Determinants of College Major Choice: Identification using an Information Experiment*

Matthew Wiswall Basit Zafar†

First Version: June 2011
This Version: March 2012

Abstract

This paper studies the determinants of college major choice using an experimentally generated panel of beliefs, obtained by providing students with information on the true population distribution of various major-specific characteristics. Students logically revise their beliefs in response to the information, and their subjective beliefs about future major choice are associated with beliefs about their own earnings and ability. We estimate a rich model of college major choice using the belief data. While earnings are a significant determinant of major choice, tastes – which are heterogeneous – are the dominant factor in the choice of major. We also investigate gender differences in major choice.

JEL Codes: D81, D84, I21, I23, J10.
Keywords: college majors; information; uncertainty; risk; subjective expectations; marriage market returns; gender differences.

1 Introduction

Understanding the determinants of occupational choices is a classic question in the social sciences: How much do occupational choices depend on expected future earnings versus tastes for various non-pecuniary aspects of an occupation? Among college graduates, occupational choices are strongly associated with college major choices as the choice of major–whether in humanities, business, science or engineering fields–represents a substantial investment in occupation-specific human capital. Underscoring the importance of college major choices, a number of studies have documented that choice of post-secondary field is a key determinant of future earnings, and

*We thank the NYU Center for Experimental Social Sciences (CESS) for providing assistance in conducting the information survey and experiment. We thank participants at presentations at the NY Fed BBL, NYU Experimental Economics Working Group, Clemson, and the 2012 ASSA meetings. Da Lin and Scott Nelson provided outstanding research assistance. All errors that remain are ours. The views expressed in this paper do not necessarily reflect those of the Federal Reserve Bank of New York or the Federal Reserve System as a whole.

†Wiswall: New York University, Department of Economics; E-mail: mwiswall@nyu.edu. Zafar: Research and Statistics, Federal Reserve Bank of New York; E-mail: Basit.Zafar@ny.frb.org

that college major composition can help explain long-term changes in inequality and earnings differences across racial groups and between men and women (Grogger and Eide, 1994; Brown and Corcoron, 1997; Weinberger, 1998; Arcidiacono, 2004; Wiswall, 2006).

This paper studies the determinants of college major choices using a unique survey and experimental design. We conduct an experiment on undergraduate college students of New York University (NYU), where in successive rounds we ask respondents their *self* beliefs about their own expected earnings and other major-specific aspects were they to major in different majors, their beliefs about the population distribution of these outcomes, and the subjective belief that they will graduate with each major. After the initial round in which the baseline beliefs are elicited, we provide students with accurate information on population characteristics of the major and observe how this new information causes respondents to update their self beliefs and their subjective probabilities of graduating with each particular major. Our experimental design creates unique panel data for major choice, which is otherwise a one-time decision. By comparing the experimental *changes* in subjective probabilities of majoring in each field with the *changes* in subjective expectations about earnings and other characteristics of the major, we can measure the relative importance of each of these various characteristics in the choice of major, without bias stemming from the correlation of fixed preferences with characteristics. Underscoring the importance of this bias, we compare cross-sectional OLS estimates of major choice to expectations about earnings with our panel fixed effects estimates, and find that the OLS estimates are severely biased upward due to positive correlation of unobserved tastes with earnings expectations.

Our approach is motivated by previous research which has found that many college students have biased beliefs about the population distribution of earnings among current graduates (Betts, 1996), and that students tend to be misinformed about returns to schooling (Jensen, 2010; Nguyen, 2010). We test whether students update their beliefs if given accurate information on the current population earnings, and find heterogeneous errors in population beliefs, and substantial and logical updating in response to our information treatment. We show how the experimental variation alone identifies a rich model of college major choice, and we use this model to understand the importance of earnings and earnings uncertainty on the choice of college major relative to other factors such as ability to complete coursework, spousal characteristics, and tastes for majors.

The standard economic literature on decisions made under uncertainty, such as occupational and educational choices, generally assumes that individuals, after comparing the expected outcomes from various choices, choose the option that maximizes their expected utility (Altonji, 1993). Given the choice data, the goal is to infer the parameters of the utility function. Because one does not typically observe expectations about future choice-specific outcomes, such as the student's expectations of earnings and ability in a major, assumptions have to be made on expectations to infer the decision rule. This approach requires a mapping between objective measures (such as realized earnings) and beliefs about them. Moreover, assumptions also have to be invoked about expectations for counterfactual majors, i.e., majors not chosen by the student. Several studies of college major choice use this approach (Freeman, 1971; Bamberger, 1986; Berger, 1988; Montmarquette, Cannings, and Mahseredjian, 2002; Arcidiacono, 2004; Beffy, Denis, and Maurel, 2011; Gemici and Wiswall, 2011). While these studies allow varying degrees of individual heterogeneity in beliefs about ability and future earnings, they typically assume that expectations are either myopic or rational, and use realized choices and realized earnings to identify the choice model. This approach is problematic because observed choices might be consistent with several combinations of expectations and preferences, and the underlying assumptions may not be valid (Manski, 1993).

A recent literature has evolved which collects and uses subjective expectations data to understand decision-making under uncertainty (see Manski, 2004, for a survey of this literature). In the context of schooling choices, Zafar (2009, 2011a), Giustinelli (2010), Arcidiacono, Hotz, and Kang (2011), Kaufmann (2011), and Stinebrickner and Stinebrickner (2010, 2011) incorporate subjective expectations into models of choice behavior. These studies collect data on expectations for the chosen alternative as well as counterfactual alternatives, thereby eliminating the need to make assumptions regarding expectations. However, as we show in Section 4, one cannot separately identify the tastes for each major from other aspects of the choice (earnings, ability, etc.) without imposing further modeling restrictions. Even in studies with panels of beliefs, the beliefs collected are separated by several months or years, requiring assumptions about the stability of preferences across this period.

We exploit experimental variation in information that creates *within* individual variation in beliefs, which we can then use to identify the choice model under more limited assumptions than in the previous research. More precisely, at the baseline, we collect self beliefs and beliefs about the population distribution of some college major characteristics, as well as probabilistic choices of major. We then provide students with accurate fact-based information on population characteristics. If students are mis-informed about population characteristics and perceive some link between population and self beliefs, this information should cause them to revise their beliefs and choices. There are in fact substantial errors in population beliefs. For example, male and female respondents overestimate the female population full-time average earnings in Economics/Business by around 30%. We next find that students logically revise their self beliefs about own earnings in response to the information we provide. The response, however, is inelastic: For a 1 percent error, students revise their self earnings by 0.07 percent, suggesting that self beliefs are not entirely linked to the type of public information that we provide.

Our reduced-form estimates using baseline (cross-sectional) data show that beliefs about future relative major choices are positively and strongly associated with beliefs about future self earnings. For example, a 1 percent increase in beliefs about self earnings in a major (relative to humanities/arts) increases the log odds of majoring in that field (relative to humanities/arts) by about 1.6 percent. However, using the revisions in beliefs and choices, we show that in fact the estimates using cross-sectional data are biased upwards because of the positive correlation between the unobserved individual-specific taste component and beliefs about earnings. For example, the choice elasticity with respect to beliefs about earnings is an order of magnitude lower (about 0.2 percent) using revisions in beliefs and choices, as part of an individual fixed effect analysis.

We next estimate a structural life-cycle utility model of college major choice. Unlike the existing literature on educational choices that only elicits beliefs of expected future earnings (Attanasio and Kaufmann, 2011, is an exception), we collect data on beliefs about the underlying earnings distribution, and also investigate the role that risk plays in college major choice. In addition, the model includes beliefs about ability, labor supply, and marriage market returns. Our parameter estimates imply a relative risk aversion coefficient of around 4-5, similar to that found by Nielsen and Vissing-Jorgensen (2006) in a Danish dataset on labor incomes and educational choices. Moreover, our estimate of relative risk aversion is higher for females, which is consistent with experimental studies of gender differences in risk preferences (Eckel and Grossman, 2008; Croson and Gneezy, 2009).

Our model estimates indicate that earnings are a significant determinant of major choice. However, the taste component at the time of choosing a college major is the dominant factor in the choice of field of study, a finding similar to that of Arcidiacono (2004), Beffy et al. (2011), and Gemici and Wiswall (2011). With respect to the marriage market returns to major choice, we find that they have a small positive impact on choosing high-earnings majors, but a substantial negative impact on choosing the "not graduate" category.

This paper also contributes to the literature on gender differences in schooling choices. Males and females are known to choose very different college majors (Turner and Bowen, 1999; Dey and Hill, 2007; Gemici and Wiswall, 2011). Niederle and Vesterlund (2007) speculate that women being less over-confident than men is one possible explanation for this. Zafar (2010), in his sample of Northwestern University undergraduates, finds that gender differences in tastes (and not ability) are the main source of these differences. In our sample, we find that women, on average, do have lower beliefs of ability in all fields relative to men. The gender-specific model estimates show that earnings differences across majors are a substantially smaller factor in college major choice for women than men, and that ability differences matter substantially more for women. The taste component is, however, dominant for both males and females.

While our experimental variation generates a panel that may look similar to other datasets with longitudinal information on beliefs (see Stinebrickner and Stinebrickner, 2010, 2011; Zafar, 2011a, in the context of college major choice), there is an important distinction: Beliefs in our survey are separated by only a few minutes, while in conventional panels, the gap is typically of several months or years. We can then credibly claim that the utility function, most notably the individual and major specific taste parameters, are truly time invariant in our context–the key assumption to identifying the tastes non-parametrically–and that our experimentally derived panel data satisfies the standard fixed effects assumptions. Estimating the taste parameters non-parametrically, we find that i) the distribution of tastes is bimodal, ii) average tastes of females are negative for all majors (relative to humanities/arts), and iii) male students have a strong relative taste for economics/business majors. Moreover, the fit of the estimated structural model using the experimental variation in beliefs is substantially better than when we estimate the model using cross-sectional data and impose a parametric assumption on the taste parameter, as in Arcidiacono et al. (2011).

The innovation of the current design is that it experimentally shifts beliefs to generate within-individual variation in expected earnings and probabilistic choices across majors. Such within-individual variation in earnings and choices is never available. In addition, the taste component is allowed to be correlated with other components in the model, such as ability and earnings, and can take any form. In fact we find evidence that tastes are strongly correlated with observables: For example, we estimate that male, high-SAT Math score, and Asian respondents have a stronger distaste for humanities/arts majors, net of differences in earnings, ability, and other factors. The strong correlation of the taste component with observables that we find has implications for how tastes are modeled in choice models. Generally it is assumed that tastes are orthogonal to other components in the model (for example, see Arcidiacono et al., 2011); our results then imply that such modeling assumptions would yield biased estimates.

This paper is organized as follows. Section 2, through a simple example, provides intuition for our identification strategy. Section 3 outlines the model of college major choice. In Section 4, we explore identification of the model using: i) commonly used revealed choice data, ii) cross-sectional beliefs, and iii) panel data on beliefs. The data collection methodology is outlined in Section 5. We examine heterogeneity in beliefs about earnings and revisions in self beliefs following the information treatment in Section 6.1. Section 6.2 reports reduced-form regressions on the relationship between beliefs about major choice and beliefs about elements of future post-graduation utility, while Section 7 reports estimates from a structural life-cycle utility model of major choice. Finally, Section 8 concludes.

2 A Simple Example

We first consider a simple example to provide some intuition for our experiment-based identification strategy. We collect students' subjective beliefs regarding characteristics of majors (e.g., future earnings) and subjective beliefs about completing each major in two stages: a pre-treatment stage and a post-treatment stage. The post-treatment stage is after we provide students with information about the population characteristics of graduates in each major, where the information we provide is the "treatment". For this example, we focus on only one characteristic of a major— average earnings. We collect an array of data about future events associated with majors and, in our empirical estimation, we consider a richer life-cycle specification of the utility function.

Let $\bar{w}_{k,i}$ be student i's pre-treatment belief about average earnings if she were to graduate with major k. Let $\pi_{k,i}$ be student i's pre-treatment belief she will graduate with major k. The information treatment provides new information to the student on the population distribution of earnings, and following the information treatment, student i can revise her beliefs about her future earnings in each major and her future probability of graduating with each degree. We cannot of course provide exact information about the future earnings of a student; we can only provide information about the distribution of population earnings. As discussed below, we see that this information causes logical updating of the student's own assessment of her future earnings. Let $\bar{w}'_{k,i}$ be the updated belief about future earnings in major k, and $\pi'_{k,i}$ be the updated belief about the probability of graduating with major k. The post- minus pre-treatment ratio in beliefs about completing the major relative to beliefs about future earnings is given by

$$\frac{\pi'_{k,i} - \pi_{k,i}}{\bar{w}'_{k,i} - \bar{w}_{k,i}}. \tag{1}$$

The intuition for our identification strategy is clearly seen in (1). The numerator of (1) measures the extent of the relative revision in beliefs about the probability of completing a major from the pre-treatment to the post-treatment period. The denominator of (1) measures the extent of the relative revision in self beliefs about earnings. The ratio of the revision of the self-reported major probabilities versus the revision in earnings identifies the marginal utility of earnings in major choice. If there is a large revision in probabilities relative to a small revision in earnings, then we conclude that earnings are an important factor in major choice. If however, there is little revision in probabilities relative to a large revision in earnings, then we conclude that other factors such as tastes or abilities, not earnings, are the predominant consideration in major choice. As we discuss in more detail below, our strategy of using *revisions* in expected choices relative to *revisions* in elements of expected utility allows us to robustly assess the importance of various elements of the utility function without (1) imposing parametric restrictions on the distribution of major-specific tastes, and (2) making assumptions about the expectations formation process.

3 Model

In this section we specify the model of college major choice. The next section shows how we use the information experiment to identify the model. The details of the life-cycle specification of future utility is specified in section 7.

Individuals choose one of K majors: $k = 1, \ldots, K$.[1] At the initial period $t = -1$, individuals are enrolled in college and have not chosen a particular college major. At period $t = 0$, the individual makes a college major choice and graduates from college. From period $t = 1$ onward, the college graduate makes all remaining choices, including choices regarding labor supply and marriage.[2]

We do not explicitly model any of the choices during or after college (i.e., choice to take particular courses during college, or any of the post-graduation choices). Instead we specify a preference ordering over the particular college majors. At period $t = -1$ (prior to choice of major), expected utility for each college major is given by

$$V_{-1,k} = \gamma_k + v(a_k) + EV_{0,k}, \tag{2}$$

where the $\gamma_1, \gamma_2, \ldots, \gamma_K$ components represent the preferences or tastes for each college major k at the initial pre-graduation stage. We define "tastes" at the point when students are in college. These could be tastes for major-specific outcomes realized in college, such as the enjoyability of coursework, or major-specific post-graduation outcomes, such as expected non-pecuniary aspects of jobs. Note that while we define tastes here during the college choice period, there is no loss of generality in modeling these time-invariant tastes as preferences over future events. These "tastes" also implicitly reflect the "switching costs" of changing majors while in school. As college students progress through college, they may optimally decide to change their major, and the data we collect on self reported probabilities $(0 - 1)$ about graduating with a given major reflect this. From this perspective, the γ_k "tastes" for major are then the cost to switching majors, with a large positive γ_k leading students to be less likely to switch out of major k into an alternative major.

$v(a_k)$ is the mapping of a student's *perceived* ability in each major to pre-graduation utility from each major, where $a_k \geq 0$ for all k. We assume $\partial v(a_k)/\partial a_k \geq 0$, reflecting that higher ability in a particular major improves performance in the major's coursework and reduces the effort cost of completing it. Ability in coursework and ability in the labor market can be closely correlated, but we do not explicitly model this interaction since our data allow us to measure

[1] As described below in the Data section, in order to model the complete potential choice set, one of the "majors" is a "no graduation" (college drop-out) choice.

[2] To make clear how this timing convention is reflected in our survey design, note that we survey college students (1st-3rd year students) at period $t = -1$, prior to college graduation. We do not survey 4th year and later students because they may have already chosen a particular college major.

expected earnings in each field and beliefs about ability in each field directly.[3] Expectations are formed according to the beliefs in period $t = -1$.[4]

At period $t = 0$, the student realizes some preference shock and then chooses her college major. Expected utility at the time of graduation for each major k is given by

$$V_{0,k} = \eta_k + \beta EV_{1,k}, \tag{3}$$

where $\eta_1, \eta_2, \ldots, \eta_K$ are the period $t = 0$ preference shocks that reflect any change in preferences that occur between the initial pre-major choice period $t = -1$ and the period when the college major is chosen.[5] In the Blass, Lach, and Manski (2010) taxonomy, η_k is "resolvable" uncertainty–uncertainty that is resolved at the point at which the choice of major is made.[6]

After college graduation, the expected discounted sum of future post-graduation utility from each major k is given by

$$EV_{1,k} = \sum_{t=1}^{T} \beta^{t-1} \int u(X) dG(X|k, t), \tag{4}$$

where $u(X)$ is the utility function that provides the mapping from the finite vector of events X to utility. X can include a wide range of events (e.g. earnings, labor supply, marriage, spousal earnings, and so on). $G(X|k, t)$ is the individual's beliefs about the distribution of future events in period t, conditional on choice of major k. The distributions of future events $G(X|k, t)$ represent "unresolvable" uncertainty as these events will not have occurred at the time of major choice. Beliefs are individual-specific and based on current information, which, as discussed below, can be a mixture of public and private information. In the next sections, we refer to these beliefs as "self" beliefs, e.g., 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 of some major characteristics, e.g., beliefs about the average

[3] In our data, we find that a student's self-reported ability rank in each major is highly correlated with self-reported expected future earnings in the field.

[4] For simplicity, (2) ignores any real separation of the $t = -1$ and $t = 0$ periods. We implicitly assume that the period $t = -1$ is "just" before the decision making period in $t = 0$. Alternatively, we could write: $V_{-1,k} = \gamma_k + v(a_k) + \beta EV_{0,k}$. However, this model is not separately identifiable from the model presented above since the discount rate would not be identified separately from the scale of the η_k shocks (3), and we can capture differences in utility flows from future post-graduation activities with a shift in the utility function (4).

[5] Note that it makes no difference whether one places the taste or ability components in the $t = 0$ period or in the $t = -1$ period. Given that we have no discounting for these college periods, the following model is equivalent in terms of choice probabilities to (2) and (3): $V_{-1,k} = EV_{0,k}$ and $V_{0,k} = \gamma_k + v(a_k) + \eta_k + \beta EV_{1,k}$.

[6] While we do not model it explicitly, our model does not rule out that individuals switch intended majors between the $t = -1$ and $t = 0$ periods. Our model is of expectations, *at the point of our information experiment*, regarding the probability of graduating with a given major. Most of our student respondents are freshman or sophomores. Given that most respondents place non-zero probability on all potential majors, the students are revealing that they in fact believe that switching from their intended major is indeed a possibility before graduation.

earnings in the population of individuals who graduate with a business degree.

Individuals choose the college major that maximizes expected utility at period $t = 0$: $V_0^* = \max\{V_{0,k}, \ldots, V_{0,K}\}$. At $t = -1$, each individual's expected probability of majoring in each of the k majors given beliefs is then obtained by integrating over the distribution of resolvable uncertainty:

$$\pi_k = \int 1\{V_{0,k}^* = V_0^*\} dF(\eta), \tag{5}$$

where $F(\eta)$ is the joint distribution of $\eta_1, \ldots, \eta_K$, and $\sum_{k=1}^{K} \pi_k = 1$.

4 Identification

In this section, we explore identification of the model using three types of data: i) commonly used revealed choice data in which we observe one choice of college major for each individual along with actual realized features of this major (e.g. earnings), ii) a cross-section of baseline (pre-treatment) beliefs only, and iii) panel data including both pre- and post- treatment beliefs.

4.1 Identification Using Actual Choice Data

We first consider identification with the typical revealed preference data in which we observe for each individual i their actual choice of major (i.e., the data are collected after college graduation). In revealed preference data, we typically observe a set of indicators for major choice, some measure(s) of ability, and some realizations of future events, such as future earnings in the chosen major. Let $d_{1,i}, \ldots, d_{K,i}$ be the set of indicators for these choices such that $d_{k,i} = 1\{V_{0,k,i} = V_{0,i}^*\}$ for all k. From these revealed choices, we can identify the probability that each major is chosen:

$$\begin{aligned} P_k &\equiv pr(d_{k,i} = 1) \\ &= \int \pi_{k,i} dQ(\gamma_{1,i}, \ldots, \gamma_{K,i}, a_{1,i}, \ldots, a_{K,i}, G_i(X|t, 1), \ldots, G_i(X|t, K)), \end{aligned}$$

where $\sum_{k=1}^{K} P_k = 1$. $Q(\cdot)$ is the population distribution of tastes, abilities, and beliefs about future post-graduation events. Note that P_k is distinct from $\pi_{k,i}$: P_k is the probability major k *was* chosen, which is revealed in post-graduation data, whereas $\pi_{k,i}$ is the belief about the future probability that major k *will* be chosen.

With this revealed preference data, the researcher faces the task of constructing elements of the utility function from actual observed data. In general, this requires four additional layers of assumptions:

i) an assumed mapping between revealed or actual post-graduation earnings to beliefs about earnings (or any other elements of post-graduation utility) for the major that is chosen,

ii) an assumed model for counterfactual beliefs about earnings (or any other elements of post-graduation utility) in majors not chosen,

iii) an assumed mapping between measures of ability to beliefs about ability in each major, and

iv) an assumed distribution of tastes for all majors.

The prior literature makes various types of assumptions along these dimensions.[7] This approach overlooks the fact that subjective expectations may be different from objective measures, assumes that formation of expectations is homogeneous, and uses choice data to infer decision rules conditional on maintained assumptions on expectations. This can be problematic since observed choices might be consistent with several combinations of expectations and preferences, and the list of underlying assumptions may not be valid (see Manski, 1993, for this inference problem in the context of how youth infer returns to schooling; also see Wolpin, 1999, and Manski, 2004).

4.2 Identification Using Baseline Beliefs

We next turn to considering identification if we have baseline beliefs data only, and do not have the post-treatment information from our information experiment. This is the data available, for example, in Delavande (2008), van der Klaauw and Wolpin (2008), Zafar (2009), Giustinelli (2010), Arcidiacono et al. (2011), Attanasio and Kaufmann (2011), and van der Klaauw (2011). The benefit of collecting belief information for outcomes in all possible choices is that this allows the researcher to relax assumptions about i) the mapping between realizations and beliefs for outcomes in the choice made, and ii) beliefs for outcomes in counterfactual choices not chosen.

In order to motivate our use of panel data on beliefs and make transparent the potential sources of bias in using cross-sectional data, let the vector of relevant future events X be divided into a subset of observed (to the researcher, in the data) events X^o and unobserved events X^u: $X = [X^o \; X^u]$. Also assume the utility function is additively separable in these arguments: $u(X) = u^o(X^o) + u^u(X^u)$. Note in our context "observed" means future events that the researcher asks respondents' expectations about and "unobserved" means any other events not inquired about. For any given student respondent i, we observe at the time of our survey (period $t = -1$, prior to college major choice):

D1) self-reported expectations of graduation with each of the K majors: $\pi_{1,i}, \ldots, \pi_{K,i}$,

[7] In his study of occupational choice, Freeman (1971), for example, assumes an adaptive expectations mapping between realized earnings and beliefs about earnings. In other occupational choice research, Siow (1984) and Zarkin (1985) make perfect foresight (rational expectations) assumptions. Implicitly these models also assume that earnings are the same for all individuals. Other work, including Bamberger (1986), Berger (1988), Flyer (1997), Eide and Waehrer (1998), Montmarquette et al. (2002), and Beffy et al. (2011) allow for some heterogeneity in earnings, across chosen and counterfactual majors, but assume rational expectations. Arcidiacono (2004) uses realized grade information during college and an assumed learning model in order to map grade measures to beliefs about ability in each major.

D2) individual beliefs about the distribution of post-graduation future events conditional on major choice $G_i^o(X^o|1,t), \ldots, G_i^o(X^o|K,t)$ for all $t = 1, \ldots, T$, and

D3) individual beliefs about ability in each of the majors $a_{1,i}, \ldots, a_{K,i}$.

$G_i^o(X^o|k,t)$ are the observed beliefs which are self-reported by respondents in the survey. The distribution of the unobserved events, covering those events not collected in the beliefs data, is given by $G_i^u(X^u|k,t)$.

Given this data, we next investigate how much of the underlying choice model can be identified. We assume that the resolvable uncertainty preference shocks for each major are distributed i.i.d. extreme value across major choices and across each individual. Note that while we assume i.i.d. taste *shocks* for each major, we place no restrictions on the time-invariant taste component $\gamma_{k,i}$, such that unobserved tastes for one major can be highly correlated with unobserved tastes for another major. Our estimates for the taste distribution (reported below) in fact show a high degree of correlation in major-specific tastes. Given we place no restriction on $\gamma_{k,i}$, the extreme value assumption on $\eta_{k,i}$ is without loss of generality in modeling the major choice since there is no parametric restriction on the combined error $\delta_{k,i} = \eta_{k,i} + \gamma_{k,i}$ (see McFadden and Train, 2000). The probability student i majors in major k from (5) is:

$$\pi_{k,i} = \frac{\exp\{\gamma_{k,i} + v(a_{k,i}) + \sum_{t=1}^{T} \beta^t \int u(X) dG_i(X|t,k)\}}{\sum_{j=1}^{K} \exp\{\gamma_{j,i} + v(a_{j,i}) + \sum_{t=1}^{T} \beta^t \int u(X) dG_i(X|t,j)\}}. \tag{6}$$

In the convenient log odds form, we can write the log odds of student i completing major k relative to a reference major $\tilde{k}$ as

$$\begin{aligned} r_{k,i} &\equiv \ln \pi_{k,i} - \ln \pi_{\tilde{k},i} \\ &= \gamma_{k,i} - \gamma_{\tilde{k},i} + v(a_{k,i}) - v(a_{\tilde{k},i}) + \beta EV_{1,k,i} - \beta EV_{1,\tilde{k},i}. \end{aligned} \tag{7}$$

Distinguishing between observed and unobserved events, we have

$$r_{k,i} = \gamma_{k,i} - \gamma_{\tilde{k},i} + v(a_{k,i}) - v(a_{\tilde{k},i}) + \beta EV_{1,k,i}^o - \beta EV_{1,\tilde{k},i}^o + \epsilon_{k,i}, \tag{8}$$

where $\epsilon_{k,i} = \beta EV_{1,k,i}^u - \beta EV_{1,\tilde{k},i}^u$,

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

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

$\epsilon_{k,i}$ represents the "error" associated with the missing information on beliefs about post-graduation events not collected in the survey. This is simply the belief data counterpart to omitted variable error in revealed preference data, e.g., "missing" information about earnings in counterfactual majors. Without loss of generality, we normalize $\gamma_{\tilde{k},i} = 0$ for all i and $E[\epsilon_{k,i}] = 0$ for all k.[8]

Collecting information about beliefs about earnings and ability has the advantage of obviating the need for assumptions mapping realized measures of ability to beliefs about ability in all fields. However, without any further modeling restrictions, we cannot separately identify the relative taste for each major γ_{ki} from the expected post-graduation future utility. The lack of identification holds since we can fully rationalize the data on expected choice probabilities as $u(X) = 0$ for any vector X and $r_{k,i} = \gamma_{k,i}$ for all $k \neq \tilde{k}$. Separately identifying $EV_{1,k,i}$ from tastes could be achieved through a parametric restriction on the joint distribution of taste parameters $\gamma_{k,i}$ (e.g., assuming a joint extreme value or normal distribution of tastes).[9] In the next section we propose a new strategy for identification using additional data derived from experimentally perturbed beliefs.

4.3 Identification using Experimental Variation

This section provides the basis for separately identifying tastes for majors from other utility components using experimental perturbations of beliefs. Our innovation is to note that if we can perturb the beliefs of the individuals so that at least some individuals form new beliefs $G_i'(X|k) \neq G_i(X|k)$, we could identify a parameterized utility function $u(X)$ without imposing parametric restrictions on the $\gamma_{k,i}$ taste components. We perturb individual beliefs by providing individuals information on general population characteristics regarding earnings and labor supply among those who have graduated with various majors (see Data section). To the extent that the individuals' self beliefs about earnings and other characteristics are i) linked to their beliefs about the population distribution of these characteristics, and ii) they are mis-informed about the population characteristics, this new information may cause some individuals to update their own self beliefs. We use our experimental data to test whether individuals are mis-informed and to examine the extent to which individuals update their own self beliefs based on this new information. As we discuss below, we find substantial misperceptions about population characteristics, and observe logical self belief updating in response to our information treatment.

An important distinction between our panel generated using experimental variation and other longitudinal information on beliefs is that we collect beliefs data over a (very) short

[8] To see that there is no loss of generality, note that the original model and the model with $\gamma_{\tilde{k},i} = 0$ for all i are equivalent: by adding the major $\gamma_{\tilde{k},i}$ taste parameter, we return to the original model as $\breve{u}(X) = \gamma_{\tilde{k},i} + u(X)$.

[9] For example, in our notation, Arcidiacono et al. (2011) assume that $\delta_{k,i} = (\eta_{k,i} + \gamma_{k,i})$ is distributed i.i.d. extreme value. We make the same parametric assumption about the resolvable uncertainty $\eta_{k,i}$, but relax this assumption for the permanent taste component $\gamma_{k,i}$. Given no restriction is placed on $\gamma_{k,i}$, the parametric assumption on $\eta_{k,i}$ places no restriction on $\delta_{k,i}$ (see McFadden and Train, 2000). Our model is then a mixed logit model which uses the experimental perturbation of beliefs to generate panel data to separately identify a taste component.

period of time, where the period before and after the information is provided in our experiment is separated by only a few minutes. This is in contrast to other studies (e.g., Lochner 2007; Stinebrickner and Stinebrickner, 2010, 2011; Zafar, 2011a) where the separation between beliefs is much longer, typically months or years. We can then credibly claim that the utility function, most notably the individual and major specific taste parameters, are truly time invariant in our context, and that our experimentally derived panel data satisfies the standard fixed effects assumptions.[10]

After providing information on the population distribution, we augment the baseline information on self beliefs (D1, D2, and D3) with the following post-treatment beliefs:

D1') post-treatment self-reported expectations of graduating with each of the K majors: $\pi'_{1,i}, \ldots, \pi'_{K,i}$,

D2') post-treatment individual beliefs about the distribution of post-graduation future events conditional on major choice $G_i^{o'}(X^o|1,t), \ldots, G_i^{o'}(X^o|K,t)$, and

D3') post-treatment individual beliefs about ability in each of the majors $a'_{1,i}, \ldots, a'_{K,i}$.

With this experimental data, using (7) we can write the individual post- minus pre-treatment difference in the log odds of majoring in each major (relative to a reference major $\tilde{k}$) as

$$r'_{k,i} - r_{k,i} = [\ln \pi'_{k,i} - \ln \pi'_{\tilde{k},i}] - [\ln \pi_{k,i} - \ln \pi_{\tilde{k},i}]$$

$$= v(a'_{k,i}) - v(a'_{\tilde{k},i}) - [v(a_{k,i}) - v(a_{\tilde{k},i})] + \beta[EV^{o'}_{1,k,i} - EV^{o'}_{1,\tilde{k},i}] - \beta[EV^{o}_{1,k,i} - EV^{o}_{1,\tilde{k},i}] + \epsilon'_{k,i} - \epsilon_{k,i}, \quad (9)$$

where $EV^{o'}_{1,k,i} = \sum_{t=1}^{T} \beta^{t-1} \int u(X) dG_i^{o'}(X^o|k,t)$. Given this structure and parameterized utility and ability functions $u(X,\theta)$ and $v(a_{k,i},\alpha)$, with finite dimensional unknown parameter vectors θ and α, we assume the following moment condition, which is the basis of our estimation strategy:

$$E[\Delta\epsilon_{k,i} | h(Z_i, \theta, \alpha)] = 0 \text{ for all } k \neq \tilde{k}, \quad (10)$$

where $\Delta\epsilon_{k,i} = \epsilon'_{k,i} - \epsilon_{k,i}$, $Z_i = [G_i^o(X|1,t), \ldots, G_i^o(X|K,t), G_i^{o'}(X|1,t), \ldots, G_i^{o'}(X|K,t)]$, and

$$\begin{aligned} h(Z_i,\theta,\alpha) &= v(a'_{k,i},\alpha) - v(a'_{\tilde{k},i},\alpha) - [v(a_{k,i},\alpha) - v(a_{\tilde{k},i},\alpha)] \\ &+ \sum_{t=1}^{T} \beta^t \int u(X,\theta) dG_i^{o'}(X^o|k,t) - \sum_{t=1}^{T} \beta^t \int u(X,\theta) dG_i^{o'}(X^o|\tilde{k},t) \\ &-[\sum_{t=1}^{T} \beta^t \int u(X,\theta) dG_i^{o}(X^o|k,t) - \sum_{t=1}^{T} \beta^t \int u(X,\theta) dG_i^{o}(X^o|\tilde{k},t). \end{aligned}$$

[10] The disadvantage of our approach relative to these other studies is of course that we cannot study the belief formation process over the long term.

Note that with our data collection, the vector of beliefs for each individual Z_i is *data* since we elicit these beliefs in our survey design.[11]

Our identification assumption states that any *changes* in beliefs about unobserved events, contained in the $\Delta\epsilon_{k,i}$ term, is mean-independent of the function of observed *changes* in beliefs given by $h(Z_i, \theta)$. Violations of the assumption would occur if experimental variation in earnings and labor supply information also affects beliefs about major characteristics we do not inquire about in our survey (e.g., unobserved beliefs about non-pecuniary aspects of a major). This would be the case if beliefs about earnings are correlated with beliefs about unobserved non-pecuniary aspects, as in a compensating differentials type framework. While we cannot test this assumption directly, our main strategy is to collect wide ranging data on a range of key post-graduation factors that could affect major choice, including information on beliefs about own earnings at different points in the life-cycle, earnings uncertainty, ability, beliefs about future marriage and spousal earnings, and intensive (expected hours per week) and extensive (expected probabilities of full or part-time employment) margins of future labor supply decisions. In addition, with our experiment-based data collection in which the pre- and post- information treatment periods are separated by only a few minutes, we can credibly claim that the $\gamma_{k,i}$ taste terms, the post-graduation utility function $u(X, \theta)$, and the current effort cost ability function $v(a_{k,i}, \alpha)$ are time invariant.[12]

5 Data

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

5.1 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

[11] Note as in the typical panel model with homogeneous elements, we do not require that ALL individuals update their beliefs, only that some individuals update their beliefs. This is because we restrict the post-graduation utility function to be homogeneous, but allow heterogeneity in fixed taste parameters. In general if we have many belief changes, we could identify rich patterns of heterogeneity in the utility function as well.

[12] A potential violation of this is if the provision of earnings information itself changes some other element of the utility function, as if the very act of providing information to students "primes" them to put more salience on this information than they otherwise would.

sent an online link to the survey (constructed using the SurveyMonkey software). The students could use any internet-connected computer to complete the survey. The students were given 2-3 days to start the survey before the link became inactive, and were told to complete the survey in one sitting. The survey took approximately 90 minutes to complete, and consisted of several parts. Students were not allowed to revise answers to any prior questions after new information treatments were received. Many of the questions had built-in logical checks (e.g., percent chances of an exhaustive set of events such as majors had to sum to 100). Students were compensated $30 for successfully completing the survey.

5.2 Survey Instrument

Our survey instrument consisted of three distinct stages. But for the purposes of estimating the choice model in this paper, we use only the initial stage self beliefs (pre-treatment) and the final stage (post-treatment) beliefs. The following summarizes the part of the survey/experiment design relevant for the choice model:

1. In the Initial Stage, respondents were asked about their population and self beliefs.
2. In the beginning of the Final Stage, respondents were given information about various statistics about the earnings and labor supply of the US population (e.g., mean earnings for all male college graduates with a degree in business or economics). Appendix Table A1 lists the information.[13] At the conclusion of the Final Stage, after having seen this information, respondents were then re-asked about their self beliefs.[14]

Our goal was to collect information on consequential life activities that would plausibly be key determinants of the utility gained from a college major. Because of time constraints, we aggregated the various college majors to 5 groups: 1) Business and Economics, 2) Engineering and Computer Science, 3) Humanities and Other Social Sciences, 4) Natural Sciences and Math, and 5) Never Graduate/Drop Out. Conditional on graduating in each of these major groups, and for different future points in time (immediately after graduation, at age 30, and at age 45), students were asked for the distribution of self earnings, the probability of marriage, labor supply, and spouse's earnings and labor supply. In addition, we collected data on the probability a student believed she would graduate with a major in each of these fields. We discuss below the specific format of some of the questions, and Section B in the Appendix provides additional information.

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

[14] In the intermediate Stage 2–not used in this paper–respondents were randomly selected to receive one of four possible information treatments shown in Table A1.

5.3 Sample Selection and Descriptive Statistics

A total of 501 students participated in the study. Our sample is constructed using the following steps. First, we drop 6 students who report that they are in the 4th year of school or higher, violating the recruitment criteria. Second, we exclude 7 individuals who report a change in graduation probabilities of greater than 0.75 in magnitude (on a 0-1 scale) in any of the 5 major categories, under the presumption that they either made errors in filling out the survey or simply did not take the survey seriously. We 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 $500,000 are recorded as $500,000. In addition, 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.

The final sample consists of 488 individual observations and 488 x 5 x 2 = 4,880 total (person x major x pre and post treatment) responses. Sample characteristics are shown in Table 1. 36 percent of the sample (176 respondents) is male, 38 percent is white and 45 percent is Asian. The mean age of the respondents is about 20, with 40 percent of respondents freshmen, 36 percent sophomores, and the remaining juniors. The average grade point average of our sample is 3.5 (on a 4.0 scale), and the students have an average Scholastic Aptitude Test (SAT) math score of 700, and a verbal score of 683 (with a maximum score of 800). These correspond to the 93rd percentile of the population score distributions. Therefore, our sample represents a high ability group of college students.

6 Reduced-Form Analysis

In this section, we describe the heterogeneity in beliefs about population average earnings and self expected earnings at age 30. We present reduced-form analysis of revisions in self beliefs following the information treatment, and document a strong and logical causal effect of our information treatment on earnings revisions. We also examine how (changes in) beliefs about own future earnings relate to (changes in) self-reported beliefs about majoring in the different fields. In the following section, we report estimates from a structural life-cycle utility model which incorporates additional elements of utility such as ability and spousal earnings.

6.1 Earnings Beliefs and Belief Updating

6.1.1 Population Beliefs About Earnings

Beliefs about population earnings were elicited as follows: "*Among all male (female) college graduates currently aged 30 who work full time and received a Bachelor's degree in each of the* *following major categories, what is the average amount that you believe these workers currently earn per year?*".

Columns (1) and (4) of Table 2 report the mean and standard deviation of respondents' beliefs about US population earnings of women and men by the 5 major fields, including the college drop-out, no degree "major". Examining first the beliefs among male students in the top panel of the table, we see that the mean male belief about age 30 female full-time earnings varies from $33,300 for college drop-outs to $79,000 for graduates with degrees in economics or business. Students believe humanities and arts majors have the lowest average earnings among the graduating majors ($54,100). Engineering and computer science graduates are believed to have earnings close to economics and business, followed by natural science majors. There is considerable heterogeneity in beliefs as indicated by the large standard deviation in beliefs about the population mean. For example, for the economics and business field, the 5th percentile of the male belief distribution in our sample is $50,000, the 50th percentile is $75,000, and the 95th percentile is $200,000.

Based on responses of students who reported population earnings for both males and females, we can construct the perceived gender gap (female - male) in earnings. This is reported in column (7) of the table. Males expect a wage gap in their favor in each of the five major fields, with the gap varying from -2.36% for natural sciences to -6.80% in college drop-out.

The lower panel of Table 2 shows that female students have beliefs similar to those of male students about relative earnings in the majors, and expect the highest average earnings in economics or business, followed by engineering and computer science, and the lowest earnings in humanities and arts among the graduating majors. However, relative to male students, female students believe average earnings to be higher in all fields for both females and males (except for male earnings for college drop-outs). Female students, like their male counterparts, perceive a wage gap in favor of men in all the fields.

Errors in Population Beliefs Columns (2) and (5) of Table 2 report the percent "error" in these beliefs relative to the information treatment "truth" we provided (see Table A1 for true population earnings that were revealed in the information treatments). We calculate errors as truth minus belief, so that a positive (negative) error indicates that the student under-estimates (over-estimates) the truth. We report both actual percent errors and the absolute value of the error. Importantly, since errors can be both positive and negative, a mean actual error close to zero may not indicate a homogenous low level of error.

Table 2 reports that the mean percent error is negative in certain categories, such as economics/business and humanities/arts, and positive in others such as engineering/computer sciences for male respondents. The errors in many categories are substantial: for example, students over-estimate full time earnings for economics and business graduates by 31.6 and 4.8 percent, depending on sub-group and sample. Reflecting the dispersion in baseline beliefs, there is considerable heterogeneity in errors, with non-trivial numbers of students making both positive and negative errors in all categories (as shown by the significantly larger mean absolute errors in columns (3) and (6) of the table).

The top two panels of Table A2 show the distribution of errors regarding full time women's and men's earnings, respectively. The heterogeneity in errors is quite striking: for example, the median error regarding full time females' earnings in engineering/computer science is +10% (that is, under-estimation of 10 percent), while the 10th percentile is -33.2% and the 90th percentile is +46.7%.

The last two columns of Table 2 show that, while both male and female students correctly perceive the wage gap to be negative, i.e., in favor of males in all fields, they substantially underestimate the wage gender gap.

6.1.2 Self Beliefs About Earnings

Next, we turn to self beliefs about *own* earnings at age 30 if the respondent were to graduate in each major.[15] The first column of Table 3 provides the average and standard deviation of the distribution of reported self earnings in our sample before the information treatment was provided. The second column of Table 3 provides the percent revision in self earnings after the information treatment. Unsurprisingly, given our high ability sample of students, the students believe their self earnings will exceed the population earnings for the US, with the average self earnings across all of the major fields higher than the corresponding average population belief about earnings reported in Table 2. Looking across majors in column (1), we see that self earnings beliefs follow the same pattern as the population beliefs, with students believing their earnings will be highest if they complete a major in the economics/business and engineering/computer science categories, and lowest if they do not graduate or graduate in a humanities and arts field.[16] There is a clear pattern of a perceived gender gap in self earnings as the average beliefs about self earnings for men exceeds those for women in most categories.

Like the population beliefs, there is substantial heterogeneity in self beliefs, as seen in the large standard deviations (relative to the means). The third panel of Table A2 shows the distribution of self earnings. Median self earnings, for example, in economics/business are

[15] For all respondents, we asked "*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?*"

[16] Table A3 provides the baseline, pre-treatment, correlation in earnings across fields. We see that for both male and female students, there is a generally high correlation in self earnings across fields: Individuals who believe they will have high earnings in one field also believe they will have high earnings in other fields. Comparing the correlations across fields, we see a higher correlations in earnings beliefs across technical or mathematical intensive fields like economics/business and engineering/computer science, compared to economics/business and humanities/arts.

$90,000, while the 10th percentile is $60,000 and the 90th percentile is $200,000.

Revisions of Self Beliefs The second column of Table 3 reports the mean and standard deviation of the distribution of percent post- minus pre- treatment changes in self beliefs about earnings. There is considerable heterogeneity in the revisions of self beliefs, with the average percent revision for the college graduation majors varying from about -12 percent (downward revision) to +28 percent (upward revision). Average revisions in the two highest earning categories –economics/business and engineering/computer science– are negative for male respondents, while average revisions in the lowest earning field –the not graduate category– are positive and large for both male and female students. As indicated by the standard deviations, within categories there is also considerable heterogeneity.[17] The third column of Table 3 shows that mean absolute revisions are substantially larger than mean revisions, varying between 20 and 59 percent (for graduating majors).

6.1.3 Self Beliefs and Population Beliefs

In the previous section, we have documented that students revise their self beliefs in response to our information treatment. The revisions we observe could be because of simple measurement error or because the students react causally to the new information the experiment provides.[18] A measurement error explanation implies no systematic relationship between the revision of individual self beliefs and individual errors in population beliefs, whereas a causal explanation implies a systematic relationship. In particular, if self earnings beliefs are based in part on the individual's beliefs about the population distribution of earnings, and if respondents are misinformed about the distribution of population earnings (of which we find evidence above in section 6.1.1), then the *sign* of the self earnings revision should match the sign of the error: positive errors (underestimation of population earnings) should cause an upward self earnings revision and negative errors should cause a downward self earnings revision. We next examine this relationship and find evidence for this type of logical updating.

Table 4 estimates a series of reduced form regressions. In the first 3 columns, we use only the baseline, pre-treatment data, and the dependent variable is the individual's (log) expected self earnings in each field. We pool all of the majors together, and in some specifications include

[17] This is further illustrated in the fourth panel of Table A2. For example, the median percentage earnings revision in economics/business for the full sample is -14.29 percent (downward revision), while the 10th percentile is -50 percent and the 90th percentile is +20 percent.

[18] Another possibility is that repeatedly asking respondents about their self earnings may prompt them to think more carefully about their responses and may lead them to revise their beliefs. See Zwane et al. (2011) for a discussion of how surveying people may change their subsequent behavior.

In addition, there could be a pure experimenter demand effect, i.e., respondents revising their beliefs upon receipt of information simply because they believe doing so constitutes appropriate behavior (Zizzo, 2010). However, in our setting this should not be a factor since the survey is anonymous and online, and respondents do not have any explicit incentive to revise their beliefs.

separate intercepts or major-specific fixed effects (dummy variables). We regress self earnings in each field on the individual's (log) belief about the population average earnings in that field. The estimates indicate that population beliefs are strongly and statistically significantly related to beliefs about self earnings. The log-log form of the regressions gives the coefficient estimates an "elasticity" interpretation: the coefficient of 0.45 in column (1) indicates that a 1 percent increase in population beliefs about average earnings increases beliefs about own earnings by 0.45 percent. The R-squared reported for the regression in the first column indicates that nearly 25 percent of the variation in self earnings beliefs is explained by population earnings beliefs. The estimated relationship is reduced slightly as we add major-specific fixed effects and covariates for individual characteristics, but continues to be precise and significantly different from zero.

Columns (4) and (5) of Table 4 examine whether the *revisions* in self-earnings are related to *errors* in population beliefs. These regressions indicate the extent to which the information treatments we provide influence individual beliefs about earnings. We regress log earnings revision in self earnings (post minus pre-treatment) on the log relative error about population earnings (log(truth/belief)). Causal revisions in response to information would imply a positive relationship between the two. In fact, the coefficient estimates are positive and statistically significant at the 1 percent level. The coefficient estimate of 0.079 indicates that a 1 percent error (under-estimation of population earnings) is associated with a 0.079 percent upward revision of self earnings. The relatively "inelastic" response of revisions in self beliefs to population errors suggests that self beliefs about earnings are not entirely linked to the type of public population information we provide. Heterogeneous private information on the abilities and future earnings prospects of individuals may cause individuals to have an inelastic response to population information. At the same time, the very precise coefficient estimates indicate that self beliefs are at least in part based on population beliefs.

As a robustness check, columns (6) and (7) report the same specifications as in columns (4) and (5) respectively, but drop outliers. More specifically, we drop observations where respondents revise their self beliefs by more than $50,000, allowing for the possibility that these may be instances where respondents made errors filling out the survey or did not take the survey seriously enough. The results are similar and continue to be significant at the 1% level.

6.2 Major Choice and Post-Graduation Utility

6.2.1 College Major Beliefs

Along with beliefs about future earnings associated with each major, respondents were also asked for their belief about the probability they would graduate with a major in each major category.[19] Table 5 provides descriptive statistics of the expected major field probabilities for male and female students. For male students, the most likely major is economics/business at 40 percent, followed by humanities/arts at 30 percent. For women, the most likely major is humanities at 50 percent followed by economics/business at 25 percent. The probability of not graduating is less than 3 percent for both men and women.[20]

Figure 1 provides the post minus pre- treatment change in log beliefs for male and female students about majoring in each field (relative to humanities): $r_{k,i} - r'_{k,i}$. The mean of the distribution of log odds changes is positive for all fields and for both male and female students (see last two columns of Table 5), indicating that after the information treatment, students on average revised upward their expected probability of majoring in non-humanities/arts fields relative to humanities/arts. However, as indicated by Figure 1, there were a substantial number of male and female respondents who revised their expected relative major choice downward, and believed they were more likely to major in humanities/arts relative to the other majors. About 1/3 of the sample reported no change in the probability of majoring in any of the fields following the information treatment. The largest upward changes occurred for the high earning fields (economics/business and engineering/computer science), especially for women. For example, for male students, the average log odds of majoring in economics/business increased by 4.6 percentage points, from pre-treatment odds of 88 percent more likely to major in economics/business relative to humanities to 92.6 percent post-treatment. For women, the log odds of majoring in economics/business relative to humanities increased 69 percentage points from -143 percent to -74 percent (negative odds indicate more likely to major in humanities/arts than economics/business). After the information treatment, women are still more likely to major in humanities/arts than economics/business, but the difference in expected probabilities declines substantially.

6.2.2 College Major Beliefs and Self Beliefs about Own Earnings

We next examine the relationship between beliefs about college major choices and future earnings. The first three columns of Table 6 estimate a series of reduced form regressions using log expected probability of majoring in each field (relative to humanities/arts) as the dependent variable and log self beliefs about earnings at age 30 (relative to humanities/arts) as the

[19] Self beliefs about the probability of graduating with a major in each of the categories were elicited as follows: "*What do you believe is the percent chance (or chances out of 100) that you would either graduate from NYU with a major in the following major categories or that you would never graduate/drop-out (i.e., you will never receive a Bachelor's degree from NYU or any other university)?*" Percent chance was converted to $(0-1)$ probabilities.

[20] Figure A1, which presents the distribution of (log) expected major field probabilities for male and female students, shows there is considerable dispersion in beliefs about future degrees. The distributions are bi-modal for most majors, with a considerable mass of individuals reporting a small or no chance of majoring in each field and another mass of individuals reporting a large or near perfect certainty of graduating in the field.

independent variable. The regressions take the form:

$$(\ln \pi_{k,i} - \ln \pi_{\tilde{k},i}) = \beta_0 + \beta_1(\ln \bar{w}_{k,i} - \ln \bar{w}_{\tilde{k},i}) + C_i'\delta + \nu_k + \omega_{k,i}, \tag{11}$$

where $\pi_{k,i}$ is i's subjective probability of graduating with major k, $\bar{w}_{k,i}$ is belief about age 30 earnings in major k, C_i is a vector of individual-specific characteristics, and ν_k is a major k fixed effect. $\tilde{k}$, the reference major in these regressions, is humanities/arts. The residual error in this cross-sectional regression ($\omega_{k,i} = \gamma_{k,i} - \gamma_{\tilde{k},i} + \epsilon_{k,i}$) consists of unobserved relative taste differences $\gamma_{k,i} - \gamma_{\tilde{k},i}$, and a component $\epsilon_{k,i}$, which reflects all other residual components.

The log-log format of these regressions gives the estimates of β_1 a "choice elasticity" interpretation. We estimate that a 1 percent increase in beliefs about self earnings in a major (relative to self earnings in humanities/arts) increases the log odds of majoring in that field (relative to humanities/arts) by about 1.6 percent. This estimate is robust to the inclusion of a wide array of individual characteristics and major fixed effects. The estimates indicate that beliefs about future relative self earnings are strongly associated with beliefs about future relative major choices: individuals appear to select into majors that they believe will provide them with the highest earnings. Importantly, because we have beliefs about earnings for all fields, this type of regression avoids the selection issue inherent in using actual major choice and the actual earnings in the chosen major since we have beliefs about earnings for all majors not chosen.

The regressions in columns (1)-(3) of Table 6 are cross-sectional based regressions using only the baseline pre-treatment beliefs. As described in the identification section, the major drawback to using only baseline beliefs is that one cannot separately identify the taste component from earnings components. In these reduced form regressions, the residual contains individual components reflecting individual variation in tastes for each of the majors. Therefore, a concern is the cross-sectional estimates of the relationship between choices and earnings could be biased if beliefs about earnings are correlated with beliefs about tastes for the majors. To resolve this problem, column (4) of Table 6 estimates the reduced form model (11) in individual (within) differences to net out the individual taste components ($\gamma_{k,i} - \gamma_{\tilde{k},i}$):

$$\begin{aligned}&[(\ln \pi'_{k,i} - \ln \pi'_{\tilde{k},i}) - (\ln \pi_{k,i} - \ln \pi_{\tilde{k},i})] \\ &= \beta_0 + \beta_1[(\ln \bar{w}'_{k,i} - \ln \bar{w}'_{\tilde{k},i}) - (\ln \bar{w}_{k,i} - \ln \bar{w}_{\tilde{k},i})] + \nu_k + \epsilon'_{k,i} - \epsilon_{k,i},\end{aligned} \tag{12}$$

where $\pi'_{k,i}$ and $\bar{w}'_{k,i}$ are post-treatment observations of choice probabilities and expected earnings. The estimates of this model are equivalent to adding individual fixed effects (FE) as individual dummy variable indicators to (11).

Using the post- and pre- treatment panel data with individual FE, we estimate the choice elasticity, with respect to beliefs about earnings, at 0.15. The FE estimate is an order of a magnitude smaller than the estimate of around 1.6 using the cross-sectional OLS estimator. The FE estimate is statistically significant at the 15 percent level (p-value of 0.144). As a robustness check, column (5) reports the FE estimate for the sample that excludes outliers – observations where respondents revise their self beliefs by more than $50,000. The FE estimate is 0.275 (statistically significant at the 5% level), and still significantly smaller than the cross-sectional OLS estimate. The FE estimates are significantly different from the cross-sectional/OLS estimates in Columns (1)-(3) at the 1 percent level. The difference between the FE/panel and OLS/cross-sectional estimates suggests that the individual tastes components are positively correlated with beliefs about earnings, and this positive correlation is severely upwardly biasing the estimates in the cross-section.

6.3 Measurement Error

Subjective data, like most data, suffer from measurement error. Therefore, one concern in using these panel estimators is that measurement error would be exacerbated using differences. Zafar (2011b) finds that most measurement error in subjective data is classical, which would tend to attenuate the coefficient estimate toward zero. However, even reasonably large measurement error would not be able to account for the very different estimates we obtain with the experimental-based FE versus the cross-sectional OLS estimates (in both Tables 4 and 6). We have also conducted sensitivity analysis for our results using a truncated sample, created by removing outliers which may represent high measurement error observations. Our results are robust to the exclusion of these outliers.

Moreover, as explained in section 6.1.3, the systematic relationship between self earnings revisions and population errors that we observe suggests that measurement error alone cannot be driving the revisions. If the responses we are receiving are purely measurement error, we would expect no systematic relationships among key beliefs. On the contrary, estimates in columns (4)-(7) of Table 4 present strong evidence of a "first stage"– that is, the revision in beliefs that we observe are a direct consequence of the information treatments. In addition, the strong relationship between beliefs about earnings and expected major choice pre-treatment that we document in section 6.2.2, and the non-zero and logical pattern in updating that we observe, where revisions (post - pre treatment) in relative earnings are *positively* correlated with major choice probability, also cast doubt on measurement error being a serious issue in the data.

Interested readers may refer to Wiswall and Zafar (2011), which presents a detailed analysis of the revision in self earnings beliefs. The companion paper shows that students revise their self earnings beliefs meaningfully in the sense that they: (1) revise their self earnings up (down) if they under-(over-) estimate population earnings, (2) revise their self beliefs more when the population errors are greater, and (3) are more responsive to the information when they have greater uncertainty about their self earnings beliefs–as would be predicted in a Bayesian updating model. While we find substantial heterogeneity in the updating heuristics that students use, the logical systematic patterns we observe provide evidence that these rich patterns of heterogeneity are not solely measurement error.

7 Structural Estimates

We next turn to estimating a structural model of major choice. In the previous sections, our analysis centered on expected future earnings at age 30. The motivation for the structural model estimation is that we can incorporate a rich set of beliefs about earnings at different points in the life-cycle, earnings uncertainty, labor supply, and spousal characteristics into a single coherent utility maximization model. The additional beliefs data we incorporate into the model include:

Lifetime Earnings Motivated by the possibility that student believe some careers have high earnings growth, we ask about full time earnings beliefs for each major at three ages: immediately after graduation, age 30, and age 45.

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 empirical literature elicits only the average returns to schooling choices (Attanasio and Kaufmann, 2011, is an exception that collects 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.[21]

Ability Since ability in each major could be a factor in expectations about future earnings, and may affect the likelihood of a student completing required coursework necessary to graduate in each major (Arcidiacono, 2004; Zafar, 2009), we ask respondents about their ability beliefs in each of the majors.[22] Appendix section C.1 provides descriptive statistics for ability rank beliefs, and revisions in ability beliefs after the information treatment.

Labor Supply To capture potential differences in work hours across majors, in addition to information about full time earnings, we also asked respondents about the expectations regarding

[21] The 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?*"

[22] 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?*"

future labor supply. For each major, we asked beliefs regarding the probability of being unemployed, working part-time, or working full time. We also asked about beliefs regarding typical full time hours for each major. The labor supply information provides additional information about potential future consumption uncertainty.

Marriage and Spousal Characteristics Motivated by recent theoretical models which have emphasized that investment in education generates returns in the marriage market (Iyigun and Walsh, 2007; Chiappori, Iyigun, and Weiss, 2009), we also collect data to investigate whether marriage market returns are a determinant of field of study. More precisely, we collect data on students' beliefs about the probability of marriage, potential spouse's earnings, and potential spouse's labor supply, conditional on *own* field of study. This allows us to provide direct evidence on whether marriage market returns are a determinant of field of study. The data are described in Appendix section C.2.

7.1 Empirical Model of Post-Graduation Utility

Our empirical specification of post-graduation utility uses discrete time and a finite horizon (periods $t = 1, \ldots, T$). Each individual from college graduation to retirement makes a series of decisions regarding labor supply and marriage. At college graduation, we assume each individual is single and has obtained a degree in particular field $k = 1, \ldots, K$.

In defining the utility function, we distinguish between two states: married and single. The flow utility in period t if the agent is single is given by $U_{S,t} = u_S(c_{S,1,t})$, where $c_{S,1,t}$ is the individual's period t consumption. The own utility for an individual if married is given by $U_{M,t} = u_M(c_{M,1,t}, c_{M,2,t})$, where $c_{M,1,t}$ is consumption of the individual and $c_{M,2,t}$ is the consumption of the individual's spouse. $U_{M,t}$ defines the *own* utility flow in period t from being married, not the household total utility for both spouses. Our specification of the utility function allows for the possibility that the individual may derive utility from the consumption of his or her spouse. Flow utility over the two states is then given by $U_t = m_t U_{M,t} + (1 - m_t)U_{S,t}$, where $m_t = 1$ indicates marriage, and $m_t = 0$ indicates single status at period t.

We specify the utility functions with CRRA forms. When single, the utility function is given by $u_S(c_{S,1,t}) = \phi_1 \frac{c_{S,1,t}^{1-\rho_1}}{1-\rho_1}$, with $\phi_1 \in (0, \infty)$ and $\rho_1 \in (0, \infty)$. $1/\rho_1$ is the intertemporal elasticity of substitution (IES) for own consumption and ρ_1 is the coefficient of relative risk aversion. When married, we specify a commonly used specification where utility is a sum of own and spouse's utility: $u_M(c_{M,1,t}, c_{M,2,t}) = u_{M,1}(c_{M,1,t}) + u_{M,2}(c_{M,2,t})$.

Own utility while married uses the same preference structure while single (although the consumption level may be different under marriage, as we describe below): $u_{M,1}(c_{M,1,t}) = \phi_1 \frac{c_{M,1,t}^{1-\rho_1}}{1-\rho_1}$. Since we are modeling only the utility of a given individual, we specify the utility of the individual over her spouse's consumption, i.e. we allow the individual to be altruistic toward his spouse. The preferences of the individual over his spouse's consumption are allowed to be different from his preferences over his own consumption: $u_{M,2}(c_{M,2,t}) = \phi_2 \frac{c_{M,2,t}^{1-\rho_2}}{1-\rho_2}$, with $\phi_2 \in (0,\infty)$ and $\rho_2 \in (0,\infty)$. $1/\rho_2$ provides the IES for spouse's consumption.[23]

We use the individual's self beliefs about own earnings and labor supply and use the individual's self beliefs about potential spousal earnings and labor supply to define consumption levels under the single and married states. We do not model borrowing and savings and assume consumption in each period is equal to current period earnings.[24] Because we ask individuals about full time equivalent earnings, we combine the beliefs about labor supply and full time earnings to define earnings in any given period. Own and spousal earnings are modeled as $y_{1,t} = w_{FT,1,t}FT_{1,t} + w_{FT,1,t}(h_{PT,1,t}/h_{FT,1,t})PT_{1,t}$ and $y_{2,t} = w_{FT,2,t}FT_{2,t} + w_{FT,2,t}(h_{PT,2,t}/h_{FT,1,t})PT_{2,t}$, where $w_{FT,q,t}$ are full time earnings ($q = 1$ own, $q = 2$ spouse), $FT_{q,t} \in \{0,1\}$ is an indicator if working full-time, $PT_{q,t} \in \{0,1\}$ is an indicator for working part-time, $h_{FT,q,t}$ is full time hours, and $h_{PT,q,t}$ is part time hours. For each potential major, we ask respondents for their beliefs about the probability of working full or part-time, if single or married, the probability their potential spouse works full or part-time if married, and beliefs about average hours of work for each major. We allow an individual's beliefs about the future distribution of full-time and part-time probabilities to depend on marriage, and therefore earnings and consumption also depend on marriage.

Consumption conditional on marriage is then given by $c_{S,1,t} = y_{1,t}$ (own consumption when single), $c_{M,1,t} = \kappa_1(y_{1,t} + y_{2,t})$ (own consumption when married), and $c_{M,2,t} = (1-\kappa_1)(y_{1,t} + y_{2,t})$ (spousal consumption when married). $\kappa_1 \in (0,1)$ is the share parameter which indicates how much of total household earnings is consumed by each spouse.[25]

7.2 Estimation

We estimate the parameters of the utility function using the pre- and post- information beliefs. Because of time limitations, we were forced to ask a limited set of questions: we cannot ask respondents to report full time earnings for all post-graduation periods and we cannot ask an

[23] We have experimented with utility specifications that also include a term for leisure and have estimated these functions using our data on beliefs about future own labor supply and future spouse's labor supply. We have found that the parameters of this specification are only weakly identified and the estimation is generally unstable.

[24] We have two alternatives in adding borrowing and savings behavior to a model such as this. First, following the earnings and labor supply questions, one could directly ask respondents about future consumption, borrowing, savings, or asset levels. However, framing these types of questions in a meaningful way for respondents may be quite difficult. Second, one could use traditional observational data to estimate a model of borrowing and saving and combine this model with the current model allowing consumption to be endogenous given earnings and labor supply.

[25] We have also experimented with functions that allow public goods, such that consumption of each spouse when married can exceed total resources. In some preliminary estimation, we found that these more general models were at best only weakly identified.

infinite number of questions in order to provide a non-parametric estimate of the distribution of beliefs. Section D in the Appendix describes our approximations of the full life-cycle beliefs from the given data. It is important to emphasize that these approximations of beliefs are entirely individual-specific: we make no assumption regarding the distribution of beliefs in the population.

We calculate expected utility from (8) using simulation. For computing the expected utility for a parameterized utility function $u(X, \theta)$ defined over X events and finite parameter vector θ, we take R draws from each individual's belief distribution and compute expected utility for individual i as

$$EV_{1,k,i} = \sum_{t=1}^{T} \beta^{t-1} \frac{1}{R} \sum_{r=1}^{R} u(x_r, \theta)$$

where $x_1, \ldots, x_R$ are R draws from individual i's distribution of observed beliefs $G_i^o(X|k,t)$.

The estimator is based on the moment condition (10). Using the within post-pre treatment difference, the non-linear least squares (NLS) estimator for θ and α is given by:

$$(\hat{\theta}, \hat{\alpha}) = \arg\min \sum_{i=1}^{N} \sum_{k=1}^{K} [(r'_{k,i} - r_{k,i}) - \{h(Z_i, \theta, \alpha)\}]^2 \tag{13}$$

where $h(Z_i, \theta, \alpha)$, defined in (10), is a non-linear function of parameters. The utility function parameters to be estimated include $[\rho_1, \psi_1, \rho_2, \psi_2]$. We set $\kappa_1 = 1/2$ as we found it difficult to separately identify the consumption share parameter from parameters governing the marginal utility of consumption. The ability function is parameterized as $v(a) = \alpha \ln a$. β is assumed to be 0.95 and $T = 55$. The combined parameters then consist of the taste for each major $\gamma_1, \ldots, \gamma_K$ and the post-graduation utility function parameters θ. We estimate the model separately for male and female students and allow for entirely different utility function parameters for males and females.[26]

7.3 Model Estimates

Table 7 provides the parameter estimates for two versions of the structural model. Model 1 is our main model. The marginal utility of own consumption (when single) is given by $\phi_1 c_{S,1,t}^{-\rho_1}$. We estimate ϕ_1 to be 0.22 for male students and 0.20 for female students, and the curvature parameter (relative risk aversion) ρ_1 to be 4.48 for males and 5.51 for females. Both estimates are on the high end of previous estimates, but similar to the estimate in Nielsen and Vissing-Jorgensen (2006). The larger estimate of relative risk aversion for females (statistically different from the male coefficient at the 10% level) is consistent with several studies that conclude that

[26] In the estimation we also include a vector of revision fixed effects/intercepts that capture any mean differences in revisions by major.

women are more risk averse than men in their choices (Eckel and Grossman, 2008; Croson and Gneezy, 2009). The high ρ estimates, especially for women, could be driven by the fact that our sample reports very high probabilities of completing a degree in humanities (Table 5), and humanities is one of the fields with the lowest reported uncertainty in earnings. Own value of spouse's consumption has values of ϕ_2 and ρ_2 which indicate the utility value of spouse's consumption has less curvature than own consumption. The coefficient on log ability rank is around 0.10 for both male and female students.

With the estimated parameters of the utility and ability functions, we can use the pre- and post- treatment choices to estimate each individual's taste for each major (relative to humanities/arts), given by $\gamma_{k,i}$. Table 8 provides statistics for the distribution of the estimated $\gamma_{k,i}$ taste parameters (relative to humanities/arts which is normalized to 0). We see a distinct gender difference in tastes: On average, male students have a strong taste for economics/business majors over humanities/arts (positive $\gamma_{k,i}$), but average tastes for female students are negative for all majors, indicating a strong preference for humanities/arts over all other fields. Interestingly, the median male taste for economics/business majors is negative and close to zero, indicating a skewed taste distribution.[27]

Next, we assess the fit of the estimated models and compare the estimates to the reported major choice probabilities in the data. Table 9 computes the predicted probabilities of major choice using the estimated parameters from our main model. The model fits the choice probabilities quite well, for both males and females, with only slight deviations between predicted model probabilities and those from the actual data.

7.4 Using Cross-Sectional Data Only

We also estimate a second model using only the cross-sectional data and assuming a parametric distribution for college major tastes. The estimates of this model are intended to illustrate the "value added" of our panel data information experiment which allows us to flexibly estimate the distribution of unobserved tastes. For this restricted model, we assumed that the college major taste terms γ_k are distributed Type 1 extreme value with gender and major specific means. We estimated this model using only the pre-treatment data, thereby forming a cross-sectional dataset. This is essentially the same type of parametric taste restriction and data structure as Arcidiacono et al. (2011), although we use our life-cycle consumption utility specification and our data on own earnings and hours, ability, marriage, and spousal earnings and hours. The estimates for this model are reported in the last column of Table 7. We estimate a larger degree of relative risk aversion for males, but not for females, and a much higher own marginal utility

[27] Figure A2 provides a direct look at the distribution of tastes for majors for men and women, respectively. Both distributions show some bimodality, but the most frequent mode for the male students' tastes distribution is near 0, whereas the mode for the female students' tastes distribution is negative.

of earnings for women. Another key difference is that the cross-sectional model has several times larger estimates for the ability component than with the unrestricted, panel data model. It is interesting to note that while the key differences in the models is how flexibly the taste component is modeled, this modeling difference also substantially affects the estimates of the other parameters. We further explore this below.

7.5 Choice Elasticities

The estimates are most interpretable in terms of what the estimated models imply about the responsiveness of major choices to changes in self earnings. For each major, we increase beliefs regarding own earnings by 1 percent in every period. How much more likely would individuals be to major in each major due to this increase in earnings? We compute choice elasticities given by

$$\xi_{k,i} = \frac{\partial \pi_{k,i}}{\partial w_{FT,1,t}} \frac{w_{FT,1,t}}{\pi_{k,i}} \times 100.$$

Note that these choice elasticities depend on the estimated utility function parameters, and given the non-separability of tastes, abilities, and $u(X, \theta)$, also depend on the distribution of tastes and abilities.

Figure 2 graphs the distribution of the $\xi_{k,i}$ choice elasticities in our samples of male and female students. The first two columns of Table 10 report the mean of this distribution using Model 1 (panel data). A value of $\xi_{k,i} = 0.1$ indicates that individual i would increase her probability of majoring in major k by 0.1 percent for a 1 percent increase in own earnings each period. From Figure 2 it is clear that there is substantial heterogeneity in the responsiveness of individuals to changes in earnings. While some individuals would have a near zero response to the change in earnings, other individuals would have a substantial, albeit inelastic, response.

Table 10 reveals that the average response to earnings changes is higher for male students in all majors than for women. The overall mean elasticity is considerably higher in the not graduating field than in the other fields. This may be due to the relatively low level of expected earnings in this major and the estimated concavity of the utility function with respect to consumption. Our results of a relatively low response to changes in earnings is consistent with other studies using observational data (Arcidiacono, 2004; Beffy et al., 2011). Beffy et al. (2011), using data on French students, estimate earnings elasticities of between 0.09-0.12 percentage points, depending on the major. This compares favorably to the mean earnings elasticities we estimate (excluding drop-out alternative) of between 0.037-0.094, depending on major and gender.

Table 10, in the last two columns, also shows the estimated choice elasticities under the alternative model specification using only the pre-treatment, cross-sectional data, with the assumed parametric distribution for tastes. Consistent with the simple reduced form results above, the choice elasticities for most majors here are several times larger than when using the panel data with an unrestricted taste component. This emphasizes one of our main conclusions: Cross-sectional data, even incorporating rich belief data on a wide variety of beliefs, would substantially over-state how sensitive individuals are to changes in earnings.

7.6 Correlates of Tastes

In the preceding analysis, the $\gamma_{k,i}$ taste components are essentially a "black box." Tastes are inferred or "backed out" from expectations data and model estimates and allowed to have any relationship with other model components, such as earnings. We next investigate the correlates of major-specific tastes. Table 11 reports the OLS estimates of a series of regressions of tastes for each major major (relative to humanities/arts) onto various demographic characteristics and ability measures.

Four patterns are of note. First, tastes for all the fields are positively (negatively) correlated with SAT Math (Verbal) scores. This is consistent with the ability sorting patterns documented in for example, Arcidiacono (2004), who finds that natural science majors have the highest SAT Math scores, and that SAT Verbal scores are very high for humanities majors. This indicates that tastes for majors are correlated with ability, and that students with higher math ability exhibit stronger tastes for the non-humanities/arts majors.

Second, relative to females, males have significantly stronger positive tastes for all the other major categories (relative to humanities/arts). While we investigate the importance of tastes in the choice of major field for the two genders below, this suggests that markedly different tastes for majors may explain gender differences in college major choice (Brown and Corcoron, 1997; Weinberger, 1998; Wiswall, 2006; Zafar, 2009). Third, the coefficient for Asian respondents is significantly positive for all major categories, indicating a dis-taste for humanities/arts. This suggests that the background factors related to Asian race also influence the formation of tastes for majors.

In our framework, tastes also reflect switching costs. As students progress through college, it may become more costly for them to switch majors. The fourth notable pattern is that, for some majors, we observe systematic patterns in the coefficients on Sophomore and Junior indicators. The coefficient on "Junior" is significantly negative for engineering/computer science and natural sciences, i.e., students in their junior year, relative to freshmen (and sophomore) respondents, have significantly more negative tastes for these fields. This is consistent with (i) evidence that suggests that learning (about ability and tastes) in college is primarily concentrated in the math/science majors (Stinebrickner and Stinebrickner, 2011), and (ii) patterns of major switches that indicate that students switch out of math, science, and engineering (Stinebrickner and Stinebrickner, 2011; Arcidiacono, 2004).[28]

Overall we find that tastes are correlated with ability, gender, race, and school year. These results have strong implications for the modeling of tastes in choice models. Under prevalent approaches, tastes are generally assumed to be orthogonal to everything else in the model. The strong correlation of tastes with observables implies that such modeling assumptions may yield biased estimates. Second, observables explain only about 20% of the variation in tastes. Therefore, our approach of allowing tastes to follow any distribution is robust relative to other approaches which restrict the distribution of tastes to a particular parametric distribution.

7.7 Decomposition of the Determinants of College Major Choices

We next use the estimated unrestricted model to decompose the college major 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 major choice is equally likely. We accomplish this by setting each respondent's beliefs (about earnings, ability, hours of work, marriage, and spousal characteristics, i.e. spousal earnings and hours) and their tastes for each major equal to the corresponding level for the humanities/arts major. Therefore, at the baseline, the odds of majoring in each of the remaining majors (relative to humanities/arts) is $\pi_{k,i}/\pi_{\tilde{k},i} = 1$. After establishing this baseline, we then progressively re-introduce each individual's major-specific beliefs and tastes into the estimated choice model in order to capture the marginal contribution of each component. The magnitude by which the relative odds of majoring in each field changes as we add a component measures the importance of this component. Table 12 reports the choice probability at each stage of the decomposition averaged over all of the sample respondents.

7.7.1 Male Students

In the first panel, we decompose major choices for male students only. Focusing on the first row, we see that re-introducing each individual's beliefs about his own earnings in each major increases the average odds of majoring in economics/business (relative to humanities/arts) from the baseline of 1 to 1.057, or a +0.057 marginal increase in odds. The increase in the average odds of majoring in economics/business reflects the earnings advantage most individuals perceive from graduating with an economics/business degree, evaluated at the estimated utility function parameters. In contrast, adding self beliefs about own earnings reduces the odds of

[28] We also estimate these regressions separately for the male and female subsamples. The negative coefficient in engineering/computer science is driven by male respondents (that is, junior male students have a dis-taste for engineering/computer science majors, relative to freshmen). This is consistent with male students being (excessively) more confident than female students at the outset (Weinberger, 2004; Niederle and Vesterlund, 2007).

not graduating from a baseline of 1 to 0.87 (-0.13 reduction). Incorporating individual earnings beliefs implies that individuals are less likely to believe they will choose "not graduate" given lower expected earnings from not graduating.

Columns (2) through (5) progressively add other model components, and the entries in Table 12 reflect the marginal gain of each component, given the other preceding components are included. Thus, adding beliefs about own ability in Column (2) only slightly increases the odds of majoring in economics/business from 1.057 (including beliefs about own earnings) to about 1.059 (including both beliefs about own earnings and own ability). One reason for this is that the high positive correlation of beliefs about earnings and ability implies that the marginal contribution of each is rather small. The marginal contribution of ability has the largest negative effect on majoring in engineering/computer science. The negative sign on the own ability components indicates that individuals perceive higher "study effort" due to either lower ability or greater difficulty in engineering/computer science relative to humanities/arts, and thus this factor reduces the odds of majoring in engineering/computer science.

Column (3) of Table 12 re-introduces beliefs about own work hours for each major. Because higher work hours increase total earnings (and there is no disutility from work), this tends to increase the odds of majoring in economics/business the most, and tends to reduce to the odds of not graduating, given beliefs of higher unemployment spells with this major.

Column (4) adds spousal characteristics, including the probability of marriage, spousal earnings, and spousal hours. The column indicates the marginal contribution of beliefs about gains in the marriage market from choosing different majors. These gains are positive and highest for economics/business but negative for not graduating.

Finally, Column (5) adds the remaining determinant of major choice, the vector of estimated major-specific tastes. Tastes have a substantial effect on choice to major in economics/business, increasing the log odds by 0.310. For males, tastes in this case then complement the other positive contributions to choosing the economics/business major. However, tastes have a large and negative effect on choosing the other majors. The negative sign on this component indicates that, on average, male students have high dis-taste for these majors (relative to humanities/arts). But the high negative taste is offset somewhat, with the exception of the not graduate category, by the positive contribution from own earnings and spousal characteristics.

7.7.2 Female Students

The second panel of Table 12 calculates the decomposition for female students. In comparing the male and female decompositions, it is clear that own earnings differences are a smaller factor in college major choice for women than men. For ability, the reverse is true as ability differences across majors are more important for women than men. For women, the negative component from ability, reflecting lower perceived ability in these majors relative to humanities/arts, more than offsets the positive earnings advantage. This was not true for men as the ability component, with the exception of engineering/computer science, is quite minor relative to the earnings component.

For the other components, own hours and spousal characteristics play relatively small marginal roles, with the exception of the not graduate category, where beliefs about poor spousal characteristics reduces the probability of not graduating for female students. As with male choices, the taste component is large. This suggests that while the other determinants of college major choices–including earnings and ability–are meaningful, the taste component at the time of college major decision-making is dominant.

Column (4) shows that including spousal characteristics does not change the log-odds for graduating majors, but decreases the log-odds for the not graduate category. This suggests that returns in the marriage market are generated by simply going to college, and the college major itself does not matter much in this aspect.

7.7.3 Gender Ratio

The last panel of Table 12 directly assesses the contribution of the model components to the ratio of female to male major choices. Women are considerably more likely to major in humanities/arts than other majors: In our sample (before the information treatment), the average female probability of majoring in humanities is 0.5, compared to 0.3 for men. The last panel of Table 12 calculates the relative odds for women versus men for each major (relative to humanities/arts):

$$\frac{\pi_{k,i}(\text{women})/\pi_{\tilde{k},i}\text{women}}{\pi_{k,i}(\text{men})/\pi_{\tilde{k},i}\text{men}}.$$

In the pre-treatment sample, this ratio for economics/business is 0.39, reflecting that women are less likely to major in economics/business relative to humanities/arts than men. As with the previous decomposition, we start with a baseline in which men and women are equally likely to choose all majors, and hence the female-male odds ratio is 1. In column (1) we see that adding beliefs about own earnings begins to create a gap between men's and women's college major choices. Adding earnings beliefs, reduces the economics/business female-male ratio from 1 to 0.980 (-0.019 marginal reduction). Similar negative reductions are evident for engineering/computer science and natural sciences. This increase in the gap between men and women occurs because men have generally higher earnings beliefs in these fields relative to humanities/arts than women (column (1) of Table 3). The exception is the not graduate category in which the female-male ratio actually increases to a female advantage from 1 at the baseline to 1.030 (+0.0297 marginal gain).

In Column (2), we see that ability differences between men and women cause a further increase in the gender gap in major choice. Differences in beliefs about ability exacerbate the tendency for men to major in non-humanities subjects more than women. This is because men have higher ability beliefs in these majors relative to humanities/arts than women (see column (1) of Table A4). On the other hand, gender differences in beliefs about own hours and spousal characteristics have only a minor effect on the gender gap. Finally, in Column (5), adding gender differences in major-specific tastes substantially increases the gender gap. This finding suggests that pre-college determinants of tastes, as distinct in our framework from beliefs about earnings, ability, hours, and spousal characteristics, causes the majority of the gender difference in college major choices.

8 Conclusion

This paper seeks to shed light on the determinants of college major choice. While there is a recent and growing literature that uses subjective expectations data to understand schooling choices, our approach is unique in several ways. First, our survey has an innovative experimental feature embedded in it, which generates a panel of beliefs. We show that this experimental variation in beliefs can be used to identify the distribution of tastes non-parametrically. Second, we collect data on earnings uncertainty, which are usually not available in observational (and for the most part, in subjective) data. Third, instead of using indirect proxies, we provide the first direct evidence of the role of marriage market returns on schooling choice.

We find that, in the context of major choice, earnings differences across majors is a more important factor for men than women, and ability differences matter more for women than men. However, tastes for majors are a dominant factor for both males and females. Even accounting for other characteristics such as earnings, labor supply, and ability, we find that females have a strong taste for humanities/arts while male students have a strong relative taste for economics/business. We also estimate substantial heterogeneity in tastes within gender, with the distribution of relative tastes estimated to be bimodal. In our framework, "tastes" are defined at the point when students are in college. These could be tastes for major-specific outcomes realized in college, such as the enjoyability of coursework, or major-specific post-graduation outcomes, such as non-pecuniary aspects of jobs. It is important to note that tastes in our framework are distinct from ability, though they may be correlated with them (which we do find to be the case). We present evidence that the dis-taste for humanities is stronger for male, Asian, and high-SAT Math score respondents. Differences in tastes may arise exogenously because of innate differences (Kimura, 1999; Baron-Cohen, 2003), or they may be endogenously determined by earlier interactions with peers and parents (Altonji and Blank, 1999). Understanding the originations of differences in tastes is not investigated in the current study, and is an important area of future research.

The innovation of our study is that we experimentally shift beliefs to generate within-individual variation in expected earnings and probabilistic choices across majors. Such within-individual variation in earnings and choices is never available. Most of the literature assumes rational expectations and other assumptions to generate variation in earnings across people and majors, which is then used to identify the importance of earnings. Moreover, papers that do have within-individual variation in earnings (such as Stinebrickner and Stinebrickner, 2010 and 2011) have it at different points in time separated by several months, so it is unlikely that other determinants of major choice have remained fixed over that horizon. Therefore, while it would be useful to follow-up on students to observe the impact of information on actual choices (as in Jensen, 2010) as a validation exercise, what we could learn from actual choice data is not a substitute for our study. Given that major choice is a one-time decision – once individuals enter the labor market, their choice of major is generally irreversible – we would have to invoke certain assumptions to generate variation in earnings across individuals and majors (in addition to making assumptions on the expectations process), in order to estimate earnings elasticities of fields of study from choice data. Thus, we believe that the approach used in this study has certain advantages over choice data. Moreover, expectations data have been shown to be strong predictors of actual choices (Jacob and Wilder, 2010).

Our survey respondents, despite consisting of a group of high ability students enrolled at an elite university, have biased beliefs about the distribution of earnings in the population, but revise their self beliefs and choices sensibly when provided with accurate information. These results suggest a policy role for information campaigns focused on providing accurate information on returns to schooling.[29] While such campaigns have been conducted in developing countries (Jensen, 2010; Nguyen, 2010), our results make a case for such interventions in developed countries as well.[30] However, in order to understand the underlying determinants of choice behavior and the channels through which such interventions affect behavior, our results also suggest that such interventions should be accompanied with collection of rich data on subjective expectations.

References

[1] Altonji, Joseph. 1993. "The Demand for and Return to Education when Education Outcomes are Uncertain." Journal of Labor Economics, 11: 48-83.

[29] However, it should be pointed out that how students revise their beliefs and choices in an experimental framework like ours where the information is presented to them may be very different from the change in their behavior where they acquire the information themselves. While it is challenging to identify changes in information sets in actual panels (Zafar, 2011a), an important question for future research is to explore how students' beliefs and choices evolve over longer time horizons, and how persistent the impact of revealed information is on students' behavior.

[30] One study that we are aware of in a developed setting is that of Bettinger et al. (2011) who find that providing information on financial aid and assistance in filling out federal financial aid forms improves college access.

[2] Altonji, Joseph, and Rebecca Blank. 1999. "Race and Gender in the Labor Market." In Handbook of Labor Economics, Volume 3c, ed. Orley Ashenfelter and David Card, 3144-3259. Elsevier Science.

[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. 2011. "Modeling College Major Choices using Elicited Measures of Expectations and Counterfactuals." Journal of Econometrics, forthcoming.

[5] Attanasio, Orazio, and Katja Kaufmann. 2011. "Education Choices and Returns on the Labour and Marriage Markets: Subjective Expectations, Gender and Decision Making." Working Paper.

[6] Bamberger, Gustavo. 1986. "Occupation Choice: The Role of Undergraduate Education." Ph.D. Dissertation, University of Chicago.

[7] Baron-Cohen, Simon. 2003. The Essential Difference: Men, Women, and the Extreme Male Brain. London: Allen Lane.

[8] Beffy, Magali, Denis Fougere, and Arnaud Maurel. 2011. "Choosing the Field of Study in Post-Secondary Education: Do Expected Earnings Matter?" The Review of Economics and Statistics, Forthcoming.

[9] Berger, Mark. 1988. "Predicted Future Earnings and Choice of College Major." Industrial and Labor Relations Review, 41(3): 418-29.

[10] Bettinger, Eric, Bridget Long, Philip Oreopoulos, and Lisa Sanbonmatsu. 2011. "The Role of Simplification and Information in College Decisions: Results from the H&R Block FAFSA Experiment" Working Paper.

[11] Betts, Julian. 1996. "What do Students Know about Wages? Evidence from a Survey of Undergraduates." Journal of Human Resources, 31(1): 27-56.

[12] 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.

[13] Brown, Charles, and Mary Corcoran. 1997. "Sex Based Differences in School Content and the Male-Female Wage Gap." Quarterly Journal of Economics, 99: 31-44.

[14] Chiappori, Pierre-Andre, Murat Iyigun, and Yoram Weiss. 2009. "Investment in Schooling and the Marriage Market." American Economic Review, 99(5): 1689-1713.

[15] Croson, Rachel, and Uri Gneezy. 2009. "Gender Differences in Preferences." Journal of Economic Literature, 47(2): 448–74.

[16] Dey, Judy, and Catherine Hill. 2007. "Behind the Pay Gap." American Association of University Women Educational Foundation Research Report.

[17] Delavande, Adeline. 2008. "Pill, Patch or Shot? Subjective Expectations and Birth Control Choice." International Economic Review, 49(3): 999-1042.

[18] Eckel, Catherine, and Philip Grossman. 2008. "Differences in the Economic Decisions of Men and Women: Experimental Evidence." In Handbook of Experimental Economics Results, ed. Charles Plott and Vernon Smith, 509-519. New York: Elsevier.

[19] 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.

[20] Flyer, Fredrick. 1997. "The Influence of Higher Moments of Earnings Distributions on Career Decisions." Journal of Labor Economics, 15(4): 689-713.

[21] Freeman, Richard. 1971. The Market for College-Trained Manpower: Study in the Economics of Career Choice. Cambridge: Harvard University Press.

[22] Ge, Suqin. 2010. "Women's College Decision: How Much Does Marriage Matter?" Journal of Labor Economics, forthcoming.

[23] Gemici, Ahu, and Matthew Wiswall. 2011. "Evolution of Gender Differences in Post-Secondary Human Capital Investments: College Majors at the Intensive Margin." Working Paper.

[24] Grogger, Jeff, and Eric Eide. 1994. "Changes in College Skills and the Rise in the College Wage Premium." Journal of Human Resources, 30(2): 280-310.

[25] Giustinelli, Pamela. 2010. "Uncertain Outcomes and Child-Parent Decision Making in Curriculum Choice: What Data Do We Need to Tell Them Apart?" Working Paper.

[26] 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.

[27] Jacob, Brian, and Tamara Wilder. 2010. "Educational Expectation and Attainment." NBER Working Paper 15683.

[28] Jensen, Robert. 2010. "The (Perceived) Returns to Education and the Demand for Schooling." Quarterly Journal of Economics, 125(2): 515-548.

[29] Kaufmann, Katja. 2009. “Understanding the Income Gradient in College Attendance in Mexico: The Role of Heterogeneity in Expected Returns to College.” Working Paper.

[30] Kimura, Doreen. 1999. Sex and Cognition. Cambridge: MIT Press.

[31] Lafortune, Jeanne. 2010. "Making Yourself Attractive: Pre-Marital Investments and the Returns to Education in the Marriage Market." Working Paper.

[32] Lochner Lance. 2007. "Individual Perceptions of the Criminal Justice System." American Economic Review, 97(1): 444–460.

[33] Manski, Charles. 1993. "Adolescent Econometricians: How Do Youth Infer the Returns to Schooling?" in Studies of Supply and Demand in Higher Education, ed. Charles Clotfelter and Michael Rothschild, 43-57. Chicago: University of Chicago Press.

[34] Manski, Charles. 2002. "Identification of Decision Rules in Experiments on Simple Games of Proposal and Response." European Economic Review, 46(4-5): 880-891.

[35] Manski, Charles. 2004. "Measuring Expectations." Econometrica, 72(5): 1329-1376.

[36] Mare, Robert. 1991. “Five Decades of Educational Assortative Mating.” American Sociological Review, 56(1): 15-32.

[37] McFadden, D., and Kenneth Train. 2000. ““Mixed MNL Models of Discrete Response." Journal of Applied Econometrics, 15: 447-470.

[38] Montmarquette, Claude, Kathy Cannings, and Sophie Mahseredjian. 2002 “How Do Young People Choose College Majors?” Economics of Education Review, 21(6): 543-556.

[39] Nielsen, Helena, and Annette Vissing-Jorgensen. 2006. "The Impact of Labor Income Risk on Educational Choices: Estimates and Implied Risk Aversion." Working Paper.

[40] Nguyen, Trang. 2010. “Information, Role Models and Perceived Returns to Education: Experimental Evidence from Madagascar,” MIT Working Paper.

[41] 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.

[42] Paglin, Morton, and Anthony Rufolo. 1990. "Heterogeneous Human Capital, Occupational Choice, and Male-Female Earnings Differences." Journal of Labor Economics, 8: 123-144.

[43] Pencavel, John. 1998. “Assortative Mating by Schooling and the Work Behavior of Wives and Husbands.” American Economic Review, 88(2): 326-329.

[44] Saks, Raven, and Steven Shore. 2005. “Risk and Career Choice.” Advances in Economic Analysis & Policy, 5(1).

[45] Siow, Aloysius.1984. “Occupational Choice under Uncertainty.” Quarterly Journal of Economics, 52(3), 631–45.

[46] Stinebrickner, Todd, and Ralph Stinebrickner. 2010. "Learning about Academic Ability and the College Drop-out Decision." University of Western Ontario Working Paper 20086.

[47] Stinebrickner, Todd, and Ralph Stinebrickner. 2011. "Math or Science? Using Longitudinal Expectations Data to Examine the Process of Choosing a College Major." Working Paper.

[48] Turner, Sarah, and William Bowen. 1999. “Choice of Major: The Changing (Unchanging) Gender Gap.” Industrial and Labor Relations Review, 52(2): 289-313.

[49] van der Klaauw, Wilbert, and Kenneth Wolpin. 2008. "Social Security and Savings and Retirement Behavior of Low Income Households." Journal of Econometrics, 145(1-2): 21-42.

[50] van der Klaauw, Wilbert. 2011. "On the Use of Expectations Data in Estimating Structural Dynamic Choice Models." Working Paper.

[51] Weinberger, Catherine. 1998. "Race and Gender Wage Gaps in the Market for Recent College Graduates." Industrial Relations, 37(1): 67-84.

[52] Weinberger, Catherine. 2004. "Just Ask! Why Surveyed Women Did Not Pursue Information Technology Courses or Careers?" IEEE Technology and Society, 23(2): 28-35.

[53] Wiswall, Matthew. 2006. "Skill Composition and the Returns to Post-Secondary Education in the United States." Working Paper.

[54] Wiswall, Matthew, and Basit Zafar. 2011. "Belief Updating among College Students: Evidence from Experimental Variation in Information." Staff Report No. 516, Federal Reserve Bank of New York.

[55] Wolpin, Kenneth. 1999. "Commentary on "Analysis of Choice Expectations in Incomplete Scenarios", by C. F. Manski." Journal of Risk and Uncertainty, 19(1-3): 67-69.

[56] Zarkin, Gary. 1985. “Occupational Choice: An Application to the Market for Public School Teachers.” Quarterly Journal of Economics, 100(2), 409–446.

[57] Zafar, Basit. 2009. "College Major Choice and the Gender Gap." Staff Report No. 364, Federal Reserve Bank of New York.

[58] Zafar, Basit. 2011a. "How do College Students Form Expectations." Journal of Labor Economics, 29(2): 301-348.

[59] Zafar, Basit. 2011b. "Can Subjective Expectations Data be Used in Choice Models? Evidence on Cognitive Biases." Journal of Applied Econometrics, 26(3): 520-544.

[60] Zizzo, Daniel. 2010. "Experimenter demand effects in economic experiments." Experimental Economics, 13(1): 75-98

[61] Zwane, A., J. Zinman, E.Van Dusen, W. Pariente, C. Null, E. Miguel, M. Kremer, D. Karlan, R. Hornbeck, X. Giné, E. Duflo, F. Devoto, B. Crepon and A. Banerjee. 2011. "Being surveyed can change later behavior and related parameter estimates." Proceedings of the National Academy of Sciences, 10(1073): 1-6.

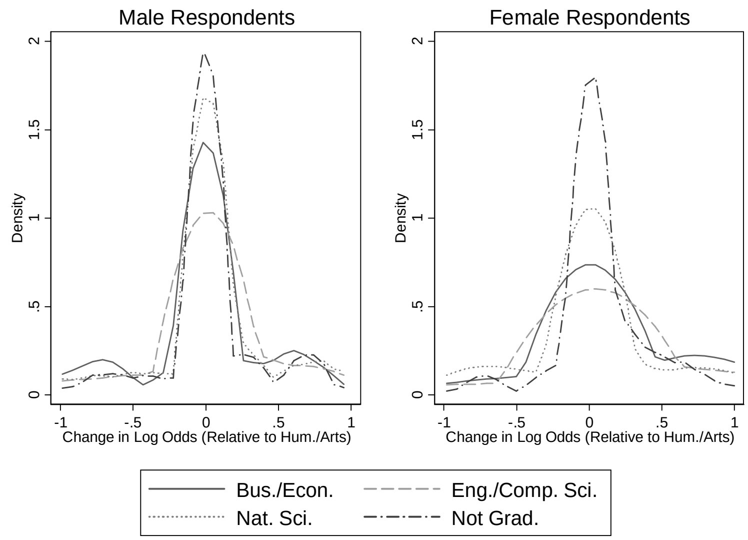

Figure 1: Distribution of Changes in Log Expected Graduation Probabilities (Relative to Humanities/Arts)

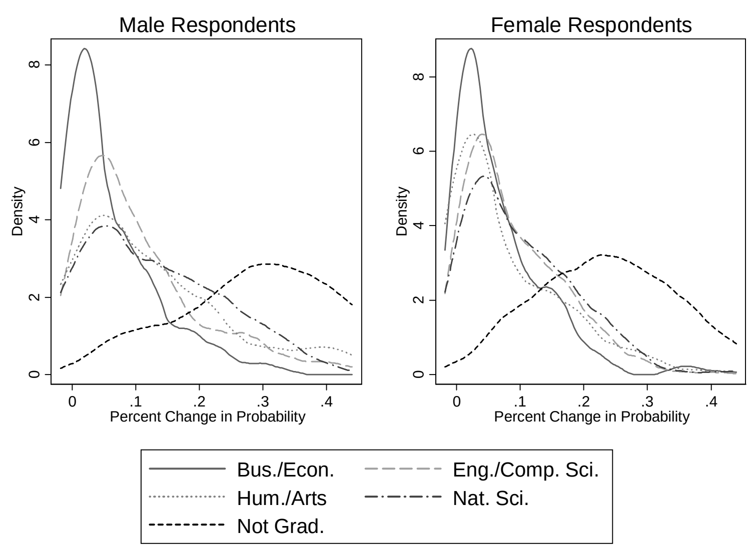

Figure 2: Distribution of Choice Elasticities

Table 1: Sample Characteristics

Number of respondents:		488
School year:		
	Freshman	40.57%
	Sophomore	35.86%
	Junior	23.56%
Age		20.13
		(1.17)
Female		63.93%
Race:		
	White	37.70%
	Non-Asian Minority	17.21%
	Asian	45.08%
Parents' Income (in $1,000)		143.84
		(123.45)
Mother has a B.A. or More		70.93%
Father has a B.A. or More		75.83%
SAT Math Score		700.57
		(76.71)
SAT Verbal Score		682.93
		(71.06)
GPA		3.48
		(0.32)
Intended/Current Major:		
	Economics	30.53%
	Engineering	4.51%
	Humanities	47.75%
	Natural Sciences	17.21%
(Intend to) Double Major		36.01%

Notes: For continuous variables, mean is reported in first row and standard deviation is reported in parentheses in second row.

Table 2: Mean and Standard Deviation of Elicited Population Beliefs

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Beliefs about Women			Beliefs about Men			Perceived Wage Gender Gap[a]	Actual Wage Gender Gap
		Belief	Percent Error $(\frac{\text{Truth - Belief}}{\text{Truth}} * 100)$		Belief	Percent Error $(\frac{\text{Truth - Belief}}{\text{Truth}} * 100)$			
			Actual	Absolute		Actual	Absolute		
Sample: Male Students									
Economics/Business	mean	7.90	-30.11	48.73	7.81	-4.82	34.86	-4.81	-18.53
	(std.)	(5.28)	(86.88)	(77.91)	(4.05)	(54.34)	(41.86)	(11.22)	
Engineering/Comp. Sci.	mean	6.85	8.82	33.68	6.89	16.33	31.81	-5.62	-8.85
	(std.)	(4.65)	(61.92)	(52.63)	(3.74)	(45.36)	(36.15)	(13.42)	
Humanities/Arts	mean	5.41	-10.16	35.92	5.48	-3.58	30.94	-5.65	-7.15
	(std.)	(4.38)	(89.2)	(82.22)	(2.85)	(53.87)	(44.16)	(17.36)	
Natural Sciences	mean	6.27	-4.54	33.48	6.26	13.80	31.25	-2.36	-17.31
	(std.)	(3.18)	(52.97)	(41.2)	(3.06)	(42.2)	(31.44)	(15.77)	
Not Graduate	mean	3.33	3.79	35.32	3.61	24.53	34.10	-6.80	-27.61
	(std.)	(1.69)	(48.81)	(33.77)	(1.63)	(34.12)	(24.46)	(15.55)	
Sample: Female Students									
Economics/Business	mean	8.00	-31.66	49.5	9.15	-22.77	46.74	-6.57	-18.53
	(std.)	(5.83)	(96.04)	(88.15)	(8.86)	(118.86)	(111.6)	(10.96)	
Engineering/Comp. Sci.	mean	7.22	3.87	31.03	8.54	-3.71	38.76	-6.84	-8.85
	(std.)	(5.23)	(69.71)	(62.51)	(9.16)	(111.19)	(104.25)	(15.48)	
Humanities/Arts	mean	5.70	-15.93	38.51	5.99	-13.23	35.40	-5.23	-7.15
	(std.)	(5.6)	(113.85)	(108.29)	(4.43)	(83.61)	(76.87)	(12.13)	
Natural Sciences	mean	7.15	-19.06	44.60	7.16	1.30	33.02	-1.34	-17.31
	(std.)	(7.68)	(128.03)	(121.49)	(4.94)	(68.13)	(59.57)	(64.44)	
Not Graduate	mean	3.54	-2.33	40.99	3.55	25.67	35.94	-0.95	-27.61
	(std.)	(3.57)	(103.23)	(94.73)	(1.84)	(38.5)	(29.12)	(72.18)	

Notes: Beliefs (columns 1, & 4) are in \$10,000's. Other columns are percentages.
[a] Wage gap is defined as 100* (female population earnings-male population earnings)/male population earnings.

Table 3: Earnings and Earnings Revisions

		(1)	(2)	(3)
		Self Earnings Pre	Self % Revision ($\frac{\text{Post-Pre}}{\text{Pre}}$*100)	Absolute Self % Revision
Sample: Male Students				
Economics/Business	mean	14.63	-11.98	23.86
	(std.)	(17.30)	(34.46)	(27.55)
Engineering/Comp. Sci.	mean	9.86	-2.18	20.16
	(std.)	(10.43)	(28.30)	(19.92)
Humanities/Arts	mean	6.86	15.58	33.73
	(std.)	(5.50)	(146.97)	(143.87)
Natural Sciences	mean	9.32	1.46	25.28
	(std.)	(10.18)	(38.75)	(29.35)
Not Graduate	mean	5.06	118.82	128.41
	(std.)	(11.00)	(830.33)	(828.89)
Sample: Female Students				
Economics/Business	mean	11.6	0.60	43.00
	(std.)	(11.95)	(130.66)	(123.36)
Engineering/Comp. Sci.	mean	9.73	7.38	40.15
	(std.)	(7.19)	(121.54)	(114.93)
Humanities/Arts	mean	6.87	8.49	33.84
	(std.)	(7.46)	(111.52)	(106.58)
Natural Sciences	mean	9.35	27.61	59.36
	(std.)	(9.79)	(336.33)	(332.19)
Not Graduate	mean	3.29	60.12	70.17
	(std.)	(4.59)	(310.02)	(307.9)

Notes: Earnings and S. d. (standard deviation) of earnings are in $10,000's.

Table 4: Population and Self Beliefs

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent Var:		Log Self Earnings			Log Earnings Revision (Post-Pre)		
Log Population Earnings Beliefs	0.451*** (0.0253)	0.308*** (0.0257)	0.309*** (0.0251)				
Log Population Earnings Errors log(Truth/Belief)				0.0786*** (0.0194)	0.0689*** (0.0195)	0.0768*** (0.0138)	0.0726*** (0.0138)
Indiv. Covariates?	NO	NO	YES	—	—	—	—
Major Dummies?	NO	YES	YES	NO	YES	NO	YES
Truncated Sample?[a]	NO	NO	NO	NO	NO	YES	YES
R-squared	0.257	0.398	0.416	0.014	0.035	0.023	0.034
Total Observations	2440	2440	2440	2440	2440	2166	2166
Individuals	488	488	488	488	488	485	485

Notes: Individual covariates include an indicator for gender; indicators for Asian, Hispanic, black, or other race (white race is omitted category), overall grade point average (GPA); scores on the verbal and mathematics SAT; indicators for whether the student's mother and father attended college; parents' income; and indicators for non-reported (missing) SAT scores, GPA, parental education or parental income. Major dummies include indicators for the remaining majors: economics/business, engineering/computer sci, natural science, and no graduation. Standard errors in parentheses. ***, **, * denote significance at 1, 5, and 10 percent, respectively.

[a] Truncated sample excludes observations where respondents revise their self beliefs by more than $50,000,

Table 5: Expected Probability of Completing a Degree in Specific Majors

		Before[a]		Before (Rel. Hum./Arts)[b]		Revisions Post-Pre Treat.		Log Odds Rev. (Rel. Hum./Arts)[c]	
		Male	Female	Male	Female	Male	Female	Male	Female
Econ/Business	mean	0.400	0.250	0.102	-0.248	-0.014	0.026	0.046	0.696
	(std.)	(0.382)	(0.337)	(0.665)	(0.657)	(0.138)	(0.108)	(1.90)	(1.99)
Eng/Comp. Sci.	mean	0.086	0.057	-0.212	-0.441	0.024	0.022	0.597	0.795
	(std.)	(0.156)	(0.135)	(0.424)	(0.448)	(0.089)	(0.082)	(2.10)	(2.25)
Humanities/Arts	mean	0.298	0.498	-	-	-0.023	-0.048	-	-
	(std.)	(0.360)	(0.389)	-	-	(0.128)	(0.145)	-	-
Natural Sciences	mean	0.192	0.176	-0.106	-0.322	0.015	-0.002	0.229	0.333
	(std.)	(0.284)	(0.273)	(0.526)	(0.569)	(0.134)	(0.102)	(2.02)	(1.91)
Not Graduate	mean	0.027	0.022	-0.271	-0.476	-0.002	0.002	0.073	0.155
	(std.)	(0.077)	(0.064)	(0.366)	(0.400)	(0.076)	(0.04)	(1.99)	(1.90)

Notes: This table reports the mean self belief about completing each of the majors. Probabilities are reported on a 0 - 100 scale, and then normalized to 0 - 1. The standard deviation is in parentheses.

[a] Reported before receiving info treatments.

[b] Probability in major - Probability in Humanities.

[c] Log(Post Probability in major / Post Probability in Humanities) - Log(Pre Probability in major / Pre Probability in Humanities).

Table 6: Graduation Expectations and Expected Earnings

	(1)	(2)	(3)	(4)	(5)
Dep. Variable:	Log Odds of Major Rel. to Hum.			Log Odds Revision (Post-Pre)	
Log Self Earnings (Rel to Hum/Arts)	1.68***	1.57***	1.61***		
	(0.113)	(0.152)	(0.140)		
Log Self Earnings Rev (Post - Pre)				0.146^{+}	0.275**
				(0.0996)	(0.140)
Indiv. Covariates?	NO	NO	YES	-	-
Major Dummies?	NO	YES	YES	YES	YES
Truncated Sample?[a]	NO	NO	NO	NO	YES
R-squared	0.096	0.121	0.270	0.013	0.012
Total Observations	1952	1952	1952	1952	1710
Individuals	488	488	488	488	485

Notes: Heteroskedastic cluster robust standard error in parentheses. Standard errors are adjusted for clustering at the individual level for the models which include individual covariates. Individual covariates are the same as in Table 4.

***, **, *, + denote significance at 1, 5, 10, and 15 percent, respectively.

[a] Truncated sample excludes observations where respondents revise their self beliefs by more than $50,000,

Table 7: Structural Model Post-Graduation Parameter Estimates

	Model 1 (Panel Data)		*Model 2* (Cross-Sectional Data Only)	
	Males	**Females**	**Males**	**Females**
Own Utility				
ϕ_1	0.2223***	0.2034***	0.1463***	0.6431***
	(0.0296)	(0.0165)	(0.0284)	(0.0992)
ρ_1	4.4846***	5.5085***	5.0592***	5.4059***
	(0.3595)	(0.2265)	(0.9005)	(0.3993)
Spouse Utility				
ϕ_2	0.3274***	0.3277***	0.1131***	0.5818***
	(0.0265)	(0.0293)	(0.0231)	(0.0179)
ρ_2	3.7876***	4.0326***	1.3585**	2.9086***
	(0.3213)	(0.3732)	(0.6150)	(0.6457)
Ability α	0.0982***	0.1090***	0.5298***	0.7565***
	(0.0305)	(0.0212)	(0.0912)	(0.0656)

Notes: Bootstrapped standard errors in parentheses calculated from 50 bootstrap repetitions. Standard errors in parentheses. ***, **, * denote significance at 1, 5, and 10 percent, respectively.

Table 8: Distribution of Estimated Taste Parameters (Relative to Humanities/Arts)

	Econ./Bus.	Eng./Comp.Sci	Nat. Sci.	No Grad.
Male Students				
Mean	0.507	-1.38	-0.764	-2.07
(Std.)	(4.47)	(3.71)	(3.90)	(3.01)
Median	-0.0381	-0.464	-0.198	-1.59
Female Students				
Mean	-1.36	-3.13	-2.06	-3.53
(Std.)	(4.21)	(3.28)	(3.67)	(2.80)
Median	-1.55	-2.83	-1.61	-3.96

Table 9: Sample Fit

	Data	Model
Male Students Prob. of Majoring in...		
Economics/Business	0.3995	0.4028
Engineering/Comp. Sci.	0.0860	0.0883
Humanities/Arts	0.2976	0.2902
Natural Sciences	0.1919	0.1997
Not Graduate	0.0268	0.0191
Female Students Prob. of Majoring in...		
Economics/Business	0.2498	0.2638
Engineering/Comp. Sci.	0.0567	0.0603
Humanities/Arts	0.4977	0.4829
Natural Sciences	0.1757	0.1740
Not Graduate	0.0219	0.0189

Table 10: Own Earnings Choice Elasticities: Average Percent Change in Probability of Graduating in Each Major with a 1% Increase in Own Earnings in that Major

	Unrestricted Model		**Cross-Sectional Data Only**	
	Male Students	Female Students	Male Students	Female Students
% Δ Prob Bus/Econ	0.0395	0.0367	0.1471	0.1500
% Δ Prob Eng/Comp. Sci.	0.0703	0.0486	0.2574	0.1886
% Δ Prob Hum./Arts	0.0935	0.0508	0.2099	0.1359
% Δ Prob Nat. Sci.	0.0769	0.0610	0.2240	0.2165
% Δ Prob No Grad.	0.2290	0.2063	0.2808	0.6839

Table 11: Correlates of Major-specific Tastes (Relative to Humanities/Arts)

	Bus/Econ.	Eng/Comp	Nat. Sci.	No Grad.
Male	1.76***	1.44***	1.03***	1.27***
	(.365)	(.296)	(.335)	(.256)
Sophomore	.139	-.069	-.364	.542*
	(.386)	(.314)	(.351)	(.278)
Junior	-.625	-1.16***	-1.60***	-.339
	(.452)	(.348)	(.419)	(.302)
Asian	2.23***	1.52***	.794**	.910***
	(.446)	(.331)	(.384)	(.292)
Hispanic	.380	.335	.0050	-.641
	(.691)	(.528)	(.598)	(.455)
Black	.011	-.066	-.0354	.725
	(1.07)	(.871)	(.935)	(.692)
SAT Math	.0091***	.0082***	.013***	.004***
	(.0022)	.(0020)	.(0021)	(.002)
SAT Verbal	-.0078***	-.0066***	-.0104***	-.0034**
	(.0021)	(.0018)	.(0022)	(.0016)
R-squared	0.2039	0.2099	0.1860	0.1610
Num. Obs.	488	488	488	488

Notes: Linear predictors of tastes (relative to Humanities/Arts). Standard errors in parentheses. ***, **, * denote significance at 1, 5, and 10 percent, respectively.

Table 12: Decomposition of the Determinants of College Major Choices

		(1)	(2)	(3)	(4)	(5)	(6)
		Change in Odds Relative to Humanities/Arts					
	Baseline Equal Odds	Add Own Earnings	Add Own Ability	Add Own Hours	Add Spousal Charact.	Add Own Tastes	Actual (Predicted) Odds
Male Students							
Econ./Bus.	1.0000	0.0568	0.0021	0.0100	0.0129	0.3099	1.3917
Eng./Comp. Sci.	1.0000	0.0477	-0.0302	0.0078	0.0048	-0.7301	0.3000
Nat. Sci.	1.0000	0.0258	0.0001	0.0055	0.0049	-0.3564	0.6799
Not Grad.	1.0000	-0.1263	-0.0291	-0.0192	-0.0701	-0.6881	0.0672
Female Students							
Econ./Bus.	1.0000	0.0366	-0.0417	0.0086	0.0029	-0.4619	0.5445
Eng./Comp. Sci.	1.0000	0.0319	-0.0880	0.0039	0.0040	-0.8265	0.1253
Nat. Sci.	1.0000	0.0227	-0.0463	0.0056	0.0016	-0.6238	0.3598
Not Grad.	1.0000	-0.1003	-0.1033	-0.0138	-0.0663	-0.6782	0.0382
Female/Male Ratio							
Econ./Bus.	1.0000	-0.0192	-0.0412	-0.0008	-0.0085	-0.5390	0.3913
Eng./Comp. Sci.	1.0000	-0.0151	-0.0573	-0.0033	-0.0004	-0.5064	0.4176
Nat. Sci.	1.0000	-0.0030	-0.0452	0.0004	-0.0029	-0.4200	0.5292
Not Grad.	1.0000	0.0297	-0.0868	0.0053	0.0003	-0.3800	0.5684

APPENDICES (NOT FOR PUBLICATION)

A Appendix

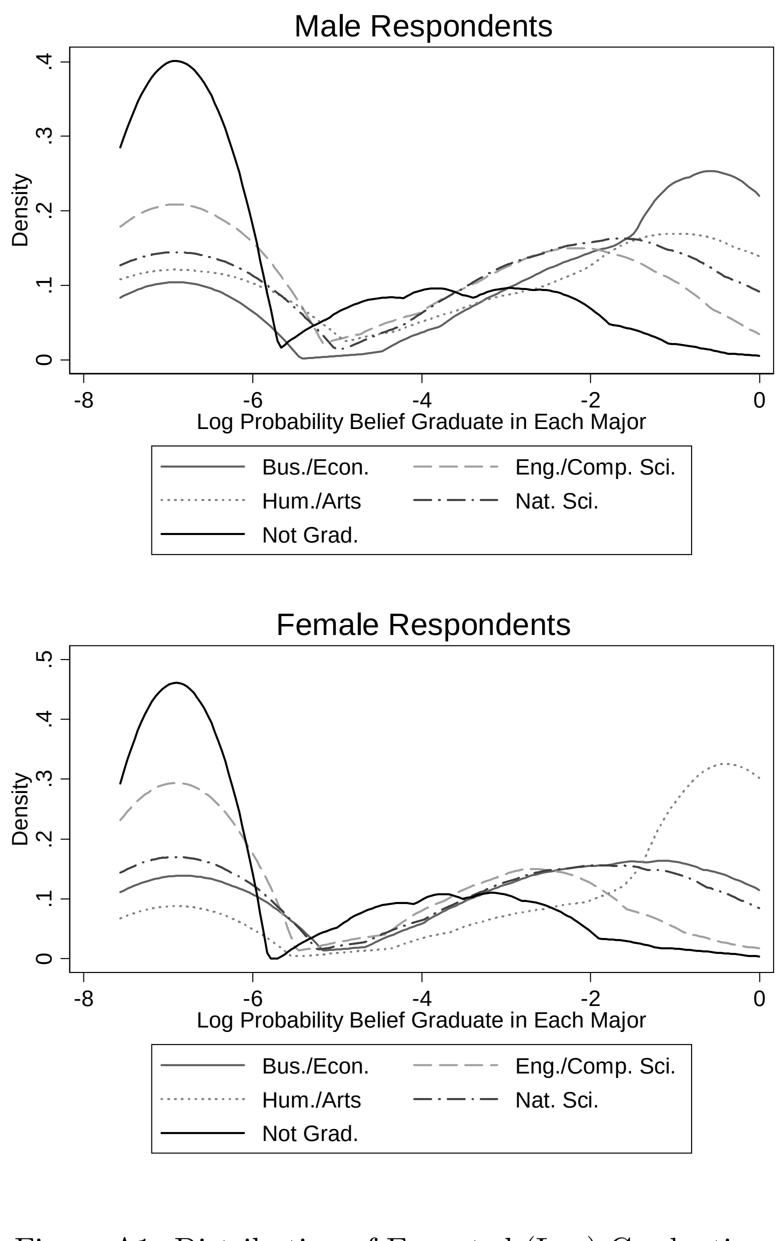

Figure A1: Distribution of Expected (Log) Graduation Probabilities

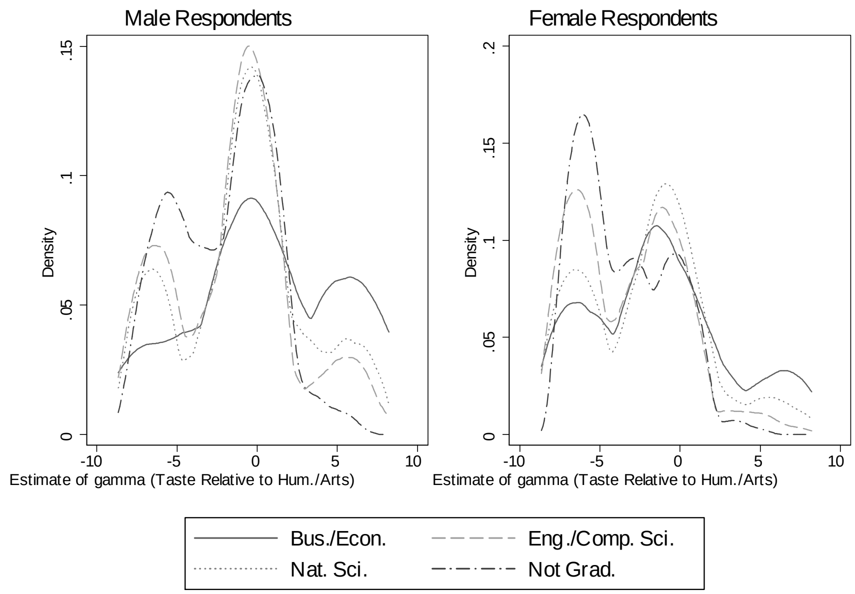

Figure A2: Distribution of Individual Fixed Taste (Rel to Hum./Arts) Component γ_{ik}

Table A1: Information revealed in the study

Labor market information about age 30 individuals

The following information is from the US Census Bureau.
Among all individuals (including college and non-college graduates) aged 30:

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

Labor market information about age 30 college-graduate individuals

The following information is from the US Census Bureau.
Among all college graduates currently aged 30:

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

Labor market information about age 30 college-graduate individuals, by gender and field of study

The following information is from the US Census Bureau.
Among all *female* college graduates aged 30 who received a Bachelor's degree in major (M):

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

The following information is from the US Census Bureau.
Among all *male* college graduates aged 30 who received a Bachelor's degree in major (M):

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

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

Table A2: Distribution of Beliefs

Percentiles:	10	25	50	75	90
Panel 1: Percent Error about Females' Earnings					
All Majors	-52.58	-23.50	-1.15	20.09	46.73
Economics/Business	-97.60	-48.20	-15.26	1.20	25.90
Engineering/Comp. Sci.	-33.18	-6.54	10.10	26.75	46.73
Humanities/Arts	-52.58	-22.07	-1.72	18.62	38.97
Natural Sciences	-49.95	-24.96	0.03	16.70	41.69
Not Graduate	-44.50	-15.60	13.30	39.31	71.10
Panel 2: Percent Error about Males' Earnings					
All Majors	-37.77	-13.34	6.09	31.11	58.16
Economics/Business	-60.98	-34.15	-7.32	19.51	32.92
Engineering/Comp. Sci.	-21.39	2.89	15.02	27.16	51.44
Humanities/Arts	-51.12	-22.79	-3.90	14.99	43.33
Natural Sciences	-37.77	-10.22	10.45	31.11	49.02
Not Graduate	-4.60	10.05	26.78	47.70	68.62
Panel 3: Self Earning Beliefs (in $10,000)					
All Majors	3.00	4.50	7.00	9.00	12.50
Economics/Business	6.00	7.25	9.00	12.00	20.00
Engineering/Comp. Sci.	5.00	6.95	8.00	10.00	15.00
Humanities/Arts	4.00	5.00	6.00	7.50	9.00
Natural Sciences	4.50	5.50	7.00	9.95	12.50
Not Graduate	1.00	2.00	3.00	4.00	5.00
Panel 4: Percent Self Earning Revision					
All Majors	-40.00	-22.22	0.00	16.67	50.00
Economics/Business	-50.00	-33.33	-14.29	0.00	20.00
Engineering/Comp. Sci.	-44.44	-25.00	-5.88	11.27	33.33
Humanities/Arts	-37.50	-16.67	0.00	16.67	42.86
Natural Sciences	-40.00	-23.61	-5.56	14.29	44.00
Not Graduate	-22.22	0.00	14.29	50.00	125.00

Table A3: Correlation in Self Earnings Across College Majors

Panel A: Male Students					
	Econ/Bus	Eng/Comp.	Hum./Arts	Nat Sci.	No Grad.
Econ/Bus	1.00				
Eng/Comp.	0.579	1.00			
Hum./Arts	0.209	0.230	1.00		
Nat Sci.	0.393	0.587	0.670	1.00	
Not Grad.	0.521	0.892	0.456	0.670	1.00
Panel B: Female Students					
	Econ/Bus	Eng/Comp.	Hum./Arts	Nat Sci.	No Grad.
Econ/Bus	1.00				
Eng/Comp.	0.596	1.00			
Hum./Arts	0.290	0.618	1.00		
Nat Sci.	0.465	0.603	0.428	1.00	
Not Grad.	0.366	0.626	0.560	0.445	1.00

Table A4: (Revisions in) Beliefs about Ability and Spousal Earnings

		(1)	(2)	(3)	(4)	(5)	(6)
		Self Ability Before	Ability Revision (Post - Pre)	Absolute Ability Revision	Spouse's Earnings Before[a]	Spouse Earnings % Revision ($\frac{\text{Post-Pre}}{\text{Pre}}$*100)	Absolute Spouse Revision
Sample: Male Students							
Economics/Business	mean	68.4	6.91	11.88	9.27	2.12	34.11
	(std.)	(28.46)	(20.02)	(17.52)	(8.93)	(89.75)	(83)
Engineering/Comp. Sci.	mean	56.13	12.28	18.26	7.78	13.88	34.63
	(std.)	(29.33)	(24.85)	(20.83)	(7.31)	(111.16)	(106.51)
Humanities/Arts	mean	68.64	8.51	14.61	7.07	3.43	26.07
	(std.)	(29.14)	(24.8)	(21.75)	(8.93)	(50.39)	(43.21)
Natural Sciences	mean	66.71	4.45	13.68	7.97	12.93	38.71
	(std.)	(27.23)	(22.65)	(18.57)	(7.89)	(115.49)	(109.54)
Not Graduate	mean	70.13	1.13	22.41	4.64	34.46	58.4
	(std.)	(38.99)	(38.41)	(31.17)	(5.57)	(168.15)	(161.36)
Sample: Female Students							
Economics/Business	mean	58.78	6.01	14.76	11.48	-1.09	30.7
	(std.)	(27.1)	(23.09)	(18.73)	(9.82)	(84.16)	(78.35)
Engineering/Comp. Sci.	mean	45.81	10.2	20.5	9.61	12.49	35.36
	(std.)	(29.09)	(27.68)	(21.19)	(7.8)	(107.22)	(101.97)
Humanities/Arts	mean	74.2	-0.21	12.75	7.85	15.75	38.33
	(std.)	(23.8)	(21.9)	(17.8)	(7.75)	(138.8)	(134.31)
Natural Sciences	mean	58.08	5.53	16.6	9.59	8.68	34.72
	(std.)	(28.64)	(24.25)	(18.51)	(8.76)	(104.91)	(99.36)
Not Graduate	mean	56.31	11.28	28.31	5.57	35.97	56.25
	(std.)	(43.21)	(43.08)	(34.34)	(9.17)	(128.73)	(121.22)

Notes: Ability ranking is measured on a 100 point scale, with 100 being top rank and 1 lowest rank.
[a] Spouse's earnings pre are beliefs about expected earnings of the student's spouse, conditional on the student's own major (not the spouse's major).

B Data

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

B.1 Survey Instrument

Because we wanted to approximate life cycle utility from each major, we collected beliefs about both initial earnings- just after college graduation, and for later periods, when earnings might be believed to be much higher. We collected post-graduation beliefs for three periods: i) first year after college graduation (when most respondents would be aged 22-24), 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 (including measures of dispersion), work status (not working, part time, full time), probability of marriage, and spouse's earnings. An example question on expected earnings at age 30: "*If you received a Bachelor's degree in each of the following major categories and you were working FULL TIME when you are 30 years old what do you believe is the average amount that you would earn per year?*"[31] The instructions emphasized to the respondents that their answers should reflect their own beliefs, and not use any outside information.[32]

Our questions on earnings were intended to elicit beliefs about the distribution of future earnings. We asked three questions on earnings: beliefs about expected (average) earnings, beliefs about the percent chance earnings would exceed \$35,000, and percent change earnings would exceed \$85,000. As detailed below, we use this information to estimate individual-specific distribution of earnings beliefs. Beliefs about spouse's earnings conditional on *own* major were also elicited in a similar way.

The probability of marriage was elicited as follows: "*What do you believe is the percent chance that you will be married by age 30 if you received a Bachelor's degree in each of the following?*"

Beliefs about labor supply were elicited conditional on marriage. For example, labor supply conditional on being not married at age 30 was asked 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 NOT married and you*

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

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

received a Bachelor's degree in each of the following?"

Respondents were also asked about their spouse's labor supply and field of study, conditional on own field of study. Beliefs about average hours of work for each major were also asked. The full survey questionnaire is available from the authors upon request.

B.2 Information on Survey Design and Information Treatments

Description of data sources provide to survey respondents:

Sources:

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

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

Methodology:

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

2) NSCG: We calculate inflation adjusted earnings using the Consumer Price Index. The salary figures we report are therefore equivalent to CPS figures in 2009 March real dollars. Full time status is defined as in the CPS sample. Given the need to make precise calculations for each field of study group, we use the combined sample of 30-35 year old respondents and age adjust the reported statistics for 30 year olds. This sample consists of 14,116 individuals. To calculate average earnings, we use an earnings regression allowing for separate age intercepts, one each for 6 ages 30-35. The predicted value of earnings from the regression is used as the estimate of average earnings for 30 year olds. For the percent full time employed, and percent with earnings greater than $35,000 and $85,000, we use a logit model to predict these percentages for 30 year olds and include a separate coefficient for each of the 6 ages 30-35.

C College Major Beliefs and Self Beliefs

In this section, we describe data on two other potential elements of post-graduation utility: perceived ability and spousal earnings.

C.1 College Major Beliefs and Self Beliefs about Ability

Ability in each major could be a factor in expectations about future earnings, and may affect the likelihood of a student completing required coursework necessary to graduate in each major. We asked the following question: "*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?*" To provide easier interpretation, we re-scaled the ability beliefs such that 100 represents highest ability and 1 represents lowest ability. The first column of Table A4 provides descriptive statistics for the ability rank beliefs. In general, male students believe they have higher relative ability than female students (except in the category of humanities and arts)- this is consistent with evidence that women tend to be less confident than men (Weinberger, 2004; Niederle and Vesterlund, 2007). For both male and female students, lowest believed ability is in engineering and computer science (56 for male students and 46 for female students). The highest average beliefs about ability for women are in humanities, whereas for male students it is in the not graduate category.

The second column of Table A4 reports the ability revisions after the information treatment.[33] For almost all categories, the average ability revision is upward: After receiving the earnings and labor supply information, the students believe they are more able than they were before. The only exception to the positive ability updating was humanities/arts for female students where the average ability rank fell somewhat following the information treatment. The third column shows that absolute average ability revisions are substantially larger than average ability revisions, indicating that a non-trivial proportion of students revise their ability beliefs both up and down.

C.2 College Major Beliefs and Self Beliefs about Spouse's Earnings

One potentially important consideration of major choice may be the types of potential spouses one might marry. Recent empirical papers suggest that investment in education generates

[33] In general, the information treatments we provide can shift perceptions of own ability in a field if individuals perceive some link between the difficulty of completing a task to the reward provided for that task.

returns in the marriage market, but this is inferred indirectly in existing studies.[34] We investigate this in a direct way, and asked respondents about the earnings of their potential spouse *if* they were to be married at age 30 and their spouse worked 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?*" 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. Column (4) of Table A4 reports the mean and standard deviation of beliefs about spouse's earnings. Compared to beliefs about own earnings in column (1) of Table 3, male students believe their spouse's earnings will be below their own earnings in every major category (except humanities/arts), while female students believe their spouse's earnings will exceed their own earnings. There are substantial 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 majored in economics/business, and lowest if they graduated in humanities/arts (among graduating majors). The relative spousal earnings for own major are similar to the relative self earnings for own major. These patterns indicate that students perceive sorting of spouses by own major choice, and is suggestive of assortative mating by field of study.[35]

Column (5) of Table A4 indicates that the information treatment induced considerable revisions in beliefs about spousal earnings, with the mean of the distribution of spousal beliefs shifting upward in almost all cases. The huge standard deviations in revisions of spousal earnings indicate that there is large heterogeneity in revisions of spousal average earnings. This is further highlighted by the large absolute revisions of spousal average earnings, shown in column (6) of Table A4.

[34] Ge (2010) 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 (2010) 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.

[35] The fact that there is assortative mating by education (more precisely, years of schooling) in the US is well documented (Mare, 1991; Pencavel, 1998).

D Estimation Details

This Appendix describes the approximation of beliefs we use to construct expected lifetime utility from each major. To make clear the relationship between the beliefs questions, which are conditioned on future ages of the respondents, we index age $q = 22, \ldots, 55$, rather than use time. At period $t = 1$ (first post-graduation period) in the lifecycle model we assume individuals are aged 22.

D.1 Beliefs about Own Earnings

For each individual, for each major, and for both the pre- and post- treatment periods, we have 7 data points: i) expected earnings immediately after graduation, ii) expected earnings at age 30, iii) belief that own earnings would exceed \$35,000 at age 30, iv) belief that own earnings would exceed \$85,000 at age 30, v) expected earnings at age 45, vi) belief that own earnings would exceed \$35,000 at age 45, vii) belief that own earnings would exceed \$85,000 at age 45. With 5 major categories, this provides $5x7x2 = 70$ data points on beliefs about own earnings for each individual respondent.

From this data, we estimate a Normal distribution approximation to individual beliefs about the distribution of earnings for all periods. For each individual i, we assume beliefs about earnings in major k follow

$$\ln w_{FT,1,q,i,k} \sim N(\mu_{1,q,i,k}, \sigma^2_{1,q,i,k}),$$

where

$$\mu_{1,q,i,k} = \mu^0_{1,i,k} + \mu^1_{1,i,k} q + \mu^2_{1,i,k} q^2,$$

$$\sigma_{1,q,i,k} = \sigma^0_{1,i,k} + \sigma^1_{1,i,k} q.$$

This parameterization allows beliefs in earnings to grow with age q, following the standard concave pattern. We also allow the variance in beliefs about own earnings to vary over time by allowing the variance parameter to depend on age. The individual-specific beliefs parameters consist of $\omega_{i,k} = [\mu^0_{i,k}, \mu^1_{i,k}, \mu^2_{i,k}, \sigma^0_{i,k}, \sigma^1_{i,k}]$. We compute the best fitting parameters to approximate the assumed distribution using simulation. For any given parameter vector $\omega_{i,k}$, we form a sequence of simulated earnings beliefs draws. From this sequence of earnings draws, we construct the simulated counterpart to the 7 statistics detailed above. We then chooses the $\omega_{i,k}$ parameters that minimize the quadratic distance between the simulated and actual data beliefs. Note that we compute $\omega_{i,k}$ for all individual, majors, and for the pre- and post- treatment states separately.[36]

D.2 Beliefs about Spouse's Earnings

For self beliefs about future spouse's earnings, we use a similar approximation method. For beliefs about spouse's earnings we economized on data question given the length of survey collection and only asked about the equivalent i)-v) beliefs for spouses. We follow the same model and approximation procedure for spouse's earnings beliefs as with own earning beliefs and compute a potentially different vector $\omega_{i,k}$ of parameters for spouses.

$$\ln w_{FT,2,q,i,k} \sim N(\mu_{2,q,i,k}, \sigma^2_{2,q,i,k}),$$

where

$$\mu_{2,q,i,k} = \mu^0_{2,i,k} + \mu^1_{2,i,k} q + \mu^2_{2,i,k} q^2,$$

$$\sigma_{2,q,i,k} = \sigma^0_{2,i,k} + \sigma^1_{2,i,k} q.$$

D.3 Beliefs about Own Labor Supply

For labor supply, we asked respondents to report their beliefs about the probability they would work either full-time, part-time, or not all, conditional on marriage. We asked this information for two time periods: age 30 and age 45. We also asked population beliefs by major about the average hours each individual believes a full time individual works in each major. To conserve on time, this question was only asked in the final post-treatment part of the survey, but the full/part/no work probability question was asked both in the pre- and post- treatment periods. The average hours beliefs by major, which were asked only in the pre-treatment period, are assumed to remain the same following the treatment. Our information treatments provided no information on average hours by major, and only provided information on full time probability.

We construct the hours distribution (conditional on marriage $m_{q,i,k} \in \{0, 1\}$) as

[36] In order to remove outliers that can happen by chance in the simulated wages, we enforce an earnings ceiling and floor as in the original data. We replace all simulated full-time earnings exceeding \$500,000 with \$500,000 and all simulated earnings less than \$10,000 with \$10,000.

$$h_{1,q,i,k} = \begin{cases} \bar{h}_{1,i,k} & \text{w/ prob. } pr(FT_{1,q,i,k} = 1|m_{q,i,k}) \\ 20 & \text{w/ prob. } pr(PT_{1,q,i,k} = 1|m_{q,i,k}) \\ 0 & \text{w/ prob. } 1 - (pr(FT_{1,q,i,k} = 1|m_{q,i,k}) + pr(PT_{1,q,i,k} = 1|m_{q,i,k})). \end{cases},$$

where $\bar{h}_{i,k} = \bar{h}_{30,i,k}1\{q \leq 35\} + \bar{h}_{45,i,k}1\{q > 35\}$ is individual i's belief about average full time hours in major k, which depends on age. Beliefs about part-time hours are assumed to be 20 hours for all individuals and majors.

D.4 Beliefs about Spouse's Labor Supply

The distribution of spouse's hours is modeled symmetrically with own labor supply. We therefore set full time hours for spouse's labor supply to 40.

$$h_{2,q,i,k} = \begin{cases} \bar{h}_{2,i,k} & \text{w/ prob. } pr(FT_{2,q,i,k} = 1) \\ 20 & \text{w/ prob. } pr(PT_{2,q,i,k} = 1) \\ 0 & \text{w/ prob. } 1 - (pr(FT_{2,q,i,k} = 1) + pr(PT_{2,q,i,k} = 1)). \end{cases},$$

where $\bar{h}_{i,k} = \bar{h}_{30,i,k}1\{q \leq 35\} + \bar{h}_{45,i,k}1\{q > 35\}$ is individual i's belief about opposite gender's average full time hours in major k, which depends on age. $pr(FT_{2,q,i,k} = 1)$ and $pr(PT_{2,q,i,k} = 1)$ are the beliefs of individual i about her spouse's probability of working full or part-time at age t if individual i graduates with major k.

D.5 Beliefs about Marriage

For marriage, we elicited beliefs about the probability the individual is married for 3 time periods: i) first year upon graduation ($q = 22$), ii) age 30, and iv) and age 40. We use a linear function to interpolation beliefs for all years as follows:

$$pr(m_{q,i,k} = 1) = \begin{cases} pr(m_{22,i,k} = 1) & \text{for } q = 22 \\ pr(m_{22,i,k} = 1) + \frac{pr(m_{30,i,k}=1) - pr(m_{22,i,k}=1)}{30-22}(q - 22) & \text{for } 30 < q < 22 \\ pr(m_{30,i,k)} = 1) & \text{for } q = 30 \\ pr(m_{30,i,k} = 1) + \frac{pr(m_{45,i,k}=1) - pr(m_{30,i,k}=1)}{45-30}(q - 30) & \text{for } 30 \leq q < 45 \\ pr(m_{45,i,k} = 1) & \text{for } q \geq 45. \end{cases}.$$