What Jobs Come to Mind?
Stereotypes about Fields of Study

Please click here for the latest version.

John J. Conlon Dev Patel

September 28, 2022*

Abstract

We show that college freshmen in the U.S. hold systematically incorrect beliefs about how their choice of major affects their future occupation. 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). To explain this pattern, 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 have systematic, and at times counterintuitive, effects on their beliefs. We embed stereotypical thinking into a simple model of college major choice to show that it boosts demand for "risky" majors: ones with rare stereotypical careers and low-paying alternative jobs. In a field experiment, providing statistical information on career frequencies to first-year college students has modest but significant effects on their intended fields of study, with students especially moving away from risky majors.

*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, Emily Breza, Katie Coffman, Lucas Coffman, 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. Sometimes students are too optimistic on average, sometimes too pessimistic, and almost always there is wide disagreement between decision-makers even about objective facts.[1] Similar puzzles—pervasive errors in beliefs 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 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

[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).

that job: 65% of prospective art majors expect to be artists (only 17% are), 63% of biology majors expect to be doctors (23% are), 42% of communications/journalism majors expect to be writers or journalists (4% are), 62% of psychology majors expect to be counselors (21% are), and so on. In turn, few students expect to work in business or to teach (unless they are pursuing business or education majors), and almost none expects to be non-employed. By themselves, of course, these results are consistent with mechanisms other than stereotyping. For instance, they could be driven by students exaggerating their own abilities, overestimating the demand for certain jobs, or selecting fields of study about which they 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 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 assessing the likelihood of career c after majoring in M try to think of people with c and M. Their beliefs then depend on the careers and majors of the people they manage to think of: the more lack either c or M, the lower her belief about $P(c|M)$. Because a student's belief depends on who comes to mind, regularities in human recall play a key role (Kahana, 2012; Bordalo, Conlon, et al., 2022; Bordalo, Burro, et al., 2022). First, recall is limited, meaning that memories must compete for retrieval. Thus, when one person becomes easier to recall, others thereby become 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 naturally predicts stereotyping. 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 common careers 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 possibility is that students neglect all other outcomes, but the model makes a more subtle 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, absent this mechanism, 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), 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 journalism.

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.

[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).

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 personal experiences are especially available in memory (Malmendier & Wachter, 2022). We show that the careers and majors of people students know particularly well—their parents and other role models—systematically affect their beliefs in 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]

In addition to these more straightforward role-model effects, the model also predicts that knowing someone with one major should affect students' beliefs about other majors. For example, when a career c is stereotypical of a major M, knowing someone with that career but a *different* major should reduce beliefs about $P(c|M)$. Intuitively, knowing a teacher who majored in biology might dampen one's stereotype that all education majors become teachers. In contrast, when c is *not* stereotypical of M, knowing someone with that profession but a different major can have the opposite effect. For example, a student may decide that more journalism majors work in business if they know many businesspeople, even if none of them majored in journalism. Both of these patterns appear in our data.

Though the model makes numerous predictions that appear to be borne out in the survey data, stereotyping is by far the dominant pattern in students' beliefs.[6] We therefore focus on this channel when analyzing the welfare consequences of mistaken beliefs. We describe a stylized model of major choice in which students care about salary but also have heterogeneous non-pecuniary preferences for stereotypical jobs. To motivate this approach, we first show that beliefs about careers drive heterogeneity in beliefs about salaries: students who

[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]More precisely, we use a Shapley-Shorrocks decomposition of R^2 (Shorrocks, 1982) to quantify the relative contribution of various mechanisms in explaining variation in the difference between students' beliefs and the true frequency of careers. We find that a career being stereotypical of a major explains more of this variance that any other mechanism we look at (including role model effects, overestimation of rare careers, selection into majors, and biases about careers unconditional on majors).

think higher-paying careers are particularly common among a major's graduates expect to earn a particularly high salary if they pursue that major. But students also care about careers over and above their impact on salaries: in hypothetical choices, they report being willing to sacrifice a substantial amount in earnings to work in their preferred career.

We focus our welfare analysis on students considering a field of study with lower expected earnings in the hopes of attaining an appealing stereotypical career. For students on the margin between majors, two factors 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 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 those majors' stereotypical careers are objectively quite common or because wages are high even for those who end up in non-stereotypical careers. 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.[7]

Finally, 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. These effects are larger among students who originally planned to pursue risky majors and who particularly overestimated the likelihood of their preferred major's stereotypical outcome. However, the size of these effects are somewhat modest: a six percentage point shift for risky majors and a three percentage point shift for less risky majors. We also find some corroborating evidence that the information affected students' choices of classes in subsequent semesters, though these estimates are often imprecise. The modest size of these effects may in part reflect the fact that students substantially discount the information we provided about the careers of others when updating their beliefs about their own future career.

Our study is related to a rich literature on beliefs and human capital investment.[8] In

[7]Such welfare effects could be smaller if, for instance, exaggerated beliefs about the likelihood of stereotypical careers motivate students to work harder in school. In our OSU survey data, however, we find no evidence of this channel: students report that they would expect to study slightly *more* if they could not enter their preferred career.

[8]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.

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 (Arcidiacono et al., 2012, 2020; Wiswall & Zafar, 2018, 2021). These studies tend to focus on beliefs about average salary conditional on major (e.g., Conlon 2021), though a recent notable exception is Arcidiacono et al. (2020) who decompose Duke undergraduates' expectations about the wage returns of majors into their anticipated effects on wages conditional on occupation choice and on their likelihood of entering each occupation. Closely related to their findings, we show that the occupation beliefs in our sample are highly predictive of students' earnings expectations, highlighting one potential channel underpinning salary beliefs. This could help to explain the puzzling finding of relatively small elasticities of major choice with respect to average wages (Wiswall & Zafar, 2015a; Beffy et al., 2012). Our results suggest that students' may be unaware of the true mapping between majors and occupations, which could attenuate such estimates (Long et al., 2015).

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.[9] Our conception of stereotyping as a byproduct of associative memory speaks to the broader literature studying the origins of beliefs and their biases.[10] The underlying mechanics of memory and mental simulation in our model build off literatures in psychology, and recent work in economics increasingly incorporates many of these assumptions.[11]

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

[9] 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).

[10] 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).

[11] 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).

2015.[12] 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.[13] In that sense, we call this sample "nationally representative" of incoming college freshmen.

We focus on a subset of questions related to students' expectations about the future. First, students are asked to mark their "probable field of study" from a list of around 80 options, including "Other" and "Undecided."[14] 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.

To compare students' expected careers to the actual distribution of occupations, we use the Annual Social and Economic Supplement (ASEC) to the Current Population Survey (CPS) data from 1976 onward (Flood et al., 2021). We restrict the data to those aged 33 to 37, because by this time the vast majority of people are no longer students and have started their career.[15] 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 while 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

[12] 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.

[13] 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.

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

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

do (0.7% and 2.8%).[16] 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.[17] 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.[18]

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.

[16]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.

[17]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.

[18]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.

3.1 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 work 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 Alternatives to Stereotypical Thinking

The patterns described in Sections 2 and 3.1, while consistent with stereotypical thinking, by themselves could reflect several over potential mechanisms. For instance, they could be the result of students exaggerating their ability to achieve difficult careers, simply overestimating common or salient jobs, or selecting fields of study about which they have particularly extreme beliefs. Finally, the Freshman Survey asks qualitative questions (asking students to mark one job as their "probable" career), which complicates interpreting differences between expected and actual outcomes as errors. To distinguish between these possibilities, we designed administered surveys of first-year undergraduates at The Ohio State University (OSU). In this section, we describe each of these alternative explanations in more detail along with how our OSU surveys distinguish them from stereotypical thinking. We then describe the data in more detail and the empirical results in the following section.

Qualitative vs Quantitative Expectations: First, one might reasonably worry that the qualitative nature of the Freshman Survey questions make 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 may be overstating the extent of bias in students' true beliefs. For this reason, in our OSU surveys described below, we instead asked students quantitative questions about the percent chance that they will work in each job and about the percent of others doing so, allowing students to express uncertainty more precisely.

Overestimation of Small (or Large) Probabilities: Next, students could systematically over- or underestimate unlikely careers. For example, if stereotypical jobs tend to be common and students simply exaggerate or latch onto more common outcomes, this could superficially look like stereotypical thinking. We can separate this force from stereotyping by comparing students' beliefs about stereotypical careers to their beliefs about *similarly likely* but non-stereotypical outcomes: in short, by controlling for true frequencies.

Confidence: Third, students' beliefs could deviate from the truth because they believe their outcomes will be systematically different from population outcomes, for example due to overconfidence. We can eliminate this channel by asking students their beliefs about the outcomes of *others*, rather than themselves. These are what we call the "population beliefs" in the OSU surveys (described below).

Selection: Fourth, students may systematically select into majors depending on their 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 that major* despite no underlying bias in the population at large. We can rule out this explanation by eliciting students' beliefs about potential outcomes conditional on pursuing (a random sample of) majors other than the actual major they are pursuing.

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 "wishful thinking" in which students hold mistaken beliefs in order to *ex post* justify their chosen major. Similarly, 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. By asking students about *all* majors—not just their intended majors—we can distinguish stereotypical thinking from any such mechanisms.

Unconditional Biases: Fifth, the role of stereotyping could be confounded if students exaggerate the unconditional likelihood of careers that happen to be stereotypical of common majors, perhaps because certain occupations are simply more salient. For example, if most people majored in business and students exaggerated the number of people working in business (contrary to what we actually find), this could spuriously be interpreted as stereotyping. We can eliminate this source of bias by asking students about multiple majors and controlling for career fixed effects.

3.3 Isolating Stereotypical Thinking

To isolate stereotypical thinking from these other potential mechanisms, we designed and administered surveys among first-year undergraduates in 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. Column 2 of Table 1 gives self-reported demographic information about 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.[19] It is also similar to the Freshman survey along gender, ethnicity, first-generation status, and self-reported family income.[20] See Appendix B for more details about the survey and implementation.

The survey began by displaying the ten major groups and asking students to rank them by how likely they thought they were to graduate with a degree in each. For the sake of survey length, we asked student detailed questions about only four of these ten majors. These included students' two top ranked majors, as well as two additional majors chosen randomly from the remaining eight. When applicable, we use inverse probability weights to estimate average beliefs unconditional on major ranking.[21] 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 results.

The primary questions that we focus on asked students about the distribution of careers by major. First, we asked students to report 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 nine careers in a random order, plus "working in any other job?" and "not working for pay?" We call these students' "population beliefs," in that they ask students about the population of people who already have each major, rather than their own outcomes if they pursued that major. Note that these population questions were worded to allow us to compare students' answers with the true conditional frequency of occupations in the American Community Survey (ACS). Next, the survey asked what we call the "self beliefs" version of the same question, which asked, "Now 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..." The same eleven options as in the population question were

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

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

[21]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).

then listed. In both questions, students' answers had to add up to 100%. The order of the questions—population beliefs then self beliefs—was intentional: we did not want students to anchor their population beliefs on their self beliefs, because our empirical strategy ultimately rests on the former to test for stereotypical thinking.

First, we simply report average beliefs about each major's most stereotypical career. For completeness, Table A.VII in the Appendix shows average beliefs about each major-career pair. The light blue bars in Figure 2 show average self beliefs restricting attention to students' most likely field of study. We take these data, on self-beliefs about students' own expected major, to be the quantitative (i.e., probabilistic) analogue of the qualitative questions in the Freshman survey. We again 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, the average OSU student's belief about their own likelihood of having their major's stereotypical career is 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 despite differences in time period (1970s-2010s vs 2020), sample (students around the country vs only Ohio State), and elicitation method (qualitative vs quantitative expectations).

The gray bars in Figure 2 show the average *population* belief, expanding the sample to include all four majors that each student in the 2020 sample was asked about, rather than restricting it to each student's 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.[22] These differences are again quite large and comparable to both the 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 (as defined in Section 3.2).

To more formally test for stereotypical thinking, Table 2 shows OLS estimates of variants of the regression specification in equation 1, where $\pi^i_{c|M}$ is student i's 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, which we only include in the final

[22] 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%.

column of Table 2.

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

All regressions cluster standard errors at the student level and at the career-by-major level.

Column 1 of Table 2 restricts the sample to students' top-ranked major. The dependent variable is student i's self beliefs about their chances of having career c conditional on graduating with the major with which they said they are most likely to graduate, or what we will refer to as their top major. The coefficient of 0.51 on $p_{c|M}$ show that students' beliefs are responsive to true population frequencies, though the relationship is far from one-for-one. Second, the coefficient of 0.43 on the dummy for being the most stereotypical career shows that students perceive themselves to be 43 percentage points ($p < 0.01$) more likely to be working in their top major's stereotypical career than other similarly frequent careers.

The next three columns of Table 2 then change this baseline specification to isolate the role of stereotyping. Column 2 is the same as Column 1 except that the dependent variable is now students' *population* beliefs (i.e., their beliefs about the fraction of others with this major who have each career), shutting down the "Confidence" channel from Section 3.2. Column 3 then expands the sample to include all four majors that each student was asked about, not just their top ranked major, to shut down the "Selection" channel.[23] Finally, column 4 adds career-by-individual fixed effects (μ^i_c from equation 1), shutting down the "Unconditional Biases" channel. As we move from column 1 to column 4, the estimated effect of stereotypical thinking gets somewhat smaller but remains large and highly statistically significant even in column 4. In particular, 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 less stereotypical outcomes.

These results are highly robust to different specifications and ways of constructing the relevant variables. For example, we see nearly identical results when disaggregating the sample by gender and using gender-specific conditional probabilities, as shown in Table A.VIII. Table A.IX shows similar regressions but where the dependent variable is the *difference* between students' beliefs and the actual fraction of people with a career, in essence setting γ to one in equation 1. Table A.X shows similar regressions but using a continuous measure of stereotypical-ness rather than the indicator function we use in the main text. We see large and statistically significant results across all these specifications.

[23] As described above, we use inverse probability weights in these regressions to account for the differential probability of answering for a given major across respondents.

3.4 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.[24] 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.[25] We can then calculate what we call the "implied error" about the probability of working in career c as shown in equation 2, where p_c is the true fraction working in c.

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

We can then compare these implied errors to a corresponding notion of "error" from the Freshman Survey data: the difference between the fraction of students who expect to have a career in each occupation minus the true proportion of college graduates with that occupation. Intuitively, this analysis is asking whether the Freshman Survey respondents' expectations are what we would expect if they held the same (population) beliefs as the OSU sample. As Figure A.I shows, there is a robust positive relationship between actual "errors" in the Freshman Survey data and ImpliedError_c. The correlation between between implied and actual error is 0.81 and is highly statistically significant ($p < 0.01$). An OLS regression of the error in the Freshman Survey data on the implied error has a coefficient of 0.84 ($p < 0.01$), with an R^2 of 0.71. We conclude from this exercise that the pattern of overestimation of careers in the Freshman Survey is quite close to what we would expect from the OSU students' beliefs.

3.5 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

[24] In Appendix Section B.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.

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

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

In 2021, we administered two similar surveys among a new cohort of the same Exploration program at Ohio State (see Appendix B for more details). The first of these also asked students' beliefs about the frequency of careers conditional on majors. In addition, immediately after each such question, students were asked "And on a scale between 0 (completely uncertain) and 100 (completely certain), how confident are you that the answers above are close to correct?" 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 Where Do Beliefs Come From?

In this section, we develop a simple cognitive model of belief formation in which students form their expectations by relying on a combination of their experiences—the people they know or have heard about with certain combinations of jobs and fields of study—and extrapolation—how easy it is to imagine major-to-career paths given their experiences. Our structure takes the basic "similarity and interference" framework of Bordalo, Conlon, et al. (2022) to model retrieval from memory and incorporates simulation from retrieved experiences as in Bordalo, Burro, et al. (2022).

4.1 Setup

We assume students form beliefs about the likelihood of careers either conditional on a particular major or unconditional on major. We use lower-case letters to denote careers (i.e., $c \in \{a, b, ...\}$) and upper-case letters to denote majors (i.e., $M \in \{A, B\}$). For simplicity, we assume there are only two majors. 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 being a businessperson.

The student assesses the "plausibility" $F(H)$ of each relevant "hypothesis" H. When assessing unconditional probabilities, these hypotheses H_c just correspond to each career c. When assessing probabilities of careers conditional on major M, these hypotheses $H_{c,M}$ are *combinations* of each career and that major. Their probabilistic beliefs, shown in equation 3, are then just the plausibility of H normalized such that their beliefs about all relevant hypotheses sum to one.

$$\pi_c = \frac{F(H_c)}{\sum_{z\in\{a,b,\ldots\}} F(H_z)} \qquad\qquad \pi_{c|M} = \frac{F(H_{c,M})}{\sum_{z\in\{a,b,\ldots\}} F(H_{z,M})} \tag{3}$$

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" $\mathcal{D}$, where the number of experiences in the database is D. We can think of the database as comprising encounters the student has had with people who have followed different career paths. These could be people the student knows personally like friends or family, as well as people they have merely heard about, e.g., from the media or second-hand from others.

Second, the student uses the person they have retrieved to try to imagine the hypothesis under consideration. We call this process "simulation," and the driving assumption of the model is simply that students more easily simulate a hypothesis when the retrieved person shares more features with the hypothesis. For example, suppose the student is assessing the plausibility of becoming an artist after majoring in art. This outcome will seem very plausible if the student thinks of someone with that particular major-career pair, less plausible if that person followed a similar but distinct path (e.g., an artist who majored in something else), and least plausible if that person did nothing remotely similar.

4.1.1 Stage 1: Retrieval

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 (i.e., similarity-based) and frequency-based (Kahana, 2012). These two forces correspond to the two components of equation 4, which we describe in turn.

$$a(e, H) = N(e)S(e, H) \tag{4}$$

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 4, 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.

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 can then define similarity between a person and a hypothesis as simply the average pair-wise similarity between that person and all experiences consistent with the hypothesis (we use $\mathcal{H}$ to denote the set of experiences consistent with H), as shown in equation 5.

$$S(e, H) = \frac{1}{|\mathcal{H}|} \sum_{u \in \mathcal{H}} N(u) s(e, u) \tag{5}$$

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.

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 6.

$$r(e, H) = \frac{a(e, H)}{\sum_u a(u, H)} \tag{6}$$

The numerator of equation 6 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 personal acquaintance) do so at the expense of others.

4.1.2 Stage 2: Simulation

After a person e comes to mind, the student then uses them to try to imagine the hypothesis under consideration. Following Kahneman & Tversky (1981) and Bordalo, Burro, et al. (2022), we call this process "simulation" and assume it is also governed by similarity. Let $\sigma(e, H)$ be the ease of simulating H after retrieving e. Recall that hypotheses are defined as either careers or career-major pairs, and we assume that $\sigma(e, H)$ is decreasing in the number of these features that differ between e and H.

We assume the student repeats this process of retrieval-and-simulation many times, averaging the ease of simulation across all the people they retrieve.[26] Their assessed plausibility

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

then converges in probability to equation 7 where the expectation operator is with respect to frequencies in the student's database $\mathcal{D}$. All proofs are in Appendix C.

$$F(H) = \sum_{e \in \mathcal{D}} r(e, H)\sigma(e, H) = \frac{E\Big[a\big(e, H\big)\sigma\big(e, H\big)\Big]}{E\Big[a\big(e, H\big)\Big]} \tag{7}$$

The plausibility of H is thus the availability-weighted average ease-of-simulation of that hypothesis.

This model naturally nests the rational-expectations benchmark. In particular, if $a\big(e, H\big)$ is constant, $\sigma\big(e, H\big)$ is a dummy variable indicating whether e is consistent with H, and the database $\mathcal{D}$ is representative of true outcomes, then the student's beliefs will be correct. This corresponds to the case where the student takes an (unbiased) random sample of people in their database, without overweighting people they have encountered more times, and counts the number that are consistent with each hypothesis.

4.1.3 Simplifying Assumptions

To analyze students' beliefs, we employ a series of simplifying assumptions. First, we assume each person $e \in \mathcal{D}$ is characterized only by their career $c(e)$ and major $M(e)$. We also assume that careers and majors of people in the student's database $\mathcal{D}$ match the true joint distribution of careers and majors. We make this latter assumption merely to highlight that the model predicts biases in beliefs even in this case.

We assume the functional form for the similarity function $s(e, u)$ between people e and u given by equation 8.

$$s(e, u) = \delta_c^{\mathbb{1}(c(e) \neq c(u))} \times \delta_M^{\mathbb{1}(M(e) \neq M(u))} \tag{8}$$

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 functional form in equation 8 is standard in models with discrete features (Mack & Palmeri, 2020; Evers et al., 2021; Bordalo, Conlon, et al., 2022).

To analyze the effect of personal acquaintances on beliefs, we assume one person, whom we call x, is the student's personal acquaintance 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

---

the weight.

experiences that are with x.

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

We also assume that the database $\mathcal{D}$ is sufficiently large that the career/major of x has a negligible impact of the true frequencies of careers and majors in $\mathcal{D}$. This also allows us to take derivatives with respect to the true fraction $p_{c,M}$ of people with major M and career c.

For ease-of-simulation, we assume the functional form in equation 10, whereby ease-of-simulation decreases by a factor of $\eta_c \leq 1$ if e lacks the relevant career an by η_M if e lacks the relevant major.

$$\sigma(e, H_c) = \eta_c^{\mathbb{1}(c(e)\neq c)} \qquad\qquad \sigma(e, H_{c,M}) = \eta_c^{\mathbb{1}(c(e)\neq c)} \times \eta_M^{\mathbb{1}(M(e)\neq M)} \tag{10}$$

These functional-form assumptions also nest the rational-expectations benchmark: if $\delta_c = \delta_M = 1$, $\eta_c = \eta_m = 0$, and $\phi = \frac{1}{D}$, then beliefs will be correct.

4.2 Beliefs about Careers Conditional on Major

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_B be the fraction of people with major B. By definition, $p_{a,B} = p_{a|B}p_B$, where $p_{a|B}$ is the probability 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 and simulation despite being normatively irrelevant for the question at hand. These effects are summarized in Prediction 1.

Prediction 1: *If a career is not too unlikely, beliefs increase in stereotypical-ness: decreasing $P(a|B)$ increases beliefs about $P(a|A)$. More precisely, equation 11 holds:*

$$\frac{\partial}{\partial p_{a|B}}\left[\log \frac{\pi_{a|A}}{\pi_{b|A}}\right] \approx p_B\left(\eta_M \frac{1}{p_{a,A}} - (1-\delta_c)\right) \tag{11}$$

where the approximation is a first-order Taylor expansion around the rational benchmark ($\delta_c = \delta_M = 1$ and $\eta_c = \eta_M = \phi = 0$).

To intuitively interpret Prediction 1, consider a student forming beliefs about how many art majors go on to become artists. Equation 11 says that, if $p_{a,A}$ is not too small (such that the second term in equation 11 dominates), this 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.[27]

In addition to stereotypical thinking, we can use the model to derive additional predictions, which we then test in Section 5. First, it predicts that, besides stereotyping, the student's beliefs should be *undersensitive* to true frequencies, a result we summarize in Prediction 2.

Prediction 2: *Absent distortions from stereotyping or personal acquaintances, 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 personal acquaintances). Then equation 12 follows whenever $\delta_c < 1$ or $\eta_c > 0$:

$$\frac{\pi_{a|A}}{\pi_{b|A}} > \frac{p_{a|A}}{p_{b|A}} \iff p_{a|A} < p_{b|A} \tag{12}$$

The intuition behind this result is straightforward. When $\delta_c < 1$, meaning that retrieval is based on similarity, then 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. In addition, when $\eta_c > 0$, the student can partially imagine a hypothesis even when they do not retrieve someone exactly consistent with it. This again disproportionately benefits rarer hypotheses, boosting their assessed likelihood.

[27] The stipulation in Prediction 1 that the career be not too unlikely reflects the role of simulation. For example, suppose that almost no one becomes an artist after majoring in art, so the student finds it almost impossible to think of anyone with that major-career combination. Then increasing the share of people with major B with that job can increase the plausibility of a conditional on major A, despite reducing its stereotypical-ness. Intuitively, being able to think of *anyone* with a, even of the wrong major, can make it easier to imagine someone having a after majoring in A.

We next analyze how the increased availability of personal acquaintances can systematically bias 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 introducing a personal acquaintance x into the student's database. 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 her personal acquaintance graduated with that major. Prediction 3 summarizes the effect this has on her beliefs.

Prediction 3: *Assuming the alternative career is not too unlikely, 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, if $p_{b,A}$ is not too small or η_c is not too large, $\frac{\partial}{\partial\phi}\log\frac{\pi_{a|A}}{\pi_{b|A}} > 0$.

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 aids in simulating that hypothesis.

We can next ask how the student's beliefs about one major are affected by knowing someone with a *different* major. Prediction 4 summarizes such effects, which depend on the objective frequency of the acquaintance's career.

Prediction 4: *The effect of personally knowing someone with career a but major B has ambiguous effects on the student's beliefs about the likelihood of career a conditional on major A:*

- *If career a is sufficiently common conditional on major A, as when a is stereotypical of A, then the effect is negative.*
- *If career a is sufficiently rare conditional on major A, as when a is not stereotypical of A, then the effect is positive.*

More precisely, let $\Big(c(x), M(x)\Big) = (a, B)$. Then equation 13 follows:

$$\frac{\partial}{\partial\phi}\log\frac{\pi_{a|A}}{\pi_{b|A}} \approx \eta_M\frac{1}{p_{a,A}} - (1-\delta_c)\frac{1+p_{b,A}-p_{a,A}+p_{b,B}-p_{a,B}}{(1+\phi)^2} \tag{13}$$

where the appoximation is a first-order Taylor expansion around $\delta_c = \delta_M = 1$ and $\eta_c = \eta_M = 0$.

The effect of the personal acquaintance on beliefs is ambiguous because interference and simulation have countervailing effects. First, suppose $p_{a,A}$ is relatively large, such as when a

is stereotypical of A. In this case, but for the personal acquaintance, the student would easily retrieve people with that career-major pair. Introducing the personal acquaintance, who has the same career but different major, distracts from these people and thus lowers beliefs. This is reflected in the second term in 13. Intuitively, knowing an artist who actually majored in business might dampen the student's stereotype that all art majors become artists.

But suppose, instead, that the career the student is assessing is extremely rare, such that she struggles to think of anyone with the right career-major combination. In that case, there is (almost) no one for the personal acquaintance to distract from, and instead it may *boost* a career's plausibility to know someone with that career but the wrong major. Intuitively, knowing many artists who majored in art may boost the student's beliefs about the fraction of *business* majors who become artists. This occurs because the student does not know anyone who majored in business but who became an artist, so knowing *anyone* with that career helps to make it seem less implausible. This countervailing effect is reflected in the first term in 13.

4.3 Beliefs about Careers Unconditional on Major

The model also makes predictions for beliefs about careers *unconditional* on major. Prediction 5 summarizes the effect of the student's personal acquaintance on her unconditional beliefs.

Prediction 5: *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 $M(x)$, equation 14 follows*:

$$\frac{\partial}{\partial \phi} \log \frac{\pi_a}{\pi_b} > 0 \tag{14}$$

The intuition behind Prediction 5 is the same as for Prediction 3; personal acquaintance with someone who has career c makes that career easily retrievable, which boosts the plausibility of H_c and thus beliefs about c (relative to other careers).

A less obvious prediction of the model stems from the fact that, unlike with beliefs conditional on major, 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 6: *Unconditional beliefs about the frequency of a career increase in the extent* *to which it is concentrated within a particular major.*

More precisely, let $\phi = 0$ (shutting down personal acquaintances) and let $p_{A|c} = p_{B|c}$ for all careers $c \neq a$. Then equation 15 follows whenever $\delta_M < 1$:

$$\frac{\partial}{\partial p_{B|a}} \log \frac{\pi_a}{\pi_b} > 0 \iff p_{B|a} > p_{A|a} \tag{15}$$

The intuition behind Prediction 6 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 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 5 and 6, unconditional beliefs (like conditional beliefs) are undersensitive to true frequencies. We summarize this result in Prediction 7:

Prediction 7: *Absent effects from personal acquaintances and differential concentration of careers within majors, the student exaggerates rare careers and underestimates common careers.*

More precisely, let $\phi = 0$ (shutting down personal acquaintances) 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 16 follows whenever $\delta_c < 1$ or $\eta_c > 0$:

$$\frac{\pi_a}{\pi_b} > \frac{p_a}{p_b} \iff p_b > p_a \tag{16}$$

The intuition for Prediction 7 is the same as for conditional beliefs. When $\delta_c < 1$, meaning that retrieval is based on similarity, then the student is disproportionately able to think of people with a when she is thinking of that career. This disproportionately benefits rarer careers that would more often fail to come to mind under unbiased sampling. When $\eta_c > 0$, the student is partially able to imagine having c even when they do not retrieve someone with that career. This again disproportionately benefits rarer careers, boosting their assessed likelihood.

5 Additional Tests of the Model

We now turn to testing the additional predictions of the model. We start by looking at average biases in conditional and unconditional beliefs (Sections 5.1 and 5.2 respectively), and then we exploit differences in the careers and majors of students' role models to test the model's predictions about heterogeneity in beliefs (Section 5.3).

5.1 Which Outcomes Do Stereotypical Thinkers Neglect?

When students exaggerate the likelihood that a major leads to its stereotypical career, which outcomes do they underestimate? One possibility is that students underestimate all non-stereotypical outcomes. Prediction 2 from Section 4 suggested instead that, in addition to exaggerating stereotypical careers, students should also overestimate sufficiently *rare* careers. Panel A of Figure 3 plots the average conditional population belief from the 2020 OSU sample for each career-major pair, broken up by whether the career is stereotypical for that major or not. Panel B plots the share of students in the Freshman Survey sample who list each career as their probable career, conditional on listing each major as their probable field of study (also broken up by whether the career is stereotypical of that major or not). As in Figure 2, students exaggerate almost every stereotypical career. Focusing on non-stereotypical outcomes, however, 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 3 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 outcomes.[28] A particularly striking case is non-employment (triangles in Figure 3). Almost no students in the Freshman Survey data, no matter their major, report that they expect not to be employed (by selecting "homemaker" as their probable career). The 2020 OSU students in turn 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. In the first of our 2021 OSU surveys, we asked similar questions about the frequency of careers (plus

[28]One instructive exception to this pattern is that government majors in the Freshman Survey sample are much more likely to list an "other" job as their probable major than to actually have such a job: 70% of such students list their likely career as "Policymaker/government" or "foreign service worker (including diplomat)". We take this as further evidence that students appear to latch onto stereotypical careers even when these fall outside our nine broad groups of careers.

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.XI 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.[29]

5.2 Unconditional Beliefs: Frequency and Concentration

Section 4.3 described how the model makes two predictions for average beliefs about careers unconditional on major. First, beliefs about a career should increase in the extent to which it is concentrated within majors. That is, careers that are primarily attained through a particular career path should be relatively overestimated. Second, as with conditional beliefs, rare careers should be overestimated.

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. We also look at the career expectations of students in the Freshman Survey unconditional on their expected major.

We estimate by OLS the following regression equation:

$$\overline{BeliefError}_c = \alpha + \beta_1 TrueShare_c + \beta_2 HHI_c + \epsilon_c \tag{17}$$

In equation 17, $\overline{BeliefError}_c$ is the average error in unconditional beliefs about career c. For the OSU 2021 sample, this corresponds to the difference between the average population belief about that career and the true fraction of college graduates working in that career. For the Freshman Survey, it is the difference between the share of Freshman who list that career as their "probable" career and the actual share of graduates working in that career.

[29] Figure A.III in the Appendix breaks up survey responses and actual employment probabilities by gender. The broad patterns are very similar to those in Figure 3. 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.

$TrueShare_c$ is then the true share working in that career. Finally, HHI_c is a variation of a Herfindahl–Hirschman index, meant to capture how concentrated a career is within particular majors. In particular, $HHI_c = \sum p^2_{m|c}$.

Figure 4 plots the partial correlations between $\overline{BeliefError}_c$ and $TrueShare_c$ as well as between $\overline{BeliefError}_c$ and HHI_c, for both the 2021 OSU sample and the Freshman survey. We see clear evidence in favor of both Predictions 6 and 7: errors in beliefs decrease in the true share of people with each career ($p < 0.01$ in both samples) and increase in how concentrated the career is within majors ($p < 0.05$ in both samples).

5.3 Role Model Effects

In this section, we test the model's predictions about heterogeneity in beliefs by asking how a student's personal background—and in particular, their exposure to people with certain careers and majors—influences their beliefs.

To do so, the first 2021 OSU survey asked students to think of "three people in your life whom you might consider role models. These should be people whom you might turn to for advice about choosing your college major or other aspects of planning for your schooling and eventual career." 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.[30] The options for their role models' major and occupation were the same groups of careers and majors that we focus on throughout. 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. 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 answers on the careers/majors of the people they know personally. In addition, the Freshman Survey asks students the careers (but not the majors) of their mother and their father.

Recall that Prediction 5 from Section 4 said that having a personal acquaintance with

[30] 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. In addition, it is far from clear, without the model, in what direction we should expect beliefs about one major to change in response to knowing someone with a *different* major.

career c should boost students' beliefs about the likelihood of that career unconditional on major. To test this prediction, Table 3 shows OLS estimates of the regression specification in equation 18:

$$\pi_c^i = \alpha + \beta RM_c^i + \mu_c + \epsilon_c^i \tag{18}$$

In equation 18, π_c^i is student i's belief the likelihood of career c. In column 1, this is OSU students' unconditional population beliefs (i.e., about the distribution of jobs nationwide). In column 2, it is OSU students' unconditional self beliefs (i.e., about their own chances of having each career). In column 3, it is a dummy variable for whether students in the Freshman Survey indicated c as their "probable" career. Next, RM_c^i is the number of role models (or, in the Freshman Survey, parents) that the student listed as having career c. Finally, μ_c are career fixed effects. All regressions cluster standard errors at the student level.

Consistent with Prediction 5, Column 1 of Table 3 shows that OSU students' unconditional population beliefs about c increase by 1.8 percentage points ($p < 0.01$) for each role model they have with that career. Recall that these survey questions asked students about the frequency of jobs among college graduates in the US as a whole; these results therefore suggest that knowing someone personally with a particular career (e.g., a student's mother is a doctor) boosts their beliefs about the frequency of that career in the national economy (e.g., that there are more doctors nationwide). Column 2 shows similar (and indeed, larger) effects on students' beliefs about their own eventual career, while column 3 shows a similar pattern for the careers that students in the Freshman Survey expect to attain. We caution that the effects on self beliefs (in columns 2 and 3) would, by themselves, be consistent with many interpretations: for instance, students may have a preference for having the same career as their parents, or personal connections might increase students' access to particular jobs. The effects on population beliefs in columns 1, in contrast, are not subject to such confounds.[31] Note also that because these regressions (and all others in Table 3) control for career fixed effects, these results are not driven by differences in the actual frequency of careers.

Recall that Predictions 3 and 4 from Section 4 said that students' beliefs about careers *conditional* on major should depend on both the career and major of their personal acquaintances. Prediction 3 said that a student who knows someone with career c and major M should have higher beliefs about the likelihood of c conditional on M. Prediction 4 said that knowing someone with career c but a major *other* than M has an ambiguous effect on

[31]None of these results are driven by a particular career or by non-employment. In particular, for each outcome, similar regressions (not shown) that exclude this outcome look qualitatively identical to Table 3.

beliefs about c conditional on M. In particular, if that career-major combination is very rare (as when c is not stereotypical of M), then the effect should be positive. In contrast, if that combination is sufficiently likely (as when c is stereotypical of M), the effect should be negative.

To test these predictions, we leverage the fact that the 2021 OSU survey asked students' their role model's college major in addition to their career. Table 4 shows OLS estimates of the following regression specification:

$$\pi^i_{c|M} = \alpha + \beta_1 RM^i_{c,M} + \beta_2 RM^i_{-c,M} + \beta_3 RM^i_{c,-M} + \mu_{c,M} + \epsilon^i_{c,M} \tag{19}$$

In equation 19, $\pi^i_{c|M}$ is student i's population belief (in columns 1-3) or self belief (columns 4-6) 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, $RM^i_{-c,M}$ the number with M but not c, and $RM^i_{c,-M}$ 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. As before, we cluster standard errors at the student level.

Column 1 of Table 4 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.

To test for the ambiguous effects described in Prediction 4, we split the data by whether career c is major M's most stereotypical outcome (columns 2 and 5) or not (column 3 and 6). We see that, as predicted, for stereotypical careers the effect is significantly negative: knowing someone with m's stereotypical career but who graduated with a different major reduces beliefs $\pi^i_{c|M}$ by 4.5 percentage points ($p < 0.01$) for population beliefs and by 3.9 percentage points for self beliefs ($p = 0.04$). In contrast, if c is *not* stereotypical of m, knowing someone with that career but a different major *boosts* beliefs $\pi^i_{c|M}$ by 0.6 percentage points ($p < 0.01$) for population beliefs and by 1.9 percentage points ($p < 0.01$) for self beliefs.[32]

[32]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. Like with role-model effects on unconditional beliefs, excluding any particular outcome (career or non-employment) from these regressions yields very similar similar estimates as in Table 4. Results available upon request.

6 Welfare and Policy Implications

6.1 Implications of Belief Biases for the Returns to Major

In this section, we turn to the question of how biases in beliefs about careers affect the perceived returns to majors. We focus this analysis on role of stereotypical thinking. To motivate this choice, we use a Shapley-Shorrocks decomposition of R^2 (Shorrocks, 1982) to quantify the relative contribution of each of the patterns in conditional beliefs that we explored in Sections 3 and 5 in explaining variation in the difference between students' beliefs and the true frequency of careers. These include overconfidence, selection, overweighting of small probabilities, as well as role model effects in the 2021 OSU data.[33] Intuitively, this captures the relative contribution of each set of regressors to the model's overall explanatory power. Table A.XII shows that a single variable—an indicator for whether a career is a major's most stereotypical outcome—explains 34-35% of the variance in students' errors: a greater proportion than any variable.[34]

We incorporate stereotypical thinking into a highly stylized model of major choice where students care about both the pecuniary and non-pecuniary career consequences of their choices. To motivate this approach, we start by giving evidence that the career beliefs we focus on appear to affect both the pecuniary and non-pecuniary perceived returns to major. First, we show that beliefs about careers are tightly linked to students' beliefs about salaries conditional on major, which have been the primary focus of most previous research on major choice (e.g., Betts 1996; Arcidiacono et al. 2012; Wiswall & Zafar 2015a; Conlon 2021). In addition to asking beliefs about the likelihood of careers, the 2020 OSU survey asked respondents about the distribution of salaries by major, both among others with that major and hypothetically for themselves if they graduated with each major (see Section B for details). We take the expected value of these distributions and call these students' "direct" salary beliefs. Column 1 of Table 5 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 5 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

[33] For further details, see the notes of Table A.XII.

[34] Because our 2021 OSU data only ask students about their two most likely majors, we do not analyze the role of selection in that sample.

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.

In addition to career beliefs being closely linked to salary beliefs, students appear to care about careers over and above the incomes they generate. While the average student in our 2020 OSU sample expects to earn about $5,000 more annually (in expectation) by pursuing their top-ranked major rather than their second-ranked major, the standard deviation of this measure is $26,000, indicating large heterogeneity. Indeed, a full 39% of students expect to earn less with their top rather than second-ranked major. Of those, the average difference in salary is $18,000. Consistent with previous work (e.g., Wiswall & Zafar 2015a), this indicates that students in our sample do not appear to choose their major simply to maximize future income.[35] In addition, the first 2021 OSU survey included a hypothetical question asking students how much *more* they would have to be paid annually to work in their second-ranked vs first-ranked career (See Appendix B for details). Half (49.8%) of students are willing to give up more than $20,000 per year to have their preferred career, which we take to be more direct evidence that many students' preferences over jobs depend on more than simply income.

We now turn to a stylized choice model. As before, assume that there are two majors, A and B.[36] Assume that major A is a specific major (e.g., Art) that leads to its associated stereotypical career a with (objective) probability $p_{a|A}$ but otherwise leads to a career b in the general sector. Major B, in contrast, is a general major (e.g., Business or economics) that leads to a career in the general sector b with certainty.[37] Note that, for simplicity, we are now assuming there are only two careers, a and b. Assume agent i's perceived expected utility from pursuing major M is given by equation 20.

[35] Note that students also perceive their personal chances of failing to graduate to be extremely low, no matter which major they choose to pursue, suggesting that they do not forgo more lucrative majors because they feel unable to do so.

[36] We ignore the supply side of providing majors and occupations. However, the welfare consequences of stereotypical thinking could be larger in a setting where institutions have an incentive to design programs to exploit such biases, as in the for-profit education sector (Deming et al., 2013).

[37] We assume that this is known to the agent. Strictly speaking, the beliefs model above would predict positive beliefs even for objectively zero-probability events as long as $\lambda > 0$. This reflects the idea that, even when the agent cannot think of anyone with a career-major pair, they may deem it somewhat plausible if they can think of people with similar career paths. We abstract away from this force in this section.

$$\widehat{EU}[M] = \pi_{a|M}(w_{a,M} + \psi) + (1 - \pi_{a|M})w_{b,M} + \nu_M \tag{20}$$

In equation 20, $\pi_{a|M}$ is i's belief about their likelihood of having career a after majoring in M (we omit i subscripts for readability), $w_{c,M}$ is the wage they expect to earn with career c and major M, and ν_M indicates utility from the non-labor-market aspects of majoring in M (difficulty/enjoyability of courses, parental approval, etc.). Finally, ψ denotes i's non-wage utility from having career a rather than b. We define the objective returns $EU[M]$ analogously but using i's objective probability $p_{c|M}$ of having each career c after majoring in M.

We focus on the case where $w_{b,B} > w_{b,A}$ such that i would earn more in the general sector if they pursue the general major. We also focus on the case where the agent is on the margin between the two majors, which happens when ψ satisfies equation 21.

$$\psi = \frac{1}{\pi_{a|A}}\Big(w_{b,B} - w_{b,A} + \nu_B - \nu_A\Big) + \Big(w_{b,A} - w_{a,A}\Big) \tag{21}$$

Equation 21 states that, when major B has higher wages ($w_{b,B} > w_{b,A}$), is less desirable for non-labor-market reasons ($\nu_B > \nu_A$), or has a lower-paying stereotypical job ($w_{b,A} > w_{a,A}$), the student must place higher non-pecuniary value on the stereotypical job to be on the margin between majors.

We can then define bias in returns to majoring in A for this marginal student according to equation 22.

$$Bias = \Big(\widehat{EU}[A] - \widehat{EU}[B]\Big) - \Big(EU[A] - EU[B]\Big) = (\pi_{a|A} - p_{a|A})\frac{w_{b,B} - w_{b,A} + \mu_B - \mu_A}{\pi_{a|A}} \tag{22}$$

Equation 22 shows that, unsurprisingly, the difference between perceived and objective returns is larger when i's beliefs are more distorted (i.e., when $\pi_{a|A} - p_{a|A}$ is large). Bias is also larger the greater the "riskier" major A is. This is true when the wage penalty between A and B conditional on working in the general sector ($w_{b,B} - w_{b,A}$) is large, and when the A's stereotypical job is unlikely ($\pi_{a|A}$ is small).[38] 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

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

the stereotypical job (i.e., ψ 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 8.

Prediction 8: *Biases 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.*

Note that Prediction 8 says that students exaggerate the returns to majors whose non-stereotypical careers are lower paying, not (necessarily) majors where the *gap* in salary between stereotypical and non-stereotypical careers (conditional on that major) is larger. To see the intuition for this result, we can imagine that $w_{a,A} = w_{b,A} < w_{b,B}$, such that wages conditional on majoring in A do not depend on occupation, but are lower than in major B. Equation 21 says that for a student to be on the margin between A and B, she must have a non-pecuniary preference for career a (else she would simply pick the higher-paying major B). Furthermore, the larger the gap in salaries between majors, the greater this non-pecuniary preference must be. Equation 22 then says that her bias about the expected utility difference between the two majors is larger the greater is the salary difference *across* majors and the lower her chances are of achieving the stereotypical career. Both of these comparative statics hold despite the fact that we are assuming there is no gap in salary *within* majors.

6.2 Which Majors are Worse for Stereotypical Thinkers?

To investigate which majors are "risky" according to the definition in Prediction 8, we look at the 20 most common majors from the American Community Survey. In addition, we separate out economics from business, because economics closely resembles the "general" major described in Section 6.1. Inspired by equation 22, we calculate two 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 $\pi_{a|A}$ in equation 22. 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.[39] Tables A.XIII and A.XIV give examples of the most

[39]The broad conclusions that we draw in this section are not sensitive to this exact 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.

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 $w_{b,A}$.

Figure 5 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 "fall-back" jobs (the top-left region of Figure 5). For example, only 1.3% of economics majors are "economists and market researchers," but economics has the highest paying "fall-back" jobs of any major. Math, physics, and business are also broadly "general" in this sense. Second, moving to the top-right region of Figure 5, STEM majors such as engineering, computer science, and biology all have relatively more common stereotypical jobs and have high 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 5 we see which majors are "risky" in the sense described above: majors with rare stereotypical jobs and low-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—perhaps due to GPA restrictions (Bleemer & Mehta, 2022)—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. This finding in part motivated our use of economics/business as the example of a "general" major in section 6.1.

A second worry is that stereotypical 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 stereotypical 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.[40] 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.

6.3 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 Survey of Household Economics and Decisionmaking to provide suggestive correlational evidence on post-college outcomes depending on students' field of study (see Appendix B for further details). Table 6 regresses various outcomes on a variable indicating whether a graduate's major is risky (which we define as humanities, psychology, art, communications, and social/behavioral sciences). We see that graduates with risky majors are 5% less likely 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; and are 35% more likely to report regretting their field of study (all comparisons significant at the $p < 0.05$ level). This correlational evidence—while merely suggestive—is consistent with a story in which stereotypical thinking leads some students to make worse human capital investments.

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

These analyses suggest that there may be welfare gains from shifting marginal students (in the sense described in Section 6.1) into less risky majors. Given the potential welfare consequences of stereotypical thinking, what types of policies might achieve this?

[40] 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 major's most stereotypical career.

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 the self and population versions of the question 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 major 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.[41] Figure A.IV 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.

At the end of the survey, both treatment and control groups were then re-asked the question about the percent chance that they would graduate with their two top majors.[42] All students also gave us permission to access their OSU transcript data, from which we calculated how many classes they took in each major during the Fall 2021 semester (pre-treatment) as well as Spring and Fall 2022 semesters (post-treatment).

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

[42] 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?"

When analyzing the results of this experiment, we divide the sample by whether students' top-ranked major is is "risky" in the sense described in Prediction 8. Among our 10 groups of college majors, these include humanities, psychology, art, and communications. Note that only 21% are pursuing one of these majors, reflecting the fact that Ohio State has particularly large business, engineering, and nursing schools (two thirds of students list one of these as their most likely major).

Columns 1-4 of Table 7 show OLS estimates where the dependent variable is students' updated belief about their likelihood of graduating with their top-ranked major. Columns 1 and 3 simply regress this updated belief on the original (pre-intervention) belief and a treatment dummy, restricting the data to students with risky and non-risky top majors, respectively. We see a statistically significant coefficient of -0.06 for students with a risky top major, indicating that the average such student reduced their perceived likelihood of graduating with that major by six percentage points ($p < 0.01$). We see a smaller, but still statistically significant effect of -0.03 for students with a non-risky top major ($p = 0.01$). Columns 2 and 4 then interact treatment status with a dummy variable indicating whether students had originally overestimated the fraction of graduates with their top major who are working in its stereotypical job (i.e., their population beliefs). We see in column 2 that for students with a risky top-ranked major, these effects are driven by those who had overestimated their major's stereotypical career. We do not see a similar pattern for those pursuing a less risky major (column 4).[43]

A common concern with survey-based measures of treatment effects is that participants may shade their survey answers toward how they believe the experimenters want them to answer. We attempted to mitigate such concerns first by including language explicitly saying that we did not care whether students updated their answers compared to the previous time they were asked the question, and instead just wanted their honest opinion. Second, we avoided any normative language about how students' ought to change their intentions in response to the information.[44] However, one set of results that may be particularly susceptible to such effects are the changes in students' beliefs about their own likelihood of having each

[43] Table A.XV shows similar regressions but splitting the sample by whether the student's top-ranked major is high vs low paying, rather than risky vs non-risky. We do not see large heterogeneity in effects using just earnings, which we view as evidence that the results presented here are not simply driven by differences in average pay across majors.

[44] As evidence that we were at least partially successful, a fair number of students expressed in free-response questions that they were pleasantly surprised to see that so few graduates with their top major go on to have stereotypical careers, because it showed how wide were the opportunities for graduates with that major. Because many students also had the reaction we expected—that fewer jobs in stereotypical careers made them more pessimistic about their top major—such responses also suggest that there may be underlying heterogeneity in treatment effects (even conditional on beliefs about the frequency of careers). If so, our estimates of the average treatment effect will understate the true effect of the information intervention.

career, which were elicited directly after (indeed, on the same page as) information about the fraction of others with those careers. Here, plausibly, there was a clear "right" direction in which respondents could infer that we anticipated their answers would move. We thus view the relationship depicted in Figure A.IV to be, if anything, an upper bound on the extent to which respondents updated their beliefs about their own future careers in response to the information about other people's careers.

We also estimate treatment effects on students' choices of classes in the semesters following the experiment. Columns 5 to 8 of Table 7 present similar regressions as columns 1 to 4 but where the dependent variable is the number of classes the student took in their top major's subject, adding together classes taken in Spring 2022 (the semester immediately post-intervention) and scheduled classes for Fall 2022. We do not see statistically significant average results for students pursuing risky or non-risky top majors (columns 5 and 7). That said, these estimates are somewhat imprecise, and we cannot rule out average effects as large as -0.65 and -0.49 for those with risky and non-risky top majors. However, column 6 shows significant heterogeneous effects for those pursuing risky majors depending on whether students over- or underestimated their major's stereotypical career. Such students take 1.5 more classes ($p = 0.05$) in their top major's subject if they underestimated this career (which only a minority, 20%, do). In contrast, we see that this effect is reduced by 1.8 classes ($p = 0.04$) for students who overestimated the stereotypical outcome. We see similarly signed but smaller and insignificant estimates for students pursuing less risky majors (column 8). Thus, while estimated treatment effects from the classes data are imprecise, we see some evidence in line with students moving away from majors whose stereotypical outcomes they overestimated, especially for risky majors.

To summarize, our within-survey measures suggest that students indeed respond to population information by revising their beliefs about their own potential careers, but this elasticity is well below one. Consistent with Prediction 8, the information has a larger effect on intended major for students who are considering a risky major and who particularly overestimate its stereotypical outcome. Our estimated effects on class choices are imprecise but some specifications show qualitatively consistent conclusions.

7 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 pattern of stereotypical thinking is a natural consequence of a simple model of intuitive belief formation, wherein students form expectations by trying to call to mind examples of people with various majors and careers. In addition, the model makes additional predictions—which we confirm empirically—both about average beliefs and about how about students' beliefs should systematically depend on the careers and majors of people they know personally. Finally, we show that attempting to correct stereotypical thinking through a light-touch information intervention has statistically significant, but somewhat modest, effects on students' intended and actual choices about what to study in college.

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, the agent forms their beliefs about the question at hand "on the fly," using a combination of retrieval-from-memory and mental simulation. 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 simply fail to come to mind (especially if it 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).[45] *Ex ante*, of course, rational mechanisms could have fully explained these patterns:

[45]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.

e.g., students with correct beliefs might rationally pursue certain career paths which, though very unlikely to pan out, they feel are worth the risk (e.g., journalism or film), or students may realize that certain majors (e.g., psychology) provide a general education not intended for use in any particular sector. Our findings suggest that mistaken beliefs may also contribute to these patterns: certain fields of study may appear especially appealing because 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., 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. (2022). College Major Restrictions and Student Stratification. *Working Paper*.
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.

Deming, D., Goldin, C., & Katz, L. (2013). *For-profit colleges* (Vol. 23).

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*.

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

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*.

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.

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*.

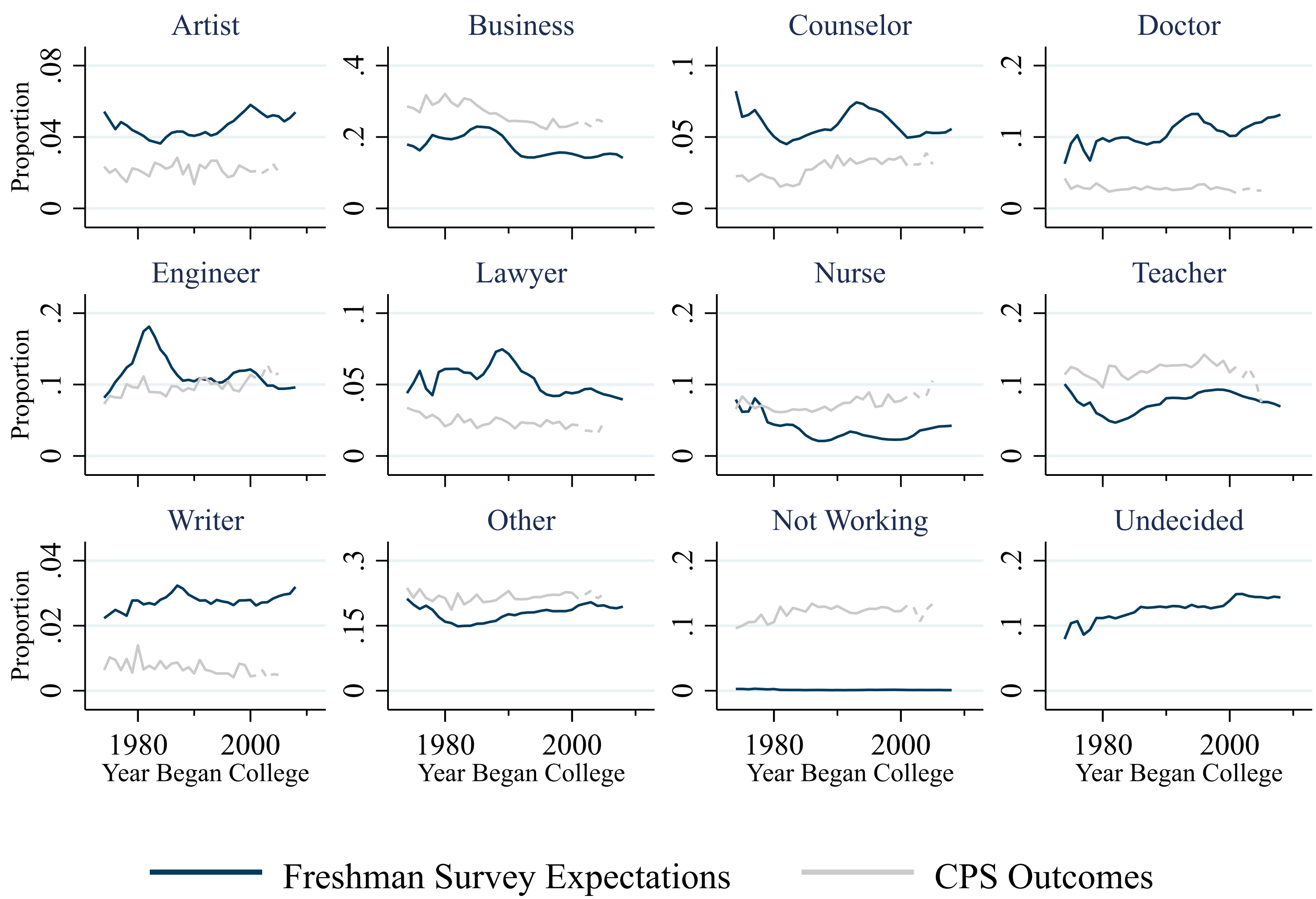
Wiswall, M., & Zafar, B. (2015a). Determinants of college major choice: Identification using an information experiment. *The Review of Economic Studies*, *82*(2), 791–824.

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.

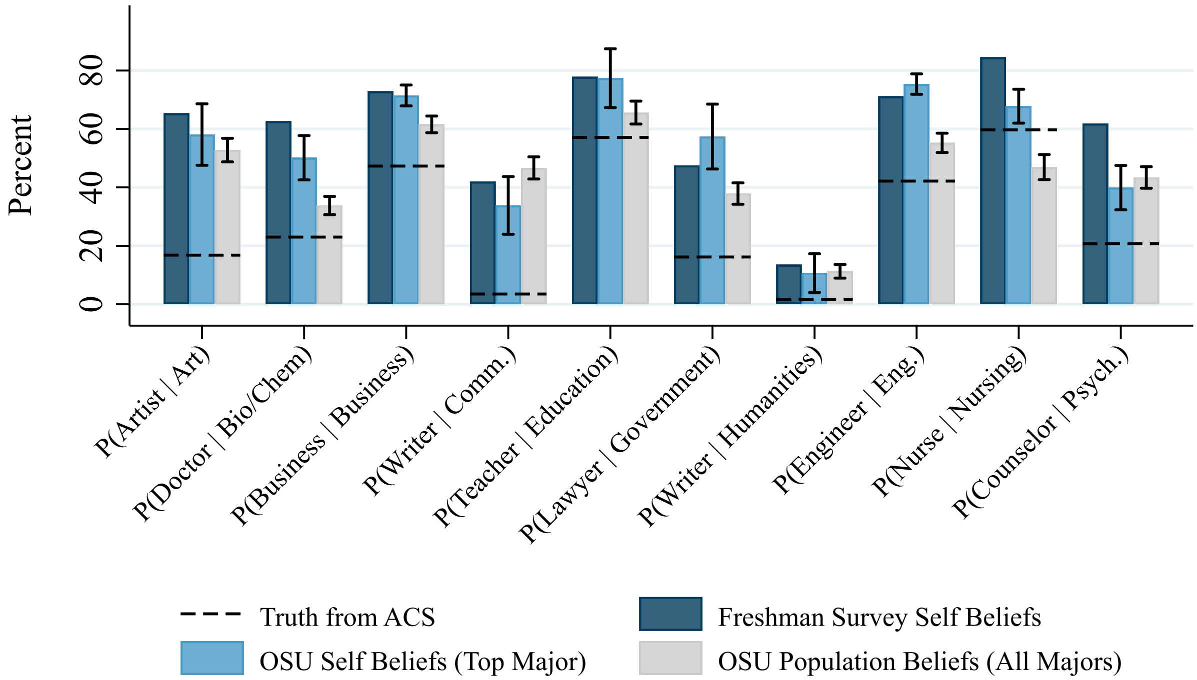
Wiswall, M., & Zafar, B. (2021). Human capital investments and expectations about career and family. *Journal of Political Economy*, *129*(5).

Figure 1: Career Expectations vs. Outcomes Over Time



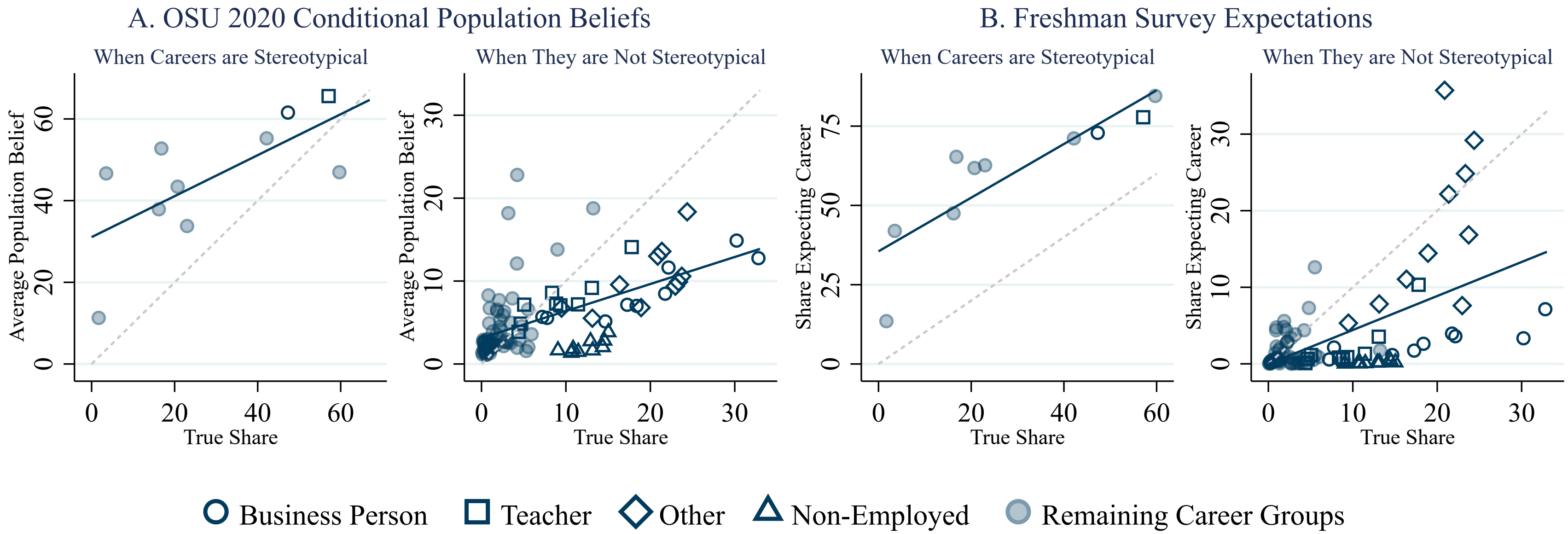
Notes: Figure 1 compares the share of first-year undergraduates by the year they began college (birth year plus 18) who expect to have a career in each occupation (blue line) according to the Freshman Survey data along with the share of college graduates (gray line) aged 33 to 37 who work in that occupation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



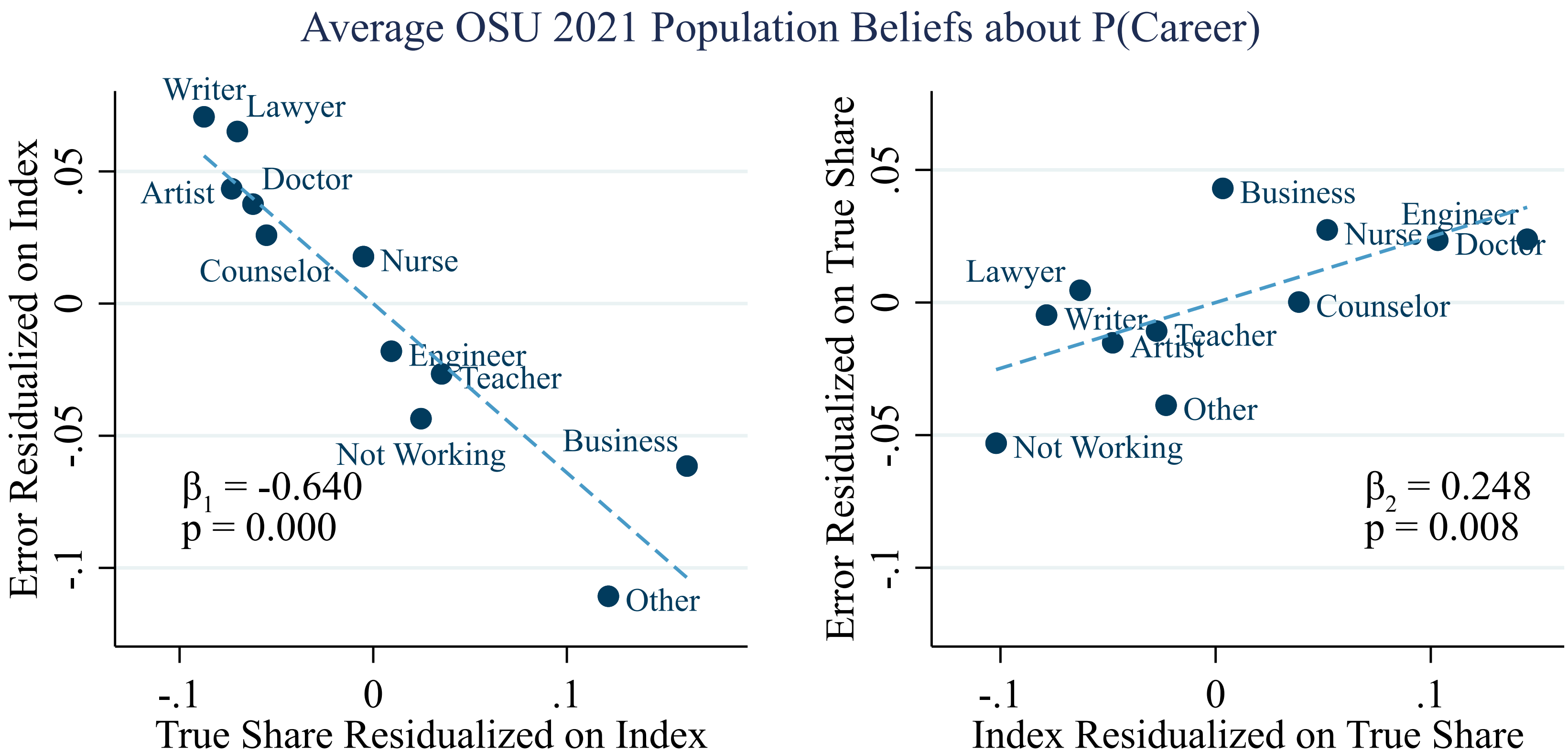
Notes: Figure 2 presents average statistics regarding the most stereotypical profession (as defined in section 3) for each major. The dashed horizontal lines denote the actual proportion of college graduates with each major between the ages of 30 and 50 that are working in that major's most stereotypical 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 sample about the probability that they would be working in each career at age 30 if they graduated from Ohio State with each 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 that are working in each occupation. Error bars show 95% confidence intervals for the mean of the OSU beliefs.

Figure 3: Which Outcomes do Stereotypical Thinkers Neglect?



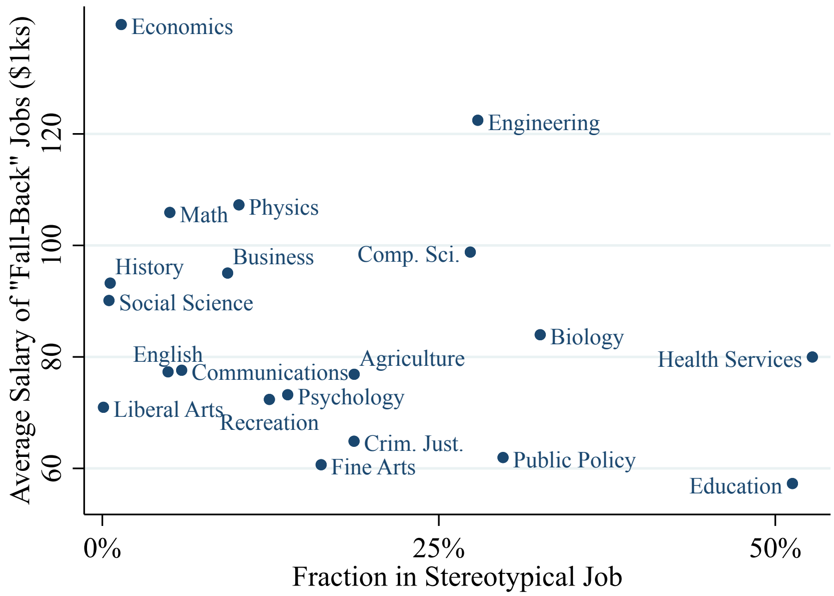
Notes: Each dot in Figure 3 represents a career-major pair (where non-employment is also one the "careers"). The left graph of each panel restricts these pairs to when the career is most stereotypical of the major, and the right graph restricts them to when the career is not stereotypical of the major. 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 of Panel A is the average population belief, among the 2020 OSU sample, about the fraction of graduates with that major who are working in that career. The y-axis of Panel B is the fraction of students in the Freshman Survey who report the career as their probable career, conditional on reporting that major as their probable major. Lines show OLS regressions including all career-major pairs within each graph.

Figure 4: Testing the Model's Predictions for Average Unconditional Beliefs



Notes: Figure 4 shows partial correlations between "errors" in unconditional beliefs and the true share of college graduates working in each career, as well as between errors in unconditional beliefs and a Herfindahl–Hirschman Index (HHI) measuring how concentrated each career is within majors. In Panel A, errors in beliefs are the difference between 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). In Panel B, errors in beliefs are the difference between the fraction of students in the Freshman Survey who expect each career and the share of college graduates of the same cohorts who are working in that career at age 33-37 (calculated using the 1976-2020 CPS). Each graph plots two of the three variables (error, true share, and HHI) residualized on the third variable.

Figure 5: Which Majors are Risky for Stereotypical Thinkers?



Notes: The x-axis of Figure 5 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.

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

	Self Beliefs	Population Beliefs		
	(1)	(2)	(3)	(4)
P(Career | Major)	0.51	0.38	0.39	0.51
	(0.08)	(0.07)	(0.09)	(0.12)
$\mathbb{1}$(Most Stereotypical)	0.43	0.39	0.32	0.29
	(0.04)	(0.03)	(0.05)	(0.04)
Constant	0.01	0.02	0.03	0.02
	(0.01)	(0.01)	(0.01)	(0.01)
Observations	8,305	8,305	33,220	33,220
Individuals	110	110	110	110
R^2	0.65	0.63	0.45	0.67
Individual X Career Fixed Effects	No	No	No	Yes
Included Majors	Top	Top	All	All

Notes: Table 2 presents OLS estimates of variants of equation 1 using the 2020 OSU sample. The dependent variable is students' beliefs about the likelihood of a career conditional on a major. The dependent variable in column 1 is students' "Self-beliefs," i.e., their belief of the percent chance that they would have that career if they graduated with that major. The dependent variable in columns 2-4 is students' "population beliefs" about the fraction of graduates with that major who are working in that occupation. Columns 1-2 include only the major that each student thought they were mostly likely to graduate with. Columns 3-4 include all four majors that students answered 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.

Table 3: Role Models and What Comes to Mind: Unconditional Beliefs

	2021 Ohio State		**Freshman Survey**
	Population Beliefs	Self Beliefs	
	(1)	(2)	(3)
RM_c^i	1.78	7.43	4.19
	(0.30)	(0.62)	(0.01)
Constant	8.80	7.87	7.64
	(0.05)	(0.11)	(0.00)
Observations	9,834	9,834	107,752,344
Individuals	894	894	8,979,362
R^2	0.19	0.17	0.04
Career-by-Major Fixed Effects	Yes	Yes	Yes

Notes: Table 3 presents OLS estimates of equation 18. The dependent variable in column 1 is students' unconditional population belief for the 2021 OSU sample: that is, the student's belief of the fraction of college graduates (unconditional on major) who are working in each career. The dependent variable in column 2 is the corresponding self beliefs for each career among the 2021 OSU sample: that is, the student's perceived percent chance that they will be working in each career (unconditional on their major). The dependent variable in column 3 is a dummy variable indicating whether a student in the Freshman Survey data listed an occupation as their probable career. RM_c^i indicates how many role models the student listed as having that career. In the Freshman Survey, these role models simply correspond to their parents. All regressions cluster standard errors at the individual level.

Table 4: Role Models and What Comes to Mind: Conditional Beliefs

	Population Beliefs			Self Beliefs		
	All	S	NS	All	S	NS
	(1)	(2)	(3)	(4)	(5)	(6)
$RM_{c,m}$	3.07	6.16	1.13	3.69	7.71	1.22
	(0.58)	(1.22)	(0.45)	(0.67)	(1.33)	(0.61)
$RM_{-c,m}$	-0.29	-0.85	-0.25	-0.29	0.20	-0.31
	(0.06)	(1.05)	(0.09)	(0.07)	(1.16)	(0.10)
$RM_{c,-m}$	0.30	-4.51	0.59	1.58	-3.93	1.92
	(0.16)	(1.58)	(0.15)	(0.26)	(1.88)	(0.25)
Constant	9.02	47.17	5.16	8.70	50.59	4.46
	(0.04)	(0.88)	(0.10)	(0.07)	(0.97)	(0.11)
Observations	19,668	1,788	17,880	19,668	1,788	17,880
Individuals	894	894	894	894	894	894
R^2	0.60	0.20	0.20	0.57	0.28	0.19
Career-by-Major Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Table 3 presents OLS estimates of equation 19. 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 4 to 6 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 1 and 4 include all career-major pairs. Columns 2 and 5 restrict the sample to career-major pairs in which the career is that major's most stereotypical career (S). Columns 3 and 6 restrict the sample to all career-major pairs where the career is not the most stereotypical (NS) of the major. $RM_{c,m}$ 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 do not have career c but do have 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.

Table 5: 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 5 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. See Section 6.1 for more details. All regressions cluster standard errors at the individual level.

Table 6: Long-Run Outcomes for Graduates with Risky vs Non-Risky Majors

	(1) Employed	(2) Dissatisfied	(3) Related Job (RJ)	(4) RJ Not Available	(5) Salary ($1ks)	(6) Debt ($1ks)	(7) Regret
Risky Major	-0.042	0.020	-0.100	0.036	-17.918	2.904	0.128
	(0.009)	(0.009)	(0.013)	(0.007)	(1.856)	(0.628)	(0.032)
Constant	0.893	0.116	0.585	0.055	80.408	10.445	0.368
	(0.004)	(0.005)	(0.007)	(0.003)	(0.967)	(0.322)	(0.014)
N	44,498	40,574	40,574	40,574	40,499	44,498	1,633

Notes: Table 6 presents OLS regressions using the 2013 round of the National Survey of College Graduates (columns 1-6) and the Survey of Household Economics and Decisionmaking (column 7). See section B.4 for further details on the data and variable construction. “Risky” is a dummy variable indicating whether the graduate majored in a risky major, following the definition in section 6.4. The risky majors are Humanities, Psychology, Art, andCommunications. The dependent variables in columns (1) through (7) correspond, respectively, to: whether the graduate is employed; whether she expresses that she is dissatisfied with her job; whether she reports that her job is related to her highest degree; whether she reports she has an unrelated job because related jobs are unavailable; her annual salary; her total student debt balance; and whether she regrets the field of her highest degree. Robust standard errors in parentheses.

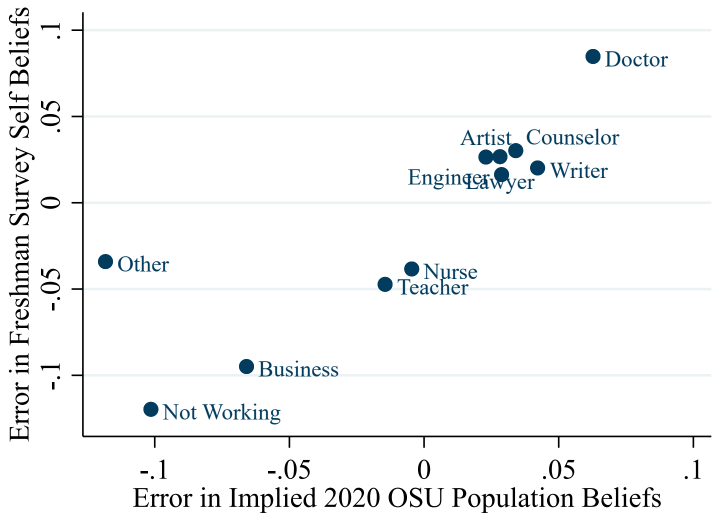
Table 7: The Effect of an Information Intervention

	Post-Treatment P(m)				Post-Treatment Classes			
	Risky		Non-Risky		Risky		Non-Risky	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Pre-Treatment P(m)	0.970	0.973	0.955	0.952				
	(0.039)	(0.038)	(0.021)	(0.022)				
Pre-Treatment Classes					0.530	0.509	0.794	0.768
					(0.195)	(0.190)	(0.112)	(0.113)
Treatment	-0.058	0.006	-0.027	-0.032	0.045	1.519	-0.160	0.082
	(0.019)	(0.056)	(0.011)	(0.017)	(0.353)	(0.772)	(0.170)	(0.252)
Overestimated Stereotypical Career		-0.004		0.007		1.111		0.740
		(0.027)		(0.014)		(0.382)		(0.236)
Treatment X Overestimated		-0.081		0.011		-1.811		-0.375
		(0.060)		(0.022)		(0.864)		(0.340)
Observations	168	168	646	646	135	135	502	502
Control Group Mean	.56	.56	.68	.68	2.4	2.4	2.5	2.5

Notes: Table 7 presents OLS regressions including data from the 2021 OSU sample. The dependent variable in columns 1-4 is students' post-intervention belief about the percent chance that they will graduate with their top-ranked major. The dependent variable in columns 5-8 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 1-2 and 5-6 restrict the data to those with "risky" top-ranked majors as defined in section 6.1: that is, humanities, psychology, communications, or art. Columns 3-4 and 7-8 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. Robust standard errors in parentheses.

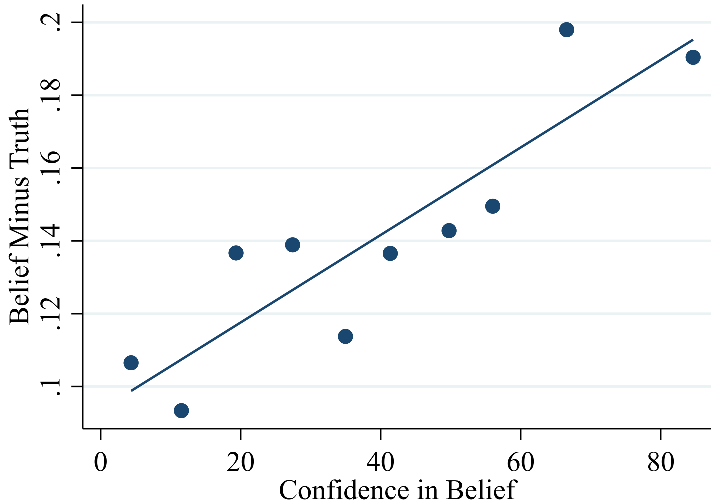
A Supplementary Figures and Tables

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



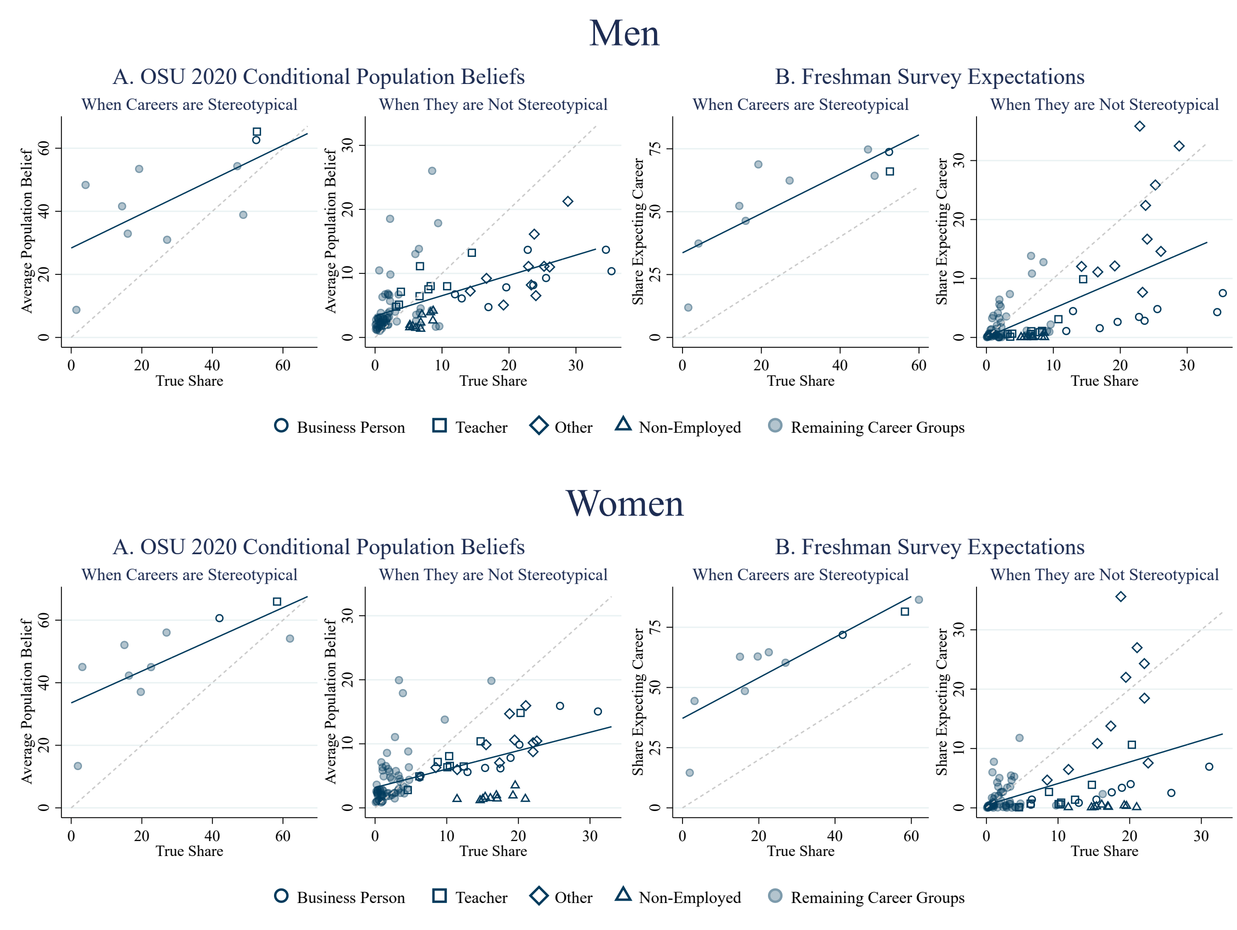
Notes: The y-axis in Figure A.I is the difference between the fraction of students in the Freshman Survey who list each occupation as their probable career and the fraction of 33-37 year old college graduates in the CPS (of the same cohorts, up to birth year 1987) who are working in each occupation. 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 3.4 for further details on the construction of this statistic.

Figure A.II: More Confident Students' Stereotype More



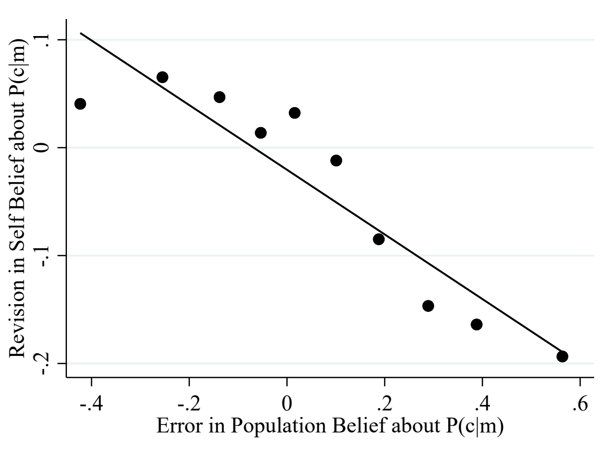
Notes: The y-axis in Figure A.I is the difference between the 2021 OSU students beliefs about the fraction of graduates (in the two majors they were asked about) working in their major's stereotypical career. The y-axis is how confident (on a 0 to 100 scale) they reported being in their answer. Data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

Figure A.III: Which Outcomes do Stereotypical Thinkers Neglect? By Gender



Notes: Each dot in Figure A.III represents a career-major pair (where non-employment is also one the "careers"). Within panels, the left graph restricts these pairs to when the career is most stereotypical of the major, and the right graph 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 of Panel A is the average population belief, among the 2020 OSU sample, about the fraction of graduates with that major who are working in that career. The y-axis of Panel B is the fraction of students in the Freshman Survey who report the career as their probable career, conditional on reporting that major as their probable major. The top panel restricts the beliefs data to men and calculates the true share in each career using only men. The bottom panel restricts the beliefs data to women and calculates the true share in each career using only women. Note that the OSU survey questions asked about college graduates of *any* gender (not just respondents' own gender). Lines show OLS regressions including all career-major pairs within each graph.

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



Notes: The x-axis of Figure A.IV is the error of students in the treatment group of the 2021 OSU survey about the fraction of graduates with their top-ranked major who are working in that major's stereotypical career. 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 error. Line shows a line of best fit estimated by OLS.

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: Testing for Stereotypical Thinking

	Female Students				Male Students			
	Self Beliefs	Population Beliefs			Self Beliefs	Population Beliefs		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
P(Career | Major)	0.45	0.28	0.36	0.53	0.46	0.41	0.38	0.53
	(0.04)	(0.04)	(0.02)	(0.04)	(0.05)	(0.04)	(0.02)	(0.06)
$\mathbb{1}$(Most Stereotypical)	0.43	0.40	0.34	0.30	0.46	0.38	0.30	0.26
	(0.02)	(0.02)	(0.01)	(0.01)	(0.03)	(0.02)	(0.01)	(0.02)
Constant	0.02	0.03	0.03	0.02	0.01	0.02	0.03	0.02
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Observations	4,301	4,301	17,204	17,204	4,004	4,004	16,016	16,016
Individuals	391	391	391	391	364	364	364	364
R^2	0.61	0.59	0.50	0.70	0.69	0.68	0.41	0.65
Individual X Career Fixed Effects	No	No	No	Yes	No	No	No	Yes
Included Majors	Top	Top	All	All	Top	Top	All	All

Notes: Table A.VIII presents OLS regressions where the dependent variable is students beliefs about the likelihood of a career conditional on a major. The dependent variable in columns 1 and 5 are students' "Self-beliefs," i.e., their belief of the percent chance that they would have that career if they graduated with that major. The dependent variable in columns 2-4 and 6-8 are students' "population beliefs" about the fraction of graduates with that major who are working in that occupation. Columns 1-2 and 5-6 include only the major that each student thought they were mostly likely to graduate with. Columns 3-4 and 7-8 include all four majors that students answered about. "P(Career | Major)" is the true fraction of graduates *of the same gender as the student* 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. Columns 1-4 include only female students, and columns 5-8 include only male students. All regressions cluster standard errors at the individual level and career-by-major level.

Table A.IX: Robustness of Test for Stereotypical Thinking

	Self Beliefs	Population Beliefs		
	(1)	(2)	(3)	(4)
$\mathbb{1}$(Most Stereotypical)	0.27	0.18	0.18	0.16
	(0.04)	(0.04)	(0.05)	(0.03)
Constant	-0.02	-0.02	-0.02	-0.01
	(0.01)	(0.01)	(0.01)	(0.00)
Observations	8,305	8,305	33,220	33,220
Individuals	110	110	110	110
R^2	0.25	0.15	0.12	0.53
Individual X Career Fixed Effects	No	No	No	Yes
Included Majors	Top	Top	All	All

Notes: Table A.IX presents OLS regressions where the dependent variable is students beliefs about the likelihood of a career conditional on a major minus the true fraction of graduates with that major who have that career. In column 1 they are students' "Self-beliefs," i.e., their belief of the percent chance that they would have that career if they graduated with that major. In columns 2 to 4 they are students' "population beliefs" about the fraction of graduates with that major who are working in that occupation. Columns 1 and 2 include only the major that each student thought they were mostly likely to graduate with. Columns 3 and 4 include all four majors that students answered about. $\mathbb{1}$(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 career-by-major level.

Table A.X: Testing for Stereotypical Thinking

	Self Beliefs	Population Beliefs		
	(1)	(2)	(3)	(4)
P(Career | Major)	1.49	1.23	1.09	0.93
	(0.17)	(0.17)	(0.13)	(0.21)
P(Career | Other Majors)	-0.91	-0.74	-0.76	-2.37
	(0.17)	(0.16)	(0.15)	(1.35)
Constant	0.04	0.05	0.06	0.22
	(0.01)	(0.01)	(0.01)	(0.14)
Observations	8,305	8,305	33,220	33,220
Individuals	110	110	110	110
R^2	0.59	0.55	0.36	0.62
Individual X Career Fixed Effects	No	No	No	Yes
Included Majors	Top	Top	All	All

Notes: Table A.X presents OLS regressions where the dependent variable is students beliefs about the likelihood of a career conditional on a major. The dependent variable in column 1 is students' "Self-beliefs," i.e., their belief of the percent chance that they would have that career if they graduated with that major. The dependent variable in columns 2-4 is students' "population beliefs" about the fraction of graduates with that major who are working in that occupation. Columns 1-2 include only the major that each student thought they were mostly likely to graduate with. Columns 3-4 include all four majors that students answered 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. "P(Career | Other Majors)" is the true fraction of graduates with *any other* major that are working in that career, also calculated from the 2017-2019 American Community Survey. All regressions cluster standard errors at the individual level and career-by-major level.

Table A.XI: 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.XI 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.

Table A.XII: 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.XII 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 23 and 24 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 23 and only the 2021 OSU sample for equation 24.

$$\begin{aligned} Y_{i,c,m,p} - T_{c,m} = {} & \psi_{c,p,Top(i,m)} + \beta_1 Self_{i,p} + \beta_2 t_{i,m} + \beta_3 Self_{i,p} \times Top_{i,m} + \beta_4 Ster_{m,c} + \beta_5 Ster_{m,c} \times Self_{i,p} + \\ & \beta_6 Ster_{m,c} \times Top_{i,m} + \beta_7 Ster_{m,c} \times Self_{i,p} \times Top_{i,m} + \beta_8 T_{c,m} + \beta_9 T_{c,m} \times Self_{i,p} + \beta_{10} T_{c,m} \times Top_{i,m} \\ & + \beta_{11} T_{c,m} \times Self_{i,p} \times Top_{i,m} + \varepsilon_{i,c,m,p} \end{aligned} \tag{23}$$

$$\begin{aligned} Y_{i,c,m,p} - T_{c,m} = {} & \psi_{c,p} + \beta_1 Self_{i,p} + \beta_2 t_{i,m} + \beta_4 Ster_{m,c} + \beta_5 Ster_{m,c} \times Self_{i,p} + \beta_8 T_{c,m} + \beta_9 T_{c,m} \times Self_{i,p} + \\ & \beta_{1,2} RM_{i,c,m} + \beta_{1,3} RM_{i,c,m} \times Self_{i,p} + \varepsilon_{i,c,m,p} \end{aligned} \tag{24}$$

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.XIII: 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.XIII and A.XIV give, for each of the 20 largest majors in the American Community Survey, 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.XIV: 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.XIII and A.XIV give, for each of the 20 largest majors in the American Community Survey, 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.XV: The Effect of an Information Nudge: Low- vs High-Paying Majors

	Post-Treatment P(m)				Post-Treatment Classes			
	Low Paying		High Paying		Low Paying		High Paying	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Pre-Treatment P(m)	0.961	0.956	0.961	0.957				
	(0.038)	(0.038)	(0.021)	(0.022)				
Pre-Treatment Classes					0.621	0.575	0.712	0.707
					(0.154)	(0.159)	(0.132)	(0.132)
Treatment	-0.025	0.007	-0.038	-0.051	0.130	0.563	-0.299	-0.064
	(0.016)	(0.029)	(0.011)	(0.019)	(0.222)	(0.296)	(0.207)	(0.354)
Overestimated Stereotypical Career		0.007		0.003		0.802		0.668
		(0.024)		(0.012)		(0.268)		(0.310)
Treatment X Overestimated		-0.055		0.023		-0.710		-0.320
		(0.034)		(0.024)		(0.427)		(0.437)
Observations	320	320	494	494	260	260	377	377
Control Group Mean	.6	.6	.69	.69	1.9	1.9	2.8	2.8

Notes: Table A.XV presents OLS regressions including data from the 2021 OSU sample. The dependent variable in columns 1-4 is students' post-intervention belief about the percent chance that they will graduate with their top-ranked major. The dependent variable in columns 5-8 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 1-2 and 5-6 restrict the data to those with lower-paying top-ranked majors, which we define as including humanities, psychology, communications, art, nursing, and education. Columns 3-4 and 7-8 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. Robust standard errors in parentheses.

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.[46]

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.[47] 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.

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 ap-

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

[47]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.

ply 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.[48] 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.

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".[49] However, in open-ended feedback the overwhelming reason given was that they took longer to fill out than multiple choice questions would have.[50] In addition, all of the main results described in Section 3.3 are nearly identical for students who did and did not report finding these questions difficult to answer.

B.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 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

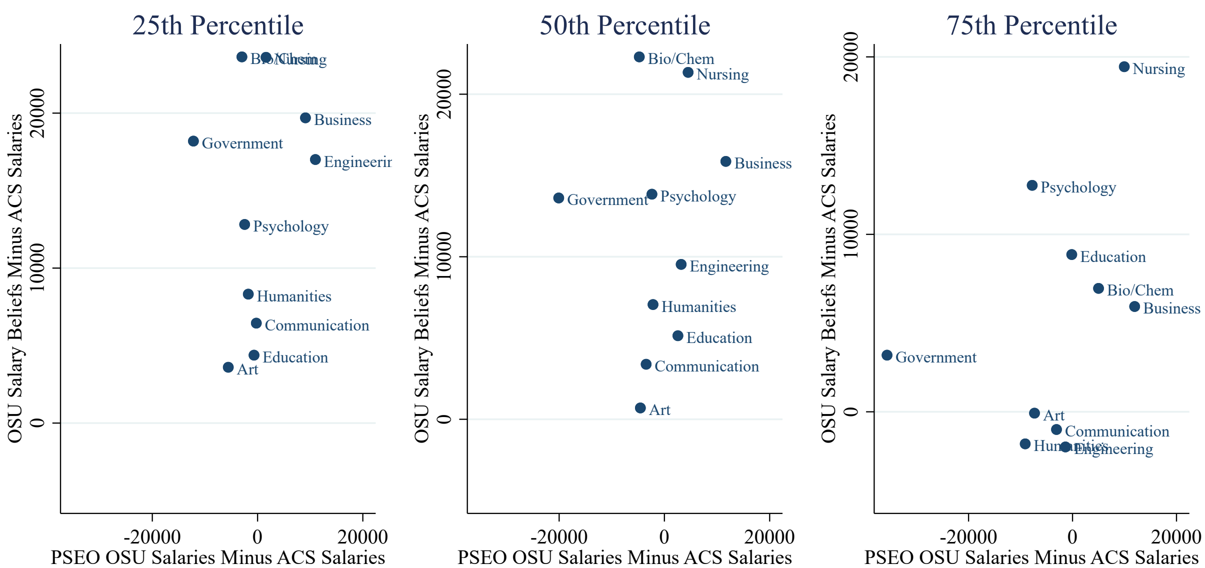
[48]For the highest bin ("greater than $150,000"), we simply assume a maximum salary of $180,000.

[49]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.

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

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.

In addition to the career questions already mentioned, the 2020 OSU sample was asked about the distribution of salaries by major, both among others with that major and hypothetically for themselves if they graduated with each major (see Appendix B for details). These questions allow us to compute the 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.[51] 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.3 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.

The first 2021 survey included questions meant to measure how much students' care about the non-pecuniary aspects of careers. In particular, it offered students a series of hypothetical choices that asked them to imagine they could choose between earning a fixed wage for their whole career in either of the two careers they said they were most likely to be working in. The option on the left (randomized to be their top or second-ranked career) always offered \$90,000 per year. The option on the right offered between \$30,000 and \$150,000, in \$20,000 increments.

B.4 National Survey of College Graduates

We use the 2013 round of the National Survey of College Graduates conducted by the U.S. Census Bureau. We assign majors to the same 10 major groups that we use elsewhere in the paper. We focus on several survey questions. First, "How would you rate your overall satisfaction with the principal job you held during the week of February 1, 2013?" We code this into a binary variable equal to 1 if the respondent answered "Very Satisfied" or "Somewhat Satisfied" as opposed to "Somewhat Dissatisfied" or 'Very Dissatisfied". Second, "To what extent was your work on your principal job related to your highest degree?" We code this into a binary variable equal to 1 if the respondent answered "Closely Related" as opposed to "Somewhat Related" or "Not Related". Those who answered "Not Related"

[51]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.XVI 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.XVI: 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.XVI 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.

were then asked "Did these factors influence your decision to work in an area outside the field of your highest degree?" We code this into a binary variable equal to 1 if the respondent answered "Job in highest degree field not available" as one of the reasons, and zero if they did not or they said their job was closely or somewhat related in the earlier question. Finally, respondents were asked among the reasons they gave to the previous question, which was the most important. We construct a binary variable equal to 1 if "Job in highest degree field not available" was the top reason given and 0 if it was not or if the respondent said their job was closely or somewhat related in the earlier question. With these outcome variables Y_i, we run the regression in equation 25, where $Risky_i$ indicates whether respondent i graduated with a BA in a risky major. Following section 6.4, low probability majors are Humanities, Biology/Chemisty, Psychology, Art, Communications, and Government. High paying majors are Business, Engineering, Biology/Chemistry, and Government.

$$Y_i = \alpha + \beta_1 HighProb_i + \beta_2 LowPay_i + \beta_3 HighProb_i \times LowPay_i + \varepsilon \tag{25}$$

C Proofs

Here we provide derivations of the results in Section 4. 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:

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

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 1, 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,

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

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

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

We can evaluate this derivative at the rational benchmark:

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

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

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:

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

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:

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

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 2, 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 $\phi = 0$:

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

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

To yield Prediction 3, 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.

To yield Prediction 4, 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,

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

Next,

$$\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,

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

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:

$$
\begin{aligned}
F(H_a) = 1/\Bigg(& (p_{a,A} + \phi)(p_{a,A} + \delta_M p_{a,B} + \phi) + p_{b,A}(\delta_c p_{a,A} + \delta_c \delta_M p_{a,B} + \delta_c \phi) + \sum_{z \notin \{a,b\}} p_{z,A}(\delta_c p_{a,A} + \delta_c \delta_M p_{a,B} + \delta_c \phi \\
& + p_{a,B}(\delta_M p_{a,A} + p_{a,B} + \delta_M \phi) + p_{b,B}(\delta_c \delta_M p_{a,A} + \delta_c p_{a,B} + \delta_c \delta_M \phi) + \sum_{z \notin \{a,b\}} p_{z,B}(\delta_c \delta_M p_{a,A} + \delta_c p_{a,B} + \delta_c \delta_M \phi) \\
& \times \Bigg((p_{a,A} + \phi)(p_{a,A} + \delta_M p_{a,B} + \phi) + \eta_c p_{b,A}(\delta_c p_{a,A} + \delta_c \delta_M p_{a,B} + \delta_c \phi) + \eta_c \sum_{z \notin \{a,b\}} p_{z,A}(\delta_c p_{a,A} + \delta_c \delta_M p_{a,B} + \delta \\
& + p_{a,B}(\delta_M p_{a,A} + p_{a,B} + \delta_M \phi) + \eta_c p_{b,B}(\delta_c \delta_M p_{a,A} + \delta_c p_{a,B} + \delta_c \delta_M \phi) + \eta_c \sum_{z \notin \{a,b\}} p_{z,B}(\delta_c \delta_M p_{a,A} + \delta_c p_{a,B} + \delta_c \delta \\
F(H_b) = 1/\Bigg(& (p_{a,A} + \phi)(\delta_c p_{b,A} + \delta_c \delta_M p_{b,B}) + p_{b,A}(p_{b,A} + \delta_M p_{b,B}) + \sum_{z \notin \{a,b\}} p_{z,A}(\delta_c p_{b,A} + \delta_c \delta_M p_{b,B}) \\
& + p_{a,B}(\delta_c \delta_M p_{b,A} + \delta_c p_{b,B}) + p_{b,B}(\delta_c p_{b,A} + p_{b,B}) + \sum_{z \neq b} p_{z,B}(\delta_c \delta_M p_{b,A} + \delta_c p_{b,B}) \Bigg) \\
& \times \Bigg(\eta_c (p_{a,A} + \phi)(\delta_c p_{b,A} + \delta_c \delta_M p_{b,B}) + p_{b,A}(p_{b,A} + \delta_M p_{b,B}) + \eta_c \sum_{z \notin \{a,b\}} p_{z,A}(\delta_c p_{b,A} + \delta_c \delta_M p_{b,B}) \\
& + \eta_c p_{a,B}(\delta_c \delta_M p_{b,A} + \delta_c p_{b,B}) + p_{b,B}(\delta_c p_{b,A} + p_{b,B}) + \eta_c \sum_{z \neq b} p_{z,B}(\delta_c \delta_M p_{b,A} + \delta_c p_{b,B}) \Bigg)
\end{aligned}
$$

First, we derive Prediction 5:

$$\frac{\partial}{\partial\phi}\log F(H_a) = \frac{2(p_{a,A}+\phi+\delta_M p_{a,B})+\eta_c\Big(\delta_c p_{b,}}{(p_{a,A}+\phi)(p_{a,A}+\delta_M p_{a,B}+\phi)+p_{a,B}(\delta_M p_{a,A}+p_{a,B}+\delta_M\phi)+\eta_c\Big(p_{b,A}(\delta_c p_{a,A}+\delta_c\delta_M p_{a,B}+\delta_c\phi)+}$$
$$-\frac{2(p_{a,A}+\phi+\delta_M p_{a,B})+\Big(\delta_c p_{b,A}}{(p_{a,A}+\phi)(p_{a,A}+\delta_M p_{a,B}+\phi)+p_{a,B}(\delta_M p_{a,A}+p_{a,B}+\delta_M\phi)+\Big(p_{b,A}(\delta_c p_{a,A}+\delta_c\delta_M p_{a,B}+\delta_c\phi)+\sum}$$

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.

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

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

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

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

$$- \frac{2(p_{a,B} - p_{a,A})(1-\delta_M) + \delta_c (1-\delta_M) \sum_{z \notin \{a\}} (p_{z,B} - \ldots}{p_{a,A}(p_{a,A} + \delta_M p_{a,B}) + p_{b,A}(\delta_c p_{a,A} + \delta_c \delta_M p_{a,B}) + \sum_{z \notin \{a,b\}} p_{z,A}(\delta_c p_{a,A} + \delta_c \delta_M p_{a,B}) + p_{a,B}(\delta_M p_{a,A} + p_{a,B}) + \ldots}$$

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.

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

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

$$- \frac{\delta_c (1-\delta_M)(p_{b,B} - p_{b,A})}{\eta_c(p_{a,A} + \phi)(\delta_c p_{b,A} + \delta_c \delta_M p_{b,B}) + p_{b,A}(p_{b,A} + \delta_M p_{b,B}) + \eta_c \sum_{z \notin \{a,b\}} p_{z,A}(\delta_c p_{b,A} + \delta_c \delta_M p_{b,B}) + \eta_c p_{a,B}(\delta_c \delta_M \ldots}$$

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 7, let $\phi = 0$ (shutting down personal acquaintances) and let $p_{A|z} =$

K for all z. Then

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

which is true when $p_b > p_a$.