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

Matthew Wiswall Basit Zafar†

First Version: June 2011
This Version: July 2014

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 panel of beliefs data. While expected earnings and perceived ability are a significant determinant of major choice, heterogeneous tastes are the dominant factor in the choice of major. Analyses that ignore the correlation in tastes with earnings expectations inflate the role of earnings in college major choices. We conclude by computing the welfare gains from the information experiment and find positive average welfare gains.

JEL Codes: D81, D84, I21, I23, J10.
Keywords: college majors; information; subjective expectations; uncertainty.

---

*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 ASU, Edinburgh, Carnegie Mellon, CIPREE Conference on Subjective Expectations, Clemson, Duke, Hunter College, Michigan-Ann Arbor, NY Fed BBL, NYU Experimental Economics Working Group, OSU, Rutgers, UCLA, UNC-Chapel Hill, Washington-St. Louis, 2012 NBER Education meetings, and the 2012 ASSA meetings. Da Lin, Victoria Gregory, 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: Arizona State University, W.P. Carey School of Business, Department of Economics; E-mail: matt.wiswall@gmail.com. Zafar: Research and Statistics, Federal Reserve Bank of New York; E-mail: Basit.Zafar@ny.frb.org

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 that college major composition can help explain long-term changes in inequality and earnings differences by gender and race (Grogger and Eide, 1994; Brown and Corcoron, 1997; Weinberger, 1998; Gemici and Wiswall, 2014).

This paper studies the determinants of college major choices using a 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 future earnings and other major-specific aspects were they to major in different majors, their beliefs about the population distribution of these outcomes, and the subjective belief that they will graduate with each major. 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 panel data for major choices, which is otherwise largely 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, free of bias stemming from the correlation of unobserved preferences with observed characteristics. Underscoring the importance of this bias, we compare cross-sectional OLS estimates of the relationship between major choice and earnings expectations with our experimental 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 individuals have biased beliefs about the population distribution of earnings (Betts, 1996; Jensen, 2010; Nguyen, 2010). We find that students in our sample also have biased beliefs about population earnings and there is considerable heterogeneity in errors, with some students over-, and other students under-, estimating average earnings in the population. We also find evidence of substantial and logical updating of their beliefs about their own future earnings if given accurate information on the current population earnings. Turning to expected major choices, we show how the experimental variation 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 and tastes.

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 (e.g. 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. Much of the past work uses this approach (Freeman, 1971, 1976a, 197b; Siow, 1984; Zarkin, 1985; Bamberger, 1986; Berger, 1988; Flyer, 1997; Eide and Waehrer, 1998; Montmarquette et al, 2002; Arcidiacono, 2004; Beffy et al, 2011; Gemici and Wiswall, 2014). 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 (Manski, 1993).

We estimate a structural life-cycle utility model of college major choice, and exploit experimental variation in information that creates *within* individual variation in beliefs to identify the model. In a decomposition of the various elements of the utility from each major, we find that beliefs about future earnings and perceived ability are a significant determinant of major choice. With the exception of drop-out (non-graduate) alternative, we estimate average elasticities of major choice to changes in future earnings in each period of between 0.03 and 0.07, which is lower but similar in magnitude to other studies using alternative identification schemes and populations. In addition, emphasizing the "value added" of our experimentally derived panel of beliefs, our estimates using the panel of beliefs, which allows us to difference out unobserved tastes for majors, yields much smaller elasticities of major choice with respect to earnings than a model estimated using only baseline beliefs in a cross-sectional analysis. Our data collection methodology also elicits students' subjective uncertainty about future earnings which is directly incorporated in the model. In fact, we estimate a large degree of risk aversion, and underscoring the need of modeling earnings uncertainty in the choice, we find that ignoring risk aversion severely inflates the responsiveness of individuals to changes in expected mean earnings.

We find that the residual unobserved taste component 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 (2014). These "tastes" for majors have a strong year in school component, and play a much larger role for older (juniors) than younger students (freshman and sophomores), indicating a large and increasing cost of switching majors as students progress through school. Indeed, in the analysis of major choice elasticities for different sub-samples, we estimate that younger students have higher average elasticities and are therefore more responsive to changes in future earnings than older students.

Finally, we compute welfare gains from the information experiment itself and find that the average change in expected major choices is equivalent to a sizable increase in earnings of between 5.6 to 6.4 percent, where the lower percentage is equivalent to $3,665 in additional income each year. It should be emphasized that our measure of welfare is in terms of expected outcomes, not realized outcomes, as our sample is still too young at the time of our analysis to have experienced many of the events we ask about in our survey of beliefs. But these non-trivial gains and the very small cost of providing information would seem to justify information interventions such as ours.

Our paper is also related to the recent and growing literature 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 (2011, 2013), Giustinelli (2010), Arcidiacono, Hotz, and Kang (2011), Kaufmann (2012), and Stinebrickner and Stinebrickner (2012, 2014) 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. We advance this literature in several ways. First, we combine data on probabilistic choices and subjective beliefs with an information experiment. As we show in Section 3, the panel data generated by the information experiment allow us to separately identify the unobserved tastes for each major from other aspects of the choice (earnings, ability, etc.) under weaker modeling restrictions than is possible with cross-sectional data. Second, we collect direct measures of earnings uncertainty and allow for a non-linear utility function in consumption. Both these innovations have implications for the choice elasticity estimates. Third, we elicit beliefs about future earnings at multiple points in time over the life-cycle, which allows us estimate a life-cycle utility model without making strong assumptions about earnings growth over the life-cycle. Fourth, we collect data on several other dimensions of future consumption uncertainty conditional on college major, such as labor supply, marriage, and spousal characteristics, and incorporate them directly into the choice model.

This paper is organized as follows. The next section outlines the model of college major choice. We explore identification of the model in Section 3, and describe the data collection methodology in Section 4. Section 5 examines heterogeneity in beliefs about earnings and revisions in self beliefs following the information treatment, and reports reduced-form regressions on the relationship between beliefs about major choice and beliefs about future earnings. Section

6 reports estimates from a structural life-cycle utility model of major choice, and related analysis. Finally, Section 7 concludes.

2 Model

In this section we specify the model of college major choice. Because the flexibility of the model is based on the particular data we collect, we reference the data collection (described below) to justify various modeling choices. The next section explicitly examines how we use the information experiment to identify the model.

2.1 Timing

Currently enrolled college students choose one of K majors: $k = 1, \ldots, K$. In order to model the complete potential choice set, one of the “majors" is a “no graduation" (college drop-out) choice. At the initial period $t = 0$, individuals are enrolled in college and have not chosen a particular college major. After realizing a shock to their utility from each major, each student then makes a college major choice and graduates from college.[1] Period $t = 1$ is the first period following college graduation. At period $t = 1$ and onward, the college graduate makes all remaining choices, including labor supply and marriage choices. At period $t = T$, the individual retires. To make clear how this timing convention is reflected in our survey design, our survey is conducted with currently enrolled students, who are therefore in period $t = 0$. While we have a single college period, we take account of the year the student is in school by discounting the future post-graduation utility according to the student's years remaining in college. Below, we discuss how our model reflects the cost of switching between major fields while in college, and how this may differ depending on the student's year in school.

2.2 Within College Preferences and Beliefs

At period $t = 0$, utility for each college major k is given by

$$V_{0,k} = \gamma_k + \alpha \ln a_k + \eta_k + EV_{1,k} \tag{1}$$

We define each of the terms as follows:

The γ_k component represents the preferences or tastes for each college major k at the initial pre-graduation stage. These could be tastes for major-specific outcomes realized in college, such as the enjoyability of coursework, or tastes for major-specific post-graduation outcomes, such

[1] This shock can be thought of as a shock to the perceived ability in each major or to simple tastes for each major. See Arcidiacono (2004), Stinebrickner and Stinebrickner (2012, 2014), and Arcidiacono, Aucejo, and Spenner (2012) for an analysis of major switching within college.

as expected non-pecuniary aspects of jobs associated with a major. 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 a large, positive γ_k leads students to be less likely to switch out of major k into an alternative major. 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 since most students report a non-zero probability of graduating with each possible major (see section 5).

The $\ln a_k$ term reflects the student's *perceived* ability in each major, where $a_k > 0$ for all k. We expect $\alpha > 0$, reflecting that higher ability in a particular major improves performance in the major's coursework and reduces the effort cost of completing it. We allow for ability in school and ability in the labor market to be correlated, since our data allow us to measure expected earnings in each field and beliefs about ability in each field directly.[2]

η_k are the period $t = 0$ preference shocks that reflect any change in the utility for a major that occurs between the initial pre-major choice period and the period when the college major is chosen. In the Blass, Lach, and Manski (2010) taxonomy, η_k is "resolvable" uncertainty–uncertainty that is resolved at the point at which the choice is made. We assume that students report their major choice probabilities prior to the realization of these shocks.

In general, we might expect considerable learning about student abilities and the characteristics of majors (e.g., future earnings) between the time of our survey and the point at which our respondents choose a major (closer to college graduation). Learning about ability, tastes, and other major specific characteristics, while enrolled in college, is central to other models of major choice (Arcidiancono, 2004; Stinebrickner and Stinebrickner, 2013). The key modeling assumption in our context is not the actual degree of updating of beliefs from the time of college to the time of major choice, but an assumption about how much updating students *believe* will occur, in particular by how much students believe their current (at the time of the survey) uncertainty will be resolved by the time they choose their major.[3] In our model, the element of uncertainty that is resolved at the time of major choice (implicitly by some unspecified learning process) is represented by the major specific $\eta_1, \ldots, \eta_K$ preference shocks. While major-specific ability beliefs are allowed to be heterogeneous and biased, we do not allow uncertainty or updating of these beliefs from the time of the survey to the time of major choice. We do allow uncertainty in post-graduation beliefs about future earnings from each major (as discussed be-

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

[3] Anticipated learning about some stochastic outcome (e.g. future earnings) should not affect the mean of the anticipated post-learning distribution, only the dispersion around this mean. If a student believes that future learning will in fact reveal that their current expectations about some outcome is biased in some particular direction, then they should revise their current expectation to what they anticipate the future post-learning expectation to be.

low in more detail), but this uncertainty is assumed to remain constant from the time of the survey until graduation. This assumption rules out students resolving some of this uncertainty as they learn about future major-specific earnings while in school. While students likely still believe that some uncertainty remains at graduation, as students would not believe they could exactly predict future earnings at all future ages, our assumption likely over-states the role of uncertainty.[4]

2.3 Post-Graduation Preferences and Beliefs

At college graduation, we assume each individual has obtained a degree in a particular field $k = 1, \ldots, K$.[5] Post-graduation utility is given by

$$EV_{1,k} = \sum_{t=1}^{T} \beta^{t-1+g} \int u(X) dG(X|k,t), \tag{2}$$

$g = \{1, 2, 3, 4\}$ is the student's years until graduation, with $g = 4$ (freshman), $g = 3$ (sophomore), $g = 2$ (junior), and $g = 1$ (senior). $\beta \in (0, 1)$ is the discount rate. The different discounting by the student's grade reflects the fact that post-graduation utility for first year students is farther in the future than for older students. $u(X)$ is the post-graduation utility function that provides the mapping from the finite vector of post-graduation events X to utility. As specified below, X includes a wide range of events (earnings, labor supply, marriage, spousal earnings). $G(X|k,t)$ is the individual's *beliefs* about the distribution of future events in period t, conditional on choice of major k. Our survey design directly elicits students' beliefs $G(X|k,t)$.[6]

Note a key difference between our specification of future utility and the standard approach in the previous literature. Previous research solves for optimal decisions and outcomes (labor supply and associated earnings and consumption) given the human capital investment. In these

[4] As discussed by Stinebrickner and Stinebrickner (2013), designing questions to elicit the proportion of uncertainty that student's believe will be resolved is quite difficult. In their paper, likely the first to tackle the issue empirically using beliefs data, they estimate the resolvable uncertainty using detailed information on *actual* updating behavior they observe in high frequency panel data. Their estimation uses a type of rational expectations assumption directly linking ex-post actual realizations of updating to ex-ante beliefs about updating. In our context, we do not have rich panel data on actual belief updating and cannot follow their approach. We could use observed updating to the information treatments in our experiment as the basis to model beliefs about updating, although, like in Stinebrickner and Stinebrickner (2013), this would require a rational expectations assumption, and our information treatments would provide at best only a partial approximation to the many types of information students might believe they would acquire while in school.

[5] At any age after graduation, individuals could be currently earning a graduate degree or have already earned a graduate degree. When we elicit beliefs in the survey for a given age, conditional on *each* major, we instruct respondents to consider the possibility that they may have earned a graduate degree by this age. The earnings beliefs we elicit therefore should include any post-graduate premium associated with graduate degrees.

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

previous approaches, the utility from the major choice is found by substituting this optimal level of labor supply and associated labor market earnings and consumption into the utility function. Our specification of the future value $EV_{1,k}$ has no optimization component since we ask each student how much they expect to work and earn given each major choice. We therefore allow students to solve for their own future decisions and simply substitute these reported beliefs into the utility function. In this way, we avoid explicitly solving for future choices, which has the advantage of both generality in allowing for non-rational expectations and reducing the time to compute the model solution.

The distribution of future post-graduation events $G(X|k,t)$ represents "unresolvable" uncertainty as these events will not have occurred at the time of major choice. Beliefs are individual-specific, and may not be consistent with rational expectations. In general, beliefs are based on current information, which, as discussed below, can be a mixture of public and private information. 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 earnings in the population for individuals who graduate with a business degree.

Marriage At each age, individuals can be either single or married, where the "single" state includes both divorced and never married states. We do not model the number or length of marriage spells, nor directly inquire in our survey about expected marriage spells. Instead, our survey elicits beliefs about the probability of being married at different future ages and the changes in these probabilities across ages indicates how an individual expects marital status to evolve over her lifecycle. Like all of the beliefs we elicit in our survey, the marriage probabilities at each age are conditioned on the individual's major, thus allowing us to see how marriage beliefs change with major choice.

Flow Utility 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 when single. 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 the individual's own consumption when married and $c_{M,2,t}$ is the individual's spouse's consumption. $U_{M,t}$ defines the individual's*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.[7] We use the individual's self beliefs about own

[7] Inclusion of marriage and spousal characteristics is motivated by recent theoretical models which emphasize that investment in education generates returns in the marriage market (Iyigun and Walsh, 2007; Chiappori,

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.

Labor Supply An individual's annual labor force status takes three values: not working, working part-time, and working full-time (defined as working at least 35 hours per week and at least 45 weeks during the year). We ask individuals their beliefs about the probability they will work full- or part-time at future ages conditional on being either single or married, and conditional on major. In addition, we ask their beliefs about their potential spouse's probability of working full or part-time at future ages conditional on their own major. Importantly, we ask about spousal characteristics conditional on the individual's own major, not the spouse's potential major, as our interest is understanding the respondent's major choice, not her spouse's. In addition to labor force status, we also ask beliefs about the expected number of hours a full-time worker works in each major by gender. This allows us to make a more precise conversion of full-time earnings to part-time earnings belief, as described below.[8] Our elicitation of labor supply beliefs makes no distinction between voluntary or involuntary sources of labor supply; an individual could believe they are unlikely to work full-time because they will choose to remain out of the labor force (e.g. to care for young children) or because they do not receive a job offer.

Earnings We collect an individual's beliefs about their own and their potential spouse's future distribution of earnings conditional on their own major. As with labor supply beliefs, we ask about spousal earnings conditional on the individual's own major, not the spouse's potential major. Because we ask individuals about full-time equivalent earnings only, we combine the beliefs about labor supply and full-time earnings to define earnings when the individual is working part-time. Own and spousal earnings are modeled as

$$y_{q,t} = w_{FT,q,t}FT_{q,t} + w_{FT,q,t}(20/h_{FT,q,t})PT_{q,t} \text{ for } q = 1, 2$$

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, and $h_{FT,q,t}$ is full time hours. Key to our modeling strategy is that we ask respondents for their potential earnings *if* they are working full-time, and ask their beliefs about the probability of working full-time as a separate question. This allows us to elicit beliefs about earnings even if the individual believes there is a zero probability of actually working full-time in the future. This aspect of our data allows us to circumvent the standard endogenous wage distribution issue where wages are only observed for individuals who work.

Iyigun, and Weiss, 2009).

[8] As described more fully in the Appendix, because of time constraints, beliefs about hours if working full- or part-time were collected as beliefs about the population average hours worked conditional on major and gender, where we used the hours for the opposite gender to construct an individual's beliefs about spouse's hours.

Consumption Since we allow an individual's beliefs about the future distribution of full-time and part-time probabilities to depend on marriage, earnings and consumption also depend on marriage. We do not model borrowing and savings and assume consumption in each period is equal to current period earnings.[9] Consumption conditional on marriage is then given by $c_{S,1,t} = y_{1,t}$ (own consumption when single), $c_{M,1,t} = \frac{1}{2}(y_{1,t} + y_{2,t})$ (own consumption when married), and $c_{M,2,t} = \frac{1}{2}(y_{1,t} + y_{2,t})$ (spousal consumption when married).[10]

Household Preferences 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, 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): $u_{M,1}(c_{M,1,t}) = \phi_1 \frac{c_{M,1,t}^{1-\rho_1}}{1-\rho_1}$. 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 her spouse. The preferences of the individual over her spouse's consumption are allowed to be different from her preferences over her 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 $\rho_2 \in (0, \infty)$. ϕ_2 and ρ_2 parametrize the individual's preferences over her spouse's consumption.[11]

Expected Post-Graduation Utility In principle, we could collect individual beliefs about the joint distribution of all post-graduation events in the model: labor supply, earnings, marriage, and spousal characteristics. In practice, due to time constraints in the survey collection, we impose a number of restrictions on the joint distribution: i) own and spousal earnings are assumed independent of employment (full or part-time) up to the hours adjustment described above, ii) own earnings are assumed independent of marriage and spousal characteristics, iii) own hours if working full-time are independent of earnings, marriage, and spousal character-

[9] In the absence of savings and borrowing, we need to make an assumption regarding a consumption floor. We assume that when the individual or spouse is not working at all, annual income is equal to $10,000 when single or $20,000 for a couple if both spouses are not working. In general, there are two alternative approaches to adding borrowing and savings to a model such as this. First, 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 beliefs about earnings and labor supply.

[10] In principle, one could generalize this model by allowing spouses to receive unequal shares of total household consumption, i.e. $c_{M,1,2} = \kappa(y_{1,t} + y_{2,t})$ and $c_{M,2,t} = (1-\kappa)(y_{1,t} + y_{2,t})$, with $\kappa \in (0, 1)$. One approach to identify κ is to collect beliefs data on individual's perceptions of future intra-household resource allocation. Another approach is to specify κ as the outcome of some intra-household bargaining process.

[11] 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.

istics. Combining these assumptions with the model structure detailed above, expected post-graduation utility (2) can be re-written as

$$\begin{aligned} EV_{1,k} &= \sum_{t=1}^{T} \beta^{t-1+g} \{ pr(m_t = 0|k,t) \sum_{l=FT,PT,NW} pr(L_{1,t} = l | m_t = 0, k, t) \int \phi_1 \frac{c_{S,1,t}^{1-\rho_1}}{1-\rho_1} dF_1(w_{FT,1,t}|k,t) \\ &+ pr(m_t = 1|k,t) [\sum_{l=FT,PT,NW} pr(L_{1,t} = l | m_t = 1, k, t) \int \phi_1 \frac{c_{M,1,t}^{1-\rho_1}}{1-\rho_1} dF_1(w_{FT,1,t}|k,t) \\ &+ \sum_{l=FT,PT,NW} pr(L_{2,t} = l|k,t) \int \phi_2 \frac{c_{M,2,t}^{1-\rho_2}}{1-\rho_2} dF_2(w_{2,FT,t}|k,t)] \}, \end{aligned} \tag{3}$$

where consumption levels given labor supply and earnings are defined above. $pr(m_t = 1|k,t)$ is the belief about the probability of being married at age t if the individual completes major k. $pr(L_{1,t} = l|m_t = j, k, t)$ for $l = FT, PT, NW$ are the beliefs about labor force status (working full-time FT, part-time PT, or not working NW), given marital state ($m_t = j$, $j = 0$ single or $j = 1$ married), major (k), and age (t). $pr(L_{2,t} = l|k,t)$ is the individual's beliefs about her spouse's labor supply conditional on the individual's own major k and own age t (*not* the spouse's major or age). $F_1(w_{FT,1,t}|k,t)$ is the individual's beliefs about own future full-time earnings conditional on major and age. $F_2(w_{FT,2,t}|k,t)$ is the individual's beliefs about potential spouse's full-time earnings conditional on the individual's own major and own age (*not* the spouse's major or age). Recall the model structure described above where we convert beliefs about full time earnings (which we explicitly ask about in the survey) to beliefs about part-time earnings (which we do not ask about) using individual's beliefs about the average hours individuals work if working full-time, all conditional on major.

In addition to restrictions on the joint distribution of events (conditional on major choice and age), there are other data limitations due to time and respondent burden considerations: we cannot ask respondents to report marital status, labor supply, and earnings for *every* year after graduation nor can we ask an infinite number of questions in order to provide a non-parametric distribution of beliefs about the future earnings distribution. We instead ask these beliefs for two or three points in time after graduation and ask for three distinct moments in the distribution of future earnings. Section E in the Appendix describes our approximations. We use various polynomial approximations to interpolate between data points by age, and use a Normal distribution to approximate the distribution of beliefs about future earnings. It is important to emphasize that these approximations are entirely individual-specific (using free parameters for each individual): we make no assumption regarding the distribution of self beliefs in the population.

2.4 Major Choice

Individuals choose the college major that maximizes expected utility at period $t = 0$: $V_0^* = \max\{V_{0,k}, \ldots, V_{0,K}\}$. Prior to the choice of major, the individual's expected probability of majoring in each of the k majors given beliefs is then obtained as follows:

$$\pi_k \equiv pr(V_{0,k} = V_0^*) = \int 1\{V_{0,k} = V_0^*\} dF(\eta), \tag{4}$$

where $F(\eta)$ is the joint distribution of the preference shocks, $\eta_1, \ldots, \eta_K$, which represent the resolvable uncertainty in the model. As discussed above, there is no other resolvable uncertainty; the uncertainty elicited regarding earnings and other major characteristics is assumed to be unresolved by the time of the major choice. Our survey elicits the individual specific expected probabilities $\pi_1, \ldots, \pi_K$, with $\pi_k \in [0, 1]$, for all k, $\sum_{k=1}^{K} \pi_k = 1$.

2.5 Major Choice Elasticities: Within College Major Switching

One of the key issues in this model is how it incorporates major switching and the sensitivity of students to changes in post-graduation outcomes, in particular future earnings. In response to an increase in the beliefs about earnings for major k, how much more likely would an individual be to complete major k? For each student i, the model yields choice elasticities given by $\xi_{k,i}$ which give the change in the percent probability of completing major k with respect to a change in the mean earnings for each period t (a change in beliefs $G_i(X|k, t)$), where the additional i subscript emphasizes that each of these objects is student i specific.[12]

The choice elasticities $\xi_{k,i}$ depend on all model elements including the post-graduation utility function and the marginal utility of earnings (i.e., the values of $\phi_1, \phi_2, \rho_1, \rho_2$). Through the non-linear utility function, the responsiveness of major choice to earnings also depends on the level of earnings, both for the individual and any potential spouse, and therefore also depends on all other beliefs about the distribution of earnings, labor supply, marriage, and spousal earnings.

An important element of the responsiveness of major choice to changes in earnings is the magnitude of the $\gamma_{1,i}, \ldots, \gamma_{K,i}$ "taste" terms for each individual. Relatively equal $\gamma_{k,i}$ terms across majors for an individual imply that this individual has a high level of responsiveness to earnings changes. Relatively different values of $\gamma_{k,i}$, as where $\gamma_{j,i} >> \gamma_{j',i}$, implies a high cost of switching from major j to major j'. While we do not model it explicitly, our model allows individuals to switch their majors. Most students believe there is at least some chance they will switch their major before graduation– as we show in the data analysis section, the majority of our sample students report uncertainty about their major choice.

[12] We could also consider choice elasticities with respect to changing other moments of the distribution of earnings beliefs $G_i(X|k, t)$. In defining this particular elasticity, we keep other moments of the distribution of earnings beliefs, e.g. the variance (uncertainty) about future earnings, the same.

In general, the responsiveness of major choice depends on the age and prior history of the respondents at that point. Our model therefore allows for heterogeneity in choice elasticities $\xi_{k,i}$, and we estimate the *distribution* of choice elasticities for different students. We expect the distribution of choice elasticities to differ on many dimensions, but in particular, based on amount of schooling the student has already completed. In general, we expect that the choice elasticities $\xi_{k,i}$ would be smaller for older students, who presumably have higher sunk investments in particular majors. At the extreme, students just prior to completing their degree are highly unlikely to change their major in response to new information, and therefore $\xi_{k,i} \rightarrow 0$. First year (freshman) students would be more responsive to changes in earnings as they have lower levels of sunk investments in particular fields, and we expect their $\xi_{k,i} >> 0$.

Another potentially important source of heterogeneity in the response of students to changes in average beliefs is that students can differ in their degree of uncertainty about future earnings. In general, with risk averse preferences, greater uncertainty reduces the marginal utility to a change in expected earnings. We would expect then that more uncertain students would have lower elasticities $\xi_{k,i}$ with respect to changes in mean earnings beliefs. This is particularly relevant to the discussion above regarding the resolvability of uncertainty: our model assumes that all uncertainty with respect to earnings is unresolved between the time of the survey and graduation. To the extent that some of this uncertainty is in fact resolved, then our model is assuming counterfactually too much uncertainty and the earnings elasticities are biased downward. This bias may also differ systematically across the age or grade level of the respondents as the assumption that no uncertainty is resolved is particularly binding for younger students who have more scope than older students to learn about post-graduate earnings while in college.

3 Identification

In this section, we discuss how the model developed above is identified with our experimentally derived panel data on beliefs. Adding subscripts for each student i to (1), the utility from each major k is given by

$$V_{0,k,i} = \gamma_{k,i} + \alpha \ln a_{k,i} + \eta_{k,i} + EV_{1,k,i},$$

where $EV_{1,k,i}$ is the discounted sum of post-graduation utility student i expects to receive if she graduates with major k. We assume $\eta_{k,i}$ are distributed i.i.d. extreme value across major choices and across individuals. Note that while we assume a particular distribution for the 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 $\delta_{k,i} = \eta_{k,i} + \gamma_{k,i}$.[13]

The log odds of student i completing major k relative to a reference major $\tilde{k}$ is then

$$\begin{aligned} r_{k,i} &\equiv \ln \pi_{k,i} - \ln \pi_{\tilde{k},i} \\ &= \alpha(\ln a_{k,i} - \ln a_{\tilde{k},i}) + EV_{1,k,i} - EV_{1,\tilde{k},i} + \psi_{k,i}, \end{aligned} \tag{5}$$

where $\psi_{k,i} = \gamma_{k,i} - \gamma_{\tilde{k},i} + \epsilon_{k,i}$ is the combined unobservable in the log odds expression that reflects individual specific relative tastes $\gamma_{k,i} - \gamma_{\tilde{k},i}$ and additional sources of error $\epsilon_{k,i}$.

Estimation of (5) directly using the cross-sectional major probabilities reported by our sample would result in biased estimates of the α parameter and the post-graduation utility parameters in $EV_{1,k,i}$ given that ability and beliefs about future major specific outcomes would be correlated with relative tastes for each major, reflected in the $\gamma_{k,i} - \gamma_{\tilde{k},i}$ term. For example, students who expect high future wages in some field k relative to field $\tilde{k}$ may also have higher tastes for field k relative to $\tilde{k}$. 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).[14]

Our innovation is to note that if we can perturb the beliefs of the individuals, we could form *panel* data on beliefs, and use a standard fixed effects identification strategy to identify model parameters without imposing a parametric assumption on the distribution of tastes. We experimentally provide an information treatment to students and then re-elicit their beliefs again, post-treatment. For any object z, let z denote pre-treatment beliefs and z' denote beliefs after receipt of the information treatment. The difference in log odds, post minus pre-treatment, is then

$$\begin{aligned} r'_{k,i} - r_{k,i} &= \alpha\{(\ln a'_{k,i} - \ln a'_{\tilde{k},i}) - (\ln a_{k,i} - \ln a_{\tilde{k},i})\} \\ &+ (EV'_{1,k,i} - EV'_{1,\tilde{k},i}) - (EV_{1,k,i} - EV_{1,\tilde{k},i}) + \epsilon'_{k,i} - \epsilon_{k,i}, \end{aligned} \tag{6}$$

With the panel data in beliefs, we eliminate the relative taste component $\gamma_{k,i} - \gamma_{\tilde{k},i}$ and therefore can form a consistent estimator of the remaining utility parameters.

Identification requires that the *change* in beliefs about unobserved events or measurement error, given by $\epsilon'_{k,i} - \epsilon_{k,i}$, is mean-independent of the *changes* in observed beliefs about ability

[13] For a discussion of these issues in discrete choice models in general, see McFadden and Train (2000).

[14] Using cross-sectional data, one approach of identifying tastes under weaker assumptions is to directly elicit beliefs from students about their tastes, for example, their beliefs about "enjoying studying" a major. Zafar (2013) uses this approach. While this approach unpacks some of the taste components that are otherwise in the residual, it is hard to elicit beliefs for all the relevant taste components. Therefore, the residual term could still include certain unobserved components of tastes.

and post-graduation outcomes. 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 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, 2012, 2014; Zafar, 2011) where the separation between beliefs is much longer, typically months or years. We can then credibly claim that the utility function, including the individual and major specific taste parameters $\gamma_{k,i}$ and the distribution of the $\eta_{k,i}$ preference shocks, are truly time invariant in our context, and that our experimentally derived panel data satisfies the standard fixed effects assumptions.[15] 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).[16] 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 prospects and spousal characteristics, and intensive (expected hours per week) and extensive (expected probabilities of full or part-time employment) margins of future labor supply decisions.

An additional advantage of the experimentally derived panel data is that we can recover a non-parametric distribution of relative tastes for each major $\gamma_{k,i} - \gamma_{\tilde{k},i}$. Using only the cross-section pre-treatment data does not allow separate identification of tastes from beliefs about ability and future post-graduation outcomes. The lack of identification holds since we can fully rationalize the data on expected choice probabilities as $\alpha = 0$ and $u(X) = 0$ for any vector X and $r_{k,i} = (\gamma_{k,i} + \gamma_{\tilde{k},i}) + \epsilon_{k,i}$. Separately identifying tastes from other model elements could be achieved through a parametric restriction on the joint distribution of taste parameters $\gamma_{k,i}$, as in, for example, Berger (1988) and Arcidiacono (2004), or Beffy et al. (2011), by assuming an extreme value or normal distribution of tastes. In our setup, we avoid making parametric assumptions about tastes, and allow correlated tastes across majors. Emphasizing the empirical importance of this generality, as discussed below, we estimate taste distributions which are quite different from commonly assumed normal or extreme value distributions.

[15] The disadvantage of this approach relative to these other studies is of course that we cannot study the belief formation process over the long term. Below we do discuss the persistence of the information treatment using a follow-up study of our original sample.

[16] 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. Another possibility 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.

4 Data

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

4.1 Administration

Our data is from an original survey instrument administered to New York University (NYU) undergraduate students over a 3-week period, during May-June 2010. NYU is a large, selective, private university located in New York City. The students were recruited from the email list used by the Center for Experimental Social Sciences (CESS) at NYU. The study was limited to full time NYU students who were in their freshman, sophomore, or junior years, were at least 18 years of age, and US citizens. Upon agreeing to participate in the survey, students were sent an online link to the survey (constructed using the SurveyMonkey software). The students could use any Internet-connected computer to complete the survey. The students were given 2-3 days to start the survey before the link became inactive, and were told to complete the survey in one sitting. The survey took approximately 90 minutes to complete, and consisted of several parts. 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.

In early 2012, we conducted a follow-up survey of a sub-sample of the initial survey participants.

4.2 Survey Instrument

Our instrument (in the initial survey) consisted of three distinct stages:

1. In the Initial Stage, respondents were asked about their population and self beliefs.

2. In the Intermediate Stage, respondents were randomly selected to receive 1 of 4 possible information treatments shown in Appendix Table A1.[17] The information was reported on the screen and the respondents were asked to read this information before they continued. Respondents were then re-asked about population beliefs (on areas they were not provided information about) and self beliefs.

[17] 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.

3. In the Final Stage, respondents were given all of the information contained in each of the 4 possible information treatments (of Table A1). After having seen this information, respondents were then re-asked about their self beliefs.

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

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.

4.3 Sample Selection and Descriptive Statistics

A total of 501 students participated in the initial 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.

4.4 Subjective Data Beliefs about Major-Specific Determinants

The next section discusses beliefs about population and self earnings at age 30, and probabilistic major choice at length. The model outlined in section 2, however, also includes several other determinants.

Individual's utility for each college major is allowed to depend on the student's perceived ability. We asked respondents about their ability beliefs in each of the majors.[18] Appendix section C.1 provides descriptive statistics for ability beliefs, and revisions in ability beliefs after the information treatment.

Post-graduation utility depends on life-cycle consumption. As explained above, consumption depends on the individual's beliefs about the future distribution of earnings, marriage, labor supply, and potential spouse's earnings and labor supply. We elicited beliefs for each of these objects for each potential major. To incorporate lifetime consumption in the model (and to allow for the possibility that students believe earnings growth may vary across majors), we asked students about full time earnings beliefs for each major at three ages: immediately after graduation, age 30, and age 45. Futhermore, since uncertainty about future earnings could play a role in educational choices (Altonji, 1993; Saks and Shore, 2005; Nielsen and Vissing-Jorgensen, 2006), our model directly incorporates it. Besides elicititing respondents' expected future earnings at various ages, we also elicited multiple points on the respondents' major-specific self earnings distribution:[19] We asked respondents about the percent chance that their own earnings would exceed \$35,000 and \$85,000 at both ages 30 and 45.[20]

[18]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?*"

[19]Most existing 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).

[20]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?*"

To capture potential differences in work hours across majors, we asked respondents about the expectations regarding 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. Finally, since consumption depends on marriage and spouse's labor supply, we also collected 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. The data are described in Appendix section C.2.

5 Reduced-Form Analysis

In this section, we examine patterns in beliefs, focusing on beliefs about the population average earnings and self expected earnings of the individual at age 30. We document a strong and logical causal effect of our information treatment on earnings revisions. The section also examines how changes in self-reported beliefs about majoring in different fields relate to changes in beliefs about own future earnings in these fields. In the following section, we report estimates from a structural life-cycle utility model which incorporates additional beliefs data, including earnings at other ages, ability, labor supply, and spousal earnings.

5.1 Earnings Beliefs and Belief Updating

5.1.1 Population Beliefs About Earnings

Columns (1a) and (2a) 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".[21] In column (1a), we see that the mean belief about age 30 female full-time earnings varies from \$34,600 for college drop-outs to \$79,600 for graduates with degrees in economics or business. Students believe humanities and arts majors have the lowest average earnings among the graduating majors (\$56,000). Engineering and computer science graduates are believed to have earnings close to economics and business, followed by natural science majors. Beliefs about males also follow a similar pattern. While the mean beliefs reported by males are higher than those reported for females for each of the five fields, the differences are not statistically significant. There is also considerable heterogeneity in beliefs as indicated by the large standard deviation in beliefs about the population mean for both women and men. For example, for the economics and business field, the 5th percentile of the male belief

[21] 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?*".

distribution in our sample is $10,000, the 50th percentile is $76,500, and the 95th percentile is $150,000.

Errors in Population Beliefs Columns (1b) and (2b) 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. Since errors can be both positive and negative, a *mean* error close to zero may not indicate a homogeneous low level of error. Therefore, we also report the absolute value of the error in columns (1c) and (2c).

Table 2 shows 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. The errors in many categories are substantial: for example, students over-estimate full-time earnings for female graduates in economics and business by 31.1 percent and for male graduates in the same field by 16.6 percent. 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 (1c) and (2c) of the table). As we show in the Appendix (Table A2), the heterogeneity in errors is quite striking: for example, the median error regarding full time females' earnings in engineering/computer science is +10.1% (that is, under-estimation of 10.1 percent), while the 10th percentile is -33.2% and the 90th percentile is +46.7%.

5.1.2 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.[22] 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. 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. Like the population beliefs, there is substantial heterogeneity in self beliefs, as seen in the large

[22]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?*"

standard deviations (relative to the means). The Appendix (Table A2) shows more information on the distribution of self earnings. Median self earnings, for example, in economics/business are $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 percent change (post- minus pre- treatment) in self beliefs about earnings. There is considerable heterogeneity in the revisions of self beliefs, with the average percent revision varying from about -12 percent (downward revision) to +33 percent (upward revision). Average revisions in the two highest earning categories –economics/business and engineering/computer science– are negative, while average revisions in the lowest earning field –the not graduate category– are positive and large. As indicated by the standard deviations, within categories there is also considerable heterogeneity.[23] The third column of Table 3 shows that mean absolute revisions are substantially larger than mean revisions, varying between 25.5 and 43 percent.

5.1.3 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 students react causally to the new information the experiment provides.[24] 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 5.1.1), 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 examine this relationship next and find evidence for this type of logical updating.

Panel A in Table 4 estimates a series of reduced form regressions. The first column, using

[23] This is further illustrated in the fourth panel of Appendix (Table A2). For example, the median percentage earnings revision in economics/business for the full sample is -14.3 percent (downward revision), while the 10th percentile is -50 percent and the 90th percentile is +20 percent.

[24] 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.

only the baseline, pre-treatment data, estimates the regression:

$$\underbrace{\ln w_{k,i}}_{\text{self-belief}} = \beta_0 + \beta_1 \underbrace{\ln \bar{w}_{k,i}}_{\text{pop. belief}} + \epsilon_{k,i},$$

where the dependent variable is individual i's (log) expected self earnings in each field, and $\ln \bar{w}'_{k,i}$ is i's (log) belief about the population average earnings in that field. We pool all of the majors together, and include separate intercepts or major-specific fixed effects (dummy variables). 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 0.31 indicates that a 1 percent increase in population beliefs about average earnings increases beliefs about own earnings by 0.31 percent. The R-squared reported for the regression in the first column indicates that nearly 42 percent of the variation in self earnings beliefs is explained by population earnings beliefs and major-specific dummies.

Columns (2) in Panel A of Table 4 examines whether the *revisions* in self-earnings are related to *errors* in population beliefs. We regress log earnings revision in self earnings (post minus pre-treatment) on the log relative error about population earnings (log(truth) - log(belief)), that is:

$$\underbrace{(\ln w'_{k,i} - \ln w_{k,i})}_{\text{beliefs revision}} = \beta_0 + \beta_1 \underbrace{(\ln \bar{w}^*_k - \ln \bar{w}_{k,i})}_{\text{error}} + \epsilon_{k,i}. \tag{7}$$

This regression indicates the extent to which the information treatments we provide influence individual beliefs about self earnings. Causal revisions in response to information would imply a positive relationship between the two. In fact, the coefficient estimate is positive and statistically significant at the 1 percent level. The 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. We obtain a qualitatively similar estimate in column (3) where the specification also includes major dummies.[25]

Panel B of Table 4 reports estimates of the same specifications as in Panel A, but restricts

[25] As a robustness check, we also estimate the specifications reported in columns (2) and (3) on the sample that drops outliers. That is, 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. We obtain estimates that are similar to those for the full sample.

the sample to underclassmen (that is, students who are freshmen or sophomores). One may expect students earlier in their college career to find information about population earnings more valuable, and hence more responsive to such information. That, however, does not seem to be the case: the estimates in Panel B are very similar to those in Panel A (and we fail to reject the equality of the coefficients in each of the specifications).

The estimates in columns (2)-(3) 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.[26] However, like most data, subjective data suffer from measurement error. Therefore, one concern in using these panel estimators is that measurement error would be exacerbated using differences. 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 Appendix D.1, we present two additional pieces of evidence that further indicate that measurement error is not a concern in our data. One, using the intermediate stage of our study design – where students were randomly provided with population earnings information that varied in its specificity (for example, labor market outcomes of all workers, versus outcomes of college-graduate workers by gender and field of study) – we show that students' self earnings beliefs are more responsive to information that is more specific. Second, using data on beliefs of a Control group – a set of students who report their self beliefs twice but are not provided with any new information – we show that the data yield a reliability ratio of 0.984, indicating that our estimated OLS coefficients are only attenuated by 1.6 percent from the true value.

5.2 Major Choice and Post-Graduation Utility

5.2.1 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.[27] The top panel of Table 5 provides descriptive statistics of the expected major field probabilities. The first column shows that the most likely major is humanities/arts at 42.6 percent, followed by economics/business at 30 percent. The probability of not graduating is less

[26] In addition, the strong relationship between beliefs about earnings and expected major choice pre-treatment that we document in the next section (section 5.2.2), and the non-zero and logical pattern in updating that we observe, where revisions (post - pre treatment) in relative self earnings are *positively* correlated with major choice probability, also cast doubt on measurement error being a serious issue in the data.

[27] 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.

than 3 percent.[28]

The next two columns of Table 5 present the average revisions in students' expected probability of majoring in each of the majors. Column (3) shows that the mean of the distribution of log odds changes is positive for all fields, 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.

The large standard deviations in revisions indicate that the response of information on choice probability revisions are quite heterogeneous. In fact, the larger absolute revisions reported in column (4) indicate that non-trivial numbers of students revise their choice probabilities both upwards and downwards.[29] This is further highlighted in Figure A2, which provides the post minus pre- treatment change in log beliefs for students about majoring in each field (relative to humanities): $r_{k,i} - r'_{k,i}$. While the mean of the distribution of log odds changes is positive for all fields (see column (3) of Table 5), Figure A2 indicates that a substantial number of respondents revised their expected relative major choice downward, and believed they were more likely to major in humanities/arts relative to the other majors. The largest upward changes occurred for the high earning fields (economics/business and engineering/computer science). For example, the average log odds of majoring in economics/business increased by 46 percentage points, and the log odds of majoring in engineering/computer science relative to humanities increased by 72 percentage points.

Column (5) of Table 5 shows that, before the information treatment, a sizable number of students provide corner probabilities (that is, a probability of zero or 100) for majoring in the field. For example, 36% of students assign a zero or 100 percent likelihood of majoring in economics/business, and 53.7% of students assign a zero or 100 percent likelihood of majoring in engineering/computer science. However, after the provision of information, column (5b) shows that the proportion of corner probabilities declines significantly for each of the graduating majors (with the differences being statistically significant in each of the cases, using a Chi-square test).

The lower panel of Table 5 restricts the sample to those cases where a corner probability was provided in the initial stage. Compared to the full sample, the revisions as well as log odds revisions of this sample are similar. This indicates that information provision led students, even those who were fairly certain about their choice probabilities at the baseline, to revise their probabilistic choices.

[28] 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.

[29] About 1/3 of the sample reported no change in the probability of majoring in any of the fields following the information treatment.

5.2.2 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 column in Panel A of Table 6 estimates a reduced form regression 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 independent variable. The regression takes the form:

$$(\ln \pi_{k,i} - \ln \pi_{\tilde{k},i}) = \beta_0 + \beta_1(\ln \bar{w}_{k,i} - \ln \bar{w}_{\tilde{k},i}) + C_i'\delta + \nu_k + \psi_{k,i}, \tag{8}$$

where $\pi_{k,i}$ is i's subjective probability of graduating with major k, $\bar{w}_{k,i}$ is i's 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 ($\psi_{k,i} = \gamma_{k,i} - \gamma_{\tilde{k},i} + \epsilon_{k,i}$) 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. The estimate indicates 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 (including those not chosen), this type of regression avoids the selection issue inherent in using actual major choice and the actual earnings in that chosen major.

The regression in the first column of Table 6 is a cross-sectional based OLS regression using only the baseline pre-treatment beliefs. As described in the identification section, the key drawback to using only baseline beliefs is that one cannot separately identify the taste component from earnings components. In this regression, the residual contains individual components reflecting individual variation in tastes for each of the majors. Therefore, a concern is the cross-sectional estimate 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 (2) of Table 6 estimates the reduced form model (8) in individual (within) differences to net out the individual taste components ($\gamma_{k,i} - \gamma_{\tilde{k},i}$):

$$\begin{aligned}&[(\ln \pi'_{k,i} - \ln \pi'_{\tilde{k},i}) - (\ln \pi_{k,i} - \ln \pi_{\tilde{k},i})] \\ &= \beta_0 + \beta_1[(\ln \bar{w}'_{k,i} - \ln \bar{w}'_{\tilde{k},i}) - (\ln \bar{w}_{k,i} - \ln \bar{w}_{\tilde{k},i})] + \nu_k + \epsilon'_{k,i} - \epsilon_{k,i},\end{aligned} \tag{9}$$

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 (8).

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 (3) 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 estimate in Columns (1) 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.

As discussed in Section 2, tastes in this framework also implicitly reflect the "switching costs" of changing majors while in school. As students progress through college, it may become more costly for them to switch majors. This could then lead us to obtain a smaller choice elasticity when differencing out the individual taste component in equation (9). In Panel B, we therefore restrict the sample to underclassmen (freshmen and sophomores), for whom arguably the switching costs are much lower. The estimate for equation (8) for this subsample is similar to that for the full sample, and we cannot reject the equality of the estimate in columns (1) for the two panels. However, consistent with switching costs being larger for students further along in their college career, the estimates based on equation (9) are larger for underclassmen: the choice elasticity based on the FE estimate for underclassmen is nearly twice that of the full sample, with the difference between the two estimates in column (2) being statistically different from zero at the 5% level. The FE estimate is, however, still significantly smaller than the cross-sectional OLS estimate. This suggests that, even when restricting the sample to individuals for whom switching costs are low, the tastes components are positively correlated with earnings beliefs, which upwardly biases the estimate in the cross-section.

Before we move to the estimation section, it is worth noting that we find strong suggestive evidence of the effect of the information on both self beliefs and major choice to be persistent. Section D.2 of the Appendix discusses the results from a follow-up survey, where we find that students' follow-up self beliefs and choices are more strongly correlated with the final stage beliefs and choices from the first survey, than with initial stage beliefs.

6 Structural Estimates

We next turn to estimating a structural model of major choice. In the previous sections, our reduced-form 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 model.

6.1 Estimation

We estimate the parameters of the utility function using the pre- and post- information beliefs. From each individual's elicited belief distributions, we calculate expected utility from (3) using simulation. We estimate the model using a non-linear least squares (NLS) estimator. With estimates of the model parameters in hand, we then "back out" the taste for each major $\gamma_{1i}, \ldots, \gamma_{Ki}$ for all i (individual and major specific fixed effects).[30]

6.2 Parameter Estimates

While the model estimates are more interpretable in terms of implied choice elasticities and decompositions of college major choices (presented below), we first briefly discuss the model parameters presented in Table 7. The marginal utility of consumption (either for the individual or potential spouses) is given by $\phi_j c^{-\rho_j}$, where $j = 1, 2$. We estimate ϕ_1 for own consumption to be 0.21 and the curvature parameter (relative risk aversion) ρ_1 to be 4.96.[31] Own value of spouse's consumption has values of ϕ_2 and ρ_2 which indicate the utility value of spouse's consumption to the individual is about the same as own consumption but has less curvature. Although this suggests a high value placed on marriage and spousal characteristics, in the decompositions we report below, marriage and spousal characteristics are not very different across major categories and are therefore only a small factor in major choice. The estimates on risk aversion are on the high end of previous estimates, but similar to the estimate in Nielsen and Vissing-Jorgensen (2006). The high ρ estimates 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.

Table 7 also provides the mean and standard deviation of the estimated non-parametric distribution of relative tastes (relative to humanities which is normalized to $\gamma_{\tilde{k},i} = 0$ for all i). The

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

[31] Note that the ϕ_1, ϕ_2 parameters on consumption are identified given that the post-graduation utility (expected discounted consumption) is only *one* part of the utility from each major, and the ϕ_1 and ϕ_2 terms measure the relative importance of post-graduation utility versus the other model components, including ability and tastes.

mean of the relative taste distribution for each major is negative, indicating that even conditioning on ability and post-graduation outcomes beliefs, students on average prefer humanities. The standard deviation of the beliefs is large indicating substantial taste heterogeneity.[32] We explore observable correlates of this heterogeneity below.

6.3 Choice Elasticities

The structural model estimates are more easily interpreted in terms of what the estimated models imply about the responsiveness of major choices to changes in self earnings. For each major and student, we compute the elasticity $\xi_{k,i}$ by increasing expected earnings by 1 percent in every period. As discussed above, the choice elasticities are in general heterogeneous across students given their different baseline beliefs, heterogeneity in major specific tastes, and their different years in school which affects the cost of switching to different majors.

Table 8 displays the average elasticity implied by the estimated model using the full sample and for a separate estimation of the structural model using only the sub-sample of freshman and sophomores.[33] With the exception of the drop-out (non-graduate) alternative, we estimate average elasticities of between 0.036 and 0.062 for the full sample, and elasticities between 0.04 and 0.07 for the sample of freshman and sophomore students.[34] 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). For example, Beffy et al. (2011), using data on French students, estimate earnings elasticities of between 0.09-0.12, depending on the major.

Elasticities by Year in School For all majors, we find that average earnings elasticities are higher for the freshman and sophomore students than for the full sample.[35] This finding for the structural model estimates mirrors those in the reduced form estimates using only age 30 earnings. However, there are two things to note when comparing the structural model estimates of earnings elasticities for younger vs. older students. First, because graduation is farther away for freshman and sophomores than for junior students, future earnings changes are discounted more heavily, hence this factor pushes the elasticity lower for freshman and sophomore students.

[32] Figure A3 provides a direct look at the distribution of tastes for underclassmen (freshmen and sophomores) and the full sample, respectively. Both distributions show some bimodality, but the most frequent mode is near 0 for the two groups.

[33] For brevity, these structural estimates are not reported but are available on request.

[34] The high elasticity for drop-out alternative is due primarily to the relatively low level of expected earnings in this major and the concavity of the utility function with respect to consumption.

[35] The two panels in Figure A4 graph the distribution of the $\xi_{k,i}$ choice elasticities for the full sample and the underclassmen. The distribution for underclassmen, relative to the full sample, is shifted to the right, indicative of more respondents among younger students with higher (but still inelastic) response to changes in earnings. From the figures, 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.

This is one reason why we find lower elasticities using the structural model estimates than we do for the reduced form estimates which do not account for the life-cycle profile of earnings. On the other hand, as discussed above, the taste component, which partly reflects the cost of switching to another major, is much smaller on average for freshman and sophomore students, and this factor pushes the elasticities higher for younger students.

Elasticities using Only Cross-Sectional Data We also estimate another set of models using only the pre-treatment cross-sectional data.[36] 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. Consistent with the simple reduced form results above, the choice elasticities for most majors using the cross-sectional data 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.

Elasticities assuming Risk Neutrality We estimate another version of the model assuming risk neutrality: $\rho_1 = \rho_2 = 0$. Recall that we elicit beliefs of earnings uncertainty and estimate a large degree of risk aversion. The earnings elasticity estimates assuming risk neutrality (reported in the last two columns of Table 8) are several times higher than the estimates in the unrestricted model in which we estimate a high degree of risk aversion (and larger than estimates in the cross-sectional data model as well). The large difference in earnings elasticity estimates indicates the importance of risk aversion and earnings uncertainty; ignoring these elements of the model greatly inflates the responsiveness of individuals to changes in expected mean earnings. The importance of uncertainty is particularly apparent in the elasticity estimates for the higher mean earning majors, business and engineering, where the students expect higher mean earnings but also higher uncertainty in earnings. Also, in general, we find that juniors have slightly less uncertainty in earnings than freshman and sophomores, hence assuming risk neutrality would have the tendency to raise the earnings elasticity more for upperclassman than for freshman and sophomores.

[36] This model includes only the $\eta_{k,i}$ preference shocks and sets $\gamma_{k,i} = 0$ for all k. The model therefore assumes that tastes for each major are independent across i and k and distributed extreme value (according to the taste shocks $\eta_{k,i}$). 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.

6.4 Correlates of Tastes

In the preceding analysis, the $\gamma_{k,i}$ taste components are essentially a "black box." We next investigate the observable correlates of major-specific tastes. Table 9 reports the OLS estimates of a series of regressions of tastes for each major (relative to humanities/arts) onto various demographic characteristics and ability measures. Three patterns are of note:

First, there are substantial demographic difference in tastes for majors, even accounting for differences in ability and post-graduation beliefs. Relative to females, males have significantly stronger positive tastes for all the other major categories (relative to humanities/arts). This indicates that even with rich data on student expectations associated with majors, we still cannot explain most of the gender gap in major choice which is the subject of considerable prior literature (Brown and Corcoron, 1997; Weinberger, 1998; Wiswall, 2006; Zafar, 2013). In addition, the coefficient for Asian respondents is significantly positive for all major categories, indicating a distaste for humanities/arts.

Second, 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.

Third, consistent with the findings of different choice responsiveness for freshman and sophomore students relative to junior students, we find that junior students have significantly more negative tastes for engineering/computer science and natural sciences (relative to humanities) than freshman or sophomore students. We find negative, but statistically insignificant, coefficients for junior students' relative tastes in the remaining majors. As these tastes represent in part the cost of switching majors, this pattern of particularly high switching costs for math and science fields 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, 2014), and (ii) patterns of major switches that indicate that students switch out of math, science, and engineering (Stinebrickner and Stinebrickner, 2014; Arcidiacono, Aucejo, and Spenner 2012; Arcidiacono, 2004).

Overall we find that tastes are correlated with gender, race, pre-college measures of ability, and school year. These results suggest that different populations can have very different distributions of major specific tastes, and therefore different responsiveness to changes in post-graduation outcomes such as future earnings. Replicating our analysis for other populations is an important area of future research to understand the external validity of our results. Given the correlation of tastes with ability in particular, our results suggest that lower ability populations than our sample could have quite different preferences for majors.

These results also 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 an unrestricted distribution for tastes is robust relative to other approaches which restrict the distribution of tastes to a particular parametric distribution depending.

6.5 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. Table 10 reports the choice probability at each stage of the decomposition averaged over all of the sample respondents.

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 1.040, or a +0.040 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 not graduating from a baseline of 1 to 0.914 (-0.096 reduction) given the expected loss in future earnings from dropping out of college.

Columns (2) through (5) progressively add other model components, and the entries in Table 10 reflect the marginal gain of each component, given the other preceding components are included. Thus, adding beliefs about own ability in Column (2) decreases the odds of majoring in economics/business from 1.040 (including beliefs about own earnings) to about 1.023 (including both beliefs about own earnings and own ability). The negative sign on the own ability component indicates that individuals perceive greater difficulty in completing other majors relative to humanities/arts.

Column (3) of Table 10 re-introduces beliefs about own work hours for each major, and Column (4) adds spousal characteristics, including the probability of marriage, spousal earnings, and spousal hours. Neither of these factors plays a substantial role in major choice, after accounting for earnings and ability differences, with the exception of marriage market "loss" if the students were to drop-out of school.

Finally, Column (5) adds the remaining determinant of major choice, the vector of estimated major-specific tastes. The negative sign on this component indicates that, on average, students have high distaste for these majors (relative to humanities/arts). The large magnitude of this component indicates, that even accounting for all other factors, the residual taste component still explains the vast majority of major choices.

6.6 Welfare Analysis

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. We find that on average they revise their self beliefs and choices logically when provided with accurate information. A common, simple, and relatively assumption-free method to assess welfare would be to use ex-post realized outcomes. In the reduced form analysis using our follow-up 2 years later, we present suggestive evidence that the information treatment affects long term beliefs. However, because our sample is still too young for us to observe many of the important post-graduation outcomes, we lack the necessary ex post outcomes to evaluate welfare in the standard way.[37]

As an alternative, we assess the welfare gains from our experiment based on the change in pre- and post-treatment *expected* utility, using the respondent's beliefs and our estimates of preference parameters. Define $\bar{V}'_{k,i} = \alpha \ln a'_{k,i} + EV'_{1,k,i}$ as the post- treatment expected utility for individual i from major k, where we omit the treatment invariant components $\gamma_{k,i}$ and the resolvable uncertainty preference shock $\eta_{k,i}$. As above, pre- and post-treatment beliefs about the probability of completing major k are given by $\pi_{k,i}$ and $\pi'_{k,i}$. Our measure of the welfare gain for student i is:

$$\Delta_i = \sum_k (\pi'_{k,i} - \pi_{k,i})\bar{V}'_{k,i}.$$

Note that the true effect of our intervention is through its impact on expected probabilistic major choices; with our measure just revising beliefs (e.g. about expected earnings) by itself would not lead to welfare gains if students did not revise their probabilistic choices.

Using the full sample, we find that 78.1 percent experience a non-negative welfare gain ($\Delta_i \geq 0$) and 21.9 percent experience a welfare loss ($\Delta < 0$). These statistics are similar

[37]Note also that we cannot directly use our structural model estimates to solve for the ex post "true" distribution of outcomes (e.g. the realized distribution of earnings). By design, our model makes no assumptions about the ex post distribution of realized outcomes. Our model and data collection is based on beliefs about post-graduate outcomes, not actual outcomes, and we can only use our model to solve for major choices under various counterfactual configurations of preferences and beliefs, as we have done in the exercises above.

for the freshman and sophomore sample, where 77.2 percent have a non-negative welfare gain. To provide a meaningful monetary measure of the welfare gains, we compare the gains from the information treatment with an alternative experiment in which we add $1,000 in expected earnings to each major in each year. This experiment is conducted using baseline beliefs to approximate the pre-treatment value of additional income to agents. Taking the ratio of the information experiment gains to the gain in welfare from this alternative experiment yields an average annual monetary gain from the information experiment of $6,267 for the full sample and $3,665 for the freshman and sophomore samples.[38] At baseline age 30 expected earnings, these gains are equivalent to a 6.4 percent gain for the full sample and a 5.6 percent gain for the freshman and sophomore sample.

We should emphasize that there is substantial heterogeneity in welfare changes. Given that 32.4 percent (full sample) and 29.8 percent (freshman and sophomores) do not update their major choice beliefs at all, the median welfare gain is 0. And, while a clear majority of the sample has non-negative gains, we still find that a non-trivial number of individuals are "worse off" following the information treatment using our measure (i.e. $\Delta_i < 0$). This result may seem at odds with the notion that providing students with accurate information can only be welfare enhancing. There are several issues to consider. First, by necessity, we approximate welfare gains using an estimated utility function which nonetheless can still be a poor approximation for utility for some students. Second, providing accurate information may have various behavioral effects on information updating. While we find that on average students update their beliefs logically in response to the information treatment, not all students do so, and those who do not sensibly update their beliefs are likely not to have positive welfare gains. However, in our sample, where we regress the level of welfare gains or losses (Δ_i) on the fraction of majors for which the respondent "logically" updates age 30 earnings beliefs, where "logical" updating is defined as revising upward (downward) self beliefs about earnings when the information treatment reveals an under- (over-) estimation of population earnings, we do not find a statistically significant relationship.[39] In an analysis in which we regress Δ_i welfare changes on respondent characteristics (using the same variables as in Table 9 including gender, race, SAT scores, and parental education), we find that in the full sample none of the variables has a significant relationship with the welfare change measure, and we cannot reject that the

[38] Note that there are some large outliers here, and the average gain in the sub-sample with gains or losses not exceeding $20,000 is $1,759 (full sample) and $1,887 (freshman and sophomore sample). Note also that differences in the marginal utility of income directly influence these welfare measures. Consistent with our results above of a higher major choice earnings elasticity for freshman and sophomores, there is a higher welfare gain from an increase in expected earnings for the freshman and sophomore samples. Therefore, the denominator of the welfare change measure, which measures how much utility increases with additional earnings, is higher for the freshman and sophomore samples. In terms of "raw" welfare differences, the average Δ_i difference in utility is actually higher for the freshman and sophomore sub-sample than in the full sample.

[39] See our companion paper, Wiswall and Zafar (2014), for a more detailed analysis of updating behaviors in response to population information.

variables are jointly insignificant at the 5 percent level (F-statistic of 1.45). In the freshman and sophomore sub-sample, men, black, and high SAT verbal scoring students have somewhat lower welfare gains than other students, but no other variables are significant from zero at the 5 percent level. The R-square in this regression is 0.05 (full sample) and 0.09 (freshman and sophomore sample). We conclude that while there is considerable heterogeneity in welfare changes to the information experiment, and most saliently that there are welfare losses for a sizable minority of students, observable characteristics explain little of this variation.

Further research is necessary to know if these expected gains will be realized, and if similar average gains are possible with other types of information interventions. But the low cost of information provision, the large misinformation about objective population returns, and sizable average welfare gains from the intervention, suggest a policy role for campaigns that provide accurate information on the returns to human capital investments. Such campaigns have been conducted in developing countries (Jensen, 2010; Nguyen, 2010), but our results make a case for such interventions in developed countries as well. Population errors in our high ability sample are sizable; in other settings, such as disadvantaged populations, errors may be even higher and hence information dissemination may have a larger impact. Furthermore, in order to understand the underlying determinants of choice behavior and the channels through which such interventions affect behavior, such interventions should be accompanied with collection of rich data on subjective expectations.

7 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 robustly identify the choice model. Second, in addition to data on beliefs about earnings and ability, we collect rich data on beliefs about earnings uncertainty, labor supply, marriage, and spousal characteristics; all of which we directly incorporate into a life-cycle framework.

We find that, in the context of major choice, earnings expectations and ability perceptions both play an important role in choice of major. Marriage, spousal characteristics, and labor supply considerations play a relatively minor role in major choice. However, even with our rich data on beliefs across a variety of pecuniary and non-pecuniary aspects of majors, major choices in our data are still largely the result of heterogeneity in major specific and unobserved "tastes." In our framework, tastes encompass preferences 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). We present evidence that the distaste for humanities is stronger for male, Asian, and high-SAT Math score respondents. In addition, upperclassman have stronger tastes, likely reflecting their higher cost of switching between majors at this later stage in college. Understanding the origins of differences in tastes is not investigated in the current study. This is a challenging task since 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); we believe this is an important area of future research.

Our results suggest several possible avenues for future work. First, the current framework does not incorporate savings and borrowing. Given the increasingly important role of student loans in financing higher education and rising student loan debt (Lee et al., 2014), and labor market returns that vary significantly by college major, a policy-relevant and useful extension would be to allow debt and consumption levels to be endogenous in our framework. Second, given the apparent importance of work flexibility in occupational choices (Goldin, 2014) and the large differences in major choices by gender, our model could be extended to incorporate additional data on students' perceptions about anticipated work arrangements associated with college majors; this would then allow us to study how these dimensions impact relative preferences for certain fields. Finally, a useful extension of our work is to combine our stylized information experiment with a longer-term panel on beliefs and choices (as in Stinebrickner and Stinebrickner, 2014), and with data on subsequent realizations. This will allow us to (i) investigate the long-term effects of information, (ii) relax our model assumption that uncertainty (with respect to earnings) remains unresolved between the time of the survey and graduation, and (iii) compare expectations data with actual realizations, providing a better measure of welfare.

References

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

[2] Altonji, Joseph, 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, Esteban Aucejo, and Ken Spenner. 2012. "What Happens After Enrollment? An Analysis of the Time Path of Racial Differences in GPA and Major Choice." IZA Journal of Labor Economics, 1(5) (October).

[4] Arcidiacono, Peter. 2004. "Ability Sorting and the Returns to College Major." Journal of Econometrics, 121(1-2): 343-375.

[5] Arcidiacono, Peter, Joseph Hotz, and Songman Kang. 2011. "Modeling College Major Choices using Elicited Measures of Expectations and Counterfactuals." Journal of Econometrics, forthcoming.

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

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

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

[9] 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.

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

[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. 2011. "Women's College Decision: How Much Does Marriage Matter?" Journal of Labor Economics, 29(4): 773-818.

[23] Gemici, Ahu, and Matthew Wiswall. 2014. "Evolution of Gender Differences in Post-Secondary Human Capital Investments: College Majors at the Intensive Margin." International Economic Review, 55 (1): 23-56.

[24] Goldin, Claudia. 2014. "A Grand Gender Convergence: Its Last Chapter" American Economic Review, 104(4): 1091-1119.

[25] 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.

[26] Giustinelli, Pamela. 2012. "Group Decision Making with Uncertain Outcomes: Unpacking Child-Parent Choice of the High School Track." University of Michigan Working Paper.

[27] 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.

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

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

[30] Kaufmann, Katja. 2012. "Understanding the Income Gradient in College Attendance in Mexico: The Role of Heterogeneity in Expected Returns to College." Working Paper.

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

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

[33] Lee, D., W. der Klaauw, A. Haughwout, M. Brown, and J. Scally. 2014. Measuring Student Debt and Its Performance. Federal Reserve Bank of New York Staff Reports, 668, April 2014.

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

[35] 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.

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

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

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

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

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

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

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

[43] 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.

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

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

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

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

[48] Stinebrickner, Todd, and Ralph Stinebrickner. 2012. "Learning about Academic Ability and the College Drop-out Decision." Journal of Labor Economics, 30(4): 707-748.

[49] Stinebrickner, Todd, and Ralph Stinebrickner. 2014. "Math or Science? Using Longitudinal Expectations Data to Examine the Process of Choosing a College Major." Review of Economic Studies, 81(1): 426-472.

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

[51] 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.

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

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

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

[55] 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.

[56] 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.

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

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

[59] Zafar, Basit. 2013. "College Major Choice and the Gender Gap." Journal of Human Resources, 48(3): 545-595.

[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.

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

		(1a)	(1b)	(1c)	(2a)	(2b)	(2c)
		Beliefs about Women			Beliefs about Men		
		Belief	Percent Error[a] $\left(\frac{\text{Truth - Belief}}{\text{Truth}}*100\right)$		Belief	Percent Error $\left(\frac{\text{Truth - Belief}}{\text{Truth}}*100\right)$	
			Actual	Abs		Actual	Abs
Economics/Business	mean	7.96	-31.1**	49.22	8.69	-16.59	42.65
	(std.)	(5.63)	(92.72)	(84.48)	(7.58)	(101.67)	(93.75)
Engineering/Comp. Sci.	mean	7.08*	5.66	31.99	7.98	3.18	36.37
	(std.)	(5.03)	(66.96)	(59.07)	(7.77)	(94.30)	(87.05)
Humanities/Arts	mean	5.60	-13.85	37.57	5.82	-9.91	33.87
	(std.)	(5.19)	(105.52)	(99.55)	(3.96)	(74.79)	(67.40)
Natural Sciences	mean	6.83	-13.81***	40.58	6.85	5.60	32.41
	(std.)	(6.44)	(107.28)	(100.25)	(4.41)	(60.70)	(51.60)
Not Graduate	mean	3.46	-0.12***	38.94	3.57	25.28	35.31
	(std.)	(3.03)	(87.52)	(78.35)	(1.77)	(37.01)	(27.58)

Notes: Beliefs (columns 1a, & 2a) are in $10,000's. Other columns are percentages.
Pairwise ttests conducted for equality of means between columns (1a) and (2a); (1b) and (2b); (1c) and (2c). ***, **, * denote significance at the 1, 5, and 10
[a] Percent Error is defined as 100* (truth-belief)/truth.

Table 3: Age 30 Earnings and Earnings Revisions

		(1a)	(1b)	(1c)
		Self earnings pre	Self % revision $\left(\frac{\text{Post-Pre}}{\text{Pre}}*100\right)$	Absolute Self % revision
Econ/Bus	mean	12.69	-12.12	27.93
	(std.)	(14.17)	(41.87)	(33.43)
Eng/Comp Sci	mean	9.78	-2.62	26.39
	(std.)	(8.49)	(40.79)	(31.19)
Hum/Arts	mean	6.87	2.70	25.45
	(std.)	(6.81)	(39.75)	(30.63)
Natural Sci	mean	9.34	-0.70	28.19
	(std.)	(9.92)	(43.11)	(32.60)
Not Graduate	mean	3.93	33.42	43.31
	(std.)	(7.59)	(59.51)	(52.74)

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

Table 4: Population and Self Beliefs

	(1)	(2)	(3)
Dependent Var:	Log Self Earnings	Log Earnings Revision (Post-Pre)	
	Indiv. Covaritates & Major Dummies Included	Major Dummies Not Included	Major Dummies Included
Panel A: Full Sample			
Log Pop Earnings Beliefs	0.309***		
	(0.0251)		
Log Pop Earnings Errors		0.0786***	0.0689***
		(0.0194)	(0.0195)
R-squared	0.416	0.014	0.035
Total Observations	2440	2440	2440
Individuals	488	488	488
Panel B: Underclassmen (freshmen and sophomores)			
Log Pop Earnings Beliefs	0.305***		
	(0.0282)		
Log Pop Earnings Errors		0.0713***	0.0624***
		(0.0216)	(0.0219)
R-squared	0.419	0.012	0.033
Total Observations	1865	1865	1865
Individuals	373	373	373

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.
Pairwise tests conducted for equality of coefficients between full sample and underclassmen. +++, ++, + denote significance at the 1, 5, and 10

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

	(1)	(2)	(3)	(4)	(5)	(5b)
	Before[a]	Revision[b]	Log Odds Rev[c]	Absolute Rev	Prop. Corner[d] before	after
Econ/Business	30.4	1.2	0.46	6.5	35.9***	25.6
	(36.1)	(12.1)	(1.98)	(10.3)		
Eng/Comp. Sci.	6.7	2.3	0.72	4.6	53.7***	40.4
	(14.4)	(8.5)	(2.20)	(7.5)		
Humanities/Arts	42.6	-3.9	-	8.1	34.2***	30.5
	(39.0)	(13.9)	-	(12.0)		
Natural Sciences	18.2	0.4	0.30	5.7	39.8***	34.4
	(27.7)	(11.5)	(1.96)	(10.0)		
Not Graduate	2.4	0.1	0.13	1.9	66.4***	65.8
	(6.9)	(5.5)	(1.93)	(5.2)		

Notes: This table reports the mean self belief about completing each of the majors. Probabilities are reported on a 0 - 100 scale. The standard deviation is in parentheses. Chi-square test conducted for equality of proportions between columns (5) and (5b). ***, **, * denote significance at the 1, 5, and 10

[a] Reported before receiving info treatments.

[b] Probability in major post-treatment - Probability in major pre-treatment.

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

[d] Proportion of corner solutions (major probability of 0 or 100).

Table 6: Graduation Expectations and Expected Earnings

	(1)	(2)	(3)
Dep. Variable:	Log Odds of Major Rel. to Hum.	Log Odds Revision (Post-Pre)	
	Indiv. Covaritates & Major Dummies Included	Full Sample (Major Dummies Included)	Truncated Sample[a]
Panel A: Full Sample			
Log Self Earnings	1.613***		
	(0.140)		
Log Self Earnings Rev		0.146#	0.275**
		(0.099)	(0.140)
R-squared	0.270	0.013	0.012
Total Observations	1952	1952	1710
Individuals	488	488	485
Panel B: Underclassmen (freshmen and sophomores)			
Log Self Earnings	1.635***		
	(0.151)		
Log Self Earnings Rev		0.262**++	0.386**
		(0.106)	(0.154)
R-squared	0.273	0.016	0.015
Total Observations	1492	1492	1310
Individuals	373	373	370

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.
All specifications in this table include major dummies.
Pairwise tests conducted for equality of coefficients between full sample and underclassmen. +++, ++, + denote significance at the 1, 5, and 10
[a] Truncated sample excludes observations where respondents revise their self beliefs by more than $50,000,

Table 7: Structural Model Parameter Estimates (Full Sample)

Utility Parameters		
Own Utility		
ϕ_1	0.210	(0.0106)
ρ_1	4.96	(0.127)
Spouse Utility		
ϕ_2	0.203	(0.0175)
ρ_2	3.23	(0.257)
Ability α	0.111	(0.0221)
Major Specific "Taste" Distribution (Taste Relative to Humanities/Arts)		
	Mean	Std.
Bus./Econ.	-0.390	4.07
Eng/Comp. Sci.	-2.12	3.26
Nat. Sci.	-1.25	3.63
No Grad	-3.03	2.73

Notes: Bootstrapped standard errors in parentheses calculated from 50 bootstrap repetitions. Standard errors in parentheses. ***, **, * denote significance at 1, 5, and 10 percent, respectively. These estimates are for the full sample including freshman, sophomores, and juniors.

Table 8: 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-Sec. Data**		**Risk Neutral Model**	
	Fresh./Soph. Students	All Students	Fresh./Soph. Students	All Students	Fresh./Soph. Students	All Students
% Δ Prob Bus./Econ.	0.0403	0.0358	0.385	0.317	0.827	0.852
% Δ Prob Eng/Comp Si	0.0603	0.0532	0.564	0.477	1.02	1.06
% Δ Prob Hum./Arts	0.0704	0.0580	0.355	0.295	0.508	0.497
% Δ Prob Nat. Sci.	0.0693	0.0618	0.528	0.466	0.807	0.835
% Δ Prob No Grad.	0.205	0.180	0.637	0.593	0.415	0.430

Table 9: Correlates of Major-specific Tastes (Relative to Humanities/Arts)

	Bus/Econ.	Eng/Comp	Nat. Sci.	No Grad.
Male	1.78***	1.47***	1.04***	1.16***
	(.365)	(.295)	(.334)	(.253)
Sophomore	.145	-.067	-.357	.531*
	(.385)	(.313)	(.351)	(.276)
Junior	-.586	-1.13***	-1.57***	-.372
	(.451)	(.347)	(.418)	(.297)
Asian	2.24***	1.54***	.797**	.956***
	(.415)	(.329)	(.383)	(.291)
Hispanic	.403	.334	.0051	-.626
	(.693)	(.525)	(.597)	(.445)
Black	.032	-.0304	-.0299	.804
	(1.07)	(.876)	(.936)	(.695)
SAT Math	.0091***	.0081***	.012***	.004***
	(.0022)	.(0020)	.(0021)	(.002)
SAT Verbal	-.0078***	-.0066***	-.0104***	-.0033**
	(.0021)	(.0018)	(.0022)	(.0016)
R-squared	0.2053	0.2124	0.1859	0.1567
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 10: 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
Bus./Econ.	1.00	0.0398	−0.0270	0.0072	0.0059	−0.266	0.759
Eng./Comp. Sci.	1.00	0.0331	−0.0696	0.0039	0.0041	−0.801	0.170
Nat. Sci.	1.00	0.0216	−0.0304	0.0040	0.0028	−0.555	0.443
No Grad.	1.00	−0.0955	−0.0769	−0.0138	−0.0596	−0.708	0.0461