo, Burro, et al., 2022). 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 does not 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.

5.4 Beliefs about Careers Unconditional on Major

Next, we turn to the model's predictions for beliefs about careers *unconditional* on major. To test these predictions, in the first of the 2021 OSU surveys, we asked students about the distribution of careers of US college graduates *unconditional* on major. In particular, 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 unconditional questions occurred before the questions that conditioned on major in order to avoid confusion.

Role models. Prediction 6 summarizes the effect of the student's role model on their unconditional beliefs.

Prediction 6: *Personally knowing someone with career c increases beliefs about the probability of career c unconditional on major. More precisely, let $c(x) = a$. Then, regardless of $M(x)$, $\frac{\partial}{\partial \phi}\pi_a > 0$.*

The intuition behind Prediction 6 is the same as for Prediction 4; having a role model with career c makes that career easily retrievable, which boosts the plausibility of H_c and thus beliefs about c.

[36] We obtain similar results if we instead cut the data by whether the career is sufficiently objectively likely, rather than by whether it is stereotypical.

Columns 4, 8, and 9 of Table 5 show OLS estimates of equation 16, where π_c^i is the student's unconditional belief about career c, RM_c^i is the number of role models they list with that career, and μ_c are career fixed effects.

$$\pi_c^i = \alpha + \beta RM_c^i + \mu_c + \epsilon_c^i \tag{16}$$

Column 4 shows that, as predicted, 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 beliefs about their own future career, though other mechanisms (e.g., networks, preferences) could drive this result. Column 9 shows analogous estimates from the Freshman Survey: 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.[37]

Concentrated and rare careers. A more novel prediction of the model, compared to those for conditional beliefs, stems from the fact that unconditional hypotheses are about groups (people with certain careers) whose members can potentially differ from each other (in that they may have graduated with different majors). Prediction 6 summarizes how this affects the student's beliefs:

Prediction 7: *Unconditional beliefs about the frequency of a career increase in the extent to which it is concentrated within particular majors.*

More precisely, let $\phi = 0$ (shutting down effects from role models) and let $p_{A|c} = p_{B|c}$ for all careers $c \neq a$. Then equation 17 follows whenever $\delta_M < 1$:

$$\frac{\partial}{\partial p_{B|a}} \pi_a > 0 \iff p_{B|a} > p_{A|a} \tag{17}$$

The intuition behind Prediction 7 lies in the fact that if everyone with a particular career is very similar to each other, because they all followed the same academic path before entering that profession, they will be easy to retrieve. In contrast, if people enter a career via many different majors, they will all be less similar to each other and therefore come to mind less easily. Thus, beliefs about a career are higher when people with that career are concentrated within a particular major.

Finally, absent the effects described in Predictions 6 and 7, unconditional beliefs (like conditional beliefs) are undersensitive to true frequencies. We summarize this result in

[37] The Freshman Survey asks the career, but not the major, of students' parents.

Prediction 8:

Prediction 8: *Absent effects from role models and differential concentration of careers within majors, the student exaggerates rare careers and underestimates common careers.*

More precisely, let $\phi = 0$ (shutting down the effect of role models) and let $p_{A|c} = p_{B|c}$ for all careers c (shutting down interference due to differing concentration of careers within majors). Then equation 18 follows whenever $\delta_c < 1$:

$$\frac{\pi_a}{\pi_b} > \frac{p_a}{p_b} \iff p_b > p_a \tag{18}$$

The intuition for Prediction 8 is the same as for conditional beliefs. When $\delta_c < 1$, meaning that retrieval is based on similarity, the student is disproportionately able to think of people with a when they are thinking of that career. This disproportionately benefits rarer careers that would more often fail to come to mind under unbiased sampling.

To test Predictions 7 and 8, we estimate by OLS the following regression equation:

$$\overline{BeliefError}_c = \alpha + \beta_1 TrueShare_c + \beta_2 HHI_c + \epsilon_c \tag{19}$$

In equation 19, $\overline{BeliefError}_c$ 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 5 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 7 and 8: 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 ($p < 0.01$).

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 show that this bias appears important for understanding students' choice of major and has potentially important welfare consequences as it boosts demand for risky academic paths. A field experiment shows that a light-touch intervention aimed at combating stereotypical thinking can have significant, though in some cases modest, effects on students intentions and choices. Finally, we use a simple model of belief formation to that show stereotyping can arise as a natural consequence of associative memory. 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.

We close with three more speculative points. First, our framework differs from traditional economic models of belief formation in that it does not assume that agents have well-formed or internally consistent beliefs about all possible states that they consult when reporting, updating, or acting on their expectations. Rather, agents form their beliefs about the question at hand "on the fly" by consulting their memory. This viewpoint may shed light on the modest effect sizes commonly found in interventions, like ours, that provide statistical information to correct even substantially mistaken beliefs (see L. C. Coffman et al. 2022 for a meta-analysis). If agents form beliefs in the way our model assumes, then statistical information may sometimes simply fail to come to mind (especially if choices are measured at a delay or in a different context).

Second, we find that the careers and majors of people close to students significantly shape their beliefs well beyond the "objective" information they convey. Common explanations behind the impact that role models or local environments have on later outcomes include providing access to networks, better resources, better advice, and changing preferences for career paths (Chetty et al., 2016; Bleemer, 2016; Bell et al., 2019; Chung et al., 2018). Our results point to a related yet distinct potential mechanism: that role models might simply change beliefs about the sorts of jobs that are out there.

Finally, 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).[38] *Ex ante*, of course, rational mechanisms could have fully explained these patterns: e.g., students with correct beliefs might rationally pursue certain career paths which, though very unlikely to pan out, they feel are worth the risk (e.g., journalism or film), or students may realize that certain majors (e.g., psychology) provide a general education not intended for

[38]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.

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 to their stereotypical jobs with exaggerated likelihoods. To the extent that these human capital investments are irreversible and costly, finding ways to help students make better informed decisions or to steer them toward less risky academic paths may thus have substantial benefits.

References

Agan, A., & Starr, S. (2017). Ban the Box, Criminal Records, and Racial Discrimination: A Field Experiment. *The Quartely Journal Of Economics*, 191–235.

Alesina, A., Ferroni, M., & Stantcheva, S. (2021). Perceptions of Racial Gaps, Their Causes, and Ways to Reduce Them. *SSRN Electronic Journal*.

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

Altmejd, A., Barrios-Fernández, A., Drlje, M., Goodman, J., Hurwitz, M., Kovac, D., ... Smith, J. (2021, 8). O brother, where start thou? sibling spillovers on college and major choice in four countries. *Quarterly Journal of Economics*, *136*, 1831-1886.

Altonji, J. G., Blom, E., & Meghir, C. (2012). Heterogeneity in human capital investments: High school curriculum, college major, and careers. *Annual Review of Economics*, *4*, 185–223.

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).

Arnold, D., Dobbie, W., & Yang, C. S. (2018). Racial bias in bail decisions. *Quarterly Journal of Economics*, *133*(4), 1885–1932.

Augenblick, N., & Rabin, M. (2021, 5). Belief movement, uncertainty reduction, and rational updating. *Quarterly Journal of Economics*, *136*, 933-985. doi: 10.1093/qje/qjaa043

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.

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.

Bell, A., Chetty, R., Jaravel, X., Petkova, N., & Van Reenen, J. (2019). Who Becomes an Inventor in America? The Importance of Exposure to Innovation. *Quarterly Journal of Economics*.

Benjamin, D. J. (2018). Errors in Probabilistic Reasoning and Judgment Biases. *SSRN Electronic Journal*, *2*.

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

Benjamin, D. J., Moore, D. A., & Rabin, M. (2017). Biased beliefs about random samples: Evidence from two integrated experiments. *NBER Working Paper Series*, 65.

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

Biderman, N., Bakkour, A., & Shohamy, D. (2020, 7). *What are memories for? the hippocampus bridges past experience with future decisions* (Vol. 24). Elsevier Ltd.

Bleemer, Z. (2016). Role Model Effects of Female Stem Teachers and Doctors on Early 20 Th Century University Enrollment in California. , 1–19.

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.

Blom, E., Cadena, B. C., & Keys, B. J. (2021). Investment over the business cycle: Insights from college major choice. *Journal of Labor Economics*, *39*(4), 1043–1082.

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. (2022). Memory and covid pessimism. *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. (2022). Memory and Probability. *Working paper.*.

Bénabou, R., & Tirole, J. (2016). Mindful economics: The production, consumption, and value of beliefs. *Journal of Economic Perspectives*, *30*, 141-164. Retrieved from http://pubs.aeaweb.org/doi/10.1257/jep.30.3.141 doi: 10.1257/jep.30.3.141

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

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

Chetty, R., Hendren, N., & Katz, L. F. (2016). The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment. *American Economic Review*, *106*, 855-902.

Chung, B., Chung, B. W., Walker, J. E., Barkowski, S., Durlauf, S., Eble, A., ... Wood, D. (2018). *Peers' Parents and Educational Attainment Peers' Parents and Educational Attainment ** (Tech. Rep.).

Coate, S., & Loury, G. C. (1993). Will Affirmative-Action Policies Eliminate Negative Stereotypes? *The American Economic Review*, *83*(5), 1220–1240.

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

Coffman, L. C., Featherstone, C. R., & Kessler, J. B. (2022). A model of information nudges. *Working paper.*.

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

Conlon, J. J., Mani, M., Rao, G., Ridley, M. W., & Schilbach, F. (2022). Failing to Learn from Others. *NBER Working Paper 28844*.

Conlon, J. J., Pilossoph, L., Wiswall, M., & Zafar, B. (2018). Labor Market Search With Imperfect Information and Learning. *NBER Working Paper 24988*.

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.

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

Dougherty, M. R., Gettys, C. F., & Thomas, R. P. (1997). The role of mental simulation in judgments of likelihood. *Organizational Behavior and Human Decision Processes*, *70*, 135-148.

Enke, B. (2020). What you see is all there is. *Quarterly Journal of Economics*, *135*, 1363-1398. doi: 10.1093/qje/qjaa012

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.*.

Enke, B., & Zimmermann, F. (2019). Correlation neglect in belief formation. *Review of Economic Studies*, *86*, 313-332. doi: 10.1093/restud/rdx081

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.

Gagnon-Bartsch, T., & Bushong, B. (2022, 7). Learning with misattribution of reference dependence. *Journal of Economic Theory*, *203*. doi: 10.1016/j.jet.2022.105473

Gagnon-Bartsch, T., Rabin, M., & Schwartzstein, J. (2017). Channeled attention and stable errors. *Working paper*.

Georganas, S., Healy, P. J., & Li, N. (2014, 4). Frequency bias in consumers' perceptions of inflation: An experimental study. *European Economic Review*, *67*, 144-158. doi: 10.1016/j.euroecorev.2014.01.014

Giglio, S., Maggiori, M., Stroebel, J., & Utkus, S. (2021). Five facts about beliefs and portfolios. *American Economic Review*.

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

Golman, R., Loewenstein, G., Molnar, A., & Saccardo, S. (2021, 12). The demand for, and avoidance of, information. *Management Science*. doi: 10.1287/mnsc.2021.4244

Hassabis, D., Kumaran, D., Vann, S. D., & Maguire, E. A. (2007). Patients with hippocampal amnesia cannot imagine new experiences. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 1726-1731.

Hastings, J. S., Neilson, C. A., Ramirez, A., & Zimmerman, S. D. (2016). (Un)informed college and major choice: Evidence from linked survey and administrative data. *Economics of Education Review*, *51*, 136–151.

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.

Jäger, S., Roth, C., Roussille, N., & Schoefer, B. (2021). Worker beliefs about rents and outside options.

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.*.

Kuchler, T., & Zafar, B. (2019). Personal experiences and expectations about aggregate outcomes. *Journal of Finance*.

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. (2021). Exposure, Experience, and Expertise: Why Personal Histories Matter in Economics. *Journal of the European Economic Association*.

Mueller, A. I., Spinnewijn, J., & Topa, G. (2021). Job seekers' perceptions and employment prospects: Heterogeneity, duration dependence and bias *. *American Economic Review*.

National Science Foundation. (2013). National Survey of College Graduates (NSCG). *National Science Foundation*.

Nordin, M., Persson, I., & Rooth, D. O. (2010). Education-occupation mismatch: Is there an income penalty? *Economics of Education Review*, *29*(6), 1047–1059.

Porter, C., & Serra, D. (2020). Gender differences in the choice of major: The importance of female role models. *American Economic Journal: Applied Economics*, *12*(3), 226–254.

Rabin, M. (2002). Inference by believers in the law of small numbers. *Quarterly Journal of Economics*, *117*, 775-816.

Riise, J., Willage, B., & Willén, A. (2022, 7). Can female doctors cure the gender stemm gap? evidence from exogenously assigned general practitioners. *The Review of Economics and Statistics*, 1-15.

Robst, J. (2007). Education and job match: The relatedness of college major and work. *Economics of Education Review*, *26*(4), 397–407.

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.

Schacter, D. L., Addis, D. R., Hassabis, D., Martin, V. C., Spreng, R. N., & Szpunar, K. K. (2012). The future of memory: Remembering, imagining, and the brain. *Neuron*, *76*, 677-694.

Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science*, *10*(1), 80–83.

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

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*.

Weber, M., D'Acunto, F., Gorodnichenko, Y., & Coibion, O. (2022). The subjective inflation expectations of households and firms: Measurement, determinants, and implications. *NBER Working Paper 30046*.

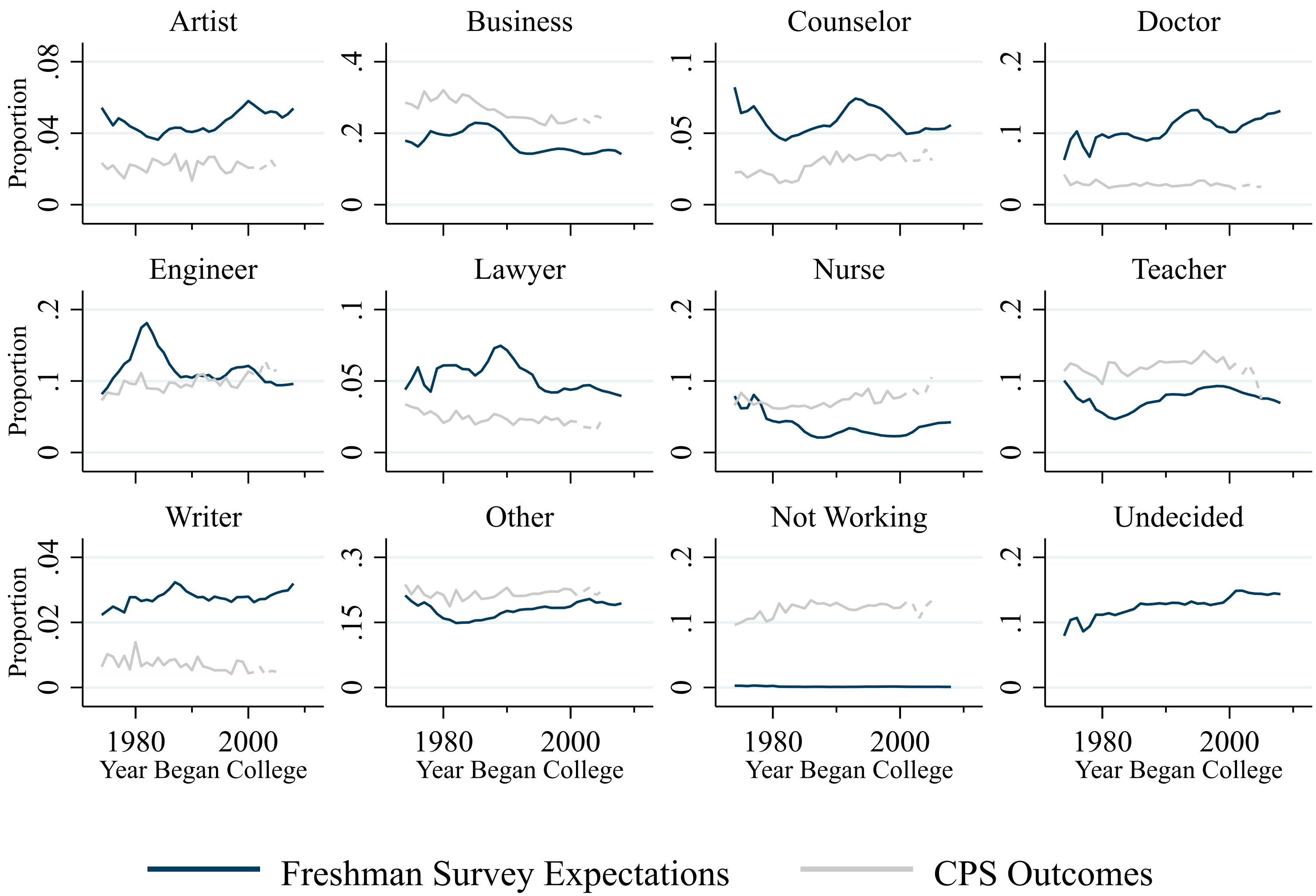
Wiswall, M., & Zafar, B. (2015a). Determinants of college major choice: Identification using an information experiment. *The Review of Economic Studies*, *82*(2), 791–824.

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

Wiswall, M., & Zafar, B. (2018). Preference for the workplace, investment in human capital, and gender. *Quarterly Journal of Economics*, *133*(1), 457–507.

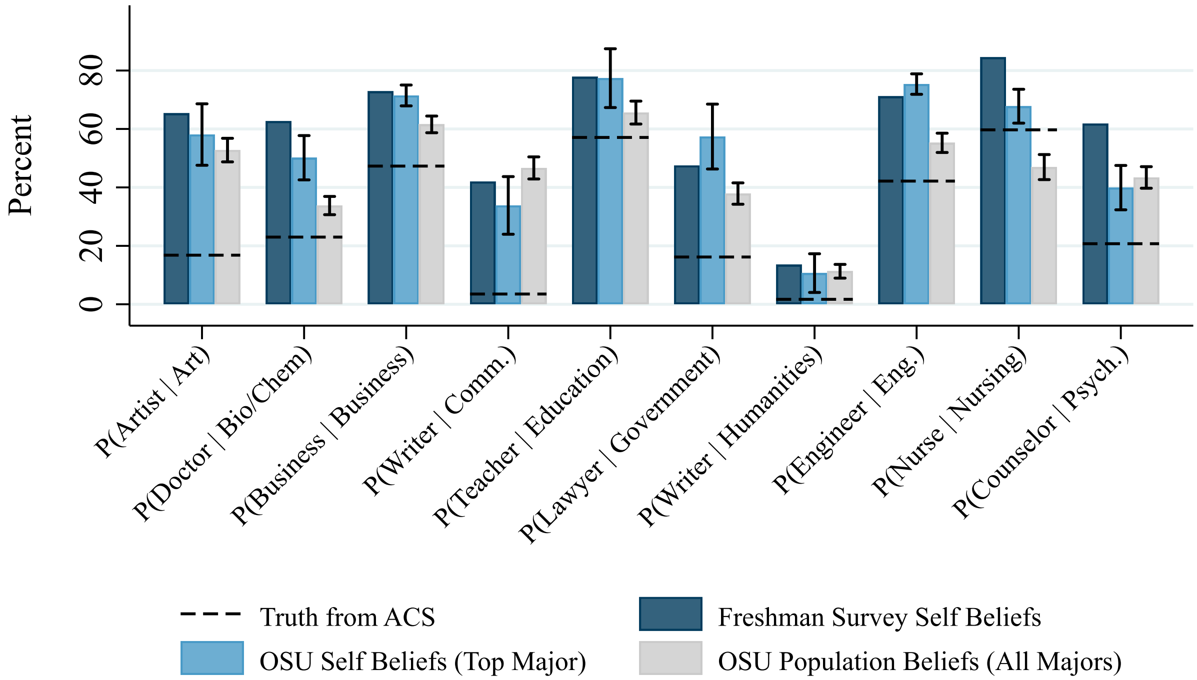
Wiswall, M., & Zafar, B. (2021). Human capital investments and expectations about career and family. *Journal of Political Economy*, *129*(5).

Figure 1: Career Expectations vs. Outcomes Over Time



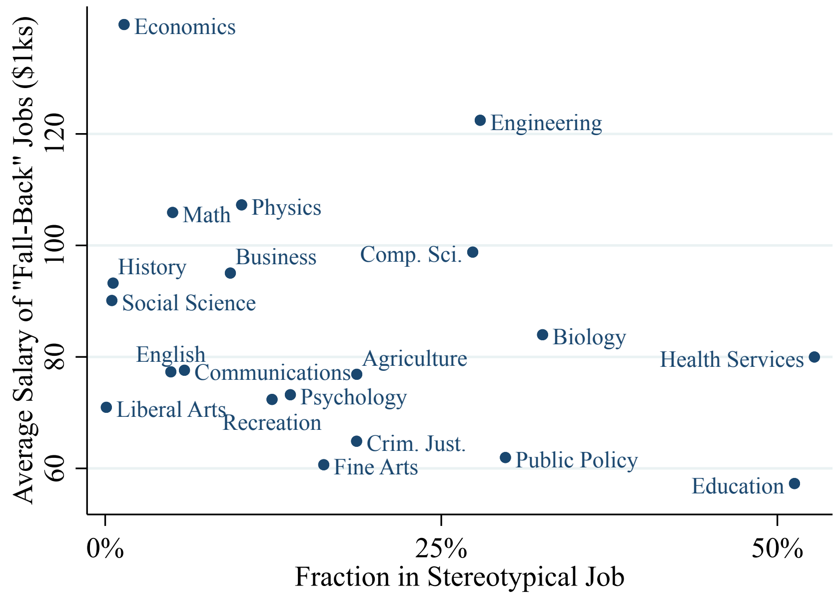
Notes: Figure 1 compares the share of first-year undergraduates by the year they began college (birth year plus 18) who expect to have a career in each occupation (blue line) according to the Freshman Survey data along with the share of college graduates (gray line) aged 33 to 37 who work in that occupation 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.

Figure 2: Exaggerating Stereotypical Careers



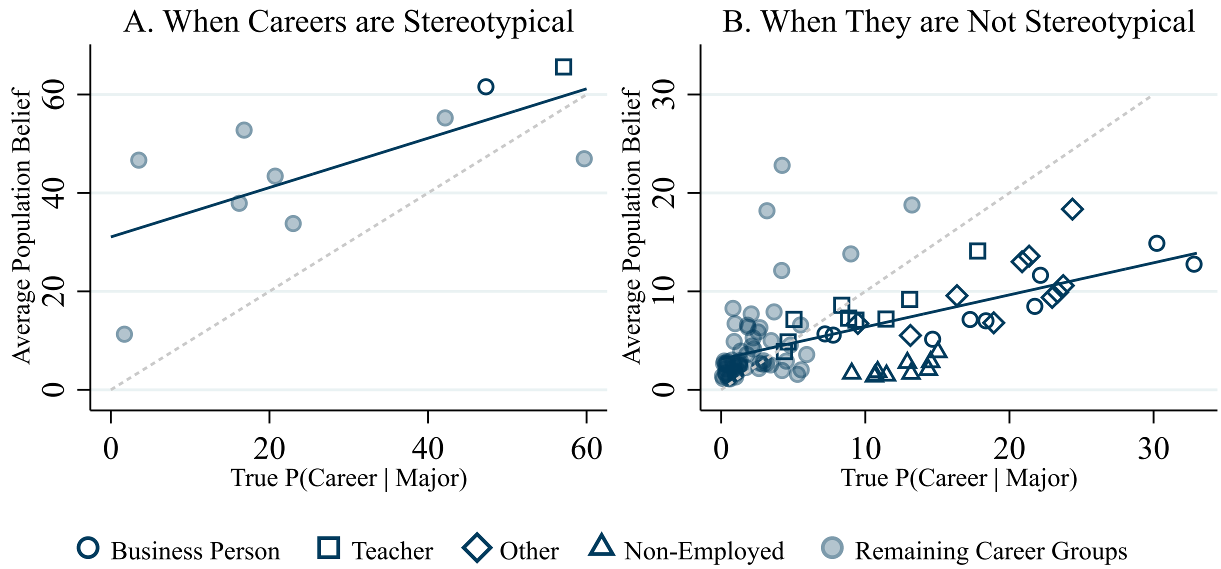
Notes: Figure 2 presents average statistics regarding the most stereotypical profession (as defined in section 3) for each major. The dashed horizontal lines denote the actual proportion of college graduates with each major between the ages of 30 and 50 that are working in that major's most stereotypical 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 profession as their probable career occupation. The light blue bar plots the average belief for the 2020 OSU sample about the probability that they would be working in each career at age 30 if they graduated from Ohio State with their top-ranked (i.e., most likely) major. The gray bars show the average belief among our 2020 OSU sample about the fraction of Americans between the ages of 30 and 50 who graduated college with each major (not only their top-ranked major) that are working in each occupation. Error bars show 95% confidence intervals for the mean of the OSU beliefs.

Figure 3: Which Majors are Risky for Stereotypical Thinkers?



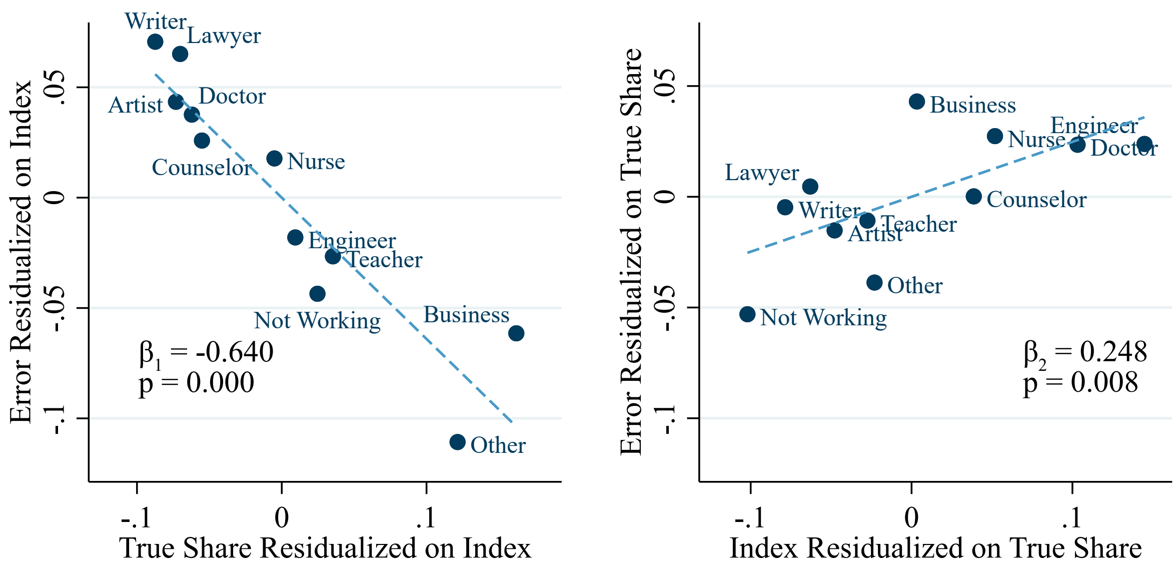
Notes: The x-axis of Figure 3 shows, for each of the 20 most common majors in the American Community Survey plus economics, the percent of graduates with that major aged 30-50 who are working in a stereotypical job for that major. We define a job as being "stereotypical" if the likelihood ratio $\frac{p_{c,m}}{p_{c,-m}}$ is greater than 10. The y-axis shows the average salary of such graduates conditional on *not* having such a stereotypical job. These statistics are calculated from the 2017-2019 American Community Survey.

Figure 4: Neglecting Common, Non-Stereotypical Outcomes



Notes: Each dot in Figure 4 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.

Figure 5: Overestimating Rare Careers and Careers That Are Concentrated within Majors



Notes: Figure 5 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 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).

Table 1: 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 1 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.

Table 2: Testing for Stereotypical Thinking

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

Notes: Table 2 presents OLS estimates of equation 1 using the 2020 OSU sample (columns 1-7) and a sample of respondents from Amazon Mechanical Turk (columns 8-9). The dependent variable is respondents' beliefs about the likelihood of a career conditional on a major. All regressions include all majors that respondents were asked about. "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 Mechanical Turk sample by whether the respondent has a four-year college degree.

Table 3: Long-Run Outcomes for Graduates with Risky vs Non-Risky Majors

	(1) Not Employed	(2) Dissatisfied	(3) Unrelated Job	(4) RJ Not Available	(5) Salary ($1ks)	(6) Debt ($1ks)	(7) Degree Not Worth Cost	(8) Regret
Risky Major	0.042	0.020	0.100	0.036	-17.918	2.904	0.135	0.128
	(0.009)	(0.009)	(0.013)	(0.007)	(1.856)	(0.628)	(0.028)	(0.032)
Constant	0.107	0.116	0.415	0.055	80.408	10.445	0.172	0.368
	(0.004)	(0.005)	(0.007)	(0.003)	(0.967)	(0.322)	(0.011)	(0.014)
N	44,498	40,574	40,574	40,574	40,499	44,498	1,633	1,633

Notes: Table 3 presents OLS regressions using the 2013 round of the National Survey of College Graduates (columns 1-6) and the 2021 Survey of Household Economics and Decisionmaking (column 7). "Risky" is a dummy variable indicating whether the graduate majored in a risky major, following the definition in section 4.4. The risky majors are Humanities, Psychology, Art, Communications, and Social/Behavioral Sciences. The dependent variables in columns (1) through (6) correspond, respectively, to: whether the graduate is not employed; whether they express that they are very or somewhat dissatisfied with their job; whether they report that their job is not closely related to their highest degree; whether they report they have an unrelated job because related jobs are unavailable; their annual salary; and their total student debt balance. The dependent variables columns 7 and 8 are, respectively, whether the respondent said the financial costs of their Bachelor's degree were somewhat or much larger than the benefits, and whether they would have chosen a different field of study if they could go back and make decisions regarding their education again. All data are restricted to college graduates between the ages of 30 and 50. Robust standard errors in parentheses.

Table 4: The Effect of an Information Intervention

	All Students		Risky Top Major		Non-Risky Top Major	
	Intentions (1)	Classes (2)	Intentions (3)	Classes (4)	Intentions (5)	Classes (6)
Overestimated Stereotypical Career	0.005	0.749	-0.004	1.111	0.007	0.740
	(0.012)	(0.212)	(0.027)	(0.382)	(0.014)	(0.236)
Treatment X Overestimated	-0.036	-0.286	-0.075	-0.292	-0.021	-0.293
	(0.011)	(0.197)	(0.018)	(0.392)	(0.013)	(0.229)
Treatment X Underestimated	-0.029	0.224	0.006	1.519	-0.032	0.082
	(0.016)	(0.238)	(0.056)	(0.772)	(0.017)	(0.252)
Pre-Treatment Belief P(M)	0.962		0.973		0.952	
	(0.019)		(0.038)		(0.022)	
Pre-Treatment Classes		0.712		0.509		0.768
		(0.099)		(0.190)		(0.113)
p-value: Over vs Under	0.7198	0.0995	0.1751	0.0380	0.6182	0.2705
Observations	814	637	168	135	646	502
Control Group Mean	0.653	2.436	0.563	2.378	0.677	2.453

Notes: Table 4 presents OLS regressions including data from the 2021 OSU sample. The dependent variable in columns 1, 3, and 5 is students' post-intervention belief about the percent chance that they will graduate with their top-ranked major. The dependent variable in columns 2, 4, and 6 is the number of classes students took in Spring 2022 plus the number they signed up to take in Fall 2022 in their top-ranked major. Columns 1 and 2 include all participants. Columns 3 and 4 restrict the data to those with "risky" top-ranked majors as defined in section 4.1: that is, humanities, psychology, communications, or art. Columns 5 and 6 include students with all other majors. "Pre-Treatment P(M)" is students' beliefs immediately pre-treatment about the percent chance they would gradaute with their top-ranked major. "Pre-Treatment Classes" is the number of classes in their top ranked major that they took during Fall 2021. "Treatment" is a dummy variable indicating whether the student was randomized into seeing the information module. "Overestimated Stereotypical Career" is a dummy variable indicating whether students' population belief about the fraction of graduates with their top major's stereotypical career was higher than the true fraction. "p-value: Over vs Under" shows the p-value testing the hypothesis that treatment effects are the same for students who overestimated vs underestimated their top-ranked major's stereotypical outcome. Robust standard errors in parentheses.

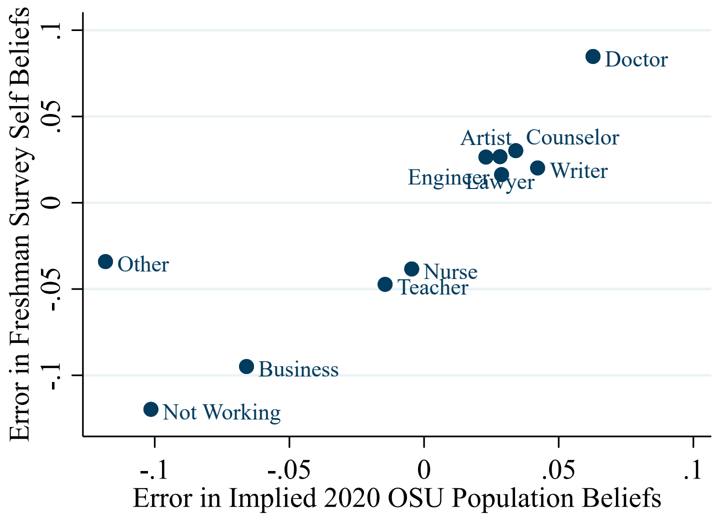
Table 5: Role Models and What Comes to Mind

	Population Beliefs				Self Beliefs				
	$P(c \mid M)$			$P(c)$	$P(c \mid M)$			$P(c)$	
	All	S	NS		All	S	NS	Ohio State	Freshman
	(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)
Observations	19,668	1,788	17,880	9,834	19,668	1,788	17,880	9,834	107,752,344
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 5 presents OLS estimates of equation 15 (columns 1-3 and 5-7) and equation 16 (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.

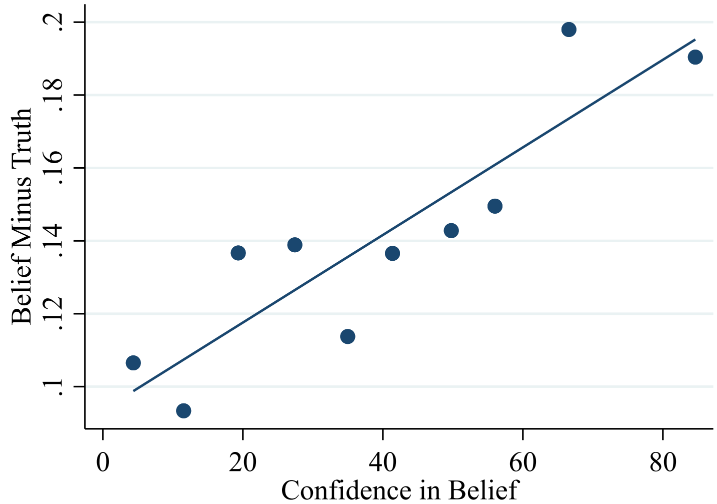
A Supplementary Figures and Tables

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



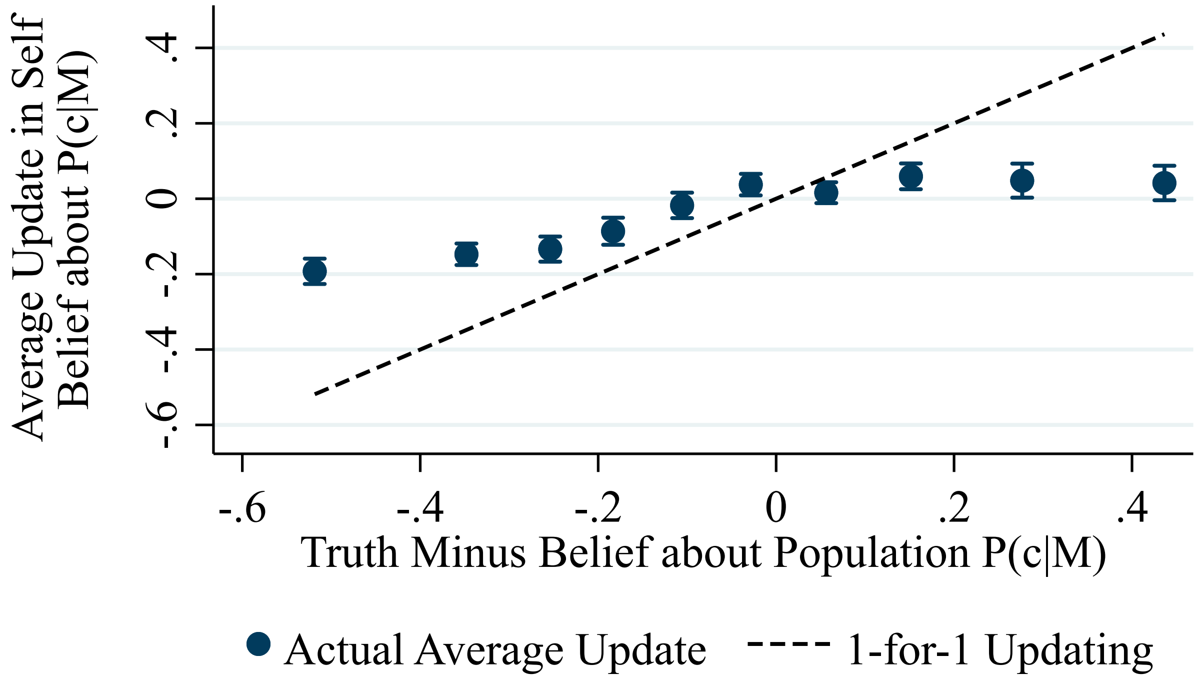
Notes: The y-axis in Figure A.I is the difference between the fraction of students in the Freshman Survey who list each occupation as their probable career and the fraction of 33-37 year old college graduates in the CPS (of the same cohorts, up to birth year 1987) who are working in each occupation. 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 3.3 for further details on the construction of this statistic.

Figure A.II: More Confident Students' Stereotype More



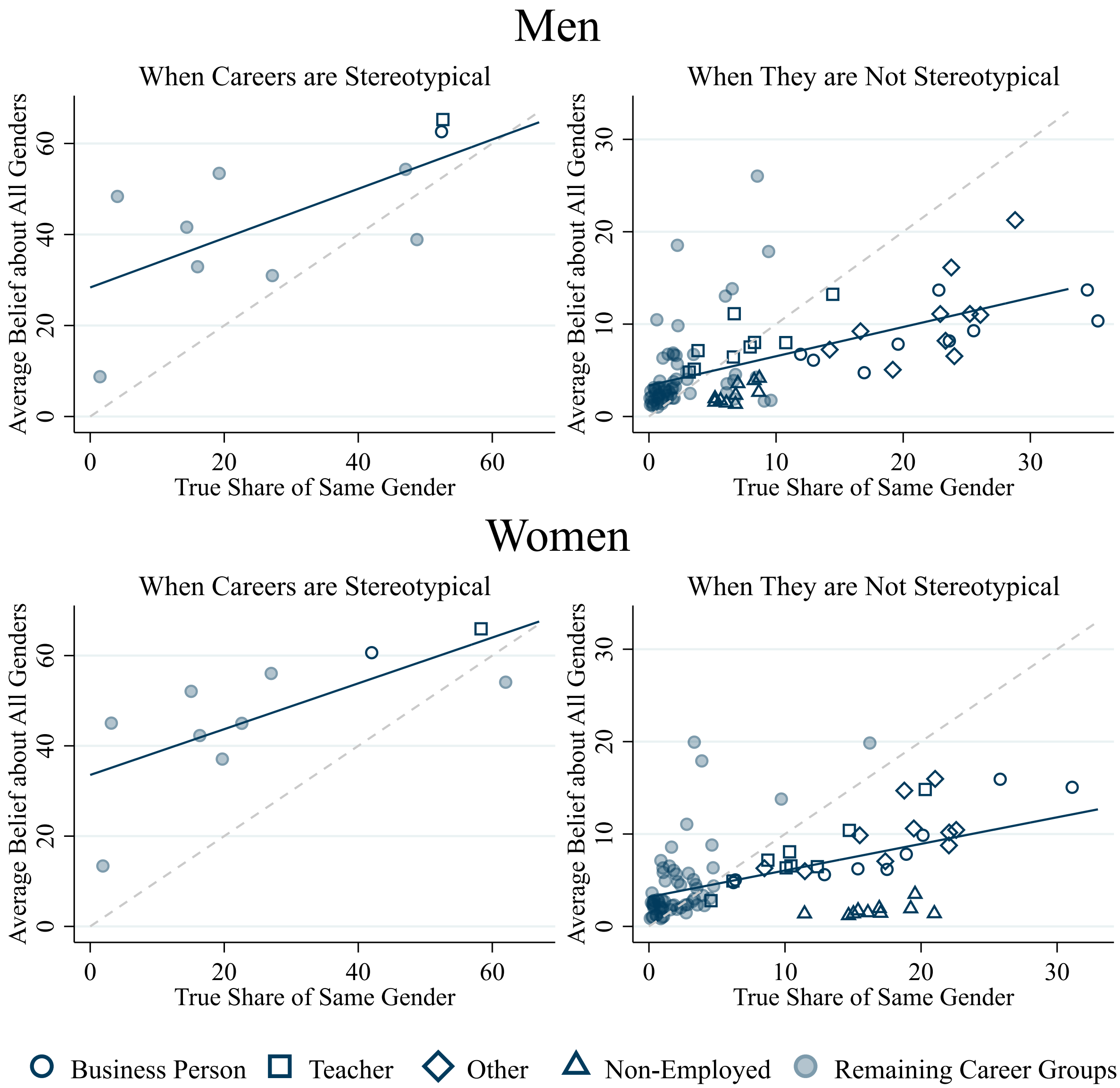
Notes: The y-axis in Figure A.I is the difference between the 2021 OSU students beliefs about the fraction of graduates (in the two majors they were asked about) working in their major's stereotypical career. The x-axis is how confident (on a 0 to 100 scale) they reported being in their answer. Data are binned into deciles of confidence. Line shows OLS regression line.

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



Notes: The x-axis of Figure A.III is the true share of college graduates working in the stereotypical career of students' top-ranked major minus the 2021 OSU sample's beliefs about that share, restricting data to the treatment group. The y-axis is the revision in such students' beliefs about their own chance of working in that profession, from before to after the information intervention. Dots show average values binned by decile of the x-axis variable. The dotted line shows the line $y = x$, the relationship that would hold if students updated "one-for-one" in response to the information (i.e., decreasing their belief about their own lileihood of having the career by one percentage point for every percentage point they overestimated its frequency among others.)

Figure A.IV: Beliefs vs Gender-Specific Outcomes



Notes: Each dot in Figure A.IV represents a career-major pair (where non-employment is also one the "careers"). The left panels restrict these pairs to when the career is most stereotypical of the major, and the right panels restrict them to when the career is not stereotypical of the major. The top panel restricts the 2020 OSU data to men, while the bottom panel includes only women. The x-axis of both panels is the share of graduates *of the same gender as the respondent* 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. Note that survey questions asked students about graduates of all majors, not just of their own majors. Lines show OLS regressions including all career-major pairs within each panel.

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

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

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

Table A.IV: Career Expectations Among College First-Years in the U.S.

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

Notes: Table A.IV shows the distribution of career 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 would fall into each group. "Outcomes" 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. p-value is from a t-test for whether the shares are equal across columns.

Table A.V: Major Expectations Among College First-Years in the U.S.

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 college 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" field of study would fall into each group. "Outcomes" 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' top-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 20 and 21 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 20 and only the 2021 OSU sample for equation 21.

$$\begin{aligned} Y_{i,c,m,p} - T_{c,m} = \psi_{c,p,Top(i,m)} + \beta_1 Self_{i,p} + \beta_2 t_{i,m} + \beta_3 Self_{i,p} \times Top_{i,m} + \beta_4 Ster_{m,c} + \beta_5 Ster_{m,c} \times Self_{i,p} + \\ \beta_6 Ster_{m,c} \times Top_{i,m} + \beta_7 Ster_{m,c} \times Self_{i,p} \times Top_{i,m} + \beta_8 T_{c,m} + \beta_9 T_{c,m} \times Self_{i,p} + \beta_{10} T_{c,m} \times Top_{i,m} \\ + \beta_{11} T_{c,m} \times Self_{i,p} \times Top_{i,m} + \varepsilon_{i,c,m,p} \end{aligned} \tag{20}$$

$$\begin{aligned} Y_{i,c,m,p} - T_{c,m} = \psi_{c,p} + \beta_1 Self_{i,p} + \beta_2 t_{i,m} + \beta_4 Ster_{m,c} + \beta_5 Ster_{m,c} \times Self_{i,p} + \beta_8 T_{c,m} + \beta_9 T_{c,m} \times Self_{i,p} + \\ \beta_{1,2} RM_{i,c,m} + \beta_{1,3} RM_{i,c,m} \times Self_{i,p} + \varepsilon_{i,c,m,p} \end{aligned} \tag{21}$$

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

Notes: Table A.IX shows parameter estimates of the model described in Section 4.1 as well as the variants thereof described B.1.4. The model is estimated by maximum likelihood, and standard errors/confidence intervals are constructed by Bayesian We omit estimates of μ_M and $c^*(i)$ for readability.

Table A.X: Stereotypical and Unstereotypical jobs

Full Major Name	Short Name	Stereotypucal Occupations (up to 5 most common)	Unstereotypucal Occupations (up to 5 most common)
Agriculture	Agriculture	farmers, ranchers, and other agricultural managers (8.0%); veterinarians (3.9%); agricultural workers, nec (2.1%); agricultural and food scientists (1.8%); first-line supervisors of landscaping, lawn service, and groundskeeping workers (1.4%)	managers, nec (including postmasters) (6.0%); sales representatives, wholesale and manufacturing (3.5%); first-line supervisors of sales workers (3.2%); elementary and middle school teachers (2.9%); postsecondary teachers (2.1%)
Biology And Life Sciences	Biology	physicians and surgeons (17.0%); physical scientists, nec (3.5%); dentists (2.8%); medical scientists, and life scientists, all other (2.4%); biological scientists (1.8%)	managers, nec (including postmasters) (4.6%); postsecondary teachers (3.9%); registered nurses (3.9%); elementary and middle school teachers (2.9%); pharmacists (2.5%)
Business	Business	accountants and auditors (9.3%)	managers, nec (including postmasters) (6.9%); first-line supervisors of sales workers (4.4%); financial managers (4.3%); managers in marketing, advertising, and public relations (2.8%); chief executives and legislators/public administration (2.5%)
Communications	Communications	editors, news analysts, reporters, and correspondents (2.3%); actors, producers, and directors (1.8%); public relations specialists (1.5%); announcers (0.3%)	managers, nec (including postmasters) (6.2%); managers in marketing, advertising, and public relations (5.9%); elementary and middle school teachers (3.7%); first-line supervisors of sales workers (3.1%); human resources, training, and labor relations specialists (2.3%)
Computer And Information Sciences	Comp. Sci.	software developers, applications and systems software (21.8%); computer programmers (5.0%); computer hardware engineers (0.5%)	computer scientists and systems analysts/network systems analysts/web developers (12.7%); managers, nec (including postmasters) (5.8%); computer and information systems managers (5.1%); computer support specialists (3.7%); first-line supervisors of sales workers (2.0%)
Criminal Justice And Fire Protection	Crim. Just.	police officers and detectives (14.6%); first-line supervisors of police and detectives (2.1%); private detectives and investigators (0.9%); first-line supervisors of correctional officers (0.6%); supervisors, protective service workers, all other (0.6%)	social workers (7.2%); managers, nec (including postmasters) (4.1%); lawyers, and judges, magistrates, and other judicial workers (2.9%); elementary and middle school teachers (2.4%); first-line supervisors of sales workers (2.2%)
Economics	Economics	economists and market researchers (1.3%)	managers, nec (including postmasters) (7.7%); lawyers, and judges, magistrates, and other judicial workers (5.8%); financial managers (5.1%); accountants and auditors (4.0%); chief executives and legislators/public administration (3.5%)
Education Administration And Teaching	Education	elementary and middle school teachers (40.2%); secondary school teachers (7.9%); special education teachers (3.2%)	education administrators (5.7%); preschool and kindergarten teachers (3.2%); postsecondary teachers (1.9%); counselors (1.8%); other teachers and instructors (1.6%)
Engineering	Engineering	engineers, nec (6.9%); civil engineers (5.3%); mechanical engineers (3.7%); electrical and electronics engineers (2.9%); architectural and engineering managers (2.3%)	software developers, applications and systems software (9.9%); managers, nec (including postmasters) (8.8%); computer scientists and systems analysts/network systems analysts/web developers (3.6%); computer and information systems managers (2.1%); postsecondary teachers (1.9%)
English Language, Literature, And Composition	English	editors, news analysts, reporters, and correspondents (2.2%); writers and authors (2.1%); technical writers (0.5%)	elementary and middle school teachers (8.5%); postsecondary teachers (5.5%); lawyers, and judges, magistrates, and other judicial workers (5.3%); managers, nec (including postmasters) (4.6%); secondary school teachers (3.7%)

Notes: Tables A.X and A.XI give, for each of the 20 largest majors in the American Community Survey plus economics, up to 5 of the most common stereotypical and unstereotypical occupations. We define a job as being "stereotypical" if the likelihood ratio $\frac{p_{c,m}}{p_{c,-m}}$ is greater than 10.

Table A.XI: Stereotypical and unstereotypical jobs (continued)

Full Major Name	Short Name	Stereotypucal Occupations (up to 5 most common)	Unstereotypucal Occupations (up to 5 most common)
Fine Arts	Fine Arts	designers (11.0%); artists and related workers (2.9%); musicians, singers, and related workers (1.8%); archivists, curators, and museum technicians (0.5%); dancers and choreographers (0.1%)	elementary and middle school teachers (4.9%); managers, nec (including postmasters) (4.9%); postsecondary teachers (3.5%); first-line supervisors of sales workers (2.8%); other teachers and instructors (2.4%)
History	History	archivists, curators, and museum technicians (0.6%)	lawyers, and judges, magistrates, and other judicial workers (9.4%); elementary and middle school teachers (7.5%); managers, nec (including postmasters) (5.5%); secondary school teachers (3.8%); postsecondary teachers (3.5%)
Liberal Arts And Humanities	Liberal Arts	None	elementary and middle school teachers (11.4%); managers, nec (including postmasters) (4.2%); first-line supervisors of sales workers (2.9%); education administrators (2.0%); secretaries and administrative assistants (1.9%)
Mathematics And Statistics	Math	mathematical science occupations, nec (3.1%); actuaries (1.9%)	postsecondary teachers (7.7%); software developers, applications and systems software (7.3%); elementary and middle school teachers (7.1%); secondary school teachers (5.9%); managers, nec (including postmasters) (5.2%)
Medical And Health Sciences And Services	Health Services	registered nurses (41.3%); pharmacists (3.8%); physical therapists (3.0%); speech language pathologists (2.6%); occupational therapists (1.6%)	medical and health services managers (3.5%); physicians and surgeons (2.7%); managers, nec (including postmasters) (1.8%); elementary and middle school teachers (1.5%); postsecondary teachers (1.4%)
Physical Fitness, Parks, Recreation, And Leisure	Recreation	physical therapists (6.0%); recreation and fitness workers (3.6%); athletes, coaches, umpires, and related workers (2.0%); chiropractors (0.7%); recreational therapists (0.1%)	elementary and middle school teachers (6.3%); managers, nec (including postmasters) (4.7%); registered nurses (3.3%); secondary school teachers (2.9%); first-line supervisors of sales workers (2.5%)
Physical Sciences	Physics	physical scientists, nec (4.4%); chemists and materials scientists (2.9%); environmental scientists and geoscientists (2.1%); atmospheric and space scientists (0.4%); astronomers and physicists (0.4%)	physicians and surgeons (5.8%); managers, nec (including postmasters) (5.8%); postsecondary teachers (5.4%); software developers, applications and systems software (3.2%); elementary and middle school teachers (3.1%)
Psychology	Psychology	counselors (7.4%); psychologists (4.7%); therapists, nec (1.6%)	social workers (6.0%); elementary and middle school teachers (5.0%); managers, nec (including postmasters) (4.3%); registered nurses (2.6%); postsecondary teachers (2.5%)
Public Affairs, Policy, And Social Work	Public Policy	social workers (27.5%); therapists, nec (2.2%)	counselors (5.8%); managers, nec (including postmasters) (3.6%); elementary and middle school teachers (3.3%); social and community service managers (3.0%); secretaries and administrative assistants (1.8%)
Social Sciences	Social Science	social scientists, nec (0.4%)	lawyers, and judges, magistrates, and other judicial workers (9.7%); managers, nec (including postmasters) (6.3%); elementary and middle school teachers (4.2%); social workers (3.6%); postsecondary teachers (2.5%)

Notes: Tables A.X and A.XI give, for each of the 20 largest majors in the American Community Survey plus economics, up to 5 of the most common stereotypical and unstereotypical occupations. We define a job as being "stereotypical" if the likelihood ratio $\frac{p_{c,m}}{p_{c,-m}}$ is greater than 10.

Table A.XII: The Effect of an Information Nudge: Majors with Rare vs Common Stereotypes

	All Students		"Rare" Top Major		"Common" Top Major	
	Intentions (1)	Classes (2)	Intentions (3)	Classes (4)	Intentions (5)	Classes (6)
Overestimated Stereotypical Career	0.005	0.749	-0.006	1.022	0.010	0.765
	(0.012)	(0.212)	(0.022)	(0.342)	(0.014)	(0.244)
Treatment X Overestimated	-0.036	-0.286	-0.053	-0.348	-0.027	-0.256
	(0.011)	(0.197)	(0.018)	(0.351)	(0.014)	(0.239)
Treatment X Underestimated	-0.029	0.224	0.004	1.410	-0.032	0.041
	(0.016)	(0.238)	(0.046)	(0.667)	(0.018)	(0.259)
Pre-Treatment Belief P(M)	0.962		0.985		0.944	
	(0.019)		(0.033)		(0.023)	
Pre-Treatment Classes		0.712		0.604		0.742
		(0.099)		(0.177)		(0.116)
p-value: Over vs Under	0.7198	0.0995	0.2430	0.0207	0.8126	0.4000
Observations	814	637	196	161	618	476
Control Group Mean	0.653	2.436	0.541	2.230	0.689	2.510

Notes: Table A.XIII presents OLS regressions including data from the 2021 OSU sample. The dependent variable in columns 1, 3, and 5 is students' post-intervention belief about the percent chance that they will graduate with their top-ranked major. The dependent variable in columns 2, 4, and 6 is the number of classes students took in Spring 2022 plus the number they signed up to take in Fall 2022 in their top-ranked major. Columns 1 and 2 include all participants. Columns 3 and 4 restrict the data to those whose top-ranked major's stereotypical outcome is rare: that is, humanities, psychology, communications, art, government/political scheince, and biology/chemistry. Columns 5 and 6 include students with all other majors. "Pre-Treatment P(M)" is students' beliefs immediately pre-treatment about the percent chance they would gradaute with their top-ranked major. "Pre-Treatment Classes" is the number of classes in their top ranked major that they took during Fall 2021. "Treatment" is a dummy variable indicating whether the student was randomized into seeing the information module. "Overestimated Stereotypical Career" is a dummy variable indicating whether students' population belief about the fraction of graduates with their top major's stereotypical career was higher than the true fraction. "p-value: Over vs Under" shows the p-value testing the hypothesis that treatment effects are the same for students who overestimated vs underestimated their top-ranked major's stereotypical outcome. Robust standard errors in parentheses.

Table A.XIII: The Effect of an Information Nudge: Low- vs High-Paying Majors

	All Students		Lower-Paying Top Major		Higher-Paying Top Major	
	Intentions (1)	Classes (2)	Intentions (3)	Classes (4)	Intentions (5)	Classes (6)
Overestimated Stereotypical Career	0.005	0.749	0.007	0.802	0.003	0.668
	(0.012)	(0.212)	(0.024)	(0.268)	(0.012)	(0.310)
Treatment X Overestimated	-0.036	-0.286	-0.048	-0.147	-0.027	-0.384
	(0.011)	(0.197)	(0.018)	(0.308)	(0.014)	(0.255)
Treatment X Underestimated	-0.029	0.224	0.007	0.563	-0.051	-0.064
	(0.016)	(0.238)	(0.029)	(0.296)	(0.019)	(0.354)
Pre-Treatment Belief P(M)	0.962		0.956		0.957	
	(0.019)		(0.038)		(0.022)	
Pre-Treatment Classes		0.712		0.575		0.707
		(0.099)		(0.159)		(0.132)
p-value: Over vs Under	0.7198	0.0995	0.1087	0.0974	0.3295	0.4636
Observations	814	637	320	260	494	377
Control Group Mean	0.653	2.436	0.596	1.912	0.690	2.807

Notes: Table A.XIII presents OLS regressions including data from the 2021 OSU sample. The dependent variable in columns 1, 3, and 5 is students' post-intervention belief about the percent chance that they will graduate with their top-ranked major. The dependent variable in columns 2, 4, and 6 is the number of classes students took in Spring 2022 plus the number they signed up to take in Fall 2022 in their top-ranked major. Columns 1 and 2 include all participants. Columns 3 and 4 restrict the data to those whose top-ranked major has low-paying alternatives to its stereotypical job: that is, humanities, psychology, communications, art, nursing, and education. Columns 5 and 6 include students with all other majors. "Pre-Treatment P(M)" is students' beliefs immediately pre-treatment about the percent chance they would gradaute with their top-ranked major. "Pre-Treatment Classes" is the number of classes in their top ranked major that they took during Fall 2021. "Treatment" is a dummy variable indicating whether the student was randomized into seeing the information module. "Overestimated Stereotypical Career" is a dummy variable indicating whether students' population belief about the fraction of graduates with their top major's stereotypical career was higher than the true fraction. "p-value: Over vs Under" shows the p-value testing the hypothesis that treatment effects are the same for students who overestimated vs underestimated their top-ranked major's stereotypical outcome. Robust standard errors in parentheses.

Table A.XIV: 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.XIV 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 Data Appendix

The section describes the datasources 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.[39]

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.[40] 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".[41] However, in open-ended feedback the overwhelming reason given was that they took longer to fill out than multiple choice questions would have.[42] In addition, all of the

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

[40] 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.

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

[42] 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

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

To elicit students' salary beliefs, the 2020 OSU survey asked students about the distribution of salaries by major. The population version of this question asked students their "best guess about the percent of Americans aged 30-50 who graduated with a major in X who are working full time that 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. The self-beliefs question asked students, "Again imagine that ***you*** successfully graduate from OSU with a major in X. Also imagine that you are working full-time. What is your best guess about the percent chance that, when you are 30 years old, you would earn an annual salary of..." The same six ranges of salaries were then listed. In both questions, students' answers had to add up to 100%.

In a similar vein, students were asked about the distribution of salaries by occupation, using the analogous language as the elicitation for majors. Each student reported expectations for four careers: the two occupations students said they were most likely to enter, and two additional randomly chosen occupations from the remaining seven. We again use inverse probability weights to account for this design when calculating average 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.[43] Any differences between beliefs about and the actual distribution of salaries therefore cannot be driven by (unobserved) beliefs about the distribution of salaries within the listed ranges.

B.1.2 Comparing OSU to Other Colleges

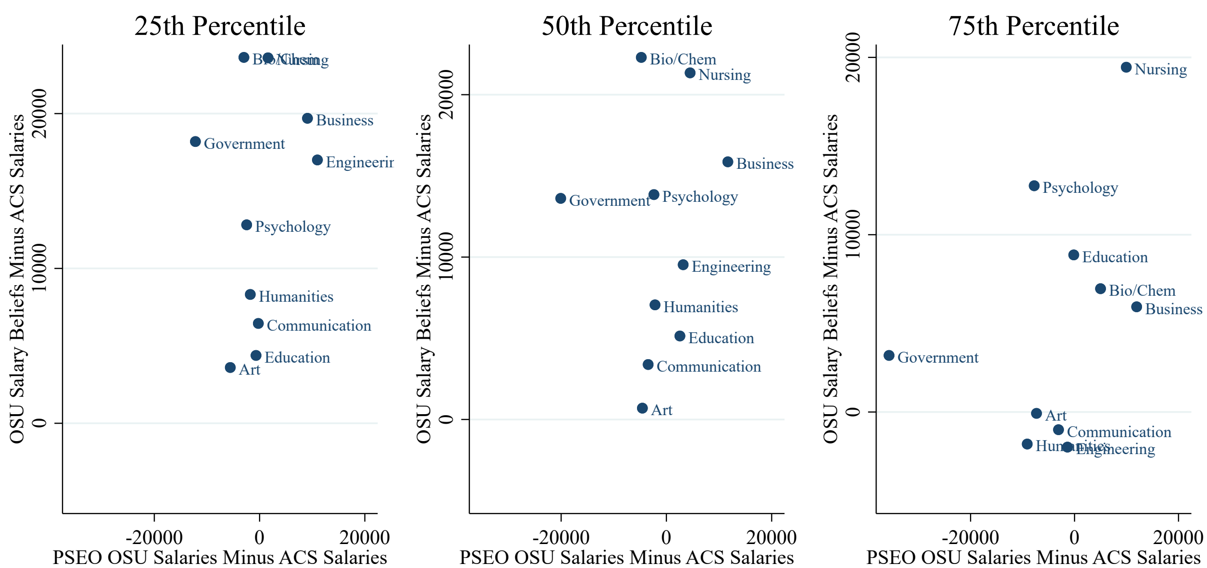
The similarities between the quantitative beliefs from the OSU sample and the qualitiative beliefs from the nationally representative sample suggest that the specific context of our surveys are not driving the results. Nevertheless, our population beliefs questions asked students about the occupations of college graduates throughout the country, and one might worry that graduates from Ohio State may tend to have stereotypical careers much more frequently than graduates from other schools. If so, then while these students' beliefs would be "biased" in that they do not match the true national occupation distributions, they might be correct about the distribution of occupations that matters most to their choice of major. Data on occupations by major for Ohio State graduates, which would allow us to directly examine this issue, are not available. However, recently released data

student, induced more careful answers than quicker multiple-choice questions would have.

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

from the post-secondary employment outcomes (PSEO) has experimental tabulations on earnings by major for OSU graduates from the Census's Longitudinal Employer-Household Dynamics. We can compare these earnings statistics to nationally representative earnings to provide evidence (albeit only suggestive) that students' beliefs do not seem to better match outcomes from OSU graduates. To do so, we consider the 25^{th}, 50^{th}, and 75^{th} percentiles of earnings for the class of 2001, 10 years after they graduated. To compare with people of a similar age range, we calculate the comparable measure for all college graduates in the ACS using 30-35 years old.

Figure A.V: Comparing OSU to All BA Holders



Notes: This figure compares the 25^{th}, 50^{th}, and 75^{th} percentiles of earnings by major. The y-axis in each panel plots the difference between the earnings expectations by major of first-years in the 2020 OSU sample and the average salaries by major among all BA holders in the 2017-2019 American Community Surveys between the age of 30 and 35. Note that students in the OSU survey were asked about 30- to 50-year-olds, which may explain why this difference tends to be positive. The x-axis plots differences in earnings among Ohio State University graduates 10 years after receiving their Bachelor's in 2001 from the Post-Secondary Employment Outcomes data and the average salary for all BA holders in the 2017-2019 American Community Surveys between the age of 30 and 35.

The 2020 OSU survey asked students the beliefs about the distribution of salaries by major, allowing us to compute average beliefs about the 25^{th}, 50^{th}, and 75^{th} percentiles of earnings. Comparing these measures with the corresponding PSEO statistics allows us to test whether differences in earnings between OSU graduates and BA-holders overall can predict the errors in students' salary expectations. Figure A.V plots the results of this exercise, comparing errors in students expectations to differences in true salaries between OSU alumni and college graduates in the ACS.[44] Across each quantile, we find no evidence that the mistakes students make about earnings expectations are correlated with the how OSU graduates' salaries differ from the national distribution. This suggests that employment outcomes for OSU graduates likely do not differ from national averages in such a way that would explain the biases in our OSU sample's beliefs.

B.1.3 Salary Beliefs Are Correlated with Career Beliefs but Are Roughly Unbiased

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

Nonetheless, students' beliefs about salaries are are tightly linked to their beliefs about the frequency of careers conditional on major. Denote the expected value of the elicited distribution of salaries students' "direct" salary beliefs. Column 1 of Table A.XVI regresses direct self beliefs on direct population beliefs. We see a robust positive relationship, which persists in Column 2 after adding major- and individual-fixed effects. These results show that students' earnings expectations, depending on major, are tightly linked to the salaries that they think others earn. Column 3 of Table A.XVI replaces direct population beliefs with what we call students' "implied" population beliefs. To construct these, we first take the actual mean salary from the ACS for each major-career pair. We then take a weighted average of these using students' beliefs about their likelihood of having each career conditional on major. Thus, their implied population beliefs are the average salaries that follow from their beliefs about the likelihood of careers, assuming they know the average salary of each career. We see in column 3 that implied population beliefs are very predictive of direct self beliefs, and this relationship persists in Column 4 after we add major- and individual-fixed effects. Intuitively, these results show that students expect to earn a high salary with a major when they think higher-paying careers are more common among that major's graduates.

[44] Note that students' salary beliefs in this figure appear misleadingly high because they were asked about ages 30-50 while this analysis restricts the ACS to those aged 30-35. Table A.XV compares the OSU sample's salary beliefs to those aged 30-50 in the ACS, where we do not see any such over-optimism.

Table A.XV: Beliefs about Salaries in 2020 OSU Survey

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

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

Table A.XVI: Connecting Career and Salary Beliefs

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

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

B.1.4 Model of Major Choice

Here we give more details on the model of major choice presented in Section **??**. Recall the paramaters of the baseline model 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 into a vector of coefficients 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 22:

$$\hat{\xi} = \underset{\xi}{\operatorname{argmax}} \sum_i \sum_{j=1}^{3} \log(L_{i,j}(\xi)) \tag{22}$$

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 of the baseline model (for readability, we omit estimates of μ_M). The next columns then show estimates of modifications to this baseline. We describe these modifications below, but across all of them we see large preferences for working in preferred careers: the WTP to increase a student's chance of working in her most preferred career by 1p.p. is never estimated to be below \$3,000 per year.

Model 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 99%.

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

Model 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.

Model 5. Identical to the baseline model, except we allow average non-pecuniary preferences for each career to be non-zero. That is, we add $\sum_c \zeta_c \pi^i_{c|M}$ to equation 4. 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.

Model 6. Identical to the baseline model, except we allow non-pecuniary preferences for two individual-specific "favorite" careers. That is, we add $\beta_2\left(\pi^i_{c_2^*(i)|M} - \pi^i_{c_2^*(i)|M'}\right)$ to equation 4, where $c_2^*(i)$ is i's second favorite career.

Model 7. Identical to the baseline model, except we allow non-pecuniary preferences for all nine careers. That is, we add $\sum_{r=2}^{9} \beta_r \left(\pi^i_{c^*_r(i)|M} - \pi^i_{c^*_r(i)|M'} \right)$ to equation 4, where $c^*_2(i)$ is i's second favorite career. To identify i's ranking of careers, we use the ranking they gave when asked to "Please rank the following careers by the likelihood you will hold that job when you are age 30." We note that this question asked students about the *likelihood* that they would have careers, not their *preference* for them, so caution is warranted when interpreting the estimates from this model.

Model 8. 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 $\alpha_2 \mathrm{E}[\mathrm{GPA} \mid M]$ to equation 4.

B.2 Fall 2021 Ohio State Surveys

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

B.3 Mechanical Turk Survey

In November of 2021, we recruited 706 participants on Amazon Mechanical Turk to take a short survey. Each participant was asked population beliefs questions about the frequency of careers conditional on a randomly selected two majors (the same career and major groups as we focus on throughout the paper). In addition to a $1 completion payment, participants were truthfully told that they would receive 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 2 are derived).

C Proofs

Here we provide derivations of the results in Section 5. Following the discussion in Section 5.2, 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 retrieving e. The plausibility of H is then the average ease of simulation among the people that the student retrieved. 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. (2022), 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 23, whereby ease-of-simulation decreases by a factor of $\eta_c \leq 1$ if e lacks the relevant career and by η_M if e lacks the relevant major.

$$\sigma(e, H_c) = \eta_c^{\mathbb{1}(c(e) \neq c)} \qquad \sigma(e, H_{c,M}) = \eta_c^{\mathbb{1}(c(e) \neq c)} \times \eta_M^{\mathbb{1}(M(e) \neq M)} \tag{23}$$

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 5.2, 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 the two hypotheses $H_{a,A}$ and $H_{b,A}$, which can be written as follows:

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

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

The agent's subjective odds ratio is given by the ratio of each hypothesis's plausibility. We can then take the log of this ratio:

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

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 the role of personal experiences. Then,

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

First, when $\eta_c = \eta_M = 0$, as in main specification, this derivative is unambiguously negative whenever $\delta_c < 1$.

We can then compute a first-order Taylor approximzation around the rational benchmark (i.e., around $\delta_c = \delta_M = 1$ and $\eta_c = \eta_M = 0$).

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

We can evaluate this derivative at the rational benchmark:

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

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

We can evaluate this derivative at the rational benchmark:

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

Next, we can take a similar derivative with respect to η_c:

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

We can evaluate this derivative at the rational benchmark:

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

Next, we can take a similar derivative with respect to η_M:

$$\frac{\partial^2}{\partial\alpha\partial\eta_M}\left[\log\frac{\pi_{a|A}}{\pi_{b|A}}\right] = p_B\frac{\delta_M-\delta_c\delta_M\eta_c}{p_{a,A}+\delta_c\eta_c p_{b,A}+\delta_c\eta_c\sum_{z\notin\{a,b\}}p_{z,A}+\delta_M\eta_M p_{a,B}+\delta_c\delta_M\eta_c\eta_M p_{b,B}+\delta_c\delta_M\eta_c\eta_M\sum_{z\notin\{a,b\}}p_z}$$
$$-\,p_B\frac{(\delta_M\eta_M-\delta_c\delta_M\eta_c\eta_M)(\delta_M p_{a,B}+\delta_c\delta_M\eta_c p_{b,B}+\delta_c\delta_M\eta_c\sum_{z\notin\{a,b\}}p_{z,B})}{(p_{a,A}+\delta_c\eta_c p_{b,A}+\delta_c\eta_c\sum_{z\notin\{a,b\}}p_{z,A}+\delta_M\eta_M p_{a,B}+\delta_c\delta_M\eta_c\eta_M p_{b,B}+\delta_c\delta_M\eta_c\eta_M\sum_{z\notin\{a,b\}}p}$$

We can evaluate this derivative at the rational benchmark:

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

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

$$\frac{\partial}{\partial\alpha}\left[\log\frac{\pi_{a|A}}{\pi_{b|A}}\right] \approx p_B\left(\eta_M\frac{1}{p_{a,A}}-(1-\delta_c)\right)$$

which is the desired expression.

Next, to yield Prediction 3, we ask when the agent's subjective odds ratio $\frac{\pi_{a|A}}{\pi_{b|A}}$ is higher than the true odds ratio $\frac{p_{a,A}}{p_{b,A}}$, letting $p_{z,B}=0$ for all careers z and setting $\phi=0$:

$$
\begin{aligned}
&\frac{\pi_{a|A}}{\pi_{b|A}} > \frac{p_{a,A}}{p_{b,A}} \iff \frac{F(H_{a|A})}{F(H_{b|A})} > \frac{p_{a,A}}{p_{b,A}} \\
\iff\ & p_{b,A}\frac{p_{a,A} + \delta_c\eta_c p_{b,A} + \delta\eta_c \sum_{z\notin\{a,b\}} p_{z,A}}{p_{a,A} + \delta_c p_{b,A} + \delta_c \sum_{z\notin\{a,b\}} p_{z,A}} > p_{a,A}\frac{\delta_c\eta_c p_{a,A} + p_{b,A} + \delta_c\eta_c \sum_{z\notin\{a,b\}} p_{z,A}}{\delta_c p_{a,A} + p_{b,A} + \delta_c \sum_{z\notin\{a,b\}} p_{z,A}} \\
\iff\ & p_{b,A}\Bigg(\delta_c p_{a,A}^2 + \delta_c^2\eta_c p_{a,A}p_{b,A} + \delta_c^2\eta_c \sum_{z\notin\{a,b\}} p_{z,A} + p_{a,A}p_{b,A} + \delta_c\eta_c p_{b,A}^2 \\
&+\delta_c\eta_c p_{b,A} \sum_{z\notin\{a,b\}} p_{z,A} + \delta_c p_{a,A} \sum_{z\notin\{a,b\}} p_{z,A} + \delta_c^2\eta_c p_{b,A} \sum_{z\notin\{a,b\}} p_{z,A} + \delta_c^2\eta_c \Big(\sum_{z\notin\{a,b\}} p_{z,A}\Big)^2\Bigg) \\
&> p_{a,A}\Bigg(\delta_c\eta_c p_{a,A}^2 + \delta_c^2\eta_c p_{a,A}p_{b,A} + \delta_c^2\eta_c \sum_{z\notin\{a,b\}} p_{z,A} + p_{a,A}p_{b,A} + \delta_c p_{b,A}^2 \\
&+\delta_c p_{b,A} \sum_{z\notin\{a,b\}} p_{z,A} + \delta_c\eta_c p_{a,A} \sum_{z\notin\{a,b\}} p_{z,A} + \delta_c^2\eta_c p_{b,A} \sum_{z\notin\{a,b\}} p_{z,A} + \delta_c^2\eta_c \Big(\sum_{z\notin\{a,b\}} p_{z,A}\Big)^2\Bigg) \\
&\iff p_{a,A}p_{b,A}(p_{b,A} - p_{a,A})(1 + \delta_c^2\eta_c - \delta_c) + \delta_c\eta_c(p_{b,A}^3 - p_{a,A}^3) \\
&+(p_{b,A} - p_{a,A})\delta_c^2\eta_c \sum_{z\notin\{a,b\}} p_{z,A} + \delta_c\eta_c(p_{b,A}^2 - p_{a,A}^2) \sum_{z\notin\{a,b\}} p_{z,A} + (p_{b,A} - p_{a,A})\delta_c^2\eta_c p_{b,A} \sum_{z\notin\{a,b\}} p_{z,A} \\
&+ (p_{b,A} - p_{a,A})\Bigg(\delta_c^2\eta_c\Big(\sum_{z\notin\{a,b\}} p_{z,A}\Big)^2\Bigg) > 0 \\
&\iff p_{b,A} > p_{a,A}
\end{aligned}
$$

So the agent relatively overestimates the rarer career (conditional on major).

To yield Prediction 4, let $\Big(c(x), m(x)\Big) = (a, A)$. Then

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

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

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

Note that the first term is larger in magnitude than the second term. The third term is smaller than the fourth term (guaranteeing that the whole derivative is negative) 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 negative 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 5, let $\Big(c(x), m(x)\Big) = (a, B)$. Then

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

Then,

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

Note that when $\eta_c = \eta_M = 0$, this derivative is unambiguously negative. To investigate this prediction allowing for simulation, 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} \\ &+ \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} \\ &+ \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,}{(\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} \\ &+ \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:

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

Then,

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

We can evaluate this derivative at the rational benchmark:

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

Then,

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

We can evaluate this derivative at the rational benchmark:

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

Then,

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

We can evaluate this derivative at the rational benchmark:

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

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

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

which is the desired expression.

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,

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

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

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

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

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

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

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

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

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

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

First, we derive Prediction 6:

$$\frac{\partial}{\partial\phi}\log F(H_a) = \frac{2(p_{a,A}+\phi+\delta_M p_{a,B}) + \eta_c\Big(\delta_c p_{b,A}+\delta_c\sum_{z\notin\{a,b\}} p_{z,A}+\delta_c\delta_M p_{b,B}+\delta_c\delta_M\sum_{z\notin\{a,b\}} p_{z,B}\Big)}{\begin{array}{l}(p_{a,A}+\phi)(p_{a,A}+\delta_M p_{a,B}+\phi)+p_{a,B}(\delta_M p_{a,A}+p_{a,B}+\delta_M\phi)\\ \quad +\eta_c\Big(p_{b,A}(\delta_c p_{a,A}+\delta_c\delta_M p_{a,B}+\delta_c\phi)+\sum_{z\notin\{a,b\}} p_{z,A}(\delta_c p_{a,A}+\delta_c\delta_M p_{a,B}+\delta_c\phi)\\ \qquad +p_{b,B}(\delta_c\delta_M p_{a,A}+\delta_c p_{a,B}+\delta_c\delta_M\phi)+\sum_{z\notin\{a,b\}} p_{z,B}(\delta_c\delta_M p_{a,A}+\delta_c p_{a,B}+\delta_c\delta_M\phi)\Big)\end{array}}$$

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

Let $X = (p_{a,A}+\phi+\delta_M p_{a,B})$, $Y = \Big(\delta_c p_{b,A}+\delta_c\sum_{z\notin\{a,b\}} p_{z,A}+\delta_c\delta_M p_{b,B}+\delta_c\delta_M\sum_{z\notin\{a,b\}} p_{z,B}\Big)$, and $Z = \Big(p_{b,A}(\delta_c p_{a,A}+\delta_c\delta_M p_{a,B}+\delta_c\phi)+\sum_{z\notin\{a,b\}} p_{z,A}(\delta_c p_{a,A}+\delta_c\delta_M p_{a,B}+\delta_c\phi)+p_{b,B}(\delta_c\delta_M p_{a,A}+\delta_c p_{a,B}+\delta_c\delta_M\phi)+\sum_{z\notin\{a,b\}} p_{z,B}(\delta_c\delta_M p_{a,A}+\delta_c p_{a,B}+\delta_c\delta_M\phi)\Big)$. Then,

$$
\begin{aligned}
\frac{\partial}{\partial \phi} \log F(H_a) &= \frac{2X + \eta_c Y}{(p_{a,A} + \phi)X + \eta_c Z} - \frac{2X + Y}{(p_{a,A} + \phi)X + Z} > 0 \\
&\iff 2Z > +(p_{a,A} + \phi)Y \\
&\iff 2\Big(\big[p_{b,A} + \sum_{z \notin \{a,b\}} p_{z,A}\big](p_{a,A} + \delta_M p_{a,B} + \phi) + \big[p_{b,B} + \sum_{z \notin \{a,b\}} p_{z,B}\big](\delta_M p_{a,A} + p_{a,B} + \delta_M \phi)\Big) \\
&\quad > (p_{a,A} + \phi)\Big(p_{b,A} + \sum_{z \notin \{a,b\}} p_{z,A} + \delta_M p_{b,B} + \delta_M \sum_{z \notin \{a,b\}} p_{z,B}\Big) \\
&\iff \Big(\big[p_{b,A} + \sum_{z \notin \{a,b\}} p_{z,A}\big](p_{a,A} + 2\delta_M p_{a,B} + \phi) + \big[p_{b,B} + \sum_{z \notin \{a,b\}} p_{z,B}\big](\delta_M p_{a,A} + 2p_{a,B} + \delta_M \phi)\Big) > 0
\end{aligned}
$$

which is true as all terms are weakly positive. So the plausibility of a increases as ϕ increases.

$$
\begin{aligned}
\frac{\partial}{\partial \phi} \log F(H_b) = &\frac{\eta_c\Big(\delta_c p_{b,A} + \delta_c \delta_M p_{b,B}\Big)}{\begin{array}{r}\eta_c(p_{a,A} + \phi)(\delta_c p_{b,A} + \delta_c \delta_M p_{b,B}) + p_{b,A}(p_{b,A} + \delta_M p_{b,B}) + \eta_c \sum_{z \notin \{a,b\}} p_{z,A}(\delta_c p_{b,A} + \delta_c \delta_M p_{b,B}) \\ + \eta_c p_{a,B}(\delta_c \delta_M p_{b,A} + \delta_c p_{b,B}) + p_{b,B}(\delta_c p_{b,A} + p_{b,B}) + \eta_c \sum_{z \neq b} p_{z,B}(\delta_c \delta_M p_{b,A} + \delta_c p_{b,B})\end{array}} \\
&- \frac{\delta_c p_{b,A} + \delta_c \delta_M p_{b,B}}{\begin{array}{r}(p_{a,A} + \phi)(\delta_c p_{b,A} + \delta_c \delta_M p_{b,B}) + p_{b,A}(p_{b,A} + \delta_M p_{b,B}) + \sum_{z \notin \{a,b\}} p_{z,A}(\delta_c p_{b,A} + \delta_c \delta_M p_{b,B}) \\ + p_{a,B}(\delta_c \delta_M p_{b,A} + \delta_c p_{b,B}) + p_{b,B}(\delta_c p_{b,A} + p_{b,B}) + \sum_{z \neq b} p_{z,B}(\delta_c \delta_M p_{b,A} + \delta_c p_{b,B})\end{array}}
\end{aligned}
$$

If we divide the numerator and denominator of the first term by η_c, we see that the numerators of the first and second terms are then the same, while the denominator of the first term is larger. Thus, the whole derivative is negative. So the plausibility of b decreases as ϕ increases. Thus, $\frac{d}{d\phi} \frac{\pi_a}{\pi_b} = \frac{d}{d\phi} \frac{F(H_a)}{F(H_b)} > 0$. Note that this result did not depend on the share of people with each major conditional on any career. Thus, an analogous derivation is possible in the case where $\Big(c(x), m(x)\Big) = (a, B)$. The prediction therefore does not depend on the major of x.

Let $p_{a,A} = p_a - p_{a,B}$ and let $\phi = 0$. Then

$$
\begin{aligned}
&\frac{\partial}{\partial p_{a,B}} \log F(H_a) = \\
&= \frac{2(p_{a,B} - p_{a,A})(1 - \delta_M) + \eta_c \delta_c (1 - \delta_M) \sum_{z \notin \{a\}} (p_{z,B} - p_{z,A})}{\begin{array}{r}p_{a,A}(p_{a,A} + \delta_M p_{a,B}) + \eta_c p_{b,A}(\delta_c p_{a,A} + \delta_c \delta_M p_{a,B}) + \eta_c \sum_{z \notin \{a,b\}} p_{z,A}(\delta_c p_{a,A} + \delta_c \delta_M p_{a,B}) \\ + p_{a,B}(\delta_M p_{a,A} + p_{a,B}) + \eta_c p_{b,B}(\delta_c \delta_M p_{a,A} + \delta_c p_{a,B}) + \eta_c \sum_{z \notin \{a,b\}} p_{z,B}(\delta_c \delta_M p_{a,A} + \delta_c p_{a,B})\end{array}} \\
&- \frac{2(p_{a,B} - p_{a,A})(1 - \delta_M) + \delta_c (1 - \delta_M) \sum_{z \notin \{a\}} (p_{z,B} - p_{z,A})}{\begin{array}{r}p_{a,A}(p_{a,A} + \delta_M p_{a,B}) + p_{b,A}(\delta_c p_{a,A} + \delta_c \delta_M p_{a,B}) + \sum_{z \notin \{a,b\}} p_{z,A}(\delta_c p_{a,A} + \delta_c \delta_M p_{a,B}) \\ + p_{a,B}(\delta_M p_{a,A} + p_{a,B}) + p_{b,B}(\delta_c \delta_M p_{a,A} + \delta_c p_{a,B}) + \sum_{z \notin \{a,b\}} p_{z,B}(\delta_c \delta_M p_{a,A} + \delta_c p_{a,B})\end{array}}
\end{aligned}
$$

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 8, let $\phi=0$ (shutting down personal acquaintances) and let $p_{A|z}=K$ for all z. Then

$$\begin{aligned}
\frac{\pi_a}{\pi_b} =& \left(p_{a,A}p_a+\eta_c\delta_c p_{b,A}p_a+\eta_c\delta_c\sum_{z\notin\{a,b\}}p_{z,A}p_a+p_{a,B}p_a+\eta_c\delta_c p_{b,B}p_a+\eta_c\delta_c\sum_{z\notin\{a,b\}}p_{z,B}p_a\right)\\
&/\left(p_{a,A}p_a+\delta_c p_{b,A}p_a+\delta_c\sum_{z\notin\{a,b\}}p_{z,A}p_a+p_{a,B}p_a+\delta_c p_{b,B}p_a+\delta_c\sum_{z\notin\{a,b\}}p_{z,B}p_a\right)\\
&/\left(\eta_c p_{a,A}\delta_c p_b+p_{b,A}p_b+\eta_c\delta_c\sum_{z\notin\{a,b\}}p_{z,A}p_b+\eta_c\delta_c p_{a,B}p_b+p_{b,B}p_b+\eta_c\delta_c\sum_{z\neq b}p_{z,B}p_b\right)\\
&\times\left(p_{a,A}\delta_c p_b+p_{b,A}p_b+\delta_c\sum_{z\notin\{a,b\}}p_{z,A}p_b+\delta_c p_{a,B}p_b+p_{b,B}p_b+\delta_c\sum_{z\neq b}p_{z,B}p_b\right)\\
=& \left(p_a+\eta_c\delta_c p_b+\eta_c\delta_c\sum_{z\notin\{a,b\}}p_z\right)/\left(p_a+\delta_c p_b+\delta_c\sum_{z\notin\{a,b\}}p_z\right)\\
&/\left(\eta_c\delta_c p_a+p_b+\eta_c\delta_c\sum_{z\notin\{a,b\}}p_z\right)\times\left(\delta_c p_a+p_b+\delta_c\sum_{z\notin\{a,b\}}p_z\right)>\frac{p_a}{p_b}\\
\iff& \eta_c\delta_c^2(p_b-p_a)\sum_{z\notin\{a,b\}}p_z+\eta_c\delta_c^2(p_b-p_a)\sum_{z\notin\{a,b\}}p_z+(1-\delta_c)p_ap_b(p_b-p_a)+\eta_c\delta_c(p_b^3-p_a^3)\\
&+\eta_c\delta_c(p_b^2-p_a^2)\sum_{z\notin\{a,b\}}p_z+\eta_c\delta_c^2p_ap_b(p_b-p_a)\\
&>0
\end{aligned}$$

which is true when $p_b>p_a$.