

# College Major Choice and the Gender Gap\*

Basit Zafar<sup>†</sup>

October 28, 2008

## Abstract

Males and females make different choices with regards to college majors. Two main reasons have been suggested for this gender gap: differences in innate abilities, and differences in preferences. This paper studies the question of how college majors are chosen with a focus on explaining the underlying gender gap. Since observed choices may be consistent with many combinations of expectations and preferences, I collect a unique dataset of Northwestern University sophomores which contains the students' subjective expectations about choice-specific outcomes. I estimate a choice model where college major choice is made under uncertainty (about personal tastes, individual abilities, and realizations of outcomes related to the choice of major). Enjoying coursework, enjoying work at potential jobs, and approval of parents are the most important determinants in the choice of college major. Males and females have similar preferences while in college, but differ in their preferences in the workplace: non-pecuniary outcomes at college are most important in the decision for females, while pecuniary outcomes realized at the workplace explain a substantial part of the choice for males. I decompose the gender gap into differences in beliefs and preferences. Gender differences in beliefs about academic ability explain a small and insignificant part of the gap; this allows me to rule out females being low in self-confidence as a possible explanation for their under-representation in the sciences. Conversely, most of the gender gap is due to differences in beliefs about enjoying coursework, and preferences.

**JEL Codes:** D8, I2, J1, Z1

**Keywords:** college majors; uncertainty; subjective expectations; preferences; gender differences; culture

---

\*Initial Draft: October 2007. I am indebted to Charles Manski, Christopher Taber, and Paola Sapienza for extremely helpful discussions and comments. I also thank Raquel Bernal, Marianne Hinds, Hilarie Lieb, Joan Linsenmeier, Carlos Madeira, Ofer Malamud, Ija Trapeznikova, Sergio Urzua and participants at Northwestern's Labor Lunches for feedback and suggestions. Financial Support from Northwestern University Graduate Research Grant, and Ronald Braeutigam is gratefully acknowledged. I thank all those involved in providing the datasets used in this paper. 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. All errors that remain are mine.

<sup>†</sup>Research and Statistics, Federal Reserve Bank of New York, 33 Liberty Street, New York, NY 10045. E-mail: Basit.Zafar@ny.frb.org

# 1 Introduction

The difference in choice of college majors between males and females is quite dramatic. In 1999-2000, amongst recipients of bachelor's degrees in the US, 13 percent of women majored in education compared to 4 percent of men, and only 2 percent of women majored in engineering compared to 12 percent of men (2001 Baccalaureate and Beyond Longitudinal Study). Figure 1 highlights the differences in gender composition of undergraduate majors of 1999-2000 bachelor's degree recipients (see also Polacheck, 1978; Turner and Bowen, 1999; Dey and Hill, 2007).

These markedly different choices in college major between males and females have significant economic and social impact. Figure 2 shows that large earnings premiums exist across majors. For example, in 2000-2001, a year after graduation in the US, the average education major employed full-time earned only 60 percent as much as one who majored in engineering (also see Eide and Grogger, 1995; Garman and Loury, 1995; Arcidiacono, 2004, for a discussion of earnings differences across majors). Paglin and Rufolo (1990), and Brown and Corcoran (1997) find that differences in major account for a substantial part of the gender gap in the earnings of individuals with several years of college education. Moreover, Xie and Shauman (2003) show that, controlling for major, the gap between men and women in their likelihood of pursuing graduate degrees and careers in science and engineering is smaller. The gender differences in choice of major have recently been at the center of hot debate on the reasons behind women's under-representation in science and engineering (Barres, 2006).

There are at least two plausible explanations for these differences. First, innately disparate abilities between males and females may predispose each group to choose different fields (Kimura, 1999, and 2006). However, studies of mathematically gifted individuals reveal differences in choices across gender, even for very talented individuals. For example, the Study of Mathematically Precocious Youth shows that mathematically talented women preferred careers in law, medicine, and biology over careers in physical sciences and engineering (Lubinski and Benbow, 1992). Moreover, the gender gap in mathematics achievement and aptitude is small and declining (Xie and Shauman, 2003; Goldin et al., 2006), and gender differences in mathematical achievement cannot explain the higher relative likelihood of majoring in sciences and engineering for males (Turner and Bowen, 1999; Xie and Shauman, 2003). These studies suggest gender differences in preferences as a second possible explanation for the gender gap in the choice of major. However, no systematic attempt has been made to study these preferences.

In this paper, I estimate a choice model of college major in order to understand how undergraduates choose college majors, and to explain the underlying gender differences. The choice of major is treated as a decision made under uncertainty—uncertainty about personal tastes, individual abilities, and realizations of outcomes related to choice of major. Such outcomes may include the associated economic returns and lifestyle as well as the successful completion of major. My choice model is closest in spirit to the theoretical model outlined in Altonji (1993), which treats education as a sequential choice made under uncertainty. In his dynamic model, the decision about attending college, field to major in, and dropping out are based on uncertain economic returns, personal tastes, and abilities. I, however, do not model the choice of college. The particular institutional setup in the Weinberg College of Arts & Sciences (WCAS) at Northwestern University allows me to estimate a choice model of college major where the decision can be treated as dynamic. However, since individuals are assumed

to maximize current expected utility, a static choice model is estimated in this paper.

The standard economic literature on decisions made under uncertainty generally assumes that individuals, after comparing the expected outcomes from various choices, choose the option that maximizes their expected utility. Given the choice data, the goal is to infer the decision rule. However, the expectations of the individual about the choice-specific outcomes are also unknown. The approach prevalent in the literature overlooks the fact that subjective expectations may be different from objective probabilities, assumes that formation of expectations is homogeneous, makes non-verifiable assumptions on expectations, and uses choice data to infer decision rules conditional on maintained assumptions on expectations. However, this can be problematic since observed choices might be consistent with several combinations of expectations and preferences, and the list of underlying assumptions may not be valid (see Manski, 1993a, for a discussion of this inference problem in the context of how youth infer returns to schooling). To illustrate this, let us assume that only two majors exist. Let us further assume that it is easier to get a college degree in the first major, but that it offers lower-paying jobs than the second major. An individual choosing the first major is consistent with two underlying states of the world: (1) she only cares about getting a college degree, or (2) she only values the job prospects but wrongly believes that the first major will get her a high-paying job. If one observes only the choice, then clearly one cannot discriminate between the two possibilities. The solution to this identification problem is to use additional data on expectations since it allows the researcher to separate the two possibilities, and that is precisely what I do.

I have designed and conducted a survey to elicit subjective expectations from 161 Northwestern University sophomores regarding choice of major. The survey collects data on demographics and background information, data relevant for the estimation of the choice model, and open-ended responses intended to explore how individuals form expectations.

In contrast to most studies on schooling choices which ignore uncertainty, I estimate a random utility model of college major choice allowing for heterogeneity in beliefs.<sup>1</sup> My approach also differs from the existing literature by accounting for the non-pecuniary aspects of the choice. Fiorito and Dauffenbach (1982) and Easterlin (1995) highlight the importance of non-price determinants in the choice of majors. However, no study has jointly modeled the pecuniary and non-pecuniary determinants of the choice. My approach allows me to quantify the contributions of both pecuniary and non-pecuniary outcomes to the choice. Moreover, the model is rich enough to explain gender differences in choices.

Responses to questions eliciting subjective expectations match up with existing statistics for several questions indicating that respondents answer meaningfully and seriously. Respondents exhibit significant heterogeneity in their responses (both between and within genders), which underscores the importance of expectations data to conduct inference in settings with uncertainty. However, I do not analyze the subjective data in the current study since that would abstract from the main goal of the paper. Interested readers are instead referred to Zafar (2008) which analyzes the subjective data in detail.

I estimate separate models for single major choice and for double major choice. The most important outcomes in the choice of single major are enjoying coursework, enjoying work at potential jobs, and approval of parents. Non-pecuniary outcomes explain about 45% of the choice behavior for males, and more than three-fourths of

---

<sup>1</sup>Literature on college majors has largely ignored the uncertainty associated with the various outcomes of the choice. Two notable empirical exceptions are Bamberger (1986), and Arcidiacono (2004). However, the former only takes into account the uncertainty about completing one's field of study. The latter estimates a dynamic model of college and major choice under highly stylized assumptions on expectations formation.

the choice for females. Males and females have similar preferences at college, but differ in their preferences regarding the workplace: males care more about the pecuniary outcomes in the workplace, females about the non-pecuniary outcomes. The results for the double major choice model are similar to those for single major. Graduating in 4 years, approval of parents, and enjoying coursework are the most important determinants of the choice. Additionally, I find evidence of individuals strategically choosing pairs of majors that allow them to specialize along certain dimensions. Females prefer pairs of majors which entail different chances of completion and getting a job upon graduation. On the other hand, males prefer major pairs that differ in their chances of completion, in the approval of parents, and in how much they would enjoy the coursework.

Besides being related to the literature on college major choice, this paper is related to three strands of literature. On the methodology side, it adds to the recent literature on subjective expectations (see Manski, 2004, for an overview of this literature). In the last decade or so, economists have increasingly undertaken the task of collecting and describing subjective data. Recently expectations data have been employed to estimate decision models. Van der Klaauw (2000) uses expectations data to improve the precision of the parameter estimates of a dynamic model of teacher career decisions. Delavande (2004) collects subjective data to estimate a choice model of birth control choice for women. The choice model used in this paper is motivated by her framework. My paper contributes to the two branches of this literature by providing an extensive description of students' expectations about major-specific outcomes, and by using subjective expectations data to estimate a choice model.

Second, this paper contributes to the recent literature on culture and economic outcomes (see Guiso et al., 2006; Alesina and Giuliano, 2007; Fernandez, 2007). In order to establish a causal link from culture to economic outcomes, I focus on the dimension of culture that is inherited by an individual from previous generations, rather than being voluntarily selected. I use information on the country of origin of the individual's parents as a cultural proxy. I find that cultural proxies bias preferences in favor of certain outcomes. Individuals with foreign-born parents value the pecuniary aspects of the choice more. In particular, males with foreign-born parents is the only sub-group in my sample for whom pecuniary outcomes explain more than 50% of the choice.

Finally, this paper is related to the literature that focuses on the underlying reasons for the gender gap in science and engineering. An interesting question is whether gender differences in choices are driven by differences in preferences or in beliefs. In the recent debate on the under-representation of women in science and engineering, some authors have claimed that the gap may be driven by the fact that women are less self-confident about their academic abilities than men. Valian (1998) argues that social prejudice against women causes them to lose self-confidence. Indeed, Solnick (1995) finds that women are more likely to shift to other majors from traditionally female majors if they attend a women's college. To check the validity of these hypotheses, I decompose the gender gap in major choice into differences in beliefs and differences in preferences. First, I find that gender differences in beliefs about ability constitute a small and insignificant part of the gap. This implies that explanations based entirely on the assumption that women have lower self-confidence relative to men (Long, 1986; Niederle et al., 2007) can be rejected in my data. Second, majority of the gender gap in majors that I consider can be explained by gender differences in beliefs about tastes for studying different fields, and preferences. For example, 60% of the gender gap in engineering is due to differences in preferences, while 30% is due to differences in how much females and males believe they will enjoy studying engineering.

Gender differences in beliefs about future earnings in engineering are insignificant and explain less than 1% of the gap. I simulate an environment in which the female subjective belief distribution about ability and future earnings is replaced with that of males; in the case of engineering, this reduces the gap by about only 14%. These results suggest that simply raising expectations for women in science, as claimed by Valian (1998), may not be enough, and that wage discrimination and social biases may not be the main reason for why women are less likely to major in science and engineering.

The paper is organized as follows: Section 2 outlines the choice model and the identification strategy. Section 3 describes the institutional setup of Weinberg College of Arts & Sciences, outlines the data collection methodology, and briefly describes the subjective data. Section 4 outlines the econometric framework used for estimation. Section 5 presents the estimation results for the single major choice model. Section 6 presents the results for the double major choice model. Section 7 undertakes a decomposition technique to understand the sources of gender differences in major choice. Finally, Section 8 concludes.

## 2 Choice Model

At time  $t$ , individual  $i$  is confronted with the decision to choose a college major from her choice set  $C_i$ . Individuals are forward-looking, and their choice depends not only on the current state of the world but also on what they expect will happen in the future. Individual  $i$  derives utility  $U_{ikt}(\mathbf{a}, \mathbf{c}, X_{it})$  from choosing major  $k$ . Utility is a function of a vector of outcomes  $\mathbf{a}$  which are realized in college, a vector of outcomes  $\mathbf{c}$  which are realized after graduating from college, and individual characteristics  $X_{it}$ . Examples of outcomes in  $\mathbf{a}$  include graduating within 4 years, enjoying the coursework, and approval of parents. Examples of outcomes in  $\mathbf{c}$  include future income, number of hours spent at the job, and ability to reconcile family and work. Both vectors,  $\mathbf{a}$  and  $\mathbf{c}$ , are uncertain at time  $t$ ; individual  $i$  possesses subjective beliefs  $P_{ikt}(\mathbf{a}, \mathbf{c})$  about the outcomes associated with choice of major  $k$  for all  $k \in C_i$ .<sup>2</sup> If an individual chooses major  $m$ , then standard revealed preference argument (assuming that indifference between alternatives occurs with zero probability) implies that:

$$m \equiv \arg \max_{k \in C_i} \int U_{ikt}(\mathbf{a}, \mathbf{c}, X_{it}) dP_{ikt}(\mathbf{a}, \mathbf{c}) \quad (1)$$

The goal is to infer the preference parameters from observed choices. However, the expectations of the individual about the choice-specific outcomes are also unknown. The most one can do is infer the decision rule conditional on the assumptions imposed on expectations. This would not be an issue if there were reason to think that prevailing expectations assumptions are correct. However, not only has the information processing rule varied considerably among studies of schooling behavior, most assume that individuals form their expectations in the same way.<sup>3</sup> First, there is little reason to think that individuals form their expectations in the same way. Second, different combinations of preferences and expectations may lead to the same choice. Manski (2002) shows that different combinations of preferences and expectations (about others' behavior) leads to same actions

<sup>2</sup>Though each major has an objective probability for  $(\mathbf{a}, \mathbf{c})$ , there's no reason to believe that subjective beliefs will be the same as the objective probabilities.

<sup>3</sup>Freeman (1971) assumed that income expectation formation of college students is myopic, that is, the youth believe that they will obtain the mean income realized by the members of a specified earlier cohort who made that choice. Arcidiacono (2004), in his dynamic model of college and major choice, makes strong assumptions about various outcomes; for example, he assumes that youth condition their expectations of future earnings on their ability, GPA, average ability of other students enrolled in that college, and some demographic variables. Similarly he assumes that all individuals have same expectations about the probability of working conditional on sex and major. The list of studies that explicitly (or implicitly) make assumptions about expectations formation is long, and there is no evidence that prevailing expectations assumptions are correct.

in the ultimatum game. To cope with the problem of joint inference on preferences and expectations, I elicit subjective probabilities directly from individuals. An additional advantage of this approach is that it allows me to account for the non-pecuniary determinants of the choice (data which does not exist otherwise).

The exact utility specification is outlined in section 4 which presents the econometric framework. I first describe the data collection methodology.

### 3 Data

I collect data on 161 Northwestern University sophomores. This section describes the institutional details at Northwestern, the data collection method, and the nature of the subjective data.

#### 3.1 Institutional Details

At time  $t$ , the individual uses available information to form subjective beliefs  $P_{ikt}(\mathbf{a}, \mathbf{c}) \forall k \in C_i$ . She then uses her subjective beliefs and preferences to choose a major that maximizes her expected subjective utility. Over time she might acquire more information about any of the major-specific outcomes. For example, she may learn about her unobserved match quality (ability and taste) in different fields by taking courses. She may also receive valuable information about the kinds of jobs and other major-related outcomes over time. This new information may prompt the individual to change her major. Northwestern University requires students to declare their major by the end of their sophomore year. Since sophomores are actively thinking about their choice of major, this study is restricted to Northwestern University sophomores.

This goal is to estimate the individual's preferences in their sophomore year. Moreover, the model allows an individual to experiment with majors until their sophomore year. Therefore, the study is restricted to schools at Northwestern University which accord students flexibility in choosing a major. For example, a student in the School of Engineering has to declare her major at the time of admission, and can only change her major by a special request to the school- she would not be eligible for the study. I further assume the choice set for an individual to be exogenous. This eliminates students in smaller schools at Northwestern since I will have to make strong assumptions about their choice set. Therefore, I restrict the study to the Weinberg College of Arts & Sciences (WCAS) at Northwestern. All sophomores with at least one major in the WCAS were eligible for the study.<sup>4</sup>

##### 3.1.1 Choice Set

WCAS offers a total of 41 majors. To estimate the choice model, one needs to elicit the subjective probabilities of the outcomes for each major in one's choice set. In order to limit the size of the choice set, I pool similar majors together. Table 1 shows the majors divided into various categories. Categories  $a$  through  $g$  span the majors offered in WCAS. Categories  $h$  through  $l$  span undergraduate majors offered by other schools at Northwestern University. There is a trade-off between the number of categories and the length of the survey. This categorization is fairly fine, and also seems reasonable.

For a student pursuing a single major in WCAS, it is assumed that her choice set includes all the categories that span WCAS majors ( $a-g$ ), and category  $k$ , the majors offered in the School of Engineering.<sup>5</sup> Therefore,

<sup>4</sup>A student could have a second major in any other school. She could take part in the study as long as she was pursuing a major in WCAS.

<sup>5</sup>This was done to elicit subjective beliefs of the outcomes associated with majoring in Engineering.

any student with a single major is assumed to have 8 categories in her choice set.

For an individual with a double major, the choice set is conditional on whether both her majors are in WCAS and the School of Engineering, or not. Conditional on the student's majors being in WCAS and the School of Engineering, the choice set is the same as that of a single major respondent except that the goal is now to select pairs of majors rather than a single one. Conditional on one of the majors being in a school other than WCAS or the School of Engineering, the choice set includes all major categories that span WCAS, category  $k$ , and the category which includes the student's non-WCAS major.<sup>6</sup>

### 3.2 Data Collection

A sample of eligible sophomores and their E-mail addresses was provided by the Northwestern Office of the Registrar. Students were recruited by E-mail, and flyers were posted on campus in schools other than WCAS.<sup>7</sup> The E-mails and flyers explicitly asked for sophomores with an intended major in WCAS. Prospective participants were told that the survey was about the choice of college majors, and that they would get \$10 for completing the 45-minute electronic survey. It was emphasized that one need not have declared their major to participate in the study. The survey was conducted from November 2006 to February 2007. Respondents were required to come to the Kellogg Experimental Laboratory to take the electronic survey.

A total of 161 WCAS sophomores were surveyed, of whom 92 were females. Table 2 shows how the characteristics of the sample compare with those of the sophomore class. The sample looks similar to the population in most aspects. However, a few differences stand out: (1) students of Asian ethnicity are over-represented in my sample; (2) 61% of the respondents had declared their major at the time of the survey, whereas the corresponding number for the sophomore population was only 18%. However, this statistic for the population was obtained at the beginning of the sophomore year. Since students may declare their major at any time during the academic year, it is very likely that this statistic was greater than 18% for the population at the time of the survey; (3) it seems that survey-takers, especially male students, have higher GPAs than their population counterparts.

Table 3 presents the distribution of WCAS majors in the sample. For comparison, the major distribution for the graduating class of 2006 is also presented.<sup>8</sup> There are a few notable features. The proportion of males who (intend to) major in Social Sciences II is twice the corresponding proportion of women in both my sample as well as the graduating class of 2006. This pattern is reversed in the case of Social Sciences I, and Literature and Fine Arts. The proportion of females who (intend to) major in Literature and Fine Arts is more than 3 times the corresponding proportion of males.

The 45-minute survey consisted of three parts. The first part collected demographic and background information (including parents' and siblings' occupations and college majors, source of college funding etc.). The second part collected data relevant for the estimation of the choice model, and is discussed in more detail in the next subsection. The third part collected responses to open-ended questions intended to explore how respondents form expectations about various major-specific outcomes, and the sources of information they used. At the end of the survey, respondents were asked if they were willing to participate in a follow-up survey in a

---

<sup>6</sup>For example, the choice set for a student with a major in WCAS and the School of Education would be categories  $a-g$ ,  $i$ , and  $k$ .

<sup>7</sup>E-mails advertising the survey were also sent out by WCAS undergraduate advisors, economics professors teaching large core classes, and Deans of some schools (other than WCAS).

<sup>8</sup>This is the most recent year for which data is available. There is no reason to believe that the distribution of majors would stay the same over time.

year's time. If the respondent agreed to the follow-up, they were asked for their name and contact information. An astounding 97% (156 out of 161) respondents agreed to the follow-up.

### 3.3 Subjective Data

The subjective beliefs,  $P_{ikt}(\mathbf{a}, \mathbf{c}) \forall k \in C_i$ , are elicited directly from the respondent. The vector  $\mathbf{a}$  includes the outcomes:

- $a_1$  successfully completing (graduating) a field of study in 4 years
- $a_2$  graduating with a GPA of at least 3.5 in the field of study
- $a_3$  enjoying the coursework
- $a_4$  hours/week spent on the coursework
- $a_5$  parents approve of the major

while the vector  $\mathbf{c}$  consists of:

- $c_1$  get an acceptable job immediately upon graduation
- $c_2$  enjoy working at the jobs available after graduation
- $c_3$  able to reconcile work and family at the available jobs
- $c_4$  hours/week spent working at the available jobs
- $c_5$  social status of the available jobs
- $c_6$  income at the available jobs

An individual's choice of major might be motivated by several pecuniary and non-pecuniary concerns. An individual motivated primarily by future earnings prospects may choose a major that is associated with large income streams ( $c_6$ ), allows a high probability of getting a job upon graduation ( $c_1$ ), and increases the possibility of getting jobs with high social status ( $c_5$ ). An individual concerned about her ability may choose a major that presents a greater probability of completion ( $a_1$ ), and allows her to graduate with a higher GPA ( $a_2$ ). On the other hand, an individual may choose a major with low-salary job prospects which allow a flexible lifestyle ( $c_3$ ,  $c_4$ ), or provide opportunities to do things she enjoys ( $c_2$ ). Similarly an individual's choice may be influenced by the kinds of courses she finds interesting ( $a_3$ ), or by how demanding the major is ( $a_4$ ). Finally, the choice may be influenced by parents and family ( $a_5$ ). Another interpretation of these outcomes is as follows:  $a_1$  and  $a_2$  are outcomes that capture ability in college;  $a_3$  can be interpreted as taste in college;  $c_2$  and  $c_3$  may be interpreted as tastes in the workplace.

Note that  $\{a_r\}_{r=\{1,2,3,5\}}$  and  $\{c_q\}_{q=\{1,2,3\}}$  are binary, while outcomes  $a_4$  and  $\{c_q\}_{q=\{4,5,6\}}$  are continuous. For all  $k \in C_i$ , the survey elicited the probability of the occurrence of the binary outcomes, i.e.  $P_{ikt}(a_r = 1)$  for  $r = \{1, 2, 3, 5\}$  and  $P_{ikt}(c_q = 1)$  for  $q = \{1, 2, 3\}$ . Expected value was elicited for the continuous outcomes, i.e.  $E_{ikt}(a_4)$  and  $E_{ikt}(c_q)$  for  $q = \{4, 6\}$ .

Questions eliciting the subjective probabilities of major-specific outcomes are based on the use of percentages. As is standard in studies that collect subjective data, a short introduction was read and handed to the respondents at the start of the survey:

*"In some of the survey questions, you will be asked about the PERCENT CHANCE of something happening. The percent chance must be a number between zero and 100. Numbers like 2 or 5% indicate*

*“almost no chance,” 19% or so may mean “not much chance,” a 47 or 55% chance may be a “pretty even chance,” 82% or so indicates a “very good chance,” and a 95 or 98% mean “almost certain.” The percent chance can also be thought of as the NUMBER OF CHANCES OUT OF 100.*

*We will start with a couple of practice questions.”*

This introduction is similar to the one in the Survey of Economic Expectations (SEE) which is described in Dominitz and Manski (1997). However, as in Delavande (2004), I do not round off the percentages. For example, I use 19% instead of 20% to encourage respondents to use the full range from zero to 100. Respondents had to answer two practice questions before starting the survey to make sure they understood how to answer questions based on the use of percentages.

The questions dealing with subjective expectations were worded as follows:

*If you were majoring in [X], what do you think is the percent chance that you will graduate with a GPA of at least 3.5 (on a scale of 4)?*

and:

*Look ahead to when you will be 30 YEARS OLD. If you majored in [X], what do you think is the percent chance that you will be able to reconcile work and your social life/ family at the kinds of jobs that will be available to you?*

The question eliciting the expected number of hours/week spent on coursework was:

*If you were majoring in [X], how many hours per week do you think you will need to spend on the coursework?*

Social status of the available jobs was elicited as follows:

*Look ahead to when you will be 30 years old. Rank the following fields of study according to your perception of the social status of the jobs that would be available to you and that you would accept if you graduated from that field of study.<sup>9</sup>*

For the expected income, the question was as follows:<sup>10</sup>

*Look ahead to when you will be 30 years old. Think about the kinds of jobs that will be available to you and that you will accept if you graduate in [X]. What is the average amount of money that you think you will earn per year by the time you are 30 YEARS OLD?*

In addition, I elicited the subjective belief of being active in the full-time labor force at the age of 30 and 40, and  $E(Y_0)$ , the expected income of dropping out from school at the age of 30.

Questions eliciting the beliefs about major-specific outcomes can be viewed in section 9.1 of the appendix. The full questionnaire is available on request from the author.

---

<sup>9</sup>This question elicits an ordinal ranking of the social status of the jobs. However, I treat these ordinal responses as cardinal in the choice model analysis. In hindsight, this question should have been asked in terms of subjective expectations of getting a high status job.

<sup>10</sup>The wording of this question is very similar to that of Dominitz and Manski (1996) who elicit student expectations of the returns to schooling from high school and college students.

### 3.4 The Data

Since the use of subjective data in economics is fairly recent, there is a genuine interest in analyzing the precision and accuracy of such data. However, given that the main goal of this paper is to estimate the preferences for college major choice by taking the subjective data as given, I don't undertake this task here since it is not possible to summarize the data in a condensed form. Moreover, to address the extent to which cognitive biases (such as respondents making little mental effort in answering questions; cognitive dissonance; social desirability) affect beliefs, one would need to observe how beliefs evolve over time. Interested readers are referred to Zafar (2008) which analyzes this set of issues by using a panel of beliefs elicited from the same sample. The companion paper shows that respondents provide meaningful answers to questions eliciting subjective expectations. In cases where responses could be compared to objective realities and statistics, survey responses match up well. For example, individuals are aware of the earnings differences across majors. The analysis in that paper does not find evidence of cognitive biases systematically affecting the way in which individuals report their beliefs.

It should be pointed out that there is substantial heterogeneity in responses both between and within gender. In order to highlight the heterogeneity in beliefs across respondents, I discuss the response to a representative question. Table 4 presents the gender-specific subjective belief distribution of graduating with a GPA of at least 3.5 in Engineering and Literature and Fine Arts. The table shows that respondents are willing to use the entire scale from zero to 100. It does seem that respondents tend to round off their responses to the nearest 5, especially for answers not at the extremes. There has been some concern that respondents might answer 50% when they want to respond to the interviewer but are unable to make any reasonable probability assessment of the relevant question.<sup>11</sup> However, the 50% response is not the most frequent one in the majority of the cases. There doesn't seem to be any evidence of anchoring since numbers that were presented in the introductory text do not occur more often than others. The substantial heterogeneity in beliefs questions the accuracy of restrictions imposed on expectations in the literature.

## 4 Econometric Model

This section outlines the econometric framework.

Recall that utility,  $U_{ikt}(\mathbf{a}, \mathbf{c}, X_{it})$ , is a function of a  $5 \times 1$  vector of outcomes  $\mathbf{a}$  realized in college, a  $6 \times 1$  vector of outcomes  $\mathbf{c}$  realized after graduating from college, and individual characteristics  $X_{it}$ . The individual maximizes her *current* subjective expected utility<sup>12</sup>; she chooses major  $m$  at time  $t$  if:

$$m = \arg \max_{k \in C_i} \int U_{ikt}(\mathbf{a}, \mathbf{c}, X_{it}) dP_{ikt}(\mathbf{a}, \mathbf{c}) \quad (2)$$

Moreover, as explained in section 3.3, the outcomes  $\{a_r\}_{r=\{1,2,3,5\}}$  and  $\{c_q\}_{q=\{1,2,3\}}$  are binary, while outcomes  $a_4$ , and  $\{c_q\}_{q=\{4,5,6\}}$  are continuous. I change the notation slightly, and define  $\mathbf{b}$  to be a  $7 \times 1$  vector of all binary outcomes, i.e.  $\mathbf{b} = \{a_1, a_2, a_3, a_5, c_1, c_2, c_3\}$ , and  $\mathbf{d}$  to be a  $4 \times 1$  vector of all continuous outcomes, i.e.  $\mathbf{d} = \{a_4, c_4, c_5, c_6\}$ . The utility can now be written as a function of outcomes  $\mathbf{b}$ ,  $\mathbf{d}$ , and characteristics  $X_{it}$ . I

<sup>11</sup>See Bruine de Bruin et. al. (2000). This is what they call "epistemic uncertainty", or the "50-50 chance".

<sup>12</sup>Under the assumption that individuals maximize current expected utility, I don't need to take into account that individuals may find it optimal to experiment with different majors. However, experimentation could be important in this context to learn about one's ability and match quality (see Manski, 1989, and Malamud, 2006). It is beyond the scope of this paper.

assume that utility is additively separable in the outcomes:

$$U_{it}(\mathbf{b}, \mathbf{d}, X_{it}) = \sum_{r=1}^7 u_r(b_r, X_{it}) + \sum_{q=1}^4 \gamma_{iq} d_q + \varepsilon_{ikt} \quad (3)$$

where  $u_r(b_r, X_{it})$  is the utility associated with the binary outcome  $b_r$  for an individual with characteristics  $X_{it}$ ,  $\gamma_{iq}$  is a constant for the continuous outcome  $d_q$ , and  $\varepsilon_{ikt}$  is a random term. The utility is same for all individuals with identical observable characteristics  $X_{it}$  up to the random term. Equation (2) can now be written as:

$$m \equiv \arg \max_{k \in C_i} \left( \sum_{r=1}^7 \int u_r(b_r, X_{it}) dP_{ikt}(b_r) + \sum_{q=1}^4 \gamma_{iq} \int d_q dP_{ikt}(d_q) + \varepsilon_{ikt} \right) \quad (4)$$

An individual  $i$  with subjective beliefs  $\{P_{ikt}(b_r), P_{ikt}(d_q)\}$  for  $r \in \{1, \dots, 7\}$ ,  $q \in \{1, \dots, 4\}$  and  $\forall k \in C_i$  chooses major  $m$  at time  $t$  with probability:

$$\Pr(m|X_{it}, \{P_{ikt}(b_r), P_{ikt}(d_q)\}_{r \in \{1, \dots, 7\}, q \in \{1, \dots, 4\}; k \in C_i}) = \Pr \left( \begin{array}{l} \sum_{r=1}^7 \int u_r(b_r, X_{it}) dP_{imt}(b_r) + \sum_{q=1}^4 \gamma_{iq} \int d_q dP_{imt}(d_q) + \varepsilon_{imt} \\ > \sum_{r=1}^7 \int u_r(b_r, X_{it}) dP_{ikt}(b_r) + \sum_{q=1}^4 \gamma_{iq} \int d_q dP_{ikt}(d_q) + \varepsilon_{ikt} \end{array} \right) \quad (5)$$

$\forall k \in C_i, m \neq k$

For the binary outcomes in  $\mathbf{b}$ ,  $P_{imt}(b_r)$  is simply  $P_{imt}(b_r = 1)$  for  $r \in \{1, \dots, 7\}$ ;  $P_{imt}(b_r = 1)$  is elicited directly from the respondents for  $\forall r \in \{1, \dots, 7\}$  and  $\forall k \in C_i$ . For the continuous outcomes in  $\mathbf{d}$ , instead of the probability distribution, the expected value of the outcome  $E_{ikt}(d_q) = \int d_q dP_{ikt}(d_q)$  is elicited  $\forall q \in \{1, \dots, 4\}$ .<sup>13</sup>

Next, I explain how I compute the expected income. Since one must successfully complete the major to gain the associated earnings,  $E_{ikt}(d_4)$ ,  $i$ 's expected earnings associated with choice  $k$  at time  $t$  are:

$$E_{ikt}(d_4) = \int w dG_{it}(w) [p_{ikt} E_{ikt}(I) + (1 - p_{ikt}) E_{it}(I_0)] \quad \text{for } k, p \in C_i \text{ and } p \neq k \quad (6)$$

where  $w$  is an indicator variable of the individual's labor force status,  $G_{it}(w)$  is the subjective belief at time  $t$  about one's labor force status at the age of 30, and  $p_{ikt}$  is individual  $i$ 's subjective probability at time  $t$  about successfully graduating in major  $k$ . The belief of being active in the labor force at the age of 30 is simply  $G_{it}(w = 1)$ .  $\int w dG_{it}(w) = G_{it}(w = 1)$ , denoted as  $g_{it}$ , is elicited directly from the respondents.<sup>14</sup> Conditional on being active in the labor force, with probability  $p_{ikt}$ , the individual's expected earnings are  $E_{ikt}(I)$ , the expected income associated with major  $k$  at the age of 30; with probability  $1 - p_{ikt}$ , her expected earnings are  $E_{it}(I_0)$ , the expected income at the age of 30 if one were to drop out of school at time  $t$ .<sup>15</sup> Equation (5) can

<sup>13</sup>A consequence of the linear utility specification is that the individual is risk-neutral, i.e.  $\int U_{it}(Y, \mathbf{b}, \mathbf{d}, X_{it}) dP_{ikt}(Y, \mathbf{b}, \mathbf{d}) = U_{it}(\int Y, \mathbf{b}, \mathbf{d}, X_{it}) dP_{ikt}(Y, \mathbf{b}, \mathbf{d})$ . Hence, I only need to elicit the expected value for the continuous outcomes.

<sup>14</sup>Note that the underlying assumption is that expectation of being active in the labor force,  $g_{it}$ , is independent of one's field of study. This is a rather restrictive assumption since one's decision of participating in the labor force may be influenced by the job opportunities available, which would be related to one's field of study. Relaxing this assumption would have required me to ask this subjective expectation for each field of study in one's choice set, and would not have been feasible.

<sup>15</sup>In an earlier version of the model, I allow the individual to change fields of study once before dropping out of school. However, the results don't seem to change much.

now be written as:

$$\Pr(m|X_{it}, \{P_{ikt}(b_r), E_{ikt}(d_q)\}_{r \in \{1, \dots, 7\}, q \in \{1, \dots, 4\}; k \in C_i}) =$$

$$\Pr \left( \begin{array}{l} \sum_{r=1}^7 \{P_{imt}(b_r = 1)u_r(b_r = 1, X_{it}) + [1 - P_{imt}(b_r = 1)]u_r(b_r = 0, X_{it})\} + \sum_{q=1}^4 \gamma_{igt} E_{imt}(d_q) + \varepsilon_{imt} \\ > \sum_{r=1}^7 \{P_{ikt}(b_r = 1)u_r(b_r = 1, X_{it}) + [1 - P_{ikt}(b_r = 1)]u_r(b_r = 0, X_{it})\} + \sum_{q=1}^4 \gamma_{ikt} E_{ikt}(d_q) + \varepsilon_{ikt} \end{array} \right) \quad (7)$$

$$\forall k \in C_i, m \neq k$$

Moreover,  $P_{imt}(b_r = 1)u_r(b_r = 1, X_{it}) + [1 - P_{imt}(b_r = 1)]u_r(b_r = 0, X_{it})$  is equivalent to  $P_{imt}(b_r = 1)\Delta u_r(X_{it}) + u_r(b_r = 0, X_{it})$ , where  $\Delta u_r(X_{it}) \equiv u_r(b_r = 1, X_{it}) - u_r(b_r = 0, X_{it})$ , i.e. it is the difference in utility between outcome  $b_r$  happening and not happening for an individual with characteristics  $X_{it}$ . The expected utility that individual  $i$  derives from choosing major  $m$  at time  $t$  is:

$$U_{imt}(\mathbf{b}, \mathbf{d}, X_{it}, \{P_{imt}(b_r = 1)\}_{r=1}^7, \{E_{imt}(d_q)\}_{q=1}^4) =$$

$$= \sum_{r=1}^7 P_{imt}(b_r = 1)\Delta u_r(X_{it}) + \sum_{r=1}^7 u_r(b_r = 0, X_{it}) + \sum_{q=1}^4 \gamma_{igt} E_{imt}(d_q) + \varepsilon_{imt} \quad (8)$$

Equation (7) can now be written as:

$$\Pr(m|X_{it}, \{P_{ikt}(b_r), E_{ikt}(d_q)\}_{r \in \{1, \dots, 7\}, q \in \{1, \dots, 4\}; k \in C_i}) =$$

$$\Pr \left( \begin{array}{l} \sum_{r=1}^7 P_{imt}(b_r = 1)\Delta u_r(X_{it}) + \sum_{q=1}^4 \gamma_{igt} E_{imt}(d_q) + \varepsilon_{imt} \\ > \sum_{r=1}^7 P_{ikt}(b_r = 1)\Delta u_r(X_{it}) + \sum_{q=1}^4 \gamma_{ikt} E_{ikt}(d_q) + \varepsilon_{ikt} \end{array} \right) \quad (9)$$

$$\forall k \in C_i, m \neq k$$

$\{\Delta u_r(X_{it})\}_{r=1}^7$ , and  $\{\gamma_{igt}\}_{q=1}^4$  are the parameters to be estimated.  $g_{it}$ ,  $\{P_{ikt}(b_r = 1)\}_{r=1}^7$ , and  $\{E_{ikt}(d_q)\}_{q=1}^4$ , and  $E_{ikt}(I) \forall k \in C_i$  are elicited directly from the respondent. In order to ensure strict preferences between choices,  $\{\varepsilon_{ikt}\}$  are assumed to have a continuous distribution. The exact parametric restrictions on the random terms required for identifying the model parameters are discussed in the next section.

## 5 Single Major Choice Model

This section deals with estimating the preferences for choice of single majors. I drop the time subscript in the analysis that follows.

### 5.1 Estimation with Homogenous Preferences

The model described in section 4 assumes that the utility function for the binary outcomes  $u_r(b_r, X_i)$ , and the constants on continuous outcomes ( $\{\gamma_{iq}\}_{q=1}^4$ ) depend on individual characteristics. I initially assume that the

utility function does not depend on individual characteristics. Under this assumption, (9) becomes:

$$\begin{aligned} & \Pr(m|P_{ik}(b_r), E_{ik}(d_q)\}_{r \in \{1, \dots, 7\}, q \in \{1, \dots, 4\}; k \in C_i) \\ &= \Pr \left( \begin{array}{l} \sum_{r=1}^7 P_{im}(b_r = 1) \Delta u_c + \sum_{q=1}^4 \gamma_q E_{im}(d_q) + \varepsilon_{imt} \\ > \sum_{r=1}^7 P_{ik}(b_r = 1) \Delta u_c + \sum_{q=1}^4 \gamma_q E_{ik}(d_q) + \varepsilon_{ikt} \end{array} \right) \end{aligned} \quad (10)$$

$$\forall k \in C_i, m \neq k$$

If I assume the random terms  $\{\varepsilon_{ikt}\}$  are independent for every individual  $i$  and choice  $k$ , and that they have a Type I extreme value distribution, then  $\{\varepsilon_{ikt} - \varepsilon_{imt}\}$  has a standard logistic distribution. Then the probability that individual  $i$  chooses major  $m$  is:

$$\begin{aligned} & \Pr(m|\{P_{ik}(b_r), E_{ik}(d_q)\}_{r \in \{1, \dots, 7\}, q \in \{1, \dots, 4\}; k \in C_i}) \\ &= \frac{\exp(\sum_{r=1}^7 P_{im}(b_r = 1) \Delta u_r + \sum_{q=1}^4 \gamma_q E_{im}(d_q))}{\sum_{k \in C_i} \exp(\sum_{r=1}^7 P_{ik}(b_r = 1) \Delta u_r + \sum_{q=1}^4 \gamma_q E_{ik}(d_q))} \end{aligned} \quad (11)$$

Under these parametric assumptions, the parameters  $\{\Delta u_r\}_{r=1}^7$ , and  $\{\gamma_q\}_{q=1}^4$  are identified. The elicited subjective probabilities described in section 3.2 are used in estimation. Column (1) of Table 5 presents the maximum likelihood estimates using stated choice data.<sup>16,17</sup>

The relative magnitudes of  $\{\Delta u_r\}_{r=1}^7$  show the importance of the binary outcomes in the choice. The difference in utility levels is positive and largest for enjoying coursework. The second most important outcome in the choice is graduating within 4 years; it has a positive coefficient that is about half of the coefficient on enjoying coursework. The third most important factor is enjoying work at the available jobs with a positive coefficient of a similar magnitude as the coefficient on graduating within 4 years. The difference in utility levels is positive for parent's approval, and (surprisingly) negative for graduating with a GPA of at least 3.5. Both coefficients are significant, and about one-fourth the coefficient on enjoying coursework. The difference in utility levels for reconciling family and work is about one-sixth in magnitude compared to that of enjoying coursework, but is surprisingly negative. The coefficient on the social status of the jobs is positive and significant. A unit increase in the social status of available jobs changes the utility by as much as a 5% increase in the probability of graduating in 4 years. The coefficient on hours/week spent at work is negative, but not significantly different from zero. Though the coefficient on income is negative, it is not significantly different from zero suggesting that it is not important in the choice.

Column (2) of Table 5 shows the maximum-likelihood estimates based on (11) with the addition of female interactions in order to get some measure of relative differences between males and females. For males, the difference in utility levels is largest for enjoying coursework, finding a job upon graduation, and the social status of the jobs in decreasing order of importance. For females, the three outcomes that matter the most are graduating in 4 years, enjoying the coursework, and enjoying work at the available jobs. Though income stays

<sup>16</sup>44 of the 83 respondents with a single major had declared their major at the time of the survey. For the remaining 39, I use their stated intended choice for estimation.

<sup>17</sup>Moreover, a respondent with an adjunct major (see Table 1) has to have another major. For the purposes of estimation, I don't differentiate between an adjunct major and a normal major. Such respondents are treated as pursuing a single major if both their majors are in the same category, and as pursuing double majors if they are in different categories.

insignificant, the coefficient on income interacted with the female dummy shows that the negative coefficient on income in Column (1) is being driven by the preferences of females; income has a positive coefficient for males now, and negative for females (though neither are significant).

In addition to stating their choice, respondents were also asked to rank the elements in their choice set. The stated preference data provides more information which can be used for estimation of the model parameters. Under the assumptions of standard logit, the probability of any ranking of alternatives can be written as a product of logits. For example, consider the case where an individual's choice set is  $\{a, b, c, d\}$ . Suppose she ranks the alternatives  $b, d, c, a$  from best to worst. Under the assumption that the  $\varepsilon_{ik}$ 's are iid and Type I distributed, the probability of observing this preference ordering can be written as the product of the probability of choosing alternative  $b$  from  $\{a, b, c, d\}$ , the probability of choosing  $d$  from  $\{a, c, d\}$ , and the probability of choosing  $c$  from the remaining  $\{a, c\}$ . If  $U_{ij} = \beta x_{ij} + \varepsilon_{ij}$  denotes the utility  $i$  gets from choosing  $j$  for  $j \in \{a, b, c, d\}$ , then the probability of observing  $b \succ d \succ c \succ a$  is simply:<sup>18</sup>

$$\Pr(b \succ d \succ c \succ a) = \frac{\exp(\beta x_{ib})}{\sum_{j \in \{a, b, c, d\}} \exp(\beta x_{ij})} \cdot \frac{\exp(\beta x_{id})}{\sum_{j \in \{a, c, d\}} \exp(\beta x_{ij})} \frac{\exp(\beta x_{ic})}{\sum_{j \in \{a, c\}} \exp(\beta x_{ij})} \quad (12)$$

Column (3) in Table 5 presents the maximum likelihood estimates using stated preference data. The difference in utility levels is still largest and positive for enjoying coursework. Graduating in 4 years, the second most important outcome using stated choice data, is now negative but not significant. Enjoying work at the jobs is the second most important outcome with a positive coefficient. Approval of parents, now the third most important outcome, has a positive coefficient that is one-half that of enjoying coursework. The difference in utility levels for graduating with a GPA of at least 3.5 is now positive and significant. Status of the jobs continues to be important: a unit increase in the social status of the jobs changes the utility by as much as a 4% increase in the probability of enjoying coursework. The difference in utility levels for other binary outcomes is not significantly different from zero. The coefficient on income is now positive, but not significant.

Column (4) allows female interaction dummies to gain further insight into gender differences in preferences. For both genders, the difference in utility levels is largest and positive for enjoying coursework. For males, graduating within 4 years is the second most important outcome, but surprisingly it has a negative sign. The third most important outcome for males is the difference in utility levels for graduating with a GPA of at least 3.5; it is positive and about half that of enjoying coursework. Status of the jobs remains important for males: a unit increase in the status of the jobs changes the utility by as much as a 10% increase in the probability of enjoying coursework. For females, two of the important outcomes are approval of parents, and enjoying work at the jobs. Both have a positive coefficient that is about two-thirds the magnitude of the coefficient on enjoying coursework. Graduating within 4 years, and graduating with a GPA of at least 3.5 have coefficients that are positive and about one-third of the coefficient on enjoying coursework.

One concern with using stated preference data is that an individual may not have complete preferences over all alternatives that are available to her. In the case that a complete ranking does not *exist*, it is possible that the lower end of her preferences is noise. To check the sensitivity of the results, the model was also estimated by using the ranking of the four most preferred choices only. The results (available upon request from the

---

<sup>18</sup>A logit on ranked data is called *exploded* logit in the literature. This is because a ranking of  $J$  alternatives explodes into  $J - 1$  pseudo-observations for estimation purposes. This expression results from the particular form of the extreme value distribution, first shown by Luce and Suppes (1965).

author) are comparable to those obtained from using the complete preference data. Therefore, I continue to use complete stated preference data in the analysis that follows.

In order to get a measure of the magnitude of the estimated parameters, the natural thing would be to do willingness to pay calculations, i.e. translate the differences in utility levels into the amount that an individual would be willing to forgo at the age of 30 in earnings in order to experience that outcome.<sup>19</sup> However, since expected income at age 30 is not significant in any of the specifications considered, the standard errors on such calculations are huge, and the results are not very meaningful. Instead of presenting the willingness to pay calculations, I outline a different decomposition method to gain insight into the relative importance of the various outcomes in the choice. For illustration, suppose that  $\Pr(\text{choice} = j) = F(\mathbf{X}_j\boldsymbol{\beta})$ , and that  $\mathbf{X}$  includes two variables,  $X_1$  and  $X_2$ . Given the parameter estimates,  $\widehat{\beta}_1$  and  $\widehat{\beta}_2$ , the contribution of  $X_1$  to the choice is defined as:

$$M_{X_1} \equiv \left\| \overline{\Pr(\text{choice} = j | \{\widehat{\beta}_1, \widehat{\beta}_2\})} - \overline{\Pr(\text{choice} = j | \{\widehat{\beta}_1 = 0, \widehat{\beta}_2\})} \right\| \quad (13)$$

$$= \sqrt{\sum_{j=1}^8 \left[ \sum_{i=1}^N \frac{\Pr(\text{choice} = j | \{\widehat{\beta}_1, \widehat{\beta}_2\})}{N} - \sum_{i=1}^N \frac{\Pr(\text{choice} = j | \{\widehat{\beta}_1 = 0, \widehat{\beta}_2\})}{N} \right]^2}$$

where the first term is the average probability of majoring in choice  $j$  predicted by the model, and the second term is the average predicted probability of majoring in  $j$  if outcome  $X_1$  were not considered. The difference of the two terms is a measure of the importance of  $X_1$  in the choice. Similarly the contribution of  $X_2$  is given as:

$$M_{X_2} \equiv \sqrt{\sum_{j=1}^8 \left[ \sum_{i=1}^N \frac{\Pr(\text{choice} = j | \{\widehat{\beta}_1, \widehat{\beta}_2\})}{N} - \sum_{i=1}^N \frac{\Pr(\text{choice} = j | \{\widehat{\beta}_1, \widehat{\beta}_2 = 0\})}{N} \right]^2} \quad (14)$$

The *relative* contribution of  $X_1$  to the choice is then  $R_{X_1} = \frac{M_{X_1}}{M_{X_1} + M_{X_2}}$ . Multiple parameters can be set to zero simultaneously to get a sense of their joint contribution to the choice. However, since the model is not linear, generally  $M_{X_1+X_2} \neq M_{X_1} + M_{X_2}$ . Table 6 presents the results of this decomposition strategy. Each cell shows the *relative* contribution ( $R$ ) of the outcome to the choice. Panel B of Table 6 presents the results of this decomposition technique using the estimates obtained from stated preference data. Column (1) shows the decomposition results of the estimates of the pooled sample: nearly three-fourths of the choice is driven by the non-pecuniary outcomes.<sup>20</sup> If the decomposition is made finer, one can see that parent's approval and enjoying coursework jointly explain about 45% of the choice. Pecuniary outcomes associated with college (hours/week spent on coursework, graduating with a GPA of at least 3.5, and graduating in 4 years), and workplace (finding a job upon graduation, hours/week spent at work, income at the age of 30, and the social status of the jobs) each account for about 20% of the choice.

The estimates of the pooled sample mask the differences between males and females. Columns (2) and (3) of Table 6 show the decomposition results using the estimates from the male sub-sample, and the female sub-sample respectively. Non-pecuniary outcomes explain about 45% of the choices for males, but more than

<sup>19</sup>For example, the amount that an individual would be willing to forgo in earnings at the age of 30 for a 2% change in the probability of outcome  $j$  is  $\frac{0.02 \times \Delta u_j}{\gamma_4}$ .

<sup>20</sup>Outcomes classified as being non-pecuniary are: parent's approval, enjoying coursework, reconciling work and family, and enjoying work at the jobs. The remaining outcomes are termed as being pecuniary.

80% of the choice for females. Parent’s approval and enjoying coursework are the most important outcomes for females explaining about 45% of their choice, while pecuniary outcomes associated with the workplace are of utmost importance to males explaining 48% of their choice. Reconciling family and enjoying work at the available jobs are second in terms of importance to females, but of least importance to males. On the whole, non-pecuniary determinants are crucial in explaining the choices for both males and females. However, males and females differ in their preferences in the workplace: males value pecuniary aspects of the workplace more, while females value non-pecuniary aspects of the workplace more.

Table 7 presents the results of various thought experiments in an attempt to assess how changes in beliefs affect the choice of majors for males and females. The baseline case is first presented. For example, the model predicts that the average probability of majoring in engineering for males is 11.7%, more than twice that for females. Experiments 1 through 3 show changes in predicted probabilities in response to changes in beliefs of outcomes that are well-defined (for example, graduating with a GPA of at least 3.5). Predicted probabilities are not very responsive to changes in beliefs in these cases. Experiments 4 through 6 shows results of *thought* experiments for outcomes that are not well-defined. For example, experiment 5 shows that the average probability of majoring in engineering increases by 20% for females, and by about 10% for males in response to a 10% increase in beliefs of enjoying coursework in engineering. The results in Table 7 indicate that outcomes like enjoying coursework and approval of parents are crucial in one’s choice of major.

## 5.2 Comparison with Stated Preferences

It would be interesting to see how the estimated preference parameters compare with stated preferences. Survey respondents were asked to state their preferences for various determinants in their choice. More specifically, respondents were asked to assign an integer between 0 and 100 to a list of reasons such that the numbers added up to a 100.<sup>21</sup> Table 8 shows the average weights assigned to the various reasons by males and females. I interpret these numbers as the relative preference for the given reason in the choice of major. Enjoying working at the jobs and learning more about things that interest me were the two most important reasons for choosing a major for both males and females. However, females, on average assign higher weights to this reason (the gender difference is significant). For males, the third most important stated reason for choosing a major is getting a high-paying job. Conversely, doing well in the coursework is the third most important reason for females. These stated preferences for various outcomes are consistent with the parameter estimates discussed above. Determinants like fraction of people of the same gender taking classes in the major or working at the jobs don’t seem to be important. One surprising finding is that males, on average, are more likely than females to have been encouraged by a mentor/ role model to choose a major (similarly females assign lower weights to peer pressure, siblings making the same choice, and parents wanting them to make the choice; the gender difference is not significant for any of these reasons though).<sup>22</sup> The result in this study could be driven by the fact that the sample is restricted to students who have at least one major in the College of Arts and Sciences (which contains majors mostly dominated by females) and, therefore, males who have a major in the College are more likely to have been encouraged by a role model than females.

<sup>21</sup>This question was asked in the follow-up survey (Zafar, 2008) which was taken by 117 of the 161 original survey-takers.

<sup>22</sup>This finding is in contrast to the literature in social psychology which finds that females are more influenceable (Eagly, 1978) and that females, in contrast to males who have mainly independent self-schemas, have interdependent ones (Cross and Madson, 1997). In light of this evidence, economic studies that empirically investigate the role model hypothesis usually focus on females only (an exception is Bettinger and Long, 2005, who do find evidence of role models for males in a few disciplines- education and business).

### 5.3 Understanding the Insignificance of Future Income in the Choice of Majors

This section discusses some robustness checks in order to determine whether income is actually insignificant in the choice of major or if the result is driven by large standard errors. One concern could be that individuals are not aware of earning differences across majors, and that could be driving the result. Table 9 presents the average and median beliefs of the respondents. Since individuals majoring in a field may have better information about their chosen field and may have beliefs different from those of individuals not majoring in it, I split survey responses by whether the respondent majors in the category about which the question is asked. Since Northwestern University does not follow its alumni, I use the 2003 average annual salaries for 1993 college graduates from selective colleges in the Baccalaureate & Beyond Longitudinal Study (B&B: 1993/2003) for comparison purposes.<sup>23</sup> These statistics are presented in columns (1) and (2) of Table 9. The average and median beliefs of respondents majoring in the field are similar to those who do not major in that field. Survey respondents, both males and females, seem to be aware of income differences across majors. However, both report median and average salaries larger than those for the B&B sample. It could be that the survey respondents are *self-enhancing* their own salary expectations.<sup>24</sup> However, there are at least three legitimate reasons why respondents' earning expectations may be different from the earnings statistics in the B&B sample. First, even though I have restricted the B&B sample to selective institutions, Northwestern graduates may work at jobs very different from those of graduates from comparable institutions. Second, respondents might think that future earnings distributions will differ from the current ones. Third, respondents may have private information (other than gender) about themselves which justifies having different expectations. Though the descriptive analysis of respondents' expectations of income in different majors in Table 9 indicates that students are aware of the income differences across majors, the variation in their responses is much larger than in actual data (for males in particular). This indicates that the insignificance of income might be driven by the noise in the reported expectations. I undertake the decomposition in equation (14) for 1000 bootstrap samples for each of the sub-samples. The bootstrap confidence interval of  $R_{\gamma_4}$  for both males and females does not include zero: the higher end of the 90% bootstrap interval for expected income is 16% and 7.5% for males and females respectively. This seems to suggest that  $\gamma_4$  is insignificant because of a large standard error, and not because it is a precise zero.<sup>25</sup>

A second reason for the insignificance of income in the choice could in part be because of the risk-neutrality assumption embedded in the model specification. This assumption was primarily made so that it would suffice to elicit the expected value for the continuous outcomes. In the absence of this assumption, I would have had to elicit multiple points on the subjective income distribution for each major in one's choice set (as in Dominitz and Manski, 1996) which would not have been feasible for the purposes of this study. Since several studies have concluded that women are more risk averse than men in their choices (see Eckel and Grossman, 1992, and Croson and Gneezy, 2008, for a summary of the literature), results in the current study regarding gender

---

<sup>23</sup> Colleges with high selectivity, and the same Carnegie Code classification as Northwestern were used for comparison. Assuming students graduate from college at the age of 22, this would be their salary at 32.

<sup>24</sup> Smith and Powell (1990) find that male college seniors report higher income expectations for themselves than they do for their college peers at the same school.

<sup>25</sup> An additional robustness check that I did was to estimate the model using the ordinal ranking of income (instead of expected income). This allows me to control for the noise in the reported income expectations. The coefficient on (ranked) income is now significant for the males, but continues to be insignificant for females. Moreover, the confidence interval of  $R_{\gamma_4}$  is [3.8%, 29.2%] for males, and [3.6%, 18.7%] for females. The overall contribution of income and social status, however, does not change since ranked income picks up a substantial part of the contribution of status towards the choice (ranked income and status are highly correlated). Therefore, none of the results change. However, this seems to suggest that income is at least significant for males.

differences in preferences for income could be a consequence of the risk-neutrality assumption.

## 5.4 Estimation with Heterogeneous Preferences

The analysis undertaken in Zafar (2008) shows that beliefs for various outcomes are associated with demographic characteristics and cultural proxies. However, it could be the case that preferences for the different outcomes also depend on individual characteristics. For example, if individuals have declining marginal utility of consumption, and preferences are separable in consumption and non-pecuniary outcomes, then the value of pecuniary outcomes will be higher for individuals from low-income households. Such heterogeneity, if not accounted for, may bias the estimates presented in section 5.1. Several empirical studies have documented the influence of family and society in the endogenous formation of preferences. For example, Fernandez et al. (2004) find that whether a male's mother worked while he is growing up is correlated with whether his wife works, and interpret this as preference transmission. Similarly, Guiso et al. (2006) present evidence of culture affecting individuals' preferences. I now relax the assumption of section 5.1 that the utility for each binary outcome  $u_r(b_r)$ , and the constants  $\gamma_q$  for the continuous outcomes do not depend on individual characteristics other than gender. Though I have relatively rich demographic information on the respondents, it is not possible to account for heterogeneity in all outcomes because of the small sample size. I, therefore, consider heterogeneity along the following dimensions:

1. An individual might care about her parent's approval for several reasons. She might be more inclined to ensure that her parents approve of her choice if she relies on them for college support. Moreover, concern for parent's approval might depend on the individual's cultural and ethnic background. I allow for heterogeneity in the utility for approval of parents by incorporating the financial support an individual receives from her parents when in college, and whether her parents are foreign-born or not.
2. Children growing up in divorced/separated households make different choices than other individuals (Gruber, 2004). Here I consider the effect of growing up in such a household on the individual's preference for being able to reconcile work and family.
3. An individual's preference for the social status of jobs may vary by her cultural background. In certain cultures, immense importance is given to the status of the jobs. This heterogeneity is accounted for by taking into account whether the individual's parents are foreign-born.
4. If non-pecuniary outcomes are a normal good, an individual from a low-income family will value the income profiles associated with the majors more relative to other individuals. I account for this heterogeneity by including information on parent's annual income. I also allow for heterogeneity by taking into account whether an individual's parents are foreign-born or not.

The enriched utility function for individual  $i$  is:

$$\begin{aligned}
& U(X_i, \{P_{im}(b_r), E_{im}(d_q)\}_{r \in \{1, \dots, 7\}, q \in \{1, \dots, 4\}\}) \\
&= \sum_{r=\{1,2,3,5,6,7\}} P_{im}(b_r = 1) \Delta u_r + \Delta u_4 [\text{parents}'\_support_i \times (1 - \text{Foreign}_i) \times P_{im}(b_4 = 1)] \\
&+ \widetilde{\Delta u}_4 [\text{parents}'\_support_i \times \text{Foreign}_i \times P_{im}(b_4 = 1)] + \widetilde{\Delta u}_7 [\text{divorced}_i \times P_{ijt}(b_7 = 1)] \\
&+ \sum_{q=1}^2 \gamma_q E_{im}(d_q) + \gamma_3 [(1 - \text{Foreign}_i) \times E_{im}(d_3)] + \widetilde{\gamma}_3 [\text{Foreign}_i \times E_{im}(d_3)] + \gamma_4 E_{im}(d_4) \\
&+ \gamma_4^{HI} [E_{im}(d_4) \times (1 - \text{low\_inc}_i) \times (1 - \text{Foreign}_i)] + \widetilde{\gamma}_4^{HI} [E_{im}(d_4) \times (1 - \text{low\_inc}_i) \times \text{Foreign}_i] \\
&+ \gamma_4^{LI} [E_{im}(d_4) \times \text{low\_inc}_i \times (1 - \text{Foreign}_i)] + \widetilde{\gamma}_4^{LI} [E_{im}(d_4) \times \text{low\_inc}_i \times \text{Foreign}_i] + \varepsilon_{im} \\
&\qquad \qquad \qquad \forall m = 1, \dots, 8
\end{aligned} \tag{15}$$

where  $\text{low\_inc}$  is a dummy variable that equals one if the individual's parents earn less than \$150,000 annually;  $\text{parents}'\_support}$  captures the financial support an individual receives from her parents,<sup>26</sup>  $\text{Foreign}$  is a dummy that equals one if either of the individual's parents is foreign-born, and  $\text{divorced}$  is a dummy that equals one if the individual's parents are either separated or divorced.

I continue to assume that the random terms  $\{\varepsilon_{ik}\}$  are independent for every individual  $i$  and choice  $k$ . Table 10 presents the maximum likelihood estimates of this model using stated preference data. Estimates from the pooled sample in Column (1) show that difference in utility levels is still largest and positive for enjoying coursework, and that the coefficient is almost unchanged from the specification with homogenous preferences. The coefficients of the outcomes for which heterogeneity is not considered stay almost the same as that in the earlier specification. With this enriched specification, the difference in utility levels for parent's approval is 0.34 for individuals with US-born parents who do not receive college support from their parents, and 2.04 for individuals who annually receive more than \$25,000 in college support from their parents. This is consistent with the hypothesis that approval of parents matters more to individuals who depend on their parents for college funding. However, I don't find support for this hypothesis for individuals with foreign-born parents. The difference in utility levels for reconciling work and family continues to be insignificant. Individuals with separated or divorced parents have a negative coefficient for reconciling work and family, but it is not significantly different from zero. Introducing heterogeneity for the status outcome gives an interesting result. Status of the available jobs, an important determinant in the choice in the earlier specifications, is not important to individuals with US-born parents. However, for individuals with foreign-born parents, a unit increase in the social status of the jobs changes the utility by as much as a 8% increase in the probability of the most important outcome, enjoying coursework. This implies that the large positive coefficient on the social status of jobs in earlier specifications is being driven by the preferences of individuals with foreign-born parents in the sample.

<sup>26</sup>It is increasing in the financial support an individual receives from her parents. Parents' support = 1 if no education expenses are paid by one's parents; equals 2 if they pay less than \$5,000; equals 3 if they pay between \$5,000- \$10,000; equals 4 if they pay between \$10,000-\$15,000; equals 5 if they pay between \$15,000- \$25,000; equals 6 if they pay \$25,000+.

The coefficient on income at age 30 is still not significantly different from zero. However, there is weak support for the hypothesis that individuals from low-income households value the future earnings profile more in their choice.

Columns (2) and (3) of Table 10 present the results of the heterogeneous choice model for the male and female sub-sample respectively. Coefficients of outcomes which are not interacted with any demographic variables are almost unchanged with respect to the corresponding specification (column 4 in Table 5). For males with US-born parents, difference in utility levels for approval of parents varies from 0.578 when receiving no support from parents to 3.47 when annually receiving more than \$25,000 in support from them. The corresponding coefficient for females with US-born parents is only half in magnitude to that for males. The coefficient on parents' approval for females with foreign-born parents is similar in magnitude to that of males with US-born parents. Surprisingly, the utility change in approval of parents for males with foreign-born parents is not significantly different from zero. On the other hand, social status of jobs only matters to males with foreign-born parents: a unit increase in the social status of the jobs changes the utility by about a 13% increase in the probability of enjoying coursework for these males. Earnings at the age of 30 are a significant determinant for males belonging to low-income families with foreign-born parents.

To gain insight into the magnitude of these parameters, Table 11 shows the results of the decomposition methodology outlined in equation (14). Except for males with foreign-born parents, non-pecuniary attributes explain more than half of the choice. For individuals with US-born parents, more than two-thirds of the choice is driven by non-pecuniary motivations; the non-pecuniary outcomes at college are of utmost importance to this group. For individuals with foreign-born parents, pecuniary outcomes at the workplace are of greatest value in the choice for males, while non-pecuniary outcomes at college continue to be of utmost importance to such females.

To recap the findings in this section, enjoying coursework and enjoying work at the available jobs are outcomes most important in the decision. Demographic characteristics bias preferences in favor of certain outcomes. Males with foreign-born parents are primarily driven by the pecuniary attributes when making their choice of college major, while the converse is true for all other groups. Parent's approval matters to all individuals except for males with foreign-born parents. One of the mechanisms through which parent's approval matters is the extent of an individual's reliance on them for college support. Finally, social status of jobs only matters to males whose parents are foreign-born.

## 5.5 Parent's Approval and Peer Effects

The estimation results in sections 5.1 and 5.4 show that approval of parents is an important determinant in the choice for males with US-born parents and for all females. The social psychology literature documents a similar finding for females: Vincent et al. (1998) find that females' perceptions of their parent's preferences for them predict their career orientation. Though section 5.4 shows that one channel through which parent's approval matters is the individual's reliance on them for college support, it is not clear what majors parents are more likely to approve, and what criteria they use for approving a major. Since only the beliefs of students are observed, I can only study the relationship between students' beliefs of parent's approval of a major and their

own beliefs of other outcomes associated with the choice.<sup>27</sup> Controlling for the individual’s major, I regress respondent  $i$ ’s beliefs about her parent’s approval for major  $j$  on her beliefs about the other outcomes associated with  $j$ . More specifically, I consider the following regression model:

$$P_{ij}(b_4 = 1) = \delta_i + \lambda_j + \boldsymbol{\alpha}' X_{ij} + \boldsymbol{\beta}' \left[ \sum_{\substack{c=1 \\ c \neq 4}}^7 P_{ij}(b_c = 1) + \sum_{q=1}^4 E_{ij}(d_q) \right] + \varepsilon_{ij} \quad (16)$$

where  $\delta_i$  is an individual fixed-effect,  $\lambda_j$  is a field-fixed effect,  $X_{ij}$  is a vector of individual-specific controls, and  $\boldsymbol{\beta}$  is the vector of interest. The results are presented in Table 12. Students’ beliefs of parent’s approval for a given major increase in their beliefs of finding a job upon graduation, enjoying work at potential jobs, and social status of jobs. Expectation of parent’s approval for a major increases by nearly 3 points (on a scale of 0-100) if the probability of finding a job upon graduation in that major increases by 10 points. This effect is even stronger for students with foreign-born parents: students believe that switching to a major with a 10 points higher probability of getting a job upon graduation is likely to increase parent’s approval by nearly 5 points. A positive and significant effect, half in magnitude to that of finding a job, is found for the social status of the jobs. Again the effect is stronger for students with foreign-born parents. The only other outcome that affects beliefs about parent’s approval is the expectation of enjoying work at the jobs for females. Another notable point is that, for females only, parent’s approval is higher by about 5 points for one’s chosen major.

Males with foreign-born parents expect approval of parents for a major to increase by about 12.5 points for a unit increase in the social status of the jobs. This result reconciles the earlier finding in section 5.4 of parent’s approval not mattering to males with foreign-born parents. Expectation of parent’s approval has a positive relationship with the perceived social status of jobs, and status of jobs is an important outcome only in the choice for males with foreign-born parents (column (2) in Table 10); hence, because of colinearity, approval of parents does not *directly* affect the choice of these individuals.

A second important issue is the extent to which one’s choice of major is influenced by their parents and peer group. The importance of peer effects in shaping individual choices has been documented in several studies within higher education (see, for example, Betts and Morell, 1999), but there is little research on peer effects in crucial decisions such as choice of college major. Sacerdote (2001) does not find evidence for (roommate) peer effects in major choice for Dartmouth College roommates. De Girogi et al. (2007) find that Bocconi undergraduates are more likely to choose a major when many of their peers make that choice. Several respondents in my survey report to have majors that are the same as that of their roommates and friends. However, there is a self-selection issue: people often select with whom they associate.<sup>28</sup> Since rooming assignments are not totally random at Northwestern and there are endogeneity issues in how friendships are being formed, I cannot analytically study the strength of peer effects in the choice of college major.

There seems to be a positive correlation between one’s major and that of their parents. In particular, students pursuing a major in Natural Sciences are more likely to have a parent who majored in that category. Of the 63 individuals with at least one sibling, 22 major in the same field as their sibling. A positive correlation

<sup>27</sup>It could be that parents have subjective beliefs about the outcomes that are very different from those of the student. However, I can only analyze the relationship the student *believes* exists between her expectation of parent’s approval and her subjective expectations of the various choice-specific outcomes.

<sup>28</sup>See Manski (1993); basically if the peers with whom a person associates share his attributes and also affect his attainment, and are not observed by the researcher, then the researcher might falsely attribute a peer effect where one does not exist.

between an individual's choice of college major with that of her parents or siblings could be consistent with either (1) her having more information about that particular choice by information acquisition of the various outcomes from her parents and siblings, and hence choosing that major through an *indirect* effect of parents, (2) *direct* parental pressure leading an individual towards a particular major choice, or (3) a utility gain by studying the same major as that of parents. For example, in Zafar (2008), I find evidence of the first: individuals are less likely to make errors in reporting starting salaries of majors if their father pursued a major in that category. Moreover, when estimating preferences which incorporate individual heterogeneity in section 5.4, demographic characteristics (like country of birth of parents) are found to bias preferences for certain outcomes. However, it is not possible to identify the channel through which this is happening, i.e. whether beliefs and preferences are subconsciously being formed as a consequence of the individual's interactions with parents, or whether parents are intentionally shaping the beliefs and preferences of their children (as in Bisin and Verdier, 2001), or both. Survey respondents were asked to explain the reasons for the similarity between their major and that of their parents and siblings. Selected responses are shown in section 9.2.2 of the Appendix. All three reasons come up as possible explanations. The responses also show instances of peer influence, but in most cases individuals seem to form friendships with similar individuals.

## 5.6 Robustness Checks

The model estimated in section 5.1 assumes that all individuals have homogeneous preferences for various outcomes. Individuals with different characteristics are very likely to have different preferences. Moreover, the assumption that the random terms  $\{\varepsilon_{ik}\}$  are independent for every individual  $i$  and choice  $k$  might be very strong. Though a model with limited heterogeneity in preferences is estimated in section 5.4, any unaccounted or unobserved heterogeneity may bias the model estimates. In this section, I specify a random parameters logit model to account for these issues (see Revelt and Train, 1997, for a discussion of mixed logit models). One could allow heterogeneity in preferences for all outcomes, but I focus on the most important outcomes: I consider a model in which the differences in utility levels for graduating with a GPA of at least 3.5, enjoying the coursework, approval of parents, enjoying work at the available jobs, and the parameter for social status of the available jobs are allowed to vary in the population with a specified distribution. The utility that individual  $i$  receives from choosing major  $m$  is:

$$\begin{aligned}
& U(X_i, \{P_{im}(b_r), E_{im}(d_q)\}_{r \in \{1, \dots, 7\}, q \in \{1, \dots, 4\}}) \\
&= \sum_{r=\{1,5,7\}} P_{im}(b_r = 1) \Delta u_r + \sum_{s=\{2,3,4,6\}} P_{im}(b_s = 1) \Delta u_{si} \\
&+ \sum_{q=\{1,2,4\}} \gamma_q E_{im}(d_q) + \gamma_{3i} E_{im}(d_3) + \varepsilon_{im}
\end{aligned} \tag{17}$$

where  $\Delta u_{si}$  for  $s = \{2, 3, 4, 6\}$ , and  $\gamma_{3i}$  are allowed to vary in the population according to a specified parametric distribution, and  $\varepsilon_{im}$  is an iid random term that is extreme value distributed. I denote the vector of parameters  $\{\Delta u_{2i}, \Delta u_{3i}, \Delta u_{4i}, \Delta u_{6i}, \gamma_{3i}\}$  by  $\beta_i$ , and the density of these parameters  $f(\beta_i | \theta)$  where  $\theta$  are the parameters

of the distribution. The probability of  $i$  choosing the major  $m$  conditional on  $\beta_i$  is:

$$\Pr(m|\beta_i) = \Pr(m|\{P_{ik}(b_r), E_{ik}(d_q)\}_{r \in \{1, \dots, 7\}, q \in \{1, \dots, 4\}; k \in C_i; \beta_i}) = \frac{\exp(\sum_{r=\{1,5,7\}} P_{im}(b_r = 1)\Delta u_r + \sum_{s=\{2,3,4,6\}} P_{im}(b_s = 1)\Delta u_{si} + \sum_{q=\{1,2,4\}} \gamma_q E_{im}(d_q) + \gamma_{3i} E_{im}(d_3))}{\sum_{k \in C_i} \exp(\sum_{r=\{1,5,7\}} P_{ik}(b_r = 1)\Delta u_r + \sum_{s=\{2,3,4,6\}} P_{ik}(b_s = 1)\Delta u_{si} + \sum_{q=\{1,2,4\}} \gamma_q E_{ik}(d_q) + \gamma_{3i} E_{ik}(d_3))} \quad (18)$$

The unconditional probability of choosing  $m$  is the integral of this conditional probability over all possible values of  $\beta_i$ , and depends on the parameters  $\theta$  of the distribution of  $\beta_i$ . The unconditional probability for  $i$  choosing  $m$  is:

$$P_{im}(\theta) = \int \Pr(m|\{P_{ik}(b_r), E_{ik}(d_q)\}_{r \in \{1, \dots, 7\}, q \in \{1, \dots, 4\}; k \in C_i; \beta_i}) f(\beta_i|\theta) d\beta_i \quad (19)$$

This integral is approximated through simulation since it cannot be calculated analytically. For a given value of the parameter vector  $\theta$ , a value of  $\beta_i$  is drawn from its distribution. Using this draw, the conditional probability is calculated. This process is repeated for  $D$  draws, and the average is taken as the approximate choice probability:

$$\widehat{P_{im}(\theta)} = \frac{1}{D} \sum_{d=1}^D \Pr(m|\{P_{ik}(b_r), E_{ik}(d_q)\}_{r \in \{1, \dots, 7\}, q \in \{1, \dots, 3\}; k \in C_i; \beta_i^d})$$

The log likelihood function  $\sum_i \ln(\Pr_i)$  is approximated by the simulated log-likelihood function  $\sum_i \ln(\widehat{P_i(\theta)})$ , and the estimated parameters are those that maximize the simulated log-likelihood function. I assume that the coefficients for graduating with a GPA of at least 3.5, enjoying the coursework, approval of parents, enjoying work at the available jobs, and social status of the available jobs are independently log-normally distributed.<sup>29</sup> The difference in utility levels for an outcome  $k$  which is assumed to vary in the population is expressed as  $\Delta u_k = \exp(\overline{\Delta u_k} + \sigma_k \mu_k)$  where  $\mu_k$  is a standard normal deviate. The parameters  $\overline{\Delta u_k}$  and  $\sigma_k$ , which represent the mean and standard deviation of  $\log(\Delta u_k)$  are estimated. The mean and standard deviation of  $\Delta u_k$  are  $\exp(\overline{\Delta u_k} + \frac{\sigma_k^2}{2})$  and  $\exp(\overline{\Delta u_k} + \frac{\sigma_k^2}{2}) * \sqrt{(\exp(\sigma_k^2) - 1)}$  respectively.

Columns (1a)-(1c) in Table 13 present the estimates of the mixed logit specification for the model with  $D = 100,000$ . Estimates of various outcomes are similar to those obtained in the corresponding model with no heterogeneity (column 3 of Table 5). The mean coefficient of enjoying coursework is still largest in absolute value and significant. The estimated standard deviations of the (random) coefficients are highly significant indicating that these parameters do indeed vary in the sample. Standard deviations for coefficients of graduating in 4 years and social status of available jobs are especially very large, indicating that there is substantial heterogeneity in how these outcomes are valued in the sample (consistent with what was also found in the previous section). Another point of note is that the mean coefficients in the mixed logit model are larger than the corresponding fixed coefficients in Table 5. This is because in the mixed logit, some of the stochastic portion of the utility is captured in  $\beta_i$  rather than in  $\varepsilon_i$ . Since the utility is scaled so that  $\varepsilon_i$  has the variance of an extreme value, the parameters are scaled down in the standard model relative to the mixed logit model (the same result is obtained by Revelt and Train, 1998, and Brownstone and Train, 1999). The fact that the mean coefficients are bigger than the fixed coefficients implies that the random parameters constitute a large share of the variance

<sup>29</sup>I use a log-normal distribution instead of a normal distribution for these parameters since these are all outcomes which one would expect to be desirable to an individual. The normal distribution allows coefficients of both signs, and implies that some share of the sample has negative coefficients for those outcomes, whether or not it is true. The log-normal assumption ensures that each respondent in the sample has a positive coefficient for these outcomes.

in unobserved utility.

One might wonder as to what extent can the variation in the parameters in the mixed logit model be explained by including demographic characteristics. Columns (2a) through (2c) in Table 13 present estimates of the mixed logit model with demographic variables that were used in the heterogeneous model described in section 5.4. The estimates are similar to those in column (1) of Table 10, though they are larger in magnitude, which is expected. The standard deviations are still large and significant which indicates that the demographic variables considered in section 5.4 only capture some of the heterogeneity that is exhibited by the individuals. Nonetheless, the fact that the relative magnitude of the estimates is similar to previous results is reassuring.

## 6 Double Major Choice Model

For reasons that will become clear shortly, a separate choice model is estimated for double majors. Nearly half of the sample respondents state that they are pursuing two majors. Anecdotal evidence suggests that about half of them will end up dropping one of their majors some time before graduation.<sup>30</sup> Since I have stated preference data from these respondents, I first estimate the same model as in section 5.1 in order to get a sense of the motivations of the choice for these individuals. The parameter estimates (available upon request) are similar to those for respondents pursuing a single major. Table 14 presents the decomposition results of equation (14) using these estimates. As before, non-pecuniary attributes explain most of the choice. It does seem that these individuals are similar to those pursuing single majors in their preferences for various outcomes.

This section outlines a model that incorporates the choice of double majors, and then deals with its estimation.

### 6.1 Estimation of Double Major Choice Model

Depending on the exact composition of the individual's major pair, the choice set of the individual now consists of either 8 or 9 categories.<sup>31</sup> For estimation, I assume that the individual may choose a single major or a pair of majors. The set of alternatives available to the individual includes all subsets of two majors in WCAS ( ${}^8C_2 = 28$ ), all possible single majors in WCAS (7), and all possible pairings of WCAS majors with non-WCAS majors for a total of 70 alternatives. The distribution of majors for individuals pursuing double majors in the sample is shown in Table 15. There's no obvious pattern in which individuals are choosing pairs.

The major-specific outcomes that appear in the utility function remain the same as before, but the form of the utility function is now different. Before specifying the structural form of the utility function, it may be useful to think about why an individual may decide to choose two majors. Respondents pursuing more than one major were asked to explain reasons for pursuing more than one major; selected responses are shown in the Appendix (section 9.2.1). Two main reasons emerge: first, two majors appropriately differentiated can provide a broader mix of options than a single major; second, it might be the case that no single major meets the needs of the individual. For example, an individual might be interested in both maximizing her income prospects

---

<sup>30</sup>According to the Registrar's Office and Northwestern Graduation Survey 2006, less than 30% of WCAS undergraduates graduate with more than one major.

<sup>31</sup>It would be the former if both majors are in WCAS and/or School of Engineering. In the event that one of the majors is in neither of the two schools, the choice set will be the latter, with the extra category including the majors offered in that school.

as well as enjoying the coursework. It could very well be the case that no single major meets her needs, but a combination of two majors does. To capture the enhanced options and specialization of function that two majors provide, I assume that the utility of a major pair depends on the attributes of each major separately, and on the attributes of a composite major combining the best of both majors. However, I only apply the idea of a composite major to outcomes associated with college. Outcomes associated with the workplace are not considered since they come as a package; for example, one does not have the option to choose the income associated with the jobs available in one major, and the lifestyle associated with the jobs in the second major. I also do not consider the composite major representation for graduating with a GPA of more than 3.5 because GPA is a composite of all coursework an individual does, and it is not possible to hedge along this dimension. The outcomes for which the composite major specification is used are: graduating in 4 years, hours/week spent on coursework, enjoying the coursework, approval of parents, and finding a job upon graduation. The utility function of a major pair  $p$  consisting of majors  $p_1$  and  $p_2$  takes the form:

$$\begin{aligned}
U_{ipt} &= U_{ip_1t}(\mathbf{b}, \mathbf{d}, X_{it}, \{P_{ip_1t}(b_r = 1)\}_{r=1}^7, \{E_{ip_1t}(d_q)\}_{q=1}^4) \\
&+ U_{ip_2t}(\mathbf{b}, \mathbf{d}, X_{it}, \{P_{ip_2t}(b_r = 1)\}_{r=1}^7, \{E_{ip_2t}(d_q)\}_{q=1}^4) \\
&+ U_{i\tilde{p}t}(\mathbf{b}, \mathbf{d}, X_{it}, \sum_{r=\{1,3,4,5\}} \max[P_{ip_1t}(b_r = 1), P_{ip_2t}(b_r = 1)], \max[E_{ip_1t}(d_1), E_{ip_2t}(d_1)])
\end{aligned} \tag{20}$$

where  $U_{ip_1t}(\cdot)$  is as defined in equation (8), and  $\tilde{p}$  refers to the composite major. Since there is no way of specifying a "primary" and a "secondary" major, I use the same functional form for the utility of each major in one's major pair, i.e.  $U_{ip_1t} = U_{ip_2t}$ . Since  $U_{ip_1t}(\cdot)$  is linear-in-parameters, the average characteristics of the two majors appear in the utility function. Assuming that the utility function does not depend on the individual characteristics,  $X_{it}$ , and dropping the time subscript, the utility function can be written as:

$$\begin{aligned}
&U_{ip}(\{P_{ip_1}(b_r), E_{ip_1}(d_q), P_{ip_2}(b_r), E_{ip_2}(d_q)\}_{r \in \{1, \dots, 7\}, q \in \{1, \dots, 4\}}) \\
&= \sum_{r=1}^7 \left\{ \frac{P_{ip_1}(b_r=1) + P_{ip_2}(b_r=1)}{2} \right\} \Delta u_{r1} + \sum_{q=1}^4 \gamma_{q1} \left\{ \frac{E_{ip_1}(d_q) + E_{ip_2}(d_q)}{2} \right\} \\
&+ \sum_{r=\{1,3,4,5\}} \max[P_{ip_1}(b_r = 1), P_{ip_2}(b_r = 1)] \Delta u_{r2} \\
&+ \gamma_{12} \min[E_{ip_1}(d_1), E_{ip_2}(d_1)] + \varepsilon_{ip} = U_{ip} + \varepsilon_{ip}
\end{aligned} \tag{21}$$

The composite major representation captures the notion of functional specialization as follows: say an individual with a major pair chooses one major with a low completion probability because of some of its other attributes, and a second major where the completion probability is the most important consideration. Given the specification above, one would expect  $\Delta u_{11} \approx 0$  and  $\Delta u_{12} > 0$  in this case of extreme specialization. On the other hand, for an individual who equally values the completion probabilities associated with both her majors,

one would expect  $\Delta u_{12} \approx 0$  and  $\Delta u_{11} > 0$ . Thus the ratio  $\Delta u_{12}/\Delta u_{11}$  ( $\{\Delta u_{r2}/\Delta u_{r1}\}_{r=\{1,3,4,5\}}$ ,  $\gamma_{12}/\gamma_{11}$ ) is a measure of the extent to which an individual desires to functionally specialize her majors along the given outcome.

I continue to assume that the random terms  $\{\varepsilon_{ip}\}$  are independent for every  $i$  and every  $p$ , and have an extreme value distribution. The maximum likelihood estimates are shown in Table 16. Panel B shows  $\{\Delta u_{r1} + \Delta u_{r2}\}_{r=1}^7$ ; the relative magnitudes of  $\{\Delta u_{r1} + \Delta u_{r2}\}_{r=1}^7$  are a measure of the importance of each outcome in choosing a major pair. For the pooled sample results presented in column (1), the difference in utility levels for graduating in 4 years is positive and largest in magnitude. The next most important outcome is enjoying the coursework with a positive coefficient. Approval of parents is the third most important outcome. Enjoying work at potential jobs, graduating with a GPA of at least 3.5, finding a job, and reconciling work and family are next in order of importance. All four are significant with positive coefficients. The coefficient on hours/week at the jobs is positive, which is rather surprising. However, an increase of 5 hours/week at work only increases the utility by as much as a 1% increase in the probability of graduating in 4 years. The coefficients on status of the jobs, hours/week spent on coursework, and earnings at 30 are not significantly different from zero.

Next I check for evidence of specialization. Estimates presented in Table 16 suggest that there is strong evidence of extreme specialization for graduating in 4 years ( $\Delta u_{12} > 0$ ,  $\Delta u_{11} \approx 0$ ), and for finding a job ( $\Delta u_{52}/\Delta u_{51} \gg 1$ ). This implies that individuals concentrate their chances of graduating in 4 years, and getting a job upon graduation in one of the majors in their major pair.<sup>32</sup> On the other hand, approval of parents and enjoying coursework are outcomes that are important in the choice of both majors (i.e.  $\Delta u_{41} > 0$ ,  $\Delta u_{42} \approx 0$  and  $\Delta u_{31} > 0$ ,  $\Delta u_{32} \approx 0$  respectively). The coefficient on hours/week spent on coursework,  $\gamma_{12}$ , is negative; this supports the specialization hypothesis, i.e. individuals prefer pairs of majors that entail different hours/week in college.

Columns (2) and (3) in Table 16 shows the estimates for the males and females sub-samples respectively. The three most important outcomes for both are the same: enjoying the coursework, graduating in 4 years, and approval of parents (though not in the same order). For males, an analysis of the ratios of  $\{\Delta u_{r2}/\Delta u_{r1}\}_{r=\{1,3,4,5\}}$ , and  $\gamma_{12}/\gamma_{11}$  reveals that they prefer to choose majors that differ in their chances of graduating in 4 years ( $\Delta u_{12} > 0$ ,  $\Delta u_{11} \approx 0$ ), in enjoying the coursework ( $\Delta u_{32}/\Delta u_{31} \gg 1$ ), and approval of parents ( $\Delta u_{52}/\Delta u_{51} \gg 1$ ). The coefficient on hours/week spent on coursework,  $\gamma_{12}$ , is negative implying that males prefer pairs of majors with different coursework levels. Females, like their male counterparts, prefer majors that entail different chances of graduating in 4 years ( $\Delta u_{12} > 0$ ,  $\Delta u_{11} \approx 0$ ). In addition, they prefer majors that differ in their chances of getting a job upon graduation ( $\Delta u_{52} > 0$ ,  $\Delta u_{51} \approx 0$ ). There is also some evidence of females preferring majors with different amounts of workload in terms of coursework ( $\gamma_{12} < 0$ ). On the other hand, approval of parents and enjoying coursework matter significantly in the choice of both majors ( $\Delta u_{41} > 0$ ,  $\Delta u_{42} \approx 0$  and  $\Delta u_{31} > 0$ ,  $\Delta u_{32} \approx 0$ ).

This model exhibits the restrictive IIA property, which is not a very realistic assumption in this particular situation. For example, one could imagine that an individual majoring in Area Studies and Literature & Fine Arts is more likely to choose Area Studies and Ethics & Values, rather than Natural Sciences and Ethics & Values. To allow flexible substitution patterns, I allow for a stochastic part for each major that is perhaps correlated over majors and heteroskedastic over individuals and majors (these appear as 12 random effects, one

<sup>32</sup>There is ample evidence of the latter in the comments submitted by the respondents (see Appendix 1).

for each of the 7 alternatives in WCAS, and the 5 categories outside WCAS), and another stochastic part that is iid over individuals and alternatives. The utility function of a major pair  $p$  is now:

$$U_{ip}(\{P_{im}(b_r), E_{im}(d_q)\}_{r \in \{1, \dots, 7\}, q \in \{1, \dots, 3\}}) = U_{ip} + \varepsilon_{ip} + c_{p1}\eta_{i,1} + c_{p2}\eta_{i,2} + \dots + c_{p12}\eta_{i,12}$$

where, as before,  $\varepsilon_{ip}$  is a random term with zero mean that is iid over alternatives of major pairs, and is normalized to set the scale of utility. The  $\eta_{i,m}$  for  $m = \{1, \dots, 12\}$  are normally distributed effects with zero mean, and  $c_{px} = 1$  if major  $x$  appears in the major pair  $p$ .<sup>33</sup> This structure allows flexible substitution patterns across alternatives. For example, the correlation between a major pair  $\kappa$  consisting of  $m = \{1, 2\}$ , and a second major pair  $\omega$  consisting of majors  $m = \{2, 3\}$  is  $E([U_{i\kappa} + \varepsilon_{i\kappa} + \eta_{i,1} + \eta_{i,2}][U_{i\omega} + \varepsilon_{i\omega} + \eta_{i,2} + \eta_{i,3}]) = Var(\eta_{i,2})$ . So utility is now correlated over alternatives. Given the vector  $\boldsymbol{\eta}_i$ , the *conditional* choice probability is simply logit, since the remaining error term is iid extreme value. The probability of individual  $i$  choosing the major pair  $p$  is:

$$\begin{aligned} \Pr(p|\boldsymbol{\eta}_i) &= \Pr(p | \{P_{im}(b_r), E_{im}(d_q)\}_{r \in \{1, \dots, 7\}, q \in \{1, \dots, 4\}}, \boldsymbol{\eta}_i) \\ &= \frac{\exp(U_{ip} + c_{p1}\eta_{i,1} + c_{p2}\eta_{i,2} + \dots + c_{p12}\eta_{i,12})}{\sum_{k \in C_i} \exp(U_{ik} + c_{k1}\eta_{i,1} + c_{k2}\eta_{i,2} + \dots + c_{k12}\eta_{i,12})} \end{aligned}$$

The *unconditional* choice probability is the expected value of the *conditional* probability over all the possible values of  $\boldsymbol{\eta}_i$ , and depends on  $\mathbf{g}(\boldsymbol{\eta}_i|\Omega)$ , the density of  $\boldsymbol{\eta}_i$ . It is:

$$P_i(\Omega) = \int \Pr(p|\boldsymbol{\eta}_i) \mathbf{g}(\boldsymbol{\eta}_i|\Omega) d\boldsymbol{\eta}_i$$

Since the integral does not have a closed form in general, it is approximated through simulation. 100,000 draws of  $\boldsymbol{\eta}_i$  for a given value of the parameters  $\Omega$  are drawn; for each draw, the  $\Pr(p|\boldsymbol{\eta}_i)$  is calculated, and the average of these probabilities is taken as the approximate choice probability:

$$\widehat{P}_i(\Omega) = \frac{1}{100,000} \sum_{d=1}^{100,000} \Pr(p|\boldsymbol{\eta}_i^d)$$

The estimated parameters from maximizing the simulated log-likelihood,  $\sum_i \ln(\widehat{P}_i(\Omega))$ , are shown in Table 18. The coefficients are similar in relative magnitude, but larger in absolute terms than the corresponding fixed coefficients in column (1) of Table 16. This is because, in the standard model, all stochastic terms are absorbed into one error term,  $\epsilon$ . The variance of this error term is larger in the standard logit model than in a mixed logit since some of the variance is now captured by the  $\eta$ 's rather than the  $\epsilon$  in the mixed logit model. Since utility is scaled so that  $\epsilon$  has the variance of an extreme value, the variance before scaling is larger in the standard logit than the mixed logit, and hence parameters are scaled down in a standard logit relative to the mixed logit. Graduating in 4 years, enjoying the coursework, and approval of parents continue to be the three most important outcomes. Individuals choose majors in their choice pair such that they enjoy coursework and have approval of parents in both majors ( $\Delta u_{31} > 0$ ,  $\Delta u_{32} \approx 0$  and  $\Delta u_{41} > 0$ ,  $\Delta u_{42} \approx 0$ ). Graduating in 4 years is an important consideration for both majors, but there is some evidence that individuals prefer majors that differ in their chances of graduating in 4 years ( $\Delta u_{12}/\Delta u_{11} > 1$ ). Individuals also prefer pairs of majors

<sup>33</sup>For example, the utility function of a major pair  $p$  that includes Natural Sciences ( $m = 1$ ), and Social Sciences II ( $m = 4$ ) would be:  $U_{ip}(\{E_i(Y_m), P_{im}(b_r), E_{im}(d_q)\}_{r \in \{1, \dots, 7\}, q \in \{1, \dots, 4\}}) = U_{ip} + \varepsilon_{ip} + \eta_{i,1} + \eta_{i,4}$

that allow them different chances of getting a job upon graduation ( $\Delta u_{52} > 0$ ,  $\Delta u_{51} \approx 0$ ). Graduating with a GPA of at least 3.5 has a positive coefficient but is not significant. The somewhat puzzling results are the positive coefficients on hours/week spent on coursework, and at the jobs (the latter is not significant).  $\gamma_{12}$ , the coefficient on  $\min[E_{ip_1}(d_1), E_{ip_2}(d_1)]$ , is negative suggesting that individuals prefer pairs of majors with different time commitments at college. However, it is not significantly different from zero.

To recap, double major individuals have preferences similar to those with single majors. Graduating in 4 years, enjoying coursework and approval of parents are the most important outcomes in the choice of a major pair. There is evidence that individuals prefer to choose pairs of majors that differ in their chances of graduating in 4 years. Females and males differ in the outcomes they specialize in. Females choose major pairs that offer different chances of finding a job, while males choose major pairs that are different in the approval of parents and enjoying coursework. On the whole, students with double majors pursue their interests at college while taking into account parents' approval, and also act strategically in their choices by choosing majors that differ in their chances of completion and finding a job upon graduation.

## 7 Understanding Gender Differences

Sections 5 and 6 show that males and females also differ in their preferences for the various outcomes. The descriptive analysis in section 3.4 and in Zafar (2008) document the heterogeneity in beliefs for various outcomes between the two genders. Though the results of the decomposition metric of equation (14) presented in Tables 6, 11, and 14 highlight the gender differences in preferences, it is not clear how much of the gender gap in the choice of college majors is driven by differences in preferences, and how much is due to differences in distributions of subjective beliefs. This distinction is important since males and females identical in their preferences will make different career choices if there are gender differences in beliefs about success in different occupations (Breen and Garcia-Penalosa, 2002). Moreover, any policy recommendations will depend on whether the gender gap exists because of innate differences, or because of social biases and discrimination. For example, if the gender gap existed because of gender differences in beliefs about ability and self-confidence, then policy interventions like single-sex classes could possibly reduce the gap.<sup>34</sup> In this section, I delve into the underlying causes for the gender gap in more detail.

### 7.1 Decomposition Analysis

As a first step, I decompose the gender gap into gender differences in beliefs and preferences. A common way to explore differences between groups in a linear framework is to express the difference in the average value of the dependent variable  $Y$  as:

$$\bar{Y}_M - \bar{Y}_F = [(\bar{X}_M - \bar{X}_F)\hat{\beta}_M] + [\bar{X}_F(\hat{\beta}_M - \hat{\beta}_F)]$$

where  $\bar{X}_j$  is a vector of average values of the independent variables and  $\hat{\beta}_j$  is a vector of the estimated coefficients for gender  $j \in \{(M)ale, (F)emale\}$ . The first term on the right hand side is the inter-group difference in mean levels of the outcome due to different observable characteristics, while the second term is the difference due

---

<sup>34</sup>However, there is mixed evidence in terms of academic achievement gap with regards to same-sex classes. See Haag's literature review in the 1998 report of the AAUW Educational Foundation.

to different effects of the characteristics. This technique is attributed to Oaxaca (1973). However, in the current context, the probability of choosing a given major,  $Y$ , is non-linear. In the case  $Y$  is nonlinear, such as  $Y = F(X\beta)$ ,  $\bar{Y}$  does not necessarily equal  $F(\bar{X}\beta)$ . The gender difference in this non-linear case can be written as:

$$\begin{aligned}\bar{Y}_M - \bar{Y}_F &= \left[ \sum_{i=1}^{N_M} \frac{F(X_{Mi}\hat{\beta}_M)}{N_M} - \sum_{i=1}^{N_F} \frac{F(X_{Fi}\hat{\beta}_M)}{N_F} \right] + \left[ \sum_{i=1}^{N_F} \frac{F(X_{Fi}\hat{\beta}_M)}{N_F} - \sum_{i=1}^{N_F} \frac{F(X_{Fi}\hat{\beta}_F)}{N_F} \right] \\ &= \overline{[F(X_M\hat{\beta}_M) - F(X_F\hat{\beta}_M)]} + \overline{[F(X_F\hat{\beta}_M) - F(X_F\hat{\beta}_F)]}\end{aligned}$$

where  $N_j$  is the sample size of gender  $j$ .<sup>35</sup> The first expression in the square brackets represents part of the gender gap that is due to gender differences in distributions of  $X$ , and the second expression represents the part due to differences in the group processes determining levels of  $Y$ . It is relatively simple to estimate the total contribution. However, identifying the contribution of group differences in specific variables/ coefficients to the gender gap is not straightforward. For this purpose, I use a decomposition method proposed by Fairlie (1999, and 2005). Contributions of a single variable/ coefficient are calculated by replacing the relevant variable of one group with that of the other group sequentially one by one. For illustration, suppose  $Y_j = F(X_j\beta_j)$  for  $j=\{F, M\}$ , and that  $X$  includes two variables,  $X_1$  and  $X_2$ . Moreover, let  $N_M = N_F = N$ , and assume there exists a natural one-to-one matching of female and male observations. The independent contribution of  $X_1$  to the gender gap is given as:

$$\frac{1}{N} \sum_{i=1}^N F(X_{1Mi}\hat{\beta}_{1M} + X_{2Mi}\hat{\beta}_{2M}) - F(X_{1Fi}\hat{\beta}_{1M} + X_{2Mi}\hat{\beta}_{2M})$$

and that of  $X_2$  is given as:

$$\frac{1}{N} \sum_{i=1}^N F(X_{1Fi}\hat{\beta}_{1M} + X_{2Mi}\hat{\beta}_{2M}) - F(X_{1Fi}\hat{\beta}_{1M} + X_{2Fi}\hat{\beta}_{2M})$$

Therefore the contribution of a variable to the gap is equal to the change in the average predicted probability from replacing the female distribution with the male distribution of that variable while holding the distributions of the other variable constant. One important thing to note is that, unlike in the linear case, the independent contributions of  $X_1$  and  $X_2$  depend on the value of the other variable. Therefore, the order of switching the distributions can be important in calculating the contribution to the gender gap.<sup>36</sup> Similarly the independent contribution of  $\beta_1$  to the gap is given by:

$$\frac{1}{N} \sum_{i=1}^N F(X_{1Fi}\hat{\beta}_{1M} + X_{2Fi}\hat{\beta}_{2M}) - F(X_{1Fi}\hat{\beta}_{1F} + X_{2Fi}\hat{\beta}_{2M})$$

and that of  $\beta_2$  is given as:

$$\frac{1}{N} \sum_{i=1}^N F(X_{1Fi}\hat{\beta}_{1F} + X_{2Fi}\hat{\beta}_{2M}) - F(X_{1Fi}\hat{\beta}_{1F} + X_{2Fi}\hat{\beta}_{2F})$$

<sup>35</sup>An equally valid expression is:  $\bar{Y}_M - \bar{Y}_F = \overline{[F(X_M\hat{\beta}_F) - F(X_F\hat{\beta}_F)]} + \overline{[F(X_M\hat{\beta}_M) - F(X_M\hat{\beta}_F)]}$ . This alternative method provides different estimates, which is the familiar index problem with the Oaxaca decomposition technique.

<sup>36</sup>Yun (2004) outlines an alternate decomposition strategy which is free from path-dependency. The method is easier to implement but I don't use it since it involves a first order Taylor approximation. Moreover, I believe that the decomposition employed in this paper is closer to what is standard in the literature.

In this illustration, I have assumed equal number of observations for females and males. However, my sample has more females than males. Since the decomposition requires one-to-one matching of female and male observations, I use the following simulation process: from the female sub-sample, I randomly draw 60 samples with the same number of observations as in the male sub-sample, and sort the female and male data by the predicted probabilities, and calculate separate decomposition estimates. The mean value of estimates from the separate decompositions is calculated and used to approximate the results from the entire female sample. As in Fairlie (2005), I approximate the standard errors using the delta method.

For the purposes of this decomposition, I treat double-major respondents as if they were pursuing a single major; I use the parameter estimates obtained from the single major choice model estimation using stated preferences of the respondents. Results of this decomposition are presented in Table 19 for four different majors.<sup>37</sup> The last row of the table shows that both expectations and preferences contribute to the gender gap for all major categories. The contributions of preferences and beliefs to the gap differ by fields: majority of the gender gap in Literature & Fine Arts and Social Sciences II is due to gender differences in beliefs, while gender differences in preferences explain most of the gap in Engineering and Social Sciences I.

If women being less overconfident than men (Niederle et al., 2007; and references therein), and women being low in self-confidence (Long, 1986; Valian, 1998) were the main explanations for the underlying gender gap, one would expect gender differences in beliefs about academic ability to be important in explaining the gender difference in major choices. However, columns (1)-(4) of Table 19 show that gender differences in beliefs about ability (more precisely beliefs about graduating in 4 years, and graduating with a GPA of at least 3.5) are insignificant and explain a small part of the gender gap. Therefore, explanations entirely based on the assumption that women are under-represented in sciences and engineering because they have lower self-confidence can be rejected in my data. Another striking observation is that gender differences in beliefs about enjoying coursework in the various fields are significant and explain a large part of the gap.

Here I discuss the decomposition results for Engineering in some detail. These results are presented in columns (1) and (5) of Table 19. The model predicts that, on average, males are nearly twice as likely as females to major in engineering (an average male probability of 0.104 versus 0.045 for females); 60% of this gap is due to gender differences in preferences for various outcomes. Moreover, nearly 27% of the gap is due to gender differences in beliefs about enjoying coursework. Interestingly, gender differences in beliefs about future earnings are insignificant and constitute less than 0.5% of the gap. Females have beliefs similar to those of males about academic ability in engineering.<sup>38</sup> These findings suggest that females are less likely to major in engineering not because they are underconfident about their academic ability, low in self-confidence, or because of beliefs about wage discrimination in the labor market. Instead this is because they believe that they won't enjoy taking courses in engineering. In other words, it's not that women think they won't be good engineers, but they think they won't enjoy studying it. The results seem to suggest that a policy that changes social attitudes might be more useful in narrowing the gap.

A question that I do not address is the source of gender differences in preferences. Gender differences in preferences could arise from differences in tastes, as well as gender discrimination. For example, parents

---

<sup>37</sup>I do not conduct this analysis for the category of Natural Sciences. This is because the category pools both life sciences and physical sciences. Traditionally, females are more likely to major in the former, and less likely to major in the latter. Since I pool them together, the decomposition analysis for the pooled category would not be very useful.

<sup>38</sup>I only observe the beliefs about academic ability, and not *actual* academic ability. However, Chemers et. al. (2001) show that confidence in one's ability is strongly related to academic performance.

who know that females would be discriminated in male-dominated majors/ occupations could try to shape the preferences of their female children so that they are more comfortable in female-dominated majors/ occupations (Altonji and Blank, 1999). The question of understanding the sources of gender differences in preferences is beyond the scope of this paper.

## 7.2 Simulations

This section simulates different environments to see how the gender gap would change under different scenarios. Column (1) of Table 20 shows the gender gap predicted by the model for the various major categories. The simulation in column (2) considers an environment where the female subjective ability distribution (beliefs about graduating within 4 years, and about graduating with a GPA of at least 3.5) is replaced with that of males.<sup>39</sup> The purpose of this simulation is to answer how much of the gap is due to females having less self-confidence in their ability (relative to men). The second simulation in column (3) replaces the female subjective earnings distribution with that of males; it is meant to answer the question of how much of the gap is due to beliefs of wage discrimination in the labor market. Columns (4) and (5) simulate an environment in which females have the same beliefs as males about enjoying coursework and enjoying work at potential jobs respectively.

I continue to focus the discussion on Engineering. The results confirm the findings obtained in Table 19. If female expectations about ability were raised to the same level as that of males through some policy intervention, the gender gap in engineering would decrease by less than 14%. The gender gap virtually stays the same if female expectations of future earnings were forced to be the same as those of males. Finally, the gender gap decreases by nearly 50% if the female beliefs about enjoying coursework in engineering were replaced with those of males. These results are in line with the findings of the previous section. The small contribution of gender differences in ability and future earnings in engineering towards the underlying gender gap in the choice of major allows me to rule out low self-confidence in women and perceived wage discrimination in the labor market as possible explanations for why women are less likely to major in fields like engineering. However, it is not clear what kind of policy would be able to bring about a change in the female beliefs about enjoying coursework and enjoying working at the jobs because these gender differences could be a consequence of innate gender differences in attitudes (Baron-Cohen, 2003), or due to social biases including discrimination (Etzkowitz et al., 1992; Valian, 1998).<sup>40</sup> This issue is pursued in more detail in the following section.

## 7.3 Understanding Beliefs About Enjoying Coursework And Work

In a quest to understand why females are less likely to enjoy studying and working in fields like Engineering, survey respondents were asked their beliefs about each gender being treated poorly at the jobs that would be available in the different major categories. The question was worded as follows<sup>41</sup>:

*"What do you think is the percent chance that X (where X = {Male, Female}) would be treated*

<sup>39</sup>I sort the female and male sub-samples according to the predicted probability of majoring in that field, and then replace the female subjective belief about ability with that of the corresponding male. Since there are more females than males, I use a simulation method similar to the one used for the Fairlie decomposition.

<sup>40</sup>An example of the latter is that women might believe that these fields are not gender-neutral but constructed in accordance with the traditional male role, and that they would be treated *poorly* in the workplace. For example, Traweek (1988) argues that an aggressive behavior is a necessary ingredient for achieving success in science, and Niederle et al. (2007) show that women tend to shy away from competitive environments. In that case, even if women perceive no gender difference in ability and compensation, their beliefs about how much they will enjoy studying engineering and science will be affected.

<sup>41</sup>This question was asked in the follow-up survey (Zafar, 2008) which was taken by 117 of the 161 original survey-takers.

*poorly in jobs that are available in each of the following fields?"*

Before analyzing the responses to this question, columns (1) and (2) of Table 21 report the fraction of females that survey respondents believe take classes in the various majors. Column (3) reports the average number of females who graduated in the various majors in 2005 and 2006 (source: IPEDS 2005 and IPEDS 2006). Survey respondents seem to be well-informed about the relative fraction of females in the various majors. The responses to the question about males and females being treated poorly are shown in columns (4)-(7) of Table 21. Several notable patterns stand out. One, male respondents believe that females are treated more poorly than males in jobs in all fields except Education, Literature & Fine Arts, and Music Studies; these three fields correspond to the three most female-dominated fields (in college) as reported by males in column (1) of the table. Second, females believe that they would be treated more poorly than males at jobs in all fields except Education- the field that females believe has the highest fraction of females. Third, for both the male and female respondents, the largest difference in females being treated poorly relative to males is for Engineering and Math & Computer Sciences- two categories with the lowest fraction of females (as reported by both males and females). Finally, both males and females believe that Education is the category in which males would be treated the worst.

Table 22 shows a significant correlation of -0.35 between the beliefs of females being treated poorly at the jobs and the fraction of females in the classes in the major. The table also shows some interesting correlation patterns between the variables described in Table 21 and beliefs about enjoying coursework and enjoying working at jobs for females and males separately. These correlations indicate that, for female students in particular, beliefs of enjoying coursework and enjoying working at the jobs are positively related to beliefs about the fraction of females taking classes in that field, and negatively correlated with the perceptions of females being treated poorly in the jobs.<sup>42</sup> Interpreting these correlations is not straightforward: it could be that females prefer fields that value female-specific attributes and where females are treated more favorably (Cejka and Eagly, 1999, find that occupations that are female-dominated are those where female-specific attributes are perceived to be essential for success), or it could be that females are treated more favorably at those jobs precisely because those are "female" occupations. Unfortunately, with the available data, it's not possible to choose between these competing causal explanations.

I re-estimate the single-major choice model initially estimated in section 5.1 to see how the inclusion of the new variables "*females treated poorly at the jobs*" and "*males treated poorly at the jobs*" affect the parameter estimates. The estimates for the determinants that were already included in the initial model stay almost the same, while the new variables are insignificant.<sup>43</sup> Moreover, the new variables do not improve the explanatory power of the model for the entire sample and for females; relative to the initial model, the Wald  $\chi^2$  (a measure of goodness-of-fit which compares the likelihood ratio chi-squared of the model to one with the null model) does not change by much.

---

<sup>42</sup>Individuals were asked to explain what "being treated poorly" meant to them. See section 9.2.3 for selective comments.

<sup>43</sup>Since beliefs of males/ females being treated poorly at the jobs are strongly correlated with beliefs of enjoying coursework and enjoying working at the jobs (Table 22), it seems that their impact on the choice is already being captured indirectly. Indeed the variable "females treated poorly at the jobs" only shows up significantly (at the 1% and 10% level respectively) for females and the entire sample in a model that excludes both enjoying coursework and enjoying working at the jobs. Model estimates are available upon request.

## 8 Conclusion

Choosing a college major is a decision that has significant social and economic consequences. Little is known about how youth choose college majors and why the observed gender gap exists. In this paper, I estimate a model of college major choice with a focus on explaining the gender gap. Gender differences in major choice are extremely complex, and no simple explanation can be provided for them. The analysis presented in this paper attempts to enhance our understanding of these issues.

On the methodology side, this paper shows that elicited expectations can be used to relax strong and often nonverifiable assumptions on expectations to infer decision rules under uncertainty. Descriptive analysis of the subjective data shows substantial heterogeneity in beliefs both within and between genders. My approach also differs from the literature on major choice by accounting for both the pecuniary and non-pecuniary determinants of the choice. I have shown that elicited subjective data can be used to infer decision rules in environments where expectations are crucial. This is particularly relevant in cases where the goal is to explain group differences in choices under uncertainty, and where expectations may differ across groups (in unknown ways).

I estimate models for single major and double major choice. Outcomes most important in choice of major are enjoying coursework, approval of parents, and enjoying work at jobs. Non-pecuniary determinants explain about half of the choice for males, and more than three-fourths of the choice for females. Males and females have similar preferences regarding choices at college, but differ in their tastes regarding the workplace; females mostly care about non-pecuniary outcomes (approval of parents, and enjoying work at jobs), while males value pecuniary outcomes (social status of the jobs, likelihood of finding a job, and earnings profiles at jobs) more. In addition, I find that students choosing double majors hedge their chances of getting a job upon graduation and completing their studies by choosing pairs of majors which differ in these two outcomes. Cultural proxies and demographic variables bias beliefs and preferences in systematic ways. Individuals with foreign-born parents value the pecuniary determinants of the choice more than individuals with US-born parents. Males with foreign-born parents are the only sub-group in my sample who value pecuniary determinants more than the non-pecuniary outcomes.

The analysis in this paper has some limitations. First, the study is based on data from Northwestern University only. The heterogeneity in subjective expectations underscores the need to elicit similar data at different undergraduate institutions, and at a larger scale in order to make policy recommendations. Second, heterogeneity in subjective responses could be driven by differential access to information, or by different information processing. Progress in understanding how people form and update expectations requires richer longitudinal data. Moreover, as Manski (2004) argues, understanding expectations formation will also require intensive probing of individuals to learn how they perceive environments and how they process new information. Third, individuals may find it optimal to experiment with different majors to learn about their ability and match quality (Manski, 1989; Altonji, 1993; and Malamud, 2006). This study does not focus on this aspect by assuming that individuals maximize current expected utility.

My results shed some light on the reasons for the gender gap in college major choice. Gender differences in beliefs about ability and future earnings are insignificant in explaining the gender gap. A policy intervention which were to raise the expectations of females about ability and future earnings in engineering to the same level as that of males would only decrease the gender gap by about 15%. This has two implications: (1) just

raising expectations of women may not be enough to eradicate the gap, and (2) hypotheses which claim that the gap could be explained by women having low self-esteem and being less overconfident than men can be rejected by my data. Most of the gender gap is due to gender differences in beliefs about enjoying coursework, and preferences for various outcomes. Gender differences in beliefs about enjoying coursework as well as preferences may exist because of differences in tastes, or due to gender discrimination. Richer data is needed to answer this question. I believe the next natural step is to explore how individuals form their beliefs.

## References

- [1] Alesina, A. and P. Giuliano (2007), "The Power of the Family", Harvard University Working Paper
- [2] Altonji, J. (1993), "The Demand for and Return to Education when Education Outcomes are Uncertain", *Journal of Labor Economics* 11: 48-83
- [3] Altonji, J. and R. Blank (1999), "Race and Gender in the Labor Market", in O. Ashenfelter and D. Card, (eds), *Handbook of Labor Economics Volume 3c* Elsevier Science B.V.: 3144-3259
- [4] Arcidiacono, P. (2004), "Ability Sorting and the Returns to College Major", *Journal of Econometrics* 121(1-2): 343-375
- [5] American Association of University Women (AAUW) Educational Foundation (1998), "Separated by Sex: A Critical Look at Single-Sex Education for Girls", Washington, DC: AAUW
- [6] Bamberger, G. (1986), "Occupation Choice: The Role of Undergraduate Education", Unpublished paper. Chicago: University of Chicago, Graduate School of Business
- [7] Baron-Cohen, S. (2003), "The Essential Difference: Men, Women, and the Extreme Male Brain", Allen Lane, London
- [8] Barres, B. (2006), "Does Gender Matter?", *Nature* 442: 133-136
- [9] Bettinger, E. and B. Long (2005) "Do Faculty Serve as Role Models? The Impact of Instructor Gender on Female Students," *American Economic Review Papers and Proceedings*, 95(2): 152-157
- [10] Betts, J. and D. Morell (1999), "The Determinants of Undergraduate Grade Point Average: The Relative Importance of Family Background, High School Resources, and Peer Group Effects", *Journal of Human Resources* 34(2): 268-293
- [11] Bisin, A. and T. Verdier (2001), "The Economics of Cultural Transmission and the Dynamics of Preference", *Journal of Economic Theory*, 97: 298-319
- [12] Breen R. and C. Garcia-Penalosa (2002), "Bayesian Learning and Gender Segregation", *Journal of Labor Economics*, 20(4): 899-922
- [13] Brown, C. and M. Corcoran (1997), "Sex Based Differences in School Content and the Male-Female Wage Gap", *Quarterly Journal of Economics*, 99: 31-44

- [14] Brownstone, D. and K. Train (1999), "Forecasting New Product Penetration with Flexible Substitution Patterns", *Journal of Econometrics*, 89: 109-129
- [15] Bruine de Bruin, W., B. Fischhoff, B. Halpern-Felsher, and S. Millstein (2000), "Expressing Epistemic Uncertainty: It's a Fifty-Fifty Chance", *Organizational Behavior and Human Decision Processes*, 81: 115-131
- [16] Canes, B. and H. Rosen (1995), "Following in Her Footsteps? Women's Choice of College Majors and Faculty Gender Composition", *Industrial and Labor Relations Review*, 48(3): 482-504
- [17] Chemers, M., L. Hu, and B. Garcia (2001), "Academic Self-Efficacy and First-Year College Student Performance and Adjustment," *Journal of Educational Psychology*, XLIII: 55-64
- [18] Croson, R. and U. Gneezy (2008), "Gender Differences in Preferences", *Journal of Economic Literature*, forthcoming
- [19] Cross, S., and L. Madson (1997), "Models of the Self: Self-Construals and Gender", *Psychological Bulletin*, 122: 5-37
- [20] De Giorgi, G., M. Pellizzari, and S. Redaelli (2007), "Be as Careful of the Books you Read as of the company you keep. Evidence on peer effects in educational choices", IGER-Bocconi and IZA Working Paper.
- [21] Delavande, A. (2004), "Pill, Patch or Shot? Subjective Expectations and Birth Control Choice", Northwestern Working Paper
- [22] Dey, J. and C. Hill (2007), "Behind the Gender Gap", American Association of University Women Educational Foundation
- [23] Dominitz J. and C. Manski (1996), "Eliciting Student Expectations of the Returns to Schooling", *Journal of Human Resources*, 31(1): 1-26
- [24] Dominitz J. and C. Manski (1997), "Using Expectations Data to Study Subjective Income Distribution", *Journal of the American Statistical Association*, 87: 855-867
- [25] Easterlin, R. (1995), "Preferences and Prices in Choice of Career: The Switch to Business, 1972-87", *Journal of Economics Behavior and Organization*, 27: 1-34
- [26] Eagly, A. (1978), "Sex Differences in Influenceability", *Psychological Bulletin*, 85: 86-116
- [27] Eckel, C. and P. Grossman (2002), "Differences in the Economic Decisions of Men and Women: Experimental Evidence," in *Handbook of Experimental Economics Results*, C. Plott, and V. Smith, eds.
- [28] Eide, E. and J. Grogger (1995), "Changes in College Skills and the Rise in the College Wage Premium", *Journal of Human Resources* 30: 280-310
- [29] Etzkowitz, H., C. Kemelgor, M. Neuschatz, and B. Uzzi (1992), "Athena Unbound: Barriers to Women in Academic Science and Engineering," *Science and Public Policy*, 19: 157-179
- [30] Fairlie, R. (1999), "The Absence of the African-American Owned Business: An Analysis of the Dynamics of Self-Employment", *Journal of Labor Economics*, 17(1): 80-108

- [31] Fairlie, R. (2005), "An Extension of the Blinder-Oaxaca Decomposition Technique to Logit and Probit Models", *Journal of Economic and Social Measurement*, 30(4): 305-316
- [32] Fernandez, R., A. Fogli, and C. Olivetti (2004), "Mothers and Sons: Preference Formation and Female Labor Force Dynamics", *Quarterly Journal of Economics*, 119(4): 249-299
- [33] Fernandez, R. (2007), "Culture and Economics", *New Palgrave Dictionary of Economics*, 2nd edition, forthcoming
- [34] Fiorito, J., and R. Dauffenbach. (1982) "Market and Nonmarket Influences in Curriculum Choice by College Students", *Industrial and Labor Relations Review*, 36(1): 88-101
- [35] Fischhoff, B., A. Parker, W. de Bruin, J. Downs, C. Palmgren, R. Dawes, and C. Manski (2000), "Teen Expectations for Significant Life Events", *Public Opinion Quarterly*, 64(2): 189-205
- [36] Freeman, R. (1971) "The Market for College-Trained Manpower", Cambridge, Mass.: Harvard University Press
- [37] Garman, D. and L. Loury (1995), "College Selectivity and Earnings", *Journal of Labor Economics*, 13: 289-308
- [38] Goldin, C., L. Katz, and I. Kuziemko (2006), "The Homecoming of American College Women: The Reversal of the College Gender Gap", *Journal of Economic Perspectives*, 20(4): 133-156
- [39] Gruber, J. (2004), "Is Making Divorce Easier Bad for Children? The Long-Run Implications of Unilateral Divorce", *Journal of Labor Economics*, 22(4): 799-833
- [40] Guiso, L., P. Sapienza, and L. Zingales (2006), "Does Culture Affect Economic Outcomes?", *The Journal of Economic Perspectives*, 20(2): 23-48
- [41] Kimura, D. (1999), "Sex and Cognition", Cambridge, Mass: A Bradford Book, MIT Press
- [42] Kimura, D. (2006), "'Under-representation" or Misrepresentation?", in *Why Aren't More Women In Science?* S.J. Ceci & W. Williams (Eds.), APA Books: 39-46
- [43] Long, V. (1986), "Relationship of Masculinity to Self-Esteem and Self-Acceptance in Female Professionals, College Students, Clients, and Victims of Domestic Violence", *Journal of Consulting and Clinical Psychology*, 54: 323-327
- [44] Lubinski, D., and C. Benbow (1992), "Gender differences in abilities and preferences among the gifted: Implications for the math/science pipeline", *Current Directions in Psychological Science*, 1: 61-66
- [45] Luce, D. and P. Suppes (1965), "Preferences, utility and subjective probability", in R. Luce, R. Bush, and E. Galanter, eds., *Handbook of Mathematical Psychology*, John Wiley and Sons, New York: 249-410
- [46] Malamud, O. (2006), "Breadth vs. Depth: The Effect of Academic Specialization on Labor Market Outcomes", *Harris School Working Paper Series 05:17*
- [47] Manski, C. (1989), "Schooling as Experimentation: a Reappraisal of the Postsecondary Dropout Phenomenon", *Economics of Education Review*, 8(4): 305-312

- [48] Manski, C. (1993a), "Adolescent Econometricians: How Do Youth Infer the Returns to Schooling?" in C. Clotfelter and M. Rothschild (editors) *Studies of Supply and Demand in Higher Education*, Chicago: University of Chicago Press: 43-57
- [49] Manski, C. (1993b), "Identification of Endogenous Social Effects: The Reflection Problem", *Review of Economic Studies*, 60: 531-542
- [50] Manski, C. (2002), "Identification of Decision Rules in Experiments on Simple Games of Proposal and Response", *European Economic Review*, 46(4-5): 880-891
- [51] Manski, C. (2004), "Measuring Expectations", *Econometrica*, 72(5): 1329-1376
- [52] Niederle, M. and L. Vesterlund (2007), "Do Women Shy away from Competition? Do Men Compete too Much?", *Quarterly Journal of Economics*, 122(3): 1067-1101
- [53] Oaxaca, R. (1973), "Male-Female Wage Differential in the Urban Labor Markets", *International Economic Review*, 14(October): 693-709
- [54] Paglin, M. and A. Rufolo (1990), "Heterogeneous Human Capital, Occupational Choice, and Male-Female Earnings Differences", *Journal of Labor Economics*, 8: 123-144
- [55] Polachek, S. (1978), "Sex Differences in College Major", *Industrial and Labor Relations Review*, 31(4): 498-508
- [56] Revelt, D. and K. Train (1998), "Mixed Logit with Repeated Choices: Households' Choices of Appliance Efficiency Level", *The Review of Economics and Statistics*, LXXX (4): 647-657
- [57] Sacerdote, B. (2001), "Peer Effects With Random Assignment: Results for Dartmouth Roommates", *Quarterly Journal of Economics*, 116 (May): 681-704
- [58] Smith, H. and B. Powell (1990), "Great Expectations: Variations in Income Expectations Among College Seniors", *Sociology of Education*, 63 (July): 194-207
- [59] Solnick, S. (1995), "Changes in Women's College Majors from Entrance to Graduation at Women's and Coeducational Colleges", *Industrial and Labor Relations Review*, 48(3): 505-514
- [60] Traweek, S. (1988), "Beamtimes and Lifetimes: The World of High Energy Physicists", Cambridge, MA: Harvard University Press
- [61] Turner, S. and W. Bowen (1999), "Choice of Major: The Changing (Unchanging) Gender Gap", *Industrial and Labor Relations Review* 52(2): 289-313
- [62] Valian, V. (1998), "Why So Slow?", MIT Press, Cambridge.
- [63] Van der Klaauw, W. (2000), "On the Use of Expectations Data in Estimating Structural Dynamic Models: An Analysis of Career Choices", UNC-Chapel Hill

- [64] Vincent, P., L. Peplau, and C. Hill (1998), "A Longitudinal Application of the Theory of Reasoned Action to Women's Career Behavior", *Journal of Applied Social Psychology*, 28: 761-778
- [65] Xie, Y. and K. Shauman (2003), "Women in Science: Career Processes and Outcomes", Cambridge, MA: Harvard University Press
- [66] Yun, M. (2004), "Decomposing Differences in the First Moment", *Economic Letters* 82(2): 275-280
- [67] Zafar, B. (2008), "How do College Students Form and Revise Expectations About Major-Specific Outcomes?", mimeo, Federal Reserve Bank of New York

## 9 Appendix

### 9.1 Survey Excerpt

**Practice Questions** In some of the survey questions, you will be asked about the PERCENT CHANCE of something happening. The percent chance must be a number between zero and 100. Numbers like 2 or 5% indicate "almost no chance," 19% or so may mean "not much chance," a 47 or 55% chance may be a "pretty even chance," 82% or so indicates a "very good chance," and a 95 or 98% mean "almost certain." The percent chance can also be thought of as the NUMBER OF CHANCES OUT OF 100.

We will start with a couple of practice questions.

1. **PRACTICE QUESTION 1: What do you think is the PERCENT CHANCE (or CHANCES OUT OF 100) that you will eat pizza for lunch next week? \_\_\_\_\_%**
2. **PRACTICE QUESTION 2: What do you think is the PERCENT CHANCE (or CHANCES OUT OF 100) that you will eat pizza for lunch on Tuesday next week? \_\_\_\_\_%**

*Once students had answered the questions, they were given the following instructions*

Note that "pizza for lunch next week" INCLUDES the possibility of "pizza for lunch on Tuesday next week". Recall that:

PRACTICE QUESTION 1: What do you think is the PERCENT CHANCE (or CHANCES OUT OF 100) that you will eat pizza for lunch next week?

PRACTICE QUESTION 2: What do you think is the PERCENT CHANCE (or CHANCES OUT OF 100) that you will eat pizza for lunch on Tuesday next week?

Since "pizza for lunch next week" INCLUDES the possibility of "pizza for lunch on Tuesday next week", your answer to PRACTICE QUESTION 2 should be SMALLER or EQUAL than your answer to PRACTICE QUESTION 1.

#### 9.1.1 Questionnaire

The following set of questions was asked for each of the relevant categories. The questions below were asked for Natural Sciences.

Q1 If you were majoring in Natural Sciences, what would be your most likely major?

Q2 If you were majoring in Natural Sciences, what do you think is the percent chance that you will successfully complete this major in 4 years (from the time that you started college)? (Successfully complete means to complete a bachelors)

NOTE: In answering these questions fully place yourself in the (possibly) hypothetical situation. For example, for this question, your answer should be the percent chance that you think you will successfully complete your major in Natural Sciences in 4 years IF you were (FORCED) to major in it.

Q3 If you were majoring in Natural Sciences, what do you think is the percent chance that you will graduate with a GPA of at least 3.5 (on a scale of 4)?

Q4 If you were majoring in Natural Sciences, what do you think is the percent chance that you will enjoy the coursework?

Q5 If you were majoring in Natural Sciences, how many hours per week on average do you think you will need to spend on the coursework?

Q6 If you were majoring in Natural Sciences, what do you think is the percent chance that your parents and other family members would approve of it?

- Q7 If you were majoring in Natural Sciences, what do you think is the percent chance that you could find a job (that you would accept) immediately upon graduation?
- Q8 If you obtained a bachelors in Natural Sciences, what do you think is the percent chance that you will go to graduate school in Natural Sciences some time in the future?
- Q9 What do you think was the average annual starting salary of Northwestern graduates (of 2006) with Bachelor's Degrees in Natural Sciences?

Now look ahead to when you will be 30 YEARS OLD. Think about the kinds of jobs that will be available for you and that you will accept if you successfully graduate in Natural Sciences.

NOTE that there are some jobs that you can get irrespective of what your Field of Study is. For example, one could be a janitor irrespective of their Field of Study. However, one could not get into Medical School (and hence become a doctor) if they were to major in Journalism.

Your answers SHOULD take into account whether you think you would get some kind of advanced degree after your bachelors if you majored in Natural Sciences.

- Q10 What kind of jobs are you thinking of?
- Q11 Look ahead to when you will be 30 YEARS OLD. If you majored in Natural Sciences, what do you think is the percent chance that you will enjoy working at the kinds of jobs that will be available to you?
- Q12 Look ahead to when you will be 30 YEARS OLD. If you majored in Natural Sciences, what do you think is the percent chance that you will be able to reconcile work and your social life/ family at the kinds of jobs that will be available to you?
- Q13 Look ahead to when you will be 30 YEARS OLD. If you majored in Natural Sciences, how many hours per week on average do you think you will need to spend working at the kinds of jobs that will be available to you?

When answering the next two questions, please ignore the effects of price inflation on earnings. That is, assume that one dollar today is worth the same as one dollar when you are 30 years old and when you are 40 years old.

- Q14 Look ahead to when you will be 30 years old. Think about the kinds of jobs that will be available to you and that you will accept if you graduate in Natural Sciences. What is the average amount of money that you think you will earn per year by the time you are 30 YEARS OLD?
- Q15 Now look ahead to when you will be 40 years old. Think about the kinds of jobs that will be available to you and that you will accept if you graduate in Natural Sciences. What is the average amount of money that you think you will earn per year by the time you are 40 YEARS OLD?

## 9.2 Debriefing

### 9.2.1 Why Choose Two Majors

I present some of the responses to the question posed to survey respondents pursuing more than one major: "**Why are you pursuing more than one major?**"

- I am unsure as to what I want to do later in life and would like to open up my options.
- To have more options, since I am not certain as to what career I want to follow
- There are plenty of econ majors in the country, doubling with Math will help me stand out. Also, the complement each other well and I enjoy them both.
- My first major, MMSS, is an adjunct major. Getting a second major allows me to broaden my horizons and also specialize in a practical field. Also, I feel it looks more impressive if you have completed more than one major
- I want to have a science major (chemistry) as well as another route (economics) for careers in life.
- One practical (MMSS) One personal interest (Linguistics). Real goal is to go to law school soon after grad. perhaps working a couple years in the consulting/finance industry
- Because Spanish is for a career and art is for a lifetime hobby.
- Multiple personal interests, having additional options later in life, stand apart from others
- I have a conflict between what is practical for the job prospect and what I truly would enjoy learning about, so I am pursuing one major which falls into each of the two categories.

- There is no single major at Northwestern which encompasses my interests;
- I want to have more fields open to me.
- To make it more easy to get a job and have a solid career
- Keep career opportunities open.
- I feel that having both majors will open up a wider range of job opportunities when I graduate. I also feel that I am interested in both subjects and am taking the opportunity to further my knowledge in them.
- Interest in subject, a more applicable major for attaining business jobs
- The Quarter system at Northwestern makes obtaining a double major very feasible. I have multiple interests so it makes sense for me to pursue multiple majors.
- I want to be a well rounded person after I graduate, and also just in case one of them does not work out.
- Because I enjoy the material, have the time, and feel like it will improve my chances of acquiring a job after I graduate

### 9.2.2 Peer Effects

The question was:

*Check all that apply*

- 1) *My (intended) major is the same as that of one of my parents*
- 2) *My (intended) major is the same as that of one of my siblings*
- 3) *My (intended) major is the same as that of my freshman-year roommate*
- 4) *My (intended) major is the same as that of my current roommate*
- 5) *My (intended) major is the same as that of the majority of my best high school friends who went to college*
- 6) *My (intended) major is the same as that of the majority of my friends in Northwestern*
- 7) *None of the above*

Next the respondent was asked: *"For each of the options (1 through 6) in Question 5 that you have marked, please explain the underlying reason for it"*

Some of the selected responses are:

- I am influenced by my father but not much by friends.
- My Integrated science major is the same as the majority of my friends, because most of the classes that I take is with Integrated science majors. Since we are in class together all the time, we have become good friends.
- My brother is majoring in Journalism but also Political Science. This played a minor influence on my decision but is mostly coincidence that we like the same sort of classes. My freshman year roommate was possibly an influence on me, but we generally had the same interests in terms of school subjects from the start.
- My dad majored in English, is passionate about the subject and is now a college professor who teaches it. He loved it, but it was never forced on me, resulting in that i grew to love it as well. And I'm good at it. When you're constantly being grammatically corrected and pushed to think loftier ideas then it kind of becomes second nature, a permanent habit. As far as my freshman year roommate, i lived in the Communications Residential College. It's 80% journalism and 19% theater. It was bound to happen.
- My brothers and I have very similar interests and strengths.
- My parents have always encouraged me to do well in school, and placed an emphasis on math and the sciences. Also, I live in a town of only 20,000 people, but there are two major research facilities in the town. Many of my peers were also children of scientists. I have a twin brother who also goes to Northwestern and studies Chemistry and German. We probably influenced each other because we're very close. We both took the German AP, which is why both of us have German as a second major (the German major is relatively light, especially if you come in already taking third year classes).
- I am interested in Psychology, and although my parents are not too keen on me studying psychology, that's what I want to to. My mom was also interested in Psych, but she never perused it

- My major is the same as my parents purely by coincidence. Somehow our interests coincide. My major is the same as the majority of my high school friends (but most of my best friends are doing medicine) because most of my high school friends who study abroad chose economics. It is also the major which most students from Hong Kong would choose when they study abroad since most jobs you can find back home is econ-related. My major is the same as the majority of my friends in Northwestern because 1) Economics is a popular major, the probability that you can find an econ major student is quite high 2) I met most of my friends and formed the friendship through classes and extracurricular activities.
- My mom is a psychologist, and even though I have no desire to pursue that career I think she might have influenced my interest in psychology
- I grew up in a household where my parents are both scientists so I became interested in medicine and science simultaneously. They never told me what to do, it was just a matter of spending more time around a certain field. Also, I live on North Campus where a majority of Northwestern science majors and engineers live so it just so happens that many people are in the same field that I intend to be in, primarily by location because the dorms up North are closer to Tech, which is where most of our classes are held.
- 1) Parental Influence 5 and 6) Social Integration with Friends of Similar Background
- For the first, my parents raised me and my siblings, and for the second, I tend to make friends with people I share classes with.
- I think they paired me with a roommate with whom I had stuff in common. My friends at Northwestern and I have the same interests and personalities and that is reflected in our majors.
- My roommate took a Psychology class last year and really enjoyed it. I had never had any exposure to Psychology classes in high school, so decided that it would be interesting to take. I took the class this fall, and really enjoyed it.
- My parents and I have similar tastes and I like the things they like. My roommate and I were best friends from high school and had very similar interests.
- I think I want to major in economics because I see how successful my dad is today and since he majored in business, I thought economics would be close enough.
- economics is something that flows for me when i learn it, maybe it's in my genes since my dad majored in it during graduate school, it's also very practical and covers many bases, so i see why my friends picked it, it's respected, it's not seen as a slacker major like psychology, and i find it very interesting as i would hope many people do since it's such a popular major
- I really think it's a coincidence. My roommate is interested in politics, too. Maybe it's because we're from similar places. We're both from coastal cities, where politics is big.
- My father has influenced me indirectly because he is an economics professor. My brother is young and wants to follow me into business. i am friends with a lot of people in my classes, which happen to be econ./MMSS classes
- My mother is terrible at math so she majored in an all-words major, Sociology, but I am OK at math so my Social Policy major incorporates a bit more economic reasoning and logic than hers

### 9.2.3 Beliefs of Being Treated Poorly

Below I present selective comments by gender to the question: *What does the phrase "treated poorly in jobs available..." mean to you*

#### **Females reported:**

- Women might be subject to jokes in strongly male-dominated fields, but men are more likely to be subject to worse treatment by female coworkers in strongly female-dominated fields. Poor treatment of women by men is much less socially acceptable than the reverse.
- It might mean that they were treated unfairly in terms of pay, or it might mean that the demands of the job didn't allow the individual to pursue his/her home/family life.
- Discriminated against in terms of salary, opportunities, and promotions
- I consider "treated poorly" to signify the chance of some form of gender discrimination present in a field (obviously would differ depending on what major/job from each field of study)
- Openly discriminated upon, and thought to be less capable or incapable of doing the work

- Looked down upon, not given respect, given bad hours, given bad assignments (women are often treated poorly more often than men.)
- Having employers expect less of them and give them less responsibility, or outright discrimination.
- Not as commonly appear in the field as the other sex
- Getting less money; not being socially accepted in that job; not having the same opportunities for promotion
- discrimination. not given the opportunity to do things because of their gender. stereotyped bc the jobs are usually occupied on the other gender.
- Preferred less when in competition with someone equally qualified; made fun of for work; must face large gender imbalance
- It means being treated differently based on gender, as in coworkers' attitudes towards you, (how seriously they take you) or even employer's treatment (like pay difference, or job expectations)
- Glass ceiling; lower expectations

**Males reported:**

- To me this means the employee is not treated fairly or simply does not feel comfortable in the work environment.
- Discrimination in salary and at the workplace, acceptance at workplace, promotion, acceptance into societies and journals, respect.
- Harassment/mistreatment from coworkers and unfair compensation when compared to the opposite sex
- It means that their employers will be biased in some way or another against them because of their gender, and will show it via some negative remark or action.
- Treated with disrespect or made to work very long hours. Or not given a fair chance for promotion.
- If a member of the opposite gender is given a higher position than you when you are more qualified for the position.
- Managers/supervisors have prejudices against the work done by members of a certain gender, judging it unnecessarily harshly.
- Not paid what they're worth, not given ample opportunity for advancement, discriminated against in hiring, not having a job that adequately takes in to account a family life and life outside of work
- To me it means discriminated against based on gender through different means such as interaction, salary, and respect.
- People may think, "She's a woman, she can't solve these types of problems."
- Jobs in which the individual is assumed to be less capable than they really are. Jobs which a small percentage of people think an individual of that gender shouldn't be doing.
- Gender discrimination based on expectations of abilities by gender (like bias against females on engineering and natural sciences)

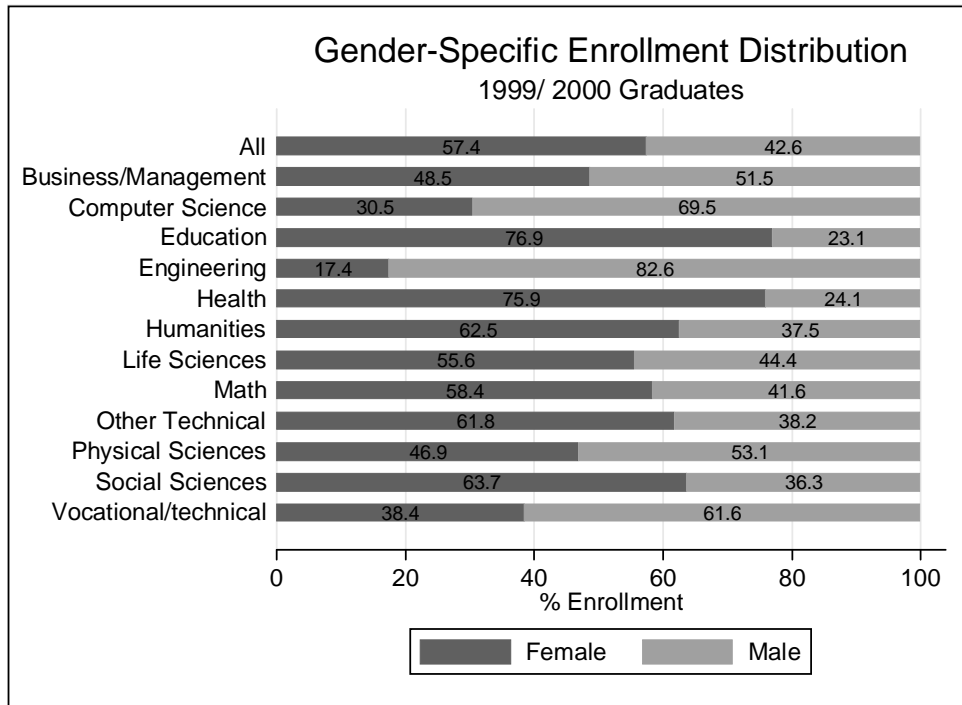


Figure 1: Gender Composition of Undergrad Majors of 1999-2000 Bachelor's Degree Recipients Employed Full-Time in 2001.

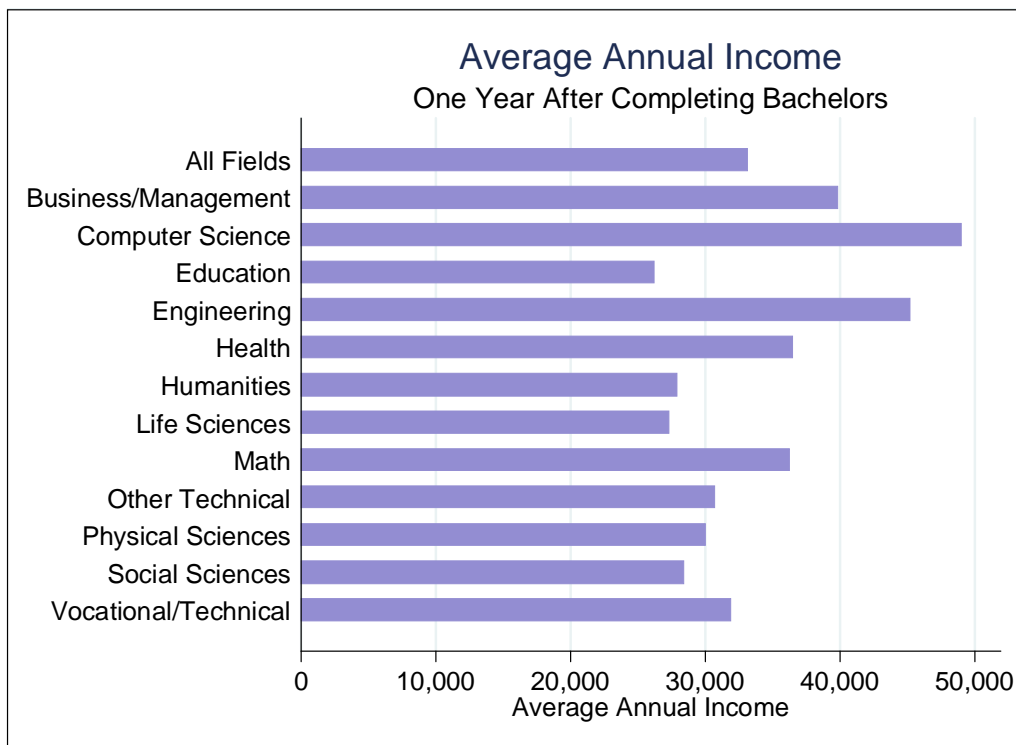


Figure 2: Average income of 1999-2000 Bachelor's Degree Recipients Employed Full-Time in 2001 by Undergraduate Major.

Table 1: List of Majors

The following is the classification of majors into categories:	<u>h Music Studies<sup>1</sup></u>
<u>a Natural Sciences</u>	Jazz Studies
Biological Sciences	Music Cognition
Chemistry	Music Composition
Environmental Sciences	Music Education
Geography*	Music Technology
Geological Sciences	Music Theory
Integrated Science	Musicology
Materials Science	Piano Performance
Physics	String Performance
	Voice and Opera Performance
	Wind and Percussion Performance
<u>b Mathematical and Computer Sciences</u>	<u>i Education and Social Policy<sup>2</sup></u>
Cognitive Science	Human Development and Psychological Services
Computing and Information Systems	Learning and Organizational Change
Mathematics	Secondary Teaching
Statistics	Social Policy
<u>c Social Sciences I</u>	<u>j Communication Studies<sup>3</sup></u>
Anthropology	Communication Studies
Gender Studies*	Dance
History	Human Communication Science
Linguistics	Interdepartmental Studies
Political Science	Performance Studies
Psychology	Radio/Television/ Film
Sociology	Theatre
<u>d Social Sciences II</u>	<u>k Engineering<sup>4</sup></u>
Economics	Applied Mathematics
Mathematical Methods in Social Sciences*	Biomedical Engineering
	Chemical Engineering
<u>e Ethics and Values</u>	Civil Engineering
Legal Studies*	Computer Engineering
Philosophy	Computer Science
Religion	Electrical Engineering
Science in Human Culture*	Environmental Engineering
	Industrial Engineering
<u>f Area Studies</u>	Manufacturing and Design Engineering
African American Studies	Materials Science& Engineering
American Studies	Mechanical Engineering
Asian & Middle East Languages & Civilization	
European Studies	
International Studies*	<u>L Journalism<sup>5</sup></u>
Slavic Languages and Literatures	Journalism
<u>g Literature and Fine Arts</u>	
Art History	
Art Theory and Practice	
Classics	
Comparative Literary Studies	
Drama	
English	
French	
German	
Italian	
Spanish	

\* *Adjunct majors. These do not stand alone*

1 Majors in the School of Music

2 Majors in the School of Education and Social Policy

3 Majors in the School of Communication

4 Majors in the McCormick School of Engineering

5 Majors in the Medill School of Journalism

Table 2: Sample Characteristics

Characteristics	Sample			Population <sup>a</sup>
	All Freq.(Percent)	Single Majors Freq.(Percent)	Double Majors Freq.(Percent)	Freq.(Percent)
<b>Gender</b>				
Male	69 (43)	33 (40)	36 (46)	465 (46)
Female	92 (57)	50 (60)	42 (54)	546 (54)
<b>Total</b>	161	83	78	1011
<b>Ethnicity</b>				
Caucasian	79 (49)	40 (48)	39 (50)	546 (54)
African American	11 (7)	7 (8.5)	4 (5)	71 (7)
Asian	56 (35)	27 (33)	29 (37)	232 (23)
Hispanic	5 (3)	2 (2)	3 (4)	61 (6)
Other	10 (6)	7 (8.5)	3 (4)	101 (10)
<b>Declared Major?<sup>b</sup></b>				
Yes	90 (56)	44 (53)	46 (59)	182 (18)
No	71 (44)	39 (47)	32 (41)	829 (82)
<b>International Std?<sup>c</sup></b>				
Yes	8 (5)	5 (6)	3 (4)	40 (4)
No	153 (95)	78 (94)	75 (96)	971 (96)
<b>Second-Gen Imm?<sup>d</sup></b>				
Yes	66 (41)	33 (40)	33 (42)	—
No	95 (59)	50 (60)	45 (58)	—
<b>Average GPA</b>				
Male	3.48	3.43	3.52	3.26
Female	3.40	3.39	3.45	3.31

<sup>a</sup> Population Statistics for the sophomore class. (Source: Northwestern Office of the Registrar)

<sup>b</sup> Whether the respondent has declared their major at the time of the survey

<sup>c</sup> Whether the respondent is an international student

<sup>d</sup> Whether at least one of the respondent's parents is foreign-born, and respondent is US-born

Table 3: Distribution of WCAS Majors

WCAS Majors <sup>a</sup>	Sample <sup>b</sup>			Class of 2006 <sup>c</sup>		
	All Freq (%)	Males Freq (%)	Females Freq (%)	All Freq (%)	Males Freq (%)	Females Freq (%)
Natural Sciences	31 (19)	15 (22)	16 (17)	156 (14)	62 (12.5)	94 (15.5)
Math & Computer Sci.	4 (2.5)	2 (3)	2 (2)	37 (3.5)	29 (6)	8 (1)
Social Sciences I	41 (25.5)	12 (17)	29 (31.5)	512 (46.5)	211 (42.5)	301 (49)
Social Sciences II	48 (30)	29 (42)	19 (21)	217 (20)	140 (28.5)	77 (13)
Ethics and Values	4 (2.5)	4 (6)	0 (0)	25 (2)	14 (3)	11 (2)
Area Studies	13 (8)	5 (7)	8 (9)	24 (2)	4 (1)	20 (3)
Literature & Fine Arts	20 (12.5)	2 (3)	18 (19.5)	132 (12)	32 (6.5)	100 (16.5)
<b>Total</b>	161 (100)	69 (100)	92 (100)	1103 (100)	492 (100)	611 (100)

<sup>a</sup> Majors that appear in each category are listed in Table 1a

<sup>b</sup> In cases where the survey respondent has more than one major in WCAS, only the first one is included

<sup>c</sup> Only includes students with a primary WCAS major (Source: Integrated Postsecondary Education Data System)

Table 4: Percent Chance of graduating with a GPA of at least 3.5 if majoring in:

Sub. Beliefs	Engineering						Lit. & Fine Arts					
	Males			Females			Males			Females		
	Freq.	Cum. %	%	Freq.	Cum. %	%	Freq.	Cum. %	%	Freq.	Cum. %	%
0	—	0	—	2	2.38	—	1	1.45	—	1	1.09	—
1	1	1.49	—	1	3.57	—	—	1.45	—	—	1.09	—
3	—	1.49	—	2	5.95	—	—	1.45	—	—	1.09	—
5	1	2.99	—	2	8.33	—	—	1.45	—	—	1.09	—
10	3	7.46	—	—	8.33	—	1	2.90	—	—	1.09	—
12	—	7.46	—	1	9.52	—	—	2.90	—	—	1.09	—
15	1	8.96	—	3	13.10	—	—	2.90	—	—	1.09	—
18	—	8.96	—	1	14.29	—	—	2.90	—	—	1.09	—
20	3	13.43	—	12	28.57	—	—	2.90	—	—	1.09	—
25	2	16.42	—	8	38.10	—	—	2.90	—	—	1.09	—
26	1	17.91	—	—	38.10	—	—	2.90	—	—	1.09	—
30	1	19.40	—	4	42.86	—	—	2.90	—	2	3.26	—
33	—	19.40	—	1	44.05	—	—	2.90	—	—	3.26	—
35	2	22.39	—	3	47.62	—	—	2.90	—	1	4.35	—
40	3	26.87	—	7	55.95	—	—	2.90	—	—	4.35	—
45	3	31.34	—	3	59.52	—	1	4.35	—	—	4.35	—
47	1	32.84	—	—	59.52	—	—	4.35	—	—	4.35	—
50	6	41.79	—	8	69.05	—	1	5.80	—	7	11.96	—
55	—	41.79	—	1	70.24	—	—	5.80	—	1	13.04	—
56	—	41.79	—	1	71.43	—	—	5.80	—	—	13.04	—
58	1	43.28	—	—	71.43	—	—	5.80	—	—	13.04	—
60	5	50.75	—	6	78.57	—	5	13.04	—	2	15.22	—
64	—	50.75	—	—	78.57	—	1	14.49	—	—	15.22	—
65	1	52.24	—	2	80.95	—	2	17.39	—	6	21.74	—
66	1	53.73	—	—	80.95	—	—	17.39	—	—	21.74	—
67	—	53.73	—	1	82.14	—	—	17.39	—	—	21.74	—
68	—	53.73	—	—	82.14	—	—	17.39	—	1	22.83	—
70	6	62.69	—	4	86.90	—	8	28.99	—	8	31.52	—
75	3	67.16	—	3	90.48	—	11	44.93	—	5	36.96	—
76	—	67.16	—	—	90.48	—	2	47.83	—	2	39.13	—
79	—	67.16	—	—	90.48	—	1	49.28	—	—	39.13	—
80	5	74.63	—	2	92.86	—	5	56.52	—	14	54.35	—
81	1	76.12	—	—	92.86	—	—	56.52	—	—	54.35	—
82	1	77.61	—	—	92.86	—	—	56.52	—	2	56.52	—
85	4	83.58	—	—	92.86	—	8	68.12	—	6	63.04	—
87	—	83.58	—	—	92.86	—	1	69.57	—	1	64.13	—
88	—	83.58	—	—	92.86	—	3	73.91	—	—	64.13	—
89	—	83.58	—	—	92.86	—	1	75.36	—	1	65.22	—
90	3	88.06	—	2	95.24	—	7	85.51	—	13	79.35	—
93	—	88.06	—	—	95.24	—	—	85.51	—	2	81.52	—
95	4	94.03	—	2	97.62	—	4	91.30	—	5	86.96	—
96	—	94.03	—	—	97.62	—	—	91.30	—	1	88.04	—
97	1	95.52	—	—	97.62	—	—	91.30	—	1	89.13	—
98	1	97.01	—	—	97.62	—	1	92.75	—	5	94.57	—
99	—	97.01	—	—	97.62	—	3	97.10	—	2	96.74	—
100	2	100	—	2	100	—	2	100	—	3	100	—
Total	67			84			69			92		

Table 5: Single Major Choice- Estimation of Homogeneous Preferences

	Using Stated Choice		Using stated Preference	
	(1)	(2)	(3)	(4)
$\Delta u_1$ for graduating within 4 years	6.84*** (1.78)	1.65 (2.93)	-0.447 (0.868)	-1.54* (0.80)
$\Delta u_1$ for graduating within 4 years $\times$ female	-	54.27*** (6.63)	-	3.15** (1.37)
$\Delta u_2$ for graduating with a GPA of at least 3.5	-3.83*** (1.11)	-1.95 (1.94)	0.903* (0.520)	1.13* (0.67)
$\Delta u_2$ for graduating with a GPA of at least 3.5 $\times$ female	-	-8.44** (4.03)	-	0.048 (1.12)
$\Delta u_3$ for enjoying the coursework	13.11*** (2.47)	9.93** (4.36)	2.69*** (0.45)	2.06*** (0.70)
$\Delta u_3$ for enjoying the coursework $\times$ female	-	11.36 (8.39)	-	1.43 (0.946)
$\gamma_1$ for hours/week spent on coursework	-0.058*** (0.017)	-0.057** (0.028)	0.012 (0.011)	0.0064 (0.0135)
$\gamma_1$ for hours/week spent on coursework $\times$ female	-	-0.045 (0.071)	-	0.0189 (0.021)
$\Delta u_4$ for approval of parents and family	3.71*** (1.16)	1.74 (3.14)	1.37** (0.56)	0.98 (0.75)
$\Delta u_4$ for approval of parents and family $\times$ female	-	1.71 (3.98)	-	1.03 (1.13)
$\Delta u_5$ for finding a job upon graduation	2.27* (1.20)	4.01* (2.17)	-0.076 (0.512)	0.279 (0.829)
$\Delta u_5$ for finding a job upon graduation $\times$ female	-	0.74 (4.15)	-	-0.863 (1.04)
$\Delta u_6$ for enjoying work at the available jobs	6.65*** (2.05)	1.80 (3.21)	1.59*** (0.384)	0.468 (0.526)
$\Delta u_6$ for enjoying work at the available jobs $\times$ female	-	18.86*** (7.01)	-	1.80** (0.817)
$\Delta u_7$ for reconciling family and work at available jobs	-1.93* (1.11)	-1.31 (2.77)	0.241 (0.539)	0.258 (0.671)
$\Delta u_7$ for reconciling family and work $\times$ female	-	-2.36 (4.66)	-	0.181 (0.946)
$\gamma_2$ for hours/week spent at work	-0.0066 (0.0166)	0.0282 (0.038)	-0.0080 (0.0099)	-0.015 (0.015)
$\gamma_2$ for hours/week spent at work $\times$ female	-	-0.073 (0.082)	-	0.024 (0.018)
$\gamma_3$ for the social status of the available jobs <sup>a</sup>	3.27*** (1.12)	4.01* (2.28)	1.09*** (0.32)	2.14*** (0.53)
$\gamma_3$ for the social status of the available jobs $\times$ female	-	-0.59 (4.08)	-	-1.696** (0.662)
$\gamma_4$ for expected Income at the age of 30	-5.25e - 7 (4.25e - 6)	9.43e - 6 (7.91e - 6)	6.43e - 7 (1.02e - 6)	1.13e - 6 (2.43e - 6)
$\gamma_4$ for expected Income at the age of 30 $\times$ female	-	-19.1e - 6 (21.8e - 6)	-	-4.40e - 7 (2.53e - 6)
Log-Likelihood	-56.58	-40.77	-733.52	-703.255
No. of Observations	83	83	83	83

\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%; robust standard errors in parentheses  
<sup>a</sup> - social status is on a scale of 1-8 (8 being the highest social status); normalized to be between 0.1-0.8  
all other variables (except income) are probabilities between 0 and 1

Table 6: Decomposition Analysis

	All	Male	Female
	(1)	(2)	(3)
<u>Panel A: Estimates Using Stated Choice Data</u>			
Attributed to:			
Pecuniary Attributes	24.30%	49.00%	33.90%
Non-Pecuniary Attributes	75.70%	51.00%	66.10%
Attributed to:			
Parents' Approval + Enjoying Coursework	44.95%	40.35%	39.95%
Coursework hrs/week + GPA + Graduating in 4 yrs	22.20%	7.10%	22.50%
Finding a job + Job hrs/week + Income at 30 + Status of Job	22.05%	47.00%	13.35%
Reconcile work & family + Enjoying Work	10.80%	5.55%	24.20%
<u>Panel B: Estimates using Stated Preference</u>			
Attributed to:			
Pecuniary Attributes <sup>a</sup>	27.90%	53.80%	18.20%
Non-Pecuniary Attributes <sup>b</sup>	72.10%	46.20%	81.80%
Attributed to:			
Parents' Approval + Enjoying Coursework	43.50%	34.00%	44.00%
Coursework hrs/week + GPA + Graduating in 4 yrs	20.40%	10.30%	11.85%
Finding a job + Job hrs/week + Income at 30 + Status of Job	20.10%	48.30%	16.05%
Reconcile work & family + Enjoying Work	16.00%	7.40%	28.10%

<sup>a</sup> Pecuniary attributes are the following outcomes pooled together: Graduating in 4 years;

Graduating with a GPA of at least 3.5; hrs/week spent on coursework; Finding a job upon graduation; Job hrs/week; Income at 30; Status of the available jobs.

<sup>b</sup> The non-pecuniary attributes include all outcomes not included in *a*

Table 7: Thought Experiments

	Natural Sciences	Math & Comp Sc	Social Sc. I	Social Sc. II	Ethics & Values	Area Studies	Lit. & Fine Arts	Eng.
<u>Baseline Model</u>								
Avg. Male Prob. for:	0.189	0.090	0.171	0.189	0.094	0.082	0.068	0.117
Avg. Female Prob. for:	0.151	0.062	0.226	0.112	0.106	0.140	0.156	0.047
<b>% Change in the probability of majoring if:</b>								
<u>Expt 1: 10% INCREASE in probability of graduating with a GPA of at least 3.5 in Engineering</u>								
Avg. Male Prob. for: <sup>a</sup>	-0.93%	-1.07%	-0.75%	-0.97%	-0.70%	-0.68%	-0.71%	6.46%
Avg. Female Prob. for:	-0.49%	-0.49%	-0.18%	-0.44%	-0.24%	-0.21%	-0.16%	5.86%
<u>Expt 2: 10% DECREASE in probability of graduating with a GPA of at least 3.5 in Literature and Fine Arts</u>								
Avg. Male Prob. for:	0.54%	0.50%	0.66%	0.50%	0.77%	0.70%	-6.56%	0.49%
Avg. Female Prob. for:	0.76%	0.88%	1.35%	1.07%	1.53%	1.61%	-7.89%	0.79%
<u>Expt 3: 10% INCREASE in probability of finding a job after graduating in Social Sciences I</u>								
Avg. Male Prob. for:	-0.27%	-0.23%	1.41%	-0.29%	-0.32%	-0.36%	-0.33%	-0.25%
Avg. Female Prob. for:	0.56%	0.60%	-2.35%	0.57%	0.81%	0.81%	0.79%	0.47%
<u>Expt 4: 10% INCREASE in probability of approval of parents for Social Sciences I</u>								
Avg. Male Prob. for:	-1.12%	-0.99%	6.00%	-1.22%	-1.47%	-1.56%	-1.51%	-1.05%
Avg. Female Prob. for:	-2.52%	-2.72%	10.50%	-2.49%	-3.69%	-3.53%	-3.60%	-2.06%
<u>Expt 5: 10% INCREASE in probability of enjoying coursework in Engineering</u>								
Avg. Male Prob. for:	-1.57%	-1.61%	-1.29%	-1.67%	-1.21%	-1.13%	-1.21%	11.04%
Avg. Female Prob. for:	-1.53%	-1.83%	-0.57%	-1.72%	-0.70%	-0.72%	-0.49%	19.23%
<u>Expt 6: 10% DECREASE in the social status of jobs available after graduating in Social Sciences II</u>								
Avg. Male Prob. for:	2.50%	3.09%	2.86%	-11.50%	2.31%	2.29%	2.47%	3.14%
Avg. Female Prob. for:	0.23%	0.41%	0.24%	-2.14%	0.26%	0.27%	0.23%	0.48%

<sup>a</sup> each cell corresponds to the percent change in the probability of majoring in that category after the intervention relative to the baseline case.

Table 8: Stated reasons for choosing a major

<b>How imp. were the following reasons in choosing a major:<sup>a</sup></b>	<b>Males</b>	<b>Females</b>
My parents wanted me to	6.02 (9.60)	5.33 (11.13)
A mentor/ role model encouraged me to	7.31* (12.00)	4.27 (7.99)
My siblings made the same choice	1.80** (5.45)	0.29 (1.17)
My high school friends and peers made the same choice	1.43 (3.51)	1.21 (5.27)
The societal reputation of the choice	7.75 (10.01)	7.71 (11.74)
To be able to get a high-paying job	14*** (11.80)	7.92 (10.43)
To be able to get a job where I could balance work & family	8.76* (8.92)	6.06 (7.64)
To be able to get a job in a field where people of my gender are not discriminated against	0**	0.80 (2.34)
To get a job that I would enjoy	18.68* (13.73)	23.15 (15.40)
To get training for a specific career	7.24 (9.69)	7.57 (9.19)
To learn more about things that interest me	18.96** (16.23)	25.44 (18.16)
To be able to do well in the coursework of the major	7.05 (7.72)	8.45 (9.48)
Fraction of ppl of my gender teaching classes in the major	0	0.18 (0.89)
Fraction of people of my gender taking classes in the major	0	0.076 (0.62)
Fraction of ppl of my gender in jobs related to the major	0.29 (1.19)	0.15 (0.86)
Other Reasons	0.69 (4.90)	1.36 (6.53)

\* gender diff is significant at 10%; \*\* sig at 5%; \*\*\* sig at 1% (2-tailed T-test)

Standard deviation in parentheses

Each cell is the AVERAGE contribution of the reasons for the choice of majors

<sup>a</sup>The exact question was: "In deciding your major, how important to you was each of the following reasons? For this question you need to assign an integer between 0 and 100 to each of the following reasons. Moreover, the responses SHOULD ALL SUM TO 100."

Table 9: Expected Annual Salary at the Age of 30

Category:	Exp Salary <sup>a</sup>		Variable: Expected salary at the age of 30 in each major category as reported by:											
	Males Fem		Respondent with Major in the category <sup>b</sup>			Respondent with Major not in the Category <sup>c</sup>								
	(1) Avg.	(2) Avg.	(3a) Avg.	(3b) Med.	(3c) N <sup>d</sup>	(4a) Avg.	(4b) Med.	(4c) N	(5a) Avg.	(5b) Med.	(5c) N	(6a) Avg.	(6b) Med.	(6c) N
Natural Sciences	82.33	61.77	101.0	100.0	[15]	97.35	70.00	[17]	96.20	80.00	[54]	87.72	65.00	[75]
Math & Comp Sc	77.70	70.40	66.00	60.00	[5]	41.00	41.00	[2]	79.50	75.00	[64]	71.01	60.00	[90]
Social Sciences I	72.15	58.80	78.75	75.00	[20]	72.22	62.50	[36]	74.76	70.00	[49]	53.30	50.00	[56]
Social Sciences II	83.43	63.00	149.6	100.0	[38]	117.5	85.00	[22]	98.87	95.00	[31]	118.54	75.00	[70]
Ethics and Values	—	—	63.00	65.00	[5]	61.50	61.50	[2]	80.58	60.00	[64]	62.19	55.00	[90]
Area Studies	76.79	53.77	87.50	77.50	[8]	55.86	55.00	[22]	62.26	60.00	[61]	54.57	50.00	[70]
Lit & Fine Arts	76.79	53.77	58.40	50.00	[5]	55.15	50.00	[20]	60.06	50.00	[64]	47.14	45.00	[72]
Music Studies	—	—	60.00	60.00	[1]	36.50	36.50	[2]						
Educ & Soc Policy	65.71	45.91	—	—	—	47.50	47.50	[2]						
Comm Studies	—	—	—	—	—	61.17	65.00	[3]						
Engineering	87.35	78.26	106.3	80.00	[4]	80.00	80.00	[1]	80.00	88.63	[63]	94.76	75.00	[83]
Journalism	—	—	67.50	67.50	[2]	55.00	55.00	[1]						

\* Response to: "Look ahead to when you will be 30 years old. Think about the kinds of jobs that will be available to you and that you will accept if you graduate in [X]. What is the average amount of money that you think you will earn per year by the time you are 30 YEARS OLD?"

The numbers presented are in thousands

<sup>a</sup> Average salary (in 2007 dollars) in 2003 of college graduates of 1993. Restricted to selective colleges with Carnegie Code 4.

Source: U.S. Department of Education National Center for Education Statistics B&B:93/03

<sup>b</sup> Answer to \* when (one of ) his/her intended major is in category X. Col (a) gives the mean, & Col (b) gives the median

<sup>c</sup> Respondent's answer to \* when his/her intended majors are in a category other than X.

<sup>d</sup> Number of respondents

Table 10: Estimation of heterogeneous preferences using Stated Preference

	Entire Sample	Males	Females
	(1)	(2)	(3)
$\Delta u_1$ for graduating within 4 years	-0.545 (0.791)	-0.958 (0.911)	1.20 (1.21)
$\Delta u_2$ for graduating with a GPA of at least 3.5	0.752 (0.575)	0.751 (0.721)	1.01 (1.01)
$\Delta u_3$ for enjoying the coursework	2.92*** (0.466)	2.49*** (0.754)	3.57*** (0.658)
$\gamma_1$ for hours/week spent on coursework <sup>a</sup>	0.0152 (0.011)	0.0098 (0.014)	0.0232 (0.016)
$\Delta u_4$ for parents approv $\times$ parents' _supp <sup>d</sup> $\times$ (1-Foreign <sup>e</sup> )	0.340** (0.150)	0.578*** (0.217)	0.262 (0.194)
$\widetilde{\Delta u}_4$ for parents approval $\times$ parents' _support $\times$ Foreign	0.0439 (0.159)	-0.147 (0.205)	0.601** (0.246)
$\Delta u_5$ for finding a job upon graduation	0.205 (0.494)	0.680 (0.759)	-0.536 (0.637)
$\Delta u_6$ for enjoying work at the available jobs	1.51*** (0.414)	0.319 (0.611)	2.24*** (0.678)
$\Delta u_7$ for reconciling family and work at available jobs	0.246 (0.579)	0.700 (0.747)	0.547 (0.847)
$\widetilde{\Delta u}_7$ for reconciling family & work $\times$ divorced <sup>f</sup>	-0.357 (0.864)	0.494 (1.26)	-0.613 (1.32)
$\gamma_2$ for hours/week spent at work <sup>c</sup>	-0.0097 (0.0100)	-0.0044 (0.016)	0.0045 (0.012)
$\gamma_3$ for social status of the available jobs <sup>b</sup> $\times$ (1-Foreign)	0.310 (0.432)	1.30* (0.76)	0.297 (0.546)
$\widetilde{\gamma}_3$ for social status of jobs $\times$ Foreign	2.28*** (0.550)	3.27*** (0.93)	0.817 (0.580)
$\gamma_4^{HI}$ for exp. Inc at 30 $\times$ (1- low_ inc <sup>g</sup> ) $\times$ (1-Foreign)	2.66e - 06 (2.75e - 06)	3.08e - 06 (2.80e - 06)	17.5e - 06 (12.5e - 06)
$\widetilde{\gamma}_4^{HI}$ for exp Inc at 30 $\times$ (1-low_ income) $\times$ Foreign	-8.16e - 07 (2.33e - 06)	-11.1e - 06 (8.07e - 06)	7.13e - 07 (7.28e - 06)
$\gamma_4^{LI}$ for exp. Income at 30 $\times$ low_ inc $\times$ (1-Foreign)	1.06e - 07 (3.39e - 06)	-3.89e - 06 (3.54e - 06)	1.02e - 06 (2.58e - 06)
$\widetilde{\gamma}_4^{LI}$ for expected Income at 30 $\times$ low_ inc $\times$ Foreign	6.64e - 06 (4.55e - 06)	11.3e - 06** (5.42e - 06)	1.40e - 06 (5.78e - 06)
Log-Likelihood	-726.19	-287.61	-401.68
No. of groups	83	33	50

† Estimates correspond to the estimation of a logit model on stated preference data

\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%; robust standard errors in parentheses

a (b) - number of hours spent per week on coursework (job) varies between 0 and 100;

c - social status is on a scale of 1-8 (8 being the highest social status); normalized to be between 0.1-0.8

all other variables (except income) are probabilities between 0 and 1

d - parents' support = 1 if no education expenses are paid by parents; = 2 if they pay less than \$5,000; = 3 if they pay between \$5,000- \$10,000; = 4 if they pay between \$10,000- \$15,000; = 5 if they pay between \$15,000-\$25,000; = 6 if they pay \$25,000+

e - Foreign is a dummy that equals 1 if either of the respondent's parents is foreign-born.

f - divorced = 1 if respondent's parents are divorced or separated; zero otherwise

g - low\_ income = 1 if parents' annual income is less than \$150,000; zero otherwise

Table 11: Decomposition Analysis

	Foreign-Born		No Foreign-Born	
	Parents		Parents	
	Males	Females	Males	Females
	(1)	(2)	(3)	(4)
Attributed to:				
Pecuniary Attributes	71.40%	35.40%	27.60%	12.20%
Non-Pecuniary Attributes	28.60%	64.60%	72.40%	87.80%
Attributed to:				
Parents' Approval + Enjoying Coursework	25.25%	46.90%	56.55%	51.80%
Coursework hrs/week + GPA + Graduating in 4 yrs	2.85%	15.30%	5.20%	8.20%
Finding a job + Job hrs/week + Income at 30 + Status of Job	65.90%	26.70%	28.95%	11.80%
Reconcile work & family + Enjoying Work	6.00%	21.10%	9.30%	28.20%

*a* Pecuniary attributes are the following outcomes pooled together: Graduating in 4 years; Graduating with a GPA of at least 3.5; hrs/week spent on coursework; Finding a job upon graduation; Job hrs/week; Income at 30; Status of the available jobs.

*b* The non-pecuniary attributes include all outcomes not included in *a*

Table 12: Best Linear Predictor of Expectation of Parent's Approval

Dependent Variable: Expectation of Parent's Approval <sup>†</sup>	Entire Sample		Males		Females	
	Estimates	Std. Error	Estimates	Std. Error	Estimates	Std. Error
	(1)	(2)	(3)	(4)	(5)	(6)
Expectation of: <sup>a</sup>						
Social Status of jobs $\times$ (1- Parents_foreign <sup>b</sup> )	0.084**	(0.035)	0.0611	(0.0622)	0.090**	(0.043)
the status of the jobs $\times$ Parents_foreign	0.188***	(0.047)	0.125*	(0.091)	0.228***	(0.064)
graduating with a GPA of at least 3.5	-0.0466	(0.0467)	-0.003	(0.078)	-0.073	(0.056)
graduating in 4 years	0.0798	(0.067)	0.068	(0.096)	0.069	(0.092)
enjoying coursework	0.0013	(0.0013)	0.00046	(0.0019)	0.0016	(0.0018)
enjoying work at the jobs	0.114***	(0.041)	0.063	(0.0660)	0.145***	(0.053)
finding a job upon graduation	0.289***	(0.067)	0.279**	(0.122)	0.303***	(0.071)
finding a job $\times$ Parents_foreign	0.207**	(0.082)	0.219*	(0.124)	0.202*	(0.110)
Income at 30 (in 10,000s)	0.000023	(0.00112)	0.0023	(0.0035)	-0.0006	(0.0009)
Income at 30 (in 10,000s) $\times$ Low_Income <sup>c</sup>	0.0018	(0.0022)	-0.00082	(0.0048)	0.0028*	(0.0015)
Mother studied given major <sup>d</sup>	0.024	(0.018)	0.051	(0.031)	0.0055	(0.02)
Father studied given major <sup>e</sup>	0.032**	(0.015)	0.0364*	(0.022)	0.024	(0.022)
Studying Given Major <sup>f</sup>	0.0357***	(0.013)	0.021	(0.021)	0.048***	(0.016)
Respondent Fixed-Effects		Yes		Yes		Yes
Major-Specific Dummies		Yes		Yes		Yes
No. of Observations		1287		551		736
No. of Clusters		161		69		92

<sup>†</sup> Dependent variable is a response 0-1 to: "If you were majoring in [X], what do you think is the percent chance that your parents and other family members would approve of it?"

All regressions include major-specific dummies, and respondent fixed effects. (Constants not shown)

Parameter estimates correspond to the estimation of OLS model. Cluster errors in parentheses

\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

*a* Expectations of outcomes except income are between 0 and 1; status is discrete on a scale of 0-0.9

*b* a dummy that equals one if either of the respondent's parents is foreign-born

*c* a dummy that equal one if respondent's parents' annual earnings are less than \$150,000

*d* a dummy that equals one if mother's field of study is the same as the relevant question

*e* a dummy that equals one if father's field of study is the same as the relevant question

*f* a dummy that equals one if the respondent's intended major category is same as category X in the question

Table 13: Mixed Logit Model Estimation

	Without Demographics			With Demographics		
	Est. (Error)	Mean, Dev		Est. (Error)	Mean, Dev	
		(1a)	(1b)		(1c)	(2a)
<b>Fixed Coefficients</b>						
$\Delta u_2$ for graduating within 4 years	-0.12 (0.69)			0.042 (0.73)		
$\gamma_1$ hours/week on coursework	0.021* (0.01)			0.024* (0.01)		
$\Delta u_4$ approval $\times$ parent_sup $\times$ For.	-			-0.51** (0.21)		
$\Delta u_5$ finding a job upon graduation	-0.13 (0.46)			0.067 (0.47)		
$\Delta u_7$ reconcile family & work at job	0.074 (0.49)			0.013 (0.54)		
$\Delta u_7$ reconcile family & work $\times$ Div	-			0.391 (1.01)		
$\gamma_2$ for hours/week spent at work	-0.01 (0.01)			-0.65 (0.77)		
$\gamma_4$ - Expected Income at 30	5e-7 (1.2e-6)			-7e-7 (1.5e-6)		
<b>Random Coefficients</b>						
$\Delta u_2$ graduating with GPA $\geq 3.5$	-0.55 (1.16)	2.31	8.14	-1.01 (1.24)	2.88	16.88
	1.69* (0.87)			2.07*** (0.67)		
$\Delta u_3$ for enjoying the coursework	1.09*** (0.20)	4.01	3.62	1.18*** (0.19)	4.23	3.52
	0.79*** (0.19)			0.74*** (0.18)		
$\Delta u_4$ for approval of parents	0.179 (0.485)	2.11	3.11	-	-	-
	1.075*** (0.37)			-		
$\Delta u_4$ approval $\times$ Parents'_supp	-	-	-	-0.662* (0.383)	0.725	0.730
	0.404 (0.388)			0.835*** (0.287)		
$\Delta u_6$ for enjoying work at jobs	0.912*** (0.396)	2.28	2.64	0.441 (0.387)	2.25	2.38
	-0.493 (0.67)			0.857** (0.408)		
$\gamma_3$ for the social status of jobs	1.59*** (0.545)	2.19	9.18	-1.66 (2.17)	2.25	31.35
	-			2.19** (1.04)		
$\tilde{\gamma}_3$ social status $\times$ Foreign	-	-	-	2.53*** (0.84)		
	-			1.98** (0.88)		
$\tilde{\gamma}_4$ Income at 30 $\times$ Low_Inc	-	-	-	1.57e-6 (2.9e-6)		
	-			2.84e-6 (5.1e-6)		
Log-Likelihood	-698.00			-690.84		
No. of Groups	83			83		
No. of Observations	2957			2957		

*a* Mean and Standard Deviation of the Log-Normally distributed coefficients are calculated at the estimated  $\Delta u$  and  $\sigma$ .  
\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%; standard errors in parentheses

Table 14: Decomposition Analysis for double major respondents using stated preference data

	All	Male	Female
	(1)	(2)	(3)
Attributed to:			
Pecuniary Attributes	24.55%	43.90%	15.90%
Non-Pecuniary Attributes	75.55%	56.10%	84.10%
Attributed to:			
Parents' Approval + Enjoying Coursework	52.25%	49.80%	54.50%
Coursework hrs/week + GPA + Graduating in 4 yrs	14.55%	11.60%	22.00%
Finding a job + Job hrs/week + Income at 30 + Status of Job	24.70%	32.00%	14.50%
Reconcile work & family + Enjoying Work	8.50%	6.60%	9.00%

Table 15: Distribution of Double Majors

First Major	Second Major						
	Natural Sci.	Math & Comp. Sci.	Social Sci. I	Social Sci. II	Ethics & Values	Area Stud.	Lit. & Fine Arts
Natural Sciences	1	—	—	—	—	—	—
Math & Computer Sci.	2	0	—	—	—	—	—
Social Sciences I	2	0	2	—	—	—	—
Social Sciences II	3	1	11	0	—	—	—
Ethics and Values	2	1	0	1	0	—	—
Area Studies	1	0	9	10	1	0	—
Literature and Fine Arts	1	1	3	2	0	5	2
Music Studies	1	0	1	1	0	0	1
Education	0	0	1	1	0	1	0
Communication Studies	1	0	1	1	0	0	0
Engineering	0	0	0	5	0	0	0
Journalism	0	0	1	0	0	0	1
Total	14	3	29	21	1	6	4

Table 16: Double Major Choice Model - Estimation Using Choice Data

Variables	All		Males		Females	
	Estimate (1a)	Error (1b)	Estimate (2a)	Error (2b)	Estimate (3a)	Error (3b)
$\Delta u_{11}$ for graduating within 4 years	-1.81	(2.73)	-5.92	(3.89)	-1.59	(5.00)
$\Delta u_{12}$ for maximum of graduating in 4 years	23.70***	(3.37)	19.77***	(6.53)	24.17***	(5.66)
$\Delta u_2$ for graduating with a GPA of $\geq 3.5$	3.53***	(1.19)	2.27	(1.98)	5.54***	(1.71)
$\Delta u_{31}$ for enjoying the coursework	11.54***	(2.74)	10.54**	(4.75)	12.15**	(4.60)
$\Delta u_{32}$ for maximum of enjoying coursework	3.41	(3.02)	24.85***	(6.42)	-1.38	(4.24)
$\gamma_{11}$ for hours/week spent on coursework	0.155***	(0.043)	0.190	(0.134)	0.157***	(0.061)
$\gamma_{12}$ for min. of hours/week on coursework	-0.176***	(0.047)	-0.101	(0.151)	-0.217***	(0.059)
$\Delta u_{41}$ for approval of parents and family	9.17***	(2.03)	6.70	(4.44)	10.00***	(3.26)
$\Delta u_{42}$ for maximum of approval of parents	1.58	(2.43)	10.02*	(6.14)	0.52	(3.64)
$\Delta u_{51}$ for finding a job upon graduation	-3.31*	(1.78)	-1.84	(3.12)	-4.36	(1.40)
$\Delta u_{52}$ for maximum of finding a job	5.67***	(2.01)	5.46	(3.70)	6.48**	(2.72)
$\Delta u_{61}$ for enjoying work at the jobs	4.51***	(1.13)	6.01**	(2.81)	4.33***	(1.40)
$\Delta u_{71}$ for reconciling family & work at jobs	1.85**	(0.91)	0.151	(2.22)	2.23**	(1.11)
$\gamma_{21}$ for hours/week spent at work	0.047***	(0.015)	0.021	(0.034)	0.033	(0.021)
$\gamma_{31}$ for the social status of the jobs	-0.644	(0.82)	0.935	(2.00)	-1.08	(1.11)
$\gamma_{41}$ for expected income at the age of 30	1.1e-6	(2.8e-6)	1.3e-6	(7.1e-6)	5.7e-7	(3.7e-6)
Log Likelihood	-154.87		-59.66		-86.43	
Number of Respondents	78		36		42	

\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

Table 17: Contributions of various outcomes

Variables	<b>All</b>	<b>Males</b>	<b>Females</b>
	Estimates	Estimates	Estimates
	(1)	(2)	(3)
$\Delta u_{11} + \Delta u_{12}$ (graduating within 4 years)	21.88***	13.84**	22.58***
$\Delta u_2$ (graduating with a GPA of at least 3.5)	3.53***	2.27	5.54***
$\Delta u_{31} + \Delta u_{32}$ (enjoying the coursework)	14.95***	35.39***	10.77***
$\gamma_{11} + \gamma_{12}$ (hours/week spent on coursework)	-0.0200	0.089**	-0.0597*
$\Delta u_{41} + \Delta u_{42}$ (approval of parents and family)	10.75***	16.73***	10.52***
$\Delta u_{51} + \Delta u_{52}$ (finding a job upon graduation)	2.36**	3.62*	2.12*
$\Delta u_{61} + \Delta u_{62}$ (enjoying work at the available jobs)	4.51***	6.01**	4.34***
$\Delta u_{71} + \Delta u_{72}$ (reconciling family and work at the jobs)	1.85**	0.151	2.23**
$\gamma_{21} + \gamma_{22}$ (hours/week spent at work)	0.047***	0.021	0.0326
$\gamma_{31} + \gamma_{32}$ (social status of the available jobs)	-0.644	0.93	-1.08
$\gamma_{41} + \gamma_{42}$ (expected income at the age of 30)	1.09e-06	1.33e-06	5.72e-07

\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

Table 18: Double Major Choice Model with Error Components

Variables	All	
	Estimates	Std. Error
	(1)	(2)
$\Delta u_{11}$ for graduating within 4 years	20.54*	(11.94)
$\Delta u_{12}$ for maximum of graduating within 4 years	28.69*	(16.92)
$\Delta u_2$ for graduating with a GPA of at least 3.5	1.64	(3.33)
$\Delta u_{31}$ for enjoying the coursework	17.73***	(6.30)
$\Delta u_{32}$ for maximum of enjoying the coursework	8.54	(5.52)
$\gamma_{11}$ for hours/week spent on coursework	0.215**	(0.100)
$\gamma_{12}$ for minimum of hours/week on coursework	-0.033	(0.024)
$\Delta u_{41}$ for approval of parents and family	20.24***	(6.57)
$\Delta u_{42}$ for maximum of approval of parents	-3.59	(4.43)
$\Delta u_{51}$ for finding a job upon graduation	-2.16	(3.99)
$\Delta u_{52}$ for maximum of finding a job	4.91*	(2.96)
$\Delta u_{61}$ for enjoying work at the available jobs	6.48*	(3.36)
$\Delta u_{71}$ for reconciling family and work at jobs	6.23	(4.31)
$\gamma_{21}$ for hours/week spent at work	0.075	(0.073)
$\gamma_{31}$ for the social status of the available jobs	1.63	(2.07)
$\gamma_{41}$ for expected income at the age of 30	-1.62e-06	(7.86e-06)
Log Likelihood	-114.46	
Number of Respondents	76	

† Includes Error Components for each major category (estimates not shown)

\* significant at 10%; \*\* significant at 5%; \*\*\* significant at 1%

Table 19: Decomposition Analysis to explain gender differences

	Eng. (1)	Lit & Arts (2)	Soc Sci II (3)	Soc Sci I (4)	Eng. (5)	Lit & Arts (6)	Soc Sci II (7)	Soc Sci I (8)
Avg. Prob for Males	0.1047	0.0681	0.2065	0.1740	0.1047	0.0681	0.2065	0.1740
Avg. Prob for Females	0.0446	0.1524	0.1158	0.2151	0.0446	0.1524	0.1158	0.2151
Gender Diff in Prob.	<b>0.0601</b>	<b>-0.0843</b>	<b>0.0907</b>	<b>-0.0411</b>	<b>0.0601</b>	<b>-0.0843</b>	<b>0.0907</b>	<b>-0.0411</b>
	<b>Contributions from gender diff in beliefs of:</b>				<b>Contributions from gender diff in coeffs of:</b>			
Graduating in 4 years	-0.00053	0.0015	-0.0042***	0.0022	0.0090***	-0.004***	0.0025**	-0.011***
	-0.89% <sup>a</sup>	-1.76%	-4.63%	-5.34%	15.07%	4.89%	2.74%	27.12%
Graduate with GPA <sub>≥</sub> 3.5	0.0028	-0.0028	-0.00084	-0.0052	0.0087***	-0.0062***	0.004***	-0.0082***
	4.63%	3.36%	-0.93%	12.51%	14.48%	7.38%	4.47%	19.95%
Enjoying coursework	0.0161***	-0.043***	0.0282***	-0.036***	0.0081***	-0.007***	0.0028***	-0.010***
	26.71%	51.25%	31.12%	86.69%	13.52%	8.24%	3.11%	24.26%
Hrs/wk on coursework	-0.0019	0.0022	-0.0011	0.0007	0.0012***	-0.00032**	0.0015***	-0.003***
	-3.14%	-2.61%	-1.18%	-1.81%	2.00%	0.37%	1.62%	6.79%
Approval of parents	0.0015**	-0.0050**	0.0059**	0.0027	0.0014***	-0.0034***	0.0024	-0.0007
	2.51%	5.96%	6.47%	-6.44%	2.35%	3.99%	2.68%	1.79%
Finding a job	0.00016	-0.00049	0.00027	0.0004	-0.00012***	0.00018***	-0.00023***	0.00012***
	0.27%	0.58%	0.30%	-0.96%	-0.20%	-0.21%	-0.25%	-0.28%
Enjoying work at jobs	0.0035	-0.0065	0.010	-0.00003	0.0030***	-0.0032***	0.004**	-0.0106***
	5.87%	7.70%	10.91%	0.63%	4.92%	3.77%	4.37%	25.85%
Reconcile family & work	0.0027	-0.0024	0.0037	0.0001	-0.0013**	0.0070***	-0.0050***	0.0046***
	4.55%	2.86%	4.04%	-0.25%	-2.21%	-8.31%	-5.52%	-11.20%
Social status of jobs	-0.0004	0.0026	0.027***	0.019***	0.0083***	-0.0244**	0.0118***	0.0023
	-1.74%	-3.11%	29.98%	-46.44%	13.81%	28.92%	13.05%	-5.63%
Hrs/week at the jobs	0.0014	-0.0007	-0.004	-0.0012	-0.0020***	0.0083***	-0.0084***	0.0066***
	2.29%	0.89%	-4.47%	2.86%	-3.32%	-9.89%	-9.27%	-16.01%
Expected Income at 30	-0.0002	0.0026***	0.006***	0.009**	0.00017	-0.0015***	0.0043	-0.0013*
	-0.27%	-3.11%	6.76%	-20.87%	0.29%	1.81%	4.82%	3.20%
<b>All included variables</b>	0.0251	-0.0523	0.0711	-0.0082	0.0350	-0.0320	0.0196	-0.0328
	<b>41.75%</b>	<b>62.01%</b>	<b>78.36%</b>	<b>20.00%</b>	<b>58.25%</b>	<b>37.99%</b>	<b>21.64%</b>	<b>80.00%</b>

Standard errors (computed by delta method) in parentheses. \*\*\* significant at 1%; \*\* significant at 5%; \* significant at 10%;

<sup>a</sup>. The contribution in % of the relevant variable to the gap

Table 20: Simulations of the Gender Gap under different Environments

Fields of Study	Base <sup>c</sup>	Ability	Income	Enjoying Coursework	Enjoying Work
	(1)	(2)	(3)	(4)	(5)
Engineering	0.0602 <sup>a</sup>	0.0517 13.92% <sup>b</sup>	0.0608 -1.06%	0.0308 48.74%	0.0534 11.18%
Natural Sciences	0.0550	0.0445 18.98%	0.0529 3.88%	0.0229 58.29%	0.0406 26.48%
Math & Computer Sci.	0.0191	0.0135 29.07%	0.0184 3.45%	0.0074 61.41%	0.0083 56.38%
Social Sciences I	-0.0412	-0.0524 -27.28%	-0.0474 -15.32%	-0.0643 -56.25%	-0.0613 -48.84%
Social Sciences II	0.0907	0.0737 18.68%	0.0881 2.88%	0.0272 69.92%	0.0608 32.92%
Ethics & Values	-0.0189	-0.0266 -40.77%	-0.0219 -15.87%	-0.0419 -122.03%	-0.0381 -101.9%
Area Studies	-0.0624	-0.0634 -1.69%	-0.0655 -4.96%	-0.0563 9.87%	-0.0721 -15.48%
Lit. & Fine Arts	-0.0843	-0.0863 -2.35%	-0.0888 -5.35%	-0.0545 35.34%	-0.0777 7.84%

<sup>a</sup> The model predicted gender gap (male prob. - female prob.) under the relevant environment

<sup>b</sup> The % decrease in the gender gap (relative to the baseline case) after the change

<sup>c</sup> The predicted gap under the baseline case, i.e. no intervention

Table 21: Perceptions of Monetary and Non-Monetary Discrimination

Category	% Females in class <sup>a</sup>			Males Poorly <sup>b</sup>		Fems Poorly	
	Males <sup>c</sup>	Fems <sup>d</sup>	Recs <sup>e</sup>	Males	Fems	Males	Fems
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Natural Sciences	40.50	39.22	57.32	8.49	7.27	24.03	22.90
Math & Comp Sci	31.50**	25.22	34.12	7.17	6.53	23.71	29.19
Social Sciences I	56.56*	60.30	61.72	8.55	11.56	12.71	14.75
Social Sciences II	43.13	42.83	34.97	8.45	7.27	19.90	27.72
Ethics and Values	55.39	55.98	39.18	9.56	11.01	12.07	15.71
Area Studies	59.84	58.15	77.27	10.52	9.87	11.07	13.19
Lit & Fine Arts	64.82	66.16	73.11	11.19	11.22	8.94	13.15
Music Studies	59.25	57.22	50.97	13.05	10.63	9.25	13.16
Educ & Social Policy	66.21	68.86	76.24	13.47	16.57	9.82	13.18
Communication Std	58.71	59.72	57.88	10.82	11.90	11.98	15.43
Engineering	30.01	27.28	27.10	6.09	5.34	25.61	30.77
Journalism	58.90	60.22	71.42	10.47	11.69	12.03	16.30

\*\*\* gender diff sig (p-value < 0.01; two-tailed t-test); \*\* diff sig at 5% level; \* diff sig at 10% level

<sup>a</sup> Fraction of students in the major who are females (on a scale of 0-100)

<sup>b</sup> The average belief that males would be treated poorly in the jobs that would be available in each of the specified categories

<sup>c(d)</sup> Response of male (female) survey respondents to the relevant question

<sup>f</sup> Fraction of females amongst graduates with that major in 2005 & 2006 (source: IPEDS)

Table 22: Correlation patterns between various beliefs

	Females treated poorly	Males treated poorly	% Females in class	Enjoy Course- work
	<u>Females</u>			
Females treated poorly	1.00			
Males treated poorly	0.3997***	1.00		
% of Females in the class	-0.3523***	0.2191***	1.00	
Enjoy coursework	-0.2616***	0.0926***	0.4654***	1.00
Enjoy working at jobs	-0.1583***	-0.0205	0.2699***	0.6196***
	<u>Males</u>			
Females treated poorly	1.00			
Males treated poorly	0.5575***	1.00		
% of Females in the class	-0.3584***	0.0692	1.00	
Enjoy coursework	-0.1604***	-0.0483	0.0832*	1.00
Enjoy working at jobs	-0.0741	-0.0915*	0.0293	0.6704***

correlation significant using the Spearman's rank correlation coefficient:

\*\*\* sig at 1%, \*\* sig. at 5%, \* sig. at 10%