College Major Choice and the Gender Gap*

Basit Zafar†

Abstract

Males and females make different choices with regard 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, focusing 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 that 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 gaining the 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: Females care more about the non-pecuniary outcomes in the workplace, while males value the pecuniary outcomes in the workplace more. 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 differences in preferences.

JEL Codes: D8, I2, J1, Z1

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

*I am indebted to Charles Manski, Christopher Taber, and Paola Sapienza for extremely helpful discussions and comments. I also thank Raquel Bernal, Meta Brown, Adeline Delavande, Marianne Hinds, Ben Jones, Hilarie Lieb, Joan Linsenmeier, Carlos Madeira, Ofer Malamud, Steve Ross, Ija Trapeznikova, Wilbert van Der Klaauw, Sergio Urzua, and seminar participants at NBER Higher Education Working Group and Northwestern University 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.

†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, among recipients of bachelor's degrees in the United States, 13% of women majored in education compared to 4% of men, and only 2% of women majored in engineering compared to 12% 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 Turner and Bowen, 1999; Dey and Hill, 2007).

These markedly different choices in college major between males and females have significant economic and social impacts. Figure 2 shows that large earnings premiums exist across majors. For example, in 2000-2001, a year after graduation in the United States, the average education major employed full-time earned only 60% as much as one who majored in engineering (also see 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). However, studies of mathematically gifted individuals reveal differences in choices across gender, even for very talented individuals (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 and/or beliefs 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 and beliefs.

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 motivated by the theoretical model outlined in Altonji (1993), which treats education as a sequential choice made under uncertainty. 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 sequential. However, since I do not have data needed to estimate a dynamic model, I assume that individuals maximize current expected utility, and estimate a static choice model.

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 parameters of the utility function. 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 nonverifiable 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, 1993, 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 relative to the second major. An individual choosing the first major is consistent with two underlying states of the world: (1) she cares only about getting a college degree, or (2) she values only 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 to allow 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. Though Northwestern University is a selective institution, the interest in understanding gender differences in major choice is driven by the under-representation of women in science and engineering, and it is precisely individuals attending elite universities who have a realistic chance of making it to the higher echelons of science and engineering. Therefore, I believe that Northwestern University is the *right* setting to explore these issues.

In contrast to most studies on schooling choices that ignore uncertainty, I estimate a random utility model of college major choice allowing for heterogeneity in beliefs.[1] My approach also differs from the existing literature by accounting for the non-pecuniary aspects of the choice. Though the importance of non-price determinants in the choice of majors has been highlighted in a few studies (Fiorito and Dauffenbach, 1982; Easterlin, 1995; Weinberger, 2004), no study has jointly modeled the pecuniary and non-pecuniary determinants of the choice. The approach in this paper 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.

Estimation of the choice model reveals that the most important outcomes in the choice of major are enjoying coursework, enjoying work at potential jobs, and gaining the approval of parents. Non-pecuniary outcomes explain about half of the choice behavior for males and more than three-fourths of the choice for

[1] 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.

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 and females about the non-pecuniary outcomes. I also present evidence that cultural proxies bias preferences in favor of certain outcomes (see Guiso et al., 2006; Fernandez et al., 2004). Individuals with foreign-born parents value the pecuniary aspects of the choice more. In particular, males with foreign-born parents are the only sub-group in my sample for whom pecuniary outcomes explain more than 50% of the choice.

On the methodology side, this paper 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. Studies have shown that subjective data tend to be good predictors of behavior (Euwals et al., 1998; Hurd et al., 2004). Recently, expectations data have been employed to estimate decision models: Wolpin (1999), van der Klaauw (2000), and van der Klaauw and Wolpin (2008) show that incorporating subjective expectations data in choice models improve the precision of the parameter estimates. Delavande (2008) collects subjective data to estimate a model of birth control choice for women. The choice model used in this paper is motivated by her framework. This paper contributes to 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.

Finally, this paper is related to the literature that focuses on the underlying reasons for the gender gap in science and engineering. For policy interventions, an important question is whether gender differences in choices are driven by differences in preferences or in beliefs. Existing studies on schooling choices have either focused on gender differences in preferences (Daymont and Andrisani, 1984), or gender differences in beliefs (Valian, 1998; Weinberger, 2004), but not both. The framework developed in this paper makes a clear distinction between preferences and beliefs. This allows me to 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 and Vesterlund, 2007) can be rejected in my data. Second, the majority of the gender gap in majors that I consider can be explained by gender differences in beliefs about enjoying studying different fields, and differences in preferences. 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 14% only (opposed to a reduction in the gap of about 50% if the simulation is done on beliefs about enjoying coursework). 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 under-confidence with regard to academic ability may not be the main reasons 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 choice model. Section 6 undertakes a decomposition technique to understand the sources of gender differences in choice of major. Finally, Section 7 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 $\mathbf{a}$ that are realized in college, a vector of outcomes $\mathbf{c}$ that are realized after graduating from college, and individual characteristics X_{it}. Examples of outcomes in $\mathbf{a}$ include graduating within four years, enjoying the coursework, and gaining 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 $k \in C_i$.[2] If an individual chooses major m, then standard revealed preference argument (assuming that indifference between alternatives occurs with zero probability) implies that:

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

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 reasons to think that prevailing expectations assumptions are correct. However, not only has the information-processing rule varied considerably among studies of schooling behavior, but most assume that individuals form their expectations in the same way.[3] 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 the same actions 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 that do not exist otherwise).

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

[2] 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. I use the terms "beliefs" and "expectations" interchangeably; both refer to beliefs about outcomes that will only be realized in the future.

[3] For example, Freeman (1971) assumes 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 several 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 the 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.

3 Data

To estimate the model of college major choice, one needs to elicit the subjective beliefs about the outcomes associated with a major, $P_{ikt}(\mathbf{a}, \mathbf{c})$, for each major ($\forall k \in C_i$) in individual i's choice set. Since the range of majors available to students and institutional details vary considerably across colleges, one standard survey cannot be used to collect data in different settings. Therefore, in this paper, as a first step towards understanding how college majors are chosen and what explains the underlying gender differences, I focus on Northwestern University. I collect data on 161 Northwestern University students. This section describes the institutional details at Northwestern, the data collection method, and the nature of the subjective data.

3.1 Institutional Details

For the purposes of this study, I focus on students who are in the process of choosing a major but have not necessarily chosen one. There are several reasons for this criteria: Students who are in the process of choosing a major are actively thinking about the occurrence of outcomes associated with the major, and hence their responses to subjective questions related to the choice of major are more likely to be meaningful. Second, interviewing students who have already chosen their major raises the issue of cognitive dissonance (Festinger, 1957). More specifically, students who have already chosen their major could rationalize their choice of major by devaluing their beliefs for outcomes associated with the majors they considered but rejected, and upgrading their beliefs for outcomes associated with the major that they chose. This systematic measurement error in elicited subjective beliefs would be problematic, and plugging in such beliefs in equation (1) would result in biased estimates of the preference parameters. Northwestern University requires students to declare their major by the end of their sophomore year. Surveying juniors and seniors would exacerbate issues arising from cognitive dissonance. On the other hand, freshmen may have little idea of what major they want to pursue when they first arrive in college, and may not have thought about the likelihood of the various outcomes conditional on the choice of major. Therefore, in order to minimize the above-mentioned biases, I restrict my sample to Northwestern University sophomores.

The study is further restricted to schools at Northwestern University that accord students flexibility in choosing a major. For example, a student in the School of Journalism has to declare her major at the time of admission and can change her major only by a special request to the school. For such a student, the choice of college and major is jointly determined. Since, I model the choice of major conditional on deciding to attend Northwestern University, such students are not eligible for the study. I further assume that the choice set for an individual is exogenous. This eliminates students in smaller schools at Northwestern since this assumption would have to be relaxed for them. 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.[4]

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 (i.e., for the major that the individual is pursuing, as well as for all the other majors in the individual'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.[5]

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; this was done precisely to elicit subjective beliefs about the outcomes associated with majoring in Engineering. Therefore, any student with a single major is assumed to have 8 categories in her choice set.

[4] 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.

Though I don't have any students in my sample with a sole major in the School of Engineering (since that choice is made jointly with that of what college to attend), I have students who have majors in both WCAS and the School of Engineering. From the perspective of trying to understand the mechanisms that drive gender differences in college major choice, it is particularly the preferences and beliefs of individuals *not* making the choice that is of interest to us.

[5] Kellogg School of Management now runs an undergraduate certificate program in Finance for undergraduates. This program was only instituted after this study was conducted, and, therefore, was not one of the choices. Students (from later cohorts) who enroll in this certificate program would most likely have majored in the category of Social Sciences II otherwise.

"Dropping out of college" is also not an element of the choice set. This is primarily because drop-out rates are very low: 93% of Northwestern University undergraduate students graduate with a degree within five years of first enrolling.

For an individual with a second 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. 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. 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*.

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.[6] 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 students need not have declared their majors to participate in the study. The survey was conducted from November 2006 to February 2007, which corresponds to the first half of the students' sophomore year. 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 overrepresented in my sample; (2) 56% of the survey-takers had declared their major, whereas the corresponding number for the sophomore population was 47%. However, the population statistic was for the end of the Fall quarter of the sophomore year, while the data collection spanned two quarters (Fall and Winter). Since students may declare their majors at any time during the academic year, these two numbers were most likely very similar by the end of the data collection period; and (3) it seems that survey-takers,

[6] 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).

especially male students, have higher GPAs than their population counterparts. However, since the focus of the paper is on gender differences, and both males and females in my sample, on average, have higher GPAs, this should not bias the results in any obvious way.[7]

Table 3 presents the distribution of WCAS majors in the sample. The major distribution for the graduating class of 2006 is also presented in the last 3 columns of the table (this is the most recent year for which data are available). There is no reason to believe that the distribution of majors would stay the same over time: In fact, Social Sciences II (which includes Economics) has been becoming more popular amongst both males and females for the last several years. The only purpose of this table is to present trends in college majors by gender. 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 both in my sample and in 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 three 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 to identify 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 year's time. If the respondents agreed to be contacted for the follow-up, they were asked for their names and contact information. Of the 161 respondents, 156 (97%) agreed to the follow-up.

[7] There might a be a concern that this sampling strategy would yield a selected sample. I don't find much evidence of this based on observables. Moreover, since gender differences are the focus of the paper, results would only be biased if one believes that factors that lead students to take the survey differ between males and females, of which there is no obvious evidence. To the extent that Asians are over-represented in the sample, all the analysis in the paper is robust to the exclusion of this group. Another concern could be that survey-takers might be motivated by pecuniary incentives to take the survey. Given that I find that non-pecuniary outcomes explain a majority of the choice, this should only bias the magnitude of the importance of non-pecuniary outcomes downward.

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 four 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 per 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 per 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 that allows a flexible lifestyle (c_3, c_4) or provides 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

captures non-pecuniary aspects in college; and c_2 and c_3 capture non-pecuniary aspects 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 each major in the individual's choice set, 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. An advantage of asking probabilistic questions relative to approaches that employ a Likert-scale or a simple binary response (yes/no; true/false) is that responses are interpersonally comparable, more informative, and allow the respondent to express uncertainty (Juster, 1966; Manski, 2004).[8] As is standard in studies that collect subjective data, a short introduction (similar to the one in Delavande, 2008), was read and handed to the respondents at the start of the survey. 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. Here, I present some of the questions that elicited the subjective expectations. For example, the belief for the binary outcome a_2 was elicited 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)?*

The question eliciting the expected number of hours per 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.*[9]

[8] Studies that examine the role of non-pecuniary influences in the choice of schooling, Fiorito and Dauffenbach (1982), Daymont and Andrisani, (1984), Easterlin (1995), and Weinberger (2004), all use questions that employ a Likert-scale.

[9] 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, since the ordinal ranking does not reveal the respondent's uncertainty about the outcome.

Wording for the question that elicited expected income was similar to that in Dominitz and Manski (1996):

> *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 at the age of 30 if one were to drop out of college.

The short introduction, practice questions, and questions eliciting beliefs about major-specific outcomes can be viewed in the Appendix. The 15 questions that elicit beliefs about major-specific outcomes were asked for *each* major category in the student's choice set. The full questionnaire (which also collected data on demographic information, formation of beliefs etc.) is available from the author on request.

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, I don't undertake this task here since it is not possible to summarize the data in a condensed form, and the main goal of the paper is to estimate a choice model of college majors. Interested readers are instead referred to Zafar (2009), available at the author's webpage, which analyzes the data in detail, shows that respondents provide meaningful answers to questions eliciting subjective expectations and that their responses match up well with objective measures. For example, comparison of students' expectations of income in different majors with objective measures (in this case, income of previous cohorts of college graduates) indicates that students are aware of the income differences across majors (this is discussed in more detail in section 5.1.1). The companion paper also addresses the various cognitive biases that may affect subjective data (Bertrand and Mullainathan, 2001), and does not find evidence of any strong biases affecting the data.

One notable feature of the subjective data is that, even with my relatively homogenous sample, there is substantial heterogeneity in responses both between and within gender. In order to highlight the heterogeneity in beliefs across respondents, I discuss the responses 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 in Literature and Fine Arts. The table shows that respondents are willing to use the entire scale from zero to 100. 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.[10] 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.

Table 4 also indicates that respondents answer seriously and meaningfully. About 60% of males think that the percent chance of graduating with a GPA of at least 3.5 in Engineering is greater than 50%. On the other hand, nearly 95% of them believe that they would be able to graduate with a GPA of at least 3.5 with a probability of more than 0.5 in Literature & Fine Arts. This is consistent with the fact that it's harder to do well in Engineering than in Literature & Fine Arts; average GPA of Northwestern Engineering graduates of 2006 was 3.43, while that of Literature & Fine Arts was 3.56. Females also exhibit substantive heterogeneity in beliefs, and seem to respond to questions in a consistent manner. Whereas only 30% of females believe that there's a greater than 50% chance of graduating with a GPA of at least 3.5 in Engineering, nearly 90% of females believe that to be the case in Literature & Fine Arts. Whereas the gender-specific belief distributions are similar for Literature & Fine Arts, that is not the case for Engineering: The male belief distribution of graduating with a GPA of at least 3.5 in Engineering first order stochastically dominates the corresponding female distribution, suggesting that females are less confident than men in their ability in Engineering. The different gender-specific belief distributions underscore the heterogeneity in beliefs between the two genders.

The substantial heterogeneity in beliefs (both within and between genders) questions the accuracy of restrictions imposed on expectations in the literature.

[10]Freshmen students were not surveyed for this study to avoid this phenomenon. This is what Bruine de Bruin et al. (2000) call "epistemic uncertainty," or the "50-50 chance."

4 Econometric Model

This section outlines the econometric framework. In this paper, I only focus on the model for single major choice.[11]

Recall that utility, $U_{ikt}(\mathbf{a}, \mathbf{c}, X_{it})$, is a function of a 5×1 vector of outcomes $\mathbf{a}$ realized in college, a 6×1 vector of outcomes $\mathbf{c}$ realized after graduating from college, and individual characteristics X_{it}. The individual maximizes her *current* subjective expected utility[12]; she chooses major m at time t if: $m = \arg\max_{k \in C_i} \int U_{ikt}(\mathbf{a}, \mathbf{c}, X_{it}) dP_{ikt}(\mathbf{a}, \mathbf{c})$. 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 7×1 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 4×1 vector of all continuous outcomes, i.e., $\mathbf{d} = \{a_4, c_4, c_5, c_6\}$.[13] The utility can now be written as a function of outcomes $\mathbf{b}$, $\mathbf{d}$, and characteristics X_{it}. Since it would be difficult to elicit the joint probability distribution $P_{ikt}(\mathbf{b}, \mathbf{d})$, I assume that utility is additively separable in the outcomes:

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

where $u_r(b_r, X_{it})$ is the utility associated with the binary outcome b_r for an individual with characteristics X_{it}, γ_{iqt} is a constant for the continuous outcome d_q, and ε_{ikt} is a random term. The utility is the same for all individuals with identical observable characteristics X_{it} up to the random term. Equation (1) can now be written as:

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

[11] 48% of the sample respondents claim to pursue more than one major. However, since each respondent submitted a preference ordering over the majors in their choice set, these respondents are included in the sample. To understand *why* individuals choose more than one major, Zafar (2010) estimates a separate model.

[12] 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 Malamud, 2006, and Stinebrickner and Stinebrickner, 2008). It is beyond the scope of this paper.

[13] The vectors $\mathbf{a}$ and $\mathbf{c}$ (as well as $\mathbf{b}$ and $\mathbf{d}$) are the set of outcomes common to all majors. It is the joint probability distribution of these outcomes $P_{ikt}(\mathbf{a}, \mathbf{c})$ (or $P_{ikt}(\mathbf{b}, \mathbf{d})$) which is indexed by major k. The vector $\mathbf{b}$ consists of b_1= graduating in 4 years, b_2= graduating with a GPA of at least 3.5, b_3= enjoying the coursework, b_4= parents approving of the major, b_5= getting a job on graduation, b_6= enjoying work at the jobs, and b_7= being able to reconcile work and family at the jobs. The vector $\mathbf{d}$ consists of d_1= average hours per week spent on coursework, d_2= average hours per week spent at the job, d_3= social status of the job, and d_4= expected income at the age of 30.

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

$$\Pr(m|X_{it}, \{P_{ikt}(b_r), P_{ikt}(d_q)\}_{r\in\{1,..,7\},q\in\{1,..,4\};\, k\in C_i}) =$$
$$\Pr\left(\begin{array}{c} \sum_{r=1}^{7}\int u_r(b_r, X_{it})dP_{imt}(b_r) + \sum_{q=1}^{4}\gamma_{iqt}\int d_q dP_{imt}(d_q) + \varepsilon_{imt} \\ > \sum_{r=1}^{7}\int u_r(b_r, X_{it})dP_{ikt}(b_r) + \sum_{q=1}^{4}\gamma_{iqt}\int d_q dP_{ikt}(d_q) + \varepsilon_{ikt} \end{array}\right) \tag{2}$$
$$\forall k \in C_i,\ m \neq k$$

For the binary outcomes in $\mathbf{b}$, $P_{imt}(b_r)$ is simply $P_{imt}(b_r = 1)$ for $r \in \{1,..,7\}$; $P_{imt}(b_r = 1)$ is elicited directly from the respondents for $\forall r \in \{1,..,7\}$ 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 $\forall q \in \{1,..,4\}$.[14]

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:

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

where w is an indicator variable of the individual's labor force status at the age of 30, $G_{it}(w = 1)$ is the subjective belief at time t about being active in the labor force at the age of 30, and p_{ikt} is individual i's subjective probability at time t about successfully graduating in major k. 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.[15] Both $E_{it}(I_0)$ and $G_{it}(w = 1)$

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

[15] 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 change qualitatively.

are elicited directly from the respondents.[16] Equation (2) can now be written as:

$$\Pr(m|X_{it},\{P_{ikt}(b_r),E_{ikt}(d_q)\}_{r\in\{1,..,7\},q\in\{1,..,4\};\,k\in C_i}) =$$
$$\Pr\left(\begin{array}{c}\sum_{r=1}^{7}\{P_{imt}(b_r=1)u_r(b_r=1,X_{it})+[1-P_{imt}(b_r=1)]u_r(b_r=0,X_{it})\}+\sum_{q=1}^{4}\gamma_{iqt}E_{imt}(d_q)+\varepsilon_{imt}\\ >\sum_{r=1}^{7}\{P_{ikt}(b_r=1)u_r(b_r=1,X_{it})+[1-P_{ikt}(b_r=1)]u_r(b_r=0,X_{it})\}+\sum_{q=1}^{4}\gamma_{iqt}E_{ikt}(d_q)+\varepsilon_{ikt}\end{array}\right) \tag{3}$$
$$\forall k\in C_i,\ m\neq k$$

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 $P_{imt}(b_r=1)\triangle u_r(X_{it})+u_r(b_r=0,X_{it})$, where $\triangle u_r(X_{it})\equiv u_r(b_r=1,X_{it})-u_r(b_r=0,X_{it})$, 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:

$$\begin{aligned}&U_{imt}(\mathbf{b},\mathbf{d},X_{it},\{P_{imt}(b_r=1)\}_{r=1}^{7},\{E_{imt}(d_q)\}_{q=1}^{4}) =\\ &\sum_{r=1}^{7}P_{imt}(b_r=1)\triangle u_r(X_{it})+\sum_{r=1}^{7}u_r(b_r=0,X_{it})+\sum_{q=1}^{4}\gamma_{iqt}E_{imt}(d_q)+\varepsilon_{imt}\end{aligned} \tag{4}$$

Equation (3) can now be written as:

$$\Pr(m|X_{it},\{P_{ikt}(b_r),E_{ikt}(d_q)\}_{r\in\{1,..,7\},q\in\{1,..,4\};\,k\in C_i}) =$$
$$\Pr\left(\begin{array}{c}\sum_{r=1}^{7}P_{imt}(b_r=1)\triangle u_r(X_{it})+\sum_{q=1}^{4}\gamma_{iqt}E_{imt}(d_q)+\varepsilon_{imt}\\ >\sum_{r=1}^{7}P_{ikt}(b_r=1)\triangle u_r(X_{it})+\sum_{q=1}^{4}\gamma_{iqt}E_{ikt}(d_q)+\varepsilon_{ikt}\end{array}\right) \tag{5}$$
$$\forall k\in C_i,\ m\neq k$$

$\{\triangle u_r(X_{it})\}_{r=1}^{7}$, and $\{\gamma_{iqt}\}_{q=1}^{4}$ are the parameters to be estimated; $\triangle u_r(X_{it})$ is the change in utility from the occurrence of outcome b_r for an individual with characteristics X_{it}, while γ_{iqt} is the parameter in the utility function for the continuous outcome d_q. $G_{it}(w=1)$, $E_{it}(I_0)$, $\{P_{ikt}(b_r=1)\}_{r=1}^{7}$, $\{E_{ikt}(d_q)\}_{q=1}^{3}$, 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

[16]Note that the underlying assumption is that the belief of being active in the labor force, $G_{it}(w=1)$, is independent of one's field of study. This is a rather restrictive assumption since one's decision to participate in the labor force may be influenced by the job opportunities available, which could 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 that would not have been feasible. I am also assuming that the expected earnings if one were to drop out of college, $E_{it}(I_0)$, are independent of one's intended field of study in college.

random terms required for identifying the model parameters are discussed in the next section.

5 Choice Model Estimation

This section deals with estimating the preferences for choice of college major. 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 coefficients 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, (5) becomes:

$$\begin{aligned}
&\Pr(m|P_{ik}(b_r), E_{ik}(d_q)\}_{r\in\{1,..,7\},q\in\{1,..,4\};\,k\in C_i}) = \\
&\Pr\begin{pmatrix} \sum_{r=1}^{7} P_{im}(b_r=1)\triangle u_c + \sum_{q=1}^{4}\gamma_q E_{im}(d_q) + \varepsilon_{imt} \\ > \sum_{r=1}^{7} P_{ik}(b_r=1)\triangle u_c + \sum_{q=1}^{4}\gamma_q E_{ik}(d_q) + \varepsilon_{ikt} \end{pmatrix} \\
&\qquad\qquad\qquad\qquad\qquad\qquad \forall k \in C_i,\ m \neq k
\end{aligned}$$

If I assume that 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:

$$\Pr(m|\{P_{ik}(b_r), E_{ik}(d_q)\}_{r\in\{1,..,7\},q\in\{1,..,4\};\,k\in C_i}) = \frac{\exp(\sum_{r=1}^{7} P_{im}(b_r=1)\triangle u_r + \sum_{q=1}^{4}\gamma_q E_{im}(d_q))}{\sum_{k\in C_i}\exp(\sum_{r=1}^{7} P_{ik}(b_r=1)\triangle u_r + \sum_{q=1}^{4}\gamma_q E_{ik}(d_q))} \tag{6}$$

The elicited subjective probabilities, $\{P_{ik}(b_r = 1)\}_{r=1}^7$, and elicited expected values, $\{E_{ik}(d_q)\}_{q=1}^4$, described in section 3.2 are used in estimation. The parameters of interest are $\{\triangle u_r\}_{r=1}^7$ and $\{\gamma_q\}_{q=1}^4$, and they are identified under these parametric assumptions.

However, in addition to stating their (intended) choice, respondents were also asked to rank the majors in their choice set. The exact question was: "*Put yourself in the hypothetical situation where you have not yet chosen a field of study to major in. Rank the following fields of study according to how likely you think* *you will major in that field of study*". The stated preference data provide more information that can be used for estimation of the model parameters.[17,18] 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 ε_{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 (Luce and Suppes, 1965):

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

Column (1) of Table 5 presents the maximum-likelihood estimates using stated preference data. The relative magnitudes of $\{\triangle u_r\}_{r=1}^{7}$ show the importance of the binary outcomes in the choice. The difference in utility levels is largest and positive for enjoying coursework. Enjoying work at the jobs and gaining approval of parents are significant determinants of major choice: Both coefficients are about one-half that of enjoying coursework. Status of the jobs is also a significant determinant in the choice: A unit increase in the social status of the jobs changes the utility by as much as a 3% 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 positive but not significantly different from zero, suggesting that it is not important in the choice.

To get a sense of gender differences in preferences, columns (2) and (3) of Table 5 present the maximum-likelihood estimates based on equation (6) for the male and female sub-samples, respectively.

[17] Kapteyn et al. (2007) use a similar approach to estimate preference parameters for retirement.

[18] 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 from the author upon request) 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.

Estimation results using choice data are not reported here and are available from the author upon request. They are quantitatively similar to the estimates obtained using preference data.

For both genders, the difference in utility levels is largest and positive for enjoying coursework. For males, approval of parents is the second most important outcome. The third most important outcome for males is the social status of the jobs: A unit increase in the status of the jobs changes the utility by as much as a 6% increase in the probability of enjoying coursework. For females, enjoying work at the jobs is the second most important outcome. Two other important outcomes for females are gaining approval of parents and graduating with a GPA of at least 3.5. Both have a positive coefficient that is about one-third the magnitude of the coefficient on enjoying coursework.

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 of earnings that an individual would be willing to forgo at the age of 30 in order to experience that outcome.[19] However, since expected income at age 30 is not significant, 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 $\Pr(choice = j) = F(\mathbf{X}_j\boldsymbol{\beta})$ 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:

$$\begin{aligned} M_{X_1} &\equiv \| \overline{\Pr(choice = j| \{\widehat{\beta_1}, \widehat{\beta_2}\}} - \overline{\Pr(choice = j| \{\widehat{\beta_1} = 0, \widehat{\beta_2}\}} \| \\ &= \sqrt{\sum_j \left[\sum_{i=1}^{N} \frac{\Pr(choice = j| \{\widehat{\beta_1}, \widehat{\beta_2}\})}{N} - \sum_{i=1}^{N} \frac{\Pr(choice = j| \{\widehat{\beta_1} = 0, \widehat{\beta_2}\})}{N} \right]^2}, \end{aligned} \tag{7}$$

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 in the two terms is a measure of the importance of X_1 in the choice. 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 using the estimates

[19] 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 $\frac{0.02 \times \triangle u_j}{\gamma_4}$.

from Table 5. Each cell shows the *relative* contribution (R) of the outcome to the choice. Column (1) in Panel A of Table 6 shows that, for the pooled sample, nearly three-fourths of the choice is driven by the non-pecuniary outcomes.[20] Once the decomposition is made finer in Panel B, one can see that gaining parents' approval and enjoying coursework jointly explain about 45% of the choice. Pecuniary outcomes associated with college (hours per week spent on coursework, graduating with a GPA of at least 3.5, and graduating in four years) and workplace (finding a job upon graduation, hours per 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 55% of the choices for males, but more than 85% of the choice for females. Gaining parents' approval and enjoying coursework are the most important outcomes for both females and males. Reconciling family and enjoying work at the available jobs are second in terms of importance to females, but of least importance to males. For males, pecuniary outcomes in the workplace are second in terms of importance. 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 for outcomes in the workplace: Males value pecuniary aspects of the workplace more (relative to the non-pecuniary outcomes in the workplace), while females value non-pecuniary aspects of the workplace more. Section A.1 of the online appendix shows that the estimated preference parameters compare favorably with respondents' stated reasons for choosing a major.[21]

The analysis in this section is based on the assumption that, conditional on gender, preferences are homogenous. This assumption is relaxed in the online appendix. Section A.2 of the appendix allows the preference parameters to depend on other demographic characteristics (for example, how much an individual depends on her parents for monetary support; whether any of the individual's parents are foreign-born). In particular, I find systematic differences in preferences of respondents conditional on whether their par-

[20]Outcomes classified as being non-pecuniary are gaining parents' approval, enjoying coursework, reconciling work and family, and enjoying work at the jobs. The remaining outcomes are termed as being pecuniary. Social status is classified as a pecuniary outcome since it is highly correlated with income.

[21]The appendix is available on the author's webpage at http://www.newyorkfed.org/research/economists/zafar/p1_appendix.pdf.

ents were born in the US or not. Individuals with a foreign-born parent value the pecuniary outcomes more in the choice of major. Since this dimension of culture is inherited by an individual from previous generations, rather than being voluntarily selected, I interpret this as a causal link from culture to preferences.

In section A.3, I relax the assumption that the random terms $\{\varepsilon_{ik}\}$ are independent for every individual i and choice k, and estimate a mixed logit model which allows for unobserved heterogeneity in preferences for some outcomes (as in Revelt and Train, 1998). Though the estimates show that there is substantial heterogeneity in the preferences for several outcomes, the relative magnitudes of the estimates are similar to those in this section. The online appendix also shows that parent's approval matters more for individuals who rely on their parents for college support, and that parents are perceived to approve of majors that offer a higher chance of finding a (high-status) job.

As additional robustness checks, the preference parameters in the model described in this section are also estimated by excluding respondents who are pursuing more than one major, as well as by excluding students who have a major outside WCAS and the School of Engineering (so that all respondents have the same choice set). The results remain qualitatively the same in both cases.

5.1.1 Understanding the Unimportance of Future Income in the Choice of Majors

This section discusses some robustness checks in order to determine whether income is actually unimportant 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 earnings differences across majors, and this is driving the result. Table 7 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 intends to major 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.[22] These statistics are presented in columns (1) and (2) of Table 7. 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.[23] Though the descriptive analysis of respondents' expectations of income in different majors in Table 7 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 (7) for 1,000 bootstrap samples for each of the sub-samples. The bootstrap confidence interval of R_{γ_4} (the relative importance of income) for both males and females does not include zero, which suggests that γ_4 is insignificant because of a large standard error and not because it is a precise zero.[24]

One concern might be that the sample contains few students choosing high-paying majors, and that is driving the result. That, however, is not the case: As shown in Table 7, students perceive Social Sciences II and Natural Sciences to be the highest paying majors, and Table 3 shows that nearly half of the sample intends to major in either of these categories.

Another reason for the insignificance of income in the choice could in part be due to the risk-neutrality assumption embedded in the model specification. This assumption was 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

[22] 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 age 32.

[23] It could be that the survey respondents are *self-enhancing* their own salary expectations. However, there are at least three legitimate reasons why respondents' earnings 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 that justifies having different expectations.

[24] As an additional robustness check, the model was also estimated 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_{γ_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 toward 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.

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 (Eckel and Grossman, 2008, and Croson and Gneezy, forthcoming), results in the current study regarding gender differences in income preferences could be a consequence of the risk-neutrality assumption.

6 Understanding Gender Differences

Section 5 shows that males and females differ in their preferences for the various outcomes. The descriptive analyses in section 3.4 documents the heterogeneity in beliefs for various outcomes between the two genders. Though the results of the decomposition metric of equation (7), presented in Tables 6, 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. In this section, I delve into the underlying causes for the gender gap in more detail.

6.1 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 my case, the two genders) in a linear framework is to express the difference in the average value of the dependent variable Y as:

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

where $\overline{X}_j$ is a vector of average values of the independent variables and $\widehat{\beta}_j$ 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 gender difference in mean levels of the outcome due to different observable characteristics (in the context of the model, the characteristics, X, correspond to the subjective beliefs), while the second term is the difference due to different effects of the characteristics, i.e., the $\widehat{\beta}$'s (this corresponds to the preference parameters estimated in section 5). 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 $Y = F(X\beta)$ and $F(.)$ is a non-linear function, $\overline{Y}$ does not equal $F(\overline{X}\beta)$. The gender difference in this non-linear case can be written as:

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

where N_j is the sample size of gender j. The first expression in the square brackets represents part of the gender gap that is due to gender differences in distributions of X (i.e., the beliefs), and the second expression represents the part due to differences in the group processes determining levels of Y (i.e., the preferences). It is relatively simple to estimate the total contribution. However, identifying the contribution of gender differences in specific variables (beliefs) and coefficients (preferences) to the gender gap is not straightforward. For this purpose, I use a decomposition method proposed by Fairlie (2005). Contributions of a single variable/coefficient are calculated by replacing the relevant variable of one gender with that of the other gender 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:

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

and that of X_2 is given as:

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

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.[25] Similarly, the independent contribution of β_1 to the gap is given by:

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

and that of β_2 is given as:

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

For the purposes of the decomposition, I use the parameter estimates shown in columns (2) and (3) of Table 5.[26] Results of this decomposition are presented in Table 8 for four different majors. 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. The majority of the gender gap in Literature & Fine Arts and in Social Sciences II is due to gender differences in beliefs, while gender differences in preferences explain majority of the gap in Engineering and in Social Sciences I.

If women being less overconfident than men (Niederle and Vesterlund, 2007, and references therein)

---

[25] Yun (2004) outlines an alternate decomposition strategy that 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.

[26] In the illustration above, I have assumed an 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. Then I 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.

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 8 show that gender differences in beliefs about ability (more precisely, beliefs about graduating in four years, and beliefs of graduating with a GPA of at least 3.5) are insignificant and explain a small part of the gender gap. Therefore, explanations based entirely 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 8. 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 and academic ability are insignificant and constitute less than 5% of the gap.[27] These findings suggest that females are less likely to major in engineering not because they are under-confident about their academic ability, low in self-confidence, or fear wage discrimination in the labor market. Instead, it is because they believe that they won't enjoy taking courses in Engineering.

6.2 Simulations

This section simulates different environments to see how the gender gap would change under different scenarios. Column (1) of Table 9 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 four years and about graduating with a GPA of at least 3.5) is

[27] I observe only the *beliefs* about academic ability, not *actual* academic ability. However, Chemers et al. (2001) show that confidence in one's ability is strongly related to academic performance. Moreover, it is the beliefs that matter when an individual is making a choice under uncertainty.

replaced with that of males.[28] The purpose of this simulation is to determine 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 in Table 8. If female expectations about ability were raised to the same level as those 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 beliefs about ability and future earnings in Engineering toward the underlying gender gap in the choice of major allows me to rule out low self-confidence 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 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 (Valian, 1998).[29] This issue is pursued in more detail in the following section.

[28] 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.

[29] 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 therefore 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 and Vesterlund (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.

6.3 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, in a follow-up survey (taken by 117 of the 161 original survey-takers), 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:

> "*What do you think is the percent chance that X (where X = {Male, Female}) would be treated poorly in jobs that are available in each of the following fields?*"

Columns (1) and (2) of Table 10 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 10. Several notable patterns stand out: First, 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.

There is a significant correlation of -0.35 between females' beliefs about being treated poorly at the jobs and the fraction of females in the major's classes. The correlation patterns between the variables described in Table 10 and beliefs about enjoying coursework and enjoying work at jobs indicate that, for female students in particular, beliefs about enjoying coursework and enjoying work at the jobs are related positively 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 (the correlations are available from the author upon request). 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.[30] Moreover, the new variables do not improve the model's explanatory power for the entire sample and for females; relative to the initial model, the Wald χ^2 (a measure of goodness-of-fit that compares the likelihood ratio chi-squared of the model to one with the null model) does not change by much.

The findings in this section suggest that females are less likely to major in fields like engineering not because they are under-confident about their academic ability, low in self-confidence, or fear wage discrimination. Instead, it is because they believe that they won't enjoy taking courses in Engineering. It seems that females enjoy fields which they believe treat them more favorably, and which happen to have a higher fraction of females. It is not clear how to interpret these correlations. Overall, the results seem to suggest that a policy that changes social attitudes might be more useful in narrowing the gap.

[30]Since beliefs of males and females being treated poorly at the jobs are strongly correlated with beliefs about enjoying coursework and enjoying work at the jobs, their impact on the choice is already being captured indirectly. Indeed, the variable "females treated poorly at the jobs" only shows up significantly for females and the entire sample (at the 1% and 10% level, respectively) in a model that excludes both enjoying coursework and enjoying work at the jobs. Model estimates are available upon request.

A question that I do not address is the source of gender differences in preferences, which could arise from differences in tastes, as well as gender discrimination. For example, parents who know that females would be discriminated against 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.[31]

7 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 about 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).

[31] As mentioned earlier, another concern that is not directly addressed in this paper is the extent to which beliefs are affected by preferences. In particular, cognitive dissonance may cause individuals to report beliefs that are consistent with their choices. One needs to see how beliefs evolve over time in order to study this. This issue is studied in detail in Zafar (2009). There are three ways that I check for this: (1) compare beliefs with objective measures in cases where it is possible (for example, future income), and check if individuals self-enhance beliefs for outcomes associated with their intended major; (2) since some individuals had officially declared their majors while others had an intended major at the time of first survey, systematic differences in estimates for the two sub-samples would indicate presence of cognitive dissonance; and (3) analyze how individuals revise their beliefs for outcomes associated with the different majors– this requires a panel of beliefs and provides the most convincing evidence on the presence of any biases. All three approaches reveal that cognitive issues do not affect the way in which individuals report their beliefs.

I find that outcomes most important in choice of major are enjoying coursework, gaining approval of parents, and enjoying work at the 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 outcomes at college, but differ in their tastes regarding the workplace. For outcomes in the workplace, females care more about non-pecuniary outcomes (enjoying work at jobs, and reconciling work and family), while males value pecuniary outcomes (social status of the jobs, likelihood of finding a job, and earnings profiles at jobs) more. Cultural proxies and demographic variables bias beliefs and preferences in systematic ways. 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. However, since the range of majors available to students and institutional details vary considerably across colleges, this is a challenging task because one cannot simply replicate the survey design employed in this study. 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. Third, individuals may find it optimal to experiment with different majors to learn about their ability and match quality (Altonji, 1993; Malamud, 2006; and Stinebrickner and Stinebrickner, 2008). Because of insufficient data, this study does not focus on this aspect, assuming instead 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 that were to raise the expectations of females about ability and future earnings in engineering to the same level as those of males would decrease the gender gap only by about 15%. This result has two implications: (1) just raising the expectations of women may not be enough to eradicate the gap, and (2) hypotheses that 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 different preferences for various outcomes– simply replacing females' beliefs about enjoying coursework with those of the males decreases the gender gap in engineering by almost half. Gender differences in beliefs about enjoying coursework as well as in preferences may exist because of differences in tastes or because of gender discrimination. Richer data are needed to answer this question. I believe the next natural step is to explore how individuals form beliefs.

Appendix

Survey Excerpt

Introduction and 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.

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? _________%

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

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.

Questionnaire

The following set of questions was asked for EACH of the relevant major categories. For example, the questions below were asked for the category of 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?

References

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

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

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

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

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

Barres, Ben. 2006. "Does Gender Matter?", *Nature*, 442: 133-136.

Bertrand, Marianne, and Sendhil Mullainathan. "Do People Mean What They Say? Implications for Subjective Survey Data", *American Economic Review Papers & Proceedings*, 91(2): 67-72.

Breen, Richard, and Cecilia Garcia-Penalosa. 2002. "Bayesian Learning and Gender Segregation." *Journal of Labor Economics*, 20(4): 899-922.

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

Bruine de Bruin, Wandi, Baruch Fischhoff, Bonnie Halpern-Felsher, and Susan Millstein. 2000. "Expressing Epistemic Uncertainty: It's a Fifty-Fifty Chance." *Organizational Behavior and Human Decision Processes*, 81: 115-131.

Cejka, Mary, and Alice Eagly. 1999. "Gender-Stereotypic Images of Occupations Correspond to the Sex Segregation of Employment." *Personality and Social Psychology Bulletin*, 25(4): 413-423.

Chemers, Martin, Li-tze Hu, and Ben Garcia. 2001. "Academic Self-Efficacy and First-Year College Student Performance and Adjustment." *Journal of Educational Psychology*, 93(1): 55-64.

Croson, Rachel, and Uri Gneezy. Forthcoming. "Gender Differences in Preferences." Journal of Economic Literature.

Daymont, Thomas, and Paul Andrisani. 1984. "Job Preferences, College Major, and the Gender Gap in Earnings." *Journal of Human Resources*, 19(3): 408-428.

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

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

Dominitz, Jeff, and Charles Manski. 1996. "Eliciting Student Expectations of the Returns to Schooling." *Journal of Human Resources*, 31(1): 1-26.

Easterlin, Richard. 1995. "Preferences and Prices in Choice of Career: The Switch to Business, 1972-87." *Journal of Economics Behavior and Organization*, 27: 1-34.

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

Euwals, Rob, Bertrand Melenberg, and Arthur van Soest. 1998. "Testing the Predictive Value of Subjective Labor Supply Data." *Journal of Applied Econometrics*, 13: 567-585.

Fairlie, Robert. 2005. "An Extension of the Blinder-Oaxaca Decomposition Technique to Logit and Probit Models." *Journal of Economic and Social Measurement*, 30(4): 305-316.

Fernandez, Raquel, Alessandra Fogli, and Claudia Olivetti. 2004. "Mothers and Sons: Preference Formation and Female Labor Force Dynamics." *Quarterly Journal of Economics*, 119(4): 249-299.

Festinger, Leon. 1957. *A Theory of Cognitive Dissonance*. Stanford: Stanford University Press.

Fiorito, Jack, and Robert Dauffenbach. 1982. "Market and Nonmarket Influences in Curriculum Choice by College Students." *Industrial and Labor Relations Review*, 36(1): 88-101.

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

Garman, David, and Linda Loury. 1995. "College Selectivity and Earnings." *Journal of Labor Economics*, 13: 289-308.

Goldin, Claudia, Lawrence Katz, and Ilyana Kuziemko. 2006. "The Homecoming of American College Women: The Reversal of the College Gender Gap." *Journal of Economic Perspectives*, 20(4): 133-156.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2006. "Does Culture Affect Economic Outcomes?" *The Journal of Economic Perspectives*, 20(2): 23-48.

Hurd, M., J. Smith, and J. Zissimopoulos (2004), "The Effects of Subjective Survival on Retirement and Social Security Claiming", Journal of Applied Econometrics, 19: 761-775

Juster, Thomas. 1964. *Anticipations and Purchases: An Analysis of Consumer Behavior*. Princeton: Princeton University Press.

Kapteyn, Arie, Arthur van Soest, and Julie Zissimopoulos. 2007. "Using Stated Preferences Data to Analyze Preferences for Full and Partial Retirement." IZA Discussion Paper 2785.

Kimura, Doreen.1999. *Sex and Cognition*. Cambridge: MIT Press.

Long, Vonda. 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.

Lubinski, David, and Camilla 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.

Luce, Duncan and Patrick Suppes. 1965. "Preferences, Utility and Subjective Probability." In *Handbook of Mathematical Psychology*, ed. Duncan Luce, Robert Bush, and Eugene Galanter, 249-410. New York: John Wiley and Sons.

Malamud, Ofer. 2006. “Breadth vs. Depth: The Effect of Academic Specialization on Labor Market Outcomes.” Harris School Working Paper Series 05:17.

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

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

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

Niederle, Muriel, and Lise Vesterlund. 2007. “Do Women Shy away from Competition? Do Men Compete too Much?” *Quarterly Journal of Economics*, 122(3): 1067-1101.

Oaxaca, Ronald. 1973. "Male-Female Wage Differential in the Urban Labor Markets." *International Economic Review*, 14(3): 693-709.

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

Revelt, David, and Kenneth Train. 1998. "Mixed Logit with Repeated Choices: Households’ Choices of Appliance Efficiency Level." *The Review of Economics and Statistics*, 80(4): 647-657.

Stinebrickner, Todd, and Ralph Stinebrickner. 2008. "Learning about Academic Ability and the College Drop-out Decision." University of Western Ontario Working Paper 20086.

Traweek, Sharon. 1988. *Beamtimes and Lifetimes: The World of High Energy Physicists*. Cambridge: Harvard University Press.

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

Valian, Virginia. 1998. *Why So Slow? The Advancement of Women*. Cambridge: MIT Press.

van der Klaauw, Wilbert. 2000. "On the Use of Expectations Data in Estimating Structural Dynamic Models: An Analysis of Career Choices", UNC-Chapel Hill Working Paper.

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

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

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

Xie, Yu, and Kimberlee Shauman. 2003. *Women in Science: Career Processes and Outcomes*. Cambridge: Harvard University Press.

Yun, Myeong-Su. 2004. "Decomposing Differences in the First Moment." *Economic Letters*, 82(2): 275-280.

Zafar, Basit. 2009. "Using Subjective Expectations Data in Choice Models." Mimeo, Federal Reserve Bank of New York.

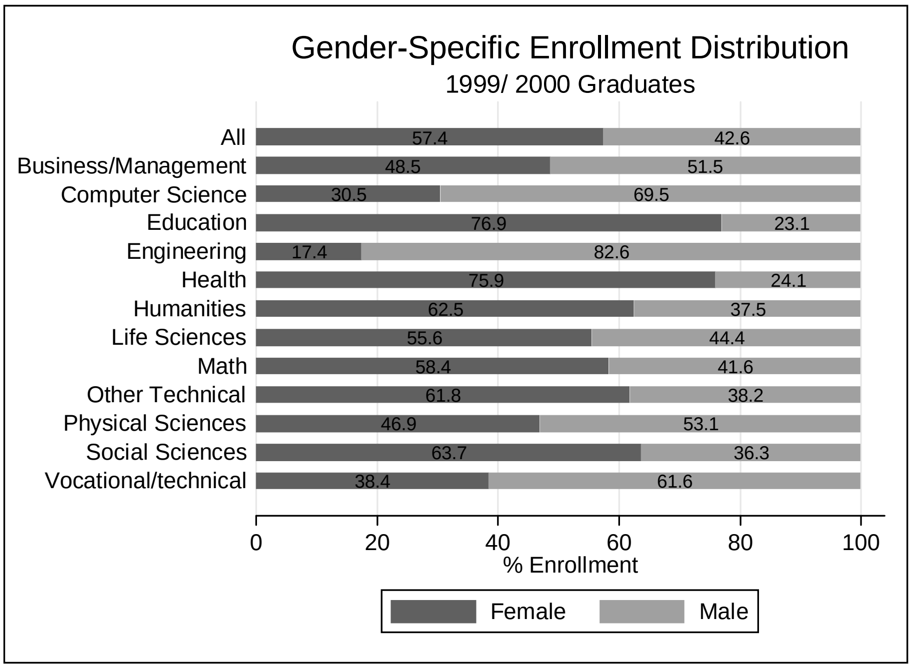

Figure 1: Gender Composition of Majors of 1999-2000 Bachelor's Degree Recipients Employed Full-Time in 2001.

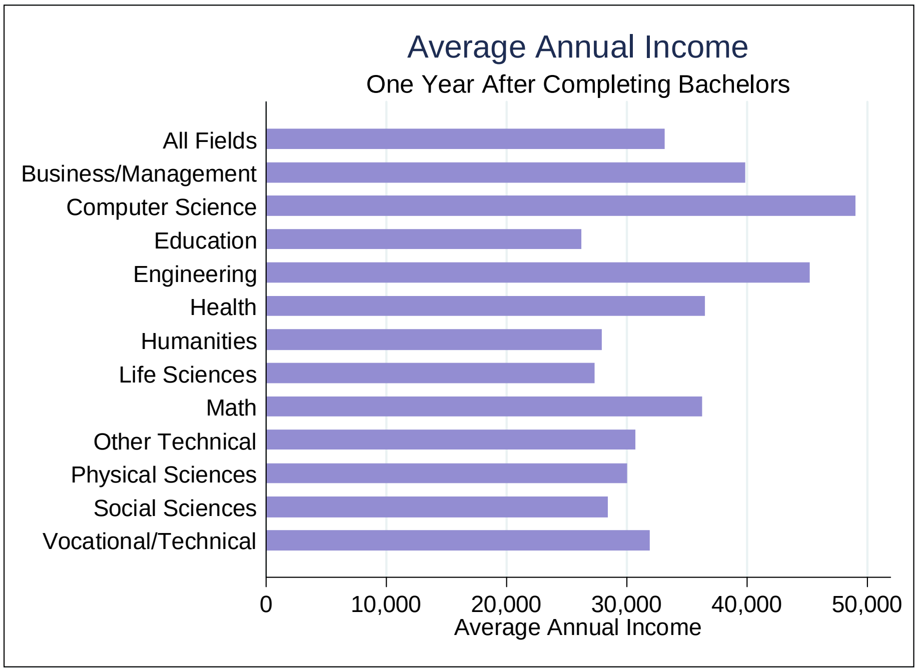

Figure 2: Average income of 1999-2000 Bachelor's Degree Recipients Employed Full-Time in 2001 by Major.

Table 1: List of Majors

The following is the classification of majors into categories:

a Natural Sciences
- Biological Sciences
- Chemistry
- Environmental Sciences
- Geography*
- Geological Sciences
- Integrated Science
- Materials Science
- Physics

b Mathematical and Computer Sciences
- Cognitive Science
- Computing and Information Systems
- Mathematics
- Statistics

c Social Sciences I
- Anthropology
- Gender Studies*
- History
- Linguistics
- Political Science
- Psychology
- Sociology

d Social Sciences II
- Economics
- Mathematical Methods in Social Sciences*

e Ethics and Values
- Legal Studies*
- Philosophy
- Religion
- Science in Human Culture*

f Area Studies
- African American Studies
- American Studies
- Asian & Middle East Languages & Civilization
- European Studies
- International Studies*
- Slavic Languages and Literatures

g Literature and Fine Arts
- Art History
- Art Theory and Practice
- Classics
- Comparative Literary Studies
- Drama
- English
- French
- German
- Italian
- Spanish

h Music Studies[1]
- Jazz Studies
- Music Cognition
- Music Composition
- Music Education
- Music Technology
- Music Theory
- Musicology
- Piano Performance
- String Performance
- Voice and Opera Performance
- Wind and Percussion Performance

i Education and Social Policy[2]
- Human Development and Psychological Services
- Learning and Organizational Change
- Secondary Teaching
- Social Policy

j Communication Studies[3]
- Communication Studies
- Dance
- Human Communication Science
- Interdepartmental Studies
- Performance Studies
- Radio/Television/ Film
- Theatre

k Engineering[4]
- Applied Mathematics
- Biomedical Engineering
- Chemical Engineering
- Civil Engineering
- Computer Engineering
- Computer Science
- Electrical Engineering
- Environmental Engineering
- Industrial Engineering
- Manufacturing and Design Engineering
- Materials Science& Engineering
- Mechanical Engineering

L Journalism[5]
- Journalism

∗ *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 All Freq.(Percent)		Population[a] Freq.(Percent)	
Gender				
Male	69	(43)	465	(46)
Female	92	(57)	546	(54)
Total	161		1011	
Ethnicity				
Caucasian	79	(49)	546	(54)
African American	11	(7)	71	(7)
Asian	56	(35)	232	(23)
Hispanic	5	(3)	61	(6)
Other	10	(6)	101	(10)
Declared Major?[b]				
Yes	90	(56)	477	(47)
No	71	(44)	534	(53)
Double Major?[c]				
Yes	78	(48)	–	
No	83	(52)	–	
International Std?[d]				
Yes	8	(5)	40	(4)
No	153	(95)	971	(96)
Second-Gen Imm?[e]				
Yes	66	(41)	–	
No	95	(59)	–	
Average GPA				
Male	3.48		3.26	
Female	3.40		3.31	

a Population Statistics for the sophomore class. (Source: Northwestern Office of the Registrar)
b Whether the respondent has declared their major at the time of the survey.
c Whether the respondent claims to pursue more than one major.
d Whether the respondent is an international student.
e 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[a]	Sample[b] All Freq	(%)	Sample[b] Males Freq	(%)	Sample[b] Females Freq	(%)	Class of 2006[c] All Freq	(%)	Class of 2006[c] Males Freq	(%)	Class of 2006[c] 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)
Total	161	(100)	69	(100)	92	(100)	1103	(100)	492	(100)	611	(100)

a Majors that appear in each category are listed in Table 1
b In cases where the survey respondent has more than one major in WCAS, only the first one is included
c Only 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:

	Engineering				**Lit. & Fine Arts**			
	Males		**Females**		**Males**		**Females**	
Sub. Beliefs	**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

	All	Males	Females
Δu_1 for graduating within 4 years	0.0826	-0.714	1.050
	(0.607)	(0.744)	(0.797)
Δu_2 for graduating with a GPA of at least 3.5	0.747^{**}	0.398	1.068^{**}
	(0.364)	(0.474)	(0.526)
Δu_3 for enjoying the coursework	3.081^{***}	2.682^{***}	3.375^{***}
	(0.333)	(0.576)	(0.409)
γ_1 for hours/week spent on coursework	0.0107	0.0164	0.0102
	(0.0089)	(0.0128)	(0.0116)
Δu_4 for approval of parents and family	1.304^{***}	1.450^{***}	1.187^{**}
	(0.360)	(0.560)	(0.491)
Δu_5 for finding a job upon graduation	0.103	0.0633	0.0793
	(0.338)	(0.508)	(0.469)
Δu_6 for enjoying work at the available jobs	1.358^{***}	0.669	1.856^{***}
	(0.288)	(0.474)	(0.405)
Δu_7 for reconciling family & work at jobs	0.295	0.640	0.0449
	(0.378)	(0.579)	(0.494)
γ_2 for hours/week spent at work	-0.0022	-0.0073	0.0084
	(0.0073)	(0.0122)	(0.0082)
γ_3 for the social status of the available jobs[a]	0.913^{***}	1.670^{***}	0.382
	(0.234)	(0.364)	(0.309)
γ_4 for expected Income at the age of 30	3.07e-07	1.09e-06	2.86e-07
	(7.05e-07)	(1.95e-06)	(5.63e-07)
Log-Likelihood	-1403.58	-615.03	-761.04
No. of individuals	161	69	92
No. of Choice Situations[b]	1287	551	736

* significant at 10%; ** significant at 5%; *** significant at 1%; robust standard errors in parentheses
a - 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 and hrs/week) are probabilities between 0 and 1
b -using the stated preference ranking, each student has n-1 choice situations for a choice set of size n

Table 6: Decomposition Analysis

	All	Males	Females
Panel A			
Attributed to:			
Pecuniary Attributes[a]	26.00%	46.40%	13.80%
Non-Pecuniary Attributes[b]	74.00%	53.60%	86.20%
Panel B			
Attributed to:			
Parents' Approval + Enjoying Coursework	45.50%	42.15%	45.80%
Coursework hrs/week + GPA + Graduating in 4 yrs	18.00%	8.80%	15.65%
Finding a job + Job hrs/week + Income at 30 + Status of Job	23.25%	40.40%	18.70%
Reconcile work & family + Enjoying Work	13.25%	8.65%	19.85%

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 7: Expected Annual Salary at the Age of 30

Variable: Expected salary (in 1000s) at the age of 30 in each major category as reported by:

	Exp Salary[a]		Respondent with Major in the category[b]						Respondent with Major not in the Category[c]					
Category:	**Males**	**Fem**	**Males**			**Females**			**Males**			**Females**		
	(1) Avg.	(2) Avg.	(3a) Avg.	(3b) Med.	(3c) N^d	(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

[a] 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

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

[c] Respondent's answer to * when his/her intended majors are in a category other than X.

d Number of respondents

Table 8: Decomposition Analysis to explain gender differences

	Eng.	Lit & Arts	Soc Sci II	Soc Sci I	Eng.	Lit & Arts	Soc Sci II	Soc Sci I
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(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.	**0.0601**	**-0.0843**	**0.0907**	**-0.0411**	**0.0601**	**-0.0843**	**0.0907**	**-0.0411**
	Contributions from gender diff in beliefs of:				**Contributions from gender diff in coeffs of:**			
Graduating in 4 years	-0.00053	0.0015	-0.0042***	0.0022	0.0090***	-0.004***	0.0025**	-0.011***
	-0.89%[a]	-1.76%	-4.63%	-5.34%	15.07%	4.89%	2.74%	27.12%
Graduate with GPA≥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%
All included variables	0.0251	-0.0523	0.0711	-0.0082	0.0350	-0.0320	0.0196	-0.0328
	41.75%	**62.01%**	**78.36%**	**20.00%**	**58.25%**	**37.99%**	**21.64%**	**80.00%**

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

a: Contribution in % of the relevant variable to the gap. A positive (negative) number means the contribution is in the (opposite) direction of the gender diff in prob. Here, -0.89% implies that replacing female beliefs about graduating in 4 years in Engineering by those of males will decrease the gap of 0.0601 by -0.89% (or increase it by 0.89%).

Table 9: Simulations of the Change in Gender Gap under different Environments

Fields of Study	Base[c]	Ability	Income	Enjoying Coursework	Enjoying Work
	(1)	(2)	(3)	(4)	(5)
Engineering	0.0602^a	0.0517 $13.92\%^b$	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%

a The model predicted gender gap (male prob. - female prob.) under the relevant environment
b The % decrease in the gender gap (relative to the baseline case) after the change
c The predicted gap under the baseline case, i.e. no intervention

Table 10: Perceptions of Monetary and Non-Monetary Discrimination

	% Females in class[a]			Males Poorly[b]		Fems Poorly	
Category	Males[c]	Fems[d]	Recs[e]	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
[a] Fraction of students in the major who are females (on a scale of 0-100)
[b] The average belief that males would be treated poorly in the jobs that would be available in each of the specified categories
[c(d)] Response of male (female) survey respondents to the relevant question
[e] Fraction of females amongst graduates with that major in 2005 & 2006 (source: IPEDS)