

Federal Reserve Bank of New York
Staff Reports

Human Capital Investments and Expectations about Career and Family

Matthew Wiswall
Basit Zafar

Staff Report No. 792
August 2016



This paper presents preliminary findings and is being distributed to economists and other interested readers solely to stimulate discussion and elicit comments. The views expressed in this paper are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System. Any errors or omissions are the responsibility of the authors.

Human Capital Investments and Expectations about Career and Family

Matthew Wiswall and Basit Zafar

Federal Reserve Bank of New York Staff Reports, no. 792

August 2016

JEL classification: D81, D84, I21, I23, J10, J12, J13, J16, J24

Abstract

This paper studies how individuals *believe* human capital investments will affect their future career and family life. We conducted a survey of high-ability currently enrolled college students and elicited beliefs about how their choice of college major, and whether to complete their degree at all, would affect a wide array of future events, including future earnings, employment, marriage prospects, potential spousal characteristics, and fertility. We find that students perceive large “returns” to human capital not only in their own future earnings, but also in a number of other dimensions (such as future labor supply and potential spouse’s earnings). In a recent follow-up survey conducted six years after the initial data collection, we find a close connection between the expectations and current realizations. Finally, we show that both the career and family expectations help explain human capital choices.

Key words: human capital, subjective expectations, college major, uncertainty, marriage, fertility, labor supply, gender

Wiswall: University of Wisconsin-Madison and Arizona State University, W.P. Carey School of Business (e-mail: matt.wiswall@gmail.com). Zafar: Federal Reserve Bank of New York (e-mail: basit.zafar@ny.frb.org). Luis Armona and Kevin Morris provided outstanding research assistance. The authors thank Pierre-André Chiappori for early discussions on this project, and Arnaud Maurel and participants at the 2015 Barcelona GSE Summer Forum, the 2015 Southern Economic Association Meetings, and the 2016 American Economic Association Meetings for valuable comments. All errors are our own. The views expressed in this paper are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System. Any errors or omissions are the responsibility of the authors.

1 Introduction

Early seminal work on human capital investments focused on “career concerns” motivations for human capital investment (Becker, 1962; Ben Porath, 1967), where the motivation for human capital investment is solely the gain in one’s own future labor income. While earnings are of course an important consideration, human capital could affect life in many ways and there could be a number of other motivations for human capital investments. For example, several recent studies have analyzed marriage market “returns” to human capital investment in which human capital affects an individual’s marriage prospects and the “quality” of potential spouses (Iyigun and Walsh, 2007; Chiappori, Iyigun, and Weiss, 2009; Ge, 2011; Lafortune, 2013; Attanasio and Kaufmann, 2011; Kaufmann, Messner and Solis, 2015). To the extent that human capital investments also likely lead to particular occupations and jobs offering various levels of workplace hours flexibility and accommodation for raising children, individuals could also consider how their human capital choices affect their future fertility and family life (Bertrand, Goldin, and Katz, 2010; Goldin and Katz, 2011; Flabbi and Moro, 2012; Goldin, 1997, 2014; Klevin, Landais, and Sogaard, 2015; Wasserman, 2015; Bronson, 2015).

Do young people actually consider these issues when making key human capital decisions? Much of the existing empirical work uses realized outcomes—observed differences by education in marriage rates, assortative mating, employment patterns, and fertility—to understand the choice of human capital investments. A well-known problem is that these observed outcomes may not represent “causal” relationships in the sense that we cannot use this information to directly infer the counterfactual outcome if the individual were to have made a different human capital choice. Even if we can ascertain that human capital causally affects these outcomes, it is still difficult to infer from these realized outcomes whether these outcomes were actually anticipated at the time the individual made their human capital choices, and whether they were drivers of their decision-making.

In this paper, we directly study how young people *believe* human capital investments will affect their future career and family life. We designed a survey to elicit beliefs from currently enrolled college students at a large selective private university, New York University (NYU). For this sample of nearly 500 high-ability students, we asked beliefs about future, unrealized outcomes, *if* the student was to complete various college majors and complete college at all (relative to dropping out and not completing a degree). Because we collected beliefs conditional on a set of human capital choices *for each student*, we can directly analyze how each student believes human capital will affect their future. Our work complements previous work on using expectations that has primarily focused on beliefs about future earnings. In addition to rich data on beliefs about earnings, earnings growth, and earnings uncertainty, we collect data on a variety of other beliefs, includ-

ing expectations about marriage, spousal earnings, fertility, and labor supply. These data allow us to analyze how young people perceive the trade-offs in career and family as they contemplate different possible human capital choices.

Our data shed direct light on the determinants of human capital because we collect beliefs around the time at which important human capital choices for college students are being made—what field to study and whether to stay in school at all. These *ex ante* beliefs need not be “correct” – they need not be exactly equal to *ex post* outcomes realized years later. But we argue that it is the beliefs at the time of choice, not the realized outcomes in later periods, which are fundamental to understanding choices. In a follow-up survey, we do however investigate whether expectations are predictive of actual future outcomes, and find a close connection between beliefs and outcomes.

We find that in addition to perceiving important career consequences, college students also anticipate that their human capital choices would substantially affect their future family life. Both men and women perceive a positive “marriage market” return to completing a degree, as women believe that on average the chance of being married at age 30 is nearly 13 percent higher if they complete a degree, and men believe their chances are over 35 percent higher. We also find that female college students believe that there is a marriage market “penalty” to completing a degree in science or business as these degrees, relative to a humanities or social sciences degree, are believed to reduce their chances of being married by nearly 15 percent by age 30. In contrast, males, on average, believe that college major choice (science/business versus humanities/social science) would have almost no effect on their probability of being married at any age.

We also elicit students’ beliefs regarding the earnings of their potential spouses if they themselves would complete different human capital levels. Overall, men believe that their spouses’ earnings will be lower than own earnings, while female students believe their spouse’s earnings will exceed their own earnings. We find that students perceive a large “spousal return” to completing higher-earning degrees indicating that they believe this investment will yield not only higher earnings for themselves but lead to matches with higher-earning spouses as well.

Our survey also included questions about future fertility, and we find that both men and women on average believe that completing a science or business degree rather than a degree in the humanities would reduce their expected number of children at age 30 by about 42 percent and 48 percent, respectively. Comparing the beliefs at age 45 reveals that students anticipate a life-cycle pattern. By age 45, the differences in expected fertility by human capital (science/business versus humanities/social sciences) have reduced by about half, suggesting that men and women anticipate that human capital investments will also have a large effect on the *timing* of fertility, and not just the level of fertility.

We next turn to anticipated labor supply (unconditional and conditional on being married). On average, male and female students believe completing any degree will increase their future labor

supply, and in particular completing a science or business degree rather than a humanities degree will increase their probability of full-time employment at age 30 by 15 percent (for males) and 9 percent (for females). Female students also believe that their labor supply will be substantially different if they are married: women believe that the probability of working full-time at age 30 will decrease by 18 percent when married versus single, and female students expect the probability of working part-time and not working to nearly double when married. In contrast, the average male student does not believe marriage will affect their labor supply at all.

We administered a follow-up survey nearly six years after the original data collection to try to ascertain whether these beliefs were related in a systematic way to later realized outcomes.¹ While some of the outcomes we asked about have yet to occur for our young sample (average age 25 in 2016), for several key areas we find a remarkably high level of correspondence between expectations and realized outcomes. The distributions of expected and realized own earnings are remarkably similar. Mean expected earnings are very close to the mean of the actual earnings six years later, \$73,500 versus \$75,000. In 2010, our sample of college students believed they would be working full-time 77.5 percent of the time when they were age 30, which compares favorably to the actual full-time employment rate realized six years later of 74 percent. Students however were too optimistic about marriage prospects, and over-predicted their marriage rates at early ages. But, similar to the own earnings expectations, we find that expectations regarding "spouse's" earnings are quite similar to actual realized earnings of their actual spouses or partners. Looking at individual relationships between beliefs and outcomes, we find a close and systematic relationship between earnings beliefs and actual reported earnings: students who expected higher future earnings are in fact more likely to be earning more. For women in particular, expectations about working full- or part- time, relationship status, and earnings of spouses are all positively related with actual outcomes six years later. The results from the follow-up survey indicate that the expectations data are in fact meaningful and predictive of future outcomes.

Finally, we attempt to quantify how important career and family considerations are to educational choices by estimating a simple stylized model of human capital investment incorporating perceptions of own earnings as well as the rich set of family beliefs we elicit. Overall, the results indicate that own earnings are a significant determinant for both genders, but most strongly for males, pushing them towards the higher-paying economics/business field. Marriage market returns – spouse's earnings – have sizable impacts on the choices of women in particular, pushing them away from not graduating and into economics/business majors. Fertility beliefs too appear to be an important consideration but mainly for female students, causing women to turn to humanities

¹The beliefs we elicited need not be correct, and the difference between beliefs and outcomes provides some measure of the uncertainty facing students when making important educational decisions. On the other hand, finding systematic relationships between self-reported beliefs and traditionally collected realized outcomes also helps to provide some evidence about the "reliability" of this data.

and social science degrees. Our results indicate that considering only own earnings would provide a substantially incomplete picture of how individuals actually make human capital choices.

Our empirical strategy follows the small but fast growing literature which uses subjective expectations data to understand decision-making under uncertainty (see Manski, 2004 for an earlier survey of this literature). In the context of schooling choices, Arcidiacono, Hotz, and Kang (2012), Stinebrickner and Stinebrickner (2012, 2014), Zafar (2013), Giustinelli (2015), Kaufmann (2015), and Wiswall and Zafar (2015a, 2016) incorporate subjective expectations into models of choice behavior. Our approach is closest to that of Arcidiacono et al. (2015), who elicit beliefs about earnings associated with counterfactual choices of college majors and occupations from Duke undergraduate students, and estimate ex ante returns to occupations. As in these previous studies, we collect data on expectations for a number of possible alternative choices (that is, both the future chosen alternative as well as the counterfactuals). This allows us to construct for each individual their *individual-specific* expected “return” to choosing one particular type of human capital over another. Our contribution is to extend this idea to “family” dimensions as well, and construct “returns” for not only own earnings, but several other potentially important dimensions.

This paper also contributes to the literature on gender differences in schooling choices. Males and females choose very different college majors, and with the prior literature unable to fully explain gender differences using standard observed variables (such as earnings and measures of abilities). A large residual remaining difference—sometimes labeled as differences in “tastes” is left as the default explanation (Turner and Bowen, 1999; Arcidiacono, 2004; Gemici and Wiswall, 2014). Some recent work has sought other explanations for the gender gap in major choices. Most recently, Bronson (2015) shows the importance of work hours flexibility and changes in divorce law and divorce risk in explaining longer term trends in major choices. Wasserman (2015) shows that changes in work hours requirements can affect women’s choices of medical specialties and fertility. Finally, our work also relates to the recent literature examining the returns to education in the marriage market (Iyigun and Walsh, 2007; Chiappori, Iyigun, and Weiss, 2009; Chiappori, Salanie, and Weiss, 2015). The relevant empirical literature, which we discuss later, infers marriage market returns indirectly by analyzing ex post outcomes (and assuming some form of rational expectations). Our approach of eliciting ex ante returns, at a point when individuals are making these choices, can be viewed as providing empirical underpinnings for this primarily theoretical literature.

This paper is organized as follows. In the next section, we provide a framework for understanding the relationship between ex ante beliefs, potential realized outcomes, and observed realized outcomes. We show that one can use ex ante beliefs data to robustly identify expected returns to human capital. Section 3 discusses the data collection methodology and provides descriptive statistics for our sample. To provide a basis of comparison with our beliefs, section 4 presents analysis

of various outcomes by education in the United States using recent data from the American Community Survey. The next sections (sections 5-8) examine patterns in beliefs data for various key outcomes: earnings, marriage, fertility, and labor supply. Section 9 presents results on the relationship between ex ante beliefs and realizations data collected several years later. Section 10, using a model of human capital investment estimated using our beliefs data, quantifies the relative importance of each factor to the human capital choice. Finally, section 11 concludes.

2 Model: Beliefs and Realized Choices

Our empirical strategy is to collect individual beliefs about how various future events (future earnings, labor supply, marriage, spousal characteristics, and fertility) are affected by human capital investments. In this section, we present a simple model of human capital investments in order to precisely define how the different types of data—ex ante beliefs data we analyze here and ex post realized choice data traditionally analyzed—relate to each other and to the human capital investment process.

2.1 Human Capital Investment

At the beginning of period $t = 1$, individual i chooses among different discrete levels or types of human capital $k = 1, 2, \dots, K$. Each level of human capital could be, for example, a particular college major or whether to complete a college degree at all, as we examine here, or any number of different kinds of human capital investments such as whether to finish high school or complete a job training program. We abstract from the issue that human capital investment is often a lengthy sequential process which lasts over many periods, but we return to this later when interpreting our results. Assuming individuals live $t = 1, 2, \dots, T$ discrete periods, we write the discounted expected utility from each human capital choice as:

$$E_i V_{i,k} = \sum_{t=1}^T \beta^{t-1} \int u(X) dG_i(X|k, t), \quad (1)$$

where $\beta \in (0, 1)$ is the discount rate. $u(X)$ is the utility function that provides the mapping from the vector of events X to utility. X includes a wide range of events associated with each human capital investment type. We leave X quite general – X could include future earnings, labor supply, and marital and family outcomes. X could also include events such as additional human capital investments as well, as in the case where completing one degree allows the individual to complete more advanced degrees.

$G_i(X|k, t) \in R^J$, where $J = \dim\{X\}$, is the individual's *beliefs* about the probability of the vector of future events X occurring in period t , *if* the individual were to complete human capital investment k . Our survey design directly elicits individual's self beliefs $G_i(X|k, t)$. We ask each individual what they believe the distribution of X would be in some future period t *if* they were complete some level of human capital (particular college majors or drop-out of college and not complete a degree).² $G_i(X|k, t)$ also reflects any individual uncertainty about the probability of events X occurring, and in our survey we elicit probability distributions of future outcomes (e.g. expected probability of working full-time and uncertainty about future earnings). $G_i(X|k, t)$ could be a degenerate distribution if it is the case that the individual has absolute certainty about the realization of some future outcome. But, in general, individual beliefs would be probabilistic, where the individual is uncertain about the distribution of future X realizations given some choice.³

Individual i chooses the type of human capital, k_i^* , that maximizes expected lifetime utility: k_i^* is such that $E_i V_{i,k^*} = \max\{E_i V_{i,1}, \dots, E_i V_{i,K}\}$. For each individual, the *ex post* realized distribution of future outcomes is observed only for the one optimal human capital choice.

2.2 Beliefs

The focus of our research is individual beliefs, given by the distribution $G_i(X|k, t)$. These beliefs represent “unresolvable” uncertainty as these events will not have occurred at the time of the human capital choice. Beliefs are individual-specific, potentially heterogeneous, and may not be consistent with rational expectations—that is, they could in fact be incorrect and biased relative to the actual distribution of conditional future outcomes. We refer to these beliefs as “self” beliefs, for example beliefs about what the individual would earn if she graduated with a business degree. Self beliefs are distinct from the “population” beliefs that students hold about the population distribution – for example, beliefs about the average earnings in the population for all individuals who graduate with a business degree.

²Note that while we explicitly write that beliefs are heterogeneous, as noted by the indexing of $G_i(X|k, t)$ by i to emphasize the individual specific belief data we collect, we do not index preferences $u(x)$ by i . We could allow preferences to be heterogeneous as well, but choose not to complicate the analysis in this section any further. In the empirical specification, we allow preferences to vary by gender.

³Note a key difference between our specification of future utility and the standard approach in much of the previous literature. Previous research solves for optimal decisions and associated expected outcomes and utility given the human capital investment. The expected utility from the human capital choice is then found by substituting these expected optimal utility levels into the utility function. For example, the expected utility from attending college is found by computing the future optimal level of labor supply and associated labor earnings given the individual has a college degree. Our specification of the future value has no optimization component since we ask each individual *directly* their expectations about future events (e.g., how much they expect to work and earn) given their human capital choice. We therefore allow students to solve for their own future decisions and simply substitute these reported beliefs into the utility function. In this way, we avoid explicitly solving for future choice and has the advantage of generality in allowing for non-rational expectations.

From the collection of self beliefs data we analyze *individual-level* differences in belief distributions:

$$\Delta_{G,i}(k, k') = G_i(X|k, t) - G_i(X|k', t) \text{ ex ante treatment effect.} \quad (2)$$

where we leave the indexing by t implicit. $\Delta_{G,i}(k, k')$ can be viewed as the *ex ante* (i.e. prior to human capital investment choice) “treatment effect” on the distribution of future events X (see Arcidiacono et al., 2015). (2) reflects the “causal effect” individual i expects if she chooses human capital type k rather than k'

As an example, consider beliefs about future earnings at age $t = 30$, where earnings w would be one element of the X vector, $w \subset X$. In our survey, we elicit each individual’s beliefs about their expected future earnings at age $t = 30$ if they were to complete college (call this human capital choice k) and, in a separate question, we elicit their beliefs about earnings if they were not to complete a degree (call this human capital choice k'). The difference in the individual’s point forecast of expected earnings is given by:

$$\delta_i = E_i(w|k = \text{college}, 30) - E_i(w|k' = \text{no college}, 30),$$

where these expectations are moments of the individual’s earnings belief distribution at future age 30:

$$E_i(w|k, 30) = \int w dG_i(X|k, 30) \text{ for } k = \text{college, no college.}$$

δ_i provides the individual’s belief about how their earnings would be affected if they were to complete a college degree versus the alternative of not completing a degree, and can be viewed as the *ex ante* “treatment effect” on earnings at age 30 of completing college. In our framework, we extend this idea to many other potential outcomes beyond earnings, including earnings uncertainty, marriage, fertility, spousal characteristics, and labor supply.

2.3 Potential Outcomes

Distinct from the *ex ante* beliefs of each individual $G_i(X|k, t)$, we define the *ex post* potential outcome distribution for each individual i as $F_i(X|k, t)$. $F_i(X|k, t)$ reflects the *actual* distribution of events which would occur *if* the individual chooses human capital level k . Like $G_i(X|k, t)$, $F_i(X|k, t)$ is some distribution of possible events which could occur, and $F_i(X|k, t)$ therefore reflects actual risk in *ex post* outcomes. $F_i(X|k, t)$ can also be degenerate in the case where there is a single vector of events which would occur with probability 1 if the individual chooses human capital level k . Importantly, $G_i(X|k, t)$ and $F_i(X|k, t)$ need not be the same distribution if the individual’s beliefs are incorrect. The common assumption that individuals’ *ex ante* beliefs are equal to the distribution of actual *ex post* outcomes imposes that $G_i(X|k, t) = F_i(X|k, t)$

(point-wise). This is often referred to as “rational expectations,” where the individual knows the *distribution* of future events conditional on choices but does not know which event will be realized. Perfect foresight, an even stronger assumption, imposes that beliefs are not only correct but also degenerate at the actual future realizations.

A common object of interest is the ex post difference in *potential* outcomes between choices k and an alternative k' :

$$\Delta_{F,i}(k, k') = F_i(X|k, t) - F_i(X|k', t) \text{ ex post treatment effect.} \quad (3)$$

These represent the individual-level potential treatment effects. Only one of the choices $k = 1, \dots, K$ can be made, and only one of the outcome distributions is realized and observed in standard realized choice data. The alternative outcomes are counterfactual and unobserved. A large econometrics and statistics literature studies how to identify these potential outcomes and moments of the potential outcomes (e.g. average treatment effects) from realized choice data (e.g. Heckman and Vytlacil, 2005; Angrist and Pischke, 2009; Imbens and Rubin, 2015). For example, assuming the population is of finite size M , we can define a vector of average treatment effects (ATE) of choosing k rather than k' as $ATE(k, k') = M^{-1} \sum_{i=1}^M \Delta_{F,i}(k, k')$.

2.4 Realized Outcomes

Given the human capital choice model, we can define the distribution of outcomes actually observed in traditional realized choice data. Assume the population is of finite size M . Let Ω_k be the subset of the population which optimally chooses human capital choice $k = k^*$, and $M_k \in (0, M]$ be the size of this sub-population choosing k . The distribution of realized events X for the population which chooses k is

$$H(X|k, t) = M_k^{-1} \sum_{i \in \Omega_k} F_i(X|k = k^*, t), \quad (4)$$

where $H(X|k, t)$ is a properly scaled distribution function of realized outcomes. Importantly, $H(X|k, t)$ is not indexed by i because it represents a population distribution for the sub-population which chooses choice k . $H(X|k, t)$ is the typical data on realized outcomes. The difference in the distribution of outcomes X observed in the population which chooses human capital k rather k' is:

$$\Delta_H(k, k') = H(X|k, t) - H(X|k', t) \text{ ex post observed difference in outcomes} \quad (5)$$

Note that $\Delta_H(k, k')$ is not indexed by i because it represents a population distribution.

2.5 Summary: “Effects” of Human Capital

In summary, we have defined three distinct “effects” of human capital choice k (rather than k') on outcomes X :

(i) *Ex ante* individual difference in beliefs: $\Delta_{G,i}(k, k') = G_i(X|k, t) - G_i(X|k', t)$,

(ii) *Ex post* individual differences in potential outcomes: $\Delta_{F,i}(k, k') = F_i(X|k, t) - F_i(X|k', t)$,

(iii) *Ex post* differences in realized (observed) outcomes: $\Delta_H(k, k') = H(X|k, t) - H(X|k', t)$.

Ex ante beliefs are the sole determinant of choices—and the focus of this study. Beliefs need not coincide with *ex post* outcomes (if beliefs are not perfectly correct). In addition, the well-known result is that *ex post* observed outcomes do not identify *ex post* potential outcomes without additional assumptions on the selection process.⁴

3 Data

This section describes the survey administration, the survey instrument, and the sample selection.

3.1 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. The 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 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. When appropriate, 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.

⁴In much of this literature, the selection models assume some form of rational expectations where it is assumed that $F_i(X|k, t) = G_i(X|k, t)$. One notable exception is Cunha, Heckman, and Navarro (2005) which estimate a model in which agents have unbiased beliefs about future outcomes: $\int XG_i(X|k, t) = \int XF_i(X|k, t)$, but agents may not know the level of uncertainty in outcomes (variance in future earnings in their case).

The survey was part of a larger experiment in which students were provided different types of information in successive rounds (see Wiswall and Zafar, 2015a, b). The data analyzed here are from questions administered before any information is given to respondents. Before the official survey began, survey respondents were also first required to answer a few simple practice questions in order to familiarize themselves with the format of the questions.

3.2 Survey Instrument

Our goal was to collect information on consequential life activities that would plausibly be key determinants of a college major. Although we aggregate the data to 3 groups to simplify the analysis below, in constructing survey questions, we conditioned each question on the choice of 1 of 5 potential “majors”: 1) Business and Economics, 2) Engineering and Computer Science, 3) Humanities and Other Social Sciences, 4) Natural Sciences and Math, and 5) Never Graduate/Drop Out. The last category was included to complete the choice set, but as revealed by the respondents, not graduating from college was a relatively low probability event.⁵ We provided the respondents a link where they could see a detailed listing of college majors (compiled from various NYU sources), which described how each of the NYU college majors maps into our aggregate major categories.

Questions about the future were conditioned on three particular future points in time: immediately after graduation (approximately age 23), at age 30, and at age 45. In the analysis sections below, we discuss the specific format of the relevant questions. The text of our survey instrument is available upon request.

3.3 Sample Selection and Descriptive Statistics

A total of 501 students participated in the study. We dropped 6 students who report that they are in the 4th year of school or higher, violating the recruitment criteria, and we further dropped 2 respondents who apparently completed the survey twice. The final sample consists of 493 individuals, 176 men and 317 women. The gender composition of our sample compares favorably with that of the NYU undergraduate population: males constituted 33% of the graduating class of 2010 at NYU, and 36 percent of our sample.

⁵We worded the last possibility as “you will never receive a Bachelor’s degree from NYU or any other university).” The question wording makes it clear that transferring to another school is not in the student’s choice set. This was done given the low transfer rates of NYU students, and to keep the scenario simple. At NYU, 8 percent of freshman leave after their first year (average from 2008-11 from US News and World Report). This figure includes both transfers to other universities and drop-outs, thus the transfer rate is at most 8 percent. For our high-ability sample, the expected transfer rate is likely much lower than this. We find in our sample that the expected probability of dropping out is between 2-3 percent.

Sample characteristics by gender are shown in Table 1. The table also reports the significance level of the t-test for differences in observed characteristics by gender. Male and female respondents are 42 and 36 percent white, respectively, and both are around 44 percent Asian. The mean age of the respondents is about 20, with 46 percent of males reporting they are in the first year (freshman), 33 percent sophomores, and 21 percent juniors. For women, 37 percent are in the first year, 38 percent sophomores, and 25 percent juniors. The average grade point average of our sample is nearly 3.5 (on a 4.0 scale), and the students have an average Scholastic Aptitude Test (SAT) math score of 716 (men) and 700 (women), and an average verbal score of 682 (men) and 692 (women), where the maximum score on both tests is 800. These averages correspond to the 90th and higher percentiles of the population score distributions. Therefore, as expected, our sample represents a high ability group of college students. Parents' characteristics of the students also suggest that they are over-represented among high socioeconomic groups. Except for an over-representation of freshman among the male sample and a higher average SAT math score for males, none of these differences in descriptive statistics by gender is large or statistically significant at the 10 percent level.

The bottom panel of Table 1 shows that men and women choose very different college majors, mirroring the national patterns. Here we report the students' current or intended major. 42 percent of males report majoring in Business/Economics, and 34 percent majoring in Humanities and Social Sciences, and 18 percent in Natural Sciences/Math. On the other hand, 56 percent of the females report majoring in Humanities and Social Sciences, followed by about 24 percent majoring in Business/Economics, and 16 percent majoring in Natural Sciences/Math. That is, female students are more than 50 percent more likely as males to major in the Humanities, and are about half as likely to major in Economics or Business. These gender-specific major distributions are statistically different ($p\text{-value} \leq 0.001$, using a Chi-square test for equality of distributions).

The substantial gender differences in major choice are similar to the national patterns from recent 2009 American Community Survey (ACS) data, a representative sample of the US population. In the ACS, among college graduates aged 22-24 in 2009, 62 percent of males completed a science, engineering or business degree compared to 66 percent in our sample. In the ACS data, 39 percent of women completed a science, engineering or business degree, compared to 44 percent in our sample. While our sample is clearly over-representative of high ability and high socioeconomic status students, the gender differences in major choice are largely similar to those we see nationally for recent college graduates.

4 Current Population Characteristics

To set the stage for our analysis of beliefs, we first present a brief overview of earnings, employment, and marriage for the US population using the 2009 American Community Survey (ACS). Given the timing of our survey administration (survey conducted in 2010), this snapshot of the US population would be approximately representative of the current US population at the time the survey respondents completed our survey.

It is important to note three key differences between the data from the ACS and our data on student ex ante beliefs. First, and perhaps most fundamentally, the ACS data provides the distributions of ex post outcomes, $H(X|k, t)$, defined in 4. These need not reflect a causal difference in ex post potential outcomes $F_i(X|k, t)$ and need not reflect the beliefs of the students $G_i(X|k, t)$ we collect in our data. Second, the ACS data is for an older generation of individuals, rather than the population outcomes for the same birth cohort in our sample; many of the future outcomes we ask our sample about (e.g. earnings when they are age 30) have yet to actually occur (our sample is aged around 25-27 in 2016). Below we compare actual realized outcomes for respondents using our recent 2016 follow-up survey. Third, the ACS is representative of the US population, rather than the high-ability sample we have in our data collection, although, as seen above, the major distribution in the two datasets by gender is quite similar.

We consider three groups in the ACS data: 1) college graduates (completed undergraduate degree or higher) in a business, economics, science, or engineering field, 2) college graduates with a degree in any other field (humanities, education, arts, and other social sciences), and 3) individuals with “some college,” with as little as 1 year of college to several years of college, but no undergraduate degree. We refer to these groups as “science/business”, “humanities”, and “no degree”, respectively. These groups were chosen to mimic as closely as possible the possible future educational status of our survey respondents, who consisted of currently enrolled freshman, sophomore, and junior students.⁶ With these caveats in mind, we argue that it is still informative to examine the ACS data to provide some comparison to our data on beliefs.

Table 2 provides information on earnings, spousal earnings, employment rates, and marriage rates for three ages (age 23, 30, and 45), matching our data collection. Following our survey design, we examine earnings for full-time workers (work 35 or more hours per week and 45 or more weeks per year). Average full-time earnings differences across major fields, and even between college graduates and those with only some college, are small at age 23, but become increasingly larger by age 30. At age 45, science/business majors on average earn nearly 30 percent more than

⁶We cannot of course construct a more exact match for the high ability students in our survey sample given that we cannot identify colleges and universities attended by the ACS sample; nor can we condition on measures of ability, such as SAT scores, since these data were not collected by the ACS. In addition, many of the “some college” individuals in the ACS never enrolled in four-year colleges, but instead enrolled in two-year post-secondary schools.

humanities majors, and humanities majors earn over 50 percent more than individuals with No Degree (some college but no four-year degree). We also note that, even by these aggregate college major categories, there are substantial gender differences in earnings, with women earning less than men in each group.⁷

We also examine the earnings of spouses, conditional on the individual's education. Note that these statistics are not by the spouse's education, but by the individual's *own* education, and therefore reflect the extent of assortative marriage patterns. The differences in the earnings of one's spouse by his or her own education type are large and of a similar magnitude as the differences by education for one's own earnings. Individuals who completed a college degree in science or business fields are married to spouses who on average earn nearly 30 percent more than the spouses of individuals who completed a degree in humanities. These differences in spousal earnings are similar for men and women, although across all fields, women's spouses have higher average earnings than men's spouses. These patterns are indicative of assortative mating by education, both on the extensive margin (years of schooling) and the intensive margin (field of study), consistent with findings of Eika, Mogstad, and Zafar (2015).

In addition to full-time earnings, the ACS data also provide information on full-time employment rates across education types. The most salient pattern is that women are less likely to be working full-time than men. There are smaller differences in employment rates across education types within gender, but, in general, at age 30 and 45, science and business graduates are more likely to be employed full-time than humanities graduates. These differences, of course, may arise because of either voluntary choices (selection) or be involuntary.

The marriage rate patterns too show a distinct overall gender pattern as women are marrying at younger ages than men. Marriage rates also vary by education. College graduates overall are more likely to be married at older ages than non-college graduates. At age 30 and 45, college graduates with science and business degrees are more likely to be married than humanities majors, and humanities majors are more likely to be married than non-college graduates.

5 Earnings Beliefs

We begin our analysis of student beliefs by examining how students perceive human capital investments relate to their own future earnings. These beliefs, which we term as "self" beliefs, are distinct from an individual's "population" beliefs—beliefs about the distribution of earnings in the population, such as the average earnings of current workers.⁸ Our survey asked for earnings be-

⁷See Altonji, Arcidiancono, and Maurel (2015) for a much more detailed analysis of earnings by college majors.

⁸See Wiswall and Zafar (2015a, b) for analysis of population beliefs and the relationship between self and population earnings beliefs.

beliefs for 3 future periods: i) the first year after college graduation (when most respondents would be around age 23), ii) when the respondent would be aged 30, and iii) when the respondent would be aged 45. At each of those periods, we ask respondents for their beliefs about their *own* earnings *if* they were to be working full-time.⁹ We elicited beliefs about future labor supply, including the students' beliefs about the probability they would be working full-time, in a separate set of questions, which we analyze below. Individuals were instructed to consider in their response the possibility they might receive an advanced/graduate degree following their undergraduate degree, and therefore the beliefs about earnings we collected should incorporate 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 and report earnings in current (i.e. 2010 dollars).

5.1 Earnings Levels

Panel A of Table 3 presents the average earnings beliefs at each of the future ages (age 23, 30, and 45). In this top panel, the numbers in parentheses are standard deviations, and not standard errors; all of these average earnings beliefs are statistically different from zero.¹⁰ Immediately after college graduation (about age 23), the average belief of female students is they will earn annual full-time earnings of \$53,900 if they complete a degree in science or business, \$39,400 if they complete a degree in the humanities or social sciences, and \$24,500 if they fail to complete a college degree. For the initial post-graduate period, male students have similar beliefs, with average beliefs of \$59,300, \$47,100, and \$35,000 in the science/business, humanities/arts, and not graduate categories, respectively. There is substantial heterogeneity in self beliefs, as seen in the large standard deviations (relative to the means).

The survey reveals that students believe they will experience rapid growth in future earnings, with especially rapid growth if they were to complete a degree in science or business. By age 30, the average annual full-time earnings expected if the student completes a major in science or business has grown to over \$108,600 for women, and over \$137,400 for men. By age 45, students average earnings beliefs grow to \$138,100 for women, and \$190,000 for men. Table 3 indicates that students expect substantially smaller earnings growth if they were to complete a degree in humanities/arts or not graduate. Figure 1 graphs the growth in average earnings beliefs. The figure

⁹The survey text was as follows: "*Employed full time is defined as working for pay at least 35 hours per week and 45 weeks per year. The remaining individuals are working part-time, unemployed, in school, raising children but not working full time, or otherwise not in the labor force.*"

An example question on expected earnings at age 30 is as follows: "*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?*"

¹⁰We censor reported beliefs about full time annual earnings (population or self earnings) so that earnings below \$10,000 are recorded as \$10,000 and earnings reported above \$1,000,000 are recorded as \$1,000,000.

shows the differences in the perception of the gender gap conditional on human capital choice. For each category (science/business, humanities/arts, and no degree) and at each future age, men expect to earn more than women. But the perceived gender gap is largest at later ages and much larger in the science/business field.

The average overall expected earnings – obtained from weighting the earnings beliefs for the major categories by the student’s reported probability of completing a degree in that field – are reported in the last row of Panel A. This represents the unconditional expected earnings of this individual over the possible human capital choices. Average expected earnings are higher for males at each age, with the differences highly significant at later ages: for example, male students on average expect to earn \$184,400 at age 45, while females expect to earn \$123,300 on average.

One may wonder about the accuracy of these expectations. In a later section we compare the earnings expectations in 2010 with actual realized earnings for these same respondents in 2016. We find that earnings expectations and realized earnings are systematically related, providing some strong evidence that expectations data we collect is not pure “noise.” In general however, we cannot test directly the accuracy of expectations because, by construction, these outcomes would be realized only in the future and, importantly, would be observed only for the one chosen education choice. It is important to note that it is the *expected* outcomes – as perceived by the respondent at the time of choosing their human capital investment – that matter in the choice decision (of which we show evidence in section 10). Whether these perceptions are biased is then arguably an irrelevant issue from the perspective of understanding the decision.¹¹

We can also examine the “reasonableness” of the beliefs data by comparing the beliefs with actual realized earnings using the ACS data, as described above. As would be anticipated given the high ability sample, it is clear that the students in our sample believe they will earn far more than the population average earnings, as seen in the ACS data. When asked their *beliefs* about the population average earnings directly in another part of our survey, on average, the sample respondents directly report that they believe they will indeed earn substantially more than what they believe is the population average (Wiswall and Zafar, 2015a). However, as discussed below, the qualitative patterns are similar in that the students in our sample expect substantially higher earnings in business and science majors relative to humanities, and higher earnings in humanities than if they were not to graduate.

¹¹Systematic biases in expectations do, however, imply a policy case for information interventions, something that is not the subject of the current study. Interested readers should look at Wiswall and Zafar, 2015a, and references therein.

5.2 Earnings Return

We next turn to the individual-level beliefs about how each student perceives the effects of human capital on their future earnings. For each individual, we construct a set of log earnings ex ante “treatment effects”:

$$\delta_i = \ln(E_i(w_{it}(k))) - \ln(E_i(w_{it}(k'))),$$

where $E_i(w_{it}(k))$ is individual i 's belief about what her average earnings would be at age t if she were to complete education type k . $E_i(w_{it}(k'))$ is then the belief of the same individual but for an alternative type of education k' . δ_i is then individual i 's belief about how her expected earnings w (at some future period t) would change if she were to complete human capital level k rather than k' . In general, δ_i can be positive if the individual perceives a positive return to human capital investment k over k' , or negative if they believe their earnings would actually decline if they were to choose k rather than k' . As we detail in the modeling framework section above, this individual-level difference, an ex ante “treatment effect,” is never directly observed in traditional observational data.¹²

For our three categories of human capital investment, we construct two differences: (i) business/science college degree versus humanities college degree, and (ii) overall college degree versus no degree. The overall college degree category is constructed by weighting the earnings beliefs for the two college major categories (business/science and humanities/arts) by the reported beliefs about the probability that the person would complete a degree in this field. This represents then the expected earnings of this individual from completing their college degree. Difference (i) reflects the individual's belief about the intensive margin return to a business or science degree over the alternative humanities degree if they were to complete a college degree. Difference (ii) reflects the individual's belief about the extensive margin return to a college degree relative to not completing any college degree.

The bottom panel of Table 3 presents statistics of the log earnings ex ante treatment effect. We see that, on average, female students perceive just over a 30 percent return in age 23 earnings to completing a science or business degree rather than a humanities degree, and, on average, male students perceive a return of nearly 27 percent. The expected return at age 23 earnings to completing a college degree rather than no degree is even larger: 64 percent and 59 percent for female and male students, respectively. At age 30, the expected returns have grown to 43 (female) and 52 (male) percent for a science/business degree versus a humanities degree, and in excess of 100 percent for both genders for a college degree versus no degree. By age 45, students expect similar percentage returns to human capital: 35 (female) and 45 (male) percent return for a science or

¹²Note that we cannot directly estimate the an individual specific log return as $E_i(\ln w_{it}(k) - \ln w_{it}(k'))$ without assuming a particular distribution for the earnings beliefs or collecting beliefs about expected log earnings. Our statistic δ_i is intended to give an approximation of the expected return from choosing one major over another.

business degree, and about 82 percent return for both genders for a college degree relative to no degree. In general, we find that men expect a higher average return to a science or business degree (relative to humanities degree) than women.

While the average returns indicate an important general pattern, there are salient patterns in heterogeneity of returns. Figures 2 and 3 present the distribution of δ_i for the college versus no degree return and the science/business versus humanities return (for full-time age 30 earnings), respectively. The figures indicate that there is substantial dispersion in the distribution of expected returns, with some individuals expecting a very high return (more than a 100 percent difference in earnings) and others a small return, and for a small minority even a negative return.¹³

Another way to view these joint distributions of expected labor market returns to human capital is to examine the individual-level correlation in earnings across fields. In Appendix Table A1, we report the correlation matrix across the three types of human capital. We find that for both male and female students, there is a generally high positive correlation in self earnings across fields: individuals who believe they will have high earnings in one type of human capital also believe they will have high earnings in other types.

5.3 Earnings Growth

In addition to levels of earnings, perceived earnings growth may also be an important consideration.¹⁴ Given that we have collected data on full-time earnings beliefs at three ages for each individual, we can also construct individual-specific beliefs about the *growth* in earnings.

In standard datasets, it is apparent that realized earnings growth over the lifecycle is large, and the differences in growth rates across human capital investments can be an important factor in choice. Consider an individual's perception of their growth in earnings from graduation (age 23) to age 30, where the perceived log growth in earnings between ages 30 and 23 for individual i is given by $\ln w_{i,30}(k) - \ln w_{i,23}(k)$. Individual i 's belief about how this growth in earnings would vary between education k and education k' is then given by:

$$\frac{\ln w_{i,30}(k) - \ln w_{i,23}(k)}{\ln w_{i,30}(k') - \ln w_{i,23}(k')} \quad (6)$$

¹³The 5th percentile of the both the men and women's distribution of business/science return is actually negative, indicating that some students in the sample believe their earnings would be *lower* if they were to graduate with a science or business degree rather than a humanities degree.

¹⁴For example, Berger (1988), Flyer (1997), and Eide and Waehrer (1998) all fit multinomial logit models to data on undergraduate major choices. Each relaxes the assumption that college graduates only consider starting wages in their choice. The authors calculate imputed measures of lifetime wages taking into account differing growth in earnings over the life-cycle, and find that these imputed measures of lifetime earnings are correlated with undergraduate major choice. More recent papers, including Arcidiacono (2004), Gemici and Wiswall (2014), and Wiswall and Zafar (2015a) also include earnings growth in their models of field of study choice.

The top panel of Table 4 shows earnings growth beliefs between age 23 and age 30 and between age 30 and age 45. The students in our sample expect large earnings growth in the early part of their career, with men believing their earnings will on average grow by 66 percent in the 7-year period from age 23 to age 30 and women believing their earnings will grow by 60 percent. Students expect earnings growth to slow considerably in their later career as their average expectations about earnings growth in the 15-year period from age 30 to age 45 are 28 percent for men, and 23 percent for women. These expectations are consistent with the concavity of the age-earnings profile, which has been a well-studied feature of realized labor market earnings (e.g. Heckman, Lochner, and Todd, 2008).

Panel B of Table 4 shows there are considerable differences in perceptions of how earnings growth would differ depending on the completed level of education. In general, students believe their early career earnings growth would be much larger if they complete a college degree, in particular a degree in science or business rather than humanities, than if they fail to complete a college degree. Males (females) on average expect earnings to increase by roughly 42 (39) percent more for a college degree versus no degree during the 7-year period from ages 23-30, and by 26 (12) percent more for a science or business degree rather than a humanities degree. The pattern reverses somewhat for later career earnings growth as students believe there will be a much larger slow down in earnings growth after age 30 if they complete a college degree than if they do not.

5.4 Earnings Uncertainty

Previous research has show that uncertainty about future earnings could play a role in educational choices (Altonji, 1993; Saks and Shore, 2005; Nielsen and Vissing-Jorgensen, 2006). Most prior empirical literature elicits only the average returns to schooling choices; Attanasio and Kaufmann (2011) and Wiswall and Zafar (2015a, b) are exceptions that collect data on risk perceptions of schooling choices. In addition to questions about expected (mean) earnings at various ages, we also asked respondents about the percent chance that their own earnings would exceed \$35,000 and \$85,000 at both ages 30 and 45.¹⁵

To compute a simple statistic reflecting uncertainty about future self earnings, we use the standard deviation obtained from fitting a beta distribution to each respondent's percentile responses. *For each individual and major*, we fit a 2-parameter beta distribution, and restrict the support of each distribution to be between zero and one million dollars, generating 2,465 distributions (493 respondents x 5 major categories).¹⁶

¹⁵The question was asked as follows: "What do you believe is the percent chance that you would earn: (1) At least \$85,000 per year; (2) At least \$35,000 per year, when you are 30 (45) years old if you worked full time and you received a Bachelor's degree in each of the following major categories?"

¹⁶Of the possible, we are unable to fit the distribution in about 8 percent of cases. This is a result of students

Panel A of Table 5 shows the average standard deviation of the fitted distributions for age 30 earnings. The average standard deviations are similar for science/business and humanities fields and for the two genders. Students perceive greater uncertainty in age 30 earnings on average for the no degree major. The average standard deviation in age 30 earnings for females is \$152,700 for no degree versus \$94,900 for science/business.

In order to examine perceptions of the differences in uncertainty by education, we construct differences in perceived standard deviation of age 30 earnings for individual i between education k and education k' as: $\ln \sigma_{w_i,30}(k) - \ln \sigma_{w_i,30}(k')$, where $\sigma_{w_i,30}(k)$ is the standard deviation in age 30 earnings as perceived by i in education k . This metric would be positive if the individual perceives higher uncertainty in returns to human capital investment k than k' . Panel B of Table 5 reports these statistics. Students on average expect substantially lower uncertainty in age 30 returns if they are to complete a college degree than if they are not: the standard deviation is lower by at least 30 percent. Average earnings uncertainty, for both males and females, is expected to be lower for a degree in science or business rather than humanities. College degrees and business/science degrees are perceived to not only bring higher expected earnings, but also lower earnings uncertainty.

6 Beliefs about Marriage and Spousal Characteristics

While future earnings are certainly a key consideration in education choices, another potentially important consideration is how human capital affects marriage prospects and the type of spouse one might marry. Recent theoretical models have emphasized that investment in education generates returns in the marriage market (Iyigun and Walsh, 2007; Chiappori, Iyigun, and Weiss, 2009; Chiappori, Salanie, and Weiss, 2015). Empirical work in this area (using ex post outcomes) finds evidence consistent with this, but the literature usually infers the returns in the marriage market indirectly.¹⁷ In this section, we examine whether college students believe human capital would

violating monotonicity in their reported probabilities of earning more than the two thresholds (that is, reporting a higher probability for the \$85k threshold than for the \$35k threshold). Such violations in responses to survey questions that elicit distributions have been documented in earlier studies (see, for example, Manski (2004) and the studies cited therein). Generally, around 10% of the responses are discarded in such cases. One approach is to enforce respondents to obey monotonicity, and provide them a warning message if their responses are inconsistent. We chose not to do so because we believed that may possibly confuse some respondents.

¹⁷Ge (2011) estimates a structural dynamic (partial equilibrium) model of college attendance using the NLSY 1979, and shows that marriage plays a significant role in a female's decision to attend college. Lafortune (2013) shows that a worsening of marriage market conditions spurs higher pre-marital investments—in particular for males—in her sample of second-generation Americans born around the turn of the twentieth century, and argues that part of this occurs through the anticipated shift in after-marriage bargaining power. Attanasio and Kaufmann (2011), using gender ratios in the locality as a proxy for returns to education in the marriage market, find that marriage market considerations are important in females' schooling choices in Mexico.

The most direct evidence on marriage market returns of (quality of) higher education comes from Kaufmann, Messner, and Solis (2015). Using rich administrative data and the unique features of the Chilean higher education admission system, they find a substantial effect of being admitted to a higher ranked program on spousal quality, but only for

affect various aspects of their future personal relationships.

6.1 Marriage Probabilities

We first examine whether individuals perceive that different human capital investments would affect their probability of being married in the future. We elicit the individual's belief about the percent chance of being married at various future ages, conditional on human capital (college majors and a not graduate category). Due to time constraints and to err on the side of simplicity, our survey includes only a married or non-married state, and the non-married state includes being never married or divorced. The top panel of Table 6 shows that the average probability of being married (unconditional on human capital, that is, "overall") grows from between 15-18 percent shortly after graduation (about age 23), to around 59-63 percent at age 30, and to about 80 percent by age 45. In general, women believe they are slightly more likely to be married than men at younger ages, but men and women have about the same average belief about the probability of being married at age 45. Comparing the beliefs in our sample to the realized marriage rates in the ACS data (Table 2) reveals that the marriage beliefs of our sample are quite similar to the national figures. Just as in the realized data, our sample believes that marriage rates grow rapidly during their 20s and peak at around 80 percent by age 45. Our sample also has beliefs about the gender difference in the timing of marriage which are consistent with the realized marriage data, as the women in our sample believe they will likely be married at younger ages than the men.

Examining beliefs by type of human capital, the average expected marriage probabilities at age 30 and 45 are generally lower for the no degree field than for the other two fields, for both genders. For male respondents, the average marriage likelihood is similar for a degree in science or business and humanities at the three ages. Females, on the other hand, reports higher marriage likelihood at age 30 for humanities relative to science/business.

We next construct individual-level differences in the probability of being married by human capital level. We construct marriage market "returns" in the same way as we constructed labor market earnings returns in section 5.2. Panel B of Table 6 indicates that women believe that completing a degree in science or business rather than in humanities or social sciences reduces their chances of being married at age 23 by nearly 10 percent, and by nearly 15 percent by age 30, with both differences significant from zero at conventional levels.¹⁸ In contrast, males, on average, believe that college major choice (science/business versus humanities) would have almost

females. They, however, find no impact on fertility or marriage.

¹⁸Note that the individual-level differences are log differences, as with the earnings differences analyzed above. A change in the probability of marriage from 0.45 to 0.5 is therefore $\ln(0.45) - \ln(0.5) = -0.1054$, or approximately a 10 percent reduction. When we construct individual-specific log differences in beliefs, we recode all reported extreme probabilities of 0 to 0.001 and 1 to 0.999. This follows Blass et al. (2010) who argue that dropping individuals with extreme probabilities would induce a sample selection bias in the resulting estimates.

no effect on their probability of being married at any age. At 30, we have enough power to reject the hypothesis that the difference in male and female average beliefs are statistically significant from zero at the 5 percent level. Interestingly, the beliefs about the effect of college major on marriage is one of timing only. At age 45, the differences in expected marriage rates by college major are small in magnitude and not statistically significant for both men and women. This finding indicates that women believe that choosing a more technical college major in science or business would likely cause them to delay marriage, but that marriage rates at age 45 would be about the same no matter what their college major.

Turning to the college versus no degree margin, we find that both men and women perceive a positive “marriage market” return to completing a degree, as women believe that on average the chance of being married at age 30 is nearly 13 percent higher if they complete a degree and men believe their chances are over 35 percent higher. This return to human capital persists to age 45 suggesting that both men and women believe there is a permanent loss in marital prospects if they do not complete a college degree. While we cannot directly separate various reasons for these beliefs, it appears that men in particular believe that higher levels of human capital—completing education associated with higher earnings—would make them much more attractive to potential spouses.

6.2 Marital Sorting: Potential Spouse’s Earnings

In addition to human capital investment affecting the probability of being married, human capital can also affect who one is married to, and this topic of assortative mating is the subject of a large body of research in the social sciences (Mare, 1991; Pencavel, 1998; Greenwood et al., 2014; Eika et al., 2015). Our survey allows us to directly examine perceptions about marital sorting among young people. We ask respondents to tell us their beliefs about the characteristics of potential spouses *if* the individual was to make various types of human capital investments. The text of the question is “*If you are X years old and married, and if your spouse is working full-time, what do you believe is the average amount that your spouse would earn per year if you received a Bachelor’s degree in each of the following major categories?*”, where $X = \{23, 30, 45\}$. Importantly, we emphasized to respondents that they were to report beliefs about their spouse’s earnings conditional on their *own* major, not the potential spouse’s major. In a separate set of questions analyzed later, we ask directly about the spouse’s major.

The top panel of Table 7 describes beliefs about spouse’s earnings conditional on the individual’s own human capital investments. Compared to beliefs about own earnings in Table 3, beliefs about spouses’ earnings are generally lower than own earnings for male students, while female students believe their spouse’s earnings will exceed their own earnings. There are substan-

tial differences in spousal earnings across own major choices, with both male and female students expecting their spouse's earnings to be the highest if they themselves graduated with a science or business degree, and lowest if they fail to graduate. These qualitative general patterns are consistent with the realized data on currently married couples as revealed in the ACS (Table 2), although the level of expected spousal earnings in our data is substantially higher than that in ACS, likely reflecting the high ability nature of our sample as documented above.

Following the analysis as before, we construct individual log differences in spousal earnings by human capital category to create a spousal earnings "return" for the individual's human capital choice. These are presented in Panel B of Table 7. Average spousal earnings returns are positive for major choice (science or business relative to humanities) and for completing college, indicating that individuals believe their investment in higher earning majors and completing a college degree will yield not only higher own earnings but lead to matches with higher earning spouses as well. The spousal earnings return expected by the students are substantial. Men and women believe their spouses (if they were to be married) would be at least 28 percent higher at age 30 if they completed a science or business degree than a humanities degree. Women perceive a particularly high return to completing a college degree: women believe that their spouse would earn at least 74 percent more if they completed a college degree than if they did not at age 30, and nearly 63 percent more at age 45.¹⁹

Because we elicit beliefs about own and spouse's earnings, we can compute the correlation in perceived own and spouse's earnings. The correlation in age 30 earnings, weighted over majors, is 0.76 (0.70) for females (males). The correlation decreases to 0.71 (0.62) for age 45 earnings, but continues to be very high and significant.²⁰

6.3 Marital Sorting: Assortative Mating by Education

The patterns in the previous section indicate that students perceive sorting of spouses by own major choice, and is suggestive of strong assortative mating by education. The expected returns to spousal earnings (in Panel B of Table 7) are sizable, but smaller than expected self earnings returns

¹⁹Figures A1 and A2 present the distribution of spousal earnings returns for science/business versus humanities, and for college versus no degree (for full-time age 30 earnings), respectively. We see that there is substantial dispersion in the distribution of expected spousal returns, with most (male and female) individuals expecting very high returns. For a small minority, however, there are negative returns. The gender-specific distributions are generally quite similar.

²⁰We can also measure *each* individual's beliefs about their earnings relative to their potential spouse. Appendix Table A2 presents statistics from the distribution of beliefs for own relative to spouse's earnings (specifically, $\log(\text{own earnings}) - \log(\text{spouse earnings})$). Overall, women expect to earn nearly 7 percent less than their spouse at age 30, and 3.5 percent less than their spouse at age 45, although the latter estimate is not significant at the 10 percent level ($p\text{-value} = 0.172$). Men, on the other hand, expect on average to earn 29 percent more at age 30 and 42 percent more at age 45 than their spouses. Clearly, these beliefs are not necessarily consistent with each other if the men in our sample expect to marry the women in our sample.

(lower panel of Table 3). This indicates that students do not expect spousal mating by education to be perfectly assortative.

Our survey allows us to directly examine perceived assortativeness by education. Students were asked: "*If you are 30 years old and married, what do you believe is the percent chance that your spouse received a Bachelor's degree in each of the following major categories if you received a Bachelor's degree in each of the following major categories?*" Table 8 shows the average responses to this question. Both female and male respondents expect science/business to be the modal spousal major regardless of the major they themselves graduate with. However, the table clearly shows that matching by education is perceived to be assortative. For example, the average likelihood of the spouse having a degree in science/business is at least 70% for both male and female respondents if they were to themselves graduate with a science/business degree. This probability drops to less than 60% if the respondent were to graduate with a humanities/social sciences degree or to not graduate. Respondents believe that they are significantly more likely to have a spouse without a degree if they were to not graduate than if they were to have a college degree. Table 8 also shows that the perceived assortativeness in our sample does not differ by gender.

Overall, the perceived patterns are consistent with the well-documented fact that there is assortative mating by education (more precisely, the extensive margin of years of schooling) in the US. Evidence on assortative mating by field of study (the intensive margin) is scant since datasets with information on fields of study of both spouses have begun to be collected for the US in recent years only. However, Eika et al. (2015) document evidence of assortative mating by field of study in Norway (with the level of assortative mating being higher than that observed on the extensive margin).

7 Beliefs about Fertility

We next explore whether students believe that human capital investments would affect their future fertility. Two mechanisms could be at work. First, because of the important connection between human capital and income, students could believe that human capital investments could change the expected future value of children, either increasing the demand for children as children are a normal good or reducing the demand for children through the child quantity-quality trade-off (Becker and Lewis, 1973).²¹ Another channel is that human capital investment could lead to working in occupations and jobs which differ in various dimensions making them more or less accommodating of childrearing (Goldin and Katz, 2015; Wasserman, 2015; Cortes and Pan, 2016). If individuals ex ante perceive differences in the mapping of majors to occupations – of which there is evidence (Arcidiacono et al., 2015; Wiswall and Zafar, 2016) – and in their amenability to childrearing

²¹See Jones, Schoonbroodt, and Tertilt (2010) for various channels linking income and fertility.

(through work hours flexibility, say), then fertility beliefs may differ by type of human capital.

We directly elicited beliefs about future fertility using the following question: "*What do you believe is the percent chance of the following: a) You having no children; b) You having one child; c) You having two children; d) You having three or more children by the time you are 30 years old if you received a Bachelor's degree in each of the following?*" This fertility question was asked for two future ages: age 30 and 45. From the answer to this question, we construct each respondent's expected number of children.²²

Panel A of Table 9 describes how the respondents believe their future fertility would be affected by their human capital choices. The expected fertility by age 45 is 1.7 children for men and 1.8 children for women. This level of fertility is reasonable and quite similar to the actual average fertility rate in the United States (1.9 births per woman). We also see that expected fertility differs across the human capital choices: for example, the average number of children is higher for both genders if they graduate with a humanities/social science college degree versus a science/business degree.

As with the previous beliefs, we construct an individual-specific fertility "return" to human capital investment. These are reported in Panel B of Table 9. Men and women believe that completing a science or business degree rather than a degree in the humanities would reduce their expected number of children at age 30 by about 42 percent and 48 percent, respectively. In contrast, completing a college degree, relative to no degree, is believed to nearly double their expected number of children. The beliefs of the students in our sample are consistent with the idea that children are a normal good, and the anticipated loss of income from not completing a college degree reduces the demand for children. However, this belief runs counter to the well-known pattern that fertility declines with income and education level. This pattern could also be due to a reduction in perceived marriage prospects if one were to not graduate versus complete a college degree (as can be seen in Table 6).

Comparing the beliefs at age 45 reveals that students anticipate a life-cycle pattern to their own fertility. By age 45, the differences in expected fertility by human capital (science/business versus humanities/social sciences) have reduced by about half. These findings suggest that men and women anticipate that human capital investments will have a much larger effect on the *timing* of fertility rather than the level of final fertility. Indeed, while women and men expect that majoring in a science or business field (relative to majoring in humanities/social sciences) would reduce their fertility by age 30 by 42-48%, their final fertility at age 45 – when we would expect fertility for both men and women to have largely been completed – is expected to be lower by only 14-18%.

²²Expected number of children is defined as $E(\text{children}) = 0*pr(\text{no children}) + 1*pr(1 \text{ child}) + 2*pr(2 \text{ children}) + 3*pr(3 \text{ or more children})$.

8 Beliefs about Future Labor Supply

We next examine beliefs about future labor supply conditional on human capital levels. Our survey elicited beliefs about future unconditional labor supply and labor supply conditional on being married. We ask for beliefs about the probability of working full-time (defined for the respondents as at least 35 hours per week and at least 45 weeks per year), part-time work, and no work. Because of time constraints, we were unable to ask questions to ascertain the reasons for the anticipated future employment state, in particular whether not working full- or part-time is anticipated to be the result of a voluntary or involuntary (layoffs/dismissals) decision. Our underlying assumption is that the students' beliefs about labor supply are in essence the product of the individual "integrating" over different possible states of the world, including voluntary quits and involuntary dismissals.²³

8.1 Unconditional Labor Supply

The top panel of Table 10 provides the average beliefs about the probability of working full-time, part-time, or not at all. The average male students' beliefs about the probability of working full time at age 30 is 0.83 if he completes a science or business degree, 0.76 if a humanities or social sciences degree, and 0.68 if he does not graduate. These are quite close to the actual realized full-time employment rates in the ACS national data (Table 2). Table 10 also reveals that the average female student's beliefs about the probability of working full-time is considerably lower than that of males. But, while the gender difference in beliefs and in the realized data is qualitatively similar, the average level of the female students' beliefs regarding labor supply are higher than the corresponding figures for older cohorts in the ACS. We cannot determine if this difference reflects the high-ability selection of our sample, incorrect beliefs by our sample, or that the college students accurately predict a new future trend in higher female labor supply.

As in all of the previous analysis, we leverage our conditional beliefs data to construct an individual-level log difference in beliefs by each type of human capital. The bottom panel of Table 10 indicates that students believe their human capital choice will substantially affect their future employment. On average, male and female students believe completing a science or business degree, rather than a humanities degree, will increase their probability of full-time employment at age 30 by 15 percent for males and 9 percent for females. Importantly, there are even larger anticipated effects on the part-time and no work margins. A science or business degree rather than a humanities degree is expected to reduce the probability of part-time employment at age 30 by 36 percent for males and 28 percent for females, and reduce the probability of not working at all by

²³In addition to employment state, we also elicited beliefs about expected number of work hours/week for current full-time male and female workers. These results are presented in Appendix Table A3. The largest differences are regarding hours worked in science/business versus humanities— at age 45, they are expected to be higher by about 6% (8%) for female (male) workers in science/business (versus humanities/social sciences).

37 percent for males and 10 percent for females. Turning to the college versus no degree margin, it is clear that students believe that labor supply will be much higher if they complete a college degree. This pattern is consistent with a strong substitution effect on labor supply given the high expected earnings premium to completing a college degree.

For male students, at age 45, the differences in expected labor supply by human capital type at many margins are now smaller in magnitude and not statistically different from zero. However, there remains evidence that female students still expect that their human capital choices will affect their future labor supply. Women expect that the probability they work part-time if they complete a science or business degree rather than a humanities or arts degree would be reduced by 14 percent. And female students anticipate that graduating from college would increase their probability of working full-time by 18 percent.

8.2 Labor Supply by Marital Status

Changes in labor supply behavior by marital state have been well documented (Killingsworth and Heckman, 1986; Eckstein and Lifshitz, 2011). We also elicited students' beliefs about their age 30 labor supply conditional on their future expected marital status.²⁴ These statistics are presented in Panel A of Table 11. The average male students' belief about the probability of working full time at age 30 (weighted across majors, shown in the overall row) varies little by marital status, 0.81 when married and 0.83 when single. On the other hand, the average female's belief about full-time work likelihood is significantly lower when married: 0.72 when married versus 0.82 when single. Women expect a higher probability of either working part time or not working at all when married versus when single. These beliefs could be the result of anticipated career interruptions due to fertility or childrearing (Bertrand et al., 2010; Bronson, 2015; Herr, 2015; Klevin, Landais, and Sogaard, 2015). We construct an individual-level log difference in (major-weighted) labor supply beliefs by marital status. These are reported in Panel B of Table 11. On average, female students believe that the probability of working full-time at age 30 will decrease by 18 percent when married versus single; the male estimate is zero. Female students expect the probability of working part-time and not working to nearly double when married.

Panel C of Table 11 shows statistics of the individual-level log difference in conditional labor supply beliefs by each type of human capital. We see that students perceive their future conditional employment to be significantly affected by their human capital choice. In addition, the average differences for females are larger when married than when single. For example, female students believe that completing a science or business degree rather than a humanities degree will increase

²⁴For example, labor supply beliefs conditional on marriage 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 are married and you received a Bachelor's degree in each of the following?"

their probability of full-time employment at age 30 by 50 percent when married, and by 36 percent when single. The extent to which these differences are driven by demand-side factors or supply-side factors (substitution/income effect in labor supply, fertility and childrearing etc.) is unclear.

9 Beliefs and Realized Outcomes: Results from a Follow-Up Survey

In 2016, we re-surveyed the respondents to our original 2010 survey and elicited respondents' current earnings, labor supply, and relationship status. One purpose for this follow-up is to provide some evidence on the "quality" of the expectations data: if the expectations data was purely "noise" we would not expect systematic relationships with future realizations. Another, perhaps more substantive, purpose for this data collection is to understand how well students can anticipate their future career and family lives. In summary, across a number of outcomes, we find a strong systematic relationship between the beliefs we elicit when the respondents were in school and the actual realizations six years later.

Of the original sample of 493 respondents, 274 participated in the follow-up, a response rate of 56 percent. The average respondent was 25 years old in the follow-up, with three-quarters working full-time, and about 5% married (but nearly half in a "relationship"). Appendix C provides additional information on the follow-up survey administration, descriptive statistics for the follow-up sample, and tests of the randomness of the follow-up response. Our follow-up sample appears to be a random subset of the original sample. We reject the joint hypothesis that a rich set of covariates available in 2010 for all respondents (including gender, race, parental education and income, GPA, SAT scores, and college major) are systematically related to responding to the follow-up survey at standard statistical levels (p-value of the joint F-test is 0.32).

For a number of outcomes, Table 12 displays summary statistics for expectations in 2010 and the equivalent realizations reported in 2016. We emphasize that at no point during the follow-up did we remind respondents of their previous survey answers, six years prior, and we provided no incentives for the respondents to provide answers which would match their prior beliefs. The expectations reported in this table are those weighted by the individual-specific probability of choosing each of the majors.²⁵

Panel A of Table 12 compares expected earnings in 2010 with actual earnings in 2016. We adjust 2016 realized earnings to be in 2010 dollars using the CPI-U series so as to match the expectations data in which respondents were instructed to ignore any expected price changes and

²⁵The conclusions are similar if instead we use the expectation reported (in the 2010 survey) for the major category the individual actually ends up graduating with in 2016 (for example, using the expectation reported conditional on graduating with a humanities major for an individual who actually graduated with that major).

report expectations in current (i.e., 2010 dollars).²⁶ The comparison of mean expected and realized earnings indicates that the two are remarkably similar: \$73,500 versus \$75,000. We cannot reject the hypothesis that the mean expectations are the same as the mean realizations at conventional levels, for both the full sample and for the gender sub-samples. In addition, the standard deviations of the expected and realized distributions are also quite similar. This comparison of beliefs and realized earnings is consistent with students having rational expectations regarding the distribution of their future earnings, 6 years later.

Panels B and C of Table 12 compare expectations and realizations for labor market status. In 2010, while in college, students believe they would be working full-time 77.5 percent of the time when they were age 30, which compares favorably to the actual full-time employment rate realized six years later of 74 percent. Female students however over-estimated the future frequency at which they would be working part-time– the average part-time rate (for age 30) was believed to be 18 percent, compared to the actual rate of 9 percent. The fact that our expectations data is for age 30 only (we do not have labor supply expectations for age 23) could be particularly key here as it could be the case that when our female respondents reach age 30, their part-time rate will be higher.

Turning to marriage (Panel D), we first note again that our sample is still quite young at the time of the 2016 follow-up survey, and only a small minority (about 5.5 percent) are actually married at the time of the survey. A higher fraction – almost half – report being in a “relationship” of some kind. However, it appears that the students’ expectations about marriage were considerably too optimistic. In 2010, the students believed, on average, that there would be a 16 percent chance they would be married by their first post-graduate year, still considerably higher than the realized marriage rate in 2016, which for most respondents is 3-5 years after graduation. Students’ expectations regarding being married are more reasonable when compared with the rate of marriage or being in any relationship. The age-weighted expectation of being married by 2016 was 34 percent for our sample, compared to a realized relationship rate of 48 percent.

Panels E and F show that expectations regarding spouse’s labor supply and earnings compare quite favorably with realizations, with the caveat that expectations are conditioned on marriages, and we examine realizations for spouses and non-married relationship partners. In Panel F, it is notable that female respondents’ average belief of spouse’s earnings is in fact lower than the realized average, while the opposite is the case for male respondents.

We next turn to the individual-level relationship between the expectations and actual realized

²⁶Recall that our 2010 survey elicited expectations for some outcomes at 3 ages (age 23, 30, and 45), and for other outcomes we did not elicit age 23 outcomes. To create maximum consistency between the expectations and realization data, we form age-weighted expectations using the current age of the respondent to linearly weight the age 23 and age 30 expectations data to form expectations for the respondent’s current age. For outcomes for which we do not have the earlier age 23 expectations, we use age 30 expectations.

outcomes. Our goal here is to test how closely individuals can predict their own future earnings, labor supply, and marriage outcomes. Each cell in Table 13 reports a regression of the realized outcome in 2016 on the expectation for that outcome as reported in 2010. Panel A of Table 13 reports the estimates of a regression of log realized earnings on log expected earnings. The positive and statistically significant coefficient indicates that earnings beliefs are systematically related to realized earnings in the expected way: students who expected higher future earnings do in fact report earning more (with a 1 percent higher expectation associated with realized earnings that are 0.39 percent higher). The R-square of the regression of 0.09 indicates that, although student expectations alone explain 9 percent of realized earnings, students cannot perfectly predict their future earnings. This finding is a reject of the hypothesis that students have “perfect foresight” regarding their own earnings. It is also interesting to note that women are able to predict their future earnings much better than men. This could be for several reasons, including men being overconfident in their expectations (Niederle and Vesterlund, 2007; Reuben, Wiswall, and Zafar, forthcoming) or possible gender differences in the “seriousness” with which respondents completed the original expectations survey.

Turning to the other non-pecuniary outcomes, we similarly find evidence of a relationship between expectations and actual realizations. For labor supply, expectations are strongly correlated with actual labor supply behavior (working full- or part- time) observed six years later, but only for the female sub-sample. Likewise, marriage expectations are significantly related to later relationship status. Beliefs about spouse’s labor supply are also predictive of spouse’s actual labor supply (with the relationship statistically significant for the full sample). Finally, expectations about future spousal earnings are also statistically significantly related to actual earnings of spouses/partners.²⁷

Overall, the follow-up survey reveals that expectations for a broad set of outcomes compare very favorably with realizations several years later. There is evidence of rational expectations for own earnings and to a more limited extent for own labor supply, but students have biased beliefs about their own future marriage prospects and future spousal earnings. Our finding that expectations tend to be predictive of actual outcomes at the individual level indicates that students can anticipate to some substantial degree future career and family outcomes, and that these ex ante expectations are quite sensible.

Finally, we investigate whether expected major is predictive of the actual major that individuals end up graduating with. Table 14 shows the mean probability of graduating with each of the majors as reported in 2010 for respondents grouped by their actual major reported in 2016. For individuals who graduate with humanities/social sciences, economics/business, or natural sciences, the mean probability of majoring in that field when surveyed in 2010 exceeds 70%, indicative of intended

²⁷Table C3 conducts the same analysis as in Table 13, with expectations constructed using the one major the individual graduates with. Results are generally similar.

major being predictive of actual major. For engineering graduates, the mean probabilities are more dispersed but still the highest for their own field. As expected, those who do not graduate did not ex ante anticipate that to be the case. The last column of the table shows that the vast majority of respondents graduate with the major to which they assign the highest probability in 2010.

10 Determinants of Human Capital Choices

The previous sections have presented extensive evidence that students believe their labor market returns, labor supply, fertility, as well as marriage market returns (likelihood of being married and spousal earnings) to be substantially affected by their human capital choices. The question then is whether these factors drive their education choices. A large and growing literature has investigated the determinants of college major choice (Freeman, 1971; Bamberger, 1986; Berger, 1988; Montmarquette, Cannings, and Mahseredjian, 2002; Arcidiacono, 2004; Beffy, Denis, and Maurel, 2012; Zafar, 2013; Gemici and Wiswall, 2014; Wiswall and Zafar, 2015a). All of these studies focus primarily on earnings and ability as possible determinants of the choice- the only exception is Wiswall and Zafar (2015a), who estimate a model that also incorporates utility from spousal consumption. However, even in this model, the individual derives no flow utility from marriage or fertility. Here, we estimate a stylized model of college major choice. Our main goal is to document the extent to which human capital choices are impacted by these dimensions, so we conclude with a simple decomposition exercise.

We assume that utility from a human capital choice is a function of a vector of major-specific characteristics as perceived by the student, including own earnings, spousal earnings, marriage, and fertility in this vector. We also allow one's perceived ability (specific to the human capital choice) to impact the utility from the choice- these beliefs were also elicited from the students and are discussed in Appendix B. Prior sections have shown that perceived differences across human capital choices are particularly stark at earlier ages, and hence we allow the flow utility from marriage and children to vary over the life-cycle with separate parameters for both marriage and fertility early in the life-cycle (age 23 to 35) and later in the life-cycle. The model is estimated separately for male and female students, allowing parameters to freely vary by gender.²⁸ Appendix A provides details on the variable definitions and estimation. Here we briefly summarize the most salient model estimates: (i) discounted own lifetime earnings and ability rank are estimated to be positive and significant determinants of education choice for both genders, (ii) spousal earnings are a significant determinant, but for female students only, and (iii) children are significant and

²⁸We use the reported probability of choosing the different fields of study when estimating the model. However, since we observe the actual major the student graduates with (in the follow-up survey taken by a subset of the original sample, that is discussed in the previous section), we can use the actual major to estimate the model as well. We discuss this in Appendix A. Results are qualitatively similar when using actual major.

positive for females only, but only at younger ages.

We use the estimated model to decompose the human capital choices into the constituent components in order to assess the importance of each of these factors. Our decomposition procedure starts by creating a baseline where every field is equally likely. We accomplish this by setting each respondent's beliefs (about earnings, ability, marriage, spouse's earnings, and fertility) for each field equal to the corresponding level for the humanities major. Therefore, at the baseline, the likelihood of majoring in each of the three fields is 0.33. After establishing this baseline, we then progressively re-introduce each individual's major-specific beliefs into the estimated choice model in order to capture the marginal contribution of each component. The magnitude by which the probability of majoring in each field changes as we add a component measures the importance of this component.

The top panel of Table 15 reports the choice probability at each stage of the decomposition averaged over the male and female respondents, separately. Panel B reports the percent change in the average choice probability relative to the previous column. Re-introducing each individual's beliefs about her own earnings in each major increases the average probability of majoring in science/business for both genders. This reflects the relative earnings advantage most individuals perceive from graduating with a science/business degree. Adding earnings beliefs also reduces the average likelihood of choosing the no degree option, given lower expected earnings from that option. In comparing the male and female responses, it is clear that own earnings differences are a smaller factor in educational choices for women than men.

Column (3) adds the beliefs about own ability. We see that it only slightly affects the average probabilities for males. For females, however, there is a sizable increase in the probability of choosing humanities/social sciences, reflecting higher perceived ability in humanities relative to other majors. Column (4) indicates the marginal contribution of beliefs about gains in the marriage market from choosing different majors is small. Re-introducing beliefs about spouse's earnings, in column (5) shows a large change for females, for whom there is a sizable decrease in the probability of choosing the no degree option, and an increase in the likelihood of the choosing the science/business field. This reflects the high levels of perceived marital assortativeness by education, and the higher (lower) relative earnings expected in science/business (no degree).

Finally, column (6) shows that introduction of fertility beliefs has a large impact on the average choices for females, for whom the average probability of choosing humanities increases by nearly 22%, and the average probability of choosing the non-humanities fields declines. For male respondents, there is also an increase in the average humanities probability and a decline in choosing to not graduate, but the changes are much smaller than those for females.

Overall, the results indicate that own earnings is a significant determinant of major choice for both genders, in particular for males, driving students towards the higher-paying economics/business

field. Marriage market returns – spouse’s earnings – have sizable impacts on the choices of females, pushing them away from not graduating and into economics/business. Beliefs about ability and, in particular, fertility push women away from science/business fields and into humanities. The impact of fertility beliefs is quite interesting- it could be a result of females perceiving careers in humanities/social sciences as being more accommodating of childrearing. While the literature has found that a large part of the gender gap can be attributed to the dynamic effects of having children on labor market outcomes (Bertrand et al., 2010; Klevorick et al., 2015), our findings indicate that individuals not only anticipate the life-cycle effects of fertility to differ across human capital choices (Section 7), but also adjust their human capital choices accordingly.

11 Conclusion

What do college students actually believe about the consequences of their education choices? We surveyed undergraduate students at New York University (NYU) about their perceptions of the career and family consequences of their educational choices, including earnings, labor supply, likelihood of marriage, characteristics of their future spouses, and the number of children they expected. Our rich data on beliefs allows us to determine the aspects that matter most to college students when making these important human capital decisions. We find that students believe there are not only important consequences for their own earnings, but for their future family life as well. Women in particular believe that science or business majors would raise their own earnings but also lead to lower rates and delay of marriage and fertility. Men and women both perceive a marriage market “return” to higher earning college majors and to completing a college degree at all. Our finding that these ex ante beliefs are systematically related to actual educational choices and to later life realized outcomes points to the key importance of both career and family concerns to the decision-making of young people.

References

- [1] Altonji, Joseph. 1993. The Demand for and Return to Education when Education Outcomes are Uncertain. *Journal of Labor Economics*, 11: 48-83.
- [2] Altonji, Joseph, Peter Arcidiacono, and Arnaud Maurel. 2015. The Analysis of Field Choice in College and Graduate School: Determinants and Wage Effects. *Handbook of the Economics of Education Volume 5*, Chapter 7.
- [3] Arcidiacono, Peter. 2004. Ability Sorting and the Returns to College Major. *Journal of Econometrics*, 121(1-2): 343-375.
- [4] Arcidiacono, Peter, Joseph Hotz, and Songman Kang. 2012. Modeling College Major Choices using Elicited Measures of Expectations and Counterfactuals. *Journal of Econometrics*, 166(1): 3-16.
- [5] Arcidiacono, Peter, Joseph Hotz, Arnaud Maurel, and Teresa Romano. 2015. Recovering Ex Ante Returns and Preferences for Occupations Using Subjective Expectations Data. Working Paper.
- [6] Angrist, Joshua, and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.
- [7] Attanasio, Orazio, and Katja Kaufmann. 2011. Education Choices and Returns on the Labor and Marriage Markets: Evidence from Data on Subjective Expectations. Working Paper.
- [8] Bamberger, Gustavo. 1986. "Occupation Choice: The Role of Undergraduate Education." Ph.D. Dissertation, University of Chicago.
- [9] Becker, Gary. 1962. Investment in Human Capital: A Theoretical Analysis. *Journal of Political Economy*, 70(5): 9-49.
- [10] Becker, Gary, and H. Lewis. 1973. On the Interaction between the Quantity and Quality of Children. *Journal of Political Economy*, 81(2): 279-288.
- [11] Beffy, Magali, Denis Fougere, and Arnaud Maurel. 2012. "Choosing the Field of Study in Post-Secondary Education: Do Expected Earnings Matter?" *The Review of Economics and Statistics*, 94 (1): 334-347.
- [12] Ben-Porath, Yoram. 1967. The Production of Human Capital and the Life Cycle of Earnings. *Journal of Political Economy*, 75(4): 352-265.
- [13] Berger, Mark. 1988. Predicted Future Earnings and Choice of College Major. *Industrial and Labor Relations Review*, 41(3): 418-29.
- [14] Bertrand, Marianne, Claudia Goldin, and Larry Katz. 2010. Dynamics of the gender gap for young professionals in the financial and corporate sectors. *American Economic Journal: Applied Economics*, 2:228-255.
- [15] Blass, Asher, Saul Lach, and Charles Manski. 2010. "Using Elicited Choice Probabilities to Estimate Random Utility Models: Preferences for Electricity Reliability." *International Economic Review*, 51(2): 421-440.
- [16] Bronson, Mary Ann. 2015. "Degrees are Forever: Marriage, Educational Investment, and Lifecycle Labor Decisions of Men and Women." Working Paper.
- [17] Chiappori, Pierre-Andre, Murat Iyigun, and Yoram Weiss. 2009. Investment in Schooling and the Marriage Market. *American Economic Review*, 99(5): 1689-1713.

- [18] Chiappori, Pierre-Andre, Bernard Salanie, and Yoram Weiss. 2015. Partner Choice and the Marital College Premium: Analyzing Marital Patterns Over Several Decades. Working Paper.
- [19] Cortes, Patricia, and Jessica Pan. 2016. Prevalence of Long Hours and Women's Job Choices: Evidence across Countries and within the U.S. Working Paper
- [20] Cunha, Flavio, James Heckman and Salvador Navarro. 2005. Separating Uncertainty from Heterogeneity in Life Cycle Earnings. *Oxford Economic Papers*, 57(2): 191-261.
- [21] Delavande, Adeline, and Basit Zafar. 2015. University Choice: The Role of Expected Earnings, Non-pecuniary Outcomes and Financial Constraints. Working Paper.
- [22] Dey, Judy, and Catherine Hill. 2007. Behind the Pay Gap. American Association of University Women Educational Foundation Research Report.
- [23] Eckstein, Zvi, and Osnat Lifshitz. 2011. Dynamic Female Labor Supply. *Econometrica*, 79(6): 1675-1726.
- [24] Eide, Eric., and Geetha Waehrer. 1998. The Role of the Option Value of College Attendance in College Major Choice. *Economics of Education Review*, 17(1): 73-82.
- [25] Eika, Lasse, Magne Mogstad, and Basit Zafar. 2015. Educational Assortative Mating and Household Income Inequality. Working Paper.
- [26] Flabbi, Luca, and Andrea Moro. 2012. The effect of job flexibility on female labor market outcomes: Estimates from a search and bargaining model. *Journal of Econometrics*, 168: 81-95.
- [27] Flyer, Fredrick. 1997. The Influence of Higher Moments of Earnings Distributions on Career Decisions. *Journal of Labor Economics*, 15(4): 689-713.
- [28] Gemici, Ahu, and Matthew Wiswall. 2014. "Evolution of Gender Differences in Post-Secondary Human Capital Investments: College Majors at the Intensive Margin." *International Economic Review*, 55 (1): 23-56.
- [29] Ge, Suqin. 2011. Women's College Decision: How Much Does Marriage Matter? *Journal of Labor Economics*, 29(4): 773-818.
- [30] Giustinelli, Pamela. 2015. Group Decision Making with Uncertain Outcomes: Unpacking Child-Parent Choice of the High School Track. *International Economic Review*, forthcoming.
- [31] Goldin, Claudia. 1997. Career and family: College women look to the past. In *Gender and Family Issues in the Workplace*, edited by F. Blau and R. Ehrenberg. New York: Russell Sage Press: 20-58.
- [32] Goldin, Claudia. 2014. A Grand Gender Convergence: Its Last Chapter. *American Economic Review*, 104(4): 1091-1119.
- [33] Goldin, Claudia, and Lawrence Katz. 2011. The Cost of Workplace Flexibility for High-Powered Professionals. *The Annals of the American Academy of Political and Social Science*, 638(1):45-67.
- [34] Goldin, Claudia, and Lawrence Katz. 2015. A Most Egalitarian Profession: Pharmacy and the Evolution of a Family-Friendly Occupation. *Journal of Labor Economics*, forthcoming.
- [35] Greenwood, J., N. Guner, G. Kocharkov, and C. Santos. 2014. Marry Your Like: Assortative Mating and Income Inequality. *American Economic Review, Papers and Proceedings*, 104(5): 348-53.

- [36] Heckman, James, and Edward Vytlacil. 2005. Structural Equations, Treatment Effects, and Econometric Policy Evaluation. *Econometrica*, 73(3): 669-738.
- [37] Heckman, James, Lance Lochner and Petra Todd. 2008. Earnings Functions and Rates of Return. *Journal of Human Capital*, 2(1): 1-31.
- [38] Herr, Jane. 2015. The Labor Supply Effects of Delayed First Birth. *American Economic Review: Papers & Proceedings*, 105(5): 630-637.
- [39] Imbens, Guido, and Donald Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences- An Introduction*. Cambridge University Press.
- [40] Iyigun, Murat, and Randall Walsh. 2007. Building the Family Nest: Premarital Investments, Marriage Markets, and Spousal Allocations. *Review of Economic Studies*, 74(2): 507-535.
- [41] Jacob, Brian, Brian McCall, and Kevin Stange. 2014. "The Consumption Value of Higher Education." Working Paper.
- [42] Jones, Larry, Alice Schoonbroodt, and Michèle Tertilt. 2010. Fertility Theories: Can They Explain the Negative Fertility-Income Relationship? in *Demography and the Economy*, 43-100, edited by John Shoven, University of Chicago Press.
- [43] Kaufmann, Katja. 2015. Understanding the Income Gradient in College Attendance in Mexico: The Role of Heterogeneity in Expected Returns to College. *Quantitative Economics*, forthcoming.
- [44] Kaufmann, Katja, Matthias Messner and Alex Solis. 2015. Elite Higher Education, the Marriage Market and the Intergenerational Transmission of Human Capital. Working Paper.
- [45] Killingsworth, Mark, and James Heckman. 1986. Female Labor Supply: A Survey. In *Handbook of Labor Economics*, 103-204, edited by O. Ashenfelter and R. Layard, Volume 1, Chapter 2. Elsevier Science Publishers.
- [46] Klevin, Henrik, Camille Landais, and Jakob Sogaard. 2015. Children and Gender Inequality: Evidence from Denmark. Working Paper.
- [47] Lafortune, Jeanne. 2013. Making Yourself Attractive: Pre-Marital Investments and the Returns to Education in the Marriage Market. *American Economic Journal: Applied Economics*, 5(2): 151-178.
- [48] Manski, Charles. 2004. Measuring Expectations. *Econometrica*, 72(5): 1329-1376.
- [49] Mare, Robert. 1991. Five Decades of Educational Assortative Mating. *American Sociological Review*, 56(1): 15-32.
- [50] Montmarquette, Claude, Kathy Cannings, and Sophie Mahseredjian. 2002 "How Do Young People Choose College Majors?" *Economics of Education Review*, 21(6): 543-556.
- [51] Niederle, Muriel, and Lise Vesterlund. 2007. "Do Women Shy away from Competition? Do Men Compete too Much?" *Quarterly Journal of Economics*, 122(3): 1067-1101.
- [52] Nielsen, Helena, and Annette Vissing-Jorgensen. 2006. "The Impact of Labor Income Risk on Educational Choices: Estimates and Implied Risk Aversion." Working Paper.
- [53] Pencavel, John. 1998. Assortative Mating by Schooling and the Work Behavior of Wives and Husbands. *American Economic Review*, 88(2): 326-329.

- [54] Reuben, Ernesto, Matthew Wiswall, and Basit Zafar. Forthcoming. Preferences and Biases in Educational Choices and Labor Market Expectations: Shrinking the Black Box of Gender. *Economic Journal*, forthcoming.
- [55] Saks, Raven, and Steven Shore. 2005. "Risk and Career Choice." *Advances in Economic Analysis & Policy*, 5(1).
- [56] Stinebrickner, Todd, and Ralph Stinebrickner. 2012. Learning about Academic Ability and the College Drop-out Decision. *Journal of Labor Economics*, 30(4): 707-748.
- [57] Stinebrickner, Todd, and Ralph Stinebrickner. 2014. "Math or Science? Using Longitudinal Expectations Data to Examine the Process of Choosing a College Major." *Review of Economic Studies*, 81(1): 426-472.
- [58] Turner, Sarah, and William Bowen. 1999. Choice of Major: The Changing (Unchanging) Gender Gap. *Industrial and Labor Relations Review*, 52(2): 289-313.
- [59] Wasserman, Melanie. 2015. "Hours Constraints, Occupational Choice and Fertility: Evidence from Medical Residents." Working Paper.
- [60] Weinberger, Catherine. 2004. "Just Ask! Why Surveyed Women Did Not Pursue Information Technology Courses or Careers?" *IEEE Technology and Society*, 23(2): 28-35.
- [61] Wiswall, Matthew, and Basit Zafar. 2015a. Determinants of College Major Choice: Identification Using an Information Experiment. *Review of Economic Studies*. 82(2): 791-824.
- [62] Wiswall, Matthew, and Basit Zafar. 2015b. How do College Students Respond to Public Information about Earnings? *Journal of Human Capital*, 9(2): 117-169.
- [63] Wiswall, Matthew, and Basit Zafar. 2016. Preferences for the Workplace, Investment in Human Capital, and Gender. Working Paper.
- [64] Zafar, Basit. 2013. College Major Choice and the Gender Gap. *Journal of Human Resources*, 48(3): 545-595.

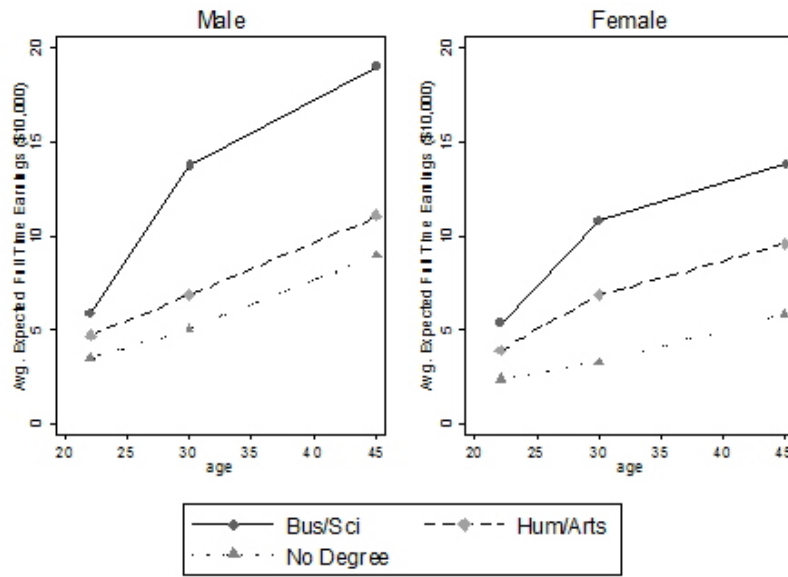


Figure 1: Growth in Self Earnings, by Major/Gender.

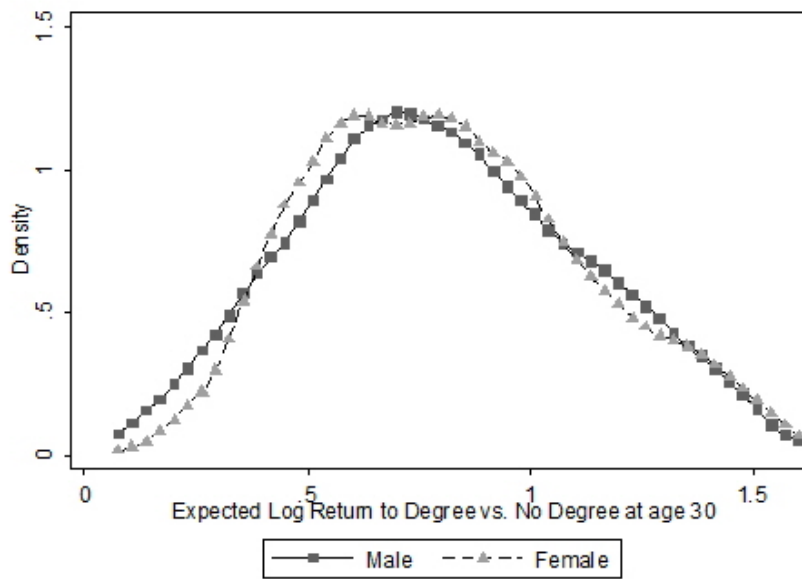


Figure 2: Distributions of Log Difference in Age 30 Earnings for Graduating versus No Degree.

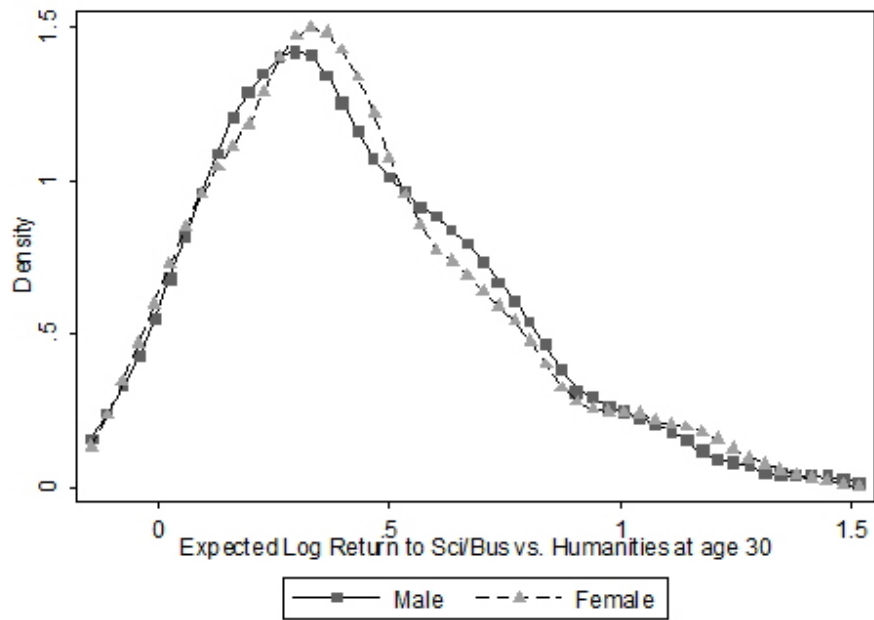


Figure 3: Distributions of Log Difference in Age 30 Earnings for Science/Business Degree versus Humanities Degree.

Table 1: Sample Characteristics

	Male	Female
Observations	176	317
Age	20.2 (1.37)	20.1 (1.04)
% Asian	44.32	44.48
% White	42.05	35.65
Ability measures:		
GPA	3.47 (.4)	3.46 (.38)
SAT Math	716.41 (82.64)	699.9 ⁺ (113.3)
SAT Verbal	682.34 (77.61)	692.9 (112.5)
Parent's Characteristics:		
% Father College-educated	73.71	76.38
% Mother College-educated	74.43	69.11
Parents' Income	152180 (125411)	144741 (120565)
School year:		
% Freshman	46.02	37.22 ⁺
% Sophomore	32.95	38.17
% Junior	21.02	24.61
Major:		
% Economics/Business	42.05	23.66 ⁺⁺⁺
% Humanities/Soc. Sciences	34.09	56.15 ⁺⁺⁺
% Engineering	5.68	3.79
% Natural Sciences/Math	18.18	16.4

Mean (Std Deviation) reported for the continuous outcomes.
⁺⁺⁺, ⁺⁺, ⁺ denote that the statistics differ by gender at the 1,
5, and 10% levels, respectively.

Table 2: Descriptive Statistics of 2009 ACS Data

	Age 23		Age 30		Age 45	
	Male	Female	Male	Female	Male	Female
Earnings (in \$10,000s)						
Science/Business	3.33 (2.15)	3.22 (2.19)	6.74 (4.81)	5.48+++ (3.15)	11.61 (9.79)	7.46+++ (6.49)
Humanities	2.51 (1.33)	2.57 (1.88)	5.40 (4.20)	4.47+++ (2.71)	9.07 (8.48)	5.93+++ (5.67)
No Degree	2.54 (1.52)	2.15+++ (1.41)	4.21 (2.50)	3.08+++ (1.59)	5.70 (4.13)	3.88+++ (2.57)
p-value ^a	0	0	0	0	0	0
Spousal Earnings (in \$10,000s)						
Science/Business	3.41 (2.09)	4.75+++ (3.11)	5.26 (3.44)	8.25+++ (5.79)	7.44 (6.69)	12.68+++ (10.15)
Humanities	2.27 (1.33)	3.49+++ (1.93)	4.30 (2.61)	6.66+++ (5.64)	5.71 (4.72)	9.85+++ (9.42)
No Degree	2.21 (1.13)	3.50+++ (1.93)	3.24 (1.86)	4.82+++ (2.92)	3.76 (2.59)	6.36+++ (4.81)
p-value	0	0.003	0	0	0	0
Full-time Employed (%)						
Science/Business	38.5	42.4+++	80.86	64.40+++	82.68	58.42+++
Humanities	30.9	36.2+++	72.96	57.92+++	75.86	52.07+++
No Degree	40.1	34.4+++	66.53	46.51+++	67.88	52.44+++
p-value	0	0	0	0	0	0
Married (%)						
Science/Business	8.2	15.9+++	61.72	67.49+++	80.79	76.14+++
Humanities	11.5	15.3+++	55.7	64.94+++	76.58	74.51+
No Degree	15.2	26.4+++	54.86	59.29+++	69.3	69.65
p-value	0	0	0	0	0	0

Earnings and spousal earnings shown in \$10,000s.

Mean (standard deviation) shown for the continuous outcomes.

+++, ++, + gender differences statistically significant at the 1, 5, and 10% levels, respectively. Symbols denoted on female column.

^a p-value of a F-test of the joint equality of means across majors. p-value of zero implies p-value < 0.001.

Table 3: Self Earnings

	Age 23		Age 30		Age 45	
	Male	Female	Male	Female	Male	Female
Panel A: Levels (in 10,000s of dollars)						
Science/Business	5.93 (7.32)	5.39 (4.66)	13.74 (16.61)	10.86++ (9.31)	19.00 (22.38)	13.81+++ (14.12)
Humanities	4.71 (7.38)	3.94 (3.51)	6.87 (5.51)	6.86 (7.4)	11.03 (13.53)	9.60 (11.75)
No Degree	3.50 (7.54)	2.45++ (1.16)	5.07 (11.0)	3.27++ (4.56)	8.97 (15.95)	5.86+++ (10.22)
Overall	5.60 (7.36)	4.68+ (3.81)	12.95 (16.35)	9.21+++ (8.45)	18.44 (22.52)	12.33+++ (13.90)
Panel B: Individual Log Differences						
Sci/Business versus. Humanities	.267*** (.019)	.304*** (.017)	.523*** (.048)	.425***++ (.025)	.446*** (.051)	.347***+ (.026)
Graduate versus. No Degree	.594*** (.033)	.642*** (.026)	1.022*** (.055)	1.038*** (.037)	.829*** (.054)	.833*** (.038)

Panel A shows the mean and standard deviations of expected earnings (in \$10,000s). +++, ++, + denote gender differences are statistically different at the 1, 5, and 10% levels, respectively.

Panel B shows the avg. log differences and standard deviations in parentheses. ***, **, * denote the means are statistically different from zero at the 1, 5, and 10% levels, respectively. +++, ++, + (shown on the female column) denote gender differences are statistically different at the 1, 5, and 10% levels, respectively.

Table 4: Earnings growth beliefs

	Age 23-30		Age 30-45	
	Male	Female	Male	Female
Panel A: Levels (in %)				
Science/Business	.67 (.72)	.63 (.65)	.25 (.47)	.19 (.54)
Humanities	.41 (.56)	.51+ (.53)	.32 (.45)	.27 (.52)
No Degree	.23 (.78)	.21 (.55)	.47 (.74)	.43 (.58)
Overall	.66 (.73)	.6 (.58)	.29 (.48)	.23 (.52)
Panel B: Individual differences				
Sci/Business versus. Humanities	.26*** (.05)	.12***+++ (.03)	-.08* (.04)	-.08*** (.03)
Graduate versus. No Degree	.42*** (.06)	.39*** (.03)	-.19*** (.06)	-.2*** (.03)

Panel A shows the mean and standard dev of beliefs about earnings growth (in %). +++, ++, + denote gender differences are statistically different at the 1, 5, and 10% levels, respectively.

Panel B shows average log differences and standard deviations in parentheses. ***, **, * denote means are statistically different from zero at the 1, 5, and 10% levels, respectively. +++, ++, + (shown on the female column) denote gender differences are statistically different at the 1, 5, and 10% levels, respectively.

Table 5: Age 30 Earnings Uncertainty - Std deviations from fitting a Beta Distribution

	Male	Female
Panel A: Levels (in \$10,000)		
Science/Business	9.17 (1.44)	9.49 (2.48)
Humanities	10.34 (27.44)	10.01 (2.32)
No Degree	14.73 (7.34)	15.27 (7.53)
Overall	9.71 (2.02)	9.68 (2.01)
Panel B: Individual differences		
Sci/Business versus. Humanities	-.11*** (.014)	-.057***+++ (.012)
Graduate versus. No Degree	-.305*** (.052)	-.335*** (.043)

Panel A shows the mean and std dev of age 30 earnings uncertainty beliefs (in \$10,000). Uncertainty is the standard deviation of the individual-specific (beta-) fitted earnings distribution.
 +++, ++, + denote gender differences statistically different at the 1, 5, and 10% levels, respectively.

Panel B shows average log differences and standard deviations in parentheses. ***, **, * denote means are statistically diff from 0 at the 1, 5, and 10% levels, respectively. +++, ++, + (shown on female column) denote gender differences are statistically different at the 1, 5, and 10% levels, respectively.

Table 6: Beliefs about Marriage

Prob Marriage:	Age 23		Age 30		Age 45	
	Male	Female	Male	Female	Male	Female
Panel A: Levels (0-1 scale)						
Science/Business	.148 (.207)	.167 (.214)	.593 (.286)	.594 (.271)	.804 (.248)	.778 (.253)
Humanities	.152 (.214)	.182 (.229)	.601 (.291)	.66++ (.268)	.797 (.253)	.800 (.246)
No Degree	.153 (.219)	.221+++ (.26)	.535 (.329)	.605++ (.29)	.727 (.302)	.743 (.287)
Overall	.149 (.213)	.179 (.225)	.589 (.288)	.634+ (.266)	.797 (.25)	.793 (.242)
Panel B: Individual Log Differences						
Sci/Business versus. Humanities	-.008 (.046)	-.096* (.053)	-.024 (.042)	-.147***++ (.039)	.013 (.014)	-.020 (.024)
Graduate versus. No Degree	.075 (.099)	-.192**+ (.091)	.354*** (.11)	.127***++ (.054)	.317*** (.09)	.161*** (.054)

Panel A shows the mean and standard deviations of marriage beliefs. +++, ++, + denote gender diffs are statistically significant at the 1, 5, and 10% levels, respectively.

Panel B shows the average log differences and standard deviations in parentheses.
 ***, **, * denote the means are statistically different from zero at the 1, 5, and 10% levels, respectively. +++, ++, + (shown on the female column) denote gender differences are statistically significant at the 1, 5, and 10% levels, respectively.

Table 7: Beliefs about Potential Spousal Earnings, Conditional on Own Major (and Own Age)

	Age 23		Age 30		Age 45	
	Male	Female	Male	Female	Male	Female
Panel A: Levels (in 10,000s of dollars)						
Science/Business	5.06 (4.12)	5.74+ (3.92)	9.00 (7.72)	10.76++ (9.14)	11.29 (13.25)	13.68+ (13.67)
Humanities	4.52 (7.35)	4.75 (3.75)	7.05 (8.93)	7.86 (7.69)	8.02 (7.95)	11.07+++ (12.90)
No Degree	4.58 (11.99)	3.46 (2.26)	4.57 (5.56)	5.54 (9.11)	6.25 (9.89)	7.76 (12.03)
Overall	5.02 (5.90)	5.30 (3.88)	8.42 (7.60)	9.74+ (8.91)	10.77 (13.20)	12.73 (13.61)
Panel B: Individual Log Differences						
Sci/Business versus. Humanities	.185*** (.019)	.198*** (.015)	.282*** (.044)	.292*** (.024)	.241*** (.04)	.221*** (.026)
Graduate versus. No Degree	.432*** (.048)	.481*** (.028)	.687*** (.05)	.741*** (.041)	.587*** (.054)	.632*** (.039)

Panel A shows the mean and standard dev of beliefs about spouse's expected earnings (in \$10,000s) conditional on own major.

+++, ++, + denote gender differences are statistically different at the 1, 5, and 10% levels, respectively.

Panel B shows avg. log differences and standard deviations in parentheses. ***, **, * denote means are statistically different from zero at the 1, 5, and 10% levels, respectively. +++, ++, + (shown on the female column) denote gender differences are statistically different at the 1, 5, and 10% levels, respectively.

Table 8: Spousal Sorting Beliefs

	Spouse Major			Spouse Major		
	Sci/Business	Hum.	< College	Sci/Business	Hum.	< College
	Female Respondents			Male Respondents		
Own Major:						
Sci/Business	78.9% (14.4%)	16.6% (12.2%)	4.5% (6.5%)	71.6% (16.6%)	22.2% (15.0%)	6.2% (7.9%)
Humanities	59.8% (23.0%)	35.2% (22.8%)	5% (7.4%)	54.7% (23.0%)	37.5% (23.0%)	7.8% (9.2%)
No Degree	52.9% (26.0%)	19.3% (14.2%)	27.8% (27.8%)	48.3% (26.7%)	18.7% (14.2%)	32.9% (29.2)

Mean beliefs in first cell. Standard deviations in parentheses.

Table 9: Beliefs about Future Fertility

	Age 30		Age 45	
	Male	Female	Male	Female
Panel A: Average Expected Fertility				
Science/Business	1.21 (.50)	1.30+ (.509)	1.60 (.476)	1.66 (.481)
Humanities	1.79 (.519)	1.88+ (.478)	1.85 (.466)	1.92 (.481)
No Degree	.957 (.614)	.972 (.639)	1.47 (.568)	1.50 (.596)
Overall	1.37 (.52)	1.57+++ (.521)	1.67 (.455)	1.78++ (.464)
Panel B: Individual Log Differences				
Sci/Business versus. Humanities	-.418*** (.095)	-.483*** (.057)	-.140* (.068)	-.180*** (.045)
Graduate versus. No Degree	.733*** (.108)	.879*** (.086)	.265** (.077)	.331** (.058)

Panel A shows the mean and standard dev of beliefs about expected fertility.

+++ , ++ , + denote gender differences are statistically different at the 1, 5, and 10% levels, respectively.

Panel B shows average log differences and standard deviations in parentheses.

***, **, * denote means are statistically different from zero at the 1, 5, and 10% levels, respectively. +++ , ++ , + (shown on the female column) denote gender differences are statistically different at the 1, 5, and 10% levels, respectively.

Table 10: Beliefs about Labor Supply:

	Prob Work FT		Prob Work PT		Prob Not Working	
	Male	Female	Male	Female	Male	Female
Panel A						
Levels- Age 30						
Science/Business	0.84 (0.16)	0.77+++ (0.18)	0.12 (0.12)	0.17+++ (0.13)	0.05 (0.07)	0.07++ (0.09)
Humanities	0.76 (0.21)	0.71+++ (0.21)	0.16 (0.14)	0.22+++ (0.16)	0.07 (0.13)	0.08 (0.10)
No Degree	0.68 (.27)	0.62++ (0.29)	0.20 (0.18)	0.23 (0.19)	0.12 (0.16)	0.15+ (0.20)
Overall	0.82 (0.18)	0.76+++ (0.18)	0.13 (0.12)	0.18+++ (0.14)	0.06 (0.10)	0.06 (0.08)
Levels- Age 45						
Science/Business	0.79 (0.22)	0.77 (0.20)	0.15 (0.16)	0.17 (0.16)	0.06 (0.10)	0.06 (0.09)
Humanities	0.78 (0.22)	0.75 (0.20)	0.16 (0.16)	0.18+ (0.15)	0.06 (0.10)	0.07 (0.10)
No Degree	0.76 (0.24)	0.71+ (0.26)	0.17 (0.17)	0.20 (0.19)	0.07 (0.11)	0.09 (0.15)
Overall	0.79 (0.22)	0.77 (0.19)	0.15 (0.16)	0.17 (0.15)	0.06 (0.11)	0.06 (0.09)
Panel B						
Individual Log Differences- Age 30						
Sci/Business versus. Humanities	.152*** (.039)	.092*** (.026)	-.363*** (.132)	-.276*** (.06)	-.381*** (.1)	-.102++ (.07)
Graduate versus. No Degree	.337*** (.07)	.414*** (.057)	-.441*** (.169)	-.018++ (.106)	-.793*** (.134)	-.597*** (.108)
Individual Log Differences- Age 45						
Sci/Business versus. Humanities	-.016 (.044)	.022* (.012)	-.078 (.12)	-.141** (.061)	-.062 (.09)	-.020 (.059)
Graduate versus. No Degree	.032 (.047)	.181***++ (.04)	-.003 (.153)	.122 (.109)	-.146 (.145)	-.205** (.104)

Panel A shows the mean and standard deviations of labor supply beliefs (FT: full-time; PT: part-time; NW: not working), on a 0-1 scale. +++, ++, + denote gender diffs are statistically significant at the 1, 5, and 10% levels, respectively. Panel B shows the average log differences and standard deviations in parentheses. ***, **, * denote the means are statistically different from zero at the 1, 5, and 10% levels, respectively. +++, ++, + (shown on the female column) denote gender differences are statistically significant at the 1, 5, and 10% levels, respectively.

Table 11: Beliefs about Labor Supply Conditional on Marriage

	FT		PT		NW	
	Married	Single	Married	Single	Married	Single
Panel A: Average Expected Probability						
Males						
Science/Business	.82 (.18)	.84 (.18)	.13 (.12)	.11 (.12)	.05 (.09)	.05 (.10)
Humanities	.76 (.21)	.77 (.23)	.18 (.16)	.16 (.17)	.07 (.10)	.07 (.14)
No Degree	.68 (.27)	.70 (.28)	.21 (.20)	.19 (.19)	.11 (.15)	.11 (.16)
Overall	.81 (.18)	.83 (.19)	.14 (.13)	.11 (.12)	.06 (.09)	.06 (.13)
Females						
Science/Business	.72+++ (.21)	.83 (.17)	.20+++ (.15)	.12 (.11)	.08++ (.11)	.05 (.09)
Humanities	.66+++ (.23)	.78 (.20)	.25+++ (.18)	.16 (.15)	.09+ (.11)	.06 (.10)
No Degree	.58+++ (.31)	.70 (.29)	.25++ (.20)	.20 (.20)	.17+++ (.22)	.10 (.17)
Overall	.72+++ (.21)	.82 (.17)	.21+++ (.16)	.13 (.12)	.074+++ (.09)	.05 (.08)
Panel B: Individual log differences- married v single						
	FT	PT	NW			
Males	0 (.06)	.47*** (.11)	.19** (.08)			
Females	-.18***+++ (.02)	1.02***+++ (.1)	.75***+++ (.09)			
Panel C: Individual Log Differences-between majors						
Males						
Sci/Business versus. Hum.	.09** (.04)	.17*** (.05)	-.32** (.13)	-.31** (.13)	-.31*** (.1)	-.31*** (.1)
Graduate versus. No Degree	.32*** (.09)	.3*** (.06)	-.27 (.19)	-.52*** (.17)	-.70*** (.14)	-.64*** (.14)
Females						
Sci/Business versus. Hum.	.1*** (.03)	.07**+ (.03)	-.24*** (.06)	-.23*** (.07)	-.07 + (.08)	-.13** (.06)
Graduate versus. No Degree	.5*** (.07)	.36*** (.06)	.11 + (.11)	-.19*+ (.11)	-.62*** (.11)	-.36***+ (.10)

Panel A shows the mean and standard dev of expected probability of labor market status (FT, PT, NW).

+++ , ++ , + (on the female means) denote gender differences are statistically different at the 1, 5, and 10% levels, respectively.

Panel B shows average log differences (and standard deviations in parentheses) for married versus single.

Panel C shows average log differences (and standard deviations in parentheses) across majors.

For Panels B and C: ***, **, * denote means are statistically different from zero at the 1, 5, and 10% levels, respectively.

+++ , ++ , + (shown on the female means) denote gender differences are statistically different at the 1, 5, and 10% levels, respectively.

Table 12: Descriptive Statistics - Expectations (Weighted by Major Probs) and Outcomes

	Expectations in 2010			Realizations in 2016		
	All	Males	Females	All	Males	Females
Panel A: Earnings Full-time (age-weighted expectation)						
Mean	7.35	9.90	6.16	7.49	10.18	6.24
SD	(8.19)	(13.52)	(3.06)	(7.74)	(12.39)	(3.46)
N	201	64	137	201	64	137
Panel B: Likelihood of full-time employment (age 30 expectation)						
Mean	77.61	82.28	75.38	73.99	73.86	74.05
SD	(19.15)	(19.51)	(18.61)	(43.95)	(44.19)	(43.95)
N	273	88	185	273	88	185
Panel C: Likelihood of part-time employment (age 30 expectation)						
Mean	16.02***	11.71	18.08***	9.16	9.09	9.19
SD	(13.1)	(12.02)	(13.12)	(28.9)	(28.91)	(28.97)
N	273	88	185	273	88	185
Panel D: Likelihood of Marriage						
Using expectation for 1-yr after graduation (and marriage for outcomes)						
Mean	16.04***	13.62*	17.16***	5.56	8.14	4.35
SD	(21.62)	(19.83)	(22.37)	(22.95)	(27.5)	(20.45)
N	270	86	184	270	86	184
Using age-weighted expectation (and marriage + cohab. for outcomes)						
Mean	34.36***	31.35***	35.77***	48.15	45.35	49.46
SD	(21.08)	(21.97)	(20.56)	(50.06)	(50.08)	(50.13)
N	270	86	184	270	86	184
Panel E: Likelihood of partner working full-time (age 30 expectation)						
Mean	73.91	62.28	78.89	76.15	69.23	79.12
SD	(21.19)	(23.28)	(18.2)	(42.78)	(46.76)	(40.87)
N	130	39	91	130	39	91
Panel F: Partner's Earnings (age-weighted expectation)						
Mean	6.52*	6.84	6.4**	7.73	5.68	8.5
SD	(2.84)	(3.24)	(2.69)	(6.14)	(3.53)	(6.73)
N	99	27	72	99	27	72

Pairwise tests conducted for equality of the means of each of the expectations and realizations for the three subgroups (All; Male; Female).

***, **, * (denoted on the expectations cells) denote mean expectations significantly different from the mean outcome at the 1, 5, and 10% levels, respectively.

Table 13: The Link between Expectations and Outcomes

	All	Males	Females
Panel A, dependent variable: Log (current earnings)			
Log(Exp Earnings, Age Weighted)	0.386*** (0.131)	0.167 (0.207)	0.521*** (0.125)
Observations	201	64	137
R^2	0.092	0.018	0.153
Mean of Dependent Variable	10.99	11.18	10.90
Panel B, dependent variable: Employed Full-time			
Expected Prob of full-time emp at 30	0.165 (0.148)	-0.189 (0.220)	0.358* (0.187)
Observations	273	88	185
R^2	0.005	0.007	0.023
Mean of Dependent Variable	0.740	0.740	0.740
Panel C, dependent variable: Employed Part-time			
Expected Prob of part-time Emp at 30	0.272* (0.161)	0.0203 (0.263)	0.392** (0.196)
Observations	273	88	185
R^2	0.015	0.000	0.032
Mean of Dependent Variable	0.0900	0.0900	0.0900
Panel D, dependent variable: Married			
Age-Weighted Exp Probability of Being Married	0.217** (0.100)	0.378* (0.217)	0.147 (0.0936)
Observations	270	86	184
R^2	0.040	0.091	0.022
Mean of Dependent Variable	0.0600	0.0800	0.0400
Panel E, dependent variable: In Any relationship			
Age-Weighted Exp Probability of Being Married	0.503*** (0.127)	0.606*** (0.209)	0.441*** (0.161)
Observations	270	86	184
R^2	0.045	0.071	0.033
Mean of Dependent Variable	0.480	0.450	0.490
Panel F, dependent variable: Spouse/Partner Working Full-time			
Expected Prob of Spouse full-time Emp at 30	0.415** (0.183)	0.458 (0.298)	0.339 (0.253)
Observations	130	39	91
R^2	0.042	0.052	0.023
Mean of Dependent Variable	0.760	0.690	0.790
Panel G, dependent variable: Log(Spouse/Partner Earnings)			
Log(Age-Weighted Expected Earnings of Spouse)	0.400** (0.173)	0.598** (0.233)	0.344* (0.206)
Observations	112	31	81
R^2	0.054	0.119	0.042
Mean of Dependent Variable	1.690	1.420	1.790

Each cell in the table presents estimates from an OLS regression of the dependent variable on the 2010 expectation and a constant. Std errors reported in parentheses. ***, **, * denote sig. at 1%, 5%, and 10% levels, respectively.

Table 14: Actual and Intended Major Choice

	Number of Resp.	2010 Mean Prob of Majoring in:					% Majoring in Category Expecting to Graduate in it ^b
		Econ/ Bus.	Eng.	Hum/ Soc Sci	Natural Sci	Not Grad	
Actual Major (2016) ^a							
Economics/Business	72	77.1	3.4	8.1	10.7	1.1	80.6
Engineering	8	19.3	39.5	25.3	12.5	3.6	50.0
Humanities/Soc Sci	145	9.1	2.8	72.7	12.9	2.7	81.3
Natural Sciences	38	8.5	7.5	12.3	71.2	.73	84.2
Not Graduate	11	31.8	8.9	40.9	15.5	3.2	2.2

^a Actual Major is the major the respondent reports in the 2016 survey as having graduated in.

^b Percent of respondents who assign the highest probability to graduating in that major in 2010, and then go on to graduate with it.

Table 15: Determinants of Major Choice - Decomposition

	Baseline		Earnings		Ability		Marriage		Spouse Earnings		Fertility		Stated Prob	
	Male	Fem	Male	Fem	Male	Fem	Male	Fem	Male	Fem	Male	Fem	Male	Fem
	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)	(5a)	(5b)	(6a)	(6b)	(7a)	(7b)
Panel A: Average Probability														
Sci/Bus.	.333	.333	0.69	0.61	0.67	0.52	0.66	0.51	0.66	0.54	0.65	0.47	0.67	0.48
Humanities	.333	.333	0.23	0.29	0.24	0.37	0.24	0.38	0.24	0.36	0.25	0.46	0.30	0.50
No Degree	.333	.333	0.08	0.10	0.10	0.11	0.11	0.11	0.10	0.10	0.10	0.07	0.03	0.02

Panel B: % change from previous col.

Sci/Bus.	105.70%	82.90%	-2.80%	-14.30%	-1.40%	-1.50%	1.10%	5.40%	-2.00%	-12.70%
Humanities	-30.30%	-12.60%	1.70%	27.10%	0.40%	1.40%	-0.40%	-3.20%	7.60%	25.90%
No Degree	-75.10%	-70.00%	18.10%	9.00%	8.20%	1.80%	-5.70%	-14.40%	-5.00%	-26.30%

Panel A shows the mean probability of choosing the major. Each column progressively adds in beliefs for the (column variable) outcome.
 Panel B shows the percent change in the mean probability relative to the previous column.

APPENDICES (NOT FOR PUBLICATION)

A Estimation of Human Capital Choice Model

Utility for student i from human capital level k , where $k = \{\text{science/business, humanities, no degree}\}$, is given by

$$\begin{aligned}
 U_{ik} &= \alpha_0 \sum_{t=23}^T \beta^{t-23} \ln(y_{ikt}) + \alpha_1 \sum_{t=23}^T \beta^{t-23} y_{ikt}^{SPOUSE} \\
 &+ \alpha_2 \sum_{t=23}^{35} \beta^{t-23} P_{ikt}^{marriage} + \alpha_3 \sum_{t=36}^T \beta^{t-23} P_{ikt}^{marriage} \\
 &+ \alpha_4 \sum_{t=23}^{35} \beta^{t-23} E_{ikt}(Children) + \alpha_5 \sum_{t=36}^T \beta^{t-23} E_{ikt}(Children) \\
 &+ \alpha_6 Ability_{ik} + \varepsilon_{ik} \\
 &= \mathbf{X}'_{ik} \alpha + \varepsilon_{ik},
 \end{aligned} \tag{A-1}$$

where y_{ikt} (y_{ikt}^{SPOUSE}) is i 's (i 's spouse's) earnings in period t if i were to graduate with major k , $P_{ikt}^{marriage}$ is the probability of i being married by period t if i graduated with major k , $E_{ikt}(Children)$ is the expected number of children i expects to have by period t if she were to graduate with major k , and $Ability_{ik}$ is i 's perceived field-specific ability. We specify $T = 60$ and the discount factor, β , to be 0.95 in the estimation. The parameters to be estimated in equation (A-1) are $\alpha = \{\alpha_0, \alpha_1, \dots, \alpha_6\}$.

Because we ask individuals about full-time equivalent earnings only, we combine the beliefs about labor supply and full-time earnings to define earnings when the individual is working part-time. Earnings are modeled as

$$\ln(y_{ikt}) = FT_{ikt} \ln(w_{FT,ikt}) + (20/h_{FT,ikt}) PT_{ikt} \ln(w_{FT,ikt}), \tag{A-2}$$

where $w_{FT,ikt}$ are full-time earnings, $FT_{ikt} \in [0, 1]$ is the probability of working full-time, $PT_{ikt} \in [0, 1]$ is the probability for working part-time, and $h_{FT,q,t}$ is full-time hours. We ask respondents for their potential earnings *if* they are working full-time, and ask their beliefs about the probability of working full-time as a separate question. This allows us to elicit beliefs about earnings even if the individual believes there is a zero probability of actually working full-time in the future. This aspect of our data allows us to circumvent the standard endogenous wage distribution issue where wages are only observed for individuals who work. We also elicited beliefs about hours if

working full-time. However, because of time constraints, these were collected as beliefs about the population average hours worked conditional on major and gender, where we used the hours for the same (opposite) gender to construct an individual's beliefs about own (spouse's) hours.

There are certain data limitations due to time and respondent burden considerations: we cannot ask respondents to report marital status, labor supply, fertility, and earnings for *every* year after graduation. We instead ask these beliefs for two or three points in time after graduation. We use various approximations to interpolate between data points by age. It is important to emphasize that these approximations are entirely individual-specific (using free parameters for each individual): we make no assumption regarding the distribution of self beliefs in the population.

For the individual's own full-time earnings, we regress $\ln(w_{FT,ikt})$ on a time quadratic and a constant. This model is estimated separately for each individual, since we have three earnings beliefs (at age 23, 30, and 45) per field of study. The probability of working full-time and part-time is elicited for ages 30 and 45 only, and so working beliefs are given as:

$$W_{ikt} = \begin{cases} W_{ik30} & \text{for } 23 < t < 30 \\ W_{ik30} + \frac{W_{ik45} - W_{ik30}}{45 - 30}(t - 30) & \text{for } 30 \leq t < 45 \\ W_{ik45} & \text{for } t \geq 45, \end{cases}$$

where $W_{ikt} = \{FT_{ikt}, PT_{ikt}\}$. The interpolated labor supply beliefs and full-time earnings are then used to compute the expected earnings at each age, as in equation (A-2). Given that we collect beliefs data on spouse's labor supply and full-time earnings, a similar approximation approach is used for spouse's earnings.

Marriage beliefs are given as

$$P_{ikt}^{marriage} = \begin{cases} P_{ik23}^{marriage} & \text{for } t = 23 \\ P_{ik23}^{marriage} + \frac{P_{ik30}^{marriage} - P_{ik23}^{marriage}}{30 - 23}(t - 23) & \text{for } 23 < t < 30 \\ P_{ik30}^{marriage} + \frac{P_{ik45}^{marriage} - P_{ik30}^{marriage}}{45 - 30}(t - 30) & \text{for } 30 \leq t < 45 \\ P_{ik45}^{marriage} & \text{for } t \geq 45. \end{cases}$$

Fertility beliefs are given as

$$E_{ikt}(Children) = \begin{cases} \max \left\{ 0, E_{ik30}(Children) + \frac{E_{ik30}(Children) - E_{ik45}(Children)}{30 - 23}(t - 23) \right\} & \text{for } 23 \leq t < 30 \\ E_{ik30}(Children) + \frac{E_{ik45}(Children) - E_{ik30}(Children)}{45 - 30}(t - 30) & \text{for } 30 \leq t < 45 \\ E_{ik45}(Children) & \text{for } t \geq 45. \end{cases}$$

Under the assumption that the random terms ε_{ik} in equation (A-1) are independent and identically distributed Type 1 extreme value across individuals and majors k , the probability that student i chooses k is

$$p_{ik} = \frac{\exp(\mathbf{X}'_{ik}\alpha)}{\sum_{k'=1}^K \exp(\mathbf{X}'_{ik'}\alpha)}, \quad (\text{A-3})$$

Applying the log-odds transformation to this yields the linear model:

$$\ln\left(\frac{p_{ik}}{p_{ik'}}\right) = (\mathbf{X}_{ik} - \mathbf{X}_{ik'})' \alpha.$$

We assume that the actual reports of field choice probabilities in our data, denoted \tilde{p}_{ik} , measure the “true” probabilities p_{ik} with error. The measurement error takes a linear in logs form such that the reported log-odds take the following form:

$$\ln\left(\frac{p_{ik}}{p_{ik'}}\right) = (\mathbf{X}_{ik} - \mathbf{X}_{ik'})' \alpha + \omega_{ik},$$

where ω_{ik} is the measurement error. We assume that the ω_{ik} ’s are i.i.d. and have mean zero, conditional on \mathbf{X} ; this is a standard assumption in analysis of stated choices. Given this, we have the following linear mean regression model:

$$E \left[\ln\left(\frac{p_{ik}}{p_{ik'}}\right) | \mathbf{X} \right] = (\mathbf{X}_{ik} - \mathbf{X}_{ik'})' \alpha. \quad (\text{A-4})$$

For this exercise, we recode all reported extreme probabilities of 0 to 0.001 and 1 to 0.999. This follows Blass et al. (2010) who argue that dropping individuals with extreme probabilities would induce a sample selection bias in the resulting estimates. Model estimates are presented in Table A4. Since we have two observations per respondent, standard errors are clustered at the individual level.

We briefly comment on the sign and significance of the estimates. Discounted own lifetime earnings and ability rank are estimated to be positive and significant for both genders. Spousal earnings are a significant determinant of field choice, but only for female students. The flow utility from marriage is found to be insignificant for both genders. Utility from children is significant and positive for females only, but only at younger ages. The model fit, presented in Table A5, shows that the model fits the relative probabilities of choosing science/business and humanities well, but over-predicts the probability of the no degree option. Since the model does not include a field-specific constant, other omitted factors – such as consumption value from college (Jacob, McCall, and Stange, 2014) – may drive students away from the no degree alternative.

In the follow-up survey, we observe the actual major the student graduates with. We can use the actual major to estimate the parameters of the utility function as well. In that case, equation (A-3) is the familiar multinomial logit (with the dependent variable being equal to 1 for the individual’s chosen major(s), and zero otherwise). The model estimates, available from the authors upon request, are qualitatively similar to those in Table A4. The decomposition results, based on these estimates, are shown in Table A6. As one can see, the patterns are very similar to those in Table 15.

B Ability Beliefs

This section examines how men and women perceive their relative ability in each field. Major-specific ability may affect the likelihood of a student completing required coursework necessary to graduate in each major, and could be a factor in expectations about future earnings (Arcidiacono, 2004; Zafar, 2013). Beliefs about ability were elicited as follows: "*Consider the situation where either you graduate with a Bachelor's degree in each of the following major categories or you never graduate/drop out. Think about the other individuals (at NYU and other universities) who will graduate in each of these categories or never graduate/drop out. On a ranking scale of 1-100, where do you think you would rank in terms of ability when compared to all individuals in that category?*" We do not further define "ability" and leave the interpretation of this open to the respondent. To provide easier interpretation, we re-scaled the ability beliefs such that 100 represents highest ability and 1 represents lowest ability.

Panel A of Table B1 indicates that the students in the sample believe they are above average; an unsurprising and logical belief given the high ability sample of students. Male and female students differ markedly in their average beliefs about their major-specific ability. Average beliefs for women are higher in humanities than in science or business; while the opposite pattern is true for men, whose average beliefs are higher in science or business than humanities. This is consistent with evidence that women tend to be less confident than men in technical tasks (Weinberger, 2004; Niederle and Vesterlund, 2007). An even larger gender difference exists for average beliefs for ability in the "no degree" category. Women believe their ability would rank them just above the average among non-college graduates, while men believe their average rank would be closer to 70. It is notable that the average overall ability (obtained by weighting the major-specific ability measures by the stated probability of graduating with that major) does not differ by gender, indicative of the genders sorting into majors by ability.²⁹

We next construct an individual-specific measure of the relative difference in perceived ability ranks. As with earnings, we examine the log relative difference in perceived ability if the individual were to graduate with a science or business degree rather than a humanities degree. Panel B of Table B1 indicates that women on average perceive a *loss* in ability rank of around 25 percent if they were to complete a business or science degree rather than a humanities degree. Men, on average, perceive an average *gain* in relative rank of 14 percent if they were to complete a science or business degree rather than a humanities degree. Figure B2 shows the expected log rank difference in science/business versus humanities. We see that the male distribution is shifted to the right, with the mode being positive (negative) for males (females). However, there is substantial hetero-

²⁹Figure B1 shows the distribution of self ranked ability by gender. We see that the male distribution is shifted to the right, with more males expecting a rank of 80 or higher than females.

geneity, with a sizable proportion of both genders having negative (comparative ability advantage in humanities/social sciences) and positive differences.

Turning to differences in perceived ability rank between college and no degree, both men and women perceive a much higher relative ability with a college degree than no degree. The difference in perceived advantage for women is much larger than that for men, with women believing their individual rank advantage would be nearly 120 percent larger with a college degree than without a degree, compared to 62 percent for men. These differences in ability rank indicate that individuals strongly perceive that they have relative abilities in certain fields.

C Follow-Up Survey

In the 2010 survey, 475 of the 493 respondents had given consent for follow-up surveys and had provided an email address. An email inviting respondents to take the follow-up survey (with a link to the online survey) was sent to these 475 individuals. Of these 475 emails, 17 immediately bounced back. We were able to locate these 17 respondents on LinkedIn (based on their name, location, age, and university listed). We also reached out on LinkedIn to all the original respondents who did not respond to the follow up survey within two weeks. Each LinkedIn respondent was sent one personalized message, inviting them to take the survey. Two network-wide posts were made on the LinkedIn account as well. The compensation was a \$15 Amazon gift card, which was increased later to \$25 for respondents who had taken both the 2010 and 2012 surveys. The data were collected over a 6-week period during January and February of 2016. Three reminder emails were sent to all respondents over the six-week period for which the survey was open.

Table C1 shows the descriptive statistics of the sample. The top panel shows the 2010 characteristics of the follow-up sample. The statistics are very similar to those for the sample that took the 2010 survey (Table 1). In fact, Table C2 shows that the propensity to take the follow-up survey is not related to a large set of demographic characteristics. No demographic variable is significantly related to the likelihood of participating in the follow-up at the 95% level or higher. In addition, the p-value of a joint F-test of the covariates of 0.32 indicates that selection into the follow-up is not based on observables.

Panel B of Table C1 shows summary statistics of some information collected in the follow-up survey. The average respondent in 2016 is 25 years old. 75% of respondents are working full-time, and nearly half are in a relationship. The gender gap in own earnings (conditional on full-time work) is sizable: average full-time earnings of \$101,800 for males versus \$62,400 for females. Likewise, partners of female respondents have substantially higher earnings. The gender gap in actual major also goes in the direction one would expect- females are nearly twice (half) as likely as males to have graduated with a major in Humanities/Social Sciences (Economics/Business).

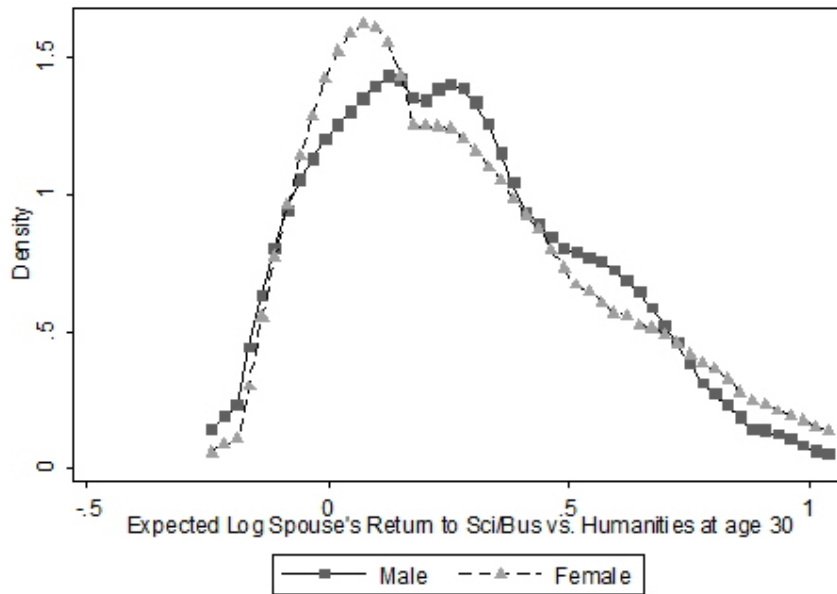


Figure A1: Distributions of Log Difference in Spousal Earnings for Own Degree in Science/Business versus Humanities.

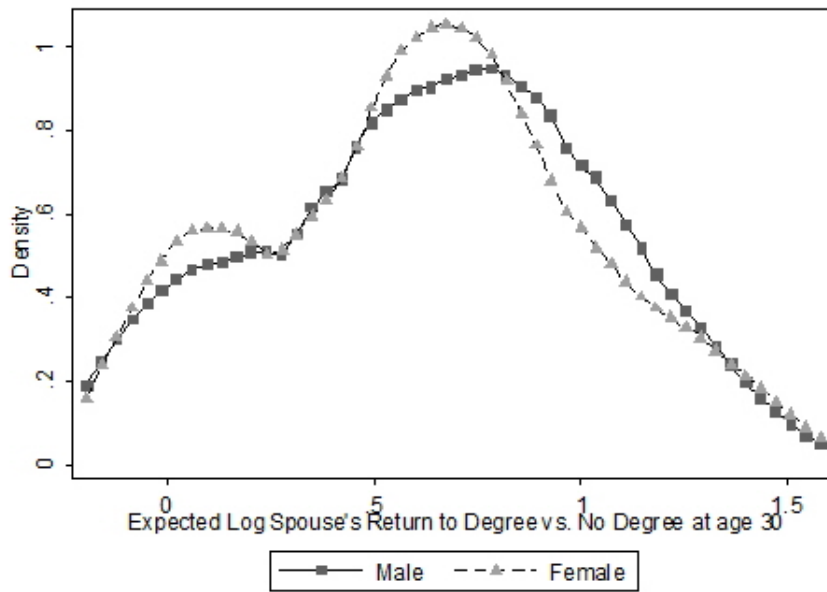


Figure A2: Distributions of Log Difference in Spousal Earnings, for Own Degree in Graduating versus No Degree.

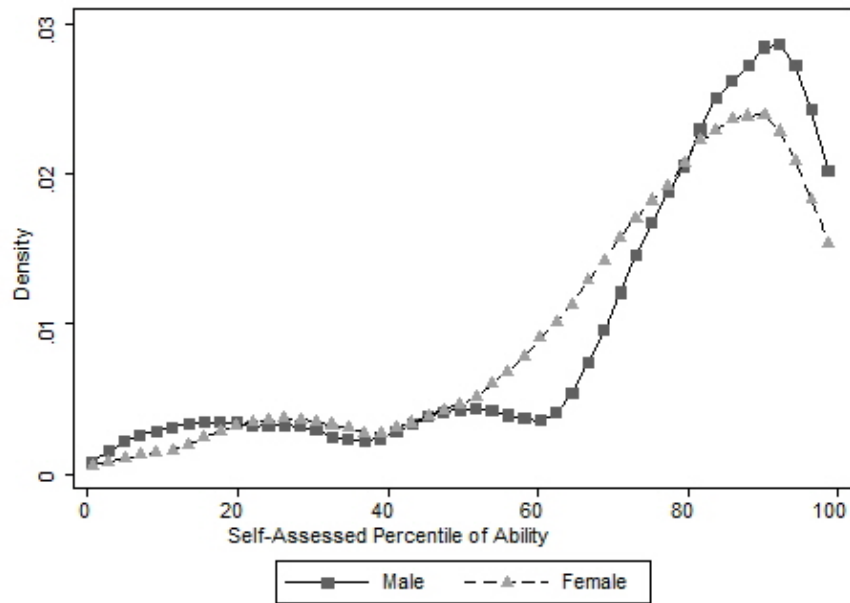


Figure B1: Distributions of Self Ranked (Percentiles) Ability, by Gender.

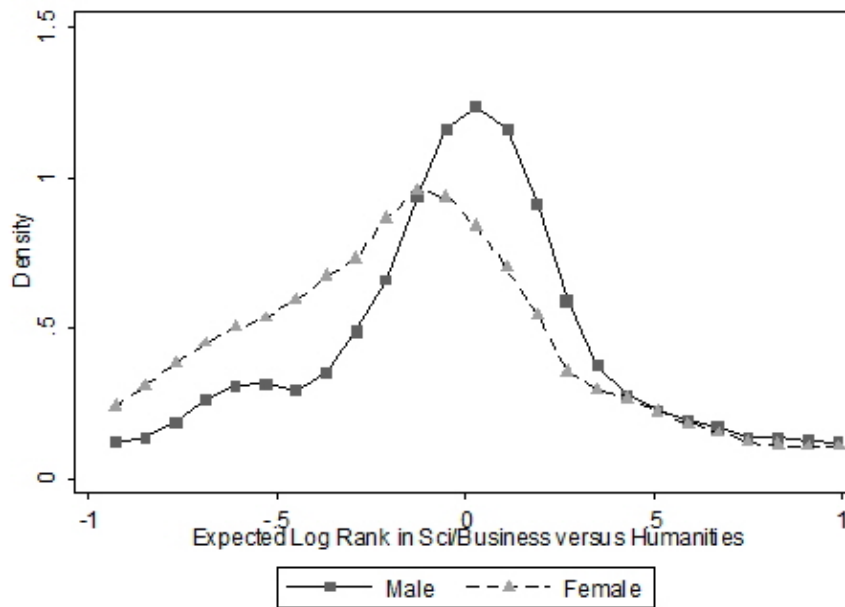


Figure B2: Distribution of Log Difference in Ability Ranking, Science/Business versus Humanities.

Table A1: Correlation in Earnings Beliefs

	Males			Females		
	Sci/Business	Humanities	No Degree	Sci/Business	Hum.	No Deg.
Age 23						
Sci/Business	1			1		
Humanities	.986	1		.849	1	
No Degree	.963	.965	1	.253	.307	1
Age 30						
Sci/Business	1			1		
Humanities	.242	1		.456	1	
No Degree	.557	.457	1	.456	.560	1
Age 45						
Sci/Business	1			1		
Humanities	.533	1		.724	1	
No Degree	.499	.463	1	.569	.72	1

Table A2: Log Differences in Own versus Spouse Earnings

	Age 23		Age 30		Age 45	
	Male	Female	Male	Female	Male	Female
Science/Business	.123 (.284)	-.063+++ (.31)	.305 (.605)	-.001+++ (.465)	.41 (.727)	.011+++ (.469)
Humanities	.041 (.368)	-.168+++ (.342)	.065 (.563)	-.135+++ (.43)	.205 (.525)	-.116+++ (.461)
Not Graduate	-.05 (.535)	-.279+++ (.441)	-.04 (.651)	-.36+++ (.638)	.182 (.61)	-.235+++ (.567)
Overall	.096 (.287)	-.12+++ (.329)	.291 (.622)	-.067+++ (.409)	.424 (.729)	-.035+++ (.456)

Table reports the means of $[\log(\text{own earnings}) - \log(\text{spouse earnings})]$. Standard deviation in parentheses.

+++, ++, + (shown on the female column) denote gender differences statistically different at the 1, 5, and 10% levels, respectively.

Table A3: Beliefs About Hours/Week for Current Full-time Workers

		Beliefs about Women		Beliefs about Men	
		Males	Females	Males	Females
Panel A: Average Hours per Week					
Age 30					
	Science/Business	48.94 (16.38)	49.04 (16.17)	51.67 (17.02)	50.88 (16.67)
	Humanities	44.47 (15.76)	45.40 (15.61)	45.74 (15.52)	45.50 (14.76)
	No Degree	44.99 (15.74)	45.84 (14.61)	47.06 (16.57)	47.12 (15.70)
	Overall	46.13 (15.37)	46.76 (13.97)	48.16 (15.48)	47.83 (14.08)
Age 45					
	Science/Business	45.85 (14.56)	46.71 (15.46)	48.61 (16.01)	48.52 (15.70)
	Humanities	43.88 (15.83)	43.93 (13.93)	44.81 (15.54)	44.78 (14.00)
	No Degree	44.34 (15.63)	44.78 (13.68)	46.08 (15.27)	46.36 (14.36)
	Overall	44.69 (14.67)	45.14 (13.17)	46.5 (15.06)	46.55 (13.18)
Panel B: Individual Log Differences					
Age 30					
	Sci/Business versus. Humanities	.095*** (.011)	.077*** (.009)	.118*** (.012)	.106*** (.011)
	Graduate versus. No Degree	.067*** (.013)	.037*** (.013)	.072*** (.015)	.039*** (.013)
Age 45					
	Sci/Business versus. Humanities	.049*** (.01)	.056*** (.008)	.081*** (.011)	.074*** (.009)
	Graduate versus. No Degree	.03** (.013)	.018 (.011)	.04*** (.012)	.02* (.012)

Panel A shows the mean and standard devs of hours beliefs. +++, ++, + denote gender diffs are statistically significant at the 1, 5, and 10% levels, respectively.

Panel B shows the average log differences and standard deviations in parentheses. ***, **, * denote the means are statistically different from zero at the 1, 5 and 10% levels, respectively.

+++, ++, + (shown on the female column) denote gender differences are statistically significant at the 1, 5, and 10% levels, respectively.

Table A4: Model Estimates

	Males	Females
Discounted Own Lifetime Earnings (α_0)	0.506*** (0.051)	0.400*** (0.043)
Discounted Spouse's Lifetime Earnings (α_1)	0.0579 (0.090)	0.115* (0.061)
Probability of marriage (age 23-35) (α_2)	-0.132 (0.447)	0.329 (0.346)
Prob of marriage (age 36-60) (α_3)	-0.719 (0.525)	-0.083 (0.346)
Expected num of children (age 23-35) (α_4)	0.186 (0.130)	0.382*** (0.102)
Expected num of children (age 36-60) (α_5)	-0.291 (0.210)	-0.357** (0.177)
Ability rank (α_6)	0.0233*** (0.005)	0.0281*** (0.005)
Number of individuals	176	317
Number of observations	352	634
R2	0.332	0.421

OLS estimates of model in A-1 (Appendix A) presented. Standard errors (clustered at the individual level) in parentheses.

***, **, * denote significance at the 1%, 5%, and 10% levels, respectively.

Table A5: Model Fit

	Observed Prob	Predicted Prob
Males		
Science/Business	66.92%	65.12%
Humanities	30.05%	25.4%
No Degree	3.22%	9.48%
Females		
Science/Business	48.08%	54.57%
Humanities	49.94%	34.22%
No Degree	2.16%	11.21%

Table shows the average likelihood of choosing each of the fields. Observed prob is the mean probability in the data. Predicted Prob is the mean probability as predicted by the model in Appendix A.

Table A6: Decomposition Results, Based on Revealed Major Choice

	Baseline		Earnings		Ability		Marriage		Spouse Earnings		Fertility		Actual	
	Male	Fem	Male	Fem	Male	Fem	Male	Fem	Male	Fem	Male	Fem	Male	Fem
	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)	(5a)	(5b)	(6a)	(6b)	(7a)	(7b)
Panel A: Average Probability														
Sci/Bus.	.333	.333	0.68	0.55	0.68	0.47	0.68	0.45	0.67	0.48	0.65	0.42	0.72	0.41
Humanities	.333	.333	0.22	0.32	0.22	0.39	0.22	0.40	0.22	0.39	0.25	0.48	0.40	0.61
No Degree	.333	.333	0.10	0.13	0.11	0.15	0.11	0.16	0.11	0.13	0.10	0.10	0.04	0.03

Panel B: % change from previous col.

Sci/Bus.	104.50%	65.20%	-0.30%	-15.10%	0.00%	-4.70%	-0.70%	7.40%	-3.10%	-12.10%
Humanities	-34.50%	-5.10%	-0.90%	22.50%	0.00%	3.10%	-0.50%	-2.00%	14.00%	23.30%
No Degree	-69.70%	-59.80%	4.00%	9.00%	1.00%	6.20%	4.70%	-15.50%	-8.10%	-25.20%

Panel A shows the mean probability of choosing the major. Each column progressively adds in beliefs for the (column variable) outcome.

Panel B shows the percent change in the mean probability relative to the previous column.

Table B1: Perceived Ability Ranking

	Male	Female
Panel A: Levels		
Sciences/Business	70.69 (25.73)	61.35+++ (25.31)
Humanities	68.49 (29.33)	74.15++ (24.02)
No Degree	69.72 (39.25)	56.53+++ (43.20)
Overall	76.23 (23.93)	74.24 (22.13)
Panel B: Individual Log Differences		
Sci/Business versus. Humanities	.138* (.073)	-.246***+++ (.052)
Graduate versus. No Degree	.624*** (.137)	1.21*** (.118)

Panel A shows the mean and standard deviations of perceived ability ranking (ability is elicited on a 1-100 scale, where 100 is the highest rank). +++, ++, + denote gender differences are statistically different at the 1, 5, and 10% levels, respectively. Panel B shows the average log differences and standard deviations in parentheses. ***, **, * denote the means are statistically different from zero at the 1, 5, and 10% levels, respectively. +++, ++, + (shown on the female column) denote gender diffs are statistically different at the 1, 5, and 10% levels, respectively.

Table C1: Descriptive Statistics of Follow-up Survey

	All	Male	Female
Number of observations	274	88	186
Panel A (2010 Survey Characteristics)			
% white	45.62	50	43.55
% Asian	43.07	40.91	44.09
Parents' Income (\$1,000s)	155.03 (125.17)	162.79 (125.76)	151.32 (125.07)
% Mother BA or More	70.22	75	67.93
% Father BA or More	76.58	75.86	76.92
SAT Math Score	700.85 (81.48)	714.7 (77.17)	694.29* (82.85)
SAT Verbal Score	589.15 (403.39)	584.15 (397.48)	591.51 (407.19)
GPA	3.46 (.43)	3.46 (.49)	3.46 (.4)
Panel B (2016 Characteristics)			
Age	25.2 (1.07)	25.33 (1.24)	25.14 (.98)
Modal Graduation Yr from NYU	2012	2012	2012
Labor force status:			
% employed full-time	73.99	73.86	74.05
% employed part-time	9.16	9.09	9.19
% not working	16.85	17.05	16.76
Earnings (in \$10,000s) Full-Time	7.49 (7.74)	10.18 (12.39)	6.24*** (3.46)
% Married	5.56	8.14	4.35
% In a Relationship	48.15	45.35	49.46
% Partner Employed Full-Time	75.37	70.00	77.66
Partner's Earnings (in \$10k) Full-Time	7.73 (6.14)	5.68 (3.53)	8.5** (6.73)
% Have Kids	1.48	2.33	1.09
% who graduate with:			
Economics/Business	26.37	44.32	17.84***
Engineering/Computer Science	2.93	3.41	2.7
Humanities/Social Sciences	52.75	32.95	62.16***
Natural Sciences/Math	13.92	14.77	13.51
Not graduate	4.03	4.55	3.78*

For the continuous outcomes, means are reported in the first cell, and standard deviations are reported in parentheses.

***, **, * (denoted on female column) indicate gender difference in means sig. at 1%, 5%, and 10% levels, respectively.

Table C2: Correlates of Participation in Follow-up

Dep Var: Dummy for Follow-up Participation	
Age	-1.79 (2.60)
Freshman	-0.017 (7.66)
Sophomore	9.60 (6.52)
Male	-7.61 (4.96)
Asian	-8.51 (5.62)
Hispanic	-1.33 (9.63)
Black	-21.86 (14.43)
Other Race	-16.84* (9.11)
Father Went to College	0.023 (0.062)
Mother Went to College	-0.059 (0.059)
Parental Income (\$1,000s)	0.032 (0.020)
Overall GPA	-2.44 (6.11)
SAT Math Score	-0.024 (0.033)
SAT Verbal Score	0.0045 (0.032)
Business/Economics	-8.85 (7.02)
Engineering/Computer Science	-15.65 (12.67)
Humanities/Other Social Sciences	-7.47 (6.63)
Constant	123.90** (60.51)
F-test (p-value) ^a	0.320
Mean of Dep Variable	55.58
R ²	0.054
Observations	493

OLS estimates presented. Std errors reported in parentheses.
 ***, **, * denote sig. at 1%, 5%, and 10% levels, respectively.
^a P-value reported for a joint F-test of sig. of all covariates.

Table C3: Link Between Expectations and Outcomes (Using Actual Major for Expectations)

	All	Males	Females
Panel A, dep variable: Log (current earnings)			
Log(Exp Earnings, Age Weighted)	0.440*** (0.120)	0.282 (0.210)	0.505*** (0.122)
Observations	200	64	136
R^2	0.121	0.049	0.154
Mean of Dependent Variable	10.99	11.18	10.90
Panel B, dep variable: Employed Full-time			
Expected Prob of FT Emp at 30	0.204 (0.134)	-0.0672 (0.224)	0.377** (0.171)
Observations	272	88	184
R^2	0.010	0.001	0.031
Mean of Dependent Variable	0.740	0.740	0.740
Panel C, dep variable: Employed Part-time			
Expected Prob of PT Emp at 30	0.198 (0.134)	0.0342 (0.204)	0.283* (0.169)
Observations	272	88	184
R^2	0.010	0.000	0.020
Mean of Dependent Variable	0.0900	0.0900	0.0900
Panel D, dep variable: Married			
Age-Weighted Exp Probability of Being Married	0.240** (0.0975)	0.411* (0.209)	0.172* (0.0975)
Observations	269	86	183
R^2	0.052	0.112	0.032
Mean of Dependent Variable	0.0600	0.0800	0.0400
Panel E, dep variable: In Any relationship			
Age-Weighted Exp Probability of Being Married	0.548*** (0.122)	0.741*** (0.196)	0.443*** (0.156)
Observations	269	86	183
R^2	0.057	0.110	0.035
Mean of Dependent Variable	0.480	0.450	0.490
Panel F, dep variable: Spouse/Partner Working Full-time			
Expected Prob of Spouse FT Emp at 30	0.372** (0.175)	0.412 (0.278)	0.286 (0.242)
Observations	130	39	91
R^2	0.036	0.048	0.017
Mean of Dependent Variable	0.760	0.690	0.790
Panel G, dep variable: Log(Spouse/Partner Earnings)			
Log(Age-Weighted Expected Earnings of Spouse)	0.370** (0.167)	0.753*** (0.260)	0.281 (0.190)
Observations	112	31	81
R^2	0.043	0.138	0.028
Mean of Dependent Variable	1.690	1.420	1.790

Each cell in the table presents estimates from an OLS regression of the dependent variable on the 2010 expectation and a constant. Std errors reported in parentheses. ***, **, * denote sig. at 1%, 5%, and 10% levels, respectively.

This table is analogous to Table 13, except that for expectations it uses the individual's expectation for the major they graduate with.