THE ROLE OF HETEROGENEOUS RISK PREFERENCES, DISCOUNT RATES,
AND EARNINGS EXPECTATIONS IN COLLEGE MAJOR CHOICE

Arpita Patnaik
Joanna Venator
Matthew Wiswall
Basit Zafar

The Role of Heterogeneous Risk Preferences, Discount Rates, and Earnings Expectations
in College Major Choice
Arpita Patnaik, Joanna Venator, Matthew Wiswall, and Basit Zafar
NBER Working Paper No. 26785
February 2020
JEL No. I23,J16,J24

ABSTRACT

In this paper, we estimate a rich model of college major choice using a panel of experimentally-derived data. Our estimation strategy combines two types of data: data on self-reported beliefs about future earnings from potential human capital decisions and survey-based measures of risk and time preferences. We show how to use these data to identify a general life-cycle model, allowing for rich patterns of heterogeneous beliefs and preferences. Our data allow us to separate perceptions about the degree of risk or perceptions about the current versus future payoffs for a choice from the individual's preference for risk and patience. Comparing our estimates of the general model to estimates of models which ignore heterogeneity in risk and time preferences, we find that these restricted models are likely to overstate the importance of earnings to major choice. Additionally, we show that while men are less risk averse and patient than women, gender differences in expectations about own-earnings, risk aversion, and patience cannot explain gender gaps in major choice.

Arpita Patnaik
Department of Economics
University of Wisconsin-Madison
1180 Observatory Drive
Madison, WI 53706
patnaik3@wisc.edu

Joanna Venator
University of Wisconsin–Madison
Department of Economics
1180 Observatory Drive
Madison, WI 53706-1393
venator@wisc.edu

Matthew Wiswall
Department of Economics
University of Wisconsin-Madison
1180 Observatory Drive
Madison, WI 53706
and NBER
matt.wiswall@gmail.com

Basit Zafar
Department of Economics
Arizona State University
P.O. Box 879801
Tempe, AZ 85287
and NBER
basitak@gmail.com

1 Introduction

Several decades of research has emphasized the importance of earnings risk and life-cycle earnings growth to understanding occupation and human capital investment choices. One strand of this past work suggests that lifetime earnings risk varies significantly across education levels, college majors, and occupations, and that these differences are an important driver of choices (Altonji, 1993; Flyer, 1997; Saks and Shore, 2005; Nielsen and Vissing-Jorgensen, 2006; Bonin et al., 2007; De Paola and Gioia, 2012; Belzil and Leonardi, 2013; Fouarge et al., 2014; Barth et al., 2017; Dillon, 2018). Another strand of this literature emphasizes the differences in earnings growth across occupations and the delays and concomitant patience required to make human capital investments (Bonin et al., 2007; Oreopoulos, 2007; Cadena and Keys, 2015; Levitt et al., 2016). The standard assumption running through this past work is that individuals, conditional on certain observables, have the same beliefs about the characteristics of the choices they face. A central challenge in this research is then separately identifying beliefs from preferences: do the observed choices reflect a different belief about the risk or growth in earnings across these occupation and education choices, or a different tolerance for risk or level of patience?

We address this identification challenge by using a unique dataset with three key features: a) individual beliefs about the level of earnings, earnings growth, and earnings uncertainty for a set of possible college major choices, b) individual measures of risk tolerance and patience, and c) an experimental component allowing us to robustly identify the importance of earnings considerations separate from other aspects of choice. We use these various components of the data to estimate a general life-cycle model of college major choice, allowing for individual-level heterogeneity in preferences and beliefs. Using our estimated model, we can then evaluate the relative importance of each model feature.

We explore the importance of risk aversion, impatience, and earnings expectations in the choice of college major, using a survey and experimental design. We collect rich data from undergraduate college students at New York University (NYU), where in successive rounds respondents were asked their *self* beliefs about their own expected future earnings and other major-specific aspects were they to major in different majors, their beliefs about the population distribution of these outcomes, and the subjective belief that they will graduate with each major. In addition, we use a set of *hypothetical* survey questions to elicit students' preferences over risk and self-control based on measures developed by Eckel and Grossman (2002) and Coller and Williams (1999),

respectively. After the initial round in which the baseline beliefs and preferences are elicited, we provide students with accurate information on population characteristics of the major and observe how this new information causes respondents to update their self beliefs and their subjective probabilities of graduating with each particular major. This experimental design creates panel data for major choices, which is otherwise largely a one-time decision. We can then use the changes in subjective expectations about major choice and earnings characteristics of majors combined with the preference parameters elicited in the survey, to identify a structural model of major choice.

Consistent with most prior literature, we find that women, on average, are more risk averse and patient than men. The average annual discount factor in our sample for women is 0.88, versus 0.86 for men, with the difference statistically significant (p-value of equality = 0.008). Likewise, the average coefficient of relative risk aversion for women is 1.82 versus 1.38 for men (p-value of equality = 0.0001). There is also substantial heterogeneity *within* gender. For example, the 10th-90th percentile range of the discount factor distribution is 0.73-0.97 for women and 0.70-0.96 for men.

Our methodological innovation is to then use these individual-specific measures on time and risk preferences in a life-cycle model of college major choice, while imposing limited structure on the choice problem. In much of the literature on dynamic discrete choice, individual-specific measures of time discounting and risk preferences are not available and are difficult to separately identify from other modeling features using observed choices. For example, Magnac and Thesmar (2002) show that the discount factor is not identified from choice data without further restrictions on functional form. Importantly, we find that models which assume standard, homogeneous levels of discounting and risk aversion are likely to overestimate how sensitive students' college major choices are to expected earnings. It is, in particular, the heterogeneity in risk preferences that seems to matter. However, gender differences in risk preferences and patience cannot explain the pattern of women being less likely to select into high-paying, but risky majors such as business/economics.[1] Instead, tastes over major characteristics other than own-earnings explain most of the gender differences in major choice.

Historically, research on college major choice typically allows for limited degrees of individual

---

[1]There are well-documented differences in risk aversion and patience across gender (see Niederle (2014) for a review). Men are, on average, both less risk averse and less patient, prompting the hypothesis that some of the gender differences we see in labor market outcomes may stem from differences in these preferences. For example, women's lower preference for risk may result in them selecting out of high-risk, high-return majors and occupations, such as finance, which may contribute to gender differences in lifetime earnings.

heterogeneity in beliefs about future earnings, assumes that expectations are either rational or myopic, and then uses observed choices and earnings to identify parameters (Freeman, 1971; Siow, 1984; Zarkin, 1985; Bamberger, 1988; Berger, 1988; Flyer, 1997; Eide and Waehrer, 1998; Montmarquette et al., 2002; Arcidiacono, 2004; Beffy et al., 2012; Brodaty et al., 2014). This approach is limited: observed choices can be consistent with many different combinations of expectations and preferences (Manski, 1993).

The experimental design of this study, combined with data on probabilistic choices and subjective beliefs about major choice and earnings, allows us to separately identify the unobserved tastes for each major from individual heterogeneity in beliefs about the earnings distribution associated with a major under weaker modeling restrictions than is possible with cross-sectional data. Additionally, we elicit beliefs about future earnings at multiple points in time over the life-cycle, which allows us estimate a life-cycle utility model without making strong assumptions about earnings growth over the life-cycle. These innovations are in line with past work using this same survey (e.g., Wiswall and Zafar, 2015) as well as a more recent literature which uses subjective expectations data rather than choice data to understand decision-making under uncertainty in the context of schooling choices (e.g., Attanasio and Kaufmann, 2014; Zafar, 2011; Zafar, 2013; Giustinelli, 2016; Arcidiacono et al., 2012; Stinebrickner and Stinebrickner, 2012; Stinebrickner and Stinebrickner, 2013; Kaufmann, 2014; Ruder and Van Noy, 2017; Attanasio and Kaufmann, 2017).

Although the more recent literature allows researchers more flexibility to estimate models without making strong assumptions about student's expectations at the time of the decision, the data available to researchers typically do not allow them to incorporate individual-specific risk and discounting components into utility. Past research instead assumes a parametric functional form for utility that implies a specific coefficient of risk aversion, such as the natural log of expected earnings (e.g., Arcidiacono et al., 2012; Arcdiacono et al., 2014; Kaufmann, 2014; Attanasio and Kaufmann, 2017) or linear in expected earnings (e.g. Zafar, 2013). Papers that allow for risk aversion typically estimate a single parameter for the whole population (e.g., Beffy et al., 2012; Brodaty et al., 2014; Nielsen and Vissing-Jorgensen, 2006) rather than allowing it to vary flexibly across individuals. Estimates of the risk coefficient vary widely across samples, ranging from 0.6 to 5. Moreover, most studies fix the annual discount factor to be $\beta = 0.95$. In fact, none of the past literature that uses subjective expectations data estimates a discount factor. In a previous paper using parts of the same dataset considered here (Wiswall and Zafar, 2015), direct measures of risk and patience preferences were not used, and instead, following the previous literature, a homogeneous risk aver-

sion parameter (by gender) was assumed. Our findings indicate that ignoring the heterogeneity in time and risk preferences leads to biased inference (more specifically, an overestimation) regarding the importance of earnings in college major choice. However, qualitative conclusions regarding the drivers of the underlying gender gap remain unchanged.

This paper is organized as follows. We first describe the data collection methodology in Section 2. In Section 3, we discuss our measures of patience and risk aversion and motivate the role that these characteristics may play in college major choices. In Section 4, we outline our model of college major choice and then discuss identification of the model in Section 5. Section 6 presents model estimates, and the next two sections report counterfactual estimates from a structural life-cycle utility model of major choice and explore how gender differences in structural parameters are related to gender differences in major choice. Finally, Section 9 concludes.

2 Data

In this section, we describe the data collection process, including the administration of the survey, the information treatment provided to students, the questions used to recover subjective expectations about earnings and major choice, and the two games used to elicit preferences over risk and discounting.

2.1 Administration

Our data is from an original survey instrument administered to New York University (NYU) undergraduate students over a 3-week period, during May-June 2010. NYU is a large, selective, private university located in New York City. Students were recruited from the email list used by the Center for Experimental Social Sciences (CESS) at NYU. The study was limited to full-time NYU students who were in their freshman, sophomore, or junior years, were at least 18 years of age, and were US citizens. Upon agreeing to participate in the survey, students were sent an online link to the survey (constructed using the SurveyMonkey software). The students could use any Internet-connected computer to complete the survey. The students were given 2-3 days to start the survey before the link became inactive and were told to complete the survey in one sitting. The survey took approximately 90 minutes to complete and consisted of several parts. Students were not allowed to revise answers to any prior questions after new information treatments were received. Many of the questions had built-in logical checks (e.g., percent chances of an exhaustive set of events such as

majors had to sum to 100). Students were compensated $30 for successfully completing the survey.

2.2 Survey Instrument

Our instrument consisted of three distinct stages:

1. In the 'Initial Stage', respondents were asked about their beliefs about major-specific characteristics for both themselves and the population.

2. In the 'Intermediate Stage', respondents were randomly selected to receive 1 of 4 possible information treatments shown in Appendix Table A-1.[2] The information was reported on the screen and the respondents were asked to read this information before they continued. Respondents were then re-asked about population beliefs (on areas they were not provided information about) and self beliefs.

3. In the 'Final Stage', respondents were given all of the information contained in each of the 4 possible information treatments (of Table A-1). After having seen this information, respondents were then re-asked about their self beliefs. At the end of this stage, we elicited respondents' time and risk preferences through two hypothetical games (described below).

The information treatment consisted of statistics about the earnings and labor supply of the US population. Some of the information was general (e.g., mean earnings for all US workers in the All Individuals Treatment), while other information was specific to individuals who had graduated in a specific major (e.g., mean earnings for all male college-graduates with a degree in business or economics, in the Male Major Specific Treatment). For the purposes of estimating the choice model in this paper, we use the initial stage self beliefs as the pre-treatment and the final stage beliefs as the post-treatment observations.

Our goal was to collect information on consequential life activities that would plausibly be key determinants of the utility gained from a college major. Because of time constraints, we aggregated the various college majors to 5 groups:

1. Business and Economics

---

[2]The information was calculated by the authors using the Current Population Survey (for earnings and employment for the general and college educated population) and the National Survey of College Graduates (for earnings and employment by college major). Details on the calculation of the statistics used in the information treatment are in Appendix Section A.1; this information was also provided to the survey respondents.

2. Engineering and Computer Science

3. Humanities and Other Social Sciences

4. Natural Sciences and Math

5. Never Graduate/Drop Out.

Conditional on graduating in each of these major groups and for different future points in time (immediately after graduation, at age 30, and at age 45), students were asked for the distribution of self earnings. In addition, we collected data on the probability a student believes she will graduate with a major in each of these fields. This probability – the subjective expectation of graduating in a given major – is the primary outcome of interest that we investigate in this paper. Throughout the paper, when we refer to major choice, we are referencing currently enrolled college students' *beliefs* about what they plan to major in, not their realized major upon graduation.

2.3   Subjective Beliefs about Major

Our survey instrument provides us with a panel of self beliefs about earnings at 22, 30, and 45 years. Students were asked to report the earnings they expect to receive for working full-time, conditional on graduating in each of the four major groups or not graduating at ages 22, 30 and 45. They were also asked to report the likelihood of earning more than $35,000 and the likelihood of earning more than $85,000 at age 30 and 45. By asking about the likelihood of earning more than certain earnings levels, we are able to elicit beliefs about earnings uncertainty.[3] Appendix Table A-1 provides details about the information treatment and shows the variation in objective likelihoods associated with these values.

From these measures, we can recover subjective expectations about both the first and second order moments over the life cycle conditional on some functional form assumptions. As we elicit students' beliefs conditional on major choice, we do not need to rely on the restrictive assumption that major choices are the product of rational and fully informed choices. While Appendix section A.2 provides some sample survey questions, interested readers are referred to Wiswall and Zafar (2015) for a detailed discussion of the survey instrument.

---

[3]The choice of the specific values of $85,000 and $35000 was somewhat arbitrary. We could have used any values to identify the earnings uncertainty for which there was variation in the likelihood of achieving that earnings level across majors and there was not a mass of individuals who believed that they were certain to earn more or less than that amount (i.e., likelihoods equal to 0 or 1).

2.4   Risk and Time Preferences

The survey also included two hypothetical games that elicit risk and time discount preferences.

In the first, we employ a method developed by Eckel and Grossman (2002), henceforth referred to as the 'EG Game'. The question wording is presented in Section A.3. Students are presented with 6 different games of chance where the payoff depends on the toss of a coin, with the riskiness increasing with each game. The student is asked to choose one game that they would like to participate in. This choice corresponds to a range of values for the risk aversion parameter, bounded by the values that would make the student indifferent between that game and the other games that have the next highest and next lowest variance in payments.

In the second game, we employ a method developed by Coller and Williams (1999), henceforth referred to as the 'CW Game'. In this game, respondents were given the choice between receiving 3000 dollars in one month (Option A) or receiving a higher payment in 7 months (Option B). Students choose between Option A and B in ten such games, each with a different payment schedule with increasing amounts delayed to the later date (see Section A.4). The point at which a person switches from preferring the smaller amount in a month's time to the larger amount in seven months corresponds to a specific range of discount factors.

Using both of these games, we recover the risk and time preference parameters for each of the respondents corresponding to a utility specification that has functional form consistent with constant relative risk aversion and geometric discounting. See Appendix A.5 for the steps for recovering these individual-specific parameters.

Note that we did not provide students with payments corresponding to these schedules, and this was made clear to them before they were asked to make their decisions (that is, there was no deception). A natural question to ask is whether hypothetical rewards versus real rewards leads to some bias. With regards to risk preferences, Holt and Laury (2002) find that choices are similar for hypothetical and real gambles when stakes are not too high, as is the case in our scenarios. With regards to discount rates, one may suspect that individuals exhibit higher discount factors (more patience) when rewards are hypothetical, but Frederick et al. (2002), in their review article, conclude that there is no clear evidence that hypothetical rewards are discounted differently than real rewards.

2.5 Sample Definition

We restrict our sample based on the following criteria.

Of the 501 students who participated in the survey, we drop 6 students who report that they are in the 4th year of school or higher as they are more likely to have already committed to the major they will graduate in.

We then restrict the sample to students who report statistically sensible expectations about their future earnings. We drop 78 students who report a higher likelihood of earning more than $85,000 than of earning more than $35,000 in either the pre- or post- treatment sample for any of the 5 major categories. We do not observe the dropped students responding incorrectly for all probability responses (i.e., misreading the question to be asking about earnings being less than 35,000) nor for only a single response (i.e., a singular error). On average, those who made at least one probability error made 5.8 mistakes out of the 20 instances in which they were asked about the likelihood of making more than $85,000 and $35,000. The mistakes were somewhat more common for higher-paying majors (economics/business and engineering). Nonetheless, the dropped students do not differ from the remaining sample on observable characteristics (i.e., race, gender, SAT score, class year, parental income), expected likelihood of graduating in each major category, or expected earnings in each major category at age 30; we therefore drop these students for all remaining analyses.[4]

For individuals who report the exact same probability for both these questions, we recode the probability of earning more than $85,000 as 0.01 lower than the original value. There are 126 total respondents who report the exact same likelihood of making $35,000 and $85,000 at age 30 for at least one major either pre- or post- treatment. These responses are more common in the pre-treatment expectations, with 67 students reporting equal likelihood for at least one major only prior to receiving the information treatment, 27 reporting equal likelihood for at least one major only post receiving the treatment, and 32 reporting equal likelihood in both the pre- and post- treatment responses. When we compare the sample of students who ever report equal likelihoods to those who do not, we see significant differences in the observable characteristics of this group: those who report

---

[4]We conduct a t-test of equality of means across the estimation sample and the dropped respondents for all demographics characteristics, likelihood of choosing each major at age 30, and expected earnings at age 30. No differences are significant at $p < 0.05$ for a two-sided t-test of equality of means. One difference is marginally significant (p = 0.09): dropped students are less likely to expect to major in the natural sciences (13.3 %) than the full sample (19.2).

equal likelihoods have significantly lower SAT verbal scores on average, are less likely to be White and more likely to be Asian, significantly less likely to expect to major in humanities/social sciences, and report significantly higher expected earnings at age 30 in economics/business, engineering, and natural sciences.[5]

We also impose a ceiling and floor on beliefs about full time annual earnings such that any value above $500,000 is coded as $500,000 and any value below $10,000 is coded as $10,000. We apply this coding for earnings expectations for every major, treatment status, and age. The major specific earnings that were most often truncated were for economics and not graduating.The largest proportion of students top-coded were the five percent of the students with expected earnings above $500,000 at age 45 after graduating from economics (pre-treatment) whereas the largest proportion of bottom-coded responses were approximately six percent of the sample with expected earnings below 10,000 after not graduating at age 22. The reader may refer to Appendix Table A-5 for details on the proportion of the sample for whom we truncate earnings beliefs. We also recode all reported probabilities of 0 to 0.001 to allow for calculating log-odds ratios. Lastly, for all calculations, we scale earnings in terms of $100,000 to make computation of the present-discounted value of lifetime earnings computationally tractable.

The resulting final sample consists of 417 respondents, for whom we have 10 observations: 5 majors × 2 (pre-treatment and post-treatment). Of the 417 students in our sample, 63 percent of the respondents are female. The sample is 38.6 percent White and 44.6 percent Asian. Most of the students are underclassmen (41.7 percent freshmen and 36.5 percent sophomores) with a high average family income of $148,430. Our sample consists of high-ability individuals with an average Scholastic Aptitude Test (SAT) Score of 700 for Math and 683 for Verbal tests. Women have slightly higher SAT verbal scores and lower SAT math scores then men, though only the difference in SAT math scores is statistically significant. Table 1 describes the summary statistics for the sample.

---

[5]We conduct a t-test of equality of means between those who ever report equal likelihoods and those who do not. The characteristics mentioned are the only ones that the samples differ at a $p < 0.10$ level; we cannot reject the null that the samples are similar on all other characteristics (i.e., gender, SAT math score, class year, parental income, expected likelihood of non-humanities majors, and expected earnings in humanities and for non-graduates).

3 Measures of Future Expectations, Risk Aversion, and Patience

In this section, we discuss descriptive results about students' expectations about their future college major choice, expectations about the earnings associated with these choices, and their risk and discounting preferences.

3.1 Expectations about Earnings and Major Choice

Table 2 describes self-beliefs about expected earnings by major separately by gender, both pre- and post- the information treatment. Panel A shows that both genders expect average earnings to be the highest in economics, followed by engineering; this the case both pre- and post- treatment. Expectations about earnings at age 45 (reported in Panel D) are unsurprisingly higher. Pre-treatment, the average earnings beliefs are quite similar by gender for the graduating majors.

After receiving information about the average earnings in these majors at age 30, men and women's earnings expectations adjust. For almost all college majors, men and women revise earnings beliefs downwards, meaning that both men and women are on average adjusting towards the population average provided in the treatment. Humanities/social sciences is the exception, with men revising age 30 beliefs upwards (though the difference is not statistically significant). As we show later, women over-estimate population earnings in humanities/social sciences more than men (12.6 percent, on average, versus a 0.7 percent overestimation by men), and also tend to have larger absolute errors in their population beliefs (that is, they are ex-ante less informed). This explains why women are more responsive to the treatment.

The other panels of Table 2 report students' expectations about the likelihood that they will be making more than $85,000 and $35,000. Both men and women respond similarly to these questions. Around 80-90% of both genders think they will make more than $35,000 at both age 30 and age 45 if they graduate from college, regardless of major. There is more variation by major in the expected likelihood of making more than $85,0000, with students thinking it is more likely with an economics or engineering major and extremely unlikely if they do not graduate from college. Students were also more responsive to the information treatment in how they respond to the likelihood of making $85,000, with the average likelihood dropping significantly across all majors.

Though students' beliefs about their future earnings potential in different majors responded to the information treatment, their beliefs about their future major choice were less responsive. Table

[3](#) provides descriptive statistics of beliefs about own probability of graduating with the different majors. Pre-treatment, women are significantly more likely than men to report that they will graduate with humanities (average probability of 49.9 versus 31.4 for men) and significantly less likely to report majoring in economics (23.8 percent versus 36.7 percent for men) or engineering (5.6 percent versus 8.7 percent for men). These gender differences remain even after the information treatment, though they are slightly smaller. Post-treatment, both men and women report a lower likelihood of choosing a humanities/social science major, with women reporting a higher likelihood of majoring in economics or engineering and men reporting a higher likelihood of majoring in engineering or the natural sciences.

3.2 Heterogeneity in Risk Aversion and Patience

We next look at how students' risk aversion and patience are correlated with gender, expected major, and other demographic characteristics, using the parameters calculated based on the responses to the EG game and the CW game. Tables [4](#) and [5](#) provide summary statistics for the risk and time discount parameters in our sample, by gender and by the major reported at the time of the survey, respectively. Note that a higher time discount factor implies a more patient individual, and a higher coefficient of relative risk aversion implies a higher degree of risk aversion.

Panel A of Table [4](#) shows that the annual discount factor is lower, on average, than the β typically used in dynamic choice models, but is consistent with other estimates of discount factors in hypothetical choice exercises. For example, the mean value of 0.874 in the NYU sample is close to estimates in [Andersen et al.](#) ([2008](#)), which finds a discount factor of 0.90 in a population of German adults. We find that men are slightly more impatient, with an average discount factor for men of 0.86 versus 0.88 for women (test of equality p-value = 0.008). This is consistent with some past work using similar elicitation methodologies (e.g., [Coller and Williams](#), [1999](#); [Dittrich and Leipold](#), [2014](#)) as well as past work eliciting discount factors for college students ([Brunello et al.](#), [2004](#)). However, the literature is not conclusive on the relationship between gender and patience. Other research finds no significant relationship between gender and patience ([Becker and Mulligan](#), [1997](#); [Andersen et al.](#), [2008](#)) and more recent cross-country analysis by [Falk and Hermle](#) ([2018](#)) finds that men are more patient on average. However, they find a great deal of heterogeneity in this parameter across countries.[6]

---

[6]Men are more patient in 68 percent of the countries they sample and this association is only significantly different from zero in 32 percent of countries. For comparison, women are more risk averse than men in 95% of the countries

Additionally, our results suggest there is substantial heterogeneity *within* gender, as indicated by the sizable standard deviation. For example, for females, the 10-90 percentile of the estimated discount factor distribution is 0.73-0.97. The corresponding range for males is 0.70-0.96. While only 50 percent of males have a discount factor that is 0.88 or higher, 65 percent of females have a discount factor in that range.

Turning to risk preferences, Panel B of Table 4 shows that the average coefficient of risk aversion is 1.7. The distribution of estimates is similar to Dave et al. (2010) with most of the respondents choosing Gamble 3, a gamble that is associated with a CRRA coefficient ranging from 0.71 to 1.16. The average coefficient of risk aversion is significantly higher for women in our sample (1.82) compared to men (1.38) and this is true for the median values as well. This is consistent with past literature (e.g., Niederle, 2014, Eckel and Grossman, 2008, Croson and Gneezy, 2009, Falk and Hermle, 2018) although some of these studies adopt a different experimental design, which makes it difficult to compare the magnitude of our estimates to theirs. The studies that are closest to our study design are Eckel and Grossman (2002) and Dave et al. (2010). Eckel and Grossman (2002) find that more than one third of the men and only 13 percent of the women in their sample choose the most risky gamble. In our sample, 11 percent of the men and 9 percent of the women chose the most risky game. Our sample, particularly the sample of men, is therefore more risk averse than the sample in Eckel and Grossman (2002). Additionally, we see that the variation *within* gender is also sizable, as indicated by the large standard deviations of 1.2 points.

Figure 1 plots the distributions of risk and time preference measures by gender. To test whether the underlying distributions differ, we use a Kolmogorov- Smirnov test and can reject the null hypothesis that the distribution of parameters are the same across genders for both the coefficient of risk aversion ($p = 0.001$) at the 1 percent level and the discount factor ($p = 0.059$) at the 10 percent level .

Panel A1 of Table 5 shows that discount factors also vary by the students' intended majors. For example, students who report intending to major in humanities, on average, have a higher discount factor, and an F-test of equality across majors marginally rejects the null hypothesis that the time preference parameter is equal across majors. When we condition on gender, as shown in Panels B1 and C1 of the table, we see sorting by females into majors based on their discount factors. We also find lower discount factors for non-economics/business majors that have reported taking at least

─────────────────────

and this relationship is statistically significant in 82% of the countries. In the country most comparable to our sample, the United States, the authors find a non-significant relationship between gender and patience.

one course in economics/business relative to non-economics/business majors that have not taken any economics/business courses (0.86 versus 0.89; p-value = 0.012).

Panel A2 of Table 5 shows that the average risk aversion parameters also differ by the students' intended majors, with students intending to major in humanities the most risk averse, on average. However, this seems to be driven by the gender sorting into majors. Conditional on gender, the variation in risk preferences across majors does not show any systematic patterns (panels B2 and C2).

We next explore other correlates of these preference parameters. Table 6 regresses each measure onto gender, ability (SAT Math and Verbal Scores), race (binary indicators for White and Asian, omitted category: Black, Hispanic, and Other Races), and parental income. Gender continues to be significantly correlated with these measures of risk and time preference measures, even after controlling for a rich set of other observable characteristics. This is consistent with past work (Dave et al., 2010) which finds similar significant correlations between risk preference and gender.

We also see a significant correlation between quantitative skills and the risk coefficient in the full sample: a 100-point increase in the SAT Math score is associated with a 0.187 lower CRRA coefficient.[7] We also look at this relationship separately by gender and find that this relationship is largely driven by the subgroup of women. We also see that women's SAT Math score is negatively related with patience. In addition, we find that race is significantly correlated with time preferences, with White students having significantly higher discount factors relative to the omitted racial category (non-Whites and non-Asians).[8] It is worth noting that these covariates explain a fairly small amount of variation in these preference parameters, as indicated by the low R-square values reported at the bottom of the Table.

---

[7] Recall that our sample is primarily high-ability students in a selective university, meaning that the relationship we observe is for the right tail of the SAT score distribution with half of the students having a 700 or higher on the SAT math test. Since there is a large mass of students very close to the ceiling of the score distribution, we re-run the regression from above and find non-linearity in this relationship, with only the top quartile of SAT scores being associated with less risk aversion.

[8] Benjamin et al. (2010) find that the average discount factor for a population of Black college students is lower than White college students, but that making students' ethnicity salient makes both Asian Americans and native Black Americans more patient. Our sample is too small to analyze Black students separately, but our results are broadly consistent with their findings in an un-primed sample of college students.

3.3 Relating Risk Aversion and Patience Parameters to Expectations about Major Choice and Earnings

In this Section, we explore the relationship between our measures of risk aversion and patience and a student's beliefs about the likelihood of choosing a given major and their expected earnings for each major.

For comparability to our model (that we describe in the next section), our outcome of interest is the log relative probability of graduating in major k for an individual i ($\tilde{P}_{ik}$), which is the natural log of the likelihood of choosing major k relative to the likelihood of choosing Humanities / Social Sciences. We regress this measure on the individual specific discount factor β_i, CRRA risk parameter θ_i, a dummy for gender (Male$_i$), and demographic controls (X_{ik}) which include SAT math and verbal scores, race dummies, parental income and expected earnings at age 22 in the major k relative to humanities. We replace all cases where students report zero as the percent chance of graduating in a major with 0.001. We standardize the risk and time preference parameters to have mean of zero and a standard deviation of one, which eases the interpretation of the coefficients.

Our specification is given by:

$$\tilde{P}_{ik} = \alpha_0 + \alpha_1 Male_i + \alpha_2 \beta_i + \alpha_3 \theta_i + \alpha_4 X_{ik} + \varepsilon_i.$$

Table 7 presents the results from this reduced-form regression. The choice of major is significantly correlated with gender, with women reporting a lower likelihood than men of choosing all majors relative to humanities/social sciences. A higher discount factor (more patience) is associated with a lower probability of majoring in economics/business relative to the humanities and this relationship is statistically significant. A higher risk coefficient (higher degree of risk aversion) is associated with a lower probability of majoring in any of the majors relative to humanities, although the relationship is not statistically significant. These findings are consistent with the pattern of unconditional averages by intended major reported in Table 5, demonstrating that the relationship between majoring in economics/business and patience is not driven by selection into the major on observable characteristics.

The lack of a significant relationship between these measures and intended major could reflect the fact that these preferences alone are not enough to explain major choice. Not all students have the same expectations about earnings – the level, the extent of uncertainty, and growth – across majors. For example, if some students see humanities degrees as risky and others see them as sources

of steady income, students with the same level of risk aversion would report different likelihoods of selecting this major. Moreover, students' beliefs about future earnings could be affected by their risk aversion or patience if those characteristics change how they acquire information about earnings associated with a major.

To address these concerns, we look at whether the risk and patience parameters are associated with expectations about *own* earnings in a major and one's expectations about *population* earnings in a major relative to US average earnings within that major (i.e., the information given in the treatment). The first exercise tests whether more patient or risk averse students perceive their own earnings potentials in each major differently than the less patient or risk averse students. The second analysis tests whether more patient or risk averse students have different beliefs about the population distribution of earnings associated with each major.

First, we evaluate the correlation between the relationship between expected earnings at various points in the life cycle and our measures of risk aversion and impatience. We do this by pooling our data for all majors (so that there are 5 observations for each individual corresponding to the 5 major categories) and then look at the binned scatter plot of expectations of earnings at age 22 and the expectations of earnings growth between ages 22 and 30 against risk and time preferences separately. The expected earnings growth for each individual and major is defined as the percent change in expected earnings between ages 22 and 30. Figure 2 reports the results of this binned scatter plot by gender. On average, more patience and more aversion to risk are associated with lower initial earnings, especially for male students. The coefficients obtained from the regressions of these measures of own expected earnings on the risk aversion and patience parameters also reflect a similar relationship (though the estimates are not precisely estimated). Further, Panel B of Figure 2 shows that more patience is also associated with higher earnings growth (again, this is particularly true for men). There seems to be little systematic relationship between risk preferences and expected earnings growth. Together, these scatter plots suggest that more patient (male) students expect higher earnings further in the future rather than immediately after graduation. Similarly, we find that more risk averse students also have more conservative expectations of earnings at age 22. Primarily due to data limitations, the standard approach in the literature generally assumes that preference parameters are not systematically related with expectations. However, these patterns suggest that might not be a good assumption. This also makes the case for collecting such data directly from respondents.

Next, we explore whether there is a relationship between risk and time preference and the

accuracy of beliefs about beliefs about *population* earnings. We construct a measure of accuracy of beliefs for individual i and gender g for major k using the percent difference between beliefs and true earnings: $100 \times \frac{\text{Earnings Beliefs}_{igk} - \text{Truth}_{gk}}{\text{Truth}gk}$, where Belief_{igk} is the belief about full time earnings at age 30 for individual i of gender g graduating from major k, and Truth_{gk} is the true full time earnings for an individual of the same gender g graduating from major k in the US (as was revealed in our information treatment). Table 8 reports the mean values of these error terms by major and gender. We report both the actual values as well as the absolute values of the error since the mean of the actual error may be close to 0 even if there are large positive and negative values of similar magnitudes. We find that there is substantial heterogeneity in belief error, both by gender as well as by major. Students overestimate the average earnings for an economics/business major by much more than they do any other major, although there are substantial differences in the degree of error by gender. Women overestimate population earnings in economics/business by 27 percent, on average, which is much larger than the degree to which men overestimate (4 percent). On the other hand, both men and women tend to underestimate the population earnings for an engineering major. Average errors are significantly larger for women, though the gender difference in average absolute errors is smaller. We find that the largest mean absolute errors for the full sample are for Economics/Business majors and for not graduating.

Similar to our earlier exercise, next, we pool the error data across majors for individuals and plot the individual specific measures of patience and risk aversion against the *absolute* percent error in beliefs, as shown in Figure 3. We also report the coefficients obtained from regressing the *absolute* percent error on each of the risk aversion and patience measures. We find that there is a positive relationship between the discount factor and the absolute error for both genders (though the relationship is not precisely estimated). That is, more patient students seem to have less accurate population beliefs. We find no systematic relationship between the risk preferences and the absolute error in population beliefs.

Taken together, these results suggest that patience and risk aversion are associated, in varying degrees, with how students form expectations about earnings in a major, but it less clear how these beliefs impact major choice. To better understand the relationship between these preferences and major choice, in the next Section, we therefore incorporate students' expectations about the earnings path associated with each major and their preferences into a unified structural model of major choice.

4 Model

In this Section, we specify the model of college major choice. Throughout, we refer to aspects of our survey design that inform the modeling decisions. As mentioned above, the survey provides us with students' subjective expectations about future income conditional on *all* possible major choices, allowing us to be more flexible in the structure we place on their beliefs and decisions.

4.1 Within-College Preferences and Beliefs

Individuals start in period $t = 0$ as college students who have not yet chosen a college major. After realizing a shock to their utility from each major, each student then chooses a major from the choice set $k = 1, ..., K$, where each option represents graduating in one of $K - 1$ majors or not graduating from college. This timing is consistent with our survey design; the survey is conducted with currently-enrolled college students, who are therefore in period $t = 0$ and have not realized their graduation/major outcome yet.

Utility for person i for each college major k at time $t = 0$ is given as follows:

$$V_{ik0} = \alpha_{ik} + \epsilon_{ik} + \mathbb{E}[V_{ik1}] \tag{1}$$

There are three components: a) known tastes α_{ik}, b) a mean-zero taste "shock" ϵ_{ik}, and c) expected future labor market earnings $\mathbb{E}[V_{ik1}]$.

The α_{ik} component represents the student's known taste for major k during the initial period. This component captures all major-specific outcomes that are known to the student during college, excluding expectations about future earnings in the major which are contained in $\mathbb{E}[V_{ik1}]$. α_{ik} includes major characteristics that are realized during college, such as interest in the coursework, as well as non-pecuniary outcomes that are realized post-college, such as knowledge about the work/life balance associated with careers within that major. Because we do not impose a parametric structure on this taste parameter, there is no loss of generality in modeling the time-invariant tastes as preferences over both current and future events.

The ϵ_{ik} term represents the preference shock for a major that is unknown to a student before they choose their major. This "shock" can be thought of as either a shock to perceived ability in each major or taste for the major that is unknown before the student commits to a major. Since the choice of k is the decision to *graduate* in a given major, the inclusion of this type of shock allows us the flexibility to address the fact that many students come into college planning on choosing one major and then later switch majors after realizing that the courses are harder or less

17

interesting than anticipated.[9] Because we survey students about their likelihood of choosing each major prior to graduation, we assume that these reports are occurring prior to the realization of the ϵ_{ik} shock. The invariant taste parameter, α_{ik}, reflects the fact that some students have stronger a priori preference for a given major than other students, making it more likely that they will persist in their initial major choice rather than switching majors upon realizing the ϵ_{ik} preference shock. We might, in fact, expect a great deal of change in how much a freshman in college knows about the career opportunities available to economics majors versus a senior in college. In our model, these taste parameters, α_{ik}, can vary by the student's age and time until graduation. But we do not explicit model any particular learning process, as in previous work (Arcidiacono, 2004 and Stinebrickner and Stinebrickner, 2013), but do assume that the major-specific preference shocks, $\epsilon_{i1}, \ldots, \epsilon_{iK}$, will be known to the student when they choose their major, resolved through some implicit, unspecified learning process.

4.2   Post-College Preferences and Beliefs

In period $t = 1$, individuals leave college and enter the labor market, where they work and consume income until period $t = T$.[10] Each period, the individual receives utility from consumption and consumes their entire income for that period. Individuals have standard constant relative risk aversion (CRRA) preferences, governed by their coefficient of risk aversion, θ_i. Section 3 describes how we elicited this parameter for each individual. Students discount future period's utility at rate β_i, a parameter also elicited in the survey through a series of hypothetical choices about the timing of a monetary reward receipt. Students have heterogeneous beliefs about the distribution of earnings in major k at time t, denoted $G_{ikt}(w)$. We also scale utility from consumption with a free parameter ϕ, which we estimate. This parameter scales utility from earnings to be in the same units as the taste shock parameter.[11]

---

[9]For an overview on the prevalence of switching majors, see Arcidiacono (2004), Stinebrickner and Stinebrickner (2012), Stinebrickner and Stinebrickner (2013), and Arcidiacono et al. (2012).

[10]We specify $T = 55$ in the estimation and do not model utility flows after this age. While individual's beliefs about labor supply and earnings may depend in part on their expectations regarding the end of life period, given the long horizon between college and age 55, omitting any explicit modeling of the period after age 55 has negligible consequences for approximating the utility from major choice.

[11]The scale of utility is in the same units as this shock rather than in dollars. We can make one standardization of scale and do that when we later impose structure on the taste shock parameter, assuming students receive this as a draw from a type I extreme value distribution. Therefore, we need ϕ to convert the scale of EV_1 from dollars to the scale of the ϵ shock.

Post-graduation utility for student i in major k is thus given by:

$$\mathbb{E}[V_{ik1}] = \phi \sum_t \int \beta_i^t \frac{w^{1-\theta_i}}{1-\theta_i} dG_{ikt}(w). \tag{2}$$

Following Wiswall and Zafar (2015), our specification of future value deviates from the standard approach to human capital models. In a typical human capital investment problem, one must solve for the optimal decisions and outcomes (e.g., labor supply, etc.) given that human capital investment. Because we ask each student how much they *expect* to earn at different ages conditional on major choice, we are able to avoid explicitly solving for the optimal choice path post- college and instead substitute the students' reported beliefs about their choices into the utility function. This is both more general, as it allows for students to have heterogeneous and non-rational expectations about their future earnings, and it also makes the model more tractable in terms of computation.

Note that we do not incorporate information about subjective beliefs about future likelihood of marriage, fertility, etc. into our utility specification, though students were asked about their expectations about likelihood of marriage and their expectations about their future spouses' earnings conditional on their own-major choice. Attanasio and Kaufmann (2017) and Wiswall and Zafar (2018b) find that educational choices are in fact impacted by marriage market considerations. We abstract away from those factors in this paper; marriage market and family considerations will be subsumed in the taste term.

Our key innovation is to allow for the discount factor and the coefficient of relative risk aversion to be individual-specific, while imposing limited structure on the choice problem. In much of the literature on dynamic discrete choice, measures of θ_i and β_i are not available and are difficult to separately identify from other modeling features using observed choices.[12] The standard approach is therefore to make ad hoc functional form assumptions that allow for identification from standard data, set values for these parameters, or typically both. For the discount rate, a usual approach is to set it to some reasonable value, such as 0.95. For the risk aversion parameter, the standard approach is to either choose a simple functional form for utility, such as the natural log of consumption (i.e., the limit of CRRA as θ goes to 1) or linear utility (i.e., $\theta = 0$), or assume a homogeneous θ conditional on some observables (for example, assuming that all females have a homogeneous risk preference). In principle, one could allow for heterogeneity in β and θ, but this typically only identified from standard data if one imposes some parametric restriction on the population

---

[12]See, for example, Rust (1994) and Magnac and Thesmar (2002) who show that the discount factor is not identified from choice data without further restrictions on functional forms.

distribution of the parameter. Our survey questions allow us to relax these types of parametric assumptions, and recover a unique individual-specific measure of time and risk preferences, without imposing any assumptions on how they are distributed in the population.

Note that the distribution of future post-graduation earnings $G_{ikt}(w)$ describes "unresolvable" uncertainty as students will not know their true post-college earnings path in each major at the time of major choice. We allow for heterogeneity in these beliefs across students and do not need to assume that students are either fully informed or that they behave consistently with rational expectations. Instead, these beliefs should be thought of as deriving from a mix of public and private information available to students during college as they are making the major choice decision. They are the beliefs that the student holds about their own earnings potential in a major, which may be quite distinct from their beliefs about the population distribution of earnings in that major. A student may know that engineering majors on average have high earnings, but believe that he will earn less than average due to private knowledge about his own abilities in engineering.

If we combine equations (1) and (2), a student's utility from choosing major k is given by

$$V_{ik0} = \alpha_{ik} + \phi \sum_t \int \beta_i^t \frac{w^{1-\theta_i}}{1-\theta_i} dG_{ikt}(w) + \epsilon_{ik}. \tag{3}$$

This is the general structure for choice-specific utility we take to the data.

5   Identification and Estimation

In this section, we describe how the parameters of the model are identified and estimated. For this purpose, we use our panel data on beliefs about earnings and major choice. The model parameters include the parameters that define the earnings distribution for each student i for each major k pre- and post- treatment, the risk parameter (θ_i), the discount factor (β_i), and the utility parameters (ϕ_i and α_{ik}).

5.1   Earnings Parameters

Although we have rich data on earnings beliefs for each individual in our sample, we do not of course have data about the perceived distribution of earnings at *all* possible future dates. As described above, for each sample member, we have 7 unique moments about each of 5 possible majors (including not graduate) and the pre- and post- experiment periods, a total of 70 earnings moments for each individual. To "fill-in" the remaining information, we need to assume some parametric

structure to earnings beliefs, although we allow this structure to be individual specific and impose no global restrictions on the population distribution. We assume a flexible wage structure that allows individual-specific beliefs about the mean earnings as well as beliefs about the dispersion of earnings by major to grow over the life cycle.

Individual i's belief about earnings (w_{ijkt}) in major k at age t during treatment period j are drawn from a distribution $G_{ijkt}(w)$. We impose a log Normal structure on $G_{ijkt}(w)$ such that,

$$
\begin{aligned}
&\log w_{ijkt} \sim N(\mu_{ijkt}, \sigma_{ijkt}), \\
&\text{where } \mu_{ijkt} = \mu_{ijk0} + \mu_{ijk1}(t-22) + \mu_{ijk2}(t-22)^2 \\
&\sigma_{ijkt} = \sigma_{ijk0} + \sigma_{ijk1}(t-22)
\end{aligned}
\tag{4}
$$

Here the mean of the log of earnings for individual i in major k at age t during treatment period j is a quadratic function of age whereas the dispersion of log of earnings is governed by a linear function of age. Therefore, for each individual, there are $5 \times 5 \times 2$ unknown parameters (five parameters for each of the five majors and pre-/post- treatment).

Our data provides information on various moments of the individual-major specific expectations, including beliefs about average earnings at age 22, age 30, and age 45 and the probability of earning more than $35,000$ and $85,000$ at ages 30 and 45. Under the parametric assumptions on the distribution of $\log w_{ijkt}$ we are able to identify the 50 ($5 \times 5 \times 2$) parameters of the earnings distribution using these 50 moments (5 moments $\times$ 5 majors $\times$ 2 treatment). Appendix B provides details of the estimation method and identification of the μ and σ terms.

5.2 Identification of Preference Parameters

Identification of the risk and time preferences has already been discussed. This leaves us with the parameters of the utility function. We now add a subscript j for pre- and post- treatment value of the major to the value function in equation (3):

$$
V_{ijk0} = \alpha_{ik} + \phi \sum_{t} \int \beta_i^t \frac{w^{1-\theta_i}}{1-\theta_i} dG_{ijkt}(w) + \epsilon_{ijk}.
$$

We begin by assuming that the preference shocks, ϵ_{ijk}, are distributed i.i.d. Extreme Value across major choices and across individuals. Note that while this imposes a particular distribution on the taste shock, we place no restriction on the permanent component of tastes α_{ik}. Because the latter is unrestricted, the distributional restriction on ϵ_{ijk} is without loss of generality since there is no parametric restriction on the combined individual taste term $\delta_{ijk} = \epsilon_{ijk} + \alpha_{ik}$.

21

Then, the probability a student chooses major k can be written as:

$$P_{ijk} = \frac{\exp\left(\alpha_{ik} + \phi \sum_t \beta_i^t \int \left[\frac{1}{1-\theta_i} w_{ijkt}^{1-\theta_i}\right] dG_{ijkt}\right)}{\sum_s \exp\left(\alpha_{is} + \phi \sum_t \beta_i^t \int \left[\frac{1}{1-\theta_i} w_{ijst}^{1-\theta_i}\right] dG_{ijst}\right)}.$$

Taking humanities/social sciences as the reference major (henceforth denoted with the subscript $k = 0$), we can re-write the log relative probability of choosing major k as:

$$\ln \frac{P_{ijk}}{P_{ij0}} = \alpha_{ik} - \alpha_{i0} + \phi \sum_t \beta_i^t \left(\int \left[\frac{1}{1-\theta_i} w_{ijkt}^{1-\theta_i}\right] dG_{ijkt} - \int \left[\frac{1}{1-\theta_i} w_{ij0t}^{1-\theta_i}\right] dG_{ij0t} \right).$$

We then evaluate the integral using simulation. We approximate a student's expectations over earnings by taking draws $r = 1, \ldots, R$ (with $R = 100$) from the earnings distribution $G_{ijk}(\hat{\mu}_{ijkt}, \sigma_{ijk})$ defined previously and computing the average of utility over those draws:

$$\underbrace{\ln \frac{P_{ijk}}{P_{ij0}}}_{\tilde{P}_{ijk}} = \alpha_{ik} - \alpha_{i0} + \phi \underbrace{\sum_t \beta_i^t \left(\frac{1}{R} \sum_r \left[\frac{1}{1-\theta_i} w_{itrjk}^{1-\theta_i}\right] - \frac{1}{R} \sum_r \left[\frac{1}{1-\theta_i} w_{itrj0}^{1-\theta_i}\right] \right)}_{f_{ijk}} \tag{5}$$

$$\tilde{P}_{ijk} = \phi f_{ijk} + \eta_{ijk}. \tag{6}$$

By combining terms, we now have a linear relationship between the known quantities in the data, the log relative probability of choosing major k, $\tilde{P}_{ijk}$ and the expected relative utility from earnings, f_{ijk}, and an unknown component η_{ijk} made up of the relative preference parameters α and a measurement error term e_{ijk} which varies across treatments, $\eta_{ijk} = \alpha_{ik} - \alpha_{i0} + e_{ijk}$.

Because we would expect f_{ijk} to be correlated with η_{ijk} through the correlation of expected earnings with tastes for a major, a linear regression of $\tilde{P}_{ijk}$ on f_{ijk} will likely yield biased estimates of ϕ and α. Previous estimates of the relationship between subjective expectations about earnings and major choice suggest we should expect these estimates to be biased upwards (Wiswall and Zafar, 2015): individuals with a high relative taste for a major typically also expect to earn more in that major. The information experiment in our setup creates panel data, allowing us to use a fixed effects identification strategy to identify the preference parameters.

For each student, we have data for their expected earnings for each major and the likelihood that they will choose each major, both pre- and post- the information treatment. Because there is a short time frame between reporting their initial preferences, receiving information about the true earnings distribution, and reporting their post-treatment preferences, we assume that the only part of the utility that changes between the pre- and post- stages are students' expectations about

earnings, not their persistent taste for a major. Thus, any changes in the reported likelihood of choosing a major k would show up in f_{ijk} or in the idiosyncratic measurement error e_{ijk}; α_{ik} would be unchanged across pre- vs. post- treatment. This assumption is quite plausible in our setting, where the pre- and post- treatment elicitation of beliefs are minutes apart.[13] On the other hand, in panels separated by several months or years (e.g., Lochner, 2007; Zafar, 2011; Stinebrickner and Stinebrickner, 2012; Stinebrickner and Stinebrickner, 2013), this assumption could be less defensible: students may learn more about themselves or the majors over extended periods of time, and their tastes for majors may evolve.

This implies that for each person we have two data points for each major k:

$$\tilde{P}_{i1k} = \phi(f_{i1k}) + \eta_{i1k}$$
$$\tilde{P}_{i2k} = \phi(f_{i2k}) + \eta_{i2k}.$$

Taking differences between the two, we can rewrite the relationship as:

$$\begin{aligned}
\tilde{P}_{i2k} - \tilde{P}_{i1k} &= \phi\big[f_{i2k} - f_{i1k}\big] + \eta_{i2k} - \eta_{i1k}, \\
&= \phi\big[f_{i2k} - f_{i1k}\big] + (\alpha_{ik} - \alpha_{i0}) - (\alpha_{ik} - \alpha_{i0}) + e_{i2k} - e_{i1k}, \\
&= \phi\big[f_{i2k} - f_{i1k}\big] + e_{i2k} - e_{i1k}.
\end{aligned} \tag{7}$$

With panel data, we can eliminate the time-invariant taste for major component, α_{ik}, allowing us to estimate an unbiased coefficient on expected utility from earnings. These taste components contain individual-specific major "tastes" that we expect would not change in the short period of time in which they are taking the survey. For example, upperclassmen likely have higher costs associated with switching majors than freshmen, but these costs are the same at the beginning of the survey and the end of the survey. This final equation is equivalent to regressing the log relative probability onto the utility from earnings, with individual and major fixed effects. Identification of ϕ depends on the assumption that the change in beliefs about unobserved components on utility, or $e_{i2k} - e_{i1k}$, are mean-independent of the change in observed beliefs about earnings. The major fixed effects control for any effects of the information treatment that may vary systematically by major.

Using this estimate of ϕ, we can back out the persistent taste component of utility $\alpha_{ik} - \alpha_{i0}$ without needing to impose any structure on the distribution of this component (e.g. that it is

[13]Any beliefs about major characteristics that change in response to the treatment are thus contained in e_{ijk}. For example, to the extent to which beliefs about spouses' earnings conditional on own major change, this is contained in this error term.

Normally distributed). Following from equation 5, we can estimate $\hat{\alpha}_{ik} = \tilde{P}_{i1k} - \hat{\phi}f_{i1k}$. Without the experimental panel data, we cannot separately identify tastes from beliefs about the earnings potential associated with each major. Past research relies on parametric restrictions on the joint distribution of taste parameters (e.g., Berger, 1988; Arcidiacono, 2004; Beffy et al., 2012. This could be problematic: using a more restrictive model, Wiswall and Zafar (2015) show the distributions of major tastes implied by our experimental data suggest that the commonly-assumed forms of Normal or extreme value distributions are not good approximations of the true distributions. The current study extends that finding to non-parametrically identify tastes in a panel setting under more general assumptions about the functional form for utility that incorporate heterogeneity in risk and time preferences.[14]

6 Estimates

6.1 Earnings Parameters

We have already outlined the earnings expectations process in Section 5.1. Table A-6 presents the average estimated parameters of the earnings process, averaged by gender, estimated using the method outlined in Appendix B. The estimates are consistent with individuals expecting an increasing but concave life cycle earnings profile, as is typically found using actual realized earnings. Appendix Figure A-1 shows the median expected earnings path for each major before and after the information treatment. We simulate 100 draws from the Log-Normal distribution using the parameters estimated to plot median earnings at each age along with a [0.25, 0.75] confidence interval.[15] We also compare the evolution of these beliefs about the earnings distribution with the actual life cycle growth of earnings for respondents in the American Community Survey (ACS). This allows us to characterize the difference in beliefs relative to the "true" age-earnings profile. This provides some indication that the beliefs reported are sensible, although given our sample composition, and the fact that these are beliefs, not realized earnings, we do not expect a perfect match. Appendix Figure A-2 also shows that the average fitted earnings profile for each major qualitatively matches the one calculated from the ACS quite well.

---

[14]In contrast to that earlier paper (Wiswall and Zafar, 2015), we do not consider labor supply or spousal earnings.

[15]Note that we do not have any moments after age 45, and assume that the earnings profile is flat between 45 and 55 years of age.

6.2   Preference Parameters

We estimate ϕ and α_{ik}, using two approaches. In the first ("cross-sectional approach"), we regress $\tilde{P}_{i1k}$ onto f_{i1k} and a set of dummies for major in the cross-sectional pre-treatment data:

$$\tilde{P}_{i1k} = \phi(f_{i1k}) + \gamma_k + e_{i1k}.$$

This estimation method implies that $\alpha_{ik} = \gamma_k +$ residual. If f_{i1k} is correlated with α_{ik}, this will result in a biased estimate of ϕ, with a positive correlation implying an overestimate of ϕ in this estimation method.

Rather than restrict α_{ik} in this way, our second approach ("Pre/Post Treatment Diff. Data") takes advantage of the variation induced by the information treatment, which gives us two observations of an individual's choices and expectations, pre- and post- treatment (denoted with j). As described in equation (7), we can use the experiment to relax the assumption that f_{ijk} is uncorrelated with α_{ik} by differencing the data across treatments and regressing $\tilde{P}_{i2k} - \tilde{P}_{i1k}$ onto $f_{i2k} - f_{i1k}$ and a set of major dummies:

$$\tilde{P}_{i2k} - \tilde{P}_{i1k} = \phi\big[f_{i2k} - f_{i1k}\big] + \gamma_k + \nu_{ik}. \tag{8}$$

The richness of our data then allows us to test how functional form restrictions can lead to biased inference and different estimates. To do this, we estimate five different versions of the discrete choice problem described above. Model 1 is the most general version of the model in which we allow for individual heterogeneity in both β_i and θ_i. In Model 2, we restrict θ to be constant across all students and equal to the mean value of θ_i in the sample, 1.7. In Model 3, we add a second restriction to Model 2 and set $\beta = 0.95$, allowing us to compare a specification with no individual heterogeneity in both the risk and time parameters to the more general version of utility in Model 1. We choose to use $\beta = 0.95$ to mirror the value often used in the literature, usually chosen without consideration for whether this represents the actual time preferences of a sample. In practice, our sample has a lower discount rate than 0.95; the median value of β in the sample is 0.89. We therefore again use homogeneous values of β and θ in Model 4, but use the median values ($\beta = 0.89$ and $\theta = 0.94$), rather than the literature-consistent value of β and the mean value of θ used in Model 3. Finally, in Model 5, we allow β to vary across individuals again, but put in place a more stringent restriction on risk preference: risk neutrality or $\theta = 0$.

Table 9 shows the parameter estimate of ϕ for each of the five models, using both estimation approaches. The estimate of ϕ is significantly different than zero in the cross-sectional data with

no controls (Column 1) for all models. We see that the coefficients are similar when we control for demographic characteristics that might be correlated with the unobserved component of preferences for a major (Column 2). When we use the differenced pre- and post- treatment data (in the last two columns of the table), the estimate of ϕ is similar for Model 1 but smaller for all other models. With the exception of Model 5 (which assumes risk neutrality), the estimates are significantly different from 0 at conventional levels.[16]

Because we have major-specific observations for each person, we are also able to estimate individual-specific ϕ values by re-estimating equation 8 with the utility from income component $(f_{i2k} - f_{i1k})$ interacted with individual identifier dummy variables. The coefficient on this interaction for person i is then ϕ_i, an individual-specific scaling parameter for utility from income. Table 10 shows the distributional moments for these ϕ for Model 1 (truncated to exclude outliers below the 5th percentile and above the 95th percentile) for the full sample (row 1), women (row 2) , men (row 3), upperclassmen (row 4), and underclassmen (row 5). We use 200 bootstrap draws to estimate standard errors for these estimates and report whether the mean, 10th, 50th and 90th percentile are significantly different than zero.

The individual-specific estimates of ϕ suggest that there is a great deal of heterogeneity in how students adjusted their expected likelihood in choosing a major in response to changes in their expectations about the future earnings because of the information treatment. The average income scaling parameter is positive, consistent with the typical assumption that a major with higher expected earnings is, all else equal, preferred to a major with lower expected earnings. Men have significantly higher average ϕ values than women (p-value = 0.051), suggesting that men are more sensitive to potential future earnings when choosing a major. We also see higher parameter estimates for freshmen than upperclassmen (i.e., sophomores and juniors), consistent with the fact that upperclassmen are more likely to be decided on a course of study and face switching costs with changing major (p=0.201). Moreover, the other moments of the distribution make it clear that not all students respond to the prospect of more earnings as we might expect. The 10th percentile estimates indicate that a non-trivial proportion of the sample has negative values of ϕ, meaning that an increase in expected earnings makes these students less interested in choosing that major.

We next use the ϕ estimates from estimation Method 2 (column 3 of table 9) to calculate the

---

[16]Wiswall and Zafar (2015), using the same data set, regress the log relative probability of each major, differenced pre- and post- treatment, on a differenced version of log expected earnings at age 30. They find a positive relationship between earnings and major choice.

"taste" residual, α, using the pre-treatment expected likelihood:

$$\alpha_{ik} - \alpha_{i0} = \ln \frac{P_{i1k}}{P_{i10}} - \phi(f_{i1k})$$

By construction, our model's estimate of the likelihood of choosing each major is then exactly equal to the reported likelihood of choosing the major pre-treatment. To assess model fit, we compare the untargeted post-treatment major likelihood with the estimated probability of choosing each major post-treatment, using the post-treatment expected earnings and the estimated ϕ and α parameters. Appendix table A-7 reports the actual reported likelihood of choosing the major post-treatment and the model predictions for each model, for the full sample (Panel A) and for freshmen alone (Panel B). We are able to match the overall trends well, with the model predicting the correct ranking of major preferences and only deviating by less than 4 percentage points in value for all majors and models. Our model is more predictive of major preferences for freshmen, with deviations from the true value by less than 1 percentage point for economics/business, natural science, and not graduating and by less then 3 percentage points for engineering and humanities/social science.

6.3 Elasticities

To compare these parameter estimates across models as well as to previous papers' findings about the relationship between expected earnings and major choice, we must first convert these structural parameters into a form that allows us to compare across utility function types. For example, in Wiswall and Zafar (2015), the authors estimate a much larger value of the parameter that scales major-specific utility from earnings but use a very different utility function, implying that the magnitude of the scaling parameter in that paper is not comparable to the scaling parameter in the current paper.

To solve this problem, we compute major choice elasticities using the estimated structural parameters reported in Table 9. Specifically, we compute the percent change in the likelihood that a student chooses major k when earnings are increased by 10% in major k and held constant in all other majors. These choice elasticities are heterogeneous across individual student and will depend on the earnings distribution we estimated, their individual risk and discounting parameters, and the taste parameters implied by our estimates of the earnings scaling parameter, ϕ. For this exercise and throughout the main text of the paper, we report results based on estimation Method 2 (column 3 of Table 9); robustness checks using the non-differenced versions as well as the individual specific ϕ's are reported in Appendix C.

Table 11 displays the average percent change in the likelihood of choosing a major implied by each of the utility specifications in response to a 10% increase in earnings for each year of the student's life, using the ϕ estimated in column 3 of Table 9. To test whether we can reject the hypothesis that these elasticities are equal to zero, we report the 90-percent confidence intervals based on 200 bootstrapped estimates in brackets. We can reject an elasticity equal to 0 for Model 1 (heterogeneous β and θ), Model 2, and Model 3 at the $p < 0.05$ level and for Model 4 at the $p < 0.10$ level. We also conduct a t-test for equality of means across the models to compare the elasticities across utility specifications (denoted by asterisks on the elasticity estimates), and find that the elasticities are significantly higher for all majors in models without heterogeneity in risk and time preferences.

For our model with full heterogeneity in risk and discounting parameters (Model 1), excluding not graduating, we estimate average percent changes in the range 0.67-0.98, indicating that a 10 percent change in earnings for major k increases the probability of majoring in k by less than 1 percent. These estimates correspond to 1 percent earnings change elasticities in the range of 0.07- 0.1. These small elasticities are comparable to estimates found in previous papers, such as Arcidiacono (2004), Beffy et al. (2012), and Wiswall and Zafar (2015), which suggests that major choice is not particularly responsive to expected earnings. For example, Wiswall and Zafar (2015) estimate average elasticities of major choice to changes in future earnings in each period of between 0.03 and 0.07, which overlaps with our range of estimates.

A key finding for this paper is that these elasticities differ across utility specifications. Consistent with past work, we find higher major choice elasticities for the risk-neutral model (Model 5) compared to models which allow for risk aversion. Under risk neutrality, the elasticities are around double that of the estimates in the models with risk aversion, particularly for high-earning majors. This is consistent with the findings in Wiswall and Zafar (2015), who also report a higher elasticity of major choice for models that assume utility is linear in earnings.

Our novel finding, however, is that assumptions about the distribution of the discount rate and the coefficient of risk aversion also change the estimates of how responsive individuals' major choices are to changes in expected earnings. Past papers using subjective expectations data assume homogeneous values of β or θ because they do not have enough information to identify individual level values for the discount rate and the coefficient of risk aversion. A common assumption is that $\beta = 0.95$. By comparing the results for Model 1 (which uses heterogeneous values of β or θ) to those of models 2-4, we are able to see how much assuming homogeneity in risk/time preferences changes

estimates of income elasticity in major choice. Notably, Table 11 shows that the models with full homogeneity (Models 2-4) result in higher elasticities. In addition, the fact that elasticity estimates are quite similar for Models 2 and 3 (with the difference being that the latter also homogenizes β) suggests that it is the heterogeneity in risk preferences that seems to matter the most.[17]

For the remainder of the paper, we focus on comparing the most general model which allows for full heterogeneity in β and θ (Model 1) with the model which assumes both these parameters are homogeneous (Model 3), to highlight the effects of incorporating individual-specific risk and discounting preferences into models of major choice. Additional results for the other models are presented in Appendix C.2.

7 Counterfactuals

We next conduct a series of counterfactual exercises to explore how much the different majors' earnings profiles drive students' choices of major and whether the conclusions are sensitive to the heterogeneity in time and risk preferences. These counterfactuals provide a meaningful metric to understand the implied magnitude of our estimates.

The first counterfactual explores how much expectations about the variance of earnings within a major can explain students' major choices. To do this, we equalize the expected riskiness of earnings across majors for a student by setting the σ values (dispersion in log earnings) for each major equal to the student's σ values for humanities/social sciences. On average, earnings dispersion is lowest in humanities/social sciences out of all four major categories. Thus, since uncertainty in earnings makes more risk averse students less likely to choose a major, we should expect that this counterfactual will push students away from humanities/social sciences.

In our second counterfactual, we look at how students respond to expected earnings growth by equalizing the parameters that govern earnings growth across majors. For each student, we set the μ_1 and μ_2 parameters in equation (4) to be equal to the individual-specific estimates for humanities/social sciences, while not changing the absolute level of earnings (governed by μ_0) and uncertainty of earnings (governed by σ_0 and σ_1). On average, values for μ_1 and μ_2 are fairly similar for all college majors in the post-treatment data, but slightly smaller for not graduating. If students value high earnings growth, we therefore would expect that this counterfactual would shift away

---

[17]We, in fact, estimate an alternative model with homogeneous β but heterogeneous θ, and find average elasticities that are similar to Model 1 (results available from the authors upon request).

from not graduating at all. We also would expect that the effects of this counterfactual should be higher in Model 3 due to the higher value of β, which makes students discount future earnings less.

The third counterfactual tests whether the differences in levels of expected earnings across majors can explain students' major choices. To do this, we equalize the starting level of earnings across majors by setting μ_0 to be equal for all majors to the humanities/social sciences μ_0, but leave earnings growth and uncertainty unchanged. Since students on average believe that earnings in this major are lower than other majors, we would expect this counterfactual to push students into humanities/social sciences. This counterfactual does, on average, increase the starting earnings level for not graduating, which should increase the weight a student puts on their likelihood of not graduating.

Finally, we combine all three of these counterfactuals to see how students' likelihood of choosing each major would change if they expected the same earnings distribution across majors and equal to that of humanities/social sciences. To the extent that we think students are selecting into high-paying majors such as economics/business or engineering due to the higher anticipated earnings rather than due to a taste for those subjects, we would expect this counterfactual to induce students to increase their likelihood of choosing lower paying majors.

The various panels of Table 12 show the results of the four counterfactuals. Note that the effects of these counterfactuals are in general heterogeneous and depend on individual preferences and earnings beliefs. We report the median percent change in the likelihood of choosing a major under the counterfactual. The 10th and 90th percentile of the changes are reported in parentheses as well. For example, the median effect of counterfactual 3 on the likelihood of choosing an economics/ business major in Model 3 is a 3.20 percent decrease from a base rate of 30.26%, which is equivalent to the average likelihood decreasing by about 1.0 percentage point.

The results of these counterfactuals demonstrate that more favorable expectations about relative earnings profiles in economics/business, engineering, and natural sciences majors are one reason why students select these majors. Though the median effect of these counterfactuals are small, there is a great deal of heterogeneity in students' responses, as shown by the 10th and 90th percentiles of the distributions of percent changes. These students represent the marginal students. Some students in our sample have already decided on a major and report a very high percent likelihood of graduating in a major; for these students, the taste parameter, α, is large and changes in expected earnings would do little to change their mind about what major to choose. Other students, however, are not as unmovable: for example, counterfactual 4 makes students at the 90th percentile about 1.8

30

percent more likely to major in humanities/social sciences in Model 1 and causes students at the 10th percentile for economics/business to decrease their likelihood by 2.2 percent.

Taking account of heterogeneity, we also see that assuming all students have the same preference parameters can severely bias the inference regarding how much earnings matters for students. For example, in our fourth counterfactual, more responsive students (i.e., 10th percentile students) reported a 2.2 percent lower likelihood of choosing economics or business as a major in the model with full heterogeneity of risk and discounting parameters compared to a 12 percent lower likelihood in a model with $\beta = 0.95$ and $\theta = 1.7$.

These differences in counterfactual results, combined with the differences in elasticities across models, suggest that the common practice of calibrating risk parameters and discount rates may be over-estimating how sensitive individuals' choices are to their future income. The average student in our sample has a β lower than 0.95. The importance of incorporating heterogeneity in discounting and risk is most apparent in low-earning majors, such as humanities/social sciences and not graduating, which is consistent with the argument that students choose low-paying majors partially because they heavily discount future income relative to the present consumption value of taking classes in a major they enjoy.

8 Gender Gaps in Major Choice

Given that most studies find that there are gender differences in risk and patience (including ours), these studies typically conclude with the argument that these disparities in preference parameters likely translate to different human capital investment decisions for men and women.[18] Specifically, men may be more drawn to high-paying, high-risk majors or occupations.

We next use our preferred unrestricted model, Model 1, to decompose major choice into different components – expectations about earnings, risk preferences, discount rates, and taste – to evaluate the importance of each of these factors in explaining gender differences in earnings. We start by setting each respondent's parameters about earnings (μ_{ijk0}, μ_{ijk1}, μ_{ijk2}, σ_{ijk0}, and σ_{ijk1}), risk (θ_i), discounting (β_i) , and residual taste (α_{ik}) equal to the average values of these parameters for men in our sample. Note that for all our decomposition exercises related to earnings, we use the earnings estimates derived from the post-treatment earnings expectations.

---

[18]For example, Eckel and Grossman (2008)'s discussion of gender differences in risk aversion concludes by stating, "Differences in risk attitudes may play a role in the documented differences in earnings between women and men... [and] contribute to lower lifetime earnings and wealth for women."

Table 13 shows the evolution of the gender gap in the likelihood of choosing each major, defined as the average likelihood for women choosing that major subtracted from the average likelihood for men. At the baseline (column 1), both genders have the same likelihood of choosing each major, equal to the sample-level average for men. We then progressively re-introduce heterogeneity in each individual's utility parameters, starting with the σ parameters which govern earnings uncertainty (column 2), then μ_1 and μ_2 parameters which govern earnings growth (column 3), then μ_0 parameters which governs initial earnings (column 4), then the coefficient of risk aversion (column 5), then the discount rate (column 6), and then finally the residual "taste" component for the major, α (column 7).

Focusing on columns 2 through 6, we can see that the elements of the model related to earnings do very little to explain gender differences in major choice. We do not see any substantial gender gap until we add the "taste" residual in column 7. In fact, adding in the risk and discounting parameter moves the gender differences in choosing natural sciences in the *opposite* direction of the actual gender gaps. Though we see heterogeneity in the student's risk and time preferences across gender, these differences cannot fully explain why students choose different majors. This suggests that gender differences in preferences for risk or patience cannot explain why women select into lower-paying majors.

9 Conclusion

This paper combines an experiment, a survey of earnings expectations, and measures of risk preferences and discounting to investigate the role that risk aversion and time discounting play in education choices. The primary contribution of this analysis is to introduce two forms of heterogeneity previously under-studied in the college major choice literature, revealing four key findings about the relationship between major choice and subjective expectations about earnings.

First, our estimates of the distribution of risk and time preference parameters in a college-student population allow us to relate these traits with students' major choices, as well as with demographic characteristics such as gender. We find that men are less risk averse and more impatient than women, which is consistent with previous analyses on more representative samples. However, while these traits systematically vary across gender, they do not seem to be systematically associated with individuals' choices of major.

Next, the richness of our data set, which combines a survey of subjective expectations about

labor market outcomes, an experiment, and a set of hypothetical games to elicit risk and time preferences, allows us to estimate a more general model of the college-major decision than is typically possible. Because we can use our data to identify each individual's discount factor and coefficient of risk aversion, we are able to compare a general model that allows for this heterogeneity to more restricted models. We show that models without heterogeneity in these parameters are likely to overstate how sensitive college major choice is to earnings. It is, in particular, the heterogeneity in risk preferences that seem to matter in drawing conclusions regarding choice elasticities.

Third, we show that earnings expectations play a role in the choice of majors, but that this role is small compared to the role of other characteristics of the major for most students. Our counterfactuals show that there are marginal students who are very responsive to changes in the expected earnings path associated with a major, with the largest changes in the estimated likelihood of choosing majors occurring in the counterfactual where we equalize the level of earnings, rather than the earnings growth or earnings uncertainty. However, major-choices in our data are still largely the result of heterogeneity in the unobserved "taste" component of major-choice. In our framework, one can think of tastes as preferences for major-specific outcomes realized in college (e.g., enjoying the courses), or major-specific post-graduation outcomes (e.g., work-life balance in a particular occupation related to major). These could be pecuniary or non-pecuniary in nature.

Lastly, our analysis of the gender gap in major choice suggests that gender differences in risk preferences and patience cannot explain the pattern of women being less likely to select into high-paying, but risky, majors such as Business/Economics. Earnings expectations differences across genders also cannot explain this gender gap. We find that the gender gap in major choice stems primarily from the unobserved taste component, not the heterogeneity in β or θ. In light of the finding that there are gender differences in risk aversion and patience in lab settings, researchers often conclude that these gender differences could contribute to broader gender inequalities in the labor market. Our results suggest that while these preferences vary across gender, there does not appear to be a robust link between gender gaps in these preferences and gender gaps in major choice. It could, however, be the case that these preferences influence the jobs and careers that students choose conditional on a major. In fact, Wiswall and Zafar (2018a) show that students perceive substantial variation in non-wage amenities of jobs they expect would be available if they were to graduate with different majors. Importantly, there is substantial heterogeneity in the kinds of workplaces – in terms of earnings, earnings risk, job stability – that students expect to enter even conditional on major. Future work that investigates the role of time and risk preferences, and

gender differences within them, on sorting into careers would be valuable.

References

J. G. Altonji. The Demand for and Return to Education When Education Outcomes are Uncertain. *Journal of Labor Economics*, 11(1, Part 1):48–83, jan 1993. ISSN 0734-306X. doi: 10.1086/ 298317. URL https://www.journals.uchicago.edu/doi/10.1086/298317.

S. Andersen, G. W. Harrison, M. I. Lau, and E. E. Rutström. Eliciting Risk and Time Preferences. *Econometrica*, 76(3):583–618, may 2008. ISSN 0012-9682. doi: 10.1111/j.1468-0262.2008.00848.x. URL http://doi.wiley.com/10.1111/j.1468-0262.2008.00848.x.

P. Arcdiacono, V. J. Hotz, A. Maurel, and T. Romano. 10.3386/w20626. Technical report, National Bureau of Economic Research, Cambridge, MA, oct 2014. URL https://www.nber.org/papers/w20626.pdf.

P. Arcidiacono. Ability sorting and the returns to college major. *Journal of Econometrics*, 121(1-2):343–375, 2004. URL https://ideas.repec.org/a/eee/econom/v121y2004i1-2p343-375.html.

P. Arcidiacono, E. M. Aucejo, and K. Spenner. What happens after enrollment? An analysis of the time path of racial differences in GPA and major choice. *IZA Journal of Labor Economics*, 1(1):5, dec 2012. ISSN 2193-8997. doi: 10.1186/2193-8997-1-5. URL http://izajole.springeropen.com/articles/10.1186/2193-8997-1-5.

O. P. Attanasio and K. M. Kaufmann. Education choices and returns to schooling: Mothers' and youths' subjective expectations and their role by gender. *Journal of Development Economics*, 109:203–216, 2014.

O. P. Attanasio and K. M. Kaufmann. Education choices and returns on the labor and marriage markets: Evidence from data on subjective expectations. *Journal of Economic Behavior & Organization*, 140:35–55, aug 2017. doi: https://doi.org/10.1016/j.jebo.2017.05.002.

G. Bamberger. Occupational choice: The role of undergraduate education. 1988.

D. Barth, S. H. Shore, and S. T. Jensen. Identifying idiosyncratic career taste and skill with income

risk. *Quantitative Economics*, 8(2):553–587, jul 2017. ISSN 17597323. doi: 10.3982/QE424. URL http://doi.wiley.com/10.3982/QE424.

G. S. Becker and C. B. Mulligan. The endogenous determination of time preference. *The Quarterly Journal of Economics*, 112(3):729–758, 1997.

M. Beffy, D. Fougère, and A. Maurel. Choosing the Field of Study in Postsecondary Education: Do Expected Earnings Matter? *Review of Economics and Statistics*, 94(1):334–347, feb 2012. ISSN 0034-6535. doi: 10.1162/REST_a_00212. URL http://www.mitpressjournals.org/doi/10.1162/REST{_}a{_}00212.

C. Belzil and M. Leonardi. Risk Aversion and Schooling Decisions. *Annals of Economics and Statistics*, No.111/112 Special Issue on Education:35–70, 2013. doi: 10.2307/23646326.

D. J. Benjamin, J. J. Choi, and A. J. Strickland. Social Identity and Preferences. *American Economic Review*, 100:1913–1928, sept 2010. doi: 10.1257/aer.100.4.1913.

M. C. Berger. Predicted Future Earnings and Choice of College Major. *ILR Review*, 41(3):418–429, apr 1988. ISSN 0019-7939. doi: 10.1177/001979398804100306. URL http://journals.sagepub.com/doi/10.1177/001979398804100306.

H. Bonin, T. Dohmen, A. Falk, D. Huffman, and U. Sunde. Cross-sectional earnings risk and occupational sorting: The role of risk attitudes. *Labour Economics*, 14(6):926–937, dec 2007. URL https://www.sciencedirect.com/science/article/pii/S0927537107000462.

T. Brodaty, R. J. Gary-Bobo, and A. Prieto. Do risk aversion and wages explain educational choices? *Journal of Public Economics*, 117:125–148, sept 2014. doi: https://doi.org/10.1016/j.jpubeco.2014.05.003.

G. Brunello, C. Lucifora, and R. Winter-Ebmer. The wage expectations of european business and economics students. *Journal of Human Resources*, 39(4):1116–1142, 2004.

B. C. Cadena and B. J. Keys. Human Capital and the Lifetime Costs of Impatience. *American Economic Journal: Economic Policy*, 7(3):126–153, aug 2015. ISSN 1945-7731. doi: 10.1257/pol.20130081. URL http://pubs.aeaweb.org/doi/10.1257/pol.20130081.

M. Coller and M. B. Williams. Eliciting Individual Discount Rates. Technical report, 1999. URL http://community.middlebury.edu/{~}jcarpent/EC499/CollerandWilliams(1999)EE.pdf.

R. Croson and U. Gneezy. Gender Differences in Preferences. *Journal of Economic Literature*, 47 (2):448–474, may 2009. ISSN 0022-0515. doi: 10.1257/jel.47.2.448. URL http://pubs.aeaweb.org/doi/10.1257/jel.47.2.448.

C. Dave, C. Eckel, C. Johnson, and C. Rojas. Eliciting risk preferences: When is simple better? *Journal of Risk and Uncertainty*, 41(3):219–243, 12 2010. ISSN 0895-5646. doi: 10.1007/s11166-010-9103-z.

M. De Paola and F. Gioia. Risk aversion and field of study choice: The role of individual ability. *Bulletin of Economic Research*, 64:s193–s209, dec 2012. URL https://doi.org/10.1111/j.1467-8586.2012.00445.x.

E. W. Dillon. Risk and Return Trade-Offs in Lifetime Earnings. *Journal of Labor Economics*, 36 (4):981–1021, oct 2018. ISSN 0734-306X. doi: 10.1086/697475. URL https://www.journals.uchicago.edu/doi/10.1086/697475.

M. Dittrich and K. Leipold. Gender differences in time preferences. *Economics Letters*, 122(3): 413–415, 2014.

C. C. Eckel and P. J. Grossman. Sex differences and statistical stereotyping in attitudes toward financial risk. *Evolution and Human Behavior*, 23(4):281–295, jul 2002. ISSN 1090-5138. doi: 10.1016/S1090-5138(02)00097-1. URL https://www.sciencedirect.com/science/article/pii/S1090513802000971.

C. C. Eckel and P. J. Grossman. Forecasting risk attitudes: An experimental study using actual and forecast gamble choices. *Journal of Economic Behavior & Organization*, 68(1):1–17, oct 2008. ISSN 0167-2681. doi: 10.1016/J.JEBO.2008.04.006. URL https://www.sciencedirect.com/science/article/pii/S0167268108000978.

E. Eide and G. Waehrer. The role of the option value of college attendance in college major choice. *Economics of Education Review*, 17(1):73–82, 1998. URL https://EconPapers.repec.org/RePEc:eee:ecoedu:v:17:y:1998:i:1:p:73-82.

A. Falk and J. Hermle. Relationship of gender differences in preferences to economic development and gender equality. *Science*, 362(6412):eaas9899, 2018.

F. A. Flyer. The Influence of Higher Moments of Earnings Distributions on Career Decisions. *Journal of Labor Economics*, 15(4):689–713, oct 1997. ISSN 0734-306X. doi: 10.1086/209842. URL https://www.journals.uchicago.edu/doi/10.1086/209842.

D. Fouarge, B. Kriechel, and T. Dohmen. Occupational sorting of school graduates: The role of economic preferences. *Journal of Economic Behavior & Organization*, 106:335–351, oct 2014. ISSN 0167-2681. URL https://www.sciencedirect.com/science/article/pii/S016726811400208X.

S. Frederick, G. Loewenstein, and T. O'Donoghue. Time discounting and time preference: A critical review. *Journal of Economic Literature*, 40(2):351–401, 2002. ISSN 00220515. URL http://www.jstor.org/stable/2698382.

R. Freeman. The market for college trained manpower. 1971.

P. Giustinelli. Group Decision Making with Uncertain Outcomes: Unpacking Child-Parent Choice of the High School Track *. *International Economic Review*, 57 (2), May 2016. URL https://economics.wustl.edu/files/economics/imce/vrptrackchoicepaper{_}sept13{_}pgiusti.pdf.

C. A. Holt and S. K. Laury. Risk Aversion and Incentive Effects. *American Economic Review*, 92: 1644–1655, dec 2002. doi: 10.1257/000282802762024700.

K. M. Kaufmann. Understanding the income gradient in college attendance in Mexico: The role of heterogeneity in expected returns. *Quantitative Economics*, 5(3):583–630, nov 2014. ISSN 17597323. doi: 10.3982/QE259. URL http://doi.wiley.com/10.3982/QE259.

S. D. Levitt, J. A. List, and S. Sadoff. The effect of performance-based incentives on educational achievement: Evidence from a randomized experiment. Working Paper 22107, National Bureau of Economic Research, March 2016. URL http://www.nber.org/papers/w22107.

L. Lochner. Individual Perceptions of the Criminal Justice System. *American Economic Review*, 97 (1):444–460, feb 2007. ISSN 0002-8282. doi: 10.1257/aer.97.1.444. URL http://pubs.aeaweb.org/doi/10.1257/aer.97.1.444.

T. Magnac and D. Thesmar. Identifying Dynamic Discrete Decision Processes. *Econometrica*, 70: 801–816, mar 2002.

C. F. Manski. Adolescent econometricians: How do youth infer the returns to schooling? In *Studies of supply and demand in higher education*, pages 43–60. University of Chicago Press, 1993.

C. Montmarquette, K. Cannings, and S. Mahseredjian. How do young people choose college majors? Technical report, 2002. URL www.elsevier.com/locate/econedurev.

M. Niederle. Gender. Working Paper 20788, National Bureau of Economic Research, December 2014. URL http://www.nber.org/papers/w20788.

H. S. Nielsen and A. Vissing-Jorgensen. The Impact of Labor Income Risk on Educational Choices: Estimates and Implied Risk Aversion *. Technical report, 2006. URL https://pdfs.semanticscholar.org/2a43/c98022c53b654566e129079a5c975d79a977.pdf?{_}ga=2.117443657.1364505584.1552779231-1306858675.1552779231.

P. Oreopoulos. Do dropouts drop out too soon? wealth, health and happiness from compulsory schooling. *Journal of Public Economics*, 91(11-12):2213–2229, 2007. URL https://EconPapers.repec.org/RePEc:eee:pubeco:v:91:y:2007:i:11-12:p:2213-2229.

A. I. Ruder and M. Van Noy. Knowledge of earnings risk and major choice: Evidence from an information experiment. *Economics of Education Review*, 57:80–90, apr 2017. doi: https://doi.org/10.1016/j.econedurev.2017.02.001.

J. Rust. Structural estimation of markov decision processes. In R. F. Engle and D. McFadden, editors, *Handbook of Econometrics*, chapter 51, pages 3082–3143. Elsevier Press, 1994.

R. E. Saks and S. H. Shore. Risk and Career Choice. *The B.E. Journal of Economic Analysis & Policy*, 5(1), jan 2005. ISSN 1935-1682. doi: 10.1515/1538-0637.1414. URL https://www.degruyter.com/view/j/bejeap.2005.5.issue-1/bejeap.2005.5.1.1414/bejeap.2005.5.1.1414.xml.

A. Siow. Occupational Choice under Uncertainty. *Econometrica*, 52(3):631–645, May 1984. URL https://ideas.repec.org/a/ecm/emetrp/v52y1984i3p631-45.html.

T. Stinebrickner and R. Stinebrickner. Learning about Academic Ability and the College Dropout Decision. *Journal of Labor Economics*, 30(4):707–748, oct 2012. ISSN 0734-306X. doi: 10.1086/666525. URL https://www.journals.uchicago.edu/doi/10.1086/666525.

T. Stinebrickner and R. Stinebrickner. Academic Performance and College Dropout: Using Longitudinal Expectations Data to Estimate a Learning Model. Technical report, National Bureau of Economic Research, Cambridge, MA, apr 2013. URL http://www.nber.org/papers/w18945.pdf.

M. Wiswall and B. Zafar. Determinants of College Major Choice: Identification using an Information Experiment. *The Review of Economic Studies*, 82(2):791–824, apr 2015. ISSN 0034-6527. doi: 10.1093/restud/rdu044. URL https://academic.oup.com/restud/article-lookup/doi/10.1093/restud/rdu044.

M. Wiswall and B. Zafar. Preference for the Workplace, Investment in Human Capital, and Gender. *The Quarterly Journal of Economics*, 133:457–507, feb 2018a. doi: https://doi.org/10.1093/qje/qjx035.

M. Wiswall and B. Zafar. Human capital investments and expectations about career and family. (22543), 2018b.

B. Zafar. How Do College Students Form Expectations? *Journal of Labor Economics*, 29(2):301–348, apr 2011. ISSN 0734-306X. doi: 10.1086/658091. URL https://www.journals.uchicago.edu/doi/10.1086/658091.

B. Zafar. College Major Choice and the Gender Gap. *Journal of Human Resources*, 48(3):545–595, jul 2013. ISSN 0022-166X. doi: 10.3368/jhr.48.3.545. URL http://jhr.uwpress.org/lookup/doi/10.3368/jhr.48.3.545.

G. A. Zarkin. Occupational Choice: An Application to the Market for Public School Teachers. *The Quarterly Journal of Economics*, 100(2):409–446, 1985. URL https://ideas.repec.org/a/oup/qjecon/v100y1985i2p409-446..html.

Tables and Figures
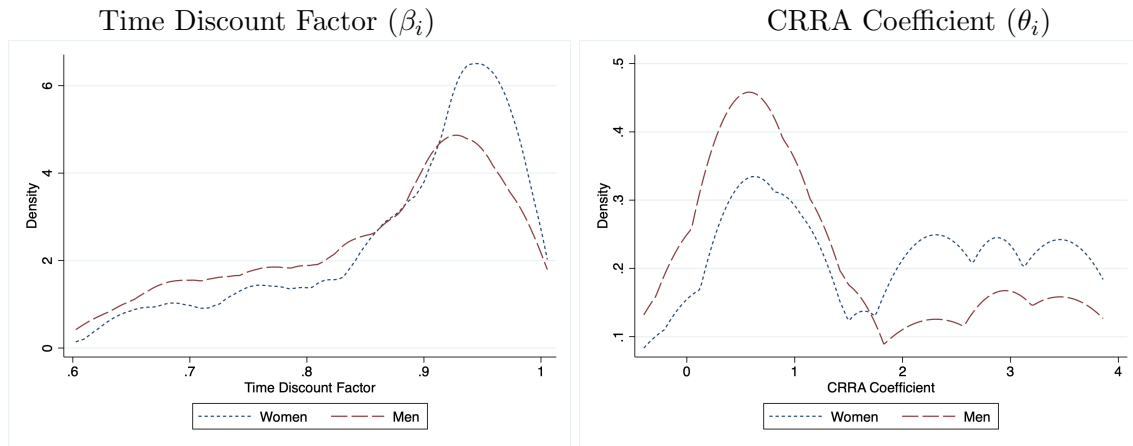
Figure 1: Kernel Density of Risk and Time Preference Parameters

Note that the Kolmogorov-Smirnov Test results show that men have β_i that is smaller than women at the 1 percent level ($p = 0.059$) and θ_i that is smaller than women at the 1 percent level ($p = 0.001$).

Panel A: Earnings Expectations at Age 22

Time Discount Factor (β_i) | CRRA Coefficient (θ_i)
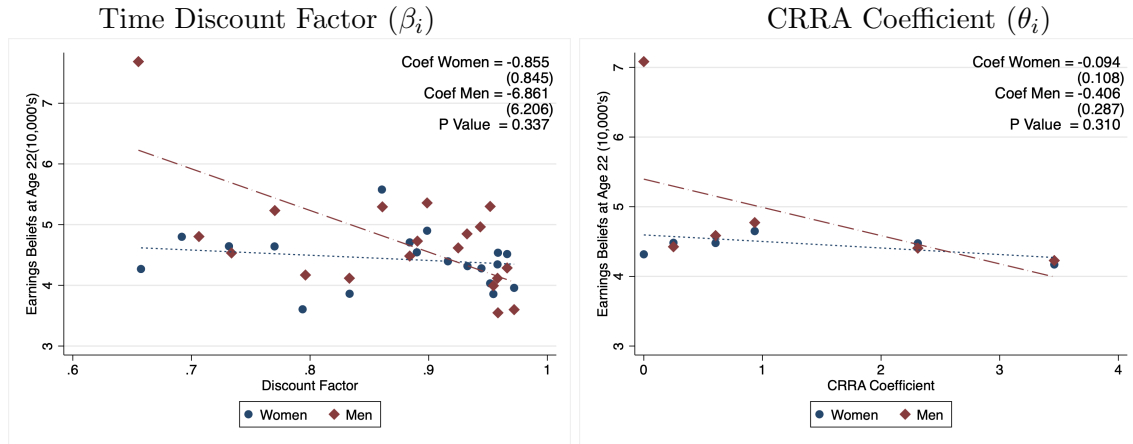
Panel B: Earnings Growth between Age 22 and Age 30

Time Discount Factor (β_i) | CRRA Coefficient (θ_i)
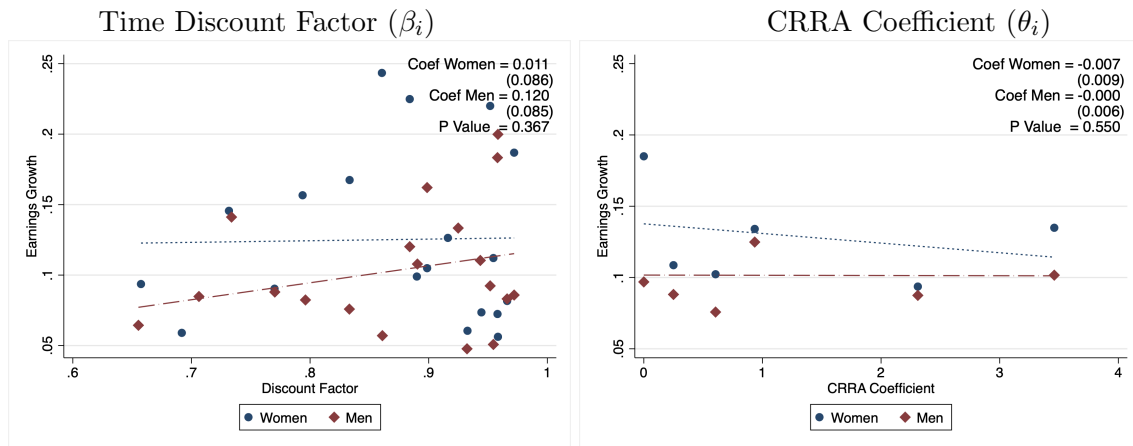
Figure 2: Bin Scatter of Earnings Expectations with Risk and Time Preference Parameters

Note. This figure shows the binned scatter plot between earnings expectations, and discount factor and CRRA risk coefficient. Panel A shows this plot for age 22 earnings and Panel B shows the same plot for earnings growth between age 22 and age 30. The earnings are top and bottom coded at \$500,000 and \$10,000 respectively. We pool the data across majors so that we have 5 observations per individual. The figure also reports the coefficient and the standard error for the linear regression of the earnings expectations on the discount factor and CRRA coefficient (with no other controls) for men and women separately. We also report the p-value of the test of difference between the coefficients for men and women. Standard errors are clustered at the respondent level for the pooled major regressions. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
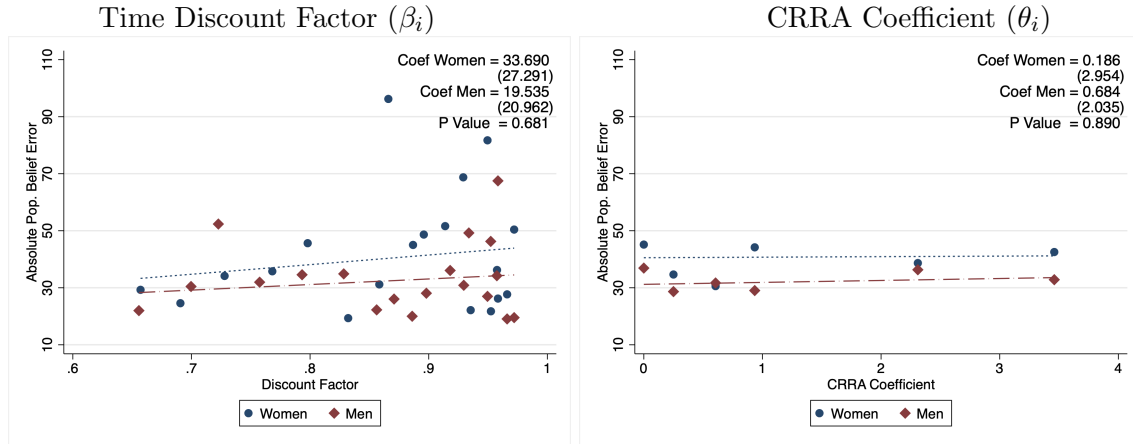
Figure 3: Bin Scatter of Absolute Value of Population Belief Error with Risk and Time Preference Parameters

Note. This figure shows the binned scatter plot of the absolute value of population belief error with discount factor and CRRA risk coefficient, separately by gender. The absolute population error for individual i of gender g for major k is defined as the $|\frac{\text{Earnings Beliefs}_{igk} - \text{Truth}_{gk}}{\text{Truth}_{gk}}|$ for major specific full time population earnings at age 30. The earnings are top and bottom coded at \$500,000 and \$10,000 respectively. We pool the data across majors so that we have 5 observations per individual. The figure also reports the coefficient and the standard error for the linear regression of the error on the discount factor and CRRA coefficient (with no other controls) for men and women separately. We also report the p-value of the test of difference between the coefficients for men and women. Standard errors are clustered at the respondent level for the pooled major regressions. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 1: Sample Statistics

	Full Sample		Men		Women	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
SAT Math	700.460	(77.442)	712.905	(81.175)	692.881***	(74.228)
SAT Verbal	682.577	(72.579)	680.426	(75.493)	683.893	(70.865)
White	0.386	(0.487)	0.429	(0.496)	0.361	(0.481)
Asian	0.446	(0.498)	0.448	(0.499)	0.445	(0.498)
Freshman	0.417	(0.494)	0.474	(0.501)	0.384 *	(0.487)
Sophomore	0.365	(0.482)	0.325	(0.470)	0.388	(0.488))
Parental Income (\$10,000)	14.843	(12.265)	15.775	(12.921)	14.294	(11.853)
Observations	417		154		263	

Note. This table reports mean values for a set of demographic characteristics for the full sample and the sample split by gender. Standard deviations are in parentheses. SAT scores have range from 0 to 800; demographic indicators are percentages with range 0 to 1; Parent income is scaled in \$10,000 increments.* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ denote significance for test of equality of means by gender.

Table 2: Beliefs about Earnings by Treatment Status

		Pre- Treatment		Post Treatment	
		Men	Women	Men	Women
Panel A :	Economics	12.27	10.91	9.947^{++}	8.062^{***+++}
	Engineering	9.016	9.765	8.428	8.190^{+++}
Average Earnings ($ 10,000) at Age	Humanities	6.441	6.712	6.751	6.076
30	Natural Sciences	8.266	9.392	7.641	7.333^{+++}
	Not Graduate	4.374	3.177***	5.234	4.123^{***++}
Panel B :	Economics	68.52	67.15	55.03^{+++}	46.21^{***+++}
	Engineering	60.34	63.22	54.26^{++}	46.72 $^{***+++}$
Percent Chance of Earnings >	Humanities	43.94	42.02	38.90^{+}	31.48^{***+++}
85,000 at Age 30	Natural Sciences	55.14	55.16	48.03^{++}	40.03^{***+++}
	Not Graduate	17.97	13.39***	21.13	11.68***
Panel C :	Economics	92.44	91.04	87.50 $^{+++}$	85.61^{+++}
	Engineering	90.10	88.54	87.38	86.83
Percent Chance of Earnings >	Humanities	83.51	78.22***	80.37	77.68
35,000 at Age 30	Natural Sciences	86.24	83.85	84.89	82.22
	Not Graduate	54.32	43.67***	61.53^{++}	47.67 $^{***+}$
Panel D :	Economics	16.22	13.58 **	13.89	9.502^{***+++}
	Engineering	11.44	11.54	11.56	9.710^{**+++}
Average Earnings ($ 10,000) at Age	Humanities	9.989	9.014	8.980	7.193 *** $^{+++}$
45	Natural Sciences	11.19	11.10	10.09	8.688^{**+++}
	Not Graduate	6.989	5.430**	6.925	4.835^{***+++}
Panel E :	Economics	77.68	76.20	63.75^{+++}	56.16^{***+++}
	Engineering	71.01	72.97	63.69^{++}	55.52^{***+++}
Percent Chance of Earnings >	Humanities	59.30	58.85	49.32^{+++}	40.15^{***+++}
85,000 at Age 45	Natural Sciences	66.61	66.94	57.65^{+++}	49.24*** $^{+++}$
	Not Graduate	30.30	23.68 ***	26.67	16.07^{***+++}
Panel F :	Economics	91.43	90.44	89.80	88.41
	Engineering	89.66	89.44	89.08	89.33
Percent Chance of Earnings >	Humanities	85.80	84.95	84.61	81.09^{*++}
35,000 at Age 45	Natural Sciences	88.33	87.94	87.36	85.32
	Not Graduate	64.73	55.11 ***	66.16	55.17***

Note. This table reports the moments used to estimate the earnings distribution, by gender and pre- and post- treatment. Panel A shows the average earnings expected at age 30 conditional on major, in $10,000 units. Panel B shows the average likelihood that students expect they will earn greater than $85,000 at age 30 conditional on major, with scale 0 to 100. Panel C shows the average likelihood that students expect they will earn greater than $35,000 at age 30 conditional on major, with scale 0 to 100. Panel D shows the average earnings expected at age 45 conditional on major, in $10,000 units. Panel E shows the average likelihood that students expect they will earn greater than $85,000 at age 45 conditional on major, with scale 0 to 100. Panel F shows the average likelihood that students expect they will earn greater than $35,000 at age 45 conditional on major, with scale 0 to 100. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ denote significance for test of equality of means by gender within treatment. + $p < 0.1$, ++ $p < 0.05$, +++ $p < 0.01$ denote significance for test of equality of means by treatment within gender.

Table 3: Beliefs about Likelihood of Graduating

	Pre- Treatment		Post Treatment	
	Men	Women	Men	Women
Economics	36.71	23.78***	36.31	26.63***
Engineering	8.714	5.624**	11.24	8.376*++
Humanities	31.39	49.87***	28.30	44.98
Natural Sciences	19.72	18.82	21.47	17.67***
Not Graduate	3.461	1.901*	2.688	2.342

Note. This table reports the average reported likelihood of graduating in each of the major categories (economics/business, engineering, humanities/ social sciences, natural sciences/ math, and not graduating) on a scale from 0 to 100. To elicit these beliefs, students were asked: "What do you believe is the percent chance (or chances out of 100) that you would either graduate from NYU with a major in the following major categories or that you would never graduate/drop-out (i.e., you will never receive a Bachelor's degree from NYU or any other university)?" * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ denote significance for test of equality of means by gender within treatment. + $p < 0.1$, ++ $p < 0.05$, +++ $p < 0.01$ denote significance for test of equality of means by treatment within gender.

Table 4: Risk and Time Discount Preference, by Gender

		Full Sample	Men	Women	p-value, H_0: Men = Women
	Mean	0.874	0.858	0.884***	0.008
Panel A:	SD	0.097	0.102	0.094	0.242
Time Discount Factor	Median	0.899	0.888	0.918**	0.032
(β_i)	10th Percentile	0.722	0.699	0.728	0.295
	90th Percentile	0.966	0.958	0.972***	0.000
	Mean	1.660	1.38	1.824***	0.0001
Panel B:	SD	1.241	1.214	1.230	0.863
Coefficient of Relative	Median	0.935	0.935	2.310***	0.000
Risk Aversion (θ_i)	10th Percentile	0.000	0.000	0.000	-
	90th Percentile	3.460	3.46	3.46	-
	Observations	417	154	263	

Note. This table reports the following statistics for the discount factor (β) in panel A and the coefficient of relative risk aversion (θ) in panel B: mean, the standard deviation, and the 50th, 10th, and 90th percentiles of the distribution for the full sample, men, and women. Details on estimation of β and θ are in appendix section A.5. The last column reports the p-value for whether we can reject the null hypothesis that these statistics are equal across gender. We conduct pairwise t-tests for the difference in means between men and women and a test on the equality of standard deviations. For the three percentiles, we run a quantile regression of the parameter on an indicator for male and a constant term; the p-value indicates whether the coefficient on 'men' was significant in these regressions. For all tests, *, **, and *** represent significance at the 10%, 5% and 1% level respectively

Table 5: Risk and Time Discount Preferences, by Reported Major and Gender

Panel A: Full Sample		All	Economics/ Business	Engineering	Humanities	Natural Science	p-value
	Mean	0.874	0.858	0.852	0.883	0.882	0.056
Panel A1:	SD	0.097	0.100	0.114	0.095	0.092	0.618
Discount Factor (β_i)	Median	0.899	0.888	0.888	0.928	0.899	-
	10th Percentile	0.707	0.680	0.728	0.707	0.728	-
	90th Percentile	0.966	0.966	0.966	0.972	0.966	-
	Mean	1.66	1.62	1.39	1.78	1.59	0.370
Panel A2:	SD	1.241	1.232	1.22	1.29	1.12	0.936
Coefficient of Relative	Median	0.935	0.935	0.935	2.31	0.935	-
Risk Aversion (θ_i)	10th Percentile	0.000	0.250	0.000	0.000	0.000	-
	90th Percentile	3.46	3.46	3.46	3.46	3.46	-
	Observations	417	117	21	204	75	
Panel B: Men		All	Economics/ Business	Engineering	Humanities	Natural Science	p-value
	Mean	0.858	0.884	0.871	0.859	0.868	0.325
Panel B1:	SD	0.102	0.097	0.107	0.105	0.106	0.715
Discount Factor (β_i)	Median	0.888	0.884	0.910	0.884	0.899	-
	10th Percentile	0.699	0.699	0.667	0.707	0.704	-
	90th Percentile	0.958	0.958	0.958	0.958	0.962	-
	Mean	1.38	1.42	1.51	1.45	1.14	0.972
Panel B2:	SD	1.22	1.27	1.15	1.32	0.925	0.945
Coefficient of Relative	Median	0.94	0.94	0.935	0.94	0.94	-
Risk Aversion (θ_i)	10th Percentile	0.000	0.000	0.000	0.593	0.125	-
	90th Percentile	3.46	3.46	3.46	3.46	2.31	-
	Observations	154	59	10	55	30	
Panel C: Women		All	Economics/ Business	Engineering	Humanities	Natural Science	p-value
	Mean	0.877	0.859	0.857	0.880	0.900	0.035
Panel C1:	SD	0.097	0.106	0.132	0.099	0.081	0.567
Discount Factor (β_i)	Median	0.899	0.888	0.924	0.930	0.932	-
	10th Percentile	0.722	0.667	0.659	0.707	0.774	-
	90th Percentile	0.966	0.972	0.972	0.966	0.972	-
	Mean	1.87	1.76	1.26	1.88	2.13	0.299
Panel C2:	SD	1.26	1.19	1.35	1.27	1.28	0.859
Coefficient of Relative	Median	2.31	2.31	0.935	2.31	2.31	-
Risk Aversion (θ_i)	10th Percentile	0.25	0.250	0.000	0.250	0.250	-
	90th Percentile	3.46	3.46	3.46	3.46	3.46	-
	Observations	263	58	11	149	45	

Note. This table reports the following statistics for the discount factor (β) and the coefficient of relative risk aversion (θ): mean, the standard deviation, and the 50th, 10th, and 90th percentiles of the distribution for the full sample, men, and women, broken down by the major students reported they were planning on choosing at the time of the survey. Details on estimation of β and θ are in appendix section A.5. The last column shows the p-statistic for an ANOVA test of the null hypothesis that the means are jointly equal across majors and the p-statistic for Bartlett's test for equal variances across major groups.

Table 6: Regression of Measures on Observable Characteristics

	Discount Factor			CRRA Coefficient		
	Overall	Men	Women	Overall	Men	Women
Male	-0.029**			-0.432***		
	(0.010)			(0.125)		
SAT Math (in 100s)	-0.004	0.014	-0.016*	-0.187**	-0.114	-0.226*
	(0.007)	(0.013)	(0.009)	(0.086)	(0.132)	(0.116)
SAT Verbal(in 100s)	0.008	-0.001	0.007	-0.131	-0.233	-0.093
	(0.008)	(0.013)	(0.010)	(0.093)	(0.149)	(0.121)
White	0.046**	-0.020	0.075***	-0.067	-0.387	0.039
	(0.014)	(0.024)	(0.016)	(0.181)	(0.336)	(0.220)
Asian	0.017	-0.047*	0.044**	-0.050	-0.523	0.142
	(0.016)	(0.027)	(0.018)	(0.178)	(0.333)	(0.218)
Parent Income ($100,000)	0.002	-0.003	0.006	0.064	0.073	0.057
	(0.004)	(0.006)	(0.005)	(0.051)	(0.081)	(0.069)
Observations	417	154	263	417	154	263
R^2	0.07	0.03	0.12	0.06	0.07	0.02
F-stat (excluding Male)	3.71***	0.77	6.22***	2.66**	2.16**	1.21

Note. This table regresses the two preference parameters, the discount factor β_i and the CRRA coefficient θ_i, on a series of demographic characteristics and also shows the regression results for subsamples of men and women in the sample. Omitted from the table are a constant, indicator variables for if the SAT scores are missing, and an indicator variable for if parental income is missing. SAT scores are scaled to be in terms of 100 point increments. Parental earnings are scaled to be in $100,000 increments. We report the F-stat for a joint test for significance of coefficients excluding the coefficient on gender. Standard errors are in parentheses,* $p < 0.10$, ** $p < 0.05$, *** $p < 0.001$

Table 7: Major Choice and Risk and Time Preferences

	Log Probability of Graduating in Major (Relative to Hum.)			
	Economics	Engineering	Natural Science	No Graduation
Men	3.344***	2.587***	1.868**	2.369***
	(0.809)	(0.775)	(0.770)	(0.688)
Std. Discount Factor	-0.971**	-0.0163	-0.267	-0.492
	(0.397)	(0.341)	(0.368)	(0.349)
Std. Risk Coefficient	-0.0181	-0.213	-0.182	0.198
	(0.407)	(0.374)	(0.375)	(0.342)
R^2	0.189	0.135	0.144	0.076
Observations	417	417	417	417
Controls	Yes	Yes	Yes	Yes

Note. The outcome variable in these regressions is the log relative probability of graduating in each major relative to the omitted major, humanities/ social sciences. We regress this on an indicator for gender and standardized versions of the discount factor and the CRRA coefficient (standardized to have mean = 0, SD = 1). We include controls for SAT math score, SAT verbal scores, dummies for missing SAT math and SAT verbal scores, dummies for White and Asian race categories (the omitted category is all other races) and expected earnings at age 22 (relative to humanities). Standard errors are in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8: Population Belief Error by Major and Gender

	Percent Error					
	Actual			Absolute		
	All	Men	Women	All	Men	Women
Economics	18.88	4.188	27.14***	42.46	34.96	46.68*
Engineering	-10.46	-19.29	-5.489**	31.64	29.96	32.58
Natural Sciences	5.916	-12.82	16.45***	39.74	32.12	44.02
Humanities	8.341	0.754	12.61	34.42	28.38	37.82
Not Graduate	-7.483	-22.99	1.234**	40.33	35.14	43.25
Observations	1570	565	1005	1570	565	1005
Number of Respondents	314	113	201	314	113	201

Note. Values indicate the average value of the parameter for the sample by gender, where percent error for individual i of gender g for major k is defined as $\frac{\text{Earnings Beliefs}_{igk} - \text{Truth}_{gk}}{\text{Truth}_{gk}}$, and absolute percent error is defined as $|\frac{\text{Earnings Beliefs}_{igk} - \text{Truth}_{gk}}{\text{Truth}_{gk}}|$ for major specific full time population earnings at age 30. Earnings referring to full-time major specific population earnings at age 30. Columns 1-3 average over the realized values of the error with positive values indicating an overestimate of the population distribution's mean and negative values indicating an underestimate; columns 4-6 average over the absolute values of the error indicating the typical magnitude of the error regardless of direction. Note that the error is not defined for men who received the treatment about female earnings and females who received the treatment about male earnings. For more details on the treatment, refer to Appendix A1. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ denote significance for test of equality of means by gender.

Table 9: Estimates of Preference Parameter: ϕ

	Method 1	Method 1 w/ controls	Method 2	Method 2, Truncated Sample
	Cross-Sectional Data		Pre/Post Treatment Diff. Data	
Model 1	0.00124*	0.00096*	0.0014*	0.0014*
	(0.000545)	(0.000485)	(0.00079)	(0.00075)
R^2	0.072	0.183	0.025	0.024
Model 2	0.0437***	0.0422***	0.0245**	0.0245**
	(0.008)	(0.008)	(0.0123)	(0.0122)
R^2	0.087	0.198	0.024	0.020
Model 3	0.0620***	0.0628***	0.0172*	0.0151*
	(0.009)	(0.008)	(0.00879)	(0.00825)
R^2	0.107	0.221	0.021	0.018
Model 4	0.310***	0.313***	0.0743*	0.0732*
	(0.034)	(0.032)	(0.0400)	(0.0415)
R^2	0.131	0.244	0.020	0.018
Model 5	0.112***	0.119***	0.0302	0.0633
	(0.025)	(0.022)	(0.0300)	(0.0490)
R^2	0.093	0.201	0.018	0.022
Number of People	417	417	417	417
Number of Observations	1668	1668	1668	1580

Note. These are the phi coefficients estimated by regressing log relative probability of choosing a major on relative expected utility from earnings and major fixed effects. Model 1 is the ϕ estimate for utility specification with individual heterogeneity in both θ_i and β_i. Model 2 is the ϕ estimate for the the utility specification with individual heterogeneity in β_i, but with homogeneous $\theta = 1.7$. Model 3 is the ϕ estimate for the utility specification with homogeneous $\theta = 1.7$ and $\beta = 0.95$. Model 4 is the ϕ estimate for the utility specification with homogeneous median values of θ (0.94) and β (0.89). Model 5 is the ϕ estimate for the utility specification with risk neutrality, $\theta = 0$ and heterogeneity in β_i. In all cases, the omitted major is humanities/social sciences and major dummies are included for the remaining majors: economics/business, engineering, natural sciences, and no graduation. Column 1 uses only pre-treatment cross-sectional data, with no demographic controls. Column 2 includes controls for race, gender, and class year. Column 3 uses the differenced pre- and post- treatment data. Column 4 excludes observations for individuals who adjusted their expected earnings by more than 100,000 following the intervention. Boot-strapped standard errors in parentheses, adjusted for clustering at the individual level; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 10: Demographic Heterogeneity in Estimates of ϕ_i

	Mean	10th percentile	50th percentile	90th percentile
Full Sample	0.126***	-0.274***	0.003	0.670***
	(0.0310)	(0.0689)	(0.0033)	(0.1306)
Women	0.081***	-0.218***	0.001	0.433***
	(0.0298)	(0.0602)	(0.0016)	(0.1361)
Men	0.210***	-0.398***	0.029	1.142***
	(0.0718)	(0.0934)	(0.0300)	(0.2609)
Upperclassmen	0.084*	-0.218	0.001	0.540***
	(0.0445)	(0.1762)	(0.0037)	(0.1426)
Freshmen	0.164***	-0.276***	0.005	0.850***
	(0.0501)	(0.0578)	(0.0064)	(0.2234)

Note. Individual ϕ calculated by regressions of log relative probability on utility component interacted with an indicator for unique identifier, all results for the most general model (heterogeneous β and θ). Sample truncated at the 5th and 95th percentile of the ϕ_i estimates to remove outlier values. Boot-strapped robust standard errors in parentheses, adjusted for clustering at the individual level * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 11: Own Earnings Choice Elasticities: Average percent change in the likelihood of choosing a major due to a 10% increase in earnings in that major

	% Δ Probability Economics/Business	% Δ Probability Engineering	% Δ Probability Humanities/SS	% Δ Probability Natural Sciences	% Δ Probability Not Graduate
Model 1	0.68	0.96	0.88	0.98	4.09
	[0.05, 1.48]	[0.05, 1.91]	[0.05, 1.27]	[0.06, 2.13]	[0.33, 10.97]
Model 2	3.33***	4.20***	2.92***	3.89***	6.14***
	[0.038,4.75]	[0.50, 5.91]	[0.36, 4.15]	[0.50, 5.97]	[1.07, 13.13]
Model 3	3.30***	4.23***	3.15***	3.98***	6.12***
	[0.35, 4.52]	[0.47, 6.14]	[0.35, 4.542]	[0.46, 6.08]	[0.98, 13.20]
Model 4	4.22***	5.53***	3.64***	4.87***	5.67**
	[1.01,8.43]	[1.28, 11.10]	[0.75, 6.47]	[1.11, 9.51]	[1.27, 10.85]
Model 5	1.77***	2.34***	1.16	1.85***	1.32***
	[-1.22,6.75]	[-1.40, 7.61]	[-0.59, 3.08]	[-1.07, 6.04]	[-0.65, 3.47]

Note. Model 1 is the ϕ estimate for utility specification with individual heterogeneity in both θ_i and β_i. Model 2 is the ϕ estimate for the the utility specification with individual heterogeneity in β_i, but with homogeneous $\theta = 1.7$. Model 3 is the ϕ estimate for the utility specification with homogeneous $\theta = 1.7$ and $\beta = 0.95$. Model 4 is the ϕ estimate for the utility specification with homogeneous median values of θ (0.94) and β (0.89). Model 5 is the ϕ estimate for the utility specification with risk neutrality, $\theta = 0$ and heterogeneity in β_i. To convert the percent change in likelihood of choosing a major to an elasticity, divide by 10%, corresponding to the change in earnings. 90% bootstrapped percentile interval for elasticity in brackets; Significance level for two-sided t-test of null hypothesis H_0: Elasticity for Model K = Elasticity for Model 1: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 12: Counterfactual Results: Equalizing Earnings Across Majors

		Econ./Business	Engineering	Hum./S.S.	Nat. Science	Not Grad.
Panel A:	**Model 1**	0.00	0.00	0.00	0.00	-0.01
Counterfactual 1		[-0.12, 0.27]	[-0.14, 0.41]	[-0.14, 0.10]	[-0.18, 0.30]	[-2.51, 1.10]
Equalizing Risk	**Model 3**	0.04	0.02	-0.03	0.01	-0.97
		[-0.77, 1.49]	[-1.51, 1.80]	[-1.35, 0.89]	[-1.36, 1.65]	[-13.55, 2.07]
Panel B	**Model 1**	0.00	0.01	0.00	0.00	-0.03
Counterfactual 2		[-0.25, 0.64]	[-0.25, 0.81]	[-0.31, 0.50]	[-0.34, 0.75]	[-5.54, 1.15]
Equalizing Growth	**Model 3**	0.04	1.38	-0.00	0.09	-3.74
		[-4.59, 8.19]	[-4.27, 9.47]	[-5.99, 6.25]	[-6.27, 8.71]	[-27.63, 14.65]
Panel C	**Model 1**	-0.10	-0.15	0.12	-0.01	1.16
Counterfactual 3		[-3.46, 0.05]	[-2.77, 0.15]	[-0.12, 1.95]	[-1.84 ,0.71]	[0.05, 27.72]
Equalizing Levels	**Model 3**	-3.20	-5.82	3.36	-0.38	32.40
		[-17.87, 1.51]	[-19.62, 3.25]	[-1.41, 22.44]	[-12.66, 10.46]	[5.22, 112.85]
Panel D	**Model 1**	-0.09	-0.10	0.11	-0.01	0.98
Counterfactual 4		[-2.18, 0.04]	[-2.13, 0.32]	[-0.05, 1.81]	[-1.24, 0.79]	[0.09, 35.30]
Equalizing All 3	**Model 3**	-2.97	-4.11	4.07	-0.19	31.72
		[-12.35, 0.58]	[-15.13, 5.24]	[-0.46, 18.56]	[-8.05, 10.49]	[8.15, 97.10]
Mean Likelihood, Model		30.26	9.22	38.86	19.11	2.54

Note. This table depicts the effects of counterfactuals 1 through 4, reporting the median percent change in choosing each major, as well as the 10th and 90th percentile of changes in brackets below. Model 1 is the utility specification with individual heterogeneity in both θ_i and β_i. Model 3 is the utility specification with homogeneous $\theta = 1.7$ and $\beta = 0.95$. 'Mean Likelihood, Model' is the likelihood of choosing each major in the absence of the counterfactual.

Table 13: The Gender Gap in Likelihood of Choosing Major, Equalization Exercise

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Baseline	**Add σ**	**Add μ_1, μ_2**	**Add μ_0**	**Add θ**	**Add β**	**Add α**
Econ/Business	0.00	-0.00	0.01	-0.00	-0.01	0.05	9.35
Engineering	0.00	-0.00	-0.00	-0.01	-0.01	0.01	3.64
Hum./S.S.	-0.00	0.00	-0.00	0.01	0.01	-0.05	-15.57
Nat. Science	0.00	-0.00	-0.01	-0.01	-0.04	-0.09	3.13
Not Grad.	0.00	-0.00	-0.00	0.01	0.04	0.07	-0.55

Note. This table shows the gender gap in the likelihood of choosing each major, where a positive value of 1 indicates women are 1 percentage point more likely to choose the major than men. The exercise uses model 1, in which each individual's ϕ and α parameters were estimated assuming heterogeneous values of β_i and θ_i.

Appendices

A   Survey Questions

This section describes the data sources used for the information treatments, the survey instrument, and the hypothetical games.

A.1   Information Treatments

Table A-1 describes the information treatments that students received between the initial survey of beliefs and the final survey of beliefs.

Sources:

1. CPS: The Current Population Survey (CPS) is a monthly survey of about 50,000 households conducted by the Bureau of the Census for the Bureau of Labor Statistics. The survey has been conducted for more than 50 years. The CPS is the primary source of information on the labor force characteristics of the U.S. population. The sample is scientifically selected to represent the civilian non-institutional population.

2. NSCG: The 2003 National Survey of College Graduates (NSCG) is a longitudinal survey, designed to provide data on the number and characteristics of individuals. The Bureau of the Census conducted the NSCG for the NSF (National Science Foundation). The target population of the 2003 survey consisted of all individuals who received a bachelor's degree or higher prior to April 1, 2000.

Methodology:

1. CPS: Our CPS sample is taken from the March 2009 survey. Full time status is defined as "usually" working at least 35 hours in the previous year, working at least 45 weeks in the previous year, and earning at least $10,000 in the previous year. Average employment rates, average earnings, and percent with greater than $35,000 or $85,000 earnings is calculated using a sample of 2,739 30 year old respondents.

2. NSCG: We calculate inflation adjusted earnings using the Consumer Price Index. The salary figures we report are therefore equivalent to CPS figures in 2009 March real dollars. Full time status is defined as in the CPS sample. Given the need to make precise calculations for

each field of study group, we use the combined sample of 30-35 year old respondents and age adjust the reported statistics for 30 year olds. This sample consists of 14,116 individuals. To calculate average earnings, we use an earnings regression allowing for separate age intercepts, one each for 6 ages 30-35. The predicted value of earnings from the regression is used as the estimate of average earnings for 30 year olds. For the percent full time employed, and percent with earnings greater than $35,000 and $85,000, we use a logit model to predict these percentages for 30 year olds and include a separate coefficient for each of the 6 ages 30-35.

A.2   Survey Instrument

Because we wanted to approximate life cycle utility from each major, we collected beliefs about both initial earnings (i.e., just after college graduation), and for later periods, when earnings might be believed to be much higher. We collected post-graduation beliefs for three periods:

1. First year after college graduation (when most respondents would be aged 22-24)

2. Age 30

3. Age 45

At each of those periods, we ask respondents for their beliefs about their own earnings (including measures of dispersion), work status (not working, part time, full time), probability of marriage, and spouse's earnings.

An example question on expected earnings at age 30: "*If you received a Bachelor's degree in each of the following major categories and you were working FULL TIME when you are 30 years old what do you believe is the average amount that you would earn per year?*"[19] The instructions emphasized to the respondents that their answers should reflect their own beliefs, and not use any outside information.[20]

---

[19]We also provided definitions of working full time ("working at least 35 hours per week and 45 weeks per year"). Individuals were instructed to consider in their response the possibility they might receive an advanced/graduate degree by age 30. Therefore, the beliefs about earnings we collected incorporated beliefs about the possibility of other degrees earned in the future and how these degrees would affect earnings. We also instructed respondents to ignore the effects of price inflation.

[20]We included these instructions: "*This survey asks YOUR BELIEFS about the earnings among different groups. Although you may not know the answer to a question with certainty, please answer each question as best you can. Please do not consult any outside references (internet or otherwise) or discuss these questions with any other people. This study is about YOUR BELIEFS, not the accuracy of information on the internet.*"

Our questions on earnings were intended to elicit beliefs about the distribution of future earnings. We asked three questions on earnings: beliefs about expected (average) earnings, beliefs about the percent chance earnings would exceed \$35,000, and percent change earnings would exceed \$85,000. As detailed below, we use this information to estimate individual-specific distribution of earnings beliefs.

Beliefs about labor supply at age 30 were elicited as follows: "*What do you believe is the percent chance of the following: (1) You are working full-time, (2) You are working part-time, (3) You are not working at all, when you are 30 years old if you received a Bachelor's degree in each of the following? Note: Your answers should sum to 100 for each major category.*"

The full survey questionnaire is available from the authors upon request.

A.3 EG Game: Survey Question

The following text is the text provided in the choice problem which elicits risk preferences.

Consider the hypothetical situation below (Table A-2) where you are presented with 6 different games of chance. The payoff in each game depends on the toss of a coin. For example, Game 2 below pays \$24 if the coin lands on Heads and \$36 if it lands on Tails. Since half of the times a coin will land on Heads and half of the time if will land on Tails, there is a 50% chance of receiving \$24 under game 2, and a 50% chance of receiving \$36. Similarly, in game 6, there is a 50% chance of receiving \$2 and a 50% chance of receiving \$70. Which game would you like to participate in? (Choose one)

A.4 CW Game: Survey Question

The following text is the text provided in the choice problem which elicits patience parameters.

Consider the hypothetical situation below (Table A-3) where you are chosen to receive \$3000, and are given the choice of two payment options: Option A and Option B . If you choose Option B, you will receive a sum of money 7 months from today. If you choose Option A, you will receive a sum of money 1 month from today, but this Option (A) will pay a smaller amount than Option B. For each of these alternatives below, please choose your preferred option.

A.5 Risk and Discounting Parameters

We use student's responses to the previously described lottery games to identify the coefficient of risk aversion, θ, and the discount factor, β. Our approach is quite similar to that of Andersen

et al. (2008), who use both time and risk price lists to jointly estimate discounting and curvature parameters.

The six choices in EG Game define regions of θ consistent with choosing a given lottery over all other lotteroies. The ex ante expected utility over a lottery is:

$$U(c_h, c_t) = \frac{1}{2}\left(\frac{c_h^{1-\theta} - 1}{1-\theta}\right) + \frac{1}{2}\left(\frac{c_t^{1-\theta} - 1}{1-\theta}\right)$$

We assume a Constant Relative Risk Aversion (CRRA) form for the utility function where θ is heterogeneous among the agents. To back out θ, we look for θ such that the ex- ante expected utility is equal from the two lotteries.

Once we have these critical values of θ for the games with the closest expected values, we can back out the intervals of θ for the choices. The choices in the game correspond to the following intervals of θ:

$$\theta = \begin{cases} (3.47, \infty) & \text{if Game 1,} \\ (1.16, 3.47) & \text{if Game 2,} \\ (0.71, 1.16) & \text{if Game 3,} \\ (0.50, 0.71) & \text{if Game 4,} \\ (0, 0.50) & \text{if Game 5,} \\ (-\infty, 0) & \text{if Game 6.} \end{cases}$$

For choice 2 through 5, we assign students the θ value that corresponds to the mid-point of the value range. For a choice of game 1, we assign $\theta = 3.47$ and for the game 6, we assign $\theta = 0$.

We then use these θ values and students' choices in CW Game to identify their discount factor. In the survey, students were given the choice of receiving Option A in one month or Option B in seven months, were told the annual interest rate, and asked to choose which option they preferred in ten different games. In all cases, the present discounted value of Option B is the same as Option A + r × Option A, where r is the effective interest rate over a six month time period, with interest compounded quarterly.

In this scenario, a person chooses option A if:

$$U(w + M_t) + \beta^\tau U(w) > U(w) + \beta^\tau U(w + M_{t+\tau})$$

w indicates one's consumption in a period in addition to the prize; M_t is the money received at time t; $M_{t+\tau}$ is the money if received at time $t+\tau$; β is one's discount factor. If we assume a CRRA

57

functional form for utility with coefficient of risk aversion given by θ, we can rewrite the above as:

$$\frac{(w + M_t)^{1-\theta}}{1-\theta} + \beta^\tau \frac{(w)^{1-\theta}}{1-\theta} \quad > \quad \frac{(w)^{1-\theta}}{1-\theta} + \beta^\tau \frac{(w + M_{t+\tau})^{1-\theta}}{1-\theta}$$

To solve for β, we must make some assumptions about what consumption would be in the absence of the prize.[21] We assume w $= 8400$, a value approximately equivalent to the earnings associated with a 20 hour per week job throughout the school year for a NYU student at the time of the experiment. Then, if we plug in $M_t = A$, $M_{t+\tau} = B$, $\tau = \frac{1}{2}$, a person is indifferent between the two options if :

$$(A + 8400)^{1-\theta} + \beta^{\frac{1}{2}}(8400)^{1-\theta} = (8400)^{1-\theta} + \beta^{\frac{1}{2}}(B + 8400)^{1-\theta}$$
$$\beta = \left[\frac{(A + 8400)^{(1-\theta)} - 8400^{(1-\theta)}}{(8400 + B)^{(1-\theta)} - 8400^{(1-\theta)}}\right]^2$$

Table A-4 shows the discount factors associated with indifference in each of the ten choices in game 1 and each of six possible θ implied by the choices in the EG Game.

For a rational student, there should be a single point where they switch from choosing A to choosing B and then all games following should choose B. We assign students who behave rationally the β value that corresponds to the midpoint of the β for the game they last pick option A and the β for the game in which they first pick option B, given θ.

B   Earnings Estimation Details

B.1   Parameter Estimation

In this section, we outline the algorithm for retrieving the moments of the earnings distribution specific to each individual, major and treatment status.

For each student i, major k, and experimental treatment status $j = 1, 2$ at age t, we estimate a log normal distribution with the following structure:

$$\log w_{ijkt} \sim N(\mu_{ijkt}, \sigma_{ijkt}),$$
$$\text{where } \mu_{ijkt} = \mu_{ijk0} + \mu_{ijk1}(t - 22) + \mu_{ijk2}(t - 22)^2 \tag{B-1}$$
$$\sigma_{ijkt} = \sigma_{ijk0} + \sigma_{ijk1}(t - 22)$$

---

[21]In the previous game, we assumed that consumption in a 'period' was equal to zero net of the prize because the dollar values of the prize were small enough that one could consume that prize in a single day. With the larger prizes in this lottery, it is more reasonable to conceive of a period as a year, rather than a single day.

Our estimation algorithm proceeds in two steps. In Step 1, we assume combinations of values of σ_{ijk0} and σ_{ijk1} over a grid of values from 0 to 4. We then calculate the closed form solutions for μ_{ijk0}, μ_{ijk1} and μ_{ijk2}. Using the distribution generated from the 5 parameters, we calculate the average earnings at age 22, 30 and 45 as well as the probability of earning more than \$85000 at age 30 and 45. These are moments that we observe in the data. In Step 2, we calculate an objective function that is the difference between the above moments in the data and those generated by the guesses of σ_{ijk0} and σ_{ijk1}. Our estimator searches for the parameters that minimizes this objective function using a grid search method.

In this section, we derive the closed form solutions for the μ_{ijk0}, μ_{ijk1}, μ_{ijk2} conditional on σ_{ijk0} and σ_{ijk1}. Under the assumption of a log normal distribution, we can write the log of expected earnings at ages 22, 30 and 45 in terms of the 5 unknown parameters μ_{ijk0}, μ_{ijk1}, μ_{ijk2}, σ_{ijk0} and σ_{ijk1}.

$$\log(\overline{w}_{ijk22}) = \mu_{ijk0} + 0.5\sigma_{ijk0}^2 \tag{B-2}$$

$$\log(\overline{w}_{ijk30}) = \mu_{ijk0} + \mu_{ijk1}(30 - 22) + \mu_{ijk2}(30 - 22)^2 + 0.5\sigma_{ijk30}^2 \tag{B-3}$$

$$\log(\overline{w}_{ijk45}) = \mu_{ijk0} + \mu_{ijk1}(45 - 22) + \mu_{ijk2}(45 - 22)^2 + 0.5\sigma_{ijk45}^2 \tag{B-4}$$

From equation B2 for a given σ_{ijk0}, we can back out μ_{ijk0} siuch that

$$\mu_{ijk0} = log(\overline{w}_{ijk22}) - 0.5\sigma_{ijk0}^2 \tag{B-5}$$

We then take the difference between equations B3 and B2 (referred to as D30 below) and rearrange to obtain the following expression:

$$D30 = \log(\overline{w}_{ijk30}) - \log(\overline{w}_{ijk22}) \;=\; \mu_{ijk1}(30 - 22) + \mu_{ijk2}(30 - 22)^2 + (30 - 22)\sigma_{ijk0}\sigma_{ijk1} \tag{B-6}$$

$$\frac{D30}{(30 - 22)} \;=\; \mu_{ijk1} + \mu_{ijk2}(30 - 22) + \sigma_{ijk0}\sigma_{ijk1} \tag{B-7}$$

Similarly, we can take the difference between equations B4 and B2 (referred to as D45 below) and obtain an analogous expression:

$$D45 = \log(\overline{w}_{ijk45}) - \log(\overline{w}_{ijk22}) = \mu_{ijk1}(45 - 22) + \mu_{ijk2}(45 - 22)^2 + 0.5(\sigma_{ijk45}^2 - \sigma_{ijk22}^2) \tag{B-8}$$

Using B7 and B6, we obtain an expression for μ_{ijk2}

$$\frac{D45}{(45 - 22)} - \frac{D30}{(30 - 22)} \;=\; \mu_{ijk2}(45 - 30) + 0.5 \times (30 - 22)\sigma_{ijk1}^2 \tag{B-9}$$

$$\mu_{ijk2} \;=\; \frac{\frac{D45}{(45-22)} - \frac{D30}{(30-22)}}{(45 - 30)} - 0.5\sigma_{ijk1}^2 \tag{B-10}$$

Finally we can write μ_{ijk1} in terms of μ_{ijk2}, σ_{ijk1} and σ_{ijk0}:

$$\mu_{ijk1} = \frac{D30}{(30-22)} - \mu_{ijk2} \times (30-22) - \sigma_{ijk1}\sigma_{ijk0} \tag{B-11}$$

We then use a grid search estimator over σ_{ijk0} and σ_{ijk1} and solve for the analytical solutions of μ_{ijk0}, μ_{ijk1} and μ_{ijk2} in terms of σ_{ijk0} and σ_{ijk1} (as derived above). We specify a grid of 200 points between the values of 0 and 4 for σ_{ijk0} and the values of -0.5 and 0.2 for σ_{ijk1} and choose the sigma vector that minimizes an objective function. The objective function that we choose is the square of the distance between the data moments and the predicted moments. For our estimation we use moments about the average earnings at ages 22, 30 and 45. We also use the probability of earning more than \$85,000 at age 30 and 45.

Note that there are a small number of individuals, who have values for mean and probability of earning more than 35000 that cannot be rationalized by a log earnings distribution for reasonable values of σ_{ijk0} and σ_{ijk1}. For these observations, we replace them by the mean of σ_{ijk0} andσ_{ijk1} in our sample.

We then simulate draws from the estimated distribution to plot the median age earnings profile and compare our estimates to the true distribution in the ACS. Note that like our data, we recode simulation draws above \$500,000 as \$500,000 and earnings below \$10,000 at \$10,000. We also set earnings past age 45 to be constant, since we do not have data moments on expectations past age 45 to identify the curvature in those years. Since the earnings profiles are fairly flat post age 45 in the ACS, we believe these profiles reflect true earnings profiles observed in the data. In Figure A-1, we plot these profiles before and after the information treatment for each major. Although the curvature is similar, there is less dispersion in the beliefs at every age after the students have been exposed to the treatment. We also report the fit for our estimates relative to the true age- earnings profiles. Figure A-2 shows fit for our mean age-earnings profiles by major relative to the mean age earnings profile in the ACS. For each major, the wage function is similarly concave reflecting the standard hump shaped age-earnings profiles.

C   Robustness checks

In our discussion of parameter identification (Section 5), we proposed two different estimation specifications for ϕ. This section reports the results of our analyses for the more restrictive approach of estimating ϕ.

C.1   Elasticities

In this section, we present estimates of the major-choice elasticities under the alternative estimation approach.

Table A-8 reports how elastic major choice is to a 10 percent increase in earnings for the specification in which the ϕ and α_{ik} are estimated using a linear regression of earnings expectations on major choice likelihoods in the cross-sectional data (column 1 of Table 9).

When compared to estimates in Table 11, the magnitude of the elasticities is much larger for Models 2-5 (it is similar for Model 1 due to the similar estimates of ϕ across the estimation methods). Given that Method 1 generally yields higher estimates of ϕ than the second method (Table 9), this is not surprising.

Across estimation specifications, however, the main takeaway of this exercise is the same: model specification matters when considering how sensitive major choice is to earnings. In all three estimation methods, the most general model of utility has the smallest elasticities.

C.2   Counterfactuals

In this section, we conduct robustness checks for the counterfactuals. First, in Table A-9, we report the results of all four counterfactuals for Model 2 ($\theta = 1.7$, heterogeneous β_i), Model 4 (median values of θ and β) and Model 5 (risk neutrality, $\theta = 0$), using Method 2 from Table 9. We also report the results for counterfactual 4 using the more restrictive method of estimating ϕ (Method 1 from Table 9); this is reported in Table A-10; we do not report the other three counterfactual results under this specification due to space, but the qualitative results are similar.

As in the elasticity robustness checks, the counterfactual results under the less general estimates of ϕ are larger, but in the same direction as our main results. Even in the least conservative model (Model 4), the effect of this counterfactual is small in real terms: it decreases the likelihood of majoring in economics by 23.64 percent, which is equivalent to 7.2 percentage points. This suggests that the results we report in our main paper are more conservative estimates of the true effect of

earnings on major choice, but that the general pattern we find of small average impacts of earnings on major choice holds across estimation methods.
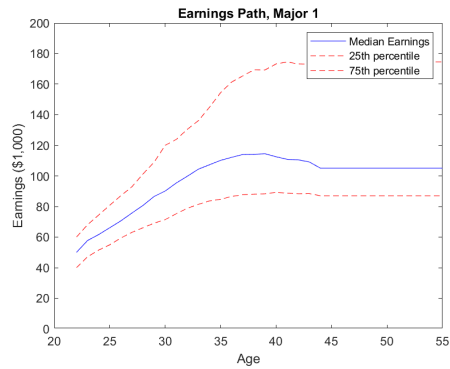
Appendix Figures and Tables

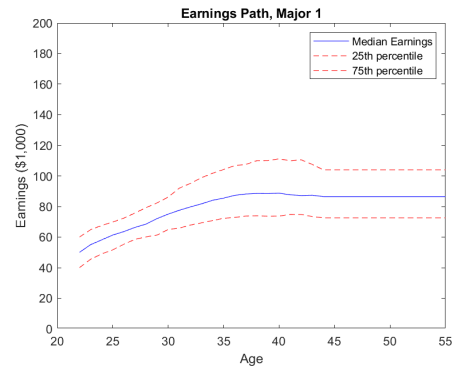Figure A-1: Median Earnings Pre and Post Treatment

Economics / Business

Engineering
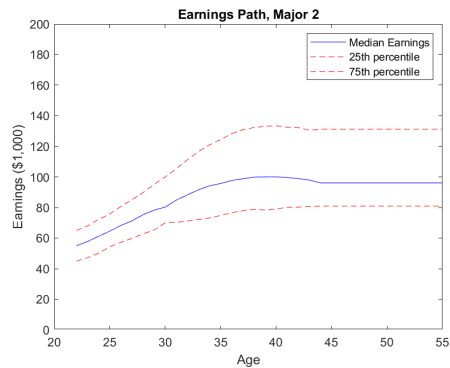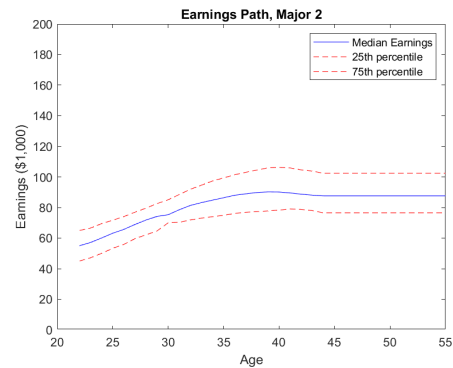
Humanities
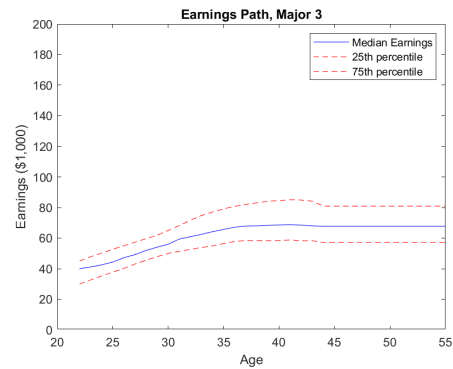
Natural Sciences
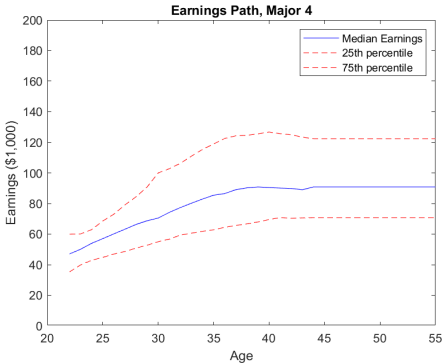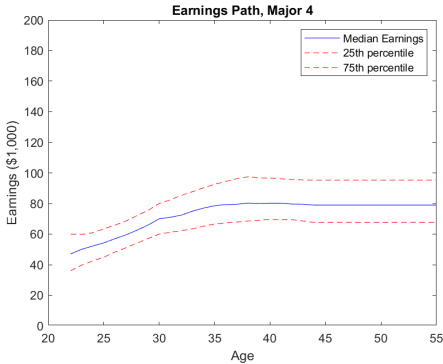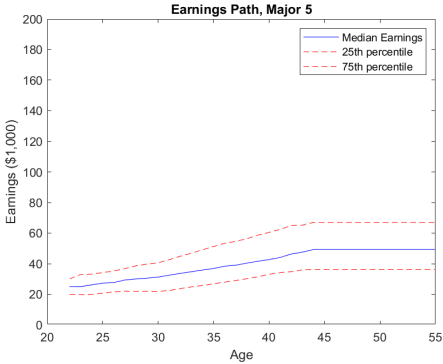
Pre-Treatment Beliefs

Post-Treatment Beliefs

No Graduation

Pre-Treatment Beliefs

Post-Treatment Beliefs
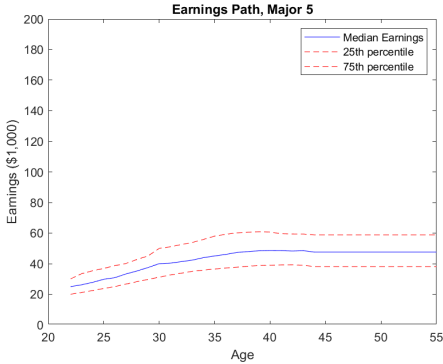
Figure A-2: Mean Earnings (Pre-Treatment) by Age Compared to Mean Earnings in ACS

Economics/ Business

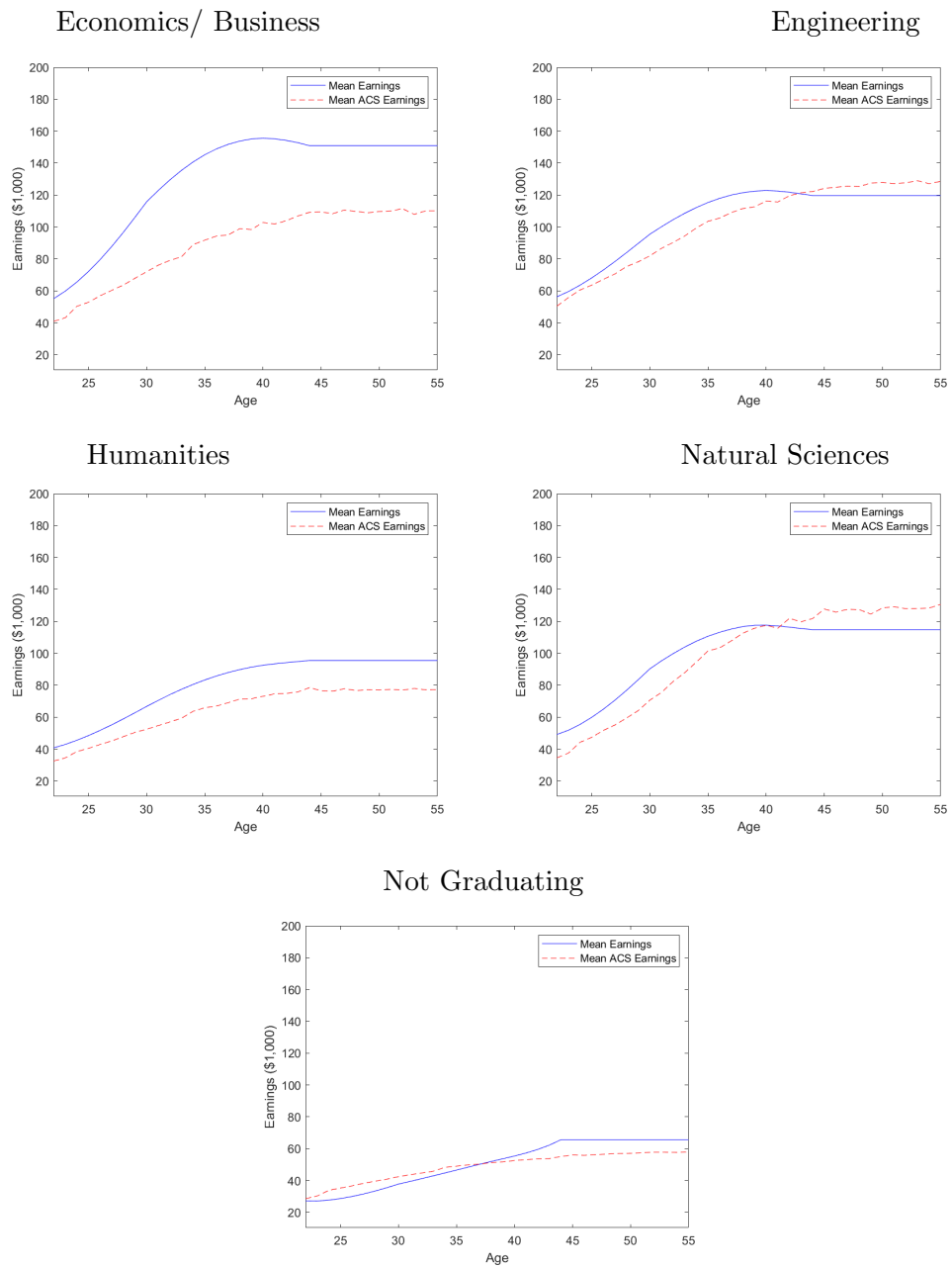Engineering

Humanities

Natural Sciences

Not Graduating

65

Table A-1: Information revealed in the study

Info Treatment 1: Information about age 30 individuals (All Individuals Treatment)

The following information is from the US Census Bureau.

Among all individuals (including college and non-college graduates) aged 30:

The percentage that are working full time is	59.80%
The percentage of those that are working full time who are women is	42.70%
The average annual earnings of those that are working full time is	$45,726
The percentage of those that are working full time that earn more than $35,000 per year is	59.00%
The percentage of those that are working full time that earn more than $85,000 per year is	7.30%

Info Treatment 2: Information about age 30 college-graduate individuals (College Treatment)

The following information is from the US Census Bureau.

Among all college graduates currently aged 30:

The percentage that are working full time is	69.80%
The percentage of those that are working full time who are women is	52.80%
The average annual earnings of those that are working full time is	$60,376
The percentage of those that are working full time that earn more than $35,000 per year is	80.70%
The percentage of those that are working full time that earn more than $85,000 per year is	14.80%

Info Treatment 3: Info about age 30 female college-graduate indivs, by major (Female Major Specific Treatment)

The following information is from the US Census Bureau.

Among all *female* college graduates aged 30 who received a Bachelor's degree in major (M):

	Econ	Eng	Hum	Nat	No Grad
The percentage that are working full time is	60.6%	72.8%	52.3%	55.3%	51.6%
The average annual earnings of those that are working full time is	$60,730	$75,086	$49,154	$60,021	$34,603
The percentage of those that are working full time that earn more than $35,000 per year is	85.5%	99.0%	72.2%	84.0%	44.9%
The percentage of those that are working full time that earn more than $85,000 per year is	27.5%	26.9%	8.0%	8.5%	1.6%

Info Treatment 4: Info about age 30 male college-graduate individuals, by major (Male Major Specific Treatment)

The following information is from the US Census Bureau.

Among all *male* college graduates aged 30 who received a Bachelor's degree in major (M):

	Econ	Eng	Hum	Nat	No Grad
The percentage that are working full time is	93.5%	91.6%	77.6%	81.9%	72.1%
The average annual earnings of those that are working full time is	$74,542	$82,377	$52,937	$72,583	$47,803
The percentage of those that are working full time that earn more than $35,000 per year is	92.4%	95.2%	78.8%	90.6%	65.2%
The percentage of those that are working full time that earn more than $85,000 per year is	31.5%	33.6%	8.7%	24.2%	5.7%

Also Revealed to All Respondents in Final Stage

	Econ	Eng	Hum	Nat	No Grad
Among all college graduates aged 30 who received a Bachelor's degree in major (M):					
The percentage of those who are women is	34.70%	18.20%	55.20%	48.00%	42.30%

Table A-2: Schedule of Payments, EG Game

Choice	Heads	Tails
Game 1	$28	$28
Game 2	$24	$36
Game 3	$20	$44
Game 4	$16	$52
Game 5	$12	$60
Game 6	$2	$70

Note. Payoff matrix shown to participants in the game which elicits risk preferences. This method of elicitation was derived from Eckel and Grossman (2002).

Table A-3: Schedule of Payments, CW Game

Game	Option A (1 mo.)	Option B (7 mo.)	Annual Interest Rate
1	3000	3075	5
2	3000	3152	10
3	3000	3229	15
4	3000	3308	20
5	3000	3387	25
6	3000	3467	30
7	3000	3548	35
8	3000	3630	40
9	3000	3713	45
10	3000	3797	50

Note. Payoff matrix shown to participants in the game which elicits time preferences. This method of elicitation was derived from Coller and Williams (1999).

Table A-4: Discount Rates Associated with Matrix of Game Choices

CW Game Payoffs	EG Gamble 1 $\theta=3.47$	EG Gamble 2 $\theta=2.31$	EG Gamble 3 $\theta=0.94$	EG Gamble 4 $\theta=0.61$	EG Gamble 5 $\theta=0.25$	EG Gamble 6 $\theta=0$
A=3000 B=3075	0.97	0.97	0.96	0.96	0.95	0.95
A=3000 B=3152	0.95	0.93	0.92	0.91	0.91	0.91
A=3000 B=3229	0.92	0.90	0.88	0.87	0.87	0.86
A=3000 B=3308	0.90	0.87	0.84	0.84	0.83	0.82
A=3000 B=3387	0.87	0.85	0.81	0.80	0.79	0.78
A=3000 B=3467	0.85	0.82	0.78	0.77	0.76	0.75
A=3000 B=3548	0.83	0.80	0.75	0.74	0.72	0.71
A=3000 B=3630	0.81	0.77	0.72	0.71	0.69	0.68
A=3000 B=3713	0.69	0.69	0.68	0.68	0.68	0.68
A=3000 B=3797	0.77	0.73	0.67	0.65	0.64	0.62
Observations	104.00	93.00	112.00	38.00	28.00	42.00

Note. This table reports the β values that would make a person indifferent between receiving payment A in one month versus payment B in six months in the CW game, conditional on their implied θ from the EG game. For more information on how these are derived, see Appendix section A.5

Table A-5: Percent of Sample with Top and Bottom Coding

			Economics	Engineering	Humanities	Natural Science	No Graduation
	Age 22	Top code	0.48	0.48	0.48	0.48	0.24
		Bottom Code	1.92	1.43	1.67	1.92	6.4
Pre-Treatment	Age 30	Top Code	3.11	0.95	0.71	0.95	0.719
		Bottom Code	1.43	1.20	1.43	1.43	5.99
	Age 45	Top Code	5.27	2.39	2.39	2.64	2.39
		Bottom Code	0.96	1.20	1.67	1.20	4.07
Post-Treatment	Age 30	Top Code	2.87	1.67	1.20	1.67	1.20
		Bottom Code	1.43	1.67	1.20	1.20	2.39
	Age 45	Top Code	1.92	1.20	0.96	1.43	1.67
		Bottom Code	1.67	1.20	1.43	1.20	3.35

Note. This table reports the percent of the sample whose expected earnings are top-coded to $500,000 or bottom-coded to $10,000. N= 417.

Table A-6: Wage Parameters, By Treatment, Major, and Gender

		Economics/Business		Engineering		Humanities		Natural Science		Not Graduate	
		Men	Women	Men	Women	Men	Women	Men	Women	Men	Women
Pre-Treatment	μ_0	-0.778	-0.932	-0.736	-0.931	-1.065	-1.303	-0.963	-1.122	-1.886	-2.183
	μ_1	0.107	0.103	0.0647	0.0890	0.0647	0.0904	0.0725	0.0965	0.0484	0.0521
	μ_2	-0.00300	-0.00287	-0.00172	-0.00251	-0.00141	-0.00229	-0.00172	-0.00265	-0.000424	-0.000453
	σ_0	0.387	0.472	0.340	0.496	0.368	0.507	0.424	0.521	0.747	0.896
	σ_1	0.00196	0.00158	0.00417	0.000288	0.00341	-0.000271	0.00102	-0.000101	-0.00273	-0.00594
Post-Treatment	μ_0	-0.842	-0.944	-0.815	-0.903	-1.142	-1.256	-0.972	-1.088	-1.709	-2.154
	μ_1	0.0777	0.0699	0.0611	0.0655	0.0787	0.0761	0.0728	0.0724	0.0877	0.107
	μ_2	-0.00185	-0.00192	-0.00143	-0.00179	-0.00210	-0.00214	-0.00189	-0.00202	-0.00223	-0.00298
	σ_0	0.482	0.499	0.449	0.492	0.437	0.449	0.453	0.496	0.610	0.747
	σ_1	-0.000274	-0.0000750	0.000381	-0.00153	0.000535	0.00256	-0.0000707	-0.00106	-0.00205	-0.00428

Note. Values indicate the average value for each gender of the parameters governing the individual wage distributions pre- and post- the information treatment. We assume that the earnings process for individual i of age t in major k during treatment period j is distributed as $\log w_{ijkt} \sim N(\mu_{ijkt}, \sigma_{ijkt})$. The parameters μ_0, μ_1 and μ_2 are parameters for evolution of mean of that distribution such that $\mu_{ijkt} = \mu_{ijk0} + \mu_{ijk1}(t-22) + \mu_{ijk2}(t-22)^2$ whereas σ_0 and σ_1 are parameters for the evolution of the dispersion of wages over the lifecycle such that $\sigma_{ijkt} = \sigma_{ijk0} + \sigma_{ijk1}(t-22)$. Earnings are estimated at the scale of \$100,000 for computational tractibility; see Appendix Section B for estimation methods.

Table A-7: Model Fit: Post- Treatment Likelihood of Major

		Economics/Business	Engineering	Humanities/SS	Natural Sciences	Not Graduate
	Data	0.302	0.095	0.388	0.191	0.025
	Model 1	0.286	0.068	0.428	0.189	0.028
Panel A:	Model 2	0.284	0.068	0.431	0.189	0.028
Full Sample	Model 3	0.283	0.068	0.432	0.189	0.028
	Model 4	0.282	0.068	0.433	0.190	0.027
	Model 5	0.283	0.069	0.433	0.190	0.026
	Data	0.307	0.093	0.390	0.185	0.026
	Model 1	0.301	0.068	0.414	0.185	0.031
Panel B:	Model 2	0.298	0.068	0.417	0.185	0.032
Freshmen	Model 3	0.298	0.068	0.417	0.185	0.032
	Model 4	0.297	0.069	0.417	0.185	0.031
	Model 5	0.300	0.069	0.417	0.185	0.030

Note. This table compares the untargeted moment, likelihood of choosing each major post- information treatment (data) and the predicted values for each model. Model 1 is the ϕ estimate for utility specification with individual heterogeneity in both θ_i and β_i. Model 2 is the ϕ estimate for the the utility specification with individual heterogeneity in β_i, but with homogeneous $\theta = 1.7$. Model 3 is the ϕ estimate for the utility specification with homogeneous $\theta = 1.7$ and $\beta = 0.95$. Model 4 is the ϕ estimate for the utility specification with homogeneous median values of θ (0.94) and β (0.89). Model 5 is the ϕ estimate for the utility specification with risk neutrality, $\theta = 0$ and heterogeneity in β_i.

Table A-8: Own Earnings Choice Elasticities: ϕ estimation method 1

	% Δ Probability Economics/Business	% Δ Probability Engineering	% Δ Probability Humanities/SS	% Δ Probability Natural Sciences	% Δ Probability Not Graduate
Model 1	0.59	0.83	0.76	0.84	3.49
Model 2	6.07	7.67	5.31	7.10	11.34
Model 3	12.45	16.07	11.87	15.11	23.90
Model 4	18.84	25.06	16.16	21.94	25.81
Model 5	6.84	9.11	4.49	7.17	5.09

Note. This table shows elasticities when ϕ is estimated using only cross-sectional data and no demographic controls, not the panel data created in the experiment. Model 1 is the ϕ estimate for utility specification with individual heterogeneity in both θ_i and β_i. Model 2 is the ϕ estimate for the the utility specification with individual heterogeneity in β_i, but with homogeneous $\theta = 1.7$. Model 3 is the ϕ estimate for the utility specification with homogeneous $\theta = 1.7$ and $\beta = 0.95$. Model 4 is the ϕ estimate for the utility specification with homogeneous median values of θ (0.94) and β (0.89). Model 5 is the ϕ estimate for the utility specification with risk neutrality, $\theta = 0$ and heterogeneity in β_i.

Table A-9: Counterfactual Results: Equalizing Earnings Across Majors

		Econ./Business	Engineering	Hum./S.S.	Nat. Science	Not Grad.
Panel A:	**Model 2**	0.04	0.01	-0.03	0.01	-0.72
Counterfactual 1		[-0.55 1.89]	[-1.26, 1.75]	[-1.14, 0.84]	[-0.94, 1.68]	[-11.22, 1.76]
Equalizing Risk	**Model 4**	2.66	4.15	-3.68	0.05	-23.78
		[-1.28, 13.75]	[-5.14, 19.32]	[-18.11, 0.68]	[-10.01, 10.38]	[-47.39, -4.74]
	Model 5	-0.00	0.00	0.01	-0.00	-0.22
		[-1.51, 0.61]	[-1.72, 1.59]	[-1.72, 1.31]	[-0.74,0.87]	[-4.29, 1.53]
Panel B	**Model 2**	0.04	0.34	-0.00	0.04	-1.01
Counterfactual 2		[-3.16, 8.17]	[-2.74, 9.31]	[-4.34, 3.91]	[-4.80, 7.88]	[-28.53, 10.61]
Equalizing Growth	**Model 4**	0.04	1.55	-0.02	0.08	-2.59
		[-6.87, 11.90]	[-5.62, 16.54]	[-8.38, 9.19]	[-7.45, 12.77]	[-19.66, 17.84]
	Model 5	0.01	0.20	-0.00	0.02	-0.17
		[-2.40, 6.02]	[-2.29, 8.98]	[-4.12, 3.45]	[-2.95,6.69]	[-8.37, 4.65]
Panel C	**Model 2**	-2.13	-3.33	2.28	-0.21	23.32
Counterfactual 3		[-20.97, 0.76]	[-22.81, 2.22]	[-1.25, 25.05]	[-14.74, 8.52]	[2.62, 167.95]
Equalizing Levels	**Model 4**	-3.20	-5.82	3.36	-0.38	32.40
		[-17.87, 1.51]	[-19.62, 3.25]	[-1.41, 22.44]	[-12.66, 10.46]	[5.22, 112.85]
	Model 5	-0.96	-1.36	0.93	-0.04	8.53
		[-9.19, 0.61]	[-10.77, 1.97]	[-0.61, 12.57]	[-6.32, 5.00]	[0.64, 66.86]
Panel D	**Model 2**	-1.90	-2.83	2.25	-0.26	23.11
Counterfactual 4		[-13.10, 0.61]	[-16.61, 3.21]	[-0.75, 18.40]	[-9.33, 8.07]	[3.28, 146.63]
Equalizing All 3	**Model 4**	-4.92	-6.43	6.34	-0.13	41.97
		[-18.05, 0.39]	[-20.27, 8.87]	[-0.18, 33.01]	[-11.27, 17.92]	[13.04, 96.24]
	Model 5	-0.94	-1.09	1.32	-0.02	6.76
		[-7.61, 0.16]	[-8.47, 2.87]	[-0.01, 12.76]	[-4.49, 6.33]	[1.08, 25.72]
Mean Likelihood, Model		30.26	9.22	38.86	19.11	2.54

Note. This table depicts the effects of counterfactuals 1 through 4 for Model 2, Model 4, and Model 5 using the ϕ calculated with estimation method 2, reporting the median percent change in choosing each major, as well as the 10th and 90th percentile of changes in brackets below. Model 2 is the utility specification with individual heterogeneity in β_i, but with homogeneous $\theta = 1.7$. Model 4 is the ϕ estimate for the utility specification with homogeneous median values of θ (0.94) and β (0.89). Model 5 is the utility specification with risk neutrality, $\theta = 0$ and heterogeneity in β_i. 'Mean Likelihood, Model' is the likelihood of choosing each major in the absence of the counterfactual.

Table A-10: Counterfactual 4: ϕ Estimation Method 1

	Econ./Business	Engineering	Hum./S.S.	Nat. Science	Not Grad.
Model 1	-0.08	-0.09	0.10	-0.00	0.85
	[-1.91, 0.04]	[-1.81, 0.28]	[-0.04, 1.58]	[-0.97, 0.69]	[0.08, 30.18]
Model 2	-3.62	-5.09	3.89	-0.60	44.83
	[-23.92, 0.90]	[-28.04, 5.64]	[-2.05, 35.09]	[-16.16, 14.74]	[5.87, 401.01]
Model 3	-11.81	-15.68	13.14	-2.16	163.14
	[-41.87, 0.98]	[-47.53, 18.72]	[-3.93, 81.23]	[-30.58, 42.80]	[31.69,943.85]
Model 4	23.64	-28.31	21.42	-4.77	305.22
	[-60.47, 0.16]	[-64.54, 36.09]	[-4.42, 209.89]	[-44.90, 79.66]	[64.60, 1471.76]
Model 5	-3.79	-4.10	4.60	-0.09	26.88
	[-26.55, 0.44]	[-28.30, 10.60]	[-0.06, 54.29]	[-16.54, 24.18]	[4.05, 133.37]
Mean Likelihood, Model	30.26	9.22	38.86	19.11	2.54

Note. This table shows counterfactual 4 when ϕ is estimated using only cross-sectional data, not the panel data created in the experiment. In counterfactual 4, students believe all majors have the same earnings parameters as a humanities major. We report the median percent change in choosing each major, as well as the 10th and 90th percentile of changes in brackets. Model 1 is the utility specification with individual heterogeneity in both θ_i and β_i. Model 2 is the utility specification with individual heterogeneity in β_i, but with homogeneous $\theta = 1.7$. Model 3 is the utility specification with homogeneous $\theta = 1.7$ and $\beta = 0.95$. Model 4 is the ϕ estimate for the utility specification with homogeneous median values of θ (0.94) and β (0.89). Model 5 is the utility specification with risk neutrality, $\theta = 0$ and heterogeneity in β_i. 'Mean Likelihood, Model' is the likelihood of choosing each major in the absence of the counterfactual.